TUTORIAL EXCEL

DATA ANALYSIS 1

Abstract

Tutorial sobre conceptos de análisis de datos, uso de gráficas, PivotTables y PivotCharts

Iván Santiago
isantiago@ogp.pr.gov
versión: 20 julio, 2018

TABLE OF CONTENTS

Sobre este tutorial 9
PRE-REQUISITOS: 9
Resumen del tutorial 9
Módulo 1: Gráficas y Filtros 10
Resumen 10
Objetivos 10
Gráficas 10
Objetivos 10
Crear gráficas 10
Estilos y colores 12
Estilos 12
Colores 14
Guardar como plantilla (template) 15
Aplicar una plantilla 15
Formateo 16
Añadir una línea de tendencia (trend line) 16
Escoger gráfica apropiada 17
Qué tipo de gráfica elegir 17
Algunas recomendaciones: 20
Pie chart: Mal aplicado y el menos apropiado… 20
Barra o columna: El más recomendado para representar magnitudes 21
Líneas 21
Mapa proporcional (*treemap*) 21
Demo: Convertir datos en gráficas 22
Formatear la gráfica 24
Modificar el título 24
Mover la gráfica a otra hoja 24
Filtros 25
Objetivos 25
Filtrar datos 25
Quitar filtros 26

Esconder columnas 27
Unhide (revelar columna) 27
Formateo de datos 28
Demo: Seleccionar datos relevantes 28
Filtrar datos 28
Revelar una columna escondida 30
Esconder columna 31
Formatear columna 31
Quitar decimales 32
Sparklines (mini-gráficas) 33
Objetivos 33
Sparklines 33
Qué es una mini-gráfica sparkline 33
Añadir sparklines 33
Remover sparklines 34
Uso de sparklines 35
Recomendaciones para sparklines: 35
Formateo de sparklines 35
Menú y herramientas Sparklines 35
Tipos de sparklines 36
Show 36
Style 36
Group 37
Axis: 37
Demo: Añadir sparklines 38
Cambiar formato a Sparkline 40
Prácticas: 41
Escenario 41
Objetivos 41
Ejercicio 1: Gráficas y exploración de datos 41
Hacer histograma 41
Ejercicio 2: Formatear datos 42

Ejercicio 3: Filtrar datos 44
Objetivos: 44
Modificar posición de columnas para hacer diagrama de dispersión 44
Cortar y pegar columna 44
Explorar si hay relación entre nivel de pobreza y desempeño alto en matemáticas 45
Añadir línea de tendencia 46
Cambiar el color de la línea de tendencia 47
OPCIONAL: Mover la gráfica a otra hoja 48
Añadir otro diagrama de dispersión: Matrícula vs Competencia avanzada en matemáticas 48
OPCIONAL Aplicar filtros 49
Módulo 2: Excel Tables y Funciones 52
Resumen 52
Objetivos 52
Tablas de Excel (Excel Tables) 52
Objetivos 52
Beneficios de las tablas 52
Qué es una tabla 52
Formateo 52
Utilidad de las tablas 53
Crear tablas 53
Formateo de la tabla 54
Ordenar y filtrar 54
Tipos de filtros 54
Ordenar (Sorting) 54
Demo: Uso de una tabla 56
Algunos conceptos antes de discutir sobre funciones 59
Objetivos 59
Named ranges: Nombrar grupos de celdas y columnas 59
Cómo hacer un named range 59
Borrar o redefinir named ranges 60
Referencias relativas y absolutas 60
Referencias relativas 60

Referencias absolutas: 60
Ejemplos de referencias absolutas: 60
Ejemplo de referencia a una celda (fila y columna fija) 60
Códigos de errores en Excel y qué significan 61
Funciones 61
AGGREGATE 61
Sintaxis 61
Opciones de AGGREGATE(function_num, options, array,[k]) 62
Funciones que usan el parámetro k 62
LOOKUP 64
Sintaxis 64
VLOOKUP 65
Sintaxis: 65
INDEX & MATCH 66
Index 66
Sintaxis 66
Match 67
Sintaxis 67
Notas 68
Uso de INDEX con MATCH 68
COUNT, COUNTA, COUNTIF, COUNTIFS 71
SUMIFS 72
DATE FUNCTIONS: Funciones temporales, fechas y horas 73
Los datos temporales se registran como números 73
Dos sistemas de fechas 73
DATE Function 74
Sintaxis 74
DATEDIF 75
Sintaxis 75
DAYS 76
Sintaxis 76
NETWORKDAYS 77

Sintaxis 77
Resumir datos 78
Objetivos 78
Propósito de resumir datos 78
Total Row 78
Funciones incluidas en comando SUBTOTAL: 79
Sintaxis 79
Demo: Añadir fila de totales 80
Prácticas: 82
Ejercicio 1: Crear tablas 82
Ejercicio 2: Trabajar con tablas 84
Ejercicio 3: Análisis de datos 86
Ejercicio 4: Uso de funciones SUMIF y COUNTIF 87
Ejercicio 5: Funciones VLOOKUP y LOOKUP 88
VLOOKUP 88
Tareas: 88
Escribir y ejecutar función VLOOKUP 89
OPCIONAL: LOOKUP 90
Tareas: 90
Insertar columna 90
Escribir la función LOOKUP 91
OPCIONAL: Sparklines para visualizar fluctuación 92
Módulo 3: PivotTables y PivotCharts 93
Resumen 93
Dimensiones y métricas en PivotTables 93
Nota de precaución: Celdas vacías en un data range o una tabla 93
Utilidad 94
Interfaz PivotTable Fields 95
Disposición de los datos para que funcionen las tablas Pivot 95
PivotTable Tools Group tabs 96
Herramientas del tab Analyze: 96
Herramientas del tab *Design* 97

Grupos: 97
Cómo funciona una tabla Pivot 98
Formatear números 99
Estilos para tablas Pivot 99
Table Layouts: Organización de la tabla 100
Forma tabular 101
Formateo condicional (Conditional Formatting) 102
Ordenar y filtrar 103
Agrupar datos 103
Slicers y Timelines 104
Cálculo de valores y de columnas 105
Summarize values by… 105
Show values as 106
Percent of Grand Total: 106
Percent of Parent Row Total 107
Otros ejemplos de Show values as 108
Show values as… Index 109
Campos calculados 110
Cálculos basados en conteos 111
Consolidar columnas (Vectorización) 112
Trasponer usando Unpivot Columns 113
Proceso 1: Importar datos: 113
Proceso 2: Usar Query Editor 114
Transform | Unpivot Columns 115
PivotCharts 116
Cómo añadir una gráfica Pivot 117
Opciones disponibles para formatear gráficas Pivot 118
Field Buttons 118
Slicers 119
Timelines 120
Prácticas: Crear PivotTables y PivotCharts 121

Ejercicio 1: Crear PivotTable 121
Mostrar valores como porcentajes (Column Total) 123
Ver participación de cada distrito dentro de la región 124
Ranks (posición) 124
Opcional: Contestar: 125
Ejercicio 2: Vectorización o consolidación (unpivot) de múltiples columnas 126
Vectorizar (unpivot) múltiples columnas 127
Cambiar nombres de las columnas 127
Aceptar cambios y cerrar la interfaz Query Editor 127
Generar PivotTable 128
Generar reporte presupuestos por agencia 129
Cambiar formato a Currency 130
Remover totales para las filas 130
Usar Conditional Formatting para resaltar valores extremos 131
Mostrar valores como porcentaje del total 132
Otras opciones Agrupar por Jerarquía: 133
Percent of Parent Row Total 134
Percent of Parent Total 135
Index 135
Difference from y % Difference from 136
Esconder valores de error en PivotTable 137
Running Total In y % Running Total In 138
Otros desgloses 139
Ejercicio 3: Crear PivotChart 139
Aplicar filtro Top 10 140
Cambiar agencias a columnas y años a Rows 141
Generar gráfica Pivot 142
Aplicar filtro en la gráfica Pivot (Top 10, Opción Percent) 143
Ejercicio 4: Campos calculados en PivotTables 145
Añadir PivotTable 145
Añadir campo calculado a la tabla Pivot 146
Cambiar formato numérico 147
Ordenar datos de forma descendente 148

Preparar PivotChart 148
Quitar botones en la gráfica Pivot 149
Fuentes 150
Funciones útiles, Excel 150
Programas Add-In Utilidades 151
Datos temporales 151
DATEDIF function 151
Visualización, gráficas 152
Recomendaciones: 152
Pivot tables 152
Otros temas: 152
Quiebras 152

SOBRE ESTE TUTORIAL

Este tutorial está diseñado para ofrecerse en persona por dos días. Proveerá instrucciones y recomendaciones sobre análisis de datos usando Excel. No es un curso básico de Excel y requiere cierta familiaridad con este programa. Los datos para prácticas serán provistos. Habrá poca o ninguna entrada manual de datos.

PRE-REQUISITOS:

- Conocimiento general del sistema operativo Windows
- Conocimiento básico de manejo de hojas de cálculo Excel

RESUMEN DEL TUTORIAL

Módulo 1: Gráficas y filtros

Módulo 2: **Tablas Excel** y algunas **funciones** para Análisis de datos

Módulo 3: **PivotTables y PivotCharts**, describe cómo producir y trabajar con PivotTables y PivotCharts

MÓDULO 1: GRÁFICAS Y FILTROS

RESUMEN

Este módulo trata sobre **cómo convertir datos en gráficas**. Excel es muy utilizado, en parte por la facilidad de generar gráficas a partir de datos. En este módulo se considerarán algunas recomendaciones generales para generar gráficas más efectivas.

OBJETIVOS

- **Crear gráficas** a partir de datos
- **Filtrar datos** y formatear gráficas
- Crear **Sparklines**
- Escoger **el tipo de gráfica más apropiada** para los datos

GRÁFICAS

Las gráficas se utilizan para resumir y entender datos extensos. Una gráfica puede resumir múltiples líneas y columnas y hacer visibles, tendencias, proporciones y patrones imperceptibles en tablas.

OBJETIVOS

Al completar esta lección, podrá ser capaz de:

- Crear gráficas usando datos provistos
- Entender los diferentes tipos de gráficas disponibles en Excel
- Seleccionar el tipo de gráfica más apropiado para los datos que quiere representar

CREAR GRÁFICAS

- Gráficas
    - Son representaciones visuales de los datos
    - Ayudan a visualizar patrones en los datos
    - Buenas para ver datos a un nivel más alto
    - No son para dar detalles de los datos
- Para crear una gráfica:
    - Sombrear (escoger) datos
    - Ir al **menú principal**, escoger **Insert** y seleccionar el tipo de gráfica
    - La gráfica se insertará en el worksheet

Una gráfica es una representación visual de los datos. Una gráfica usa líneas, barras, segmentos y otros elementos para mostrar relaciones entre los valores de los datos. Normalmente las gráficas tienen un eje horizontal X y otro vertical Y.

Ver datos gráficamente

A menudo se requiere ver los datos en diferentes **niveles de detalle**.

Nivel de detalle alto: para identificar tendencias, anomalías, diferencias entre clases y jerarquías, diferencias a través del tiempo.

Presentar datos usando gráficas ayuda a entender y ver ciertos patrones y anomalías rápidamente.

Luego de ver patrones, tendencias y anomalías, se puede ir directamente a los datos para ver ciertos detalles.

Sin embargo, las gráficas no son buenas para ver valores precisos en los datos (para esto están las tablas de valores, si son pocos). A menudo, se requerirá incluir gráficas y tablas pequeñas en los reportes. Una manera efectiva de comunicar datos complejos es presentarlos desde un enfoque de arriba hacia abajo, (top-down) usando gráficas para el nivel más alto y tablas con algunos detalles necesarios.

Cómo generar una gráfica en Excel

1. Sombrear los datos que se usarán para hacer la gráfica. Puede usar las teclas **CTRL+A** siempre y cuando esté dentro del *data range*:

	B	C
1	Municipio	POP2010
2	Culebra	1,818
3	Maricao	6,276
4	Vieques	9,301
5	Las Marías	9,881
6	Maunabo	12,225
7	Florida	12,680
8	Ceiba	13,631
9	Rincón	15,200

Qué es un *Data Range*:

Es una extensión o grupo de celdas contiguas con valores. Cualquier espacio de una o más celdas vacías (NODATA) interrumpe o separa un Data Range de otro.

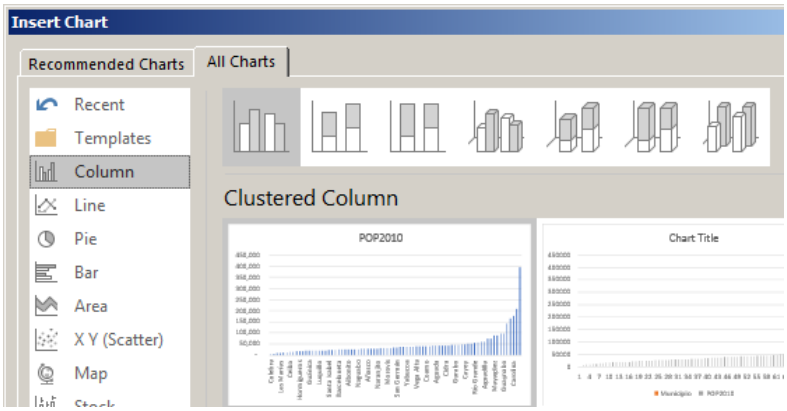
Esta es una lista de municipios con la población para el Censo de 2010.

2. Vaya al **menú principal** (ribbon) y haga **click** en el tab **Insert**

3. Haga **click** en el botón **Recommended Charts**

Dependiendo de los datos, Excel le puede devolver un listado de gráficas.

4. Puede escoger también la gráfica entre la lista recomendada o alguna otra que aparecerá en el lado izquierdo de la forma **Insert Chart**



5. Hacer **click** en el botón **OK** insertará la gráfica en la hoja de cálculo activa.

ESTILOS Y COLORES

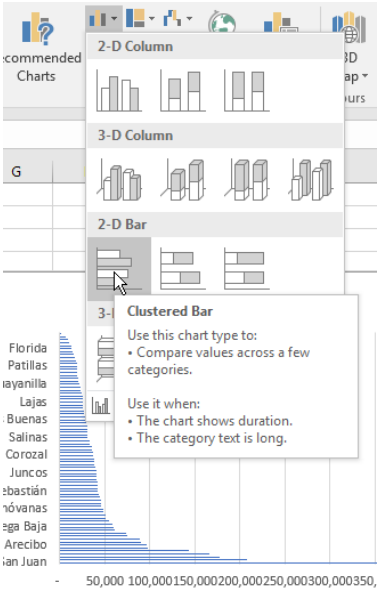
- Luego de haber decidido el tipo de gráfica, podrá:
  - Seleccionar el estilo de la gráfica
  - Combinación de colores o monocromático
  - Guardar como plantillas (templates) aquellos estilos que utilice más frecuentemente
- Como regla general, use los mismos patrones de color y estilos en las gráficas de todo el reporte.

ESTILOS

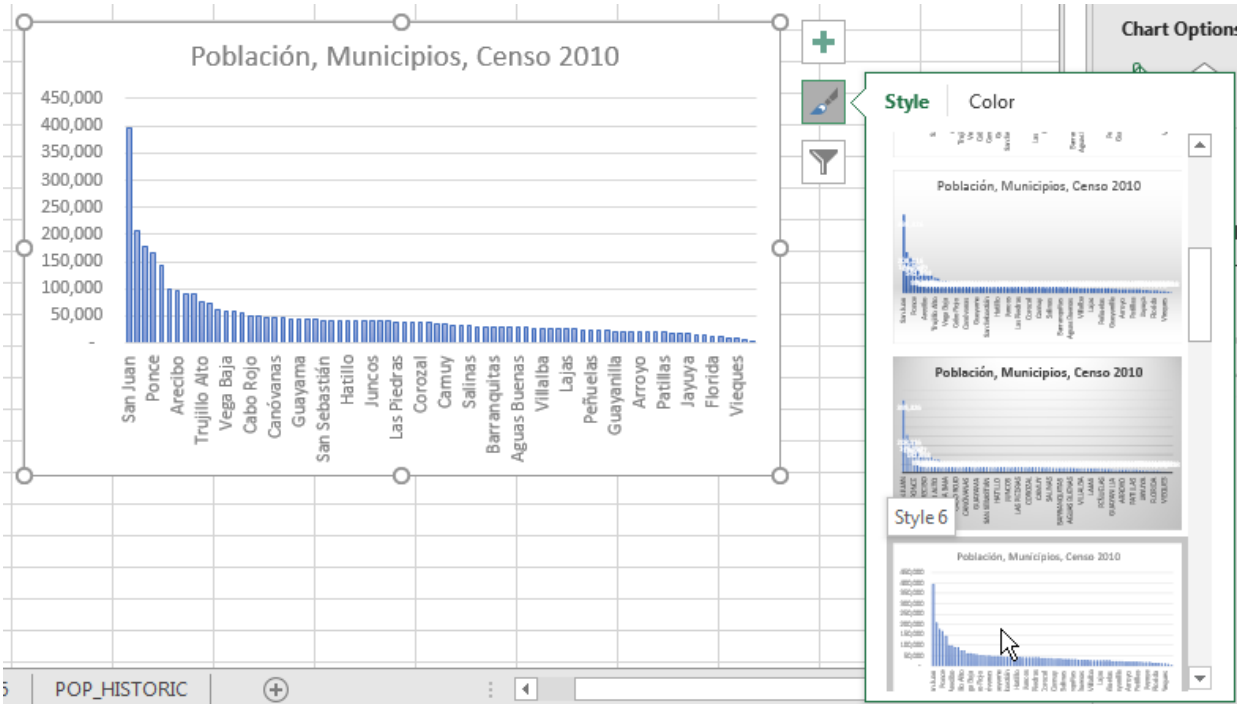
Cada tipo de gráfica ya sea pie chart, línea, columna, etcétera, tiene una variedad de estilos asociados. Hay tres maneras de ver los estilos asociados con una gráfica en particular:

1. Sombrear los datos, hacer **click** en el tab **Insert**, luego **click** en la flecha del **Charts Group** para desplegar la pantalla **Insert Chart**. Haga **click** en el tab **All Charts**, **click** en el tipo de gráfica que aparecerá en el panel izquierdo para poder ver los diferentes estilos al lado derecho de esta forma. Haga **click** en el estilo de gráfica para una visualización rápida de la gráfica resultante.
2. Sombrear los datos, hacer **click** en el tab **Insert**, luego hacer **click** en la **flecha hacia abajo**, al lado de los tipos de gráficas que aparecen en el grupo **Charts**. Podrá ver

una lista de figuras de cada estilo de gráfica

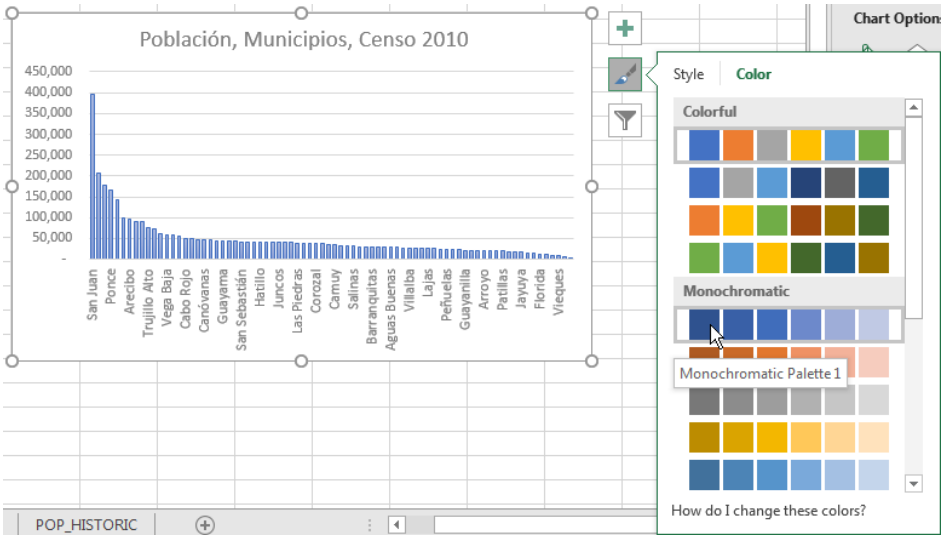


3. Escoger una gráfica existente, hacer **click** en el botón **Chart Styles** . Aparecerá una lista de opciones para cambiar el estilo de la gráfica.

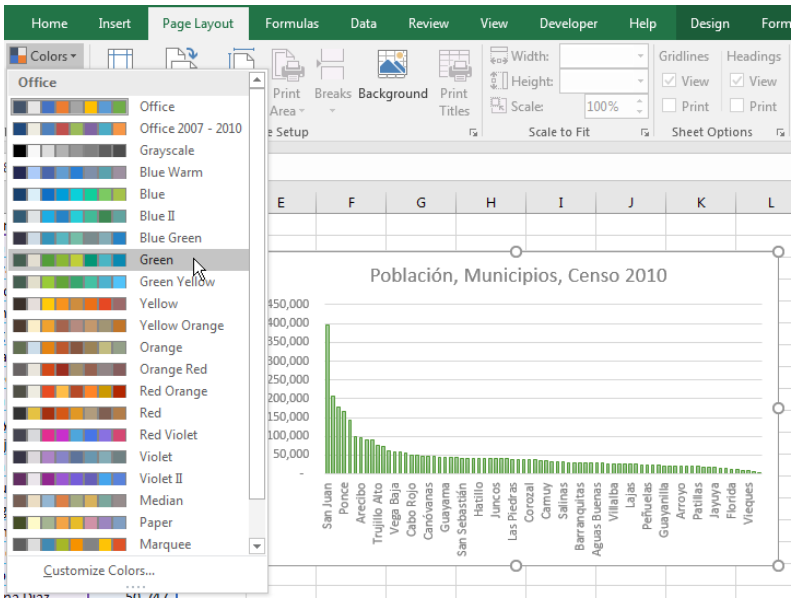


COLORES

Para seleccionar un patrón de colores para su gráfica, **selecciónela**, haga **click** en el icono **Chart Styles** y luego **click** en el tab **Color**. Se desplegarán varias opciones de patrones de colores, además de patrones monocromáticos.



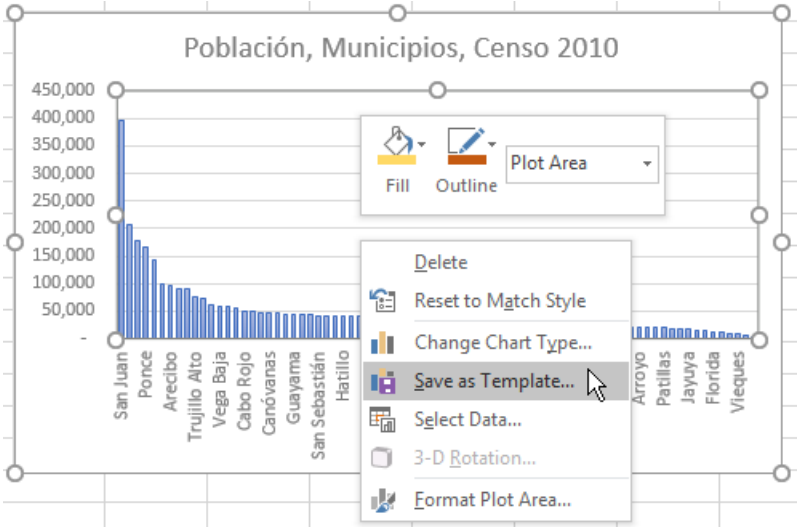
Nota: Los **patrones de colores disponibles**, **corresponden** a los patrones de color a **nivel de Workbook** Excel. Para acceder a otros patrones de colores, deberá hacer **click** en el tab **Page Layout**, hacer **click** en el botón **Colors** y escoger el patrón de colores que desee.



Guardar como plantilla (template)

Si desea estandarizar un estilo particular de estilo, patrón de colores de una gráfica, puede guardarla como una plantilla. Para crear una plantilla:

- Genere una gráfica con el estilo y colores que ud haya escogido.
- Haga **right click** en la gráfica y escoja **Save as a Template**

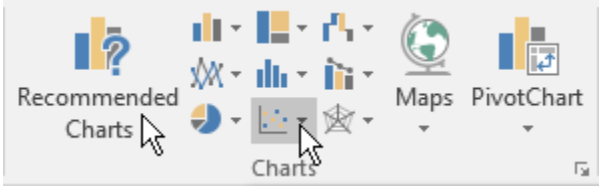


- Escriba el nombre del archivo de esta plantilla.
- Escriba el nombre del lugar a guardarla o acepte el lugar que aparece por defecto en Windows. Por ejemplo, se guarda a nivel de usuario en AppData\Roaming\Microsoft\Templates\Charts

Aplicar una plantilla

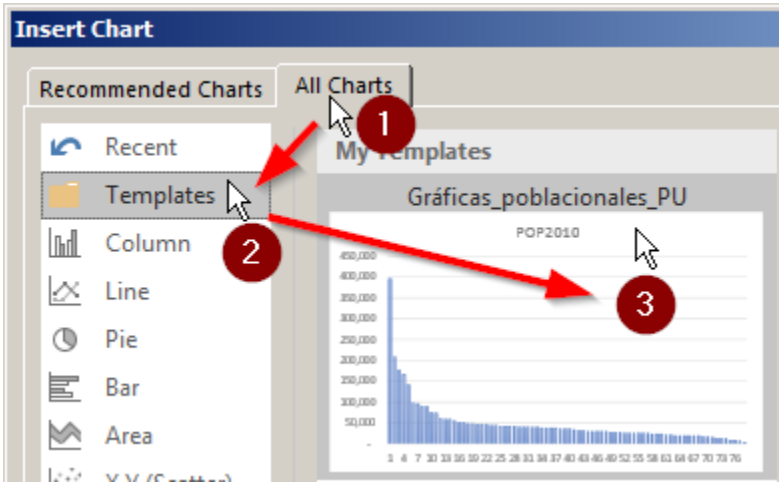
Luego de haber creado la plantilla, estará disponible para usar mediante la caja de diálogo **Insert/Change Chart Type**. Para usar una plantilla existente:

- **Seleccione** los datos (CRTL+A) **haciendo *click <u>dentro</u> del data range***
- Haga **click** en el tab **Insert**
- En el grupo de botones **Chart Group**, haga **click**, ya sea en **la flecha derecha** de uno de los botones del tipo de gráfica **o click en** el botón **Recommended Charts**



- Haga **click** en el tab **All Charts**.

- En el panel izquierdo de la forma **Insert Chart**, seleccione el folder **Templates**.



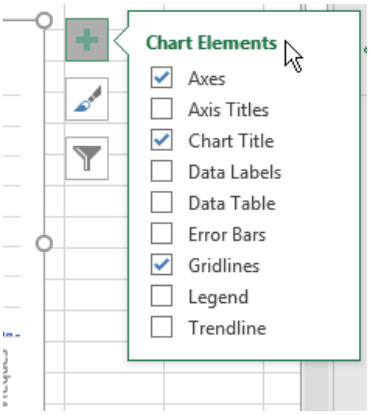
- Haga **click** en la plantilla que usará y **click** en el botón **OK**.

  Consejo: Utilice los mismos estilos de gráfica y colores en un reporte.
  Esto le da coherencia al reporte y se verá más estético y profesional.

FORMATEO

Luego de haber creado la gráfica, use la opción **Chart Elements** para cambiar sus características:

- Título, ejes, etiquetas (labels), tabla de datos, barras de error, líneas guía (grid), leyenda, línea de tendencia.
- Para añadir elementos a la gráfica, **seleccione** (**click**) la gráfica y presione el **botón**



AÑADIR UNA LÍNEA DE TENDENCIA (TREND LINE)

Una línea de tendencia utiliza los datos registrados en diferentes épocas (meses, años, días) para predecir valores en el futuro. Podemos añadirle el valor R^2 (Coeficiente de determinación) para dar una idea de cuán confiable es la línea estimada. Mientras más se acerque R^2 a 1 o -1, mayor será la certeza de la estimación.

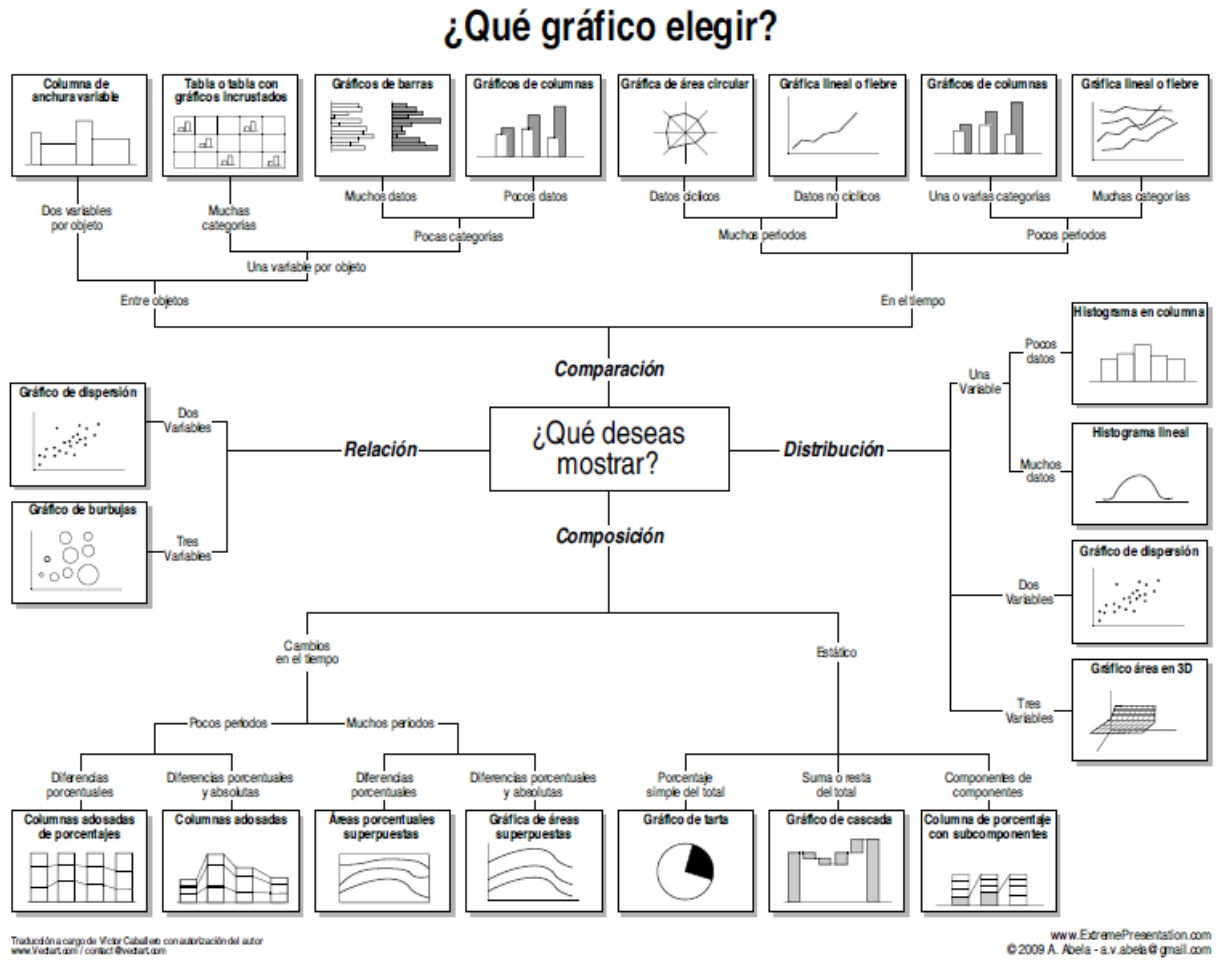
ESCOGER GRÁFICA APROPIADA

Excel tiene una gran variedad de tipos de gráficas. Producir una gráfica es fácil en Excel; sin embargo, escoger el tipo correcto de gráfica requiere cierta consideración según el tipo de conjunto de datos y la audiencia. Excel incluye los siguientes tipos de gráfica

Líneas, columnas, barras, pie charts (pastel), áreas, diagrama de dispersión, stock (gráfica de líneas para acciones en el mercado), superficie, radar, treemap, sunburst, histograma, box and whisker, cascada.

QUÉ TIPO DE GRÁFICA ELEGIR

Este diagrama muestra cómo escoger el gráfico más apropiado, según los datos y el propósito de la gráfica.



Tomado de: http://extremepresentation.typepad.com/files/que-grafico-elegir.pdf

versión en inglés: http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf

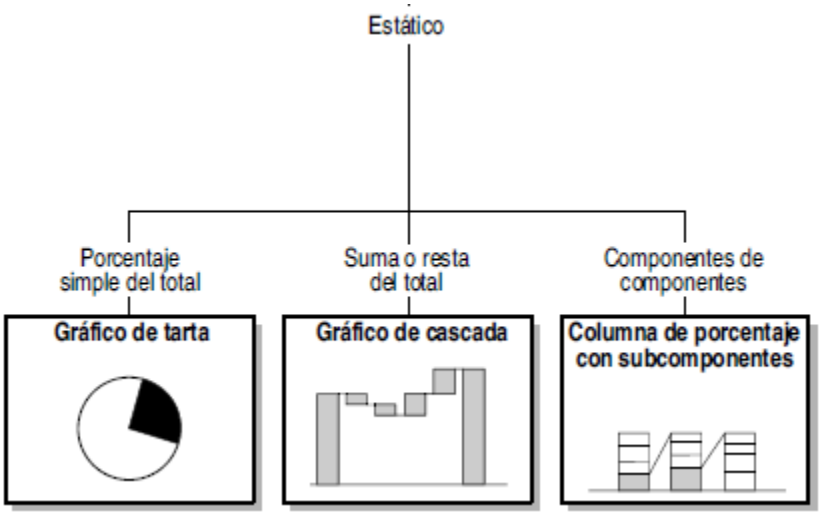
La pregunta central es **¿QUÉ DESEAS MOSTRAR?**

¿Comparación, Distribución, Composición, Relación?

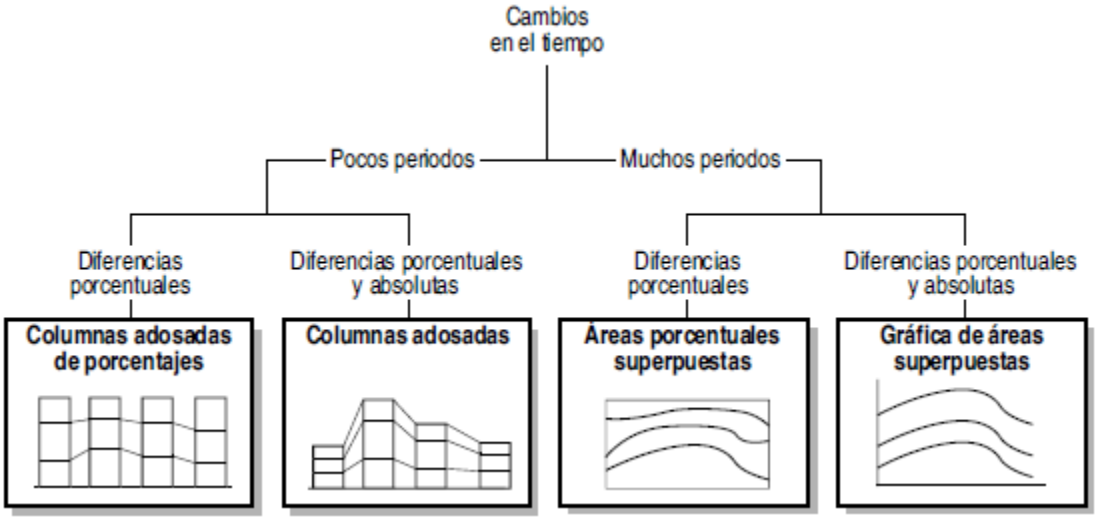
Para cada una de estas acciones, habrá diferentes opciones para crear una o más gráficas para representar los datos.

Descriptivas: **Composición y Comparación**: Para visualizar o explorar datos mediante resumen. No supone necesariamente un análisis previo de los datos.

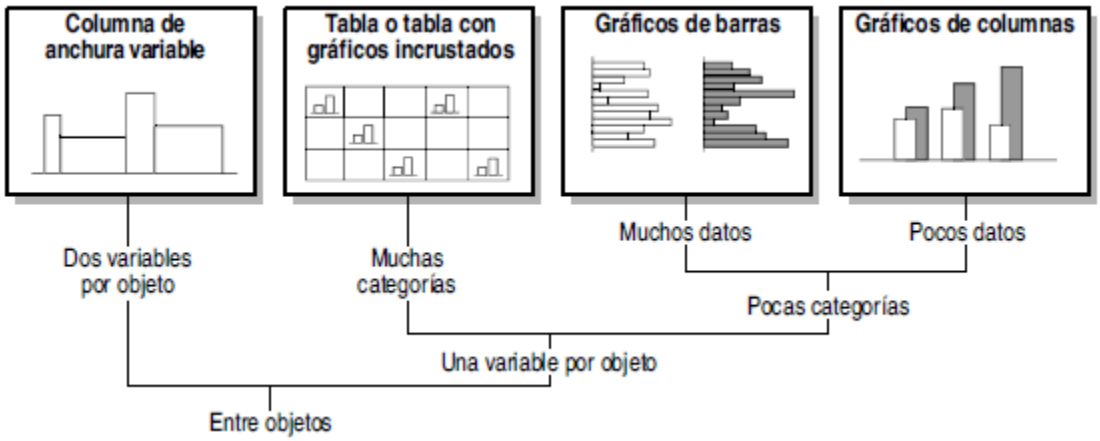
Composición: ***Estático***



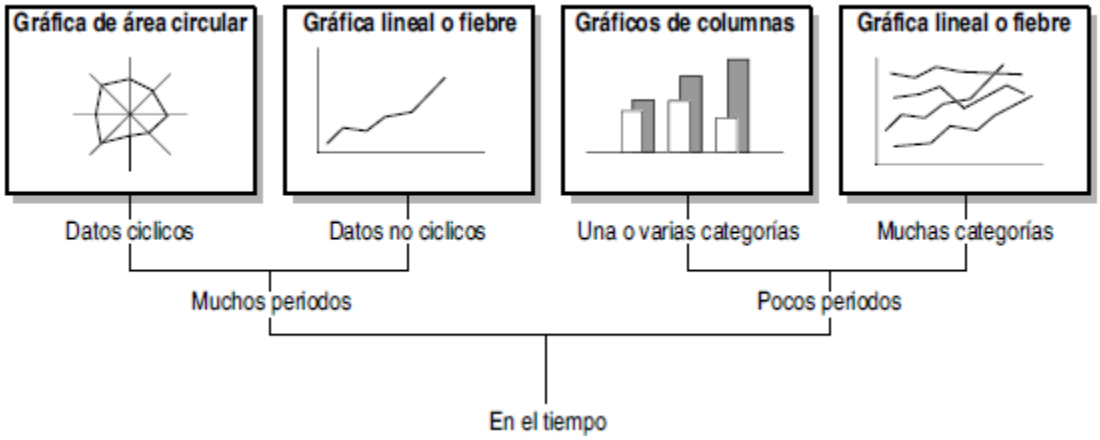
Composición: ***Cambios en el tiempo***



Comparación: entre objetos



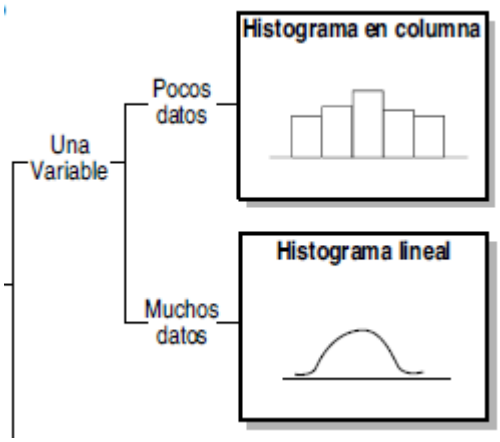
Comparación: en el tiempo



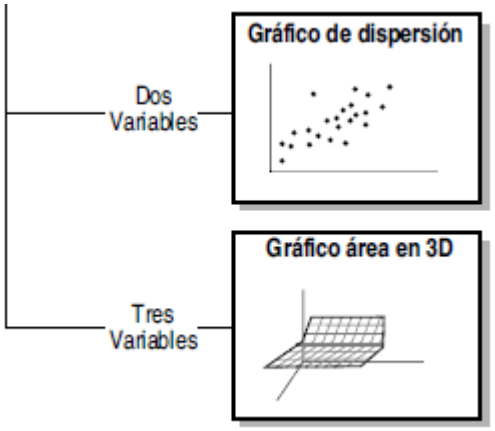
Analíticas: **Distribución y Relación**: Suponen que ha ocurrido exploración y análisis previo de los datos, mostrar patrones ocultos, reducir información.

Distribución:

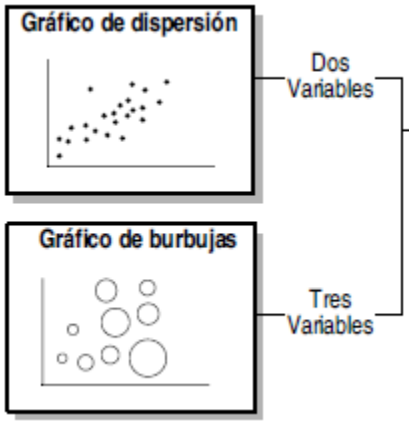
De una variable:



De dos o más variables:



Relación: entre una o más variables:



ALGUNAS RECOMENDACIONES:

Pie chart: Mal aplicado y el menos apropiado…

Este gráfico solo se recomienda para comparar entre pocas categorías (3, 4, o 5) y además las categorías no deben parecerse mucho. La vista humana no es buena distinguiendo porciones de áreas circulares (*wedges*). Si va a usar un pie chart, las recomendaciones son:

- No usar más de 5 categorías
- Asegúrese que los valores sumen 100%
- Ordene las porciones del pie chart por tamaño
- Las proporciones deben mostrar diferencias sustanciales

Barra o columna: El más recomendado para representar magnitudes

Al contrario de las gráficas pie chart, las gráficas de barras o columnas son más apropiadas para mostrar las magnitudes y proporciones. Para pocas categorías, es mejor usar un diagrama de columnas; para muchas categorías, es mejor usar un gráfico de barras.

Líneas

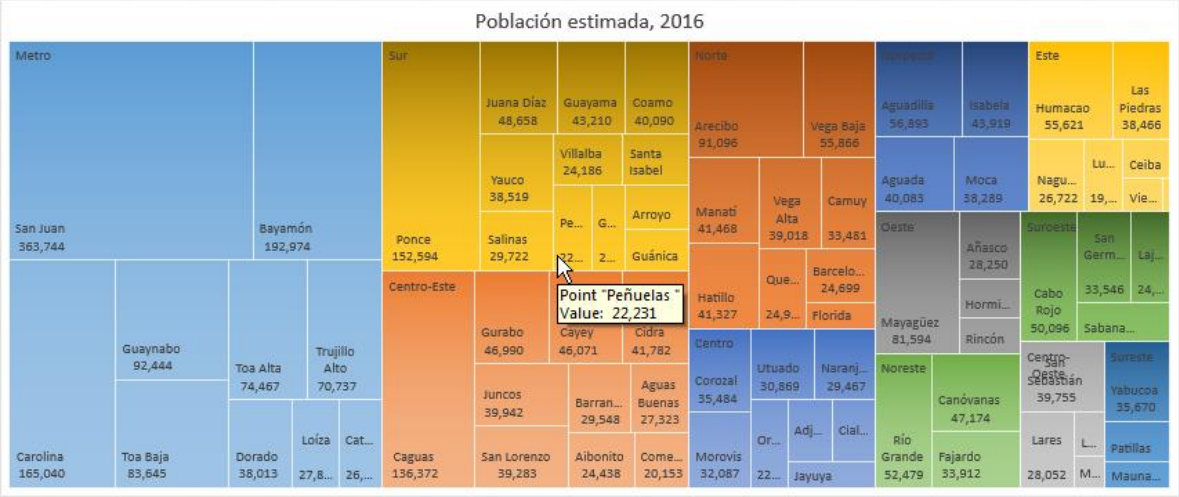
Las gráficas de líneas muestran tendencias por periodos de tiempo. También pueden mostrar varias categorías en una sola gráfica. Se recomiendan para distribuciones de datos continuos.

Se recomienda:

- Prefiera usar líneas sólidas
- Evite usar más de cuatro categorías
- Altura de línea adecuada: debe tomar más o menos 2/3 partes de la altura del eje Y.

Mapa proporcional (*treemap*)

Este tipo de gráfica ha sido utilizado generalmente para representar datos proporcionales por jerarquías (Grupos y subgrupos). Se utiliza por ejemplo para representar datos presupuestarios.



En este ejemplo, cada dato de población municipal estimada en 2016 está clasificado por zonas: Metro, Sur, Norte, etc. Los colores representan las zonas.

DEMO: CONVERTIR DATOS EN GRÁFICAS

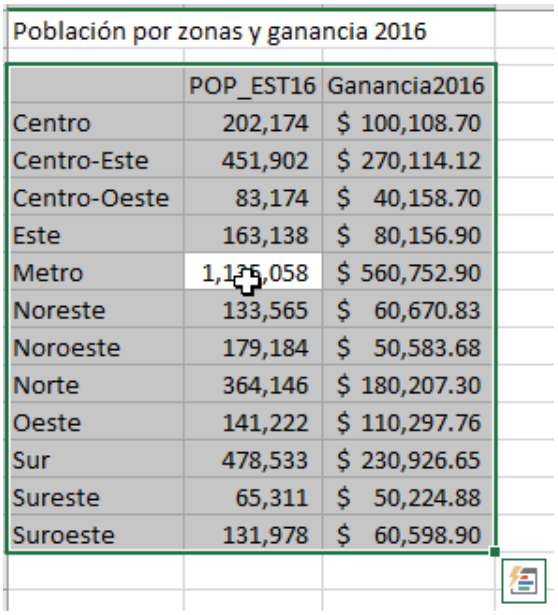
Abra Excel y busque el archivo **municipios_2016.xlsx** en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_1**

Este data range tiene datos ficticios.

Población por zonas y ganancia 2016		
	POP_EST16	Ganancia2016
Centro	202,174	$ 100,108.70
Centro-Este	451,902	$ 270,114.12
Centro-Oeste	83,174	$ 40,158.70
Este	163,138	$ 80,156.90
Metro	1,135,058	$ 560,752.90
Noreste	133,565	$ 60,670.83
Noroeste	179,184	$ 50,583.68
Norte	364,146	$ 180,207.30
Oeste	141,222	$ 110,297.76
Sur	478,533	$ 230,926.65
Sureste	65,311	$ 50,224.88
Suroeste	131,978	$ 60,598.90

Haga **click en cualquier celda *dentro*** del ***data range*** y presione **CTRL+A**.

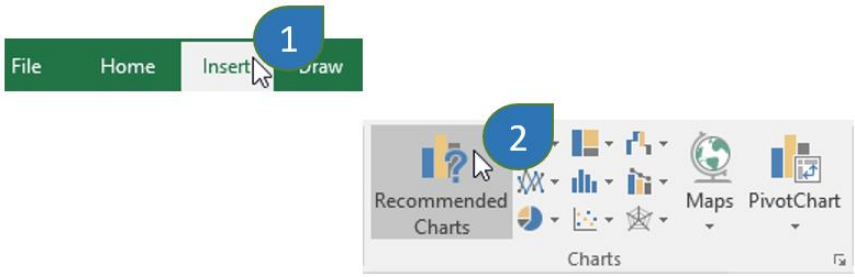
Esto sombreará el data range.



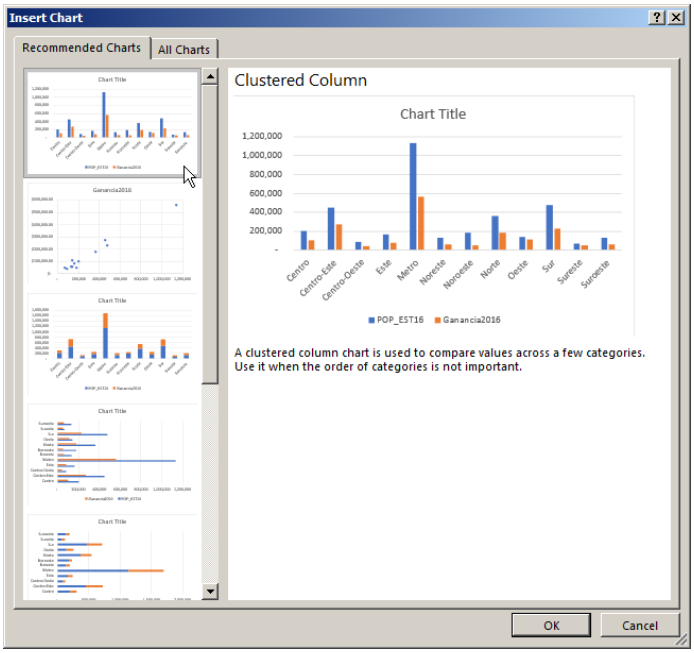
Puede usar esta combinación de teclas para evitar tener que sombrear manualmente un data range.

Recuerde que al usar **CTRL+A**, **seleccionará todas las *celdas contiguas*** que tengan algún valor. Si hay alguna celda vacía, Excel no la incluirá.

Una vez **sombreado** el data range, haga **click** en el **tab Insert**; luego en el botón **Recommended Charts**



Aparecerá la forma **Insert Chart**.



Haga **click** en las diferentes gráficas recomendadas para compararlas…

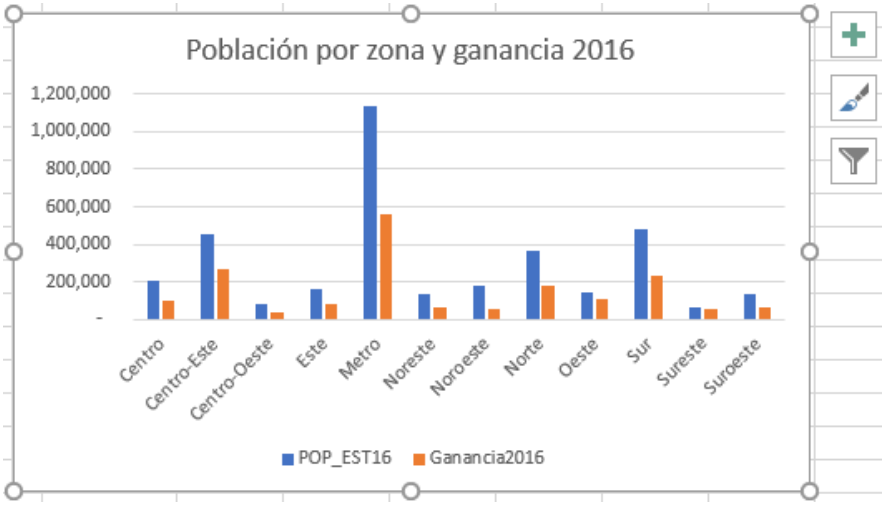
Finalmente haga **click** en la opción **Clustered column** y haga **click** en el botón **OK**.

Formatear la gráfica

Modificar el título

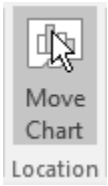
En la gráfica, haga **doble click en** el **título**.

Escriba **Población por zonas y ganancia 2016**

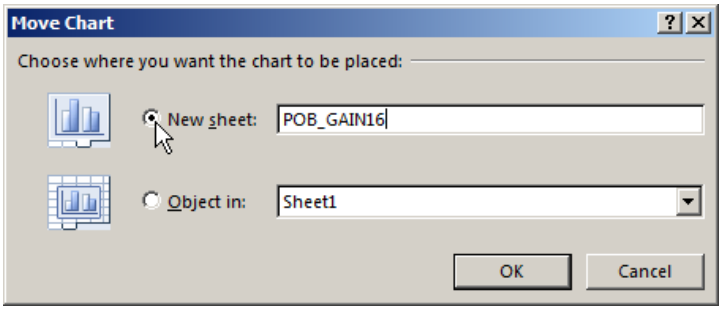


Mover la gráfica a otra hoja

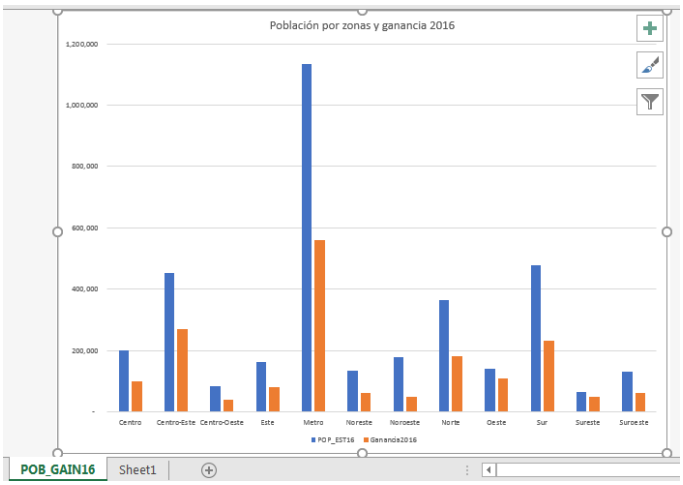
Haga **otro click** en la gráfica, pero fuera del título. La gráfica deberá estar activada. Cuando está activa, se habilita el tab **Design**. Haga **click** en el botón **Move Chart**.



Aparecerá la forma **Move Chart**. Escoja la opción **New sheet** para mover la gráfica a otra hoja de cálculo dentro de este workbook. En la caja de texto, escriba **POB_GAIN16**.



Presione el botón **OK** para aceptar las opciones y mover la gráfica a otra hoja.



Note el tab abajo (**POB_GAIN16**) correspondiente al nombre de la gráfica.

Cierre Excel. No tiene que guardar el workbook.

FILTROS

OBJETIVOS

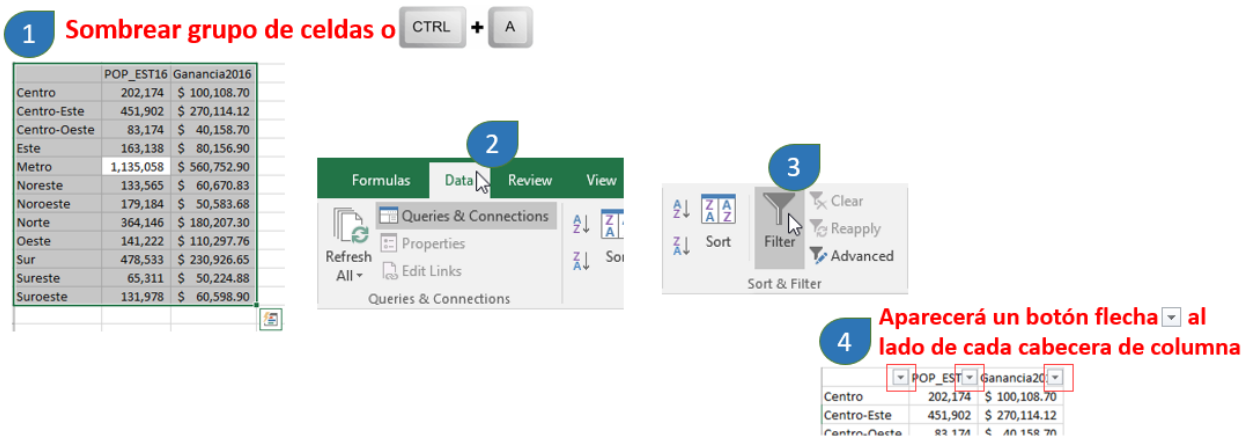
Al terminar esta lección, podrá:

- Filtrar datos para ver solamente las filas que deberían ser incluidas en una gráfica
- Formatear correctamente los datos

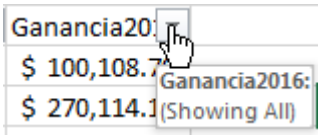
FILTRAR DATOS

- Los filtros permiten trabajar con un subconjunto de los datos
  - No cambian los datos subyacentes
  - Reducen la posibilidad de cometer errores de CUT & PASTE
- Esconder (Hide) columnas
  - Crear gráficas sólo con datos relevantes

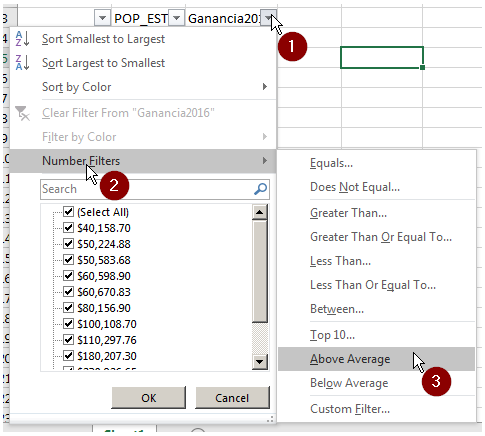
Para **añadir un filtro** a los datos:



Haga **click** en la flecha al lado derecho de la columna **Ganancia2016**

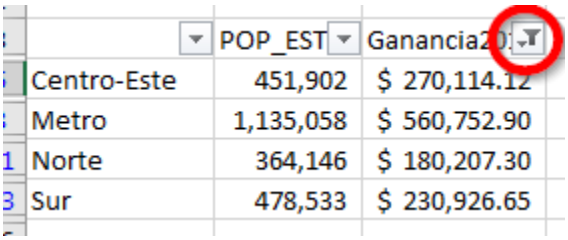


Vea las opciones para establecer filtros, en este caso, numéricos



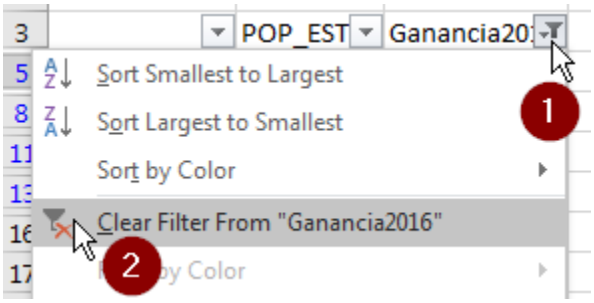
Escoja la opción **Above Average**

Note cómo cambia la lista. Estos son los récords que están por encima del valor promedio de ganancias (ficticias). Vea además que el botón de la flecha añade el símbolo del filtro. Esto indica que hay un filtro aplicado.



Quitar filtros

Para quitar filtros, haga **click** en el **botón** de **filtro**. Haga **click** en la opción **Clear Filter From**

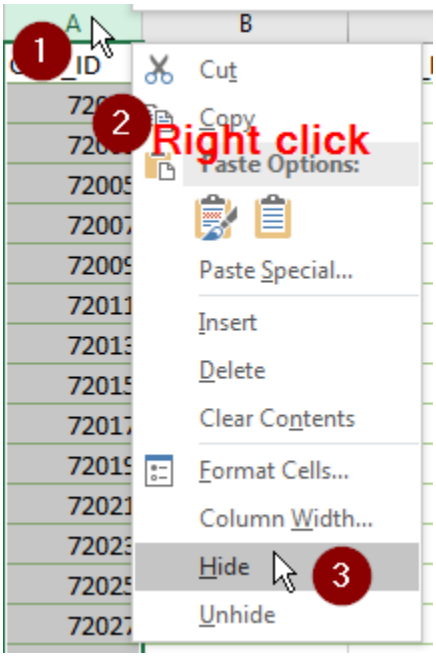


Esconder columnas

Esconder columnas permite hacer gráficas sin necesidad de borrar ni hacer *cut & paste* de los datos. Esto ayuda a reducir errores. Al esconder columnas, los datos permanecen intactos y las columnas podrán ser mostradas posteriormente, de ser necesario.

Para esconder columnas:

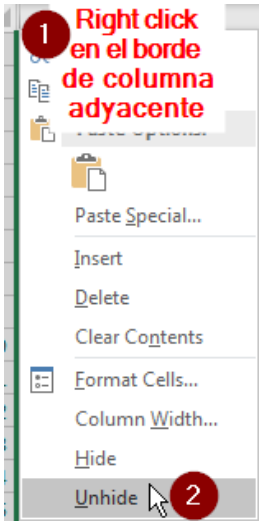
- Seleccione/**sombree** la(s) columna(s)
- Haga **right click** y escoja la opción **Hide**



Unhide (revelar columna)

Para revelar (**unhide**) columna(s), haga **right click en el borde de la cabecera de la columna adyacente**. Verá que el cursor cambiará a dos flechas:

→



Haga **right click** y escoja del menú la opción **Unhide**

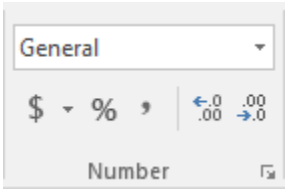
También puede seleccionar varias columnas adyacentes a la columna escondida y hacer **right click | Unhide**

Formateo de datos

Un formato correcto hace que los números y las fechas puedan entenderse mejor.

Para formatear celdas:

- Seleccione la(s) celda(s) a formatear
- Vaya al **menú principal** y escoja **Home** y vaya al **grupo Number**. Seleccione el formato numérico más apropiado.



Demo: Seleccionar datos relevantes

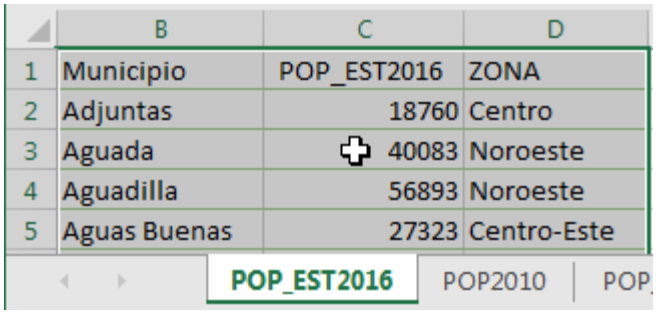
En este demo, ejecutaremos varias técnicas explicadas anteriormente: Filtrar datos, cambiar formato y generar una gráfica.

Abra una sesión de Excel si no está abierto.

Abra el archivo **municipios_poblacion.xslx**, localizado en el folder **C:\Tutorial_Excel\Analisis_1\Modulo_1**

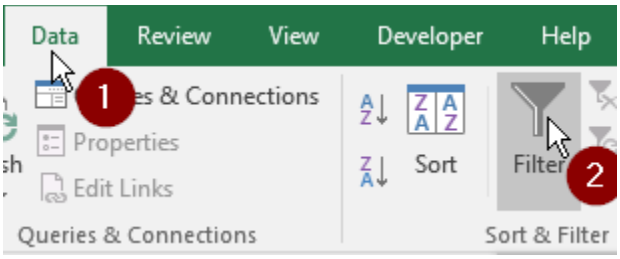
Filtrar datos

Seleccione el contenido del *data range* que aparece en la hoja llamada **POP_EST16**

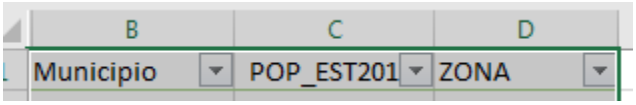


click en ***cualquiera*** de las **celdas *dentro del data range*** y usar **CTRL+A**

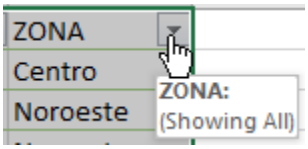
Haga **click** en el tab **Data** y luego en el botón **Filter**



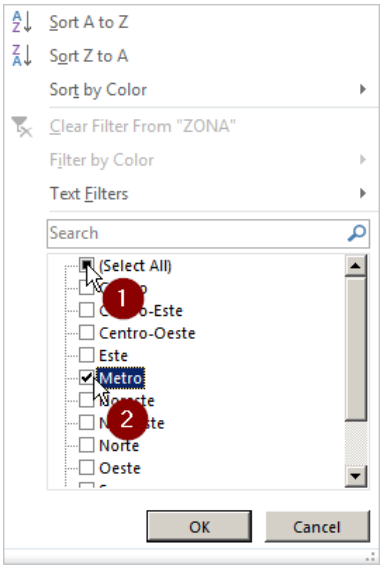
Aparecerán botones con flechas en cada cabecera de columna



Haga **click** en el botón de filtro al lado derecho de la columna **ZONA**



Inspeccione el menú de filtros.



Haga **uncheck** en la opción (**Select All**)

Haga **check** en la opción **Metro**

Esto le mostrará solamente los municipios que están dentro de la zona metropolitana de San Juan.

Datos filtrados

	B	C	D
1	Municipio	POP_EST201	ZONA
12	Bayamón	192974	Metro
17	Carolina	165040	Metro
18	Cataño	26137	Metro
27	Dorado	38013	Metro
33	Guaynabo	92444	Metro
46	Loíza	27857	Metro
66	San Juan	363744	Metro
70	Toa Alta	74467	Metro
71	Toa Baja	83645	Metro
72	Trujillo Alto	70737	Metro

Los números de las filas aparecen en azul, indicando que se les ha aplicado un filtro a los datos.

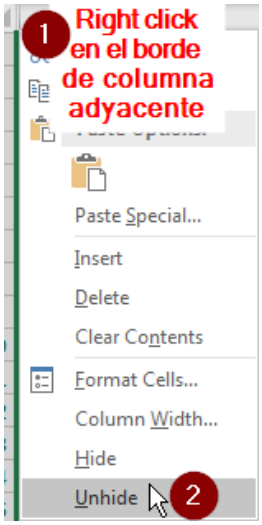
Nota:

Los filtros son persistentes. Un filtro permanece a menos que ud lo desactive. Esto tiene efecto en filtros posteriores que usted quiera aplicar.

Revelar una columna escondida

Para revelar (**unhide**) columna(s), haga **right click en el borde de la cabecera de la columna adyacente**. Verá que el cursor cambiará a dos flechas:

→



Haga **right click** y escoja del menú la opción **Unhide**

También puede seleccionar varias columnas adyacentes a la columna escondida y hacer **right click | Unhide**

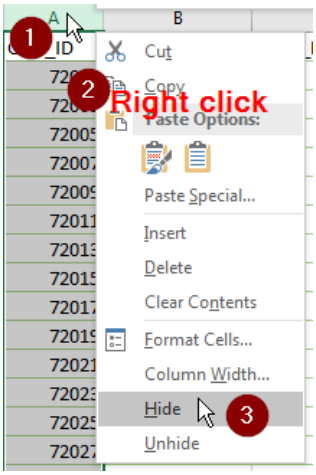
Aparecerá la columna ***GEO_ID***. Esta contiene los identificadores *FIPS* que asigna el Negociado del Censo a cada municipio, barrio y demás áreas administrativas y censales estadísticas.

GEO_ID	Municipio	POP_EST201	ZONA
72021	Bayamón	192974	Metro
72031	Carolina	165040	Metro
72033	Cataño	26137	Metro
72051	Dorado	38013	Metro
72061	Guaynabo	92444	Metro
72087	Loíza	27857	Metro
72127	San Juan	363744	Metro
72135	Toa Alta	74467	Metro
72137	Toa Baja	83645	Metro
72139	Trujillo Alto	70737	Metro

Esconder columna

Para esconder columna:

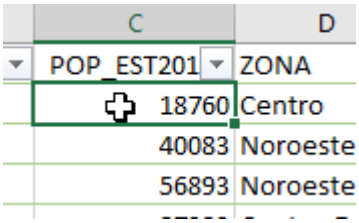
- Seleccione/**sombree** la(s) columna(s). Haga **click** en la **cabecera** de la **columna A**.
- Haga **right click** y escoja la opción **Hide**



Formatear columna

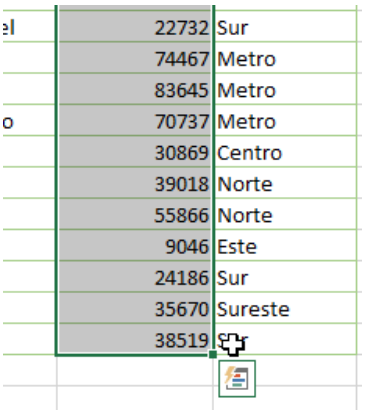
Seleccione las celdas C2 a C79. Use el siguiente método:

Click en la celda de inicio **C2** presione **CRTL +SHIFT+ Flecha hacia abajo↓**

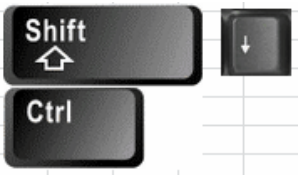


C	D
POP_EST201	ZONA
18760	Centro
40083	Noroeste
56893	Noroeste

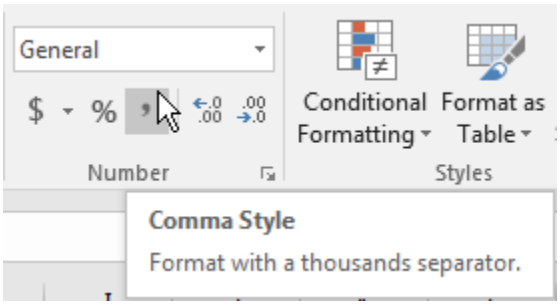
… para llegar al último récord



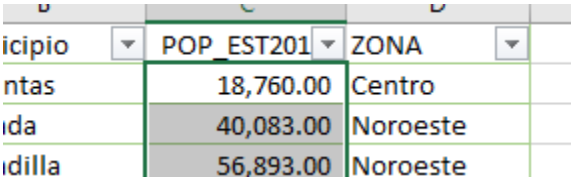
22732	Sur
74467	Metro
83645	Metro
70737	Metro
30869	Centro
39018	Norte
55866	Norte
9046	Este
24186	Sur
35670	Sureste
38519	



Haga **click** en el tab **Home**. Vaya al **grupo Number** y haga **click** en el botón **Comma Style** para añadir el separador de coma a los millares.

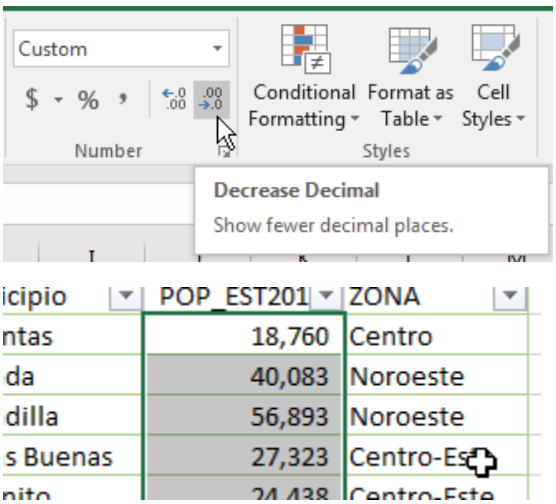


Verá que aparecen los números con comas, pero también con espacios decimales innecesarios.



QUITAR DECIMALES

Con las **celdas seleccionadas**, haga **click dos veces** en el botón **Decrease Decimal**



SPARKLINES (MINI-GRÁFICAS)

Los sparklines son mini gráficas que se generan dentro de una celda. Es una manera de visualizar tendencias en datos voluminosos, sin tomar demasiado espacio.

OBJETIVOS

Al completar esta lección, usted podrá:

- Añadir sparklines
- Escoger el sparkline más apropiado
- Remover sparklines
- Formatear sparklines

SPARKLINES

Los sparklines son mini-gráficas que caben dentro de una celda

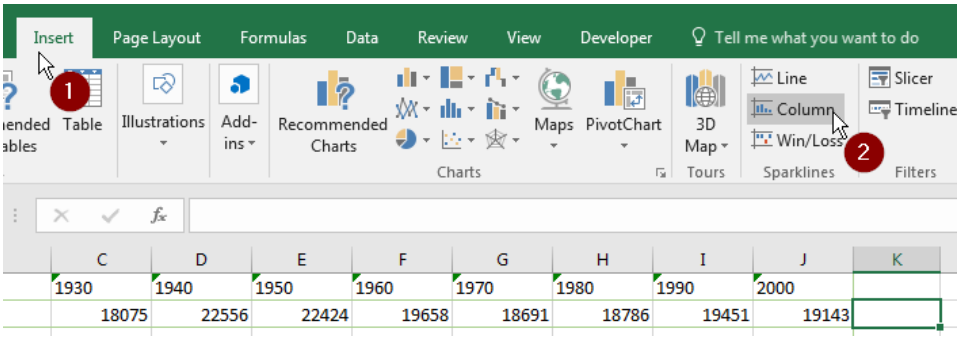
- Hay tres tipos
  - Líneas, columnas, Win/Loss
- Pueden añadirse y removerse
- Ubicarlas cerca de los datos para mejor entendimiento.

QUÉ ES UNA MINI-GRÁFICA SPARKLINE

Los sparklines no son objetos de una gráfica, aunque parecen pequeñas gráficas. Se trata de gráficos simples ubicados en el fondo de una celda. Esto significa que puede haber texto encima del espacio donde aparece el sparkline. Cada sparkline representa una fila de datos.

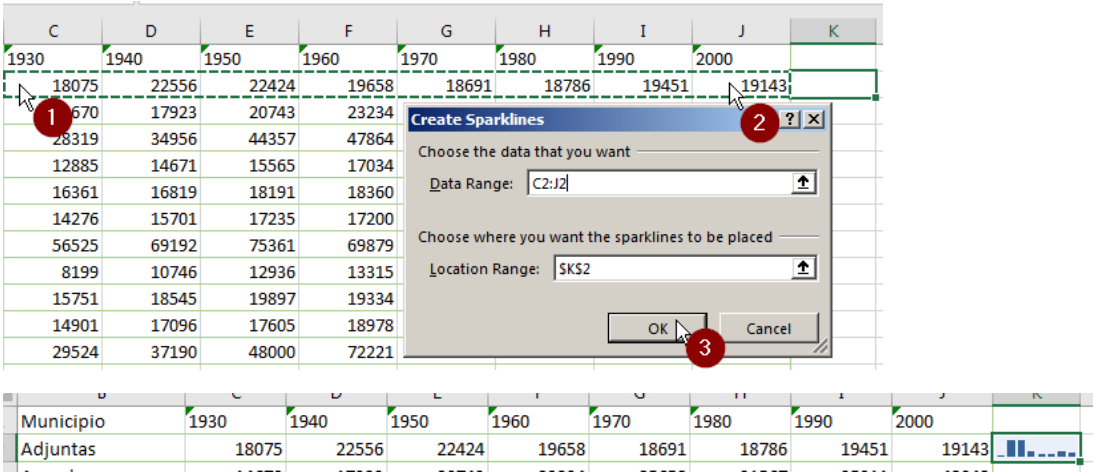
AÑADIR SPARKLINES

- Para añadir sparklines, haga **click** en el tab **Insert** y seleccione un tipo de sparkline en el grupo **Sparklines**.



Aparecerá el diálogo **Create Sparklines**

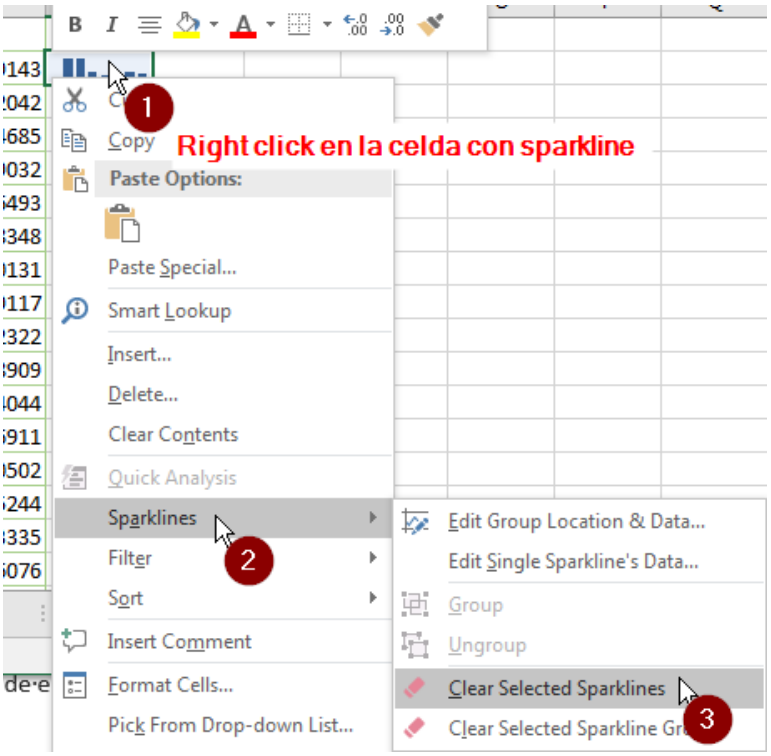
- Seleccione o escriba la extensión de celdas (data range)
- Indique la celda destino para ubicar la mini-gráfica sparkline
- Click en el botón OK



REMOVER SPARKLINES

Haga **right click** en la **celda que contiene el sparkline**

Haga **click** en **Sparklines** y **click** en la opción **Clear Selected Sparklines**



Uso de sparklines

Los sparklines muestran patrones en los datos, haciendo que la vista se concentre en puntos específicos en los datos. Estas ayudan a mostrar más fácilmente los patrones existentes en los datos y fomentan que las personas miren dónde están los números que crean estos patrones.

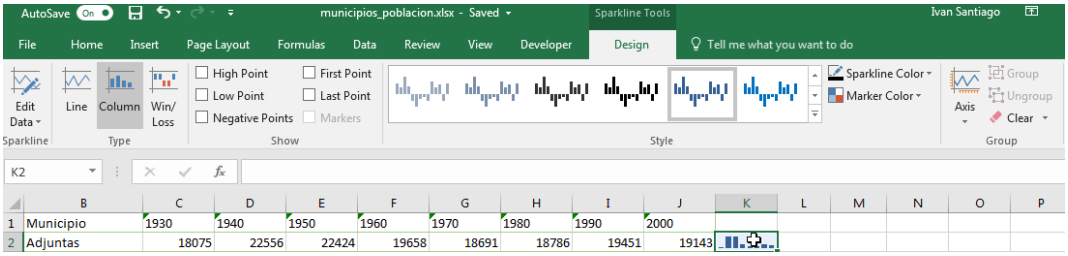
Recomendaciones para sparklines:

Ubicar los sparklines al lado de los datos que están representando.

Formateo de sparklines

Para hacer cambios a sparklines:

- **Seleccionar** la **celda** donde está el sparkline. Se activará el tab **Sparkline Tools | Design**

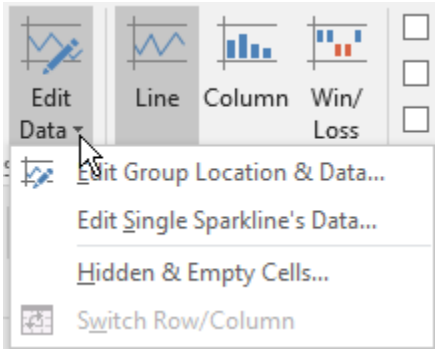


- **Edit data**: Cambiar el data range y la posición del sparkline
- **Tipo**: Línea, columna, win-loss
- **Show**: Resaltar lugares específicos de los datos
- **Style**: Colores y estilos
- **Group**: Crear o remover grupo, eliminar sparklines, definir configuración *Settings* de los ejes

Nota: Los sparklines pueden ser agrupados. Esto facilita modificar varios sparklines incluidos dentro del grupo.

Menú y herramientas Sparklines

La opción **Edit Data** le permite **cambiar el data range** utilizado para generar el sparkline, así como la **ubicación del sparkline**. Al hacer **click** en la flecha, aparecerán las opciones:



Edit Group Location and Data: Para hacer cambios en cuál es el data range a usarse, así como la localización del sparkline.

Edit Single Sparkline's Data: Para hacer cambios (data range y ubicación) para una celda sparkline.

Hidden and Empty Cells: Para especificar cómo será el manejo de celdas vacías, además de especificar la inclusión o exclusión en el sparkline de celdas escondidas.

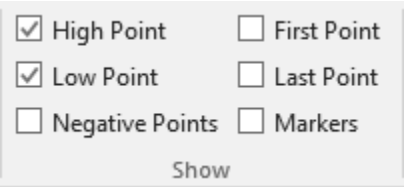
Tipos de sparklines

Línea, columna, Win/Loss

Estas pueden ser cambiadas al tipo apropiado según los datos.

Show

Para llamar la atención a diferentes puntos de los datos. Por ejemplo:



High Point: Mostrar el punto más alto.

Low Point: Mostrar el punto más bajo.

Negative Points: Mostrar puntos negativos, si existen

First Point: Primer punto

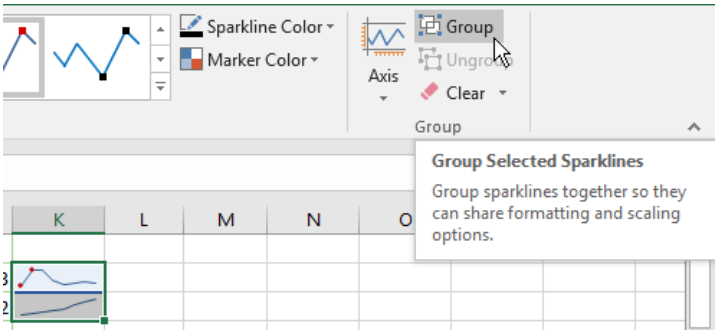
Last Point: Último punto

Markers: Marcadores (mostrar los vértices solo en líneas)

Style

Este grupo de botones sirve para cambiar los colores de las líneas de un grupo, además de los colores de los markers.

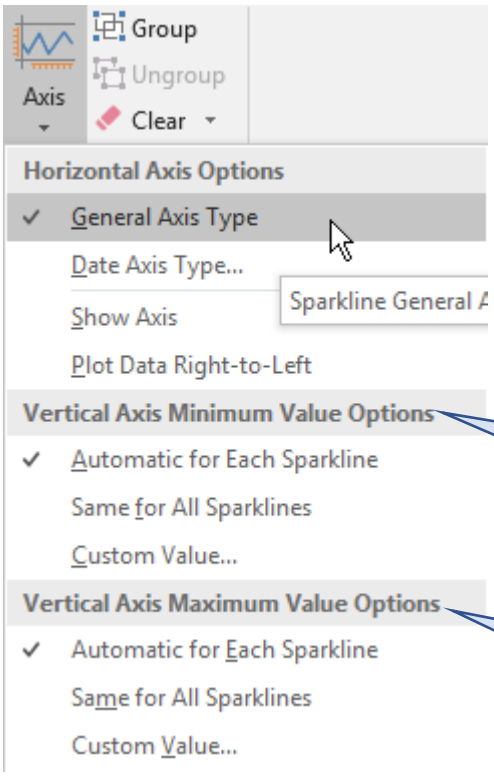
Group



El botón **Group** se usa para agrupar varios sparklines y tratarlos como una sola mini-gráfica al momento de cambiar el formato.

Axis:

El menú **Axis** contiene múltiples opciones.



Axis permite hacer **cambios en los ejes** de la mini-gráfica o en el grupo de mini-gráficas.

Por ejemplo, podemos establecer el mismo punto de origen y extensión del eje para estandarizar las escalas. De esta manera se podrán hacer comparaciones de manera más efectiva.

El eje horizontal puede cambiarse a fecha y, además invertirlo.

Opciones del **eje vertical** para **valor *mínimo***

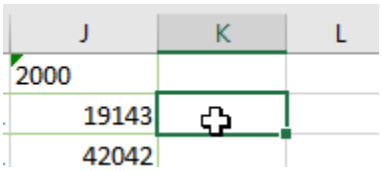
Opciones del **eje vertical** para **valor *máximo***

DEMO: AÑADIR SPARKLINES

En **Excel**, **abra** el workbook **municipios_población.xlsx**, localizado en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_1**

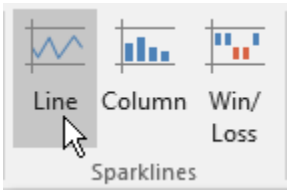
Active la hoja **POP_HISTORIC**, haciendo **click** en la pestaña (tab) con ese nombre.

Ubíquese (**click**) en la celda K2

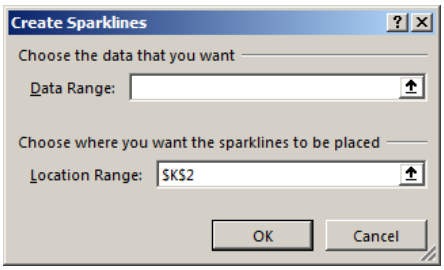


Haga **click** en el tab **Insert**,

luego **click** en el botón **Line** del grupo **Sparklines**.

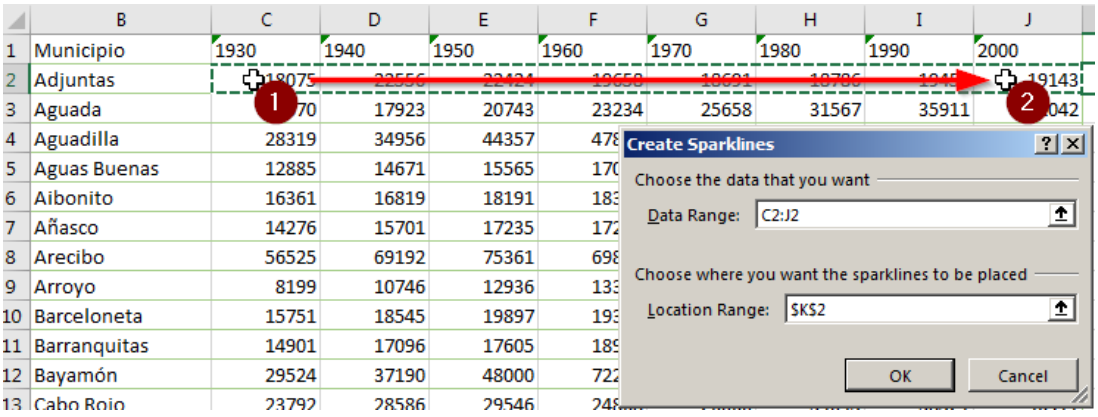


Aparecerá la forma (dialog box) **Create Sparklines**.



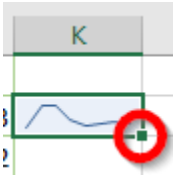
Haga **click** en el botón correspondiente a **establecer** el **data range** para este sparkline.

Haga **click** en la celda **C2** y **arrastre el cursor hasta** la celda **J2**

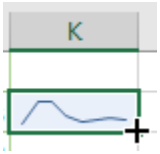


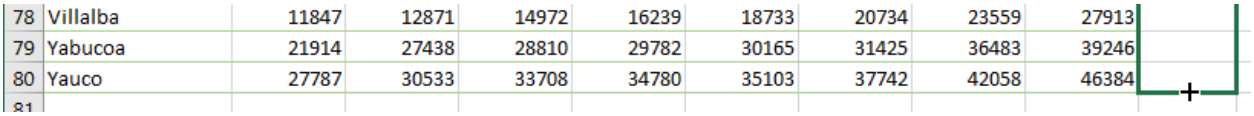
Haga **click** en el botón **OK**.

Estando ubicado en la celda K2, **fíjese** en el pequeño **cuadro** (punto) **verde** que aparece en la esquina inferior derecha de la misma:



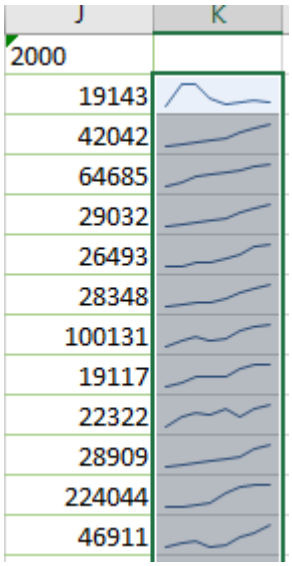
Ubique el cursor en ese punto verde y haga **click encima del punto verde**.
Mantenga el punto presionado y **arrastre** el cursor **hasta** la **celda K80**





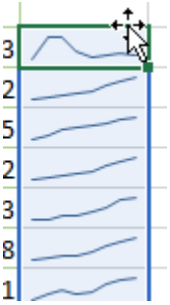
78	Villalba	11847	12871	14972	16239	18733	20734	23559	27913
79	Yabucoa	21914	27438	28810	29782	30165	31425	36483	39246
80	Yauco	27787	30533	33708	34780	35103	37742	42058	46384
81									

Podrá ver los sparklines en la columna:



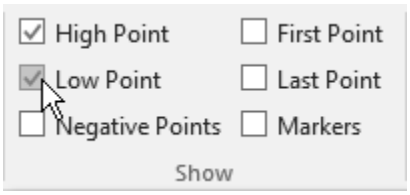
Verá que **la mayoría de los municipios *aumentaron*** su **población** de **1930 a 2000**. Además, podrá notar la pendiente de cada línea, la cual le dará idea de la magnitud del aumento o descenso de la población municipal.

Cambiar formato a Sparkline



Manteniendo la celda **K2 activada**, verá que esta celda **activa** también **el resto de las celdas contiguas** hacia **abajo** hasta la celda K80. **Esto quiere decir que los sparklines de estas celdas están en el mismo grupo**.

Vaya al grupo **Show** y **escoja** *check* las opciones **High Point** y **Low Point**.



Podrá ver los puntos mínimo y máximo en cada sparkline:

	B	C	D	E	F	G	H	I	J	K
	Municipio	1930	1940	1950	1960	1970	1980	1990	2000	
	Adjuntas	18075	22556	22424	19658	18691	18786	19451	19143	
	Aguada	14670	17923	20743	23234	25658	31567	35911	42042	
	Aguadilla	28319	34956	44357	47864	51355	54606	59335	64685	

Con estas líneas, es mucho más fácil poder identificar municipios que han tenido pérdidas constantes de población o crecimiento bien modesto. Esto se puede notar en los municipios de Vieques, Utuado y Maricao.

74	Utuado	37434	42531	46625	40449	35494	34505	34980	35336
75	Vega Alta	12333	14329	16521	17603	22810	28696	34559	37910
76	Vega Baja	20406	23105	28925	30189	35327	47115	55997	61929
77	Vieques	10582	10362	9228	7210	7767	7662	8602	9106

Aparecerán además otros municipios con puntos mínimos posteriores a la década de 1930; estos bajaron su población luego de 1930 en vez de aumentarla, contrario a la tendencia generalizada en el resto de los municipios. El caso de Vieques es atípico porque muestra mayor población en 1930 que en el censo de 2000.

PRÁCTICAS:

ESCENARIO

Imagínese que a usted le piden **mostrar gráficas sobre datos de escuelas públicas del Departamento de Educación**, **a nivel descriptivo**. Se trata de un directorio de escuelas públicas para 2017, el cual fue descargado desde el portal data.pr.gov, manejado por el Instituto de Estadísticas. El interés inicial es **saber si los datos muestran** algún tipo de **correlación** (positiva o negativa) **entre el nivel de pobreza** registrado **y el nivel** "**competente a avanzado**" **en matemáticas**.

OBJETIVOS

Con las destrezas mostradas a lo largo de este módulo usted practicará:

- Hacer gráficas de histograma y diagramas de dispersión para mostrar distribuciones de datos
- Cambiar formatos para hacerlos más legibles
- Filtrar datos

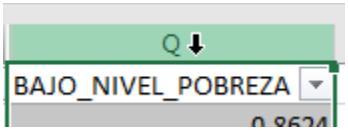
EJERCICIO 1: GRÁFICAS Y EXPLORACIÓN DE DATOS

Abra el workbook llamado **DirectorioEscuelasPublicas2017.xlsx**, localizado en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_1**

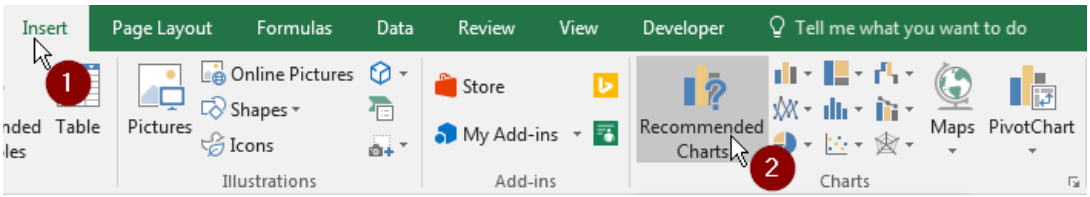
Este workbook tiene solo una hoja con un directorio de escuelas públicas, al cual se le eliminaron algunas filas que no tenían datos en las columnas de puntuaciones en las competencias español-inglés-matemáticas.

HACER HISTOGRAMA

Haga **click** en la cabecera de la columna **Q** (**BAJO_NIVEL_POBREZA**)



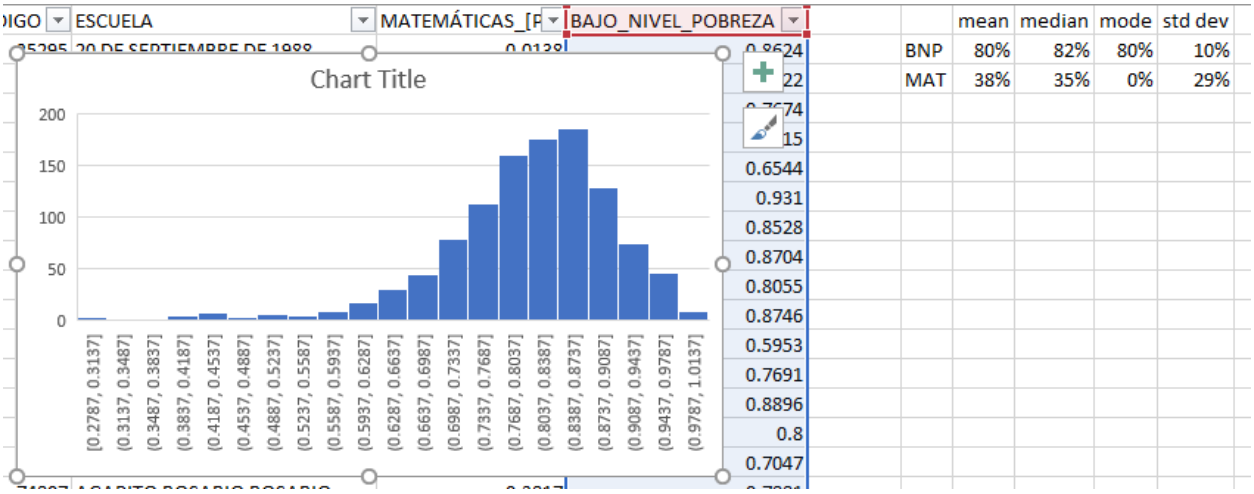
Vaya al **menú principal** y haga **click** en **Insert**, luego **click** en **Recommended Charts**



En la forma **Insert Chart**, escoja la **tercera opción** en la lista en el panel izquierdo (*Histogram*).

Haga **click** en el botón **OK**.

Aparecerá la gráfica de histograma como objeto flotante en la hoja activa:



Al lado derecho del data range, hay un pequeño data range que contiene estadísticas descriptivas de los datos de la columna **BajoNivelPobreza** (BNP) y la columna **Matemáticas** (MAT).

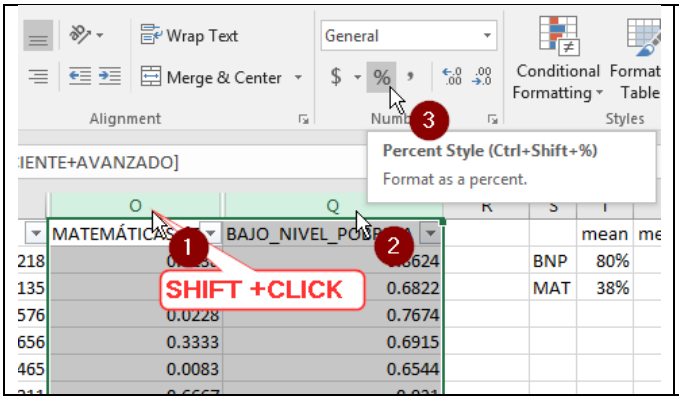
	mean	median	mode	std dev
BNP	80%	82%	80%	10%
MAT	38%	35%	0%	29%

Enseguida se puede notar la distribución de los datos. En este histograma podemos ver que la mayoría de los datos se concentra en los valores alrededor de 0.8. Este a su vez, (80%) es el valor promedio (*mean*) de las escuelas. La distribución muestra una "cola" al lado izquierdo. Estas son las escuelas con menos estudiantes bajo el nivel de pobreza.

Ejercicio 2: Formatear datos

En esta parte, vamos a cambiar el formato de los números para que indiquen que tipo de cifras son. Por ejemplo, números por cantidades mostrando comas, decimales, porcentajes, moneda, y otros. La idea aquí es presentar los números en sus formatos apropiados y hacerlos más comprensibles.

Vamos a cambiar el formato a **porcentaje** a las columnas **O** y **Q**.



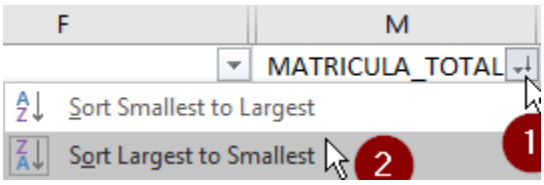
Presione la tecla **SHIFT** y haga **click** en las cabeceras de las columnas **O** y **Q**.

Una vez estén seleccionadas, haga **click** en el **botón Percent Style** % para añadirle los símbolos de porcentaje.

Haga **click** en el botón .0 .00 **Increase Decimal**, para añadir un lugar decimal.

Repita los pasos modificando para que cambie ahora el formato de la columna **M** (**MATRÍCULA_TOTAL**) a formato **numérico con comas sin números decimales**.

Ordene la columna **M** MATRICULA_TOTAL de forma **descendente**, como aparece en esta figura.



Luego de ordenar la columna M, así deben verse los números.

M	O	Q
MATRICULA_TOTAL	MATEMÁTICAS_[P	BAJO_NIVEL_POBREZA
1,309	17.8%	69.7%
1,191	18.9%	73.0%
1,185	2.8%	68.8%
1,182	3.8%	71.9%
1,148	0.3%	85.3%

Ejercicio 3: Filtrar datos

En esta parte, vamos a **hacer** una **gráfica de dispersión**. Luego de haber preparado la gráfica, **aplicaremos** algunos **filtros para explorar** los **datos** y ver si podemos **saber** si **hay** algún **patrón o relación entre** las variables (columnas) **nivel de pobreza** y **competencia avanzada en matemáticas**, así como **matrícula total** y **competencia avanzada en matemáticas**.

Objetivos:

- Primero haremos la gráfica de dispersión (*scattergram*) con todos los datos
- Luego, cambiaremos de columnas y aplicaremos los filtros.

Modificar posición de columnas para hacer diagrama de dispersión

Para hacer este diagrama de dispersión:

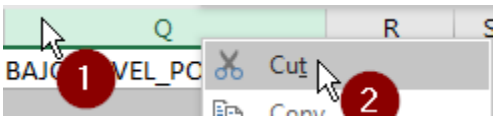
Columna	Tipo de Variable	Eje	Posición de la columna
"BAJO NIVEL POBREZA"	Independiente	X	Izquierda
"MATEMATICAS_"	Dependiente	Y	Derecha

En otras palabras, si queremos probar que el **porciento bajo nivel de pobreza incide en la competencia en matemáticas**, la columna Q debe moverse a la izquierda de la columna O.

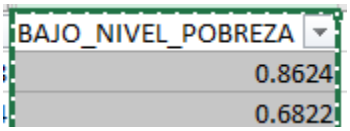
Cortar y pegar columna

Vamos a mover la columna Q al lado izquierdo de la columna O.

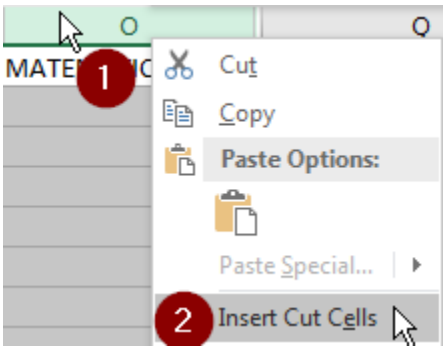
Haga **right click** en la **cabecera** de la columna **Q** y escoja **Cut**



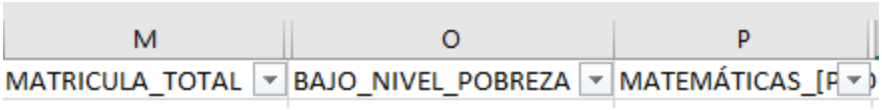
Excel representará los que datos están copiados en el *clipboard* (en memoria RAM) usando línea entrecortada de esta manera:



Haga **right click** encima de la **cabecera** de la columna **O** y escoja la opción **Insert Cut Cells**



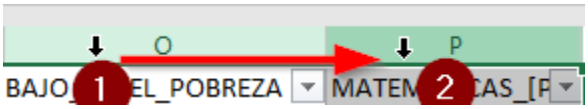
Note que la columna **Q** cambió de letra a **O**:



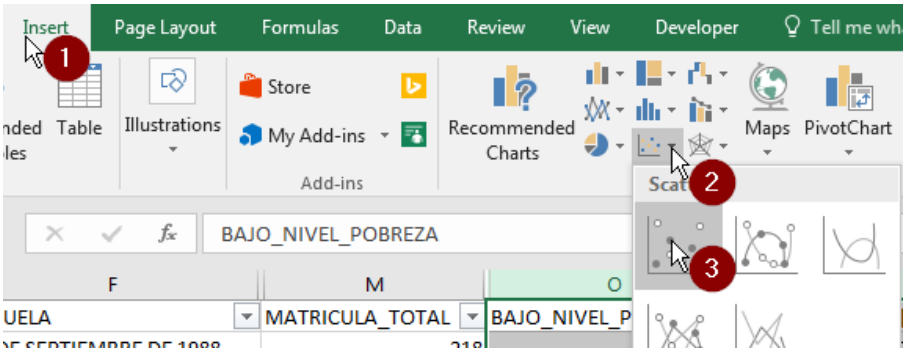
Explorar si hay relación entre nivel de pobreza y desempeño alto en matemáticas

Ya que tenemos preparada esta hoja, pasemos a hacer las gráficas de dispersión.

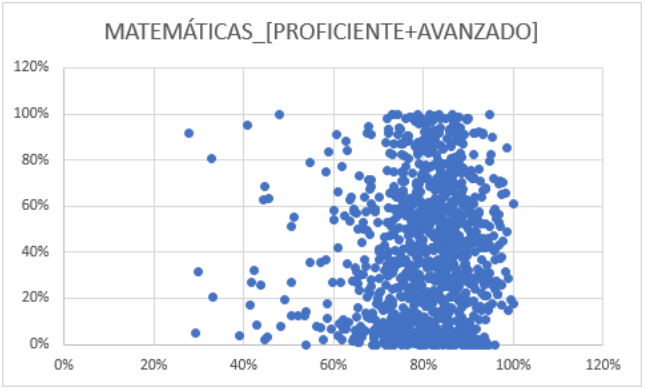
Haga **click** y **arrastre el cursor** (drag) en las cabeceras de las columnas **O** y **P**.



Vaya al **menú principal** y haga **click** en **Insert**, luego **click** en la **flecha** del botón **Scatter** y **click** en la **opción Scatter**



Así aparecerá la gráfica.



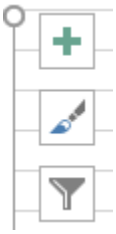
El **eje horizontal** (X) representa **porcentaje de escuelas bajo el nivel de pobreza**, el **eje vertical** (**Y**) representa **porcentajes de nivel de competencia en matemáticas avanzadas**

- Los porcentajes bajo el nivel de pobreza fluctúan entre 30-a-100%. Los niveles de competencia en Matemáticas van desde cero a 100%
- Los valores en el eje de X (%bajo nivel de pobreza) se concentran alrededor de 80%.
- Los porcentajes de competencia en matemática van desde 0 a 100
- La distribución de puntos no muestra que haya correlación que sugiera que al aumentar los niveles de pobreza, baje la competencia en matemáticas

Añadir línea de tendencia

Visualmente podemos constatar que no hay correlación significativa entre el nivel de competencia en matemáticas (en escuelas públicas) y el porcentaje bajo el nivel de pobreza. Si hubiese algún tipo de correlación, podríamos ver que los puntos se concentrarían en forma diagonal hacia arriba o hacia abajo (correlación positiva o negativa).

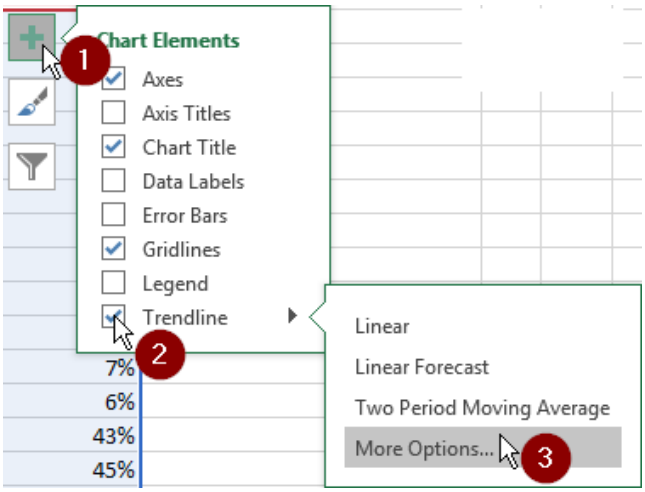
Podemos mostrar una línea de tendencia para sustentar esto y además añadir la estadística coeficiente de determinación **R^2**. Los posibles valores de R^2 van de 0 a 1. Mientras más se aproxime R^2 a 1, mayor será la correlación entre ambas variables.



Para añadir una línea de tendencia a esta gráfica, haga **click** en la gráfica para activarla. Le aparecerán tres botones (cuadros).

Haga **click** en el botón **Chart Elements** .

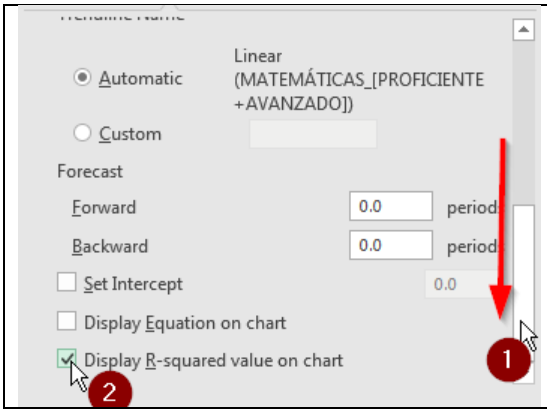
En el menú de contexto **Chart Elements** que aparecerá, haga **click** en la opción **Trendline**.



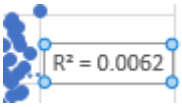
Luego **click** en **More Options…**

Vamos a añadir una etiqueta con el valor del coeficiente de determinación R^2.

En la forma **Format Trendline**, use la barra para navegar hacia abajo y haga **check** en la opción **Display R-squared value on chart**.



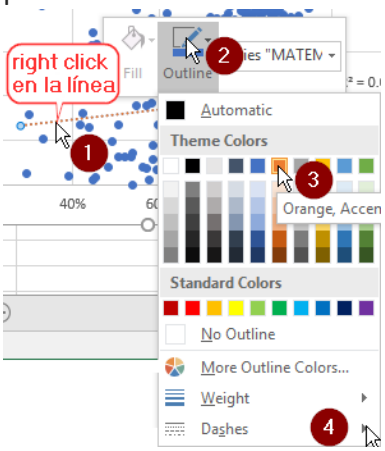
Aparecerá el valor R^2 en la gráfica. Muévalo de ser necesario.



El valor **R^2 = 0.0062** muestra que **no hay correlación significativa** entre los datos de las dos columnas.

Cambiar el color de la línea de tendencia

Cambiaremos el color de esta línea para que sobresalga un poco en relación con la nube de puntos.



Haga **right click** en la línea para activar el menú de contexto.

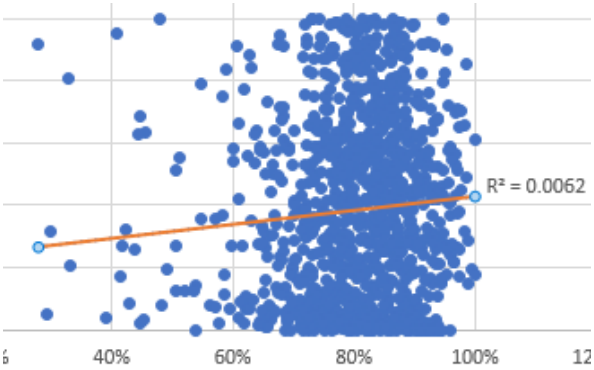
Haga **click** en el botón **Outline**.

Click en el **botón Orange, Accent**

Click en la flecha al lado de la opción **Dashes**

En la lista de líneas (***dashes***) escoja la primera opción (***solid***)

Así debe verse la línea encima del diagrama de dispersión.

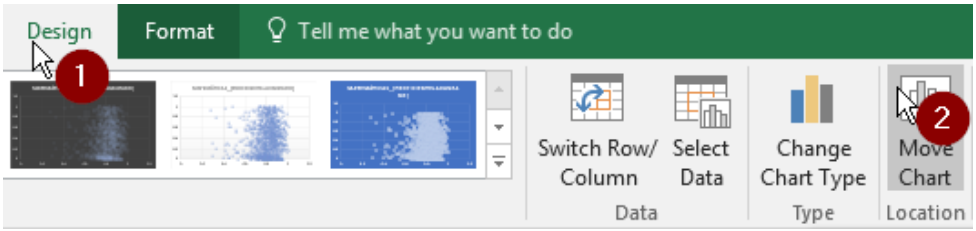


OPCIONAL: Mover la gráfica a otra hoja

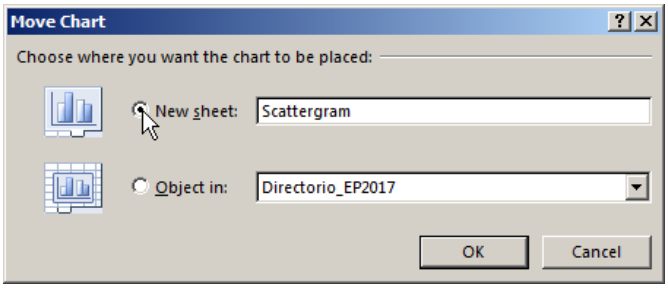
Vamos a mover la gráfica a una hoja aparte.

Para mover esta gráfica, actívela haciendo **click** encima de ella. **Deberá** tener **8 círculos** alrededor de su cuadrado externo.

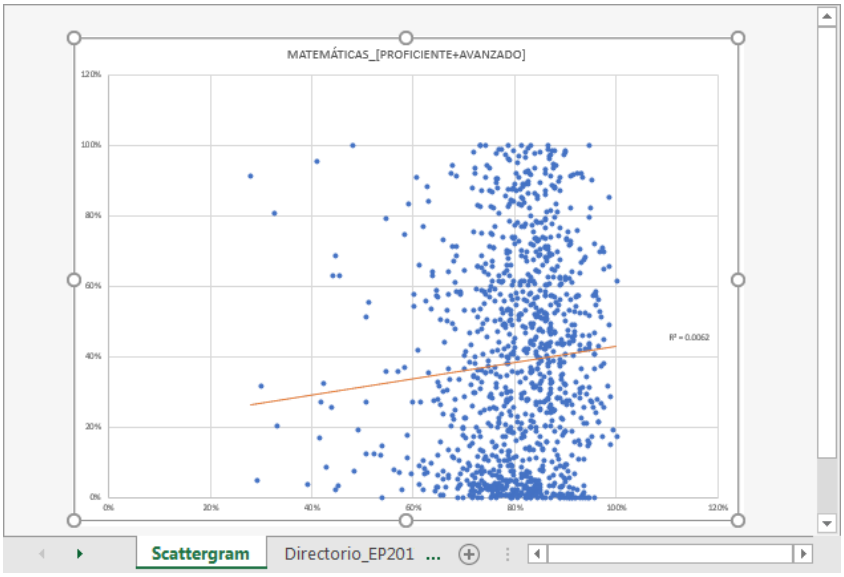
Una vez activada la gráfica, vaya al **menú principal**, haga **click** en el tab **Design** y en **Location** haga **click** en el botón **Move Chart**



En la forma **Move Chart**, escoja la opción **New Sheet** y en la caja de texto **escriba Scattergram**



Excel moverá la gráfica a otra hoja, la cual tendrá el nombre asignado.



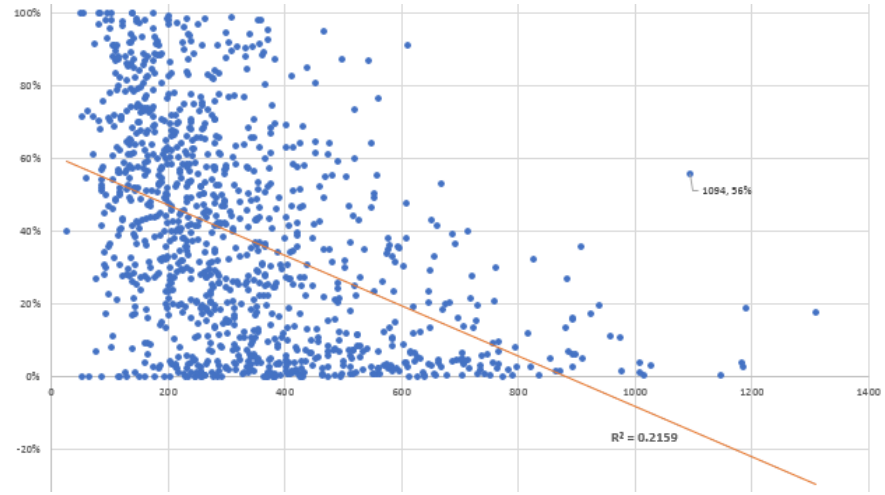
Añadir otro diagrama de dispersión: Matrícula vs Competencia avanzada en matemáticas

Vamos a poner ahora la cantidad de estudiantes matriculados versus competencia avanzada en matemáticas en otra gráfica. Entendemos que mientras mayor número de estudiantes, menor debería ser la atención del

estudiante y mayor dificultad del maestro para retener la atención de los estudiantes. Esto puede incidir en el desempeño en las pruebas.

Repita el procedimiento anterior para añadir otro diagrama de dispersión para las columnas **M** (**Matrícula**) y **P (Matemáticas)**

Al final, notará que los datos muestran una correlación débil $R^2 = 0.216$. En general, entre mayor la matrícula, menor la competencia en matemáticas. Hay pocas escuelas con matrículas grandes que a su vez tengan porcentajes altos en matemáticas. Escuela Dr. Carlos González en el municipio de Aguada con 1,094 estudiantes y 56% en matemáticas.

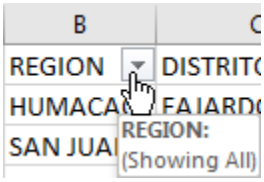


OPCIONAL APLICAR FILTROS

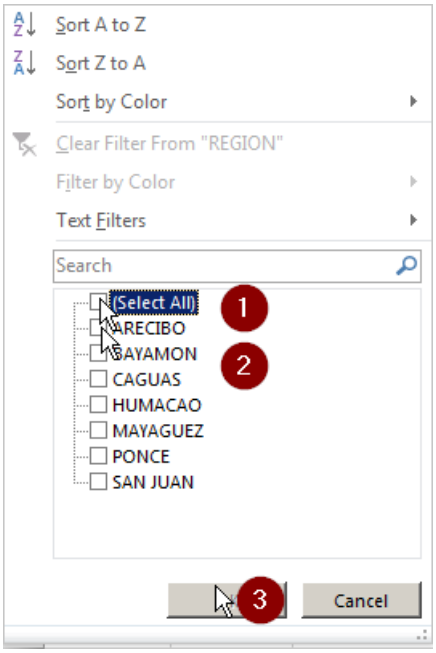
Finalmente, para aplicar los filtros, lo haremos por regiones educativas. Veremos los resultados en la hoja y en la gráfica.

Vaya a la hoja **DIRECTORIO_EP2017**.

Bajo la columna **B (REGION)**, haga **click** en el **botón**-flecha de **filtros** y **Sorting**. Verá que este lee **(Showing All)**.



En el **menú de contexto** que aparecerá, haga **click** en las opciones en el orden que aparecen en la figura.



Uncheck en **Select All**,

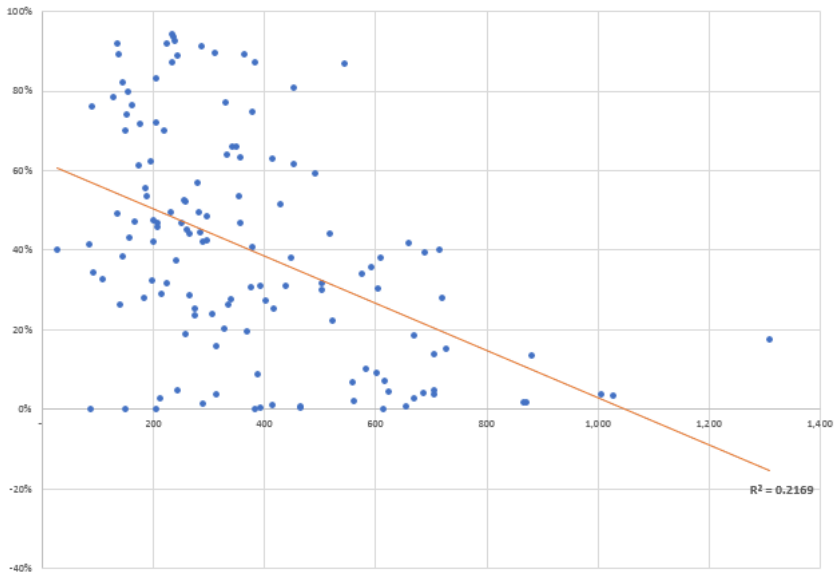
check en **ARECIBO** y

click en botón **OK**.

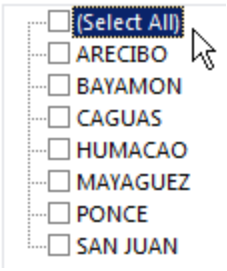
Notará que los números de las filas cambiaron a azul y que hay celdas escondidas:

Haga **click** ahora en la hoja que contiene la gráfica **Scattergram2**.

Verá que los resultados son consistentes con el resultado global:



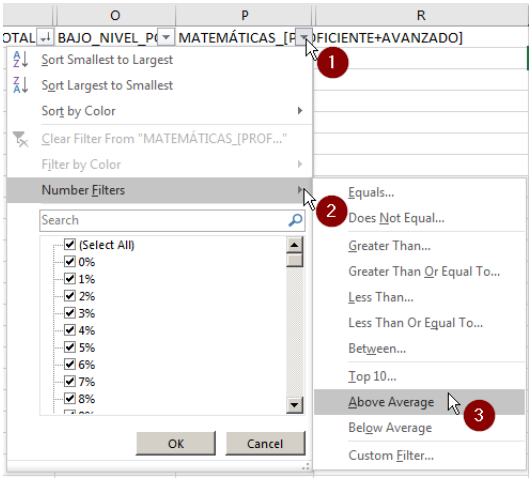
2: Repita el proceso con las demás regiones.



Recuerde quitar el filtro anterior y **aplicar filtro** a **cada región**.

¿Ve alguna diferencia?, ¿Cree que los resultados pueden ser diferentes usando otros filtros?, ¿Por municipio, zona?

3: Los filtros pueden ser aplicados a números.



Quite otros filtros e intente aplicar un filtro (**Above Average**) a la **columna P** Matemáticas.
Examine los resultados de la hoja del data range y de la gráfica.

Revisión de conceptos

Data range: _____

Filtro: _____

Sparkline: _____

MÓDULO 2: EXCEL TABLES Y FUNCIONES

RESUMEN

Según los datos van acumulándose y haciéndose más detallados, aumenta la necesidad de poder manipular eficientemente los datos. Más allá de los data ranges y los named ranges, las nuevas "***tablas***" en Excel ayudan a el manejo, análisis y resumen de los datos mediante ayudas visuales y funcionalidades diferentes aplicadas a estas ***tablas***.

OBJETIVOS

Luego de completar este módulo, usted podrá:

- Describir qué es una tabla y su utilidad
- Crear una tabla Excel
- Usar tablas para resumir datos
- Describir algunas funciones para aplicar fórmulas y agregar datos

TABLAS DE EXCEL (EXCEL TABLES)

En esta parte, se discutirá sobre las tablas y su utilidad.

OBJETIVOS

Una vez completada esta parte, será capaz de:

- Describir los beneficios de las tablas
- Crear y borrar una tabla
- Identificar los elementos que componen una tabla

BENEFICIOS DE LAS TABLAS

Una tabla es una extensión de celdas (data range) formateadas para ayudar a visualizar y analizar datos

- Todas las celdas de una tabla son tratadas en conjunto
- Se le aplican colores alternados o esquemas de color para hacer los datos más legibles
- Aparecerán botones de filtros a las columnas
- Se puede añadir una fila de totales o resúmenes numéricos
- Provee una manera rápida para comenzar el análisis de datos

QUÉ ES UNA TABLA

Una tabla de Excel es un grupo de celdas formateadas de forma particular para ayudar en la visualización y el análisis de datos.

FORMATEO

Al crear una tabla, las celdas se formatean utilizando colores por fila y columna, filtros en cada columna. El formato visual se utiliza para ayudar en el proceso de análisis de los datos; no

solamente para que las celdas luzcan bien. Esto ayuda en el proceso de análisis, ya que además de proveer botones para hacer filtros, las celdas aparecerán mejor organizadas.

Utilidad de las tablas

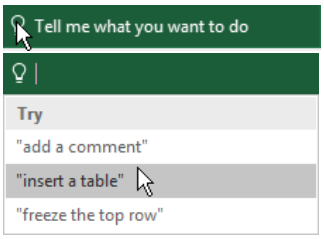
Cuando se genera una tabla, todas las celdas que la componen son tratadas como un conjunto, de manera similar a un **data range con nombre** (*named data range*) pero con mayor funcionalidad. Por ejemplo, al tratar de cambiar formatos, se cambiarán los formatos a las filas o a las columnas. Además, se aplicarán cómputos en ristra a las celdas de una columna hacia abajo automáticamente.

Crear tablas

Para crear tablas:

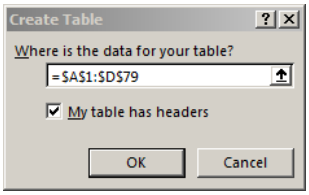
- **Sombrear** un **data range** y hacer **click** en **Insert** | **Table**

- De *otro modo*: hacer **click** en **Tell me what to do** y **click** en **Insert a Table**



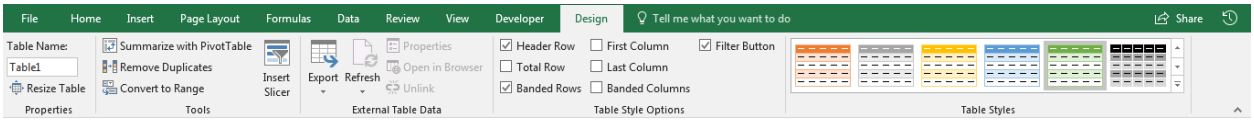
- Aparecerá el tab **Design**, el cual le permitirá las funcionalidades de:
  - Color y estilos
  - Color en filas y/o columnas
  - Ennegrecer (*bold*) los nombres de las columnas
  - Mostrar cabeceras de columnas y fila de totales
  - Remover duplicados
- Para **convertir una tabla a *data range***, haga **click** en **Design** | **Convert to Range**.

- Puede crear una tabla a partir de un data range usando **CTRL+A** *dentro* del data range y luego **CTRL+T**. Si su data range ya tiene nombres de columnas, mantenga **check** en **My table has headers**. Finalmente, **OK** en la forma **Create Table**.



FORMATEO DE LA TABLA

Luego de haber creado la tabla, aparecerá el tab **Design** con sus botones y múltiples opciones



Properties	Para nombrar la tabla
Tools	Manipulación de tablas
External Table Data	Conexión a datos en tablas externas
Table Style Options **Table Styles**	Maneras para cambiar la apariencia de las tablas

Por el momento, hablemos un poco del grupo **Tools**; especialmente los botones **Convert to Range** y **Remove Duplicates**.

Remove Duplicates: como dice el nombre, este botón se usa para eliminar récords duplicados. En ocasiones, pueden ocurrir duplicados, ya sea por error o de manera intencional. En ocasiones, se debe analizar usando una sola ocurrencia o un récord para no sobreestimar estos datos.

Convert to Range: se usa para **convertir** la tabla **en** un **data range**.

ORDENAR Y FILTRAR

Al convertir a tabla, se añaden automáticamente los botones de filtros a cada columna.

	B	C	D
1	Municipio	POP_EST2016	ZONA
5	Aguas Buenas	27,323	Centro-Este
6	Aibonito	24,438	Centro-Este
11	Barranquitas	29,548	Centro-Este
14	Caguas	136,372	Centro-Este
19	Cayey	46,071	Centro-Este
22	Cidra	41,782	Centro-Este
24	Comerío	20,153	Centro-Este
34	Gurabo	46,990	Centro-Este
41	Juncos	39,942	Centro-Este
67	San Lorenzo	39,283	Centro-Este

TIPOS DE FILTROS

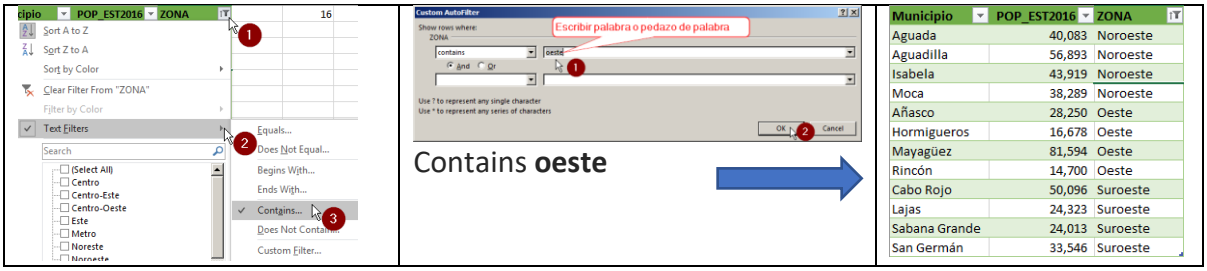
Criterios	Selección de datos de acuerdo con ciertos criterios que se escojan, ya sea por un número, nombre o código, filas antes o después de un valor, celdas entre dos valores, primeros 10, o por porcentaje
Formatos	Selección de acuerdo con formatos de las celdas: color, etc.
Lista	Seleccionar mediante una lista de valores existentes, ya sean nombres o números

Asegúrese que cada columna sea de un solo tipo de dato: texto o numérico. No mezclar números y texto en una columna numérica porque afectará los cálculos.

ORDENAR (SORTING)

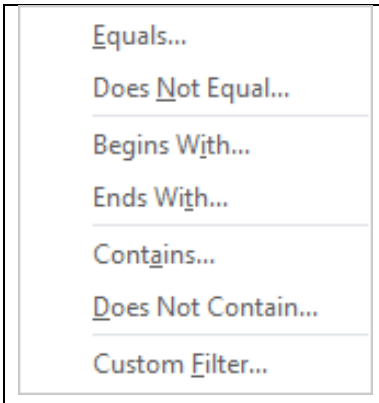
- Ascendente o descendente
- Por fechas
- Las opciones varían según el tipo de dato.
- Ordenamiento personalizado (custom sorting)

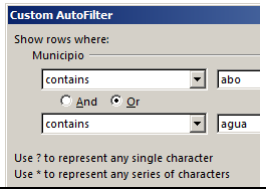
Ejemplo de tabla con un filtro aplicado y ordenado ascendente en la columna ZONA



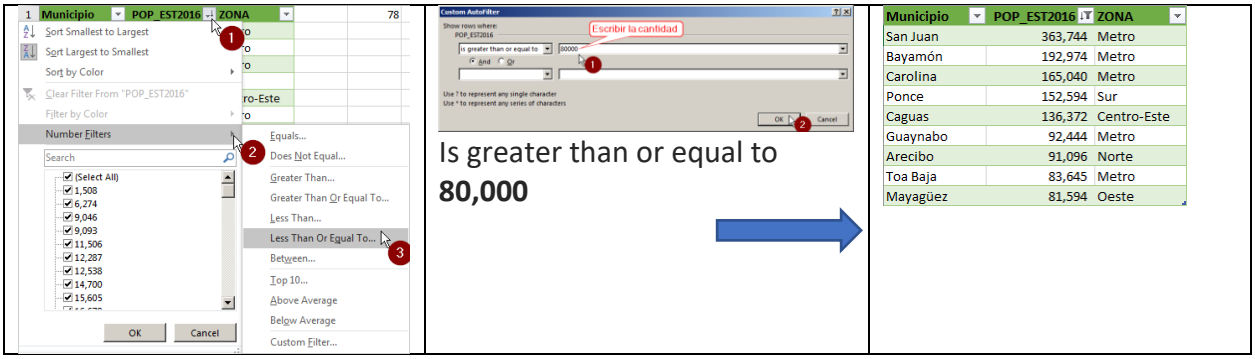
Municipio	POP_EST2016	ZONA
Aguada	40,083	Noroeste
Aguadilla	56,893	Noroeste
Isabela	43,919	Noroeste
Moca	38,289	Noroeste
Añasco	28,250	Oeste
Hormigueros	16,678	Oeste
Mayagüez	81,594	Oeste
Rincón	14,700	Oeste
Cabo Rojo	50,096	Suroeste
Lajas	24,323	Suroeste
Sabana Grande	24,013	Suroeste
San Germán	33,546	Suroeste

Comparación de cadenas alfanuméricas (strings): texto, códigos



Equals…	Pareo **exacto** del texto
Does not Equal…	El texto **no es igual a**
Begins with…	**Comienza con** texto…
Ends With…	**Termina con** texto…
Contains…	Celda(s) **que contengan la cadena de texto**…
Does not contain…	Celda(s) **que NO contengan la cadena de texto**…
Custom filter… 	Hasta dos condiciones para filtrar. En este caso, municipios **que contengan *abo*** y ***agua*** **en** las letras de **su nombre**. Por ejemplo, Guayn**abo**, Gur**abo**, C**agua**s, **Agua**da, **Agua**dilla, **Agua**s Buenas

Ejemplo: Aplicar **filtro numérico**, columna **POP_EST2016**:



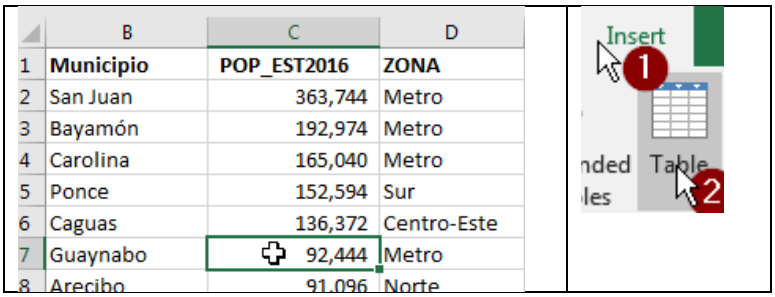
Municipio	POP_EST2016	ZONA
San Juan	363,744	Metro
Bayamón	192,974	Metro
Carolina	165,040	Metro
Ponce	152,594	Sur
Caguas	136,372	Centro-Este
Guaynabo	92,444	Metro
Arecibo	91,096	Norte
Toa Baja	83,645	Metro
Mayagüez	81,594	Oeste

Equals…	Es igual a
Does Not Equal…	No es igual a
Greater than…	Mayor que
Greater than or equal to…	Mayor o igual que
Less than…	Menor que
Less than or equal…	Igual o menor que
Between…	Entre dos valores
Top 10…	Los 10 más altos
Above Average	Sobre el promedio
Below Average	Bajo el promedio
Custom Filter…	Similar al Custom de texto, pero con las opciones numéricas

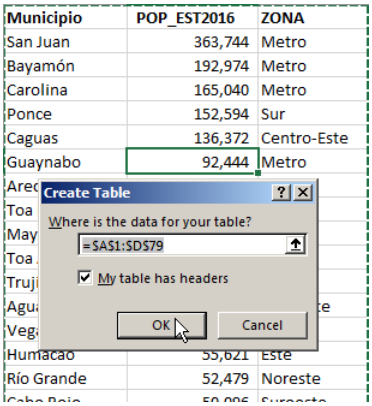
DEMO: USO DE UNA TABLA

El archivo para esta demostración se llama **municipios_poblacion.xlsx** localizado en el folder **C:\Tutorial_Excel\Analisis_1\Modulo_2**.

Para **crear** una **tabla** a partir de un data range, active la hoja (**click** en el **tab**) **POP_EST2016**. Haga **click encima del grupo de celdas**. Luego **click** en **Insert | Table.**

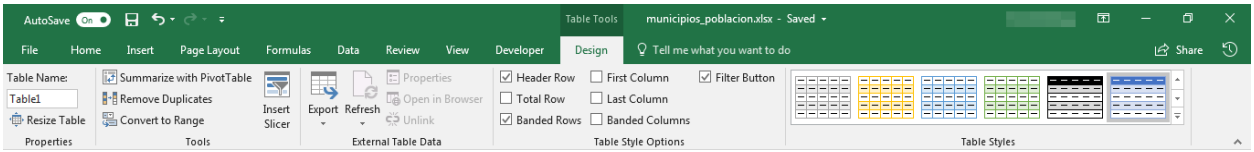


El grupo de celdas aparecerá seleccionado. También aparecerá la forma **Create Table**, en la cual aparecerá la extensión de las celdas que componen el data range. Este data range será convertido en una tabla Excel.



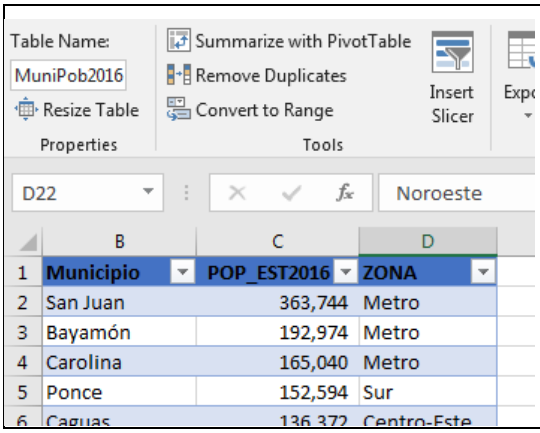
Si su data range ya tiene nombres en las columnas, (***My table has headers***) mantenga **check** en la opción **My table has headers**. Haga **click** en **OK** para aceptar los parámetros para la creación de la nueva tabla.

Inmediatamente Excel nos lleva al **menú principal**, específicamente a **Table Tools | Design** con las opciones que habíamos discutido anteriormente:



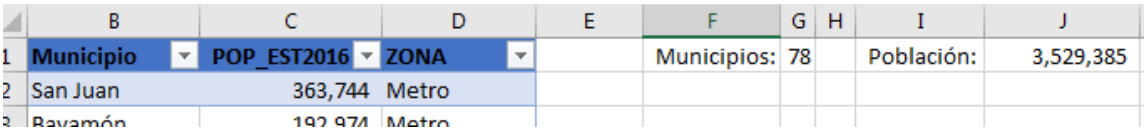
Note que, por defecto Excel le pone un nombre a la tabla: *Table1*.

Cámbiele el **nombre** de *Table1* a **MuniPob2016**



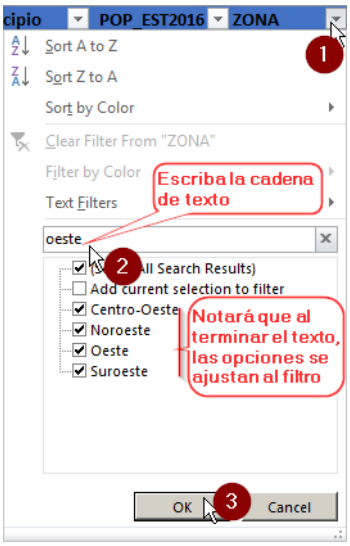
Podrá ver el **nombre de la tabla**. Además, verá que se añadieron los **botones de filtrado** y **ordenamiento** (*sorting*) a cada columna.

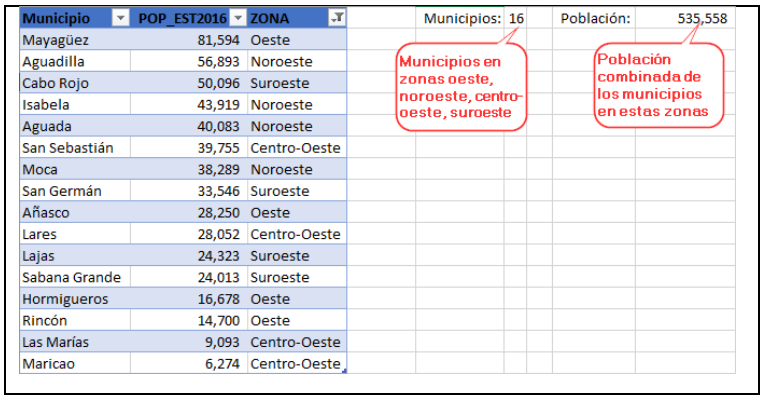
Afuera de esta tabla, he puesto dos medidas derivadas de dos variantes de una función agregada (AGGREGATE) que discutiremos un poco más adelante. Al filtrar los récords mediante el criterio que usted decida, estas medidas cambiarán: número de municipios y suma de poblaciones municipales.



	B	C	D	E	F	G	H	I	J
1	Municipio	POP_EST2016	ZONA		Municipios:	78		Población:	3,529,385
2	San Juan	363,744	Metro						
3	Bayamón	192,974	Metro						

Por ejemplo, filtrar por zona: Zonas que contengan “oeste”





Municipio	POP_EST2016	ZONA		Municipios:	16		Población:	535,558
Mayagüez	81,594	Oeste						
Aguadilla	56,893	Noroeste						
Cabo Rojo	50,096	Suroeste						
Isabela	43,919	Noroeste						
Aguada	40,083	Noroeste						
San Sebastián	39,755	Centro-Oeste						
Moca	38,289	Noroeste						
San Germán	33,546	Suroeste						
Añasco	28,250	Oeste						
Lares	28,052	Centro-Oeste						
Lajas	24,323	Suroeste						
Sabana Grande	24,013	Suroeste						
Hormigueros	16,678	Oeste						
Rincón	14,700	Oeste						
Las Marías	9,093	Centro-Oeste						
Maricao	6,274	Centro-Oeste						

ALGUNOS CONCEPTOS ANTES DE DISCUTIR SOBRE FUNCIONES

En esta sección, discutiremos algunas funciones que nos pueden ayudar en las labores diarias o para la preparación de informes que requieran expresar resúmenes por selecciones y categorías.

OBJETIVOS

Luego de terminar esta sección, usted conocerá algunos conceptos importantes sobre **nombrar grupos de celdas** (*named ranges*, y *named columns*), **referencias relativas y absolutas a celdas, filas y columnas**, además de una serie de **funciones** tales como **AGGREGATE**, **VLOOKUP**, **INDEX(MATCH)**, **COUNT**, **COUNTA**, **COUNTIFS**, **SUMIFS**, y funciones con fechas.

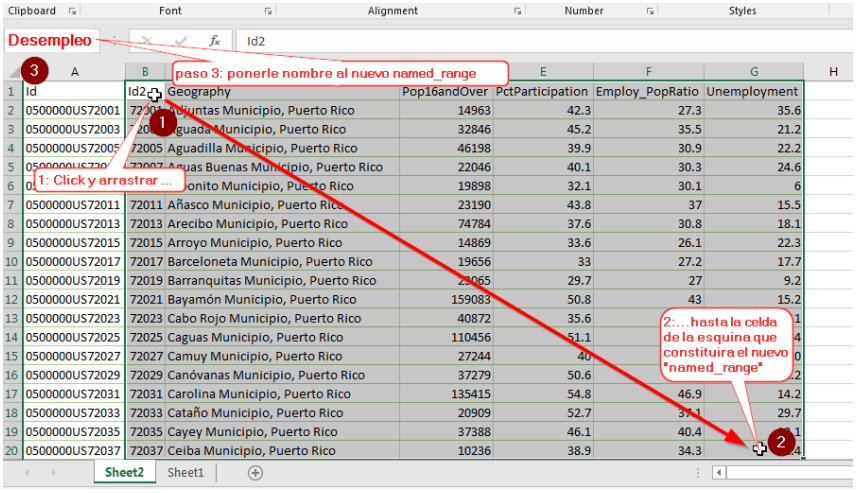
NAMED RANGES: NOMBRAR GRUPOS DE CELDAS Y COLUMNAS

En ocasiones es práctico seleccionar un subgrupo de celdas y ponerles un nombre. A esto se le conoce como *named range* y puede usarse para evitar estar sombreando frecuentemente un grupo o zona de celdas (data range) para diferentes tareas.

En ocasiones, tenemos un data range con varias columnas y solo queremos las primeras o últimas columnas, sin tener que esconderlas. Por ejemplo, la función VLOOKUP requiere que la primera columna tenga unos códigos para identificar y parear, pero en nuestro data range no está en la primera columna. En este caso, podemos asignarle nombre a un subconjunto dentro del data range de referencia que incluya como primera columna a esta columna que tiene los códigos.

CÓMO HACER UN NAMED RANGE

Seleccionar el **data range** usando **click** y **arrastrar hacia** la **celda final** hasta cubrir toda la zona de interés.

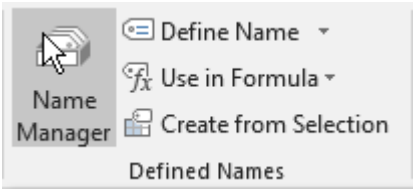


Luego **escribir el nombre** del nuevo **data range** en la **caja de texto** donde aparece el nombre de la celda. En este caso era B1.

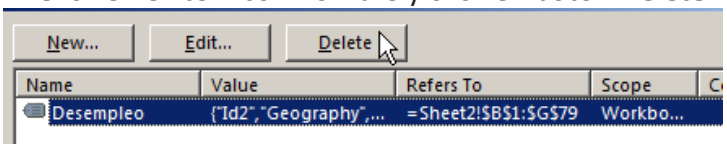
Borrar o redefinir named ranges

Para **borrar un named range**, vaya al **menú principal** en el tab **Formulas** y haga **click** en el botón **Name Manager**, localizado en el grupo **Defined Names**.

1: Click en **Name Manager**



2: Click en el ítem con nombre y **click** en botón **Delete**



Referencias relativas y absolutas

Referencias relativas

En Excel nos podemos referir a celdas de manera relativa o absoluta. Por defecto, las referencias son relativas. Las referencias relativas a celdas cambian cuando una misma fórmula se aplica a otra celda. Esto facilita el cómputo en ristra, tomando como partida a la celda siguiente para aplicar la fórmula. A1+B1, SUM(A1:R1), etc. son ejemplos de referencias relativas.

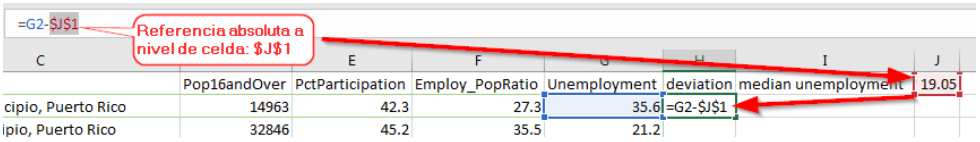
Referencias absolutas:

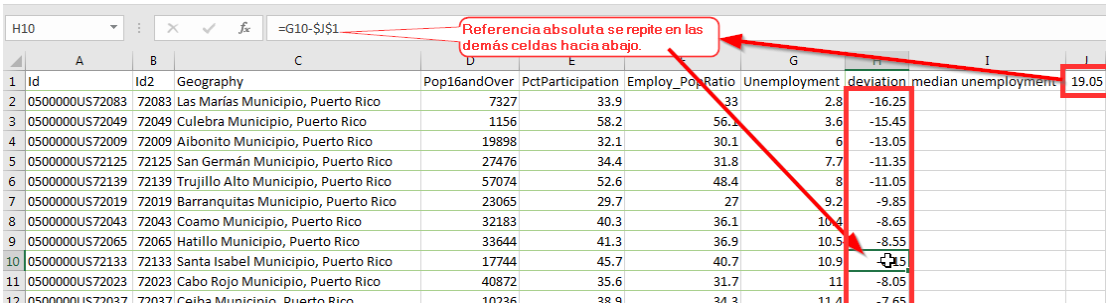
En ocasiones, solo queremos mantener la celda de origen sin cambiar su referencia, y aplicarle una fórmula a las demás que deseemos calcular. Podemos mantener referencia fija (absoluta) a una celda, a una columna o a una fila. Estas las podemos identificar mediante el signo de dólar $.

Ejemplos de referencias absolutas:

Solo a una celda	A1	La referencia a la celda (fila y columna) se mantiene fija
Solo a una fila	A$1	La referencia a la fila (row) no cambia al copiar fórmulas
Solo a una columna	$A1	La referencia a la columna no cambia al copiar fórmulas

Ejemplo de referencia a una celda (fila y columna fija)





	A	B	C	D	E	F	G	H	I	J
1	Id	Id2	Geography	Pop16andOver	PctParticipation	Employ_PopRatio	Unemployment	deviation	median unemployment	19.05
2	0500000US72083	72083	Las Marías Municipio, Puerto Rico	7327	33.9	33	2.8	-16.25		
3	0500000US72049	72049	Culebra Municipio, Puerto Rico	1156	58.2	56.1	3.6	-15.45		
4	0500000US72009	72009	Aibonito Municipio, Puerto Rico	19898	32.1	30.1	6	-13.05		
5	0500000US72125	72125	San Germán Municipio, Puerto Rico	27476	34.4	31.8	7.7	-11.35		
6	0500000US72139	72139	Trujillo Alto Municipio, Puerto Rico	57074	52.6	48.4	8	-11.05		
7	0500000US72019	72019	Barranquitas Municipio, Puerto Rico	23065	29.7	27	9.2	-9.85		
8	0500000US72043	72043	Coamo Municipio, Puerto Rico	32183	40.3	36.1	10.4	-8.65		
9	0500000US72065	72065	Hatillo Municipio, Puerto Rico	33644	41.3	36.9	10.5	-8.55		
10	0500000US72133	72133	Santa Isabel Municipio, Puerto Rico	17744	45.7	40.7	10.9	-8.15		
11	0500000US72023	72023	Cabo Rojo Municipio, Puerto Rico	40872	35.6	31.7	11	-8.05		
12	0500000US72037	72037	Ceiba Municipio, Puerto Rico	10236	38.9	34.3	11.4	-7.65		

Códigos de errores en Excel y qué significan

#DIV/0!	**División por cero**
#VALUE!	**Inclusión de valor incorrecto en un argumento en una fórmula**. Por ejemplo, entrar un texto en un argumento de una función cuando Excel espera un número.
#NAME?	**Nombre incorrecto de función**, **data_range** o **referencia a una celda**.
#REF!	**Referencia inexistente**. Por ejemplo, una fórmula escrita D2+E2 y luego borrar la columna D.
#NUM!	**Hay un problema con un número en una fórmula o una función**. Aparece cuando hay algún número no manejable por Excel (out of range). Puede ocurrir cuando Excel espera un número en una función y escribimos texto.
#N/A	Generalmente aparece cuando **Excel no encuentra algún valor en una lista**. Puede incluir otros errores.
#######	**No es error**. Significa que el tamaño (visual) de la celda no es suficiente para acomodar el valor que contiene. Estire la celda o haga doble click en el borde derecho de la columna para poder ver el valor completo.

Funciones

En esta parte, discutiremos algunas funciones de interés de Excel. La mayoría tienen que ver con resumir datos y manejo de datos temporales (fechas y horas).

AGGREGATE

Una función útil para resumir datos que expresen valores tomados de listas (arrays). Estas pueden devolver valores de toda la lista o solamente de récords filtrados, ignorando celdas escondidas, celdas con errores y celdas que contengan subtotales

Sintaxis:

AGGREGATE(function_num, options, array, [k])

function_num	Número asignado a la función a utilizarse. La lista de funciones aparece a continuación en la tabla de funciones disponibles en AGGREGATE
options	Número asignado a la opción a usarse para contar o no ciertas celdas según su condición. La lista de opciones aparecerá a continuación.
array	Un array es una lista. Esta puede ser numérica o alfanumérica, el array puede ser compuesto por una columna, una referencia a un grupo de celdas (range) o a una función que devuelva un array.
k	Opcional, dependiendo del tipo de función escogida. En algunos casos, las funciones requieren un argumento adicional. Estas aparecerán en una tabla aparte más adelante.

Lista de funciones disponibles en AGGREGATE(**function_num**, options, array, [k])

1	AVERAGE	Promedio, *media*. Medida de tendencia central.
2	COUNT	Contar número celdas
3	COUNTA	Contar número de celdas que no estén vacías. Esto no incluye celdas que tengan una cadena alfanumérica vacía "". COUNTA evalúa *(cuenta)* celdas con error
4	MAX	Devuelve el valor más alto en un conjunto
5	MIN	Devuelve el valor menor en un conjunto
6	PRODUCT	Usa el contenido de celdas aisladas o en grupos como argumentos y los multiplica
7	STDEV.S	Desviación (error) estándar de la muestra
8	STDEV.P	Desviación (error) estándar de la población
9	SUM	Sumar valores individuales, referencias a celdas, data ranges o todas las anteriores
10	VAR.S	Varianza de la muestra (error estándar al cuadrado)
11	VAR.P	Varianza de la población (error estándar al cuadrado)
12	MEDIAN	Valor central de una distribución ordenada de números
13	MODE.SNGL	Moda. Valor que más se repite
14	LARGE	Devuelve el valor n-ésimo más alto. Primer valor más alto, tercer, etc.
15	SMALL	Devuelve el valor n-ésimo más bajo. Primer valor más alto, tercer, etc.
16	PERCENTILE.INC	Valor que divide un conjunto ordenado de datos estadísticos de forma que un porcentaje de tales datos sea inferior a dicho valor. Un individuo en el percentil 80 está por encima del 80 % del grupo a que pertenece. (según RAE)
17	QUARTILE.INC	Devuelve el cuartil, que es cualquiera de los valores de percentil 25, 50 o 75. Los parámetros son 0,1,2,3,4. MIN, MAX, MEDIAN, devuelven el resultado de los parámetros *quartile* 0, 4 y 2 respectivamente.
18	PERCENTILE.EXC	Igual que percentile.inc excepto que no incluye valores que esté en el límite de la clase.
19	QUARTILE.EXC	Igual que quartile.inc pero excluyente.

Opciones de AGGREGATE(function_num, options, array,[k])

0 or omitted	Ignorar subtotales anidados y funciones AGGREGATE
1	Ignorar celdas escondidas (*hidden*), subtotales anidados y funciones AGGREGATE
2	Ignorar celdas con valor ERROR, subtotales anidados y funciones AGGREGATE
3	Ignorar celdas escondidas, celdas con ERROR, subtotales anidados y funciones AGGREGATE
4	No ignorar nada
5	Ignorar solo celdas escondidas
6	Ignorar celdas con valor ERROR
7	Ignorar celdas escondidas y celdas con valor ERROR

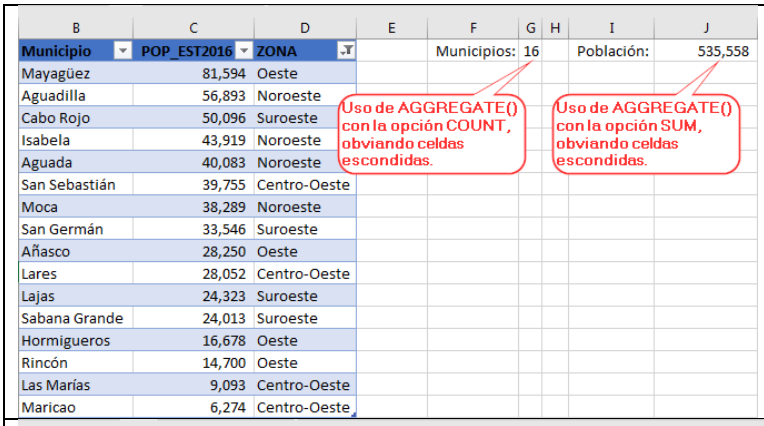
Funciones que usan el parámetro k

LARGE(array, k)
SMALL(array, k)
PERCENTILE.INC y PERCENTILE.EXC
QUARTILE.INC y QUARTILE.EXC

Veamos un ejemplo de AGGREGATE(). En el workbook **municipios_población.xlsx**, hay dos *medidas* (números) *dinámicas*. Es decir: **cambian según se apliquen filtros**. Una celda devuelve el número de municipios usando **AGGREGATE con la opción COUNT** y la otra celda usa **AGGREGATE con la opción SUM** para la suma de población de los municipios que estén en la lista. En ambas celdas se usa esta opción para que la función AGGREGATE no evalúe celdas escondidas.

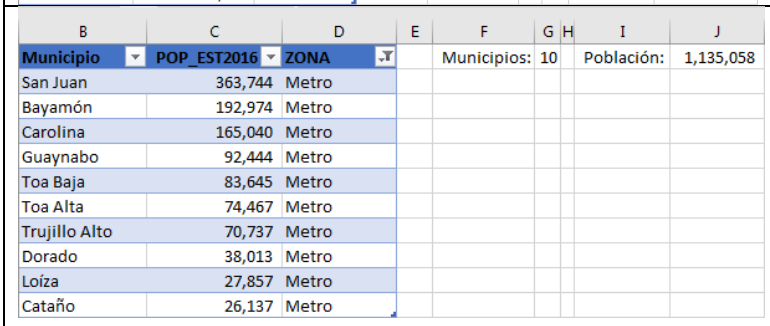
En estos casos, podemos ver que la columna **Zona** tiene un filtro aplicado

con todas las celdas que contengan la secuencia alfanumérica "*oeste*". Ambas funciones AGGREGATE usan la **opción 5** para **ignorar celdas escondidas**.



Municipio	POP_EST2016	ZONA
Mayagüez	81,594	Oeste
Aguadilla	56,893	Noroeste
Cabo Rojo	50,096	Suroeste
Isabela	43,919	Noroeste
Aguada	40,083	Noroeste
San Sebastián	39,755	Centro-Oeste
Moca	38,289	Noroeste
San Germán	33,546	Suroeste
Añasco	28,250	Oeste
Lares	28,052	Centro-Oeste
Lajas	24,323	Suroeste
Sabana Grande	24,013	Suroeste
Hormigueros	16,678	Oeste
Rincón	14,700	Oeste
Las Marías	9,093	Centro-Oeste
Maricao	6,274	Centro-Oeste

El filtro aplicado devuelve las zonas *Oeste*, *Noroeste*, *Suroeste* y *Centro-Oeste*.
Hay **16** municipios en estas zonas y suman **535,558** habitantes.

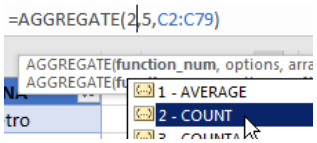


Municipio	POP_EST2016	ZONA
San Juan	363,744	Metro
Bayamón	192,974	Metro
Carolina	165,040	Metro
Guaynabo	92,444	Metro
Toa Baja	83,645	Metro
Toa Alta	74,467	Metro
Trujillo Alto	70,737	Metro
Dorado	38,013	Metro
Loíza	27,857	Metro
Cataño	26,137	Metro

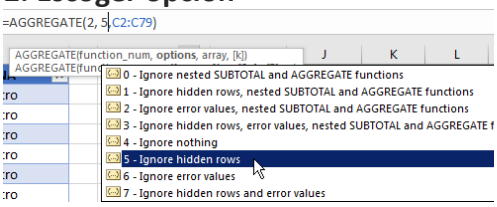
Otro ejemplo:
Hay **10** municipios y **1,135,058** habitantes **en** la **Zona Metro**

- Conteo de municipios: =AGGREGATE(2,5,C2:C79)
- Alternativa: =AGGREGATE(3,5,B2:B79) usando función **COUNTA** con *nombres* de municipios
- Sumatoria de Población: =AGGREGATE(9,5,C2:C79)

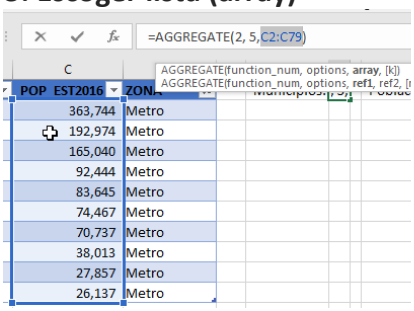
1: Escoger función



2: Escoger opción



3: Escoger lista (array)



LOOKUP

Esta función es útil para clasificar una lista de valores y clasificarlos. A diferencia de la función VLOOKUP, la función LOOKUP trabaja con una sola columna. Por ejemplo, tener una columna con valores desde 0 a 1,000. Para propósitos de agrupar, tomamos los valores y los convertimos en clases y creamos diez grupos de cien: *0-100; 100-200; 200-300; 300-400; 400-500; 500-600; 600-700; 700-800; 800-900; 900-1,000*.

LOOKUP va celda por celda y evalúa/compara el valor de la celda de inicio, con el valor del lookup_vector a asignarse en la otra celda o celdas en una columna. Puede usarse una lista existente en celdas o crear un arreglo (array).

SINTAXIS

LOOKUP(lookup_value, lookup_vector, [result_vector])

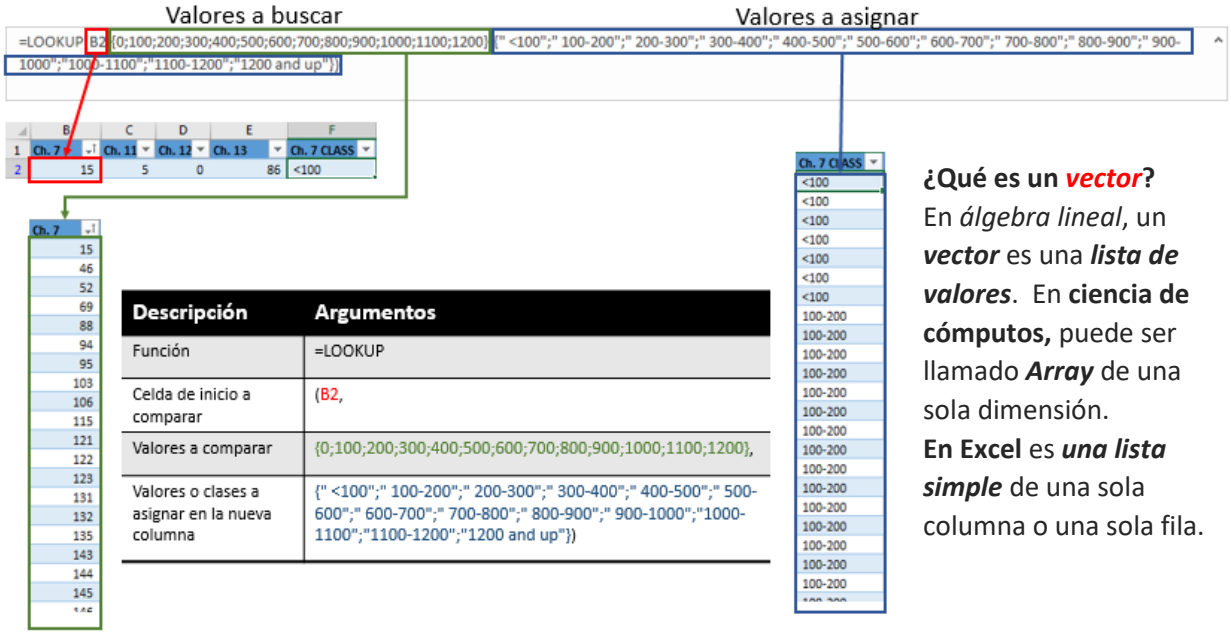
lookup_value	Celda que contiene el valor a buscar y parear.
lookup_vector	Grupo de celdas o una lista contenida en una sola columna o una sola fila. Los valores deben ser escritos en orden ascendente. De lo contrario, los resultados no serán confiables.
[result_vector]	Una sola columna o fila que contiene los valores a ser devueltos. Para cada valor de lookup_vector debe corresponder un valor en el result_vector.

Por ejemplo:

=LOOKUP(B2,

{0;100;200;300;400;500;600;700;800;900;1000;1100;1200}, *(ESTE ES EL LOOKUP VECTOR)*

{" <100";" 100-200";" 200-300";" 300-400";" 400-500";" 500-600";" 600-700";" 700-800";" 800-900";" 900-1000";"1000-1100";"1100-1200";"1200 and up"}) *(ESTE ES EL "RESULT" VECTOR)*



Descripción	Argumentos
Función	=LOOKUP
Celda de inicio a comparar	(B2,
Valores a comparar	{0;100;200;300;400;500;600;700;800;900;1000;1100;1200},
Valores o clases a asignar en la nueva columna	{" <100";" 100-200";" 200-300";" 300-400";" 400-500";" 500-600";" 600-700";" 700-800";" 800-900";" 900-1000";"1000-1100";"1100-1200";"1200 and up"})

¿Qué es un *vector*?
En *álgebra lineal*, un ***vector*** es una ***lista de valores***. En **ciencia de cómputos,** puede ser llamado ***Array*** de una sola dimensión.
En Excel es ***una lista simple*** de una sola columna o una sola fila.

VLOOKUP

Esta función se usa para **parear** (*lookup* = buscar) los **valores de la primera columna de una lista** (array) moviéndose hacia abajo (*vertical*) a través de las filas **y devolver el valor encontrado en otra celda o celdas**. VLOOKUP y HLOOKUP son parte del grupo de funciones de búsqueda y referencia (lookup and reference).

SINTAXIS:

VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

Según la documentación de Microsoft sobre VLOOKUP, la forma más simple de esta función es:

=VLOOKUP(*valor de referencia que desea buscar*, *grupo de celdas de donde se busca este valor*, el *número de la columna en el grupo de celdas que contienen el valor a devolver*, *pareo exacto o aproximado* indicado como 0/FALSE o 1/TRUE).

Recordar: Un requerimiento de VLOOKUP es que el *lookup_value* debe estar **en la primera celda** de la **primera columna**. No puede estar en otra columna, a menos que se designe de antemano, un *named range* que incluya la columna de referencia como columna #1. Los valores en el *table_array* deberán estar ordenados (sort).

Por ejemplo:

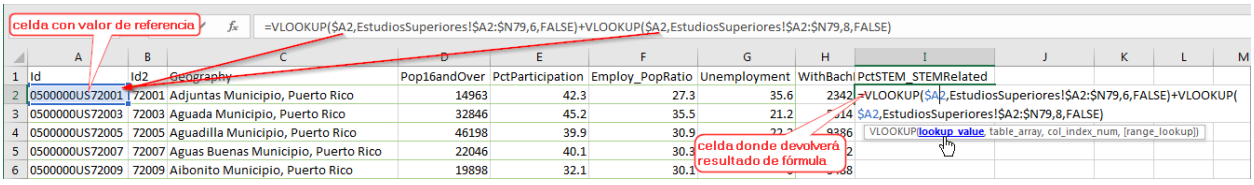
```
=VLOOKUP($A2,EstudiosSuperiores!$A2:$N79,6,FALSE)
```

lookup_value: *$A2*	Celda que contiene el valor a buscar y parear. Note que esta *referencia a celdas es fija en la columna*, como se discutió anteriormente.
table_array: *EstudiosSuperiores!$A2:$N79*	En la hoja *EstudiosSuperiores*, el grupo de celdas va desde la *$A2:$N79*. Note que esta referencia a celdas es fija *en las columnas* A a la N.
col_index: *6*	Número de la columna cuyos valores queremos transferir
range_lookup: *FALSE*	Requerimos pareo exacto.

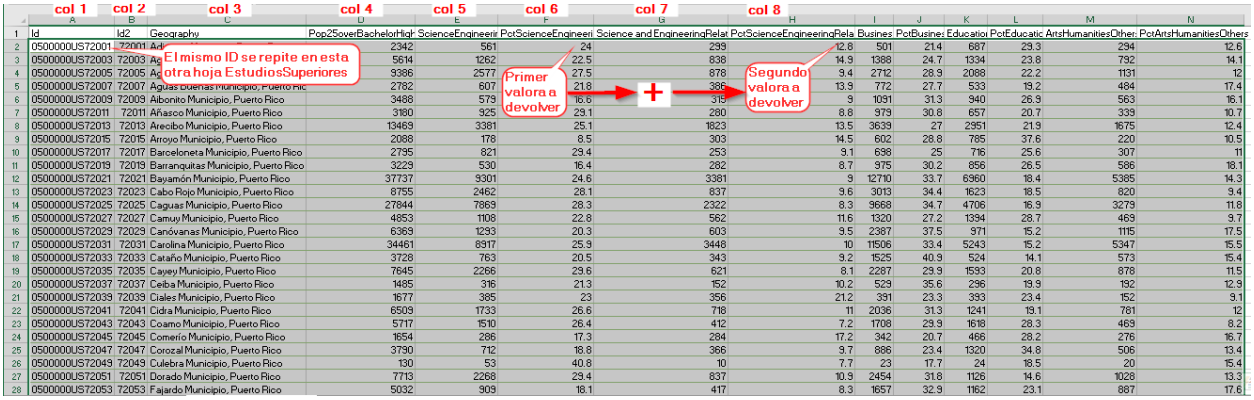
En este caso, queremos sumar dos columnas usando dos LOOKUPS:

```
=VLOOKUP($A2,EstudiosSuperiores!$A2:$N79,6,FALSE)+VLOOKUP($A2,EstudiosSuperiores!$A2:$N79,8,FALSE)
```

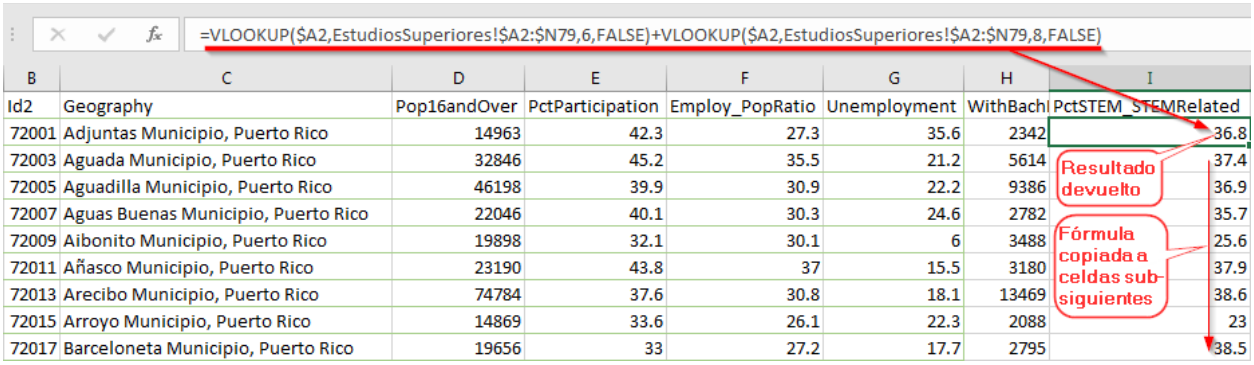
Celda con valor de referencia.



De este data range, tomar el valor de la columna 6 y sumarlo con el valor de la columna 8



Resultado devuelto:



Id2	Geography	Pop16andOver	PctParticipation	Employ_PopRatio	Unemployment	WithBachl	PctSTEM_STEMRelated
72001	Adjuntas Municipio, Puerto Rico	14963	42.3	27.3	35.6	2342	36.8
72003	Aguada Municipio, Puerto Rico	32846	45.2	35.5	21.2	5614	37.4
72005	Aguadilla Municipio, Puerto Rico	46198	39.9	30.9	22.2	9386	36.9
72007	Aguas Buenas Municipio, Puerto Rico	22046	40.1	30.3	24.6	2782	35.7
72009	Aibonito Municipio, Puerto Rico	19898	32.1	30.1	6	3488	25.6
72011	Añasco Municipio, Puerto Rico	23190	43.8	37	15.5	3180	37.9
72013	Arecibo Municipio, Puerto Rico	74784	37.6	30.8	18.1	13469	38.6
72015	Arroyo Municipio, Puerto Rico	14869	33.6	26.1	22.3	2088	23
72017	Barceloneta Municipio, Puerto Rico	19656	33	27.2	17.7	2795	38.5

INDEX & MATCH

Por diseño, la función **VLOOKUP requiere** que las **celdas de referencia** estén **en la primera columna**. A veces esto puede ser un inconveniente (a menos que se cree un *named_range* con ese fin). La combinación de funciones INDEX y MATCH pueden usarse como sustituto de VLOOKUP sin la obligación de que el *lookup_value* esté en la primera columna.

INDEX

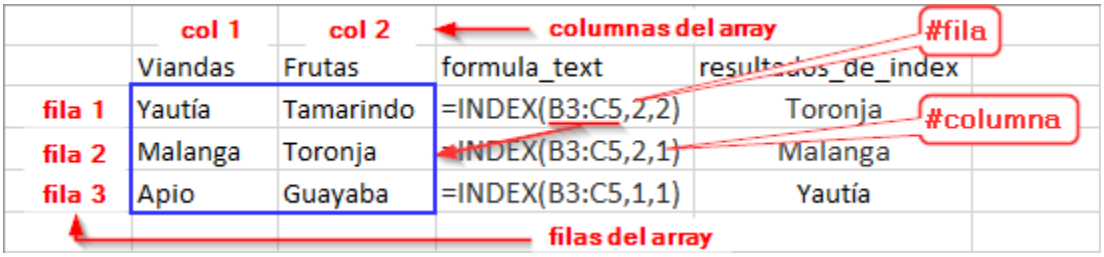
La función INDEX devuelve un valor o la referencia de un valor dentro de una tabla o un data_range.

Sintaxis

INDEX(array, row_num, [column_num])

array	Grupo de celdas en las cuales se hará la búsqueda. Puede ser un named_range, o un grupo de celdas data range sin nombre. Es preferible usar named ranges.
row_num	Fila con el valor de referencia
column_num	Número de la columna que contiene los datos que queremos copiar. Pueden usarse referencias a las columnas usando la función COLUMN().

Por ejemplo:



(la caja con borde azul es un array)

Viandas	Frutas	formula_text	resultados_de_index
Yautía	Tamarindo	=INDEX(B3:C5,2,2)	Toronja
Malanga	Toronja	=INDEX(B3:C5,2,1)	Malanga
Apio	Guayaba	=INDEX(B3:C5,1,1)	Yautía

En casos de que el array sea de una sola dimensión (de una columna o una fila), el segundo argumento es opcional.

MATCH

La función MATCH busca un ítem en un grupo de celdas (data range) y devuelve la posición relativa de ese ítem dentro de ese grupo de celdas.

Sintaxis

MATCH(lookup_value, lookup_array, [match_type])

lookup_value	Requerido. Es el valor para buscar en el *array de búsqueda* (lookup array). El valor de búsqueda (*lookup_value*) puede *ser un valor*, ya sea número, texto, valor lógico, o *ser una referencia* de celda a un valor numérico, texto, o lógico.
lookup_array	Requerido. Grupo de celdas objeto de la búsqueda.
match_type	Opcional. Escoger entre los números -1, 0, 1. Este último argumento es para especificar cómo Excel va a parear el *lookup_value* contra los valores en el lookup_array. El valor por defecto es 1. Optional. The number -1, 0, or 1. The match_type argument specifies how Excel matches *lookup_value* with values in lookup_array. The default value for this argument is 1.

Tipo de pareo

1 u omitido	MATCH buscará el valor más grande que sea igual o menor que el *lookup_value*. NOTA: En esta opción, los valores en el lookup_array tienen que estar en orden ascendente, por ejemplo: ...-2, -1, 0, 1, 2, ..., A-Z, FALSE, TRUE
0	MATCH encontrará el primer valor que sea exactamente igual al *lookup_value*. NOTA: En esta opción, los valores en el lookup_array pueden estar en cualquier orden (o desorden).
-1	MATCH encontrará el valor más pequeño que sea mayor o igual al *lookup_value*. NOTA: Los valores en el lookup array deben estar en orden descendente TRUE, FALSE, Z-A, ...2, 1, 0, -1, -2, ..., etcétera.

Notas

- MATCH devuelve *la posición* de la celda donde está el valor encontrado dentro del lookup_array, no el valor como tal.
- Por ejemplo, MATCH("pelao",{"más","pelao","que","un","chucho"},0)devolverá 2. Esta es la posición relativa de la cadena de letras "pelao", dentro del arreglo {"más","pelao","que","un","chucho"}.
- En valores de texto, MATCH no hace distinción ente mayúsculas o minúsculas.
- Si MATCH no encuentra el valor, devolverá el valor de error #N/A.
- Si el match_type es 0 y el valor a buscar es alfanumérico, se pueden usar los comodines (wildcards) ? y * en el argumento *lookup_value*.
- ?: parea solo una pieza de una cadena alfanumérica.
- : parea cualquier secuencia dentro de una cadena alfanumérica.
- Use la tilde ~ antes de la cadena alfanumérica en caso de tener que buscar signos de pregunta o asteriscos

Uso de INDEX con MATCH

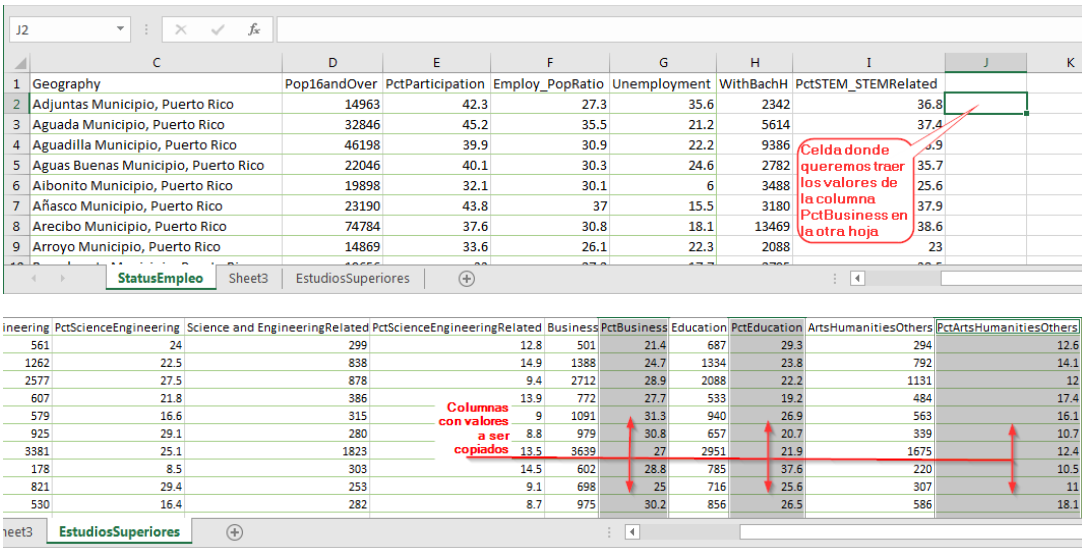
Como vimos, la función VLOOKUP requiere que el *lookup_value* esté en la primera columna del array (grupo de celdas escogidas, *named_range*, etcétera). Las funciones INDEX y MATCH juntas sirven de sustituto a la función VLOOKUP, ya que no tienen esta restricción.

En este ejemplo, usaremos estas las funciones **INDEX**, **MATCH** y añadiremos la función **COLUMN()** para referirnos a las columnas de las cuales queremos obtener valores.

Tenemos **dos hojas con datos censales** estimados de 2016 a cinco años, es decir, desde 2012 a 2016, descargados de la página http://factfinder.census.gov. Los datos vienen de tablas separadas: ***Employment Status***, Table **S2301** y la otra es la **S1502**: ***Field of bachelor's degree for first major***. Los datos están divididos **por municipio**.

El objetivo es integrar *en* la hoja *StatusEmpleo*, los porcentajes de personas con estudios superiores en grupos de disciplinas *Administración comercial*, *Educación y Artes-Humanidades*. La idea es no usar copy/paste. Esto puede traer problemas de errores y otras faltas. Cuando son muchos datos no es recomendable usar copy/paste.

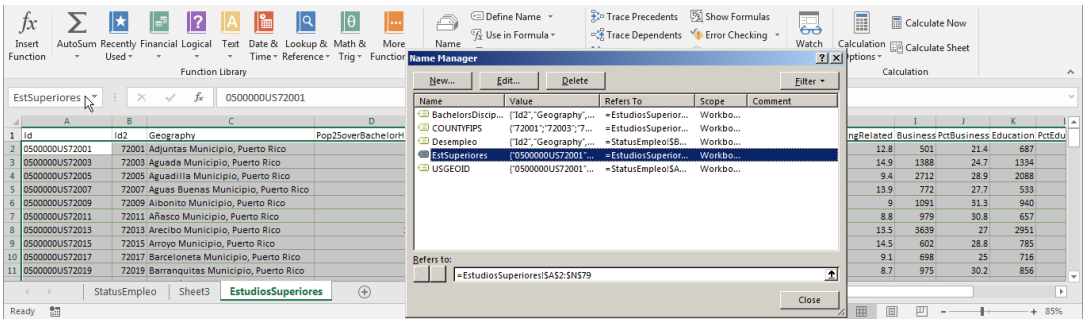
Para lograrlo, **se debe parear un récord del municipio con el que le corresponde en la otra hoja**. Disponemos de los códigos que suple el Censo para identificar los municipios. Así no tendremos que usar los nombres de los municipios para hacer el pareo. Es más práctico usar códigos identificadores en lugar de nombres.



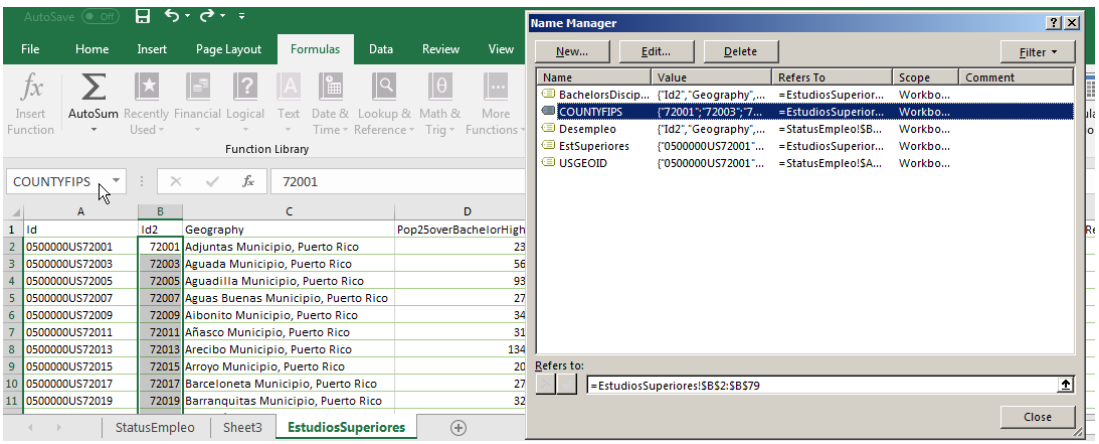
En la hoja **StatusEmpleo** se crearán tres columnas destinatarias: **PctBusiness**, **PctEducation**, **PctArts**… Copiaremos los nombres de las columnas y pasaremos a usar las funciones INDEX, MATCH y COLUMN para copiar estos valores…

Recomendaciones:

Utilizar named_range ***EstSuperiores*** para el array a usarse en la función INDEX:



Otro named_range ***COUNTYFIPS*** para obtener el row_number de la función MATCH



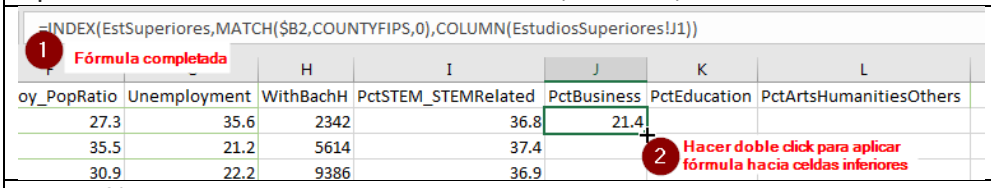
Note que la columna de pareo no es la primera del array, como requiere VLOOKUP.

La fórmula completa es:

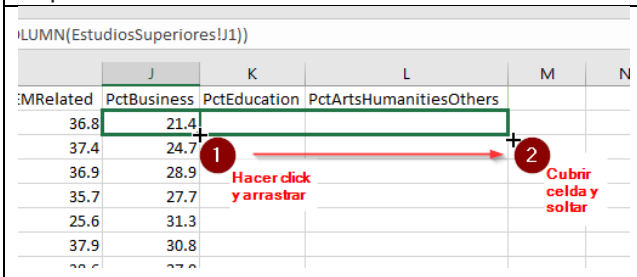
=INDEX(EstSuperiores,MATCH($B2,COUNTYFIPS,0),COLUMN(EstudiosSuperiores!J1))

INDEX	
EstSuperiores	Es el **named_range** que usaremos **para el array** requerido de la función INDEX. Este named_range está en la hoja origen
MATCH	
$B2	**Celda con el código ID a buscar**. Notar la referencia fija **$**B para que la búsqueda no cambie de columna. Esta celda está en la hoja destino.
COUNTYFIPS	Otro **named_range** que usaremos **para que devuelva el # de fila** row_number de la función MATCH. Este named_range está en la hoja origen.
0 (cero)	Pareo exacto
COLUMN(EstudiosSuperiores!J1)	Función usada **para devolver el número de la columna** que contiene los datos que queremos copiar. En este caso, la columna es J1 y está en la hoja origen

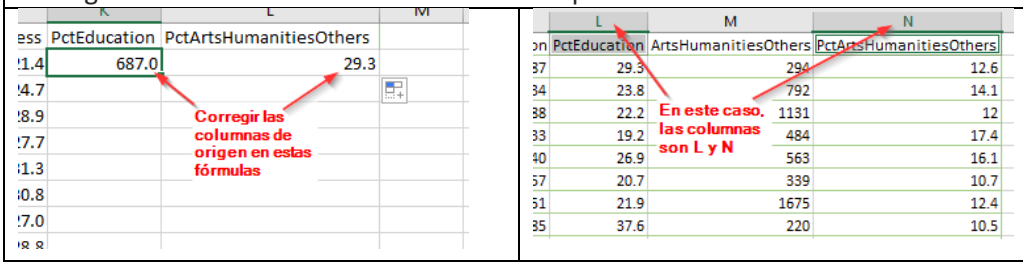
Aplicar la combinación de fórmulas INDEX, MATCH, COLUMN…

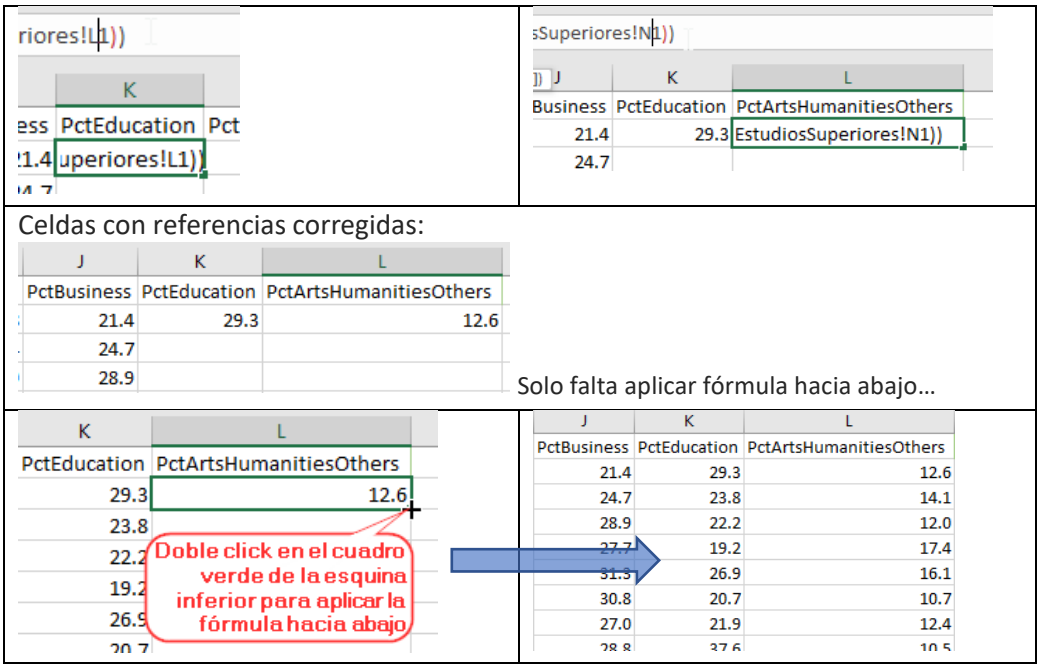


Copiar fórmula a las celdas a la derecha…



Corregir referencias a columnas erróneas al copiar…





COUNT, COUNTA, COUNTIF, COUNTIFS

Estas funciones son para conteo de celdas.

COUNT	Contar celdas que contengan números en un grupo de celdas.
Sintaxis:	COUNT(value1, [value2], ...)
	value1 puede ser un data range, celda, valor value2 es opcional e igualmente puede ser data range, celda, valor … hasta 255 argumentos (*value*)
COUNTA	Cuenta celdas que no estén vacías en un grupo de celdas.
Sintaxis:	COUNTA(value1, [value2], ...)
	value1 puede ser un data range, celda, valor value2 es opcional e igualmente puede ser data range, celda, valor … hasta 255 argumentos (*value*) COUNTA cuenta celdas que contengan *cualquier tipo de dato* incluyendo *error_values* y texto vacío ("")
COUNTIF	Conteo de celdas que cumplan una condición específica. Por ejemplo, cuántas escuelas hay en el Municipio de San Germán.
Sintaxis:	COUNTIF(range, criteria)
	Range: grupo de celdas donde se quiere buscar Criteria: qué es lo que se va a buscar. Añadir comillas ("") si se busca en cadenas alfanuméricas (texto). Puede añadir *wildcards* (*) asterisco en el argumento de búsqueda. Por ejemplo: =COUNTIF(F2:F1102,"*MUNOZ RIVERA*") busca en el data range F2:F1102, la cadena alfanumérica *MUNOZ RIVERA*. Esto incluye cualquier celda que contenga la secuencia MUNOZ RIVERA en cualquier parte de la celda.
COUNTIFS	Similar a COUNTIF. Incorpora más condiciones para seleccionar.
Sintaxis:	COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2]…)
	A diferencia de COUNTIF, COUNTIFS se usa para añadir múltiples criterios de selección. El resultado debe ser cierto para todas las condiciones incluidas en la fórmula.

SUMIFS

Sumar cantidades basado una o más condiciones.

SUMIFS	Similar a SUMIF. Incorpora más condiciones para seleccionar.	
Sintaxis:	SUMIFS(sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)	
	sum_range:	Grupo de celdas a sumar
	criteria_range1	Grupo de celdas a comparar con *la condición* (*criteria1*)
	criteria1	Condición para comparar con en el grupo de celdas *criteria_range1*.
	criteria_range2	Otro grupo de celdas a comparar con *la condición* (*criteria2*)
	criteria2	Condición para comparar con en el grupo de celdas *criteria_range2*

Ejemplos usando COUNTIFS y SUMIFS

A	B	C	D	E	F	G
Municipio	**POP_EST2016**	**ZONA**		**Formulas: COUNTIFS & SUMIFS**	**Resutado**	**Explicación**
Adjuntas	18,760	Centro		=COUNTIFS(C2:C79,"Metro",A2:A79,"Ca*")	2	el data range C2:C79 = "Metro" & A2:A79 comienza con Ca
Aguada	40,083	Noroeste		=SUMIFS(B2:B79,C2:C79,"Metro",A2:A79,"Ca*")	191,177	Sumar range B2:B79 si C2:C79 lee Metro y A2:A79 comienza con Ca
Aguadilla	56,893	Noroeste				
Aguas Buenas	27,323	Centro-Este				
Aibonito	24,438	Centro-Este				
Añasco	28,250	Oeste				
Arecibo	91,096	Norte				
Arroyo	18,778	Sur				
Barceloneta	24,699	Norte				
Barranquitas	29,548	Centro-Este				
Bayamón	192,974	Metro				
Cabo Rojo	50,096	Suroeste				
Caguas	136,372	Centro-Este				
Camuy	33,481	Norte				
Canóvanas	47,174	Noreste				
Carolina	165,040	Metro				
Cataño	26,137	Metro				

Primera fórmula: COUNTIFS(C2:C79,"Metro",A2:A79,"Ca*")

El data range C2:C79 contiene las zonas. El data range A2:A79 contiene los nombres de los municipios.

Condición	Explicación
C2:C79,"Metro"	En el data range **C2:C79** de **zonas**, seleccionar solamente los récords que pertenezcan a la zona "**Metro**" (hay 10 seleccionados)
A2:A79,"Ca*"	Que *además* entre esos 10 seleccionados, el data range **A2:A79**, de **nombres** de los **municipios** deben **comenzar en "Ca*"** (Carolina y Cataño)

Segunda fórmula =SUMIFS(B2:B79,C2:C79,"Metro",A2:A79,"Ca*")

sum_range:	Explicación
B2:B79	Data range que contiene números a sumar
Condiciones	
C2:C79,"Metro"	En el data range **C2:C79** de **zonas**, seleccionar solamente los récords que pertenezcan a la zona "**Metro**"
A2:A79,"Ca*"	Que *además* el data range **A2:A79**, de **nombres** de los **municipios** deben **comenzar en "Ca*"** (Carolina y Cataño)

Note que los datos alfanuméricos (texto) llevan comillas. Excel no devolverá error; puede devolver cero 0.

Estas fórmulas sirven para resumir datos. Recuerde que puede usar filtros y usar la función AGGREGATE para validar resultados de las funciones SUMIFS y COUNTIFS.

DATE FUNCTIONS: Funciones temporales, fechas y horas

Los datos temporales se registran como números

Las fechas en Excel no deben escribirse como texto. En realidad Excel las convierte y se registran internamente como números secuenciales. Esto quiere decir que **Excel usa un origen arbitrario: día # 1** y continúa **sumando días consecutivamente** a partir de este origen. Por ejemplo, en Excel, **el día # 1 es** el **1ro de enero de 1900**. Por otro lado, las **horas**, **minutos** y **segundos** se expresan como **fracciones**. De esta manera se pueden realizar en Excel las **operaciones aritméticas** (+-*/) **usando fechas y horas**.

Dos sistemas de fechas

Excel puede usar dos sistemas de fechas: uno de ellos tiene como fecha de origen el 1ro de enero de **1900**. Este sistema es el que se utiliza con el sistema operativo **Windows**. Por su parte las **Mac**, usan otro sistema con fecha de origen el 1ro de enero de **1904**. Ambos sistemas tienen números secuenciales diferentes para las fechas posteriores al 1º de enero de 1904. La diferencia es 1,462 días. Excel tiene una **opción** para **intercambiar** el uso de estos **sistemas**. En **Excel 2016: File | Options | Advanced | When calculating this workbook** | **Use 1904 date system**.

DATE FUNCTION

Esta es una función fundamental para el manejo de fechas. Se recomienda usar la función DATE para registrar fechas, ya que no todas las funciones de Excel reconocen las fechas escritas en texto.

Sintaxis

DATE(year,month,day)

year	Representa el año de la fecha. Excel interpreta este a partir del sistema de fecha de la computadora. Excel usa como partida el año 1900 por defecto.
	• Si escribimos años **entre 1900 y 9999** inclusive, ese será el año registrado por Excel.
	  ○ Por ejemplo, =DATE(2018,02,28) devolverá February 28, 2018.
	• Si escribimos **0 a 1899 inclusive** en el argumento *year*, Excel le suma ese número a 1900
	  ○ Por ejemplo, =DATE(68,2,29) devolverá **2/29/1968**
	  ○ =DATE(1899,2,29) devolverá **3/1/3799**
	• Si entramos un **año negativo o mayor de 9999**, Excel devolverá el código de error #NUM!.
	Recomendación: siempre escriba el año usando cuatro cifras: 2000, 2018, etcétera.

month	Representa **el mes de la fecha**: de **1 a 12**.
	• Si escribimos un **número mayor de 12**, Excel le añade el sobrante de 12 al primer mes del próximo año.
	  ○ Por ejemplo, =DATE(2018,18,28) devolverá Friday, June 28, 2019. Añade 6 meses comenzando en enero, 2019
	  ○ **Mes 0**: =DATE(2018,0,28) devolverá 12/28/2017
	  ○ **Mes negativo**: =DATE(2018,-1,28) devolverá 11/28/2017. Note que la fecha devuelta es noviembre 2017. Por lo tanto, no es restar un mes. El resultado será -1+(-n), donde -n es la cifra del mes negativo.

day	Representa **el día de la fecha**, de **1 a 31**.
	• Aplican las mismas reglas del argumento *month* para números cero y negativos
	  ○ Por ejemplo, =DATE(2018,1,31) devolverá 1/31/2018.
	  ○ =DATE(2018,1,32) devolverá 2/1/2018. Devuelve 1ro de febrero, aun cuando el mes se escribió 1.
	  ○ =DATE(2018,1,0) devolverá 12/31/2017. Devuelve 31 de diciembre de 2017
	  ○ =DATE(2018,1,-1) devolverá 12/30/2017. Note que devuelve 30 de diciembre de 2017

DATEDIF

Calcula el número de días, meses, años entre dos fechas. Esta función está escondida y existe todavía para tener compatibilidad con versiones pasadas. La documentación Microsoft advierte que esta función puede dar errores en algunas circunstancias, especialmente calculando con la opción MD: diferencia de días entre dos meses comparados.

Sintaxis

DATEDIF(start_date,end_date,unit)

DATEDIF(start_date,end_date,unit)	
start_date	Fecha inicial
end_date	Fecha final
unit	Unidad. El tipo de dato que deseamos encontrar.

unit:

Y	Número de años en el periodo
M	Número completo de meses en el periodo
D	Número de días en el periodo
YM	Diferencia en meses entre dos fechas. Con esta opción, los días y años son ignorados.
YD	Diferencia en días entre dos fechas. Con esta opción, los días y años son ignorados.

Ejemplos:

A	B	C	D	E	F
	start_date	end_date	Fórmula	Resultado	Descripción
	1/2/2015	1/1/2018	=DATEDIF(B3,C3,"Y")	2	Dos años completados entre la fecha inicial y la final (Falta un día para que sean 3 años)
	1/1/2008	1/1/2009	=DATEDIF(B4,C4,"D")	366	366 días entre la fecha inicial y final. 2008 es año bisiesto.
	6/1/2010	8/15/2012	=DATEDIF(B5,C5,"YD")	75	75 días entre junio 1 y agosto 15. Note que está ignorando los años de estas fechas.

DAYS

Devuelve el número de días entre dos fechas: final e inicio.

Sintaxis

DAYS(end_date,start_date)

Ejemplo

DAYS(end_date,start_date)				
	A	B	C	D
1	Fechas	Formula	Resultado	Explicación
2	12/31/2011	=DAYS("3/15/2011","2/1/2011")	42	Hay 42 días entre el 15 de marzo y el 1ro de febrero de 2011. Recuerde escribir las fechas con comillas para que Excel las reconozca como tales en las fórmulas.
3	1/1/2011	=DAYS(A2,A3)	364	Hay 364 días entre la fecha final A2 y la de inicio A3.

Notas:

Si las fechas no son escritas correctamente, Excel no las convertirá al formato interno secuencial numérico de fechas. Por lo tanto, los resultados serán erróneos.

NETWORKDAYS

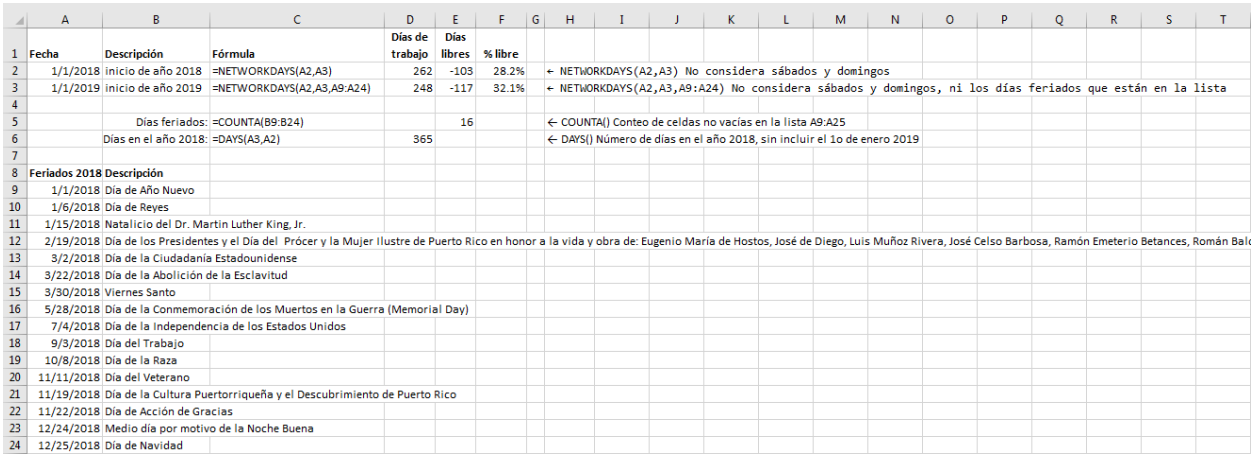
Devuelve el número de días entre dos fechas, descartando los días sábado y domingo. Opcional: Añadir lista (array) de celdas con fechas de días feriados.

Sintaxis

NETWORKDAYS(start_date, end_date, [holidays])

start_date	Fecha de inicio
end_date	Fecha final
holidays	Feriados: días libres que no sean sábado o domingo

Ejemplo: Número de días de trabajo y lista de días feriados oficiales para el año 2018



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	**Fecha**	**Descripción**	**Fórmula**	**Días de trabajo**	**Días libres**	**% libre**														
2	1/1/2018	inicio de año 2018	=NETWORKDAYS(A2,A3)	262	-103	28.2%		← NETWORKDAYS(A2,A3) No considera sábados y domingos												
3	1/1/2019	inicio de año 2019	=NETWORKDAYS(A2,A3,A9:A24)	248	-117	32.1%		← NETWORKDAYS(A2,A3,A9:A24) No considera sábados y domingos, ni los días feriados que están en la lista												
4																				
5		Días feriados:	=COUNTA(B9:B24)		16			← COUNTA() Conteo de celdas no vacías en la lista A9:A25												
6		Dias en el año 2018:	=DAYS(A3,A2)	365				← DAYS() Número de días en el año 2018, sin incluir el 1o de enero 2019												
7																				
8	**Feriados 2018**	**Descripción**																		
9	1/1/2018	Día de Año Nuevo																		
10	1/6/2018	Día de Reyes																		
11	1/15/2018	Natalicio del Dr. Martin Luther King, Jr.																		
12	2/19/2018	Día de los Presidentes y el Día del Prócer y la Mujer Ilustre de Puerto Rico en honor a la vida y obra de: Eugenio María de Hostos, José de Diego, Luis Muñoz Rivera, José Celso Barbosa, Ramón Emeterio Betances, Román Bal																		
13	3/2/2018	Día de la Ciudadanía Estadounidense																		
14	3/22/2018	Día de la Abolición de la Esclavitud																		
15	3/30/2018	Viernes Santo																		
16	5/28/2018	Día de la Conmemoración de los Muertos en la Guerra (Memorial Day)																		
17	7/4/2018	Día de la Independencia de los Estados Unidos																		
18	9/3/2018	Día del Trabajo																		
19	10/8/2018	Día de la Raza																		
20	11/11/2018	Día del Veterano																		
21	11/19/2018	Día de la Cultura Puertorriqueña y el Descubrimiento de Puerto Rico																		
22	11/22/2018	Día de Acción de Gracias																		
23	12/24/2018	Medio día por motivo de la Noche Buena																		
24	12/25/2018	Día de Navidad																		

En este ejemplo estamos usando la función NETWORKDAYS, a la cual le añadimos una lista de días feriados para 2018.

=NETWORKDAYS(A2,A3). Suma los lunes a viernes del año 2018. La fecha final es 1ro de enero de 2019 para que se pueda incluir el 31 de diciembre de 2018.

=NETWORKDAYS(A2,A3,A9:A24). Suma los lunes a viernes del año 2018. La lista A9:A24 contiene las fechas de los días feriados en Puerto Rico para el año 2018. (16 días). En realidad, son 15.5 días, considerando el medio día por motivo de Noche Buena.

RESUMIR DATOS

A medida que va pasando el tiempo, se acumulan cantidades de datos; a eso se le añade la creciente capacidad de guardar más datos que antes en las computadoras. Cuando la información es mucha, es preferible resumirla. El resumir nos da una visión general (*overview*) de las características del conjunto de datos en poco tiempo.

OBJETIVOS

- Explicar la importancia de resumir datos
- Explicar maneras apropiadas de resumir distintos tipos de datos
- Resumir *tablas* en Excel usando Total Row

PROPÓSITO DE RESUMIR DATOS

Resumir datos nos ayuda a contestar varias preguntas sobre un conjunto de datos. Por ejemplo:

- Conteo total
- Sumas totales o parciales,
- Medidas de tendencia central: media, mediana, moda,
- Medidas de dispersión: varianza, desviación estándar, coeficiente de variación
- Valores mínimo y máximo

En general, **los resúmenes nos ayudan** a **entender los datos** y **comunicar** el **significado** de ellos a otras personas.

Otras ventajas de resumir datos son:

- Validar trabajo hecho
- Probar hipótesis
- Conocer mejor los datos
- Hacer comparaciones

TOTAL ROW

Excel provee la opción de **añadir** una fila ***Total Row*** al final de una ***tabla***. Con esta opción, se usa la función compuesta **SUBTOTAL()** para añadir conteos, sumatorias, medidas de tendencia central, etcétera a cada columna de la tabla. Estas son las opciones disponibles para Total Row (**SUBTOTAL()**):

- No resumir
- AVERAGE
- COUNT
- COUNTA
- MAX
- MIN
- SUM
- STDEV
- VAR y otras…

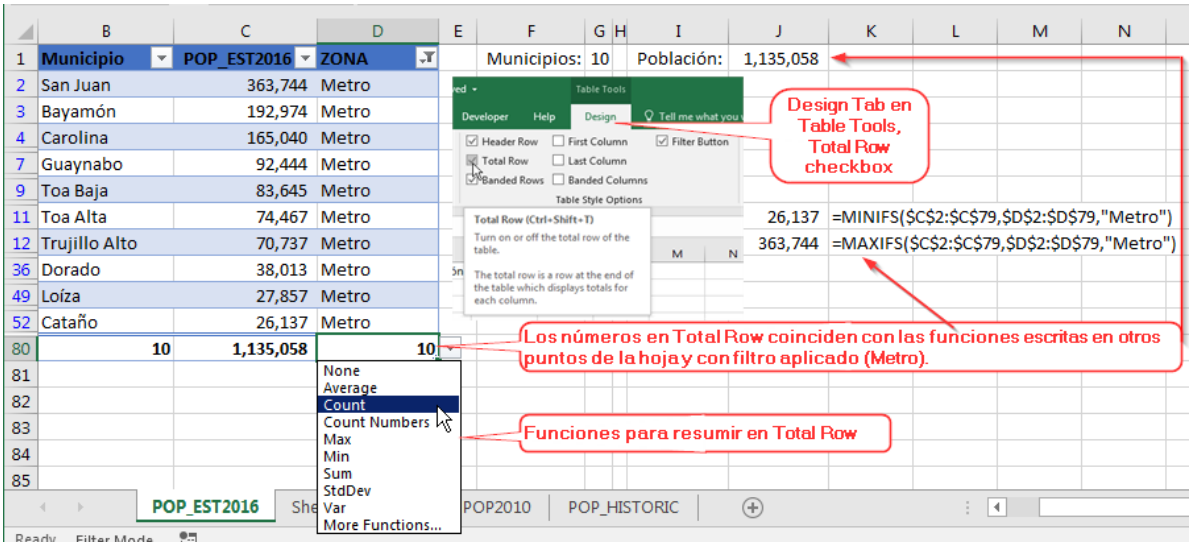
FUNCIONES INCLUIDAS EN COMANDO SUBTOTAL:

Sintaxis

SUBTOTAL(function_num, ref1,[ref2],...)

Esta función tiene los siguientes argumentos:

Función	*Número de función* **Incluye** celdas escondidas	*Número de función* **Ignora** celdas escondidas
AVERAGE	1	101
COUNT	2	102
COUNTA	3	103
MAX	4	104
MIN	5	105
PRODUCT	6	106
STDEV	7	107
STDEVP	8	108
SUM	9	109
VAR	10	110
VARP	11	111



Ejemplos usados en este caso:

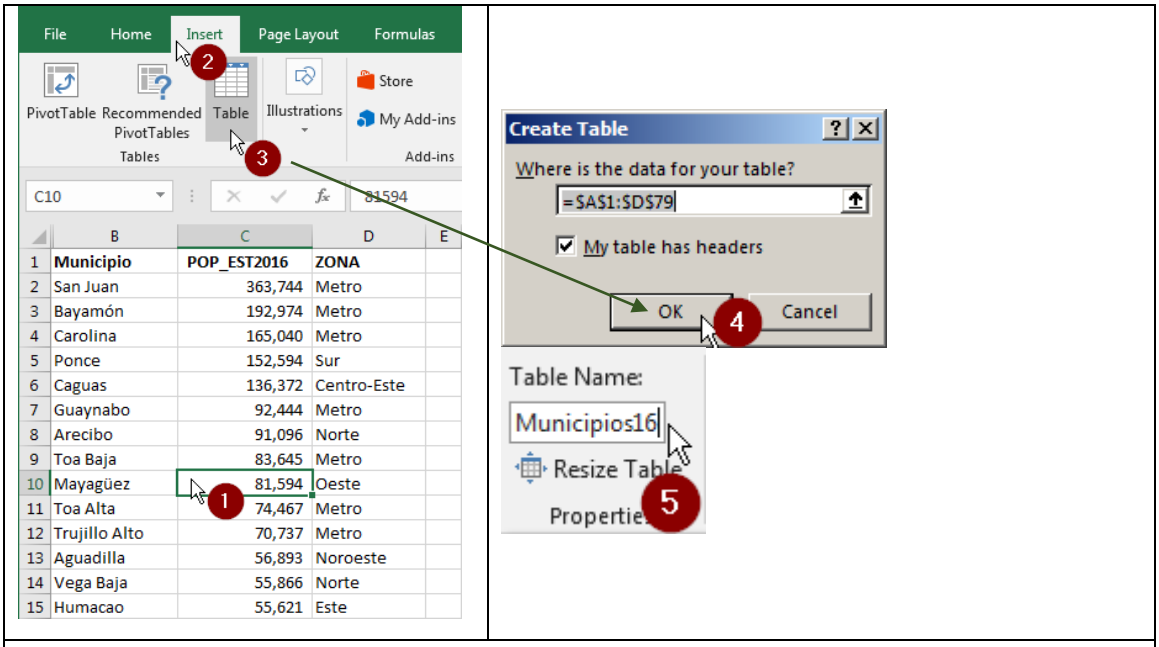
Fórmula	Función
=SUBTOTAL(103,[Municipio])	COUNTA
=SUBTOTAL(109,[POP_EST2016])	SUM

DEMO: AÑADIR FILA DE TOTALES

Para este ejemplo, usaremos el archivo **municipios_población.xlsx**, localizado en **C:\Tutorial_Excel\DataAnalysis\Modulo_2**.

Primero convertir el grupo de celdas *data_range* en una tabla Excel.

Click dentro del grupo de celdas a convertir, luego **Insert | Table**. Verificamos la extensión del data range desde A1 hasta D79 hacia abajo. Note las **referencias absolutas $** a las celdas.

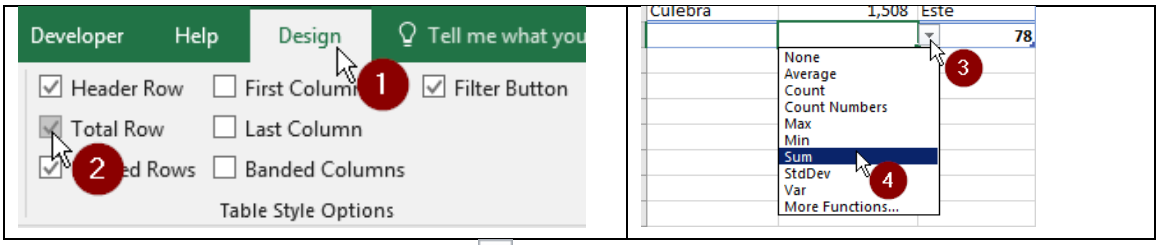


Municipio	POP_EST2016	ZONA
San Juan	363,744	Metro
Bayamón	192,974	Metro
Carolina	165,040	Metro
Ponce	152,594	Sur
Caguas	136,372	Centro-Este
Guaynabo	92,444	Metro
Arecibo	91,096	Norte
Toa Baja	83,645	Metro
Mayagüez	81,594	Oeste
Toa Alta	74,467	Metro
Trujillo Alto	70,737	Metro
Aguadilla	56,893	Noroeste
Vega Baja	55,866	Norte
Humacao	55,621	Este

Luego de aceptar la extensión de la tabla (4), escriba el nombre de la nueva tabla (5).

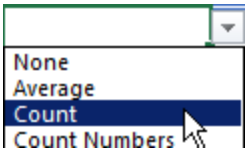
Segundo paso: añadir la fila de totales

Click en tab **Design**, luego hacer **check** en **Total Row**.

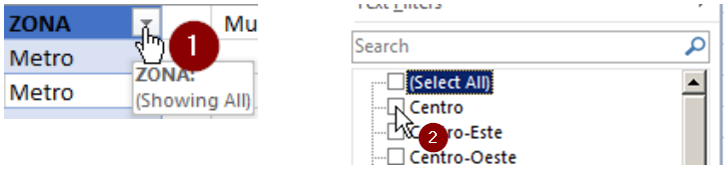


Luego hacer **click** en el **botón flecha** y escoja la función que aplique a los datos y la finalidad. En este caso, sumatoria. **Sum**

Repetimos el proceso en la **columna B** para añadir el **conteo de municipios**.



Si aplicamos filtro en la columna **ZONA**, podremos ver los subtotales. Por ejemplo, zona **Centro**.



Total Row con subtotales producto de un filtro.

	B	C	D	E
1	**Municipio**	**POP_EST2016**	**ZONA**	
38	Corozal	35,484	Centro	
42	Morovis	32,087	Centro	
43	Utuado	30,869	Centro	
46	Naranjito	29,467	Centro	
60	Orocovis	22,258	Centro	
66	Adjuntas	18,760	Centro	
69	Ciales	17,644	Centro	
71	Jayuya	15,605	Centro	
80	**8**	**202,174**	**8**	

PRÁCTICAS:

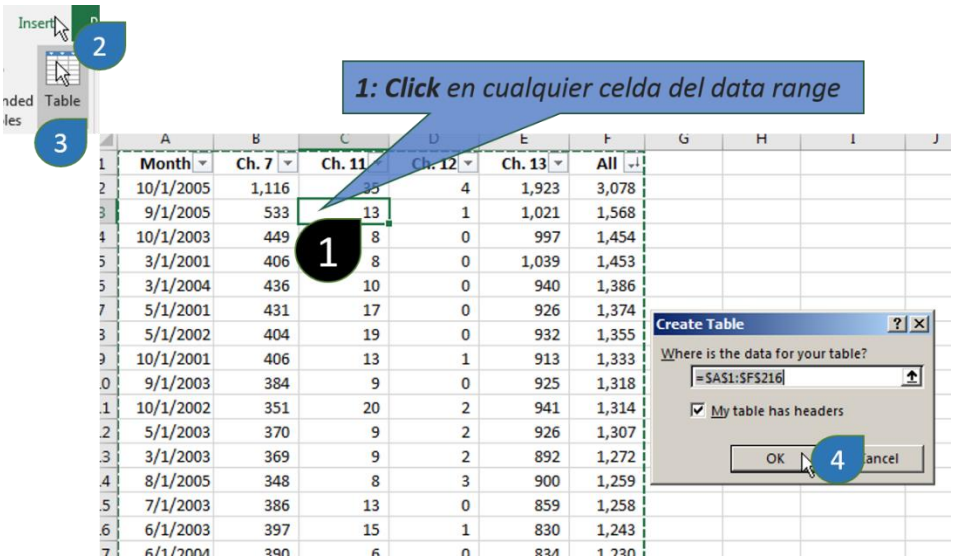
EJERCICIO 1: CREAR TABLAS

Objetivo: Derivar una tabla a partir de datos existentes en un data range.

Las tablas pueden generarse a partir de hojas de cálculo que ya tienen datos. En estas usted escoge el grupo de celdas *data range* y luego las convierte a una tabla Excel. La otra manera es mediante la importación de datos externos utilizando un grupo de instrucciones conocido como Query. Para esto Excel provee la interfaz Query Editor.

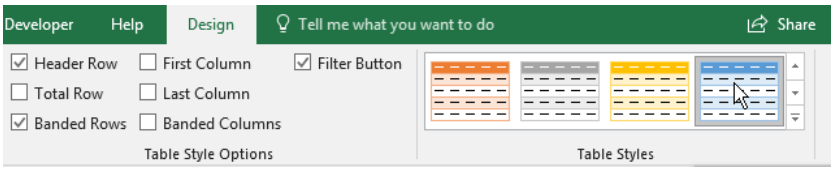
1: Para esta práctica, abra el archivo **bankruptcy.xlsx**, localizado en **C:\Tutorial_Excel\DataAnalysis\Modulo_2**. El archivo contiene los datos solicitudes de quiebras presentadas al tribunal federal. Los datos fueron descargados del portal de datos abiertos del Instituto de Estadísticas en febrero 15 de 2018.

2: Derive una tabla Excel a partir de los datos que están en la hoja de cálculo.
Estos son los pasos:

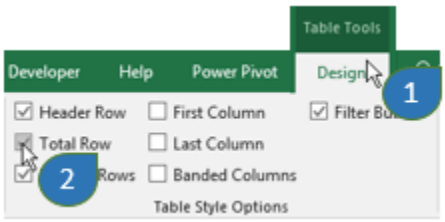


Month	Ch. 7	Ch. 11	Ch. 12	Ch. 13	All
10/1/2005	1,116	35	4	1,923	3,078
9/1/2005	533	13	1	1,021	1,568
10/1/2003	449	8	0	997	1,454
3/1/2001	406	8	0	1,039	1,453
3/1/2004	436	10	0	940	1,386
5/1/2001	431	17	0	926	1,374
5/1/2002	404	19	0	932	1,355
10/1/2001	406	13	1	913	1,333
9/1/2003	384	9	0	925	1,318
10/1/2002	351	20	2	941	1,314
5/1/2003	370	9	2	926	1,307
3/1/2003	369	9	2	892	1,272
8/1/2005	348	8	3	900	1,259
7/1/2003	386	13	0	859	1,258
6/1/2003	397	15	1	830	1,243

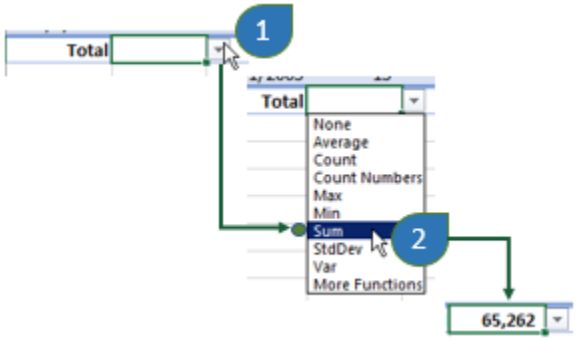
3: Seleccione diferentes estilos para visualizar la tabla



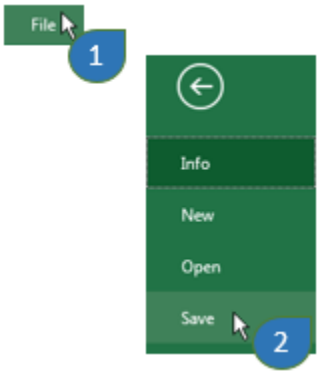
4: Añada una fila **Total Row** (al final de la tabla) que contenga las sumatorias de las columnas y el promedio de porcentajes en la última columna generada.



Añada la sumatoria a las otras columnas. Este es el procedimiento:



5: Guarde el archivo (**Save**) para la próxima práctica.



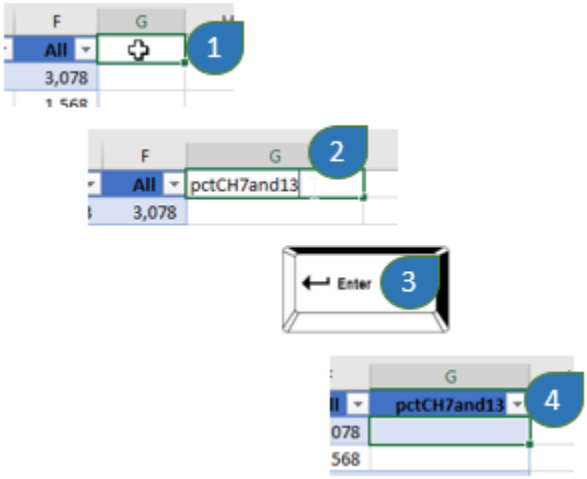
EJERCICIO 2: TRABAJAR CON TABLAS

Objetivos: En esta parte:

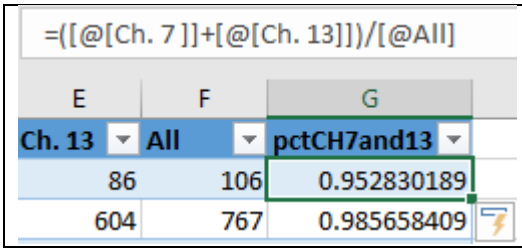
- **Añadiremos una columna**
- **Calcularemos valores** de porcentajes usando una fórmula simple
- **Cambiaremos** el **formato** de los números para representar cantidades y porcentajes.

1: Añada una columna adyacente a la última columna derecha. Llámela **pctCH7and13.**

Procedimiento:



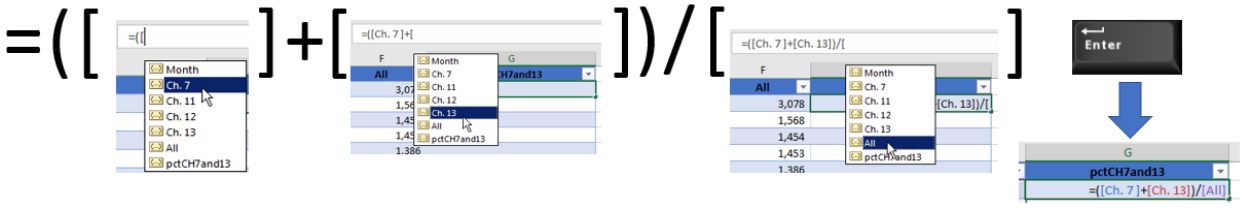
Esta debe contener el **porcentaje** de **quiebras** por **capítulo 7** y **Capítulo 13** *(fórmula =(Ch 7 + Ch 13)/All)* Ingrese los paréntesis para que no se altere el orden de las operaciones. Recuerde que con tablas Excel, las referencias no son por celdas sino por columnas…



Esto se hace añadiendo
= **abrir paréntesis** (**click** en celda **B2** signo de suma **+** **click** en celda **E2**) **cerrar paréntesis** / y **click** en celda **F2**

Note que, al generar la primera celda, automáticamente se calculan las demás hacia abajo.

Estos son los pasos:

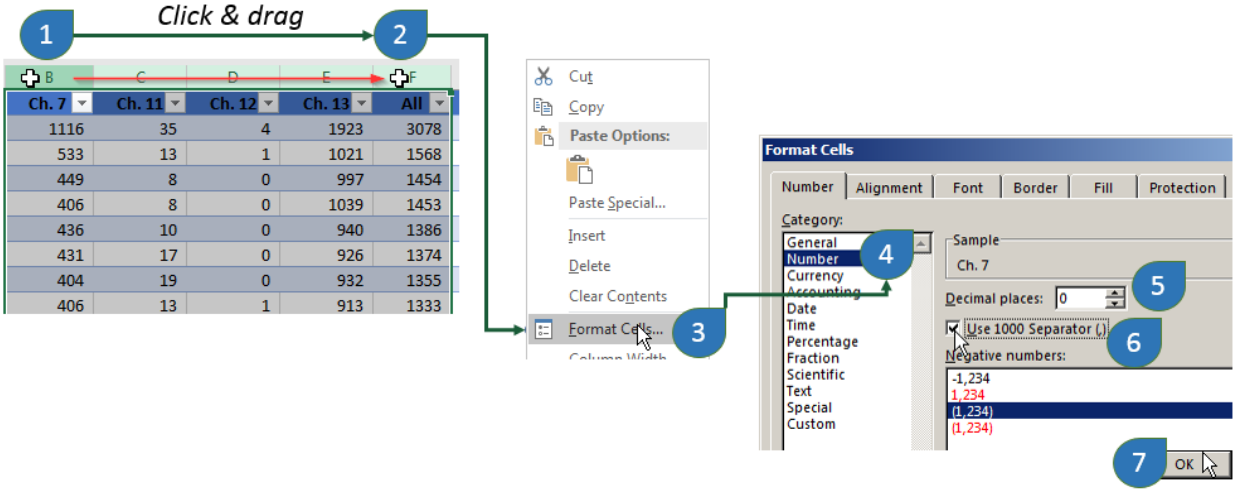


Posiblemente presente 1 en todas las celdas de la columna. Será necesario cambiar el formato de estos números para mostrar los porcentajes. Seguramente los valores están bien cercanos a 1.

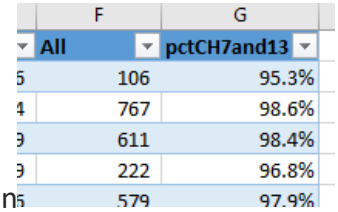
2: **Cambie el formato a las columnas**.

Seleccione las **columnas Ch 7 hasta la columna All** y haga **right click** | **Format Cells**.

Haga **check** en **Use 1000 Separator** (,) y cambie el **número de decimales** a **cero (0)**.



Seleccione la columna **pctCH7and13** y repita el proceso. Esta vez cambie el formato a porcentaje ***Percentage*** con **un lugar decimal**.



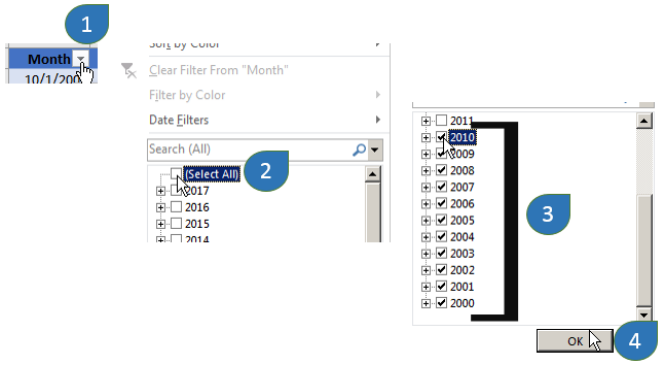
EJERCICIO 3: ANÁLISIS DE DATOS

Objetivo: Con el workbook **banckruptcy.xlsx**, usar herramientas ***Filter*** y ***Total Row*** para contestar preguntas sobre casos de quiebras entre los años 2000 y 2010.

Casos de quiebras: Usando la tabla Excel que acaba de generar:
¿Cuántos casos de quiebras hay entre los años 2000 y 2010?

Averigüe usando herramientas **Filter**, y **Total Row**. Informe solamente las **quiebras** por **capítulo 7**, **capítulo 13**, el **promedio Ch7&Ch13** y **All** (todas)

Pista/indicios:
Primero, **filtre datos por fecha**, incluyendo todos los meses de los años entre **1/1/2000 al 12/31/2010** inclusive. Hay distintas maneras de hacerlo.



Escriba las cantidades también para los meses de los años 2011 al 2017 inclusive.

Capítulo	**2000-2010**	**2011-2017**
Capítulo 7		
Capítulo 13		
Todas (All)		
Porcentaje (Caps 7&13)		

Los resultados aparecerán al final en la fila Total.
No olvide computar la fila **Total** (***average***) **en** la **columna pctCH7and13**

A	B	C	D	E	F	G
8/1/2006	144	1	0	406	551	99.82%
1/1/2007	115	14	0	354	483	97.10%
7/1/2006	94	16	2	335	447	95.97%
6/1/2006	106	4	2	303	415	98.55%
4/1/2006	95	8	0	311	414	98.07%
5/1/2006	103	7	0	302	412	98.30%
3/1/2006	88	3	2	263	356	98.60%
2/1/2006	69	5	1	223	298	97.99%
1/1/2006	52	2	1	185	240	98.75%
12/1/2005	46	4	3	169	222	96.85%
11/1/2005	15	5	0	86	106	95.28%
Total	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

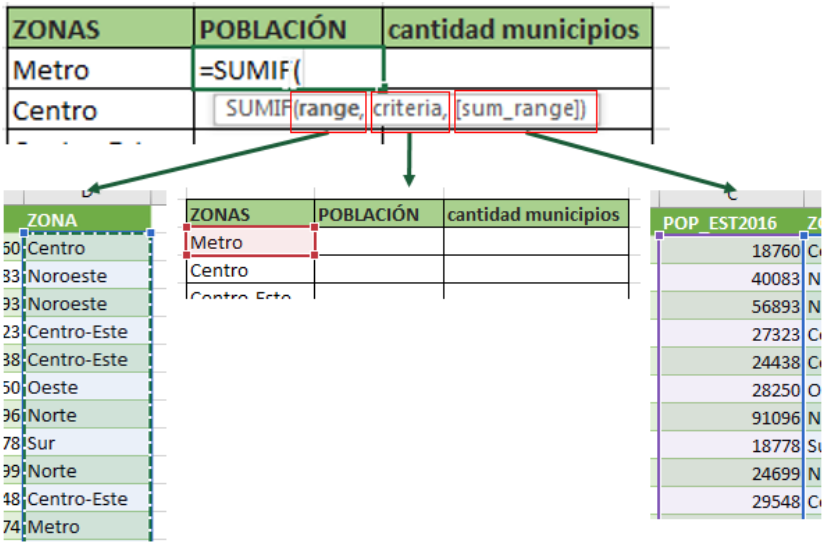
Guarde el archivo (Save). Lo vamos a utilizar en otro ejercicio.

EJERCICIO 4: USO DE FUNCIONES SUMIF Y COUNTIF

Abra el workbook **Municipios_2016.xlsx**, localizado en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_2**.

Complete la siguiente tabla. Utilice las funciones **SUMIF** para sumar habitantes por zona y **COUNTIF** para contar los municipios que pertenecen a cada región. Si va a copiar la fórmula a otras celdas, recuerde usar referencia absoluta ($) a la columna o a la fila, según sea el caso.

ZONAS	POBLACIÓN	cantidad municipios
Metro		
Centro		
Centro-Este		
Centro-Oeste		
Este		
Sureste		
Noreste		
Norte		
Noroeste		
Oeste		
Suroeste		
Sur		



range: celdas de la columna ZONA
criteria: celda que contiene el valor “Metro” en la columna “ZONAS”
sum_range: data range con los valores de población en la columna POP_EST2016

EJERCICIO 5: FUNCIONES VLOOKUP Y LOOKUP

VLOOKUP

Como se explicó anteriormente, la función VLOOKUP se usa para traer valores de un sitio a otro usando un identificador común para parear y asignar valores a una o más celdas. Trabajaremos un ejemplo con datos censales de población por décadas y municipios.

El workbook tiene dos hojas: una contiene los datos históricos de población y la otra hoja tiene el estimado de población (American Community Survey) del Censo Federal para 2017.

Tareas:

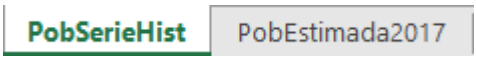
El objetivo es practicar **cómo pasar los valores del estimado de población 2017 en una hoja a la otra hoja que tiene los datos de población histórica *sin hacer copy/paste***.

Insertar columna y darle nombre:

Escribir y ejecutar la función VLOOKUP

Como vamos a trabajar con una Tabla Excel, no tendremos que copiar la fórmula al resto de las celdas de la columna.

Abra el archivo **EjercicioVLookup.xlsx**, localizado en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_2**.

El workbook tiene dos hojas. 

Deberá abrir primero la hoja ***PobSerieHist*** así:

	A	B	C	D	E	F	G	H	I	J	K	L
1	GEO_ID	Municipio	POP1930	POP1940	POP1950	POP1960	POP1970	POP1980	POP1990	POP2000	POP2010	
2	72001	Adjuntas	18,075	22,556	22,424	19,658	18,691	18,786	19,451	19,143	19,483	
3	72003	Aguada	14,670	17,923	20,743	23,234	25,658	31,567	35,911	42,042	41,959	
4	72005	Aguadilla	28,319	34,956	44,357	47,864	51,355	54,606	59,335	64,685	60,949	
5	72007	Aguas Buenas	12,885	14,671	15,565	17,034	18,600	22,429	25,424	29,032	28,659	
6	72009	Aibonito	16,361	16,819	18,191	18,360	20,044	22,167	24,971	26,493	25,900	
7	72011	Añasco	14,276	15,701	17,235	17,200	19,416	23,274	25,234	28,348	29,261	
8	72013	Arecibo	56,525	69,192	75,361	69,879	73,468	86,766	93,385	100,131	96,440	
9	72015	Arroyo	8,199	10,746	12,936	13,315	13,033	17,014	18,910	19,117	19,575	
10	72017	Barceloneta	15,751	18,545	19,897	19,334	20,792	18,942	20,947	22,322	24,816	
11	72019	Barranquitas	14,901	17,096	17,605	18,978	20,118	21,639	25,605	28,909	30,318	
12	72021	Bayamón	29,524	37,190	48,000	72,221	156,192	196,206	220,262	224,044	208,116	

Vamos a **trasladar** los **valores** de **estimados de población** para **2017** presentes en la hoja ***PobEstimada2017*** a la **columna L** de la **hoja PobSerieHist**.

	A	B
1	GEO_ID	EST_2017
2	72001	17,971
3	72003	38,118
4	72005	53,164
5	72007	25,850
6	72009	23,108

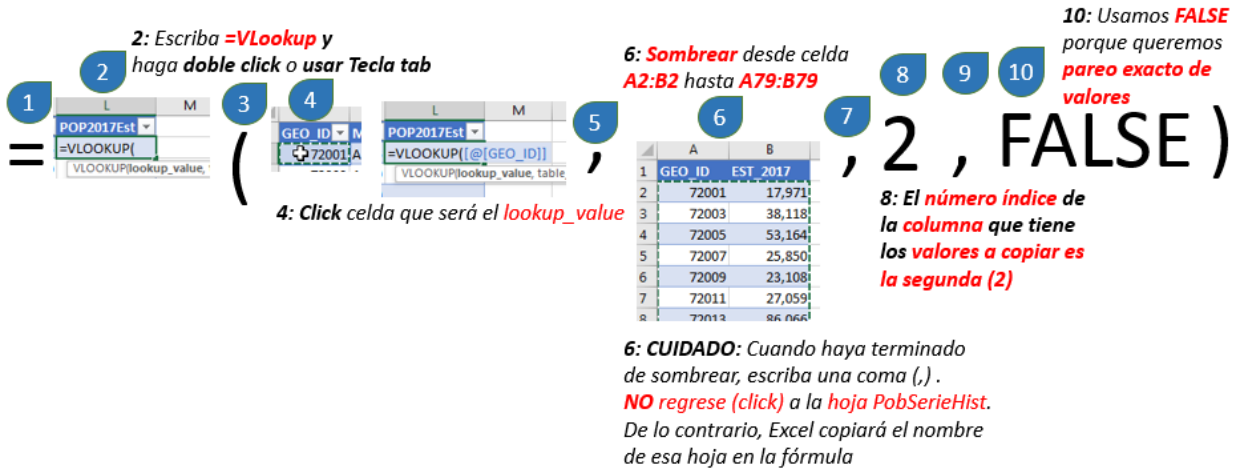
Haga **click** en la **primera celda** de la **columna L** y escriba **POP2017Est**, seguido de ***Enter***.

Verá que la columna se añadió a la tabla Excel. Se añadirá una columna a la tabla siempre y cuando la nueva columna sea adyacente a la última columna de la derecha.

GEO_ID	Municipio	POP1930	POP1940	POP1950	POP1960	POP1970	POP1980	POP1990	POP2000	POP2010	POP2017Est
72001	Adjuntas	18,075	22,556	22,424	19,658	18,691	18,786	19,451	19,143	19,483	

Escribir y ejecutar función VLOOKUP

En la celda inferior a la cabecera de la columna nueva POP2017Est, comenzará a escribir los siguientes pasos:



GEO_ID	EST_2017
72001	17,971
72003	38,118
72005	53,164
72007	25,850
72009	23,108
72011	27,059

Como estamos trabajando con una tabla Excel, verá que los valores son calculados automáticamente en las celdas inferiores:

L
POP2017Est
17,971
38,118
53,164
25,850
23,108
27,059
86,066
17,881
24,240
28,511
179,565
48,824
129,604

OPCIONAL: LOOKUP

Volvamos a abrir el workbook **bankruptcy.xlsx**, **para clasificar datos numéricos**, **agrupando por categorías**: 0 a 100, 101 a 200, etcétera.

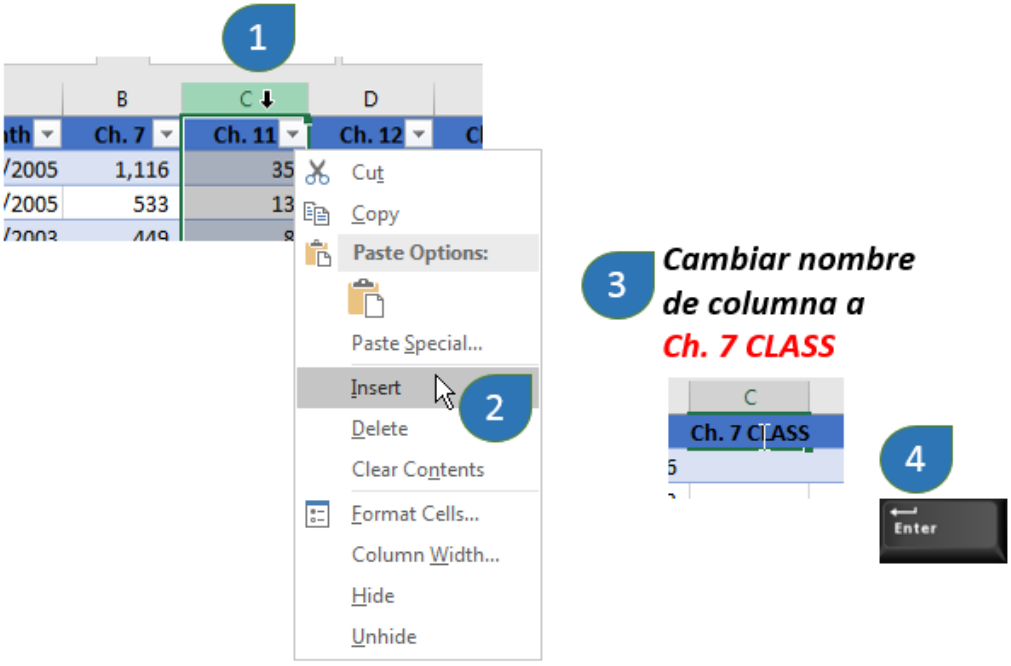
Tareas:

- **Insertar (field) columna nueva** para guardar los grupos de valores (clases) para las quiebras por capítulo 7. Estas "clases" estarán guardadas en formato alfanumérico.
- **Escribir la fórmula** que incluya la función **LOOKUP**. En este ejemplo, escribiremos (*hard coding)* los valores de inicio y final de cada clase, en lugar de tomarlos desde celdas en un data range.

Insertar columna

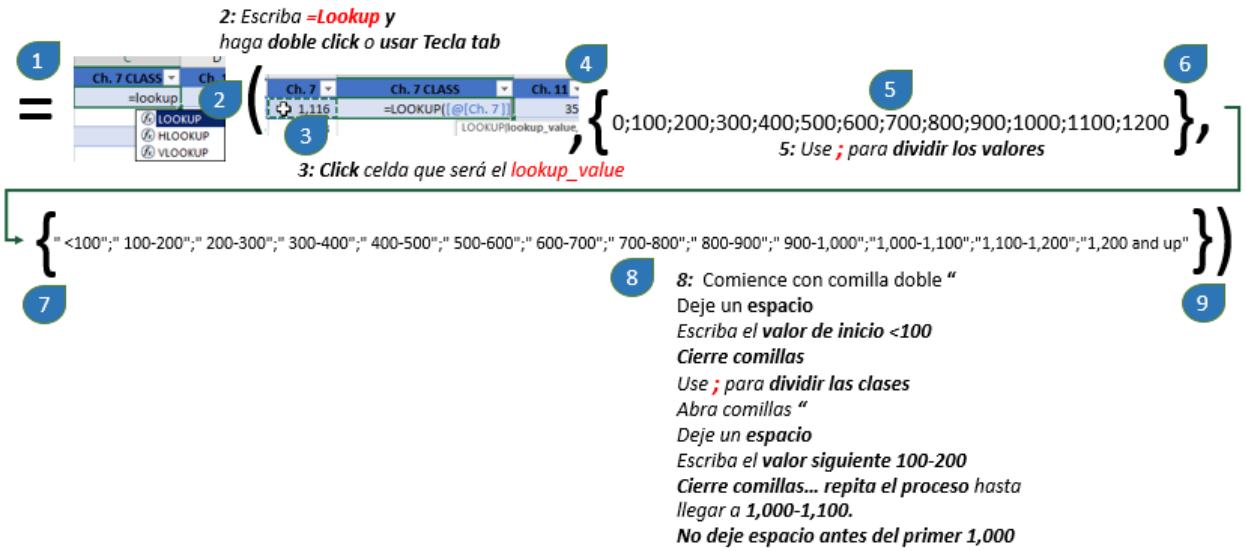
Abra el archivo **bankruptcy.xlsx**, localizado en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_2.**

Estos son los pasos: **right click** en la **columna Ch. 11**, **click Insert**.
Nombre de la nueva columna: **Ch. 7 Class**. Presione **Enter** para ejecutar el cambio.



Escribir la función LOOKUP

Estos son los pasos:



Use estas listas COPY/PASTE:

Lookup vector: `0;100;200;300;400;500;600;700;800;900;1000;1100;1200`

Result vector: `" <100";" 100-200";" 200-300";" 300-400";" 400-500";" 500-600";" 600-700";" 700-800";" 800-900";" 900-1,000";"1,000-1,100";"1,100-1,200";"1,200 and up"`

Note que **las clases resultantes tienen un espacio al inicio**. Esto se usa para que Excel ordene (Sort) correctamente las clases. Al usar alfanuméricos, el 1 y 10, 11 hasta el 19 van primero que el 2. Por lo tanto, es recomendable poner espacios en los dígitos según la cantidad de espacios que tengan las cifras. En este caso, solo falta un espacio.

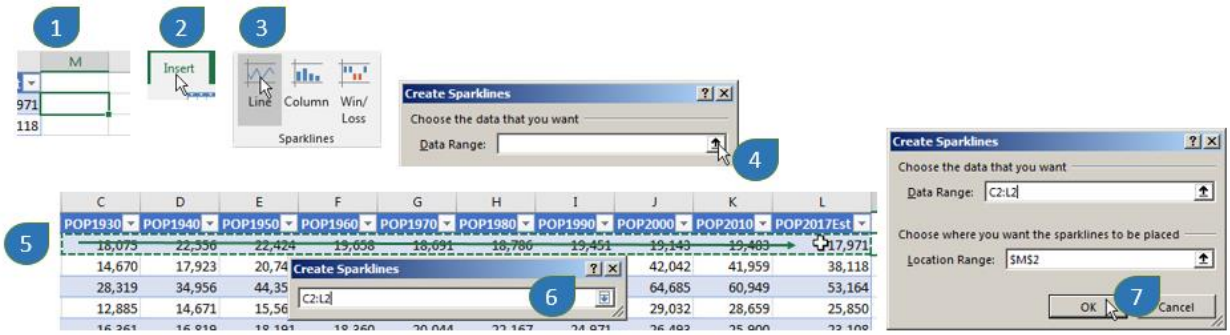
Así se verá la tabla Excel luego de aplicar la función. Automáticamente calculará los valores en las demás celdas inferiores de la columna **Ch. 7 CLASS**.

	B	C	D
h	Ch. 7	Ch. 7 CLASS	Ch. 11
2005	1,116	1,100-1,200	35
2005	533	500-600	13
2003	449	400-500	8
2001	406	400-500	8
2004	436	400-500	10
2001	431	400-500	17

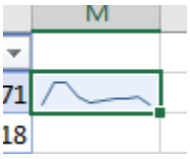
Cierre este archivo. **Guárdelo** (Save).

OPCIONAL: Sparklines para visualizar fluctuación

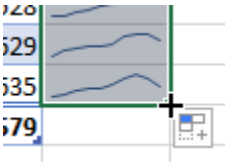
Para una visualización rápida de la fluctuación de población por décadas:



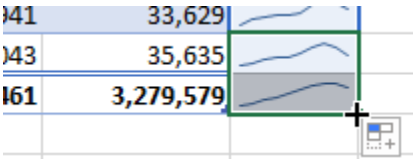
Este será el resultado para la primera celda:



Para copiar el sparkline, haga **click en el cuadrito verde** y **arrástrelo** hasta la **penúltima celda M77**.



Luego arrastre este M77 a la última celda **M79** para tener la fluctuación a nivel de Puerto Rico.



Esto concluye el ejercicio. Guarde este workbook Excel (Save).

MÓDULO 3: PIVOTTABLES Y PIVOTCHARTS

RESUMEN

Una tabla Pivot es una herramienta para resumir y analizar datos, especialmente si son voluminosos. Es una de las herramientas más útiles de Excel para el análisis de datos desde diferentes perspectivas. Una gráfica Pivot es igual a cualquier otra gráfica Excel. La diferencia es que las PivotCharts están ligadas a una tabla Pivot.

DIMENSIONES Y MÉTRICAS EN PIVOTTABLES

Para entender un poco mejor, imagínese un conjunto de datos, un data range, una Tabla Excel. A menudo los datos contienen descripciones y mediciones. Algunos llaman ***dimensiones*** o perspectivas a estas **descripciones**. Las ***métricas*** son mediciones: conteos, volúmenes, sumatorias, proporciones ***numéricas*** que componen la segunda parte de estos datos. Algunos conjuntos de datos tienen mayor número de dimensiones que métricas y viceversa.

DIMENSIONES MÉTRICAS

ESCUELA	DIRECCION_MUNICIPIO	DIRECCION_ZIPCODE	ZONA	NIVEL_ORIGINAL	GRADOS_POSTERIOR	MATRICULA	ESPAÑOL	MATEMÁTICAS	INGLÉS	BAJO_NIVEL
COTTO (ANEXO)	ARECIBO	00613-9582	URBANA	TODOS LOS NIVELES	EEE - EES	26	0.8	40%	20%	72%
JARDINES DE CAGUAS	CAGUAS	7267396	URBANA	ELEMENTAL	EEE	51	0.4286	100%	14%	73%
SU REXFORD GUY TUGWELL	CAYEY	00737-0000	RURAL	SEGUNDA UNIDAD	K - 9	52	0.7143	71%	86%	68%
PRE-VOCACIONAL CASIANO CEPEDA	RIO GRANDE	7450000	RURAL	SECUNDARIO	EEI - EES	53	0.1667	0	67%	90%
JOSE GUALBERTO PADILLA	CAYEY	7370000	RURAL	ELEMENTAL	K - 6	54	1	1	88%	87%
TOMAS VERA AYALA	PATILLAS	00723-0000	RURAL	SEGUNDA UNIDAD	K - 9 , EEE	60	0.4545	0.5455	27%	96%
JOAQUIN FERRAN LLUIS	PONCE	00732-7105	URBANA	ELEMENTAL	K - 6 , EEE	61	0.7143	0.7317	54%	90%
MANUEL CRUZ MACEIRA	COMERIO	00782-0000	URBANA	ELEMENTAL	1-6	65	0.4	0%	30%	54%
LENA M FRANCESCH - RUBIAS KM 2.1	YAUCO	-698	RURAL	ELEMENTAL	K - 6	71	0.6327	61%	43%	100%
AVELINO PENA REYES	HUMACAO	00791-0000	URBANA	SECUNDARIO	EEI - EES	72	0.2857	71%	23%	68%
PASO PALMA I	UTUADO	00641-0000	RURAL	ELEMENTAL	K - 6	75	0.8056	0.9167	100%	92%
VICTOR PARES COLLAZO	SAN JUAN	00918-0000	URBANA	TODOS LOS NIVELES	PK - EEE - EES	77	0.1538	27%	8%	62%
LAMBOGLIA	PATILLAS	00723-0000	RURAL	ELEMENTAL	K - 6	77	0.2329	0.0685	21%	92%
DAMIAN ABAJO	OROCOVIS	00720-0000	RURAL	ELEMENTAL	K - 6	81	1	1	91%	83%
INDIERA FRIA	MARICAO	00606-0000	RURAL	ELEMENTAL	PK - 6 , EEE	81	0.9118	0.7059	71%	93%
CARMEN APONTE	UTUADO	-611	RURAL	ELEMENTAL	K - 6	81	0.9429	97%	77%	78%
JUANITA RIVERA ALBERT	VIEQUES	7659712	RURAL	ELEMENTAL	PK - 1 , EEE	82				92%
MARIA C HUERTAS	COMERIO	00782-0000	RURAL	ELEMENTAL	K - 6 , EEE	84	0.9444	1	100%	81%
BRYAN	LAS MARIAS	00670-9069	RURAL	ELEMENTAL	K - 6	84	0.4545	0.6818	61%	94%
CRISTOBAL VICENS	CIALES	00638-0000	RURAL	ELEMENTAL	K - 6	85	0.2927	41%	22%	76%
LUDOVICO COSTOSO	BAYAMON	00959-5259	RURAL	TODOS LOS NIVELES	EEE - EES	86	0.2593	0.5185	21%	81%

Dimensiones y mediciones (*métricas* también) en una tabla o un data range.

Las dimensiones pueden verse como diferentes perspectivas para organizar los datos. Las PivotTables utilizan estas dimensiones para agrupar y resumir las mediciones mediante conteos, sumatorias, promedios, etcétera.

NOTA DE PRECAUCIÓN: CELDAS VACÍAS EN UN DATA RANGE O UNA TABLA

Tenga cuidado cuando haya celdas vacías dentro del data range o una tabla. Cuando hay espacios vacíos, Excel presume que los datos no son numéricos.

UTILIDAD

Además de resumir datos, las PivotTables pueden usarse para organizar, filtrar, resumir y analizar datos crudos. Por ejemplo, resumir datos de una tabla de inventario de escuelas para saber cuántas escuelas hay por región, distrito, municipio, zona, nivel, … Además, Excel permite hacer columnas de datos nuevos en las PivotTables, basadas en operaciones matemáticas.

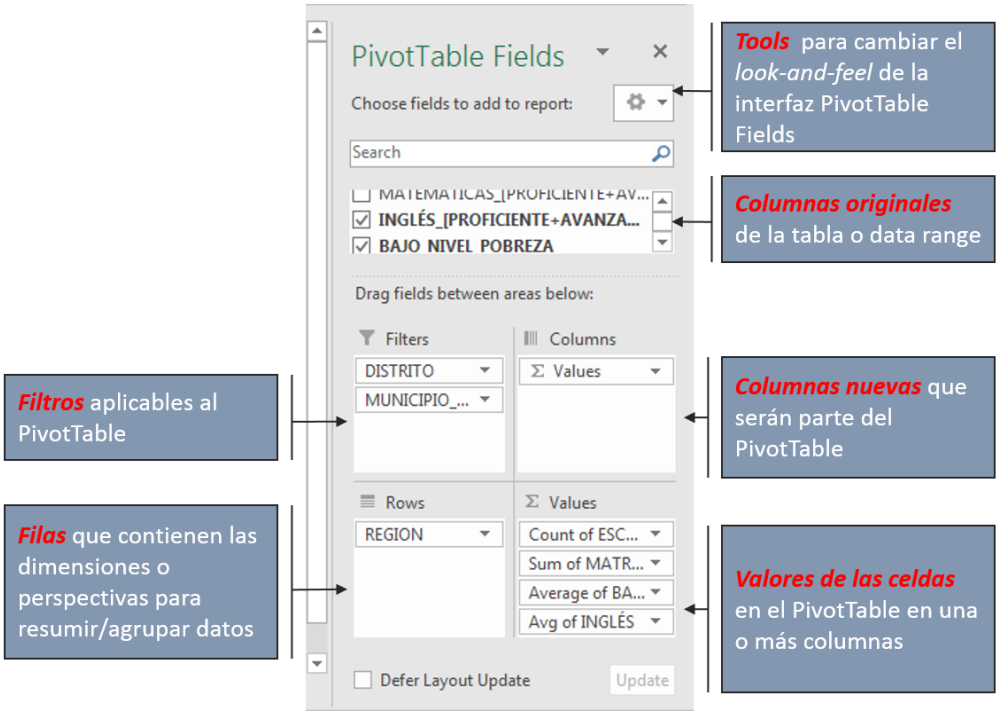
REGION	DISTRITO	MUNICIPIO_ESCOLAR	CODIGO	ESCUELA	DIRECCION_MUNICIPIO	DIRECCION_ZIP
ARECIBO	ARECIBO	ARECIBO II	10173	COTTO (ANEXO)	ARECIBO	00
CAGUAS	GURABO	CAGUAS I	23150	JARDINES DE CAGUAS	CAGUAS	
CAGUAS	CIDRA	CAYEY	21345	SU REXFORD GUY TUGWELL	CAYEY	00
HUMACAO	CANOVANAS	RIO GRANDE	35543	PRE-VOCACIONAL CASIMIRO CEPEDA	RIO GRANDE	74
CAGUAS	CIDRA	CAYEY	21303	JOSE GUALBERTO PA	CAYEY	7
HUMACAO	YABUCOA	PATILLAS	25189	TOMAS VERA AYAL	PATILLAS	0
PONCE	PONCE	PONCE II	52175	JOAQUIN FERRAN	PONCE	0
CAGUAS	BARRANQUITAS	COMERIO			COMERIO	00782-0000
PONCE	YAUCO	YAUCO			YAUCO	-698
HUMACAO	LAS PIEDRAS	HUMACAO			HUMACAO	00791-0000
PONCE	UTUADO	UTUADO			UTUADO	00641-0000
SAN JUAN	SAN JUAN I	SAN JUAN II	65953	VICTOR PARES COLLAZO	SAN JUAN	00918-0000
HUMACAO	YABUCOA	PATILLAS	25080	LAMBOGLIA	PATILLAS	00723-0000
BAYAMON	OROCOVIS	OROCOVIS	12567	DAMIAN ABAJO	OROCOVIS	00720-0000
MAYAGUEZ	MAYAGUEZ	MARICAO	41913	INDIERA FRIA	MARICAO	00606-0000
PONCE	UTUADO	UTUADO	13094	CARMEN APONTE	UTUADO	-611
HUMACAO	FAJARDO	VIEQUES	32300	JUANITA RIVERA ALBERT	VIEQUES	7659712
CAGUAS	BARRANQUITAS	COMERIO	21782	MARIA C HUERTAS	COMERIO	00782-0000
MAYAGUEZ	MAYAGUEZ	LAS MARIAS	41632	BRYAN	LAS MARIAS	00670-9069

ESCUELA	(All)					
REGION	DISTRITO	MUNICIPIO_ESCOLAR	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA	Avg Competencia INGLÉS
SAN JUAN			168	54,645	82.4%	35.4%
	CAROLINA		30	12,911	82.0%	32.7%
		CAROLINA II	17	7,476	81.1%	34.5%
		CAROLINA I	13	5,435	83.2%	30.4%
	GUAYNABO		32	10,937	80.0%	43.0%
		GUAYNABO	17	5,964	78.3%	43.9%
		TRUJILLO ALTO	15	4,973		
	SAN JUAN I		44	12,847		
		SAN JUAN II	21	6,779		
		SAN JUAN I	23	6,068		
	SAN JUAN II		62	17,950		
		SAN JUAN IV	26	7,020		
		SAN JUAN III	22	6,725	87.3%	29.7%
		SAN JUAN V	14	4,205	87.0%	24.3%
ARECIBO			134	50,011	81.3%	49.4%

Con esta herramienta podemos contestar preguntas que no pueden ser contestadas fácilmente usando solamente filtros y fórmulas en Excel.

INTERFAZ PIVOTTABLE FIELDS

Antes de comenzar a trabajar con PivotTables, debemos entender la interfaz **PivotTable Fields**. A través de esta interfaz escogemos cuáles serán las ***dimensiones*** Rows que queremos usar para resumir datos, las ***mediciones*** Σ Values que vamos a calcular y qué ***filtros*** Filters vamos a aplicar para separar unos datos de otros, si es necesario.



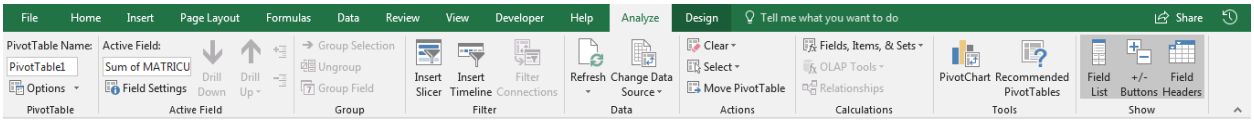
Podemos incluir además otras ***columnas*** Columns a la derecha en la tabla Pivot que vayamos a construir, aunque no es requerido para construir una tabla Pivot como en el ejemplo anterior.

DISPOSICIÓN DE LOS DATOS PARA QUE FUNCIONEN LAS TABLAS PIVOT

Apropiado	Inapropiado
• **Rectangular: variables en columnas observaciones en filas** • **Sin formateo extra** • **Contiene solo dimensiones y mediciones** • **Headers (cabeceras de columnas) claras y precisas** • **No debe tener headers extra, footers (calce), subtotales, ni campos calculados**	• Datos traspuestos (TRANSPOSE) variables como filas, observaciones como columnas • Formateo innecesario • Contiene campos calculados • Nombres de columnas/headers ambiguos • Filas extra en las cabeceras de columnas

PIVOTTABLE TOOLS GROUP TABS

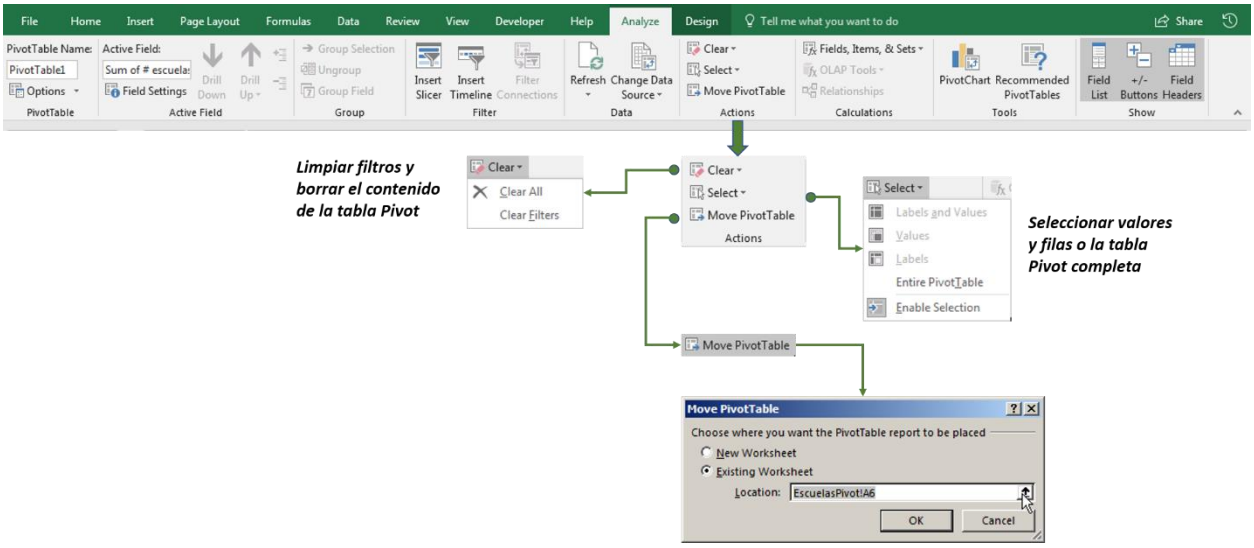
HERRAMIENTAS DEL TAB ANALYZE:



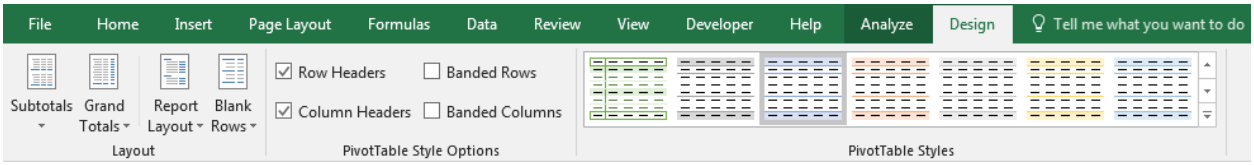
Grupos:

PivotTable	Opciones de configuración y nombre del PivotTable
Active Field	Opciones de configuración de campos o columnas (fields)
Group	Para agrupar o desagrupar datos
Filter	Seleccionar/separar datos
Data	Opciones para cambiar la fuente de datos y traer a memoria los datos (refresh)
Actions	Opciones para borrar datos en memoria, seleccionar medidas y labels y mover la tabla Pivot
Calculations	Cálculo de nuevas columnas (fields, items and sets) y relaciones entre tablas
Tools	Generar PivotCharts y recomendaciones para hacer otros PivotTables
Show	Opciones para configurar listas de campos, botones y cabeceras de columnas

Botones del grupo **Actions**: Mover tabla Pivot, limpiar contenido, seleccionar valores, filas o la tabla completa.



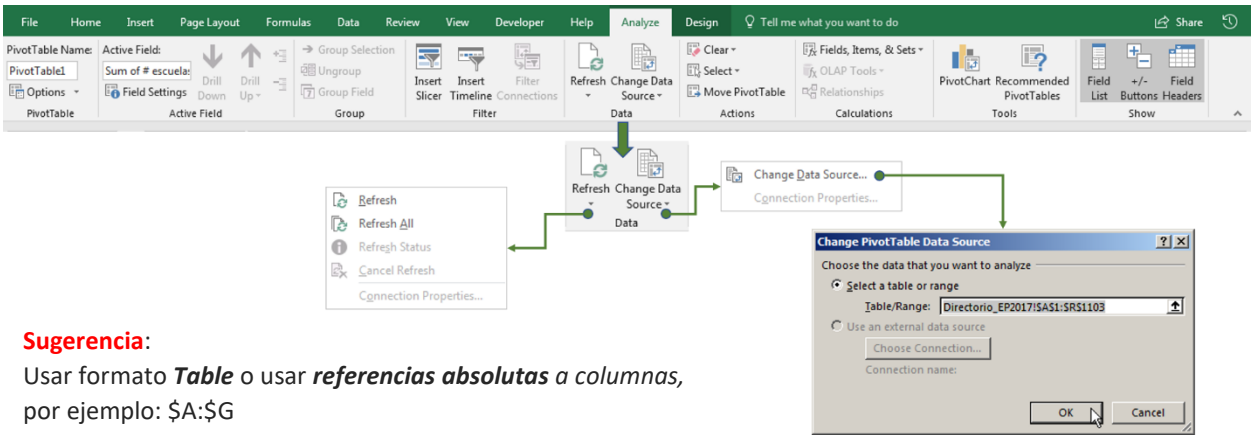
HERRAMIENTAS DEL TAB *DESIGN*



GRUPOS:

Layout	Opciones para cambiar la apariencia de la tabla Pivot en subtotales, totales, filas vacías.
PivotTable Style Options	Opciones para incluir cabeceras de filas y columnas y sombreado de filas y columnas

Botones del grupo **Data | Refresh** de los datos cuando haya cambios en los datos crudos, cambiar la fuente de datos.

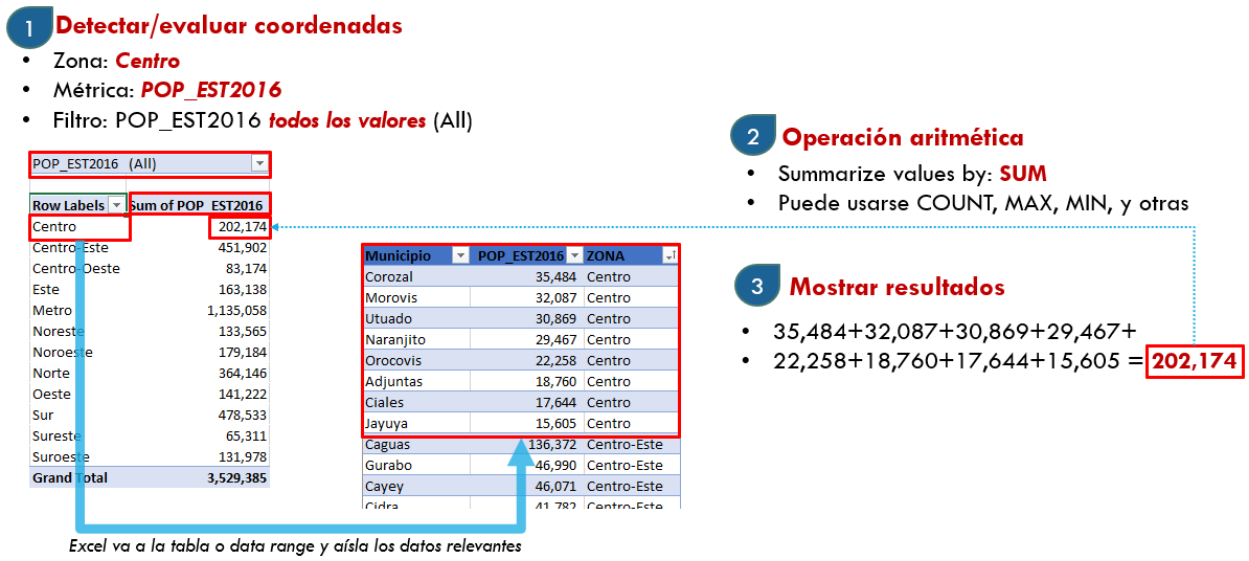


Sugerencia:
Usar formato ***Table*** o usar ***referencias absolutas*** *a columnas,*
por ejemplo: $A:$G

Refresh actualiza trayendo los datos del grupo de celdas (data range) a la tabla Pivot.

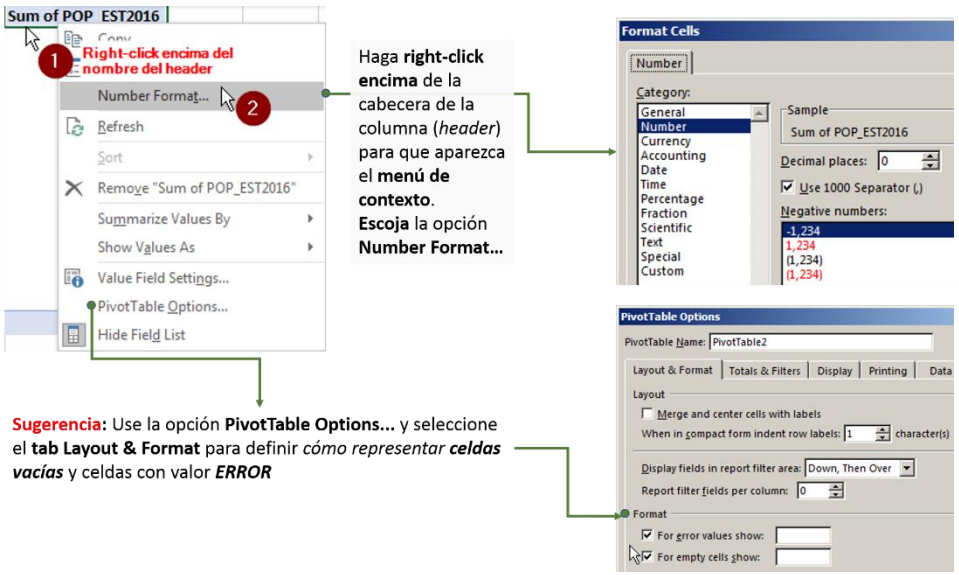
CÓMO FUNCIONA UNA TABLA PIVOT

Para generar una tabla Pivot, Excel selecciona y agrupa los datos necesarios para devolver un resultado, basado en las dimensiones que escoja el usuario. En este caso, se está resumiendo (mediante sumatoria) la población 2016 por zonas que incluyen cierto número de municipios.



FORMATEAR NÚMEROS

Como se mencionó anteriormente, es preferible darle un formato claro a los números para que se entienda qué representan, especialmente si son cantidades, moneda, porcentajes o proporciones. Escoja la opción apropiada para sus números, así como la cantidad de decimales que usted especifique.



ESTILOS PARA TABLAS PIVOT

Excel le provee múltiples opciones para cambiar los colores y las líneas de las tablas. Usted puede escoger de la lista o desarrollar un estilo particular y guardarlo.

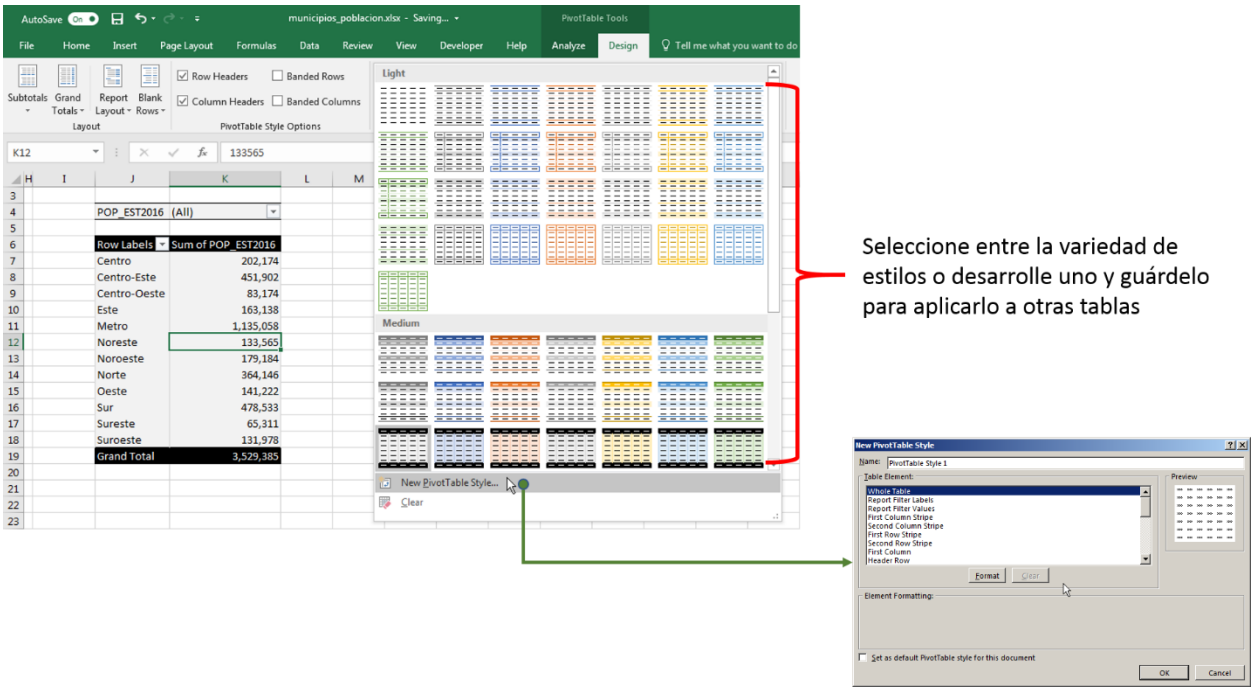


TABLE LAYOUTS: ORGANIZACIÓN DE LA TABLA

Para el análisis de datos, es preferible usar la forma extendida o de bosquejo (*outline*).

Row Labels	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA	Avg Competencia INGLÉS	Densidad Estudiantil
SAN JUAN	**168**	**54,645**	**82.4%**	**35.4%**	**325.27**
CAROLINA	**30**	**12,911**	**82.0%**	**32.7%**	**430.37**
CANOVANAS	2	750	80.7%	26.4%	375.00
CAROLINA	28	12,161	82.1%	33.2%	434.32
GUAYNABO	**32**	**10,937**	**80.0%**	**43.0%**	**341.78**
CAGUAS	2	441	86.6%	66.3%	220.50
CATANO	2	509	60.1%	48.9%	254.50
GUAYNABO	12	4,547	79.5%	40.8%	378.92
SAN JUAN	2	733	80.2%	25.8%	366.50
TRUJILLO ALTO	14	4,707	82.2%	43.3%	336.21
SAN JUAN I	**44**	**12,847**	**76.4%**	**38.8%**	**291.98**
SAN JUAN	44	12,847	76.4%	38.8%	291.98
SAN JUAN II	**62**	**17,950**	**88.0%**	**30.3%**	**289.52**
GUAYNABO	1	413	94.8%	23.9%	413.00
SAN JUAN	60	17,008	88.2%	30.2%	283.47
TRUJILLO ALTO	1	529	69.8%	43.7%	529.00
ARECIBO	**134**	**50,011**	**81.3%**	**49.4%**	**373.22**

Forma compacta: (default)

Los campos están anidados en la primera columna. Permite aplicar solo un filtro personalizado (custom filter)

REGION	DISTRITO	DIRECCION_MUNICIPIO	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA	Avg Competencia INGLÉS	Densidad Estudiantil
SAN JUAN			**168**	**54,645**	**82.4%**	**35.4%**	**325.27**
	CAROLINA		**30**	**12,911**	**82.0%**	**32.7%**	**430.37**
		CANOVANAS	2	750	80.7%	26.4%	375.00
		CAROLINA	28	12,161	82.1%	33.2%	434.32
	GUAYNABO		**32**	**10,937**	**80.0%**	**43.0%**	**341.78**
		CAGUAS	2	441	86.6%	66.3%	220.50
		CATANO	2	509	60.1%	48.9%	254.50
		GUAYNABO	12	4,547	79.5%	40.8%	378.92
		SAN JUAN	2	733	80.2%	25.8%	366.50
		TRUJILLO ALTO	14	4,707	82.2%	43.3%	336.21
	SAN JUAN I		**44**	**12,847**	**76.4%**	**38.8%**	**291.98**
		SAN JUAN	44	12,847	76.4%	38.8%	291.98
	SAN JUAN II		**62**	**17,950**	**88.0%**	**30.3%**	**289.52**
		GUAYNABO	1	413	94.8%	23.9%	413.00
		SAN JUAN	60	17,008	88.2%	30.2%	283.47
		TRUJILLO ALTO	1	529	69.8%	43.7%	529.00
ARECIBO			**134**	**50,011**	**81.3%**	**49.4%**	**373.22**

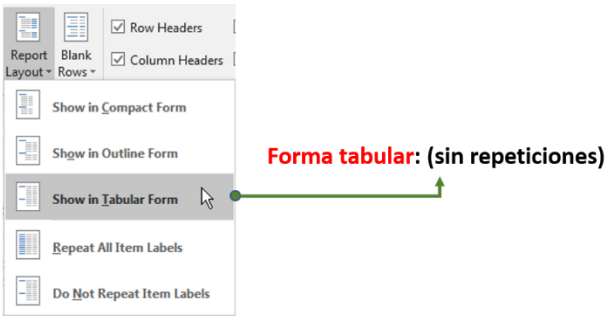
Forma Outline: Bosquejo (recomendada)

Los **campos** aparecen **cada uno en su columna**. **Permite aplicar filtros a cada columna tipo label: (las que tienen los botones de filtrado)**

Las columnas que permiten las opciones para ordenar son las que aparecen con los botones de filtrado.

FORMA TABULAR

Estas opciones son útiles para convertir una tabla Pivot en una tabla "rectangular"; primero estirando (*flattening*) la tabla, expandiendo las columnas anidadas en nuevas columnas.



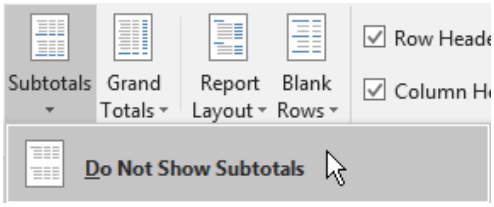
REGION	DISTRITO	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA
⊟SAN JUAN	CAROLINA	30	12,911	82.0%
	GUAYNABO	32	10,937	80.0%
	SAN JUAN I	44	12,847	76.4%
	SAN JUAN II	62	17,950	88.0%
SAN JUAN Total		**168**	**54,645**	**82.4%**
⊟ARECIBO	CAMUY	27	11,303	79.1%
	VEGA ALTA	37	14,049	82.7%
	ARECIBO	36	13,321	80.1%
	MANATI	34	11,338	82.9%
ARECIBO Total		**134**	**50,011**	**81.3%**
⊟MAYAGUEZ	SAN SEBASTIAN	40	14,054	79.0%
	AGUADILLA	46	14,876	81.9%
	CABO ROJO	40	12,731	78.9%
	MAYAGUEZ	43	11,429	83.4%
MAYAGUEZ Total		**169**	**53,090**	**80.9%**

REGION	DISTRITO	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA
⊟SAN JUAN	CAROLINA	30	12,911	82.0%
SAN JUAN	GUAYNABO	32	10,937	80.0%
SAN JUAN	SAN JUAN I	44	12,847	76.4%
SAN JUAN	SAN JUAN II	62	17,950	88.0%
SAN JUAN Total		**168**	**54,645**	**82.4%**
⊟ARECIBO	CAMUY	27	11,303	79.1%
ARECIBO	VEGA ALTA	37	14,049	82.7%
ARECIBO	ARECIBO	36	13,321	80.1%
ARECIBO	MANATI	34	11,338	82.9%
ARECIBO Total		**134**	**50,011**	**81.3%**
⊟MAYAGUEZ	SAN SEBASTIAN	40	14,054	79.0%
MAYAGUEZ	AGUADILLA	46	14,876	81.9%
MAYAGUEZ	CABO ROJO	40	12,731	78.9%
MAYAGUEZ	MAYAGUEZ	43	11,429	83.4%
MAYAGUEZ Total		**169**	**53,090**	**80.9%**

Sugerencia: Use la opción **Tabular con repeticiones** para crear una tabla nueva como "datos crudos".

Más aún, podemos rellenar los espacios vacíos repitiendo los valores "de cabecera". Esto es útil para *hacer* la *transición de un PivotTable a una tabla de una base de datos*. Deberá eliminar también las filas Subtotales, si le interesa convertir la tabla en una tabla de datos crudos *raw dataset*. Esto aparece en el menú **PivotTable Tools** en **Layout | SubTotals | Do not Show Subtotals.**

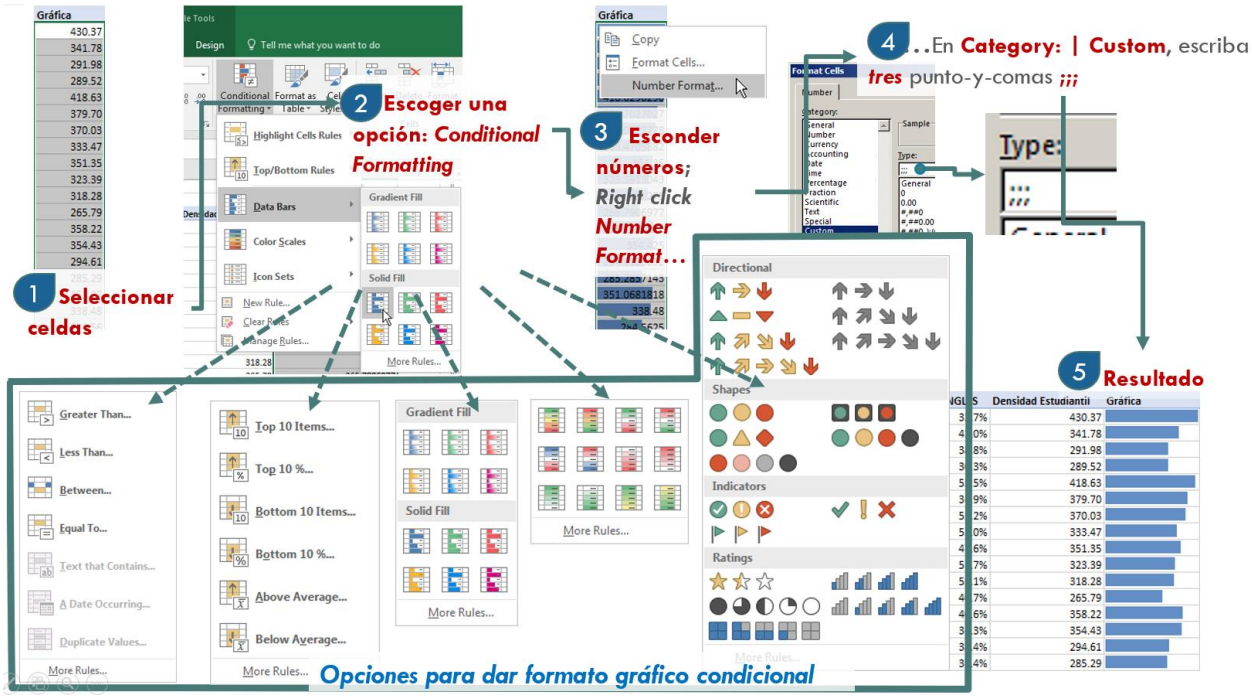
Eliminar subtotales:



REGION	DISTRITO	Sum of # escuelas	Sum of MATRICULA_TOTAL	Avg BAJO_NIVEL_POBREZA
SAN JUAN	CAROLINA	30	12,911	82.0%
SAN JUAN	GUAYNABO	32	10,937	80.0%
SAN JUAN	SAN JUAN I	44	12,847	76.4%
SAN JUAN	SAN JUAN II	62	17,950	88.0%
ARECIBO	CAMUY	27	11,303	79.1%
ARECIBO	VEGA ALTA	37	14,049	82.7%
ARECIBO	ARECIBO	36	13,321	80.1%
ARECIBO	MANATI	34	11,338	82.9%
MAYAGUEZ	SAN SEBASTIAN	40	14,054	79.0%
MAYAGUEZ	AGUADILLA	46	14,876	81.9%
MAYAGUEZ	CABO ROJO	40	12,731	78.9%
MAYAGUEZ	MAYAGUEZ	43	11,429	83.4%

FORMATEO CONDICIONAL (CONDITIONAL FORMATTING)

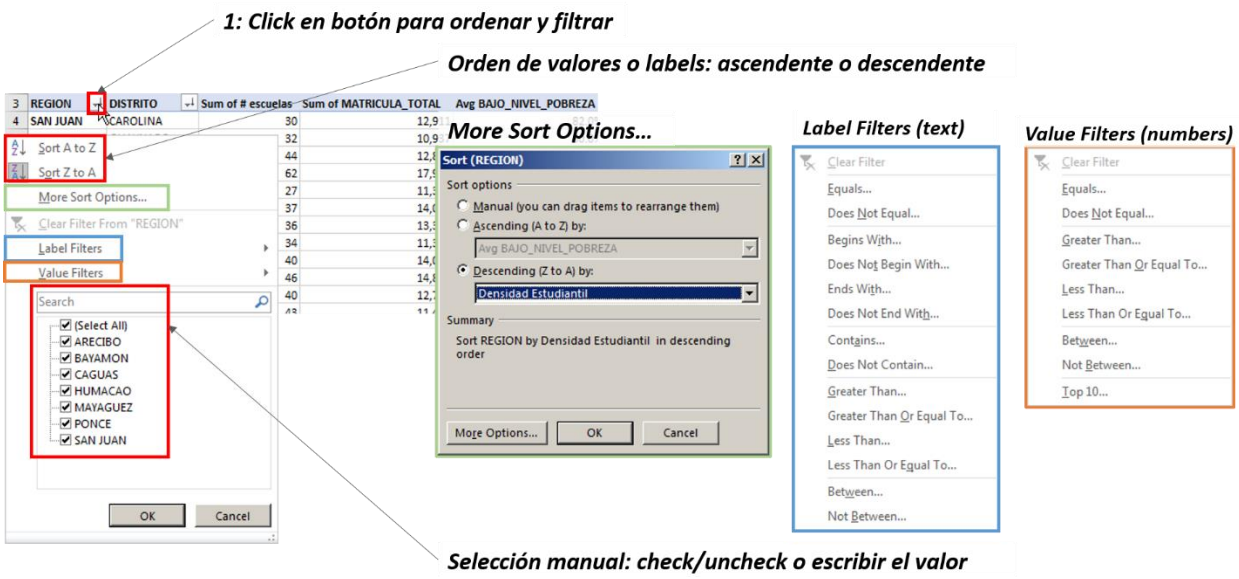
Las tablas Pivot pueden contener celdas con formateo condicional (gráfico y colores) de la misma manera que las tablas Excel y data ranges. Para representar una columna numérica usando barras, por ejemplo, se debe duplicar la columna para producir el gráfico. Estos son los pasos:



Note que en el paso #4 hacemos que los números sean transparentes, dejando ver en el paso #5, la gráfica limpia sin números.

ORDENAR Y FILTRAR

Al igual que en tablas y grupos de celdas, Excel provee herramientas para ordenar y filtrar datos en tablas Pivot. Hay opciones simples y otras más complejas para escoger (filtrar) valores, ya sean textuales (*Label Filters*) o numéricos (*Value Filters*).



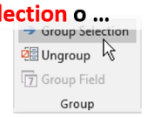
AGRUPAR DATOS

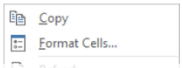
Excel permite agrupar datos para hacer cálculos a modo de subtotales con los mismos.

1: Escoger las celdas a agrupar

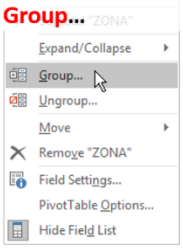
POP_EST2016	(All)
Row Labels	Sum of POP_EST2016
Centro	202,174
Centro-Este	451,902
Centro-Oeste	83,174
Este	163,138
Metro	1,135,058
Noreste	133,565
Noroeste	179,184
Norte	364,146
Oeste	141,222
Sur	478,533
Sureste	65,311
Suroeste	131,978
Grand Total	3,529,385

2: Hacer click en botón Group Selection o ...





2: ... right click y escoger la opción Group...



3: Aparecerá Group1 en la cabecera de filas (row)

POP_EST2016	(All)
Row Labels	Sum of POP_EST2016
Group1	737,250
Centro	202,174
Centro-Este	451,902
Centro-Oeste	83,174
Este	163,138
Este	163,138
Metro	1,135,058

4: Puede cambiar el nombre por defecto a uno de mayor significado

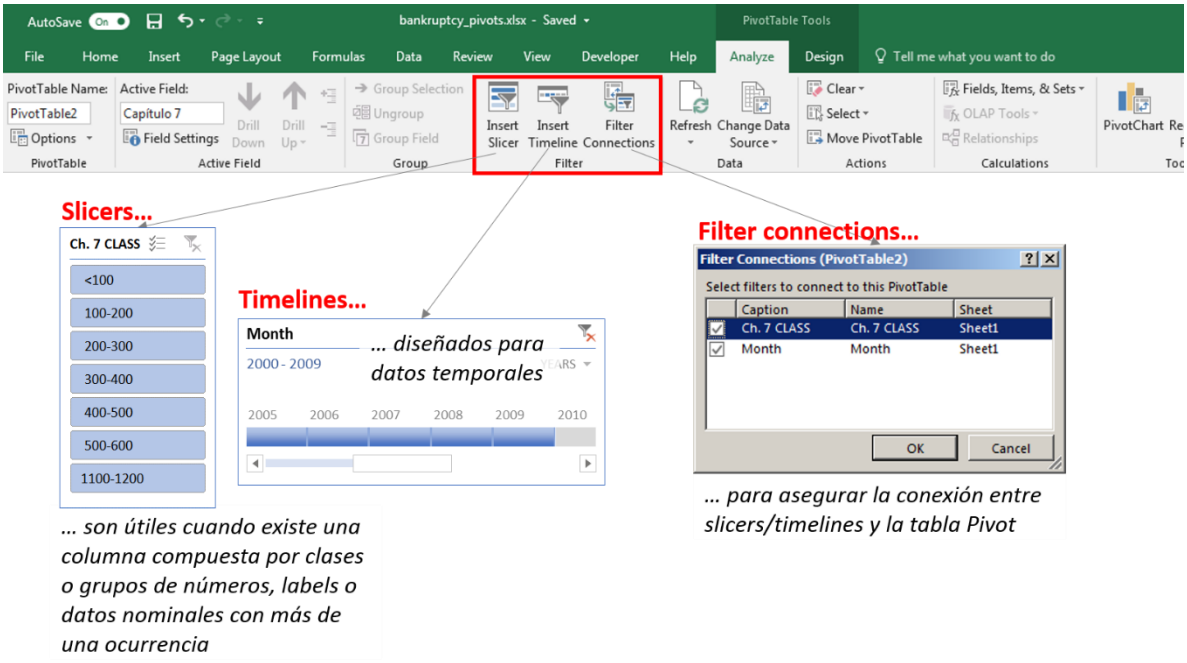
POP_EST2016	(All)
Row Labels	Sum of POP_EST2016
Interior montañoso	737,250
Centro	202,174
Centro-Este	451,902
Centro-Oeste	83,174
Este	163,138
Este	163,138
Metro	1,135,058

De manera análoga, podemos **desagrupar** (*Ungroup*) el grupo que hayamos creado usando el botón **Ungroup...**

o hacer **right click** en la celda con agrupación y escoger la opción **Ungroup...**

SLICERS Y TIMELINES

Los slicers y timelines son filtros gráficos interactivos. Como aparece en la figura siguiente, los slicers tienen utilidad cuando los datos a filtrar son discretos, son alfanuméricos o están agrupados.



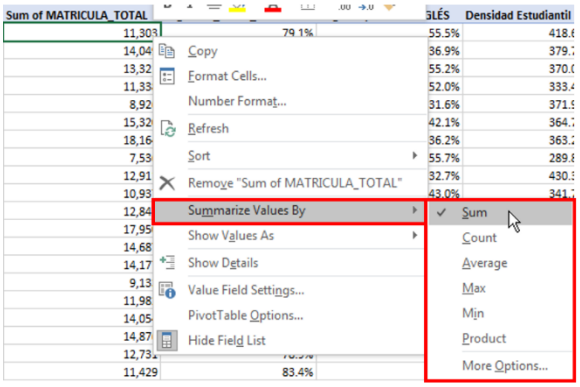
Para este ejemplo se calculó una columna adicional en la tabla Excel para guardar valores de clases que agruparan la cantidad de presentaciones de quiebras por Capítulo 7 al tribunal federal.

CÁLCULO DE VALORES Y DE COLUMNAS

En esta sección discutiremos varias de las opciones para calcular valores de las columnas, así como generar columnas nuevas en una tabla Pivot.

SUMMARIZE VALUES BY…

Resume los datos por la o las dimensiones escogidas. Los resúmenes pueden ser sumas, promedios, conteos y otros.



Summarize Values By: determina cómo se van a tratar los números y con cuál función agregarlos, ya sea mediante SUM, COUNT, AVERAGE, MAX, MIN, etc.

Advertencia: Cuando hayan celdas vacías o celdas con texto, Excel calcula por defecto el conteo de celdas (COUNT) en vez de sumar.
Esto se puede arreglar,
1) volviendo a los datos y cotejar si hay alguna falta o error.
2) También es posible "forzar" la suma mediante el uso de **Summarize Values By | Sum**

Como dice la advertencia, tenga cuidado con las sumas o conteos que Excel haga por defecto. En ocasiones, hay celdas con alguna falta o error. Puede haber una celda vacía o una celda con valores alfanuméricos donde debería haber un valor numérico. En ese caso, debe cotejar para arreglarlo si es posible. De lo contrario, estar consciente de que deberá forzar la suma cuando aplique.

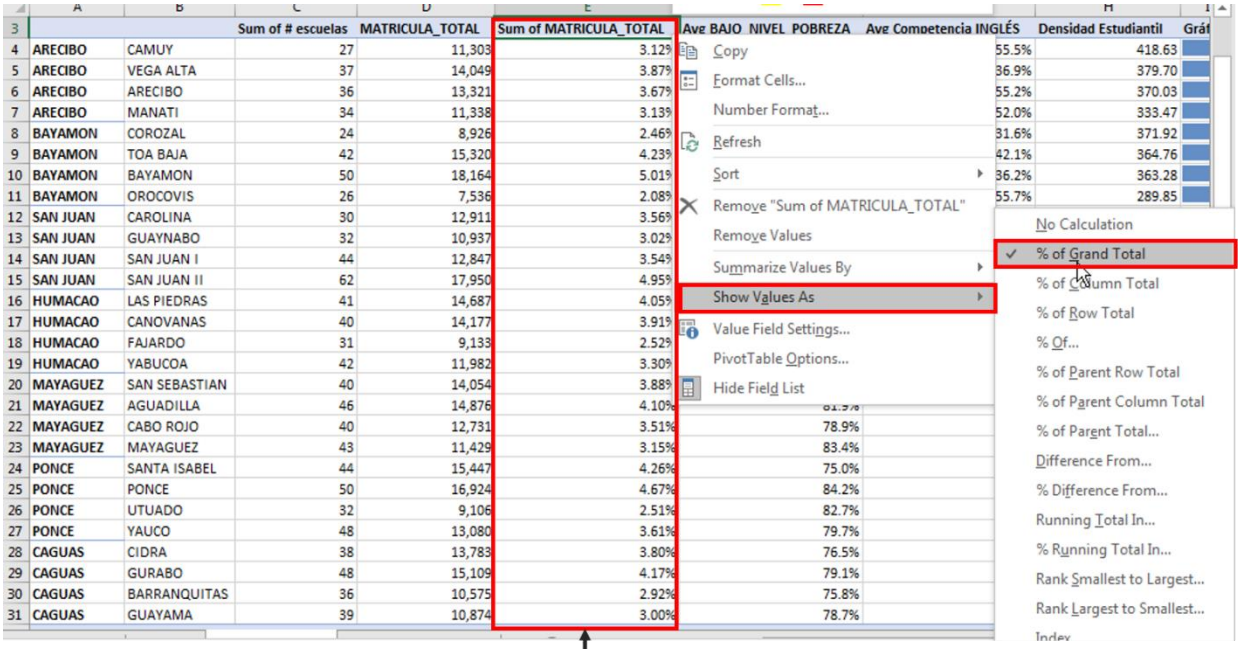
SHOW VALUES AS

Esta serie de funciones contiene funciones analíticas de gran importancia porque pueden mostrar los datos según su contribución a un gran total, al total de un grupo, diferencia entre fechas, running total (total movible), rank (posición), índices.

Se trata de una manera cómoda y rápida de analizar los datos desde diferentes ángulos sin tener que utilizar funciones ni calcular datos en el data range (datos crudos) o en la tabla Excel.

PERCENT OF GRAND TOTAL:

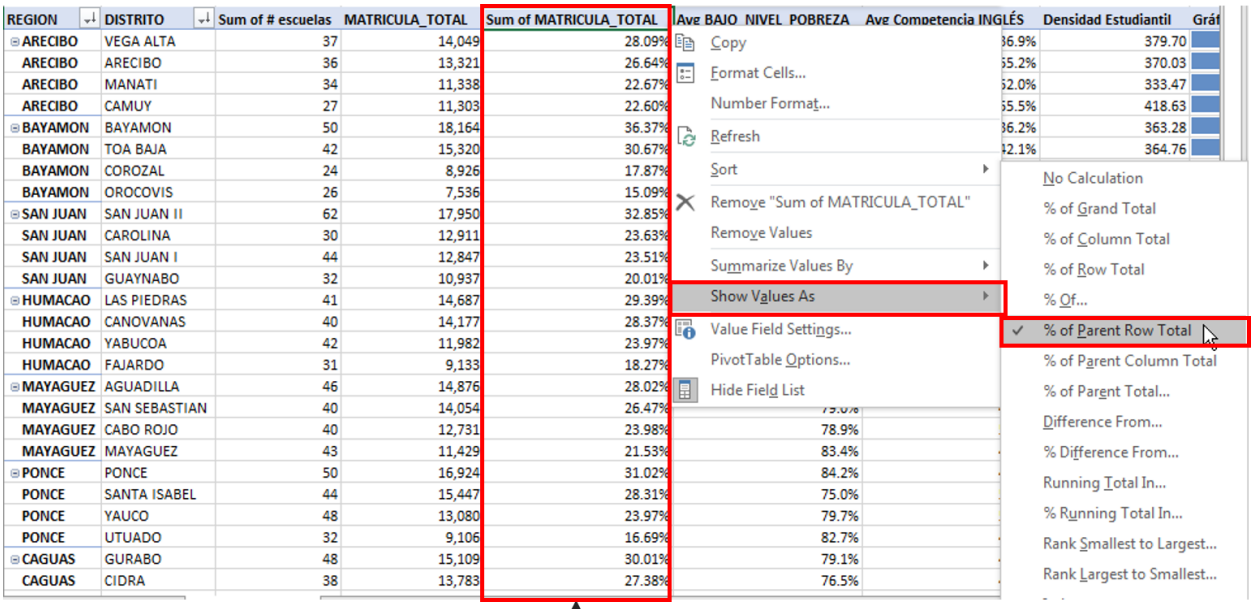
Muestra la participación de matrícula de cada distrito con relación al total. Para este ejemplo, los valores **% of Grand Total** coinciden con **% of Column Total**



Show Values As: En este caso, estamos mostrando cuál es la participación de cada escuela dentro del total de matrícula de las más de 1,000 escuelas en este directorio.

Percent of Parent Row Total

Con esta opción, Excel devuelve el cómputo de la participación de cada DISTRITO dentro de la REGIÓN educativa o *Parent Row* en este caso:



Show Values As: % of Parent Row Total

En este caso, estamos mostrando *cuál es la participación de cada DISTRITO* dentro del total de matrícula de cada REGIÓN.

Otros ejemplos de Show values as

Aquí mostramos otros ejemplos dentro de las opciones incluidas en la lista ***Show Values As***. Estamos usando datos de casos de solicitudes de permisos de construcción en varios años sometidos a la Oficina de Gerencia de Permisos (OGPe), mostrando el costo estimado del proyecto.

			No calculation	*% of Column Total*	*% of Parent Row Total*	*% of Percent Difference from… (Years)*	*Running Total In… (Years)*	*Rank Largest to Smallest*
5	Región	Years	Costo Estimado Proyectos	Costo Estimado2	Costo Estimado3	Costo Estimado4	Costo Estimado5	Costo Estimado6
6	**Aguadilla**		**$273,009,983**	**18.03%**	**18.03%**			**3**
7		2013	$103,951,243	6.87%	38.08%		$103,951,243	4
8		2014	$128,683,130	8.50%	47.13%	23.79%	$232,634,373	2
9		2015	$40,375,609	2.67%	14.79%	-68.62%	$273,009,983	3
10	**Arecibo**		**$250,253,371**	**16.53%**	**16.53%**			**4**
11		2013	$119,794,314	7.91%	47.87%		$119,794,314	3
12		2014	$95,348,841	6.30%	38.10%	-20.41%	$215,143,155	3
13		2015	$35,110,215	2.32%	14.03%	-63.18%	$250,253,371	4
14	**Central**		**$448,992,237**	**29.66%**	**29.66%**			**1**
15		2013	$200,448,853	13.24%	44.64%		$200,448,853	1
16		2014	$131,917,467	8.71%	29.38%	-34.19%	$332,366,320	1
17		2015	$116,625,916	7.70%	25.98%	-11.59%	$448,992,237	1
18	**Humacao**		**$339,860,954**	**22.45%**	**22.45%**			**2**
19		2013	$173,524,073	11.46%	51.06%		$173,524,073	2
20		2014	$92,884,783	6.14%	27.33%	-46.47%	$266,408,856	4
21		2015	$73,452,098	4.85%	21.61%	-20.92%	$339,860,954	2
22	**Ponce**		**$201,757,384**	**13.33%**	**13.33%**			**5**
23		2013	$96,313,551	6.36%	47.74%		$96,313,551	5
24		2014	$82,301,193	5.44%	40.79%	-14.55%	$178,614,744	5
25		2015	$23,142,640	1.53%	11.47%	-71.88%	$201,757,384	5

Las diferentes opciones nos muestran:

No calculation	La **suma** del Costo Estimado en dólares por región OGPe
% of Column Total	**Participación** de cada región por año **dentro del total** de la **columna** Costo Estimado.
% of Parent Row Total	**Participación** de cada región por año **dentro del total de la fila de origen** (**parent**) en este caso, los años dentro de la Región OGPe.
% Difference from…	**Diferencia porcentual** dentro de cada región entre años, usando la opción *previous*.
Running Total In…	**Totales, según se van acumulando** por años.
Rank Largest to Smallest	**Escala ordinal** de costo estimado entre regiones, **ordenado de mayor a menor.** Si se usa la opción por años, los resultados serán diferentes.

Show values as… Index

Esta opción la presentamos aparte para entender cómo trabajaría con el conjunto de datos previamente usado de Solicitudes de permisos de construcción. Le aplicamos ***Conditional formatting*** a las celdas para hacer resaltar las celdas con valores extremos más altos o más bajos.

Costo Estimado Proyectos	Región				
Years	Aguadilla	Arecibo	Central	Humacao	Ponce
2013	0.83	1.04	0.97	1.11	1.04
2014	1.34	1.09	0.84	0.78	1.16
2015	0.78	0.74	1.36	1.13	0.60

En este caso, los costos estimados se están viendo como índices por año y por región de la OGPe. Como podemos ver, los índices más altos están en la región Central (San Juan) para el año 2015 y en Aguadilla para el año 2014. Los índices más bajos están en Ponce (0.60) y en Arecibo (0.74) para el año 2015.

Los índices son estimados ponderados. Si tomáramos en cuenta solamente los costos estimados y le aplicamos ***Conditional Formatting***, veríamos algo diferente, que puede o no coincidir. Fíjese en los valores crudos de 2013 en la región Central. En cuanto al índice, es uno bajo, pero al ponderar entre filas y columnas, su efecto es mucho menor.

Costo Estimado Proyectos	Región					
Years	Aguadilla	Arecibo	Central	Humacao	Ponce	Grand Total
2013	$103,951,243	$119,794,314	$200,448,853	$173,524,073	$96,313,551	$694,032,034
2014	$128,683,130	$95,348,841	$131,917,467	$92,884,783	$82,301,193	$531,135,415
2015	$40,375,609	$35,110,215	$116,625,916	$73,452,098	$23,142,640	$288,706,479
Grand Total	**$273,009,983**	**$250,253,371**	**$448,992,237**	**$339,860,954**	**$201,757,384**	**$1,513,873,928**

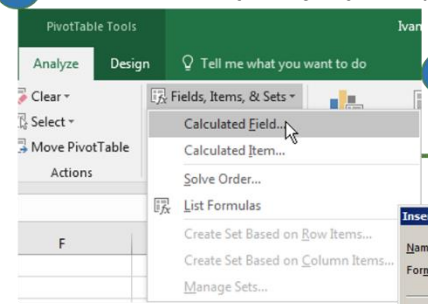
Los **índices** son calculados de la siguiente manera:

$$\frac{Valor\ de\ la\ celda * Gran\ Total}{(Total\ de\ la\ fila * Total\ de\ la\ columna)}$$

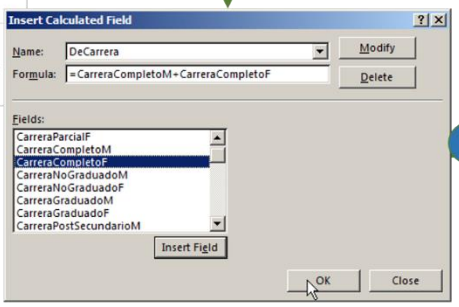
CAMPOS CALCULADOS

A menudo tenemos columnas con conteos divididos en las cuales queremos hacer ciertas operaciones aritméticas, tales como sumar, restar, multiplicar o dividir. En una tabla Pivot podemos generar nuevas columnas que contengan cómputos basados en otras columnas. En este ejemplo, tenemos el **Registro Anual de Puestos por Demografía**. En este registro se dividen los conteos por género. Podemos sumar conteos sin afectar los datos originales. A continuación, mostramos cómo producir una columna nueva para una tabla Pivot:

1 **Pivot Table Tools | Analyze | Fields, Items, & Sets | Calculated Field…** Para crear una nueva columna (field)



2 **Insert Calculated Field**
Name: Nuevo nombre de columna
Formula: =escribir nueva fórmula o modificar existente



3 **Resultados**

1		
2	AnoFiscal	2012-2013
3		
4	Row Labels	Sum of DeCarrera
5	Corporaciones públicas	35,200
6	Departamentos y Agencias	100,643
7	Municipios	26,600
8	Oficina del Gobernador	1,069
9	Otras entidades	344
10	Rama Judicial	3,568
11	Rama Legislativa	529
12	Grand Total	167,953

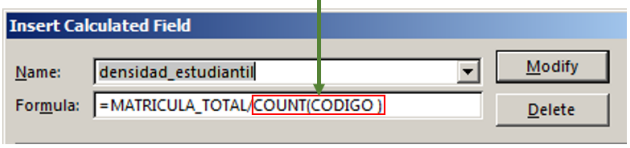
CÁLCULOS BASADOS EN CONTEOS

En PivotTables, las operaciones numéricas normales están restringidas a campos numéricos basados en sumatorias; no en conteos.

Siguiendo el tema de campos calculados, ¿qué pasaría si queremos utilizar un campo basado en conteos en lugar de estar basado en sumatorias para hacer operaciones aritméticas? Al hacer resúmenes por dimensión, (región, clase, tipo, etc.) lo hacemos generalmente mediante sumas y cómputos agregados como promedios. En ocasiones, queremos utilizar algún campo basado en conteos para calcular otros campos.

El siguiente diagrama muestra primero un ejemplo de una falta al no considerar que un campo es basado en conteos.

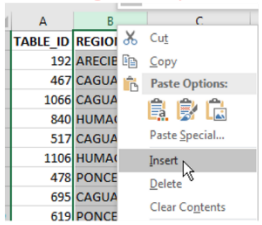
Falta al calcular columna:



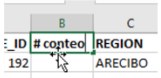
Debe existir una columna que registre los conteos (basada en una medición) para que funcione correctamente.

Este es el **procedimiento recomendado**: **crear columna nueva en la tabla de datos**.
¿Cómo hacerlo?

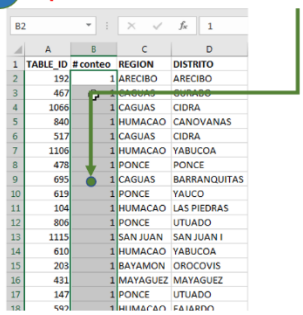
1 Crear columna nueva usando right click | Insert



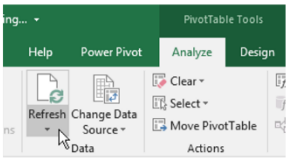
2 Ponerle nombre a la columna



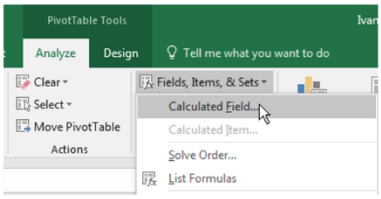
3 Copiar el número 1 … en todas las celdas de esta columna



4 ***Regresar a la tabla pivot…***
Pivot Table Tools | Analyze | Refresh



5 Pivot Table Tools | Fields, Items & Sets | Calculate Field…



6 Calcular la columna

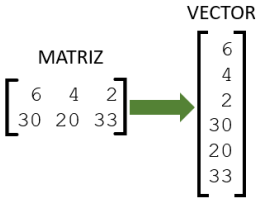
En este ejemplo se genera una columna con valores 1. Excel generará una suma y esta nueva columna podrá ser usada para cómputos aritméticos. En este caso: promedio de matrícula por escuela.

Resultados:

REGION	Count of CODIGO	Sum of MATRICULA_TOTAL	Sum of Densidad Estudiantil
ARECIBO	134	50,011	373
BAYAMON	142	49,946	352
CAGUAS	161	50,341	313
HUMACAO	154	49,979	325
MAYAGUEZ	169	53,090	314
PONCE	174	54,557	314
SAN JUAN	168	54,645	325
Grand Total	**1102**	**362,569**	**329**

CONSOLIDAR COLUMNAS (VECTORIZACIÓN)

En ocasiones, ***una misma métrica está dividida por años o alguna otra subclase***. Por lo tanto, es preferible **consolidar varias columnas que tratan sobre el mismo tema en una sola columna**. Este concepto es llamado ***vectorización***, en términos de ***Data Science***. En este proceso, como en álgebra lineal, se convierte una matriz en un vector.

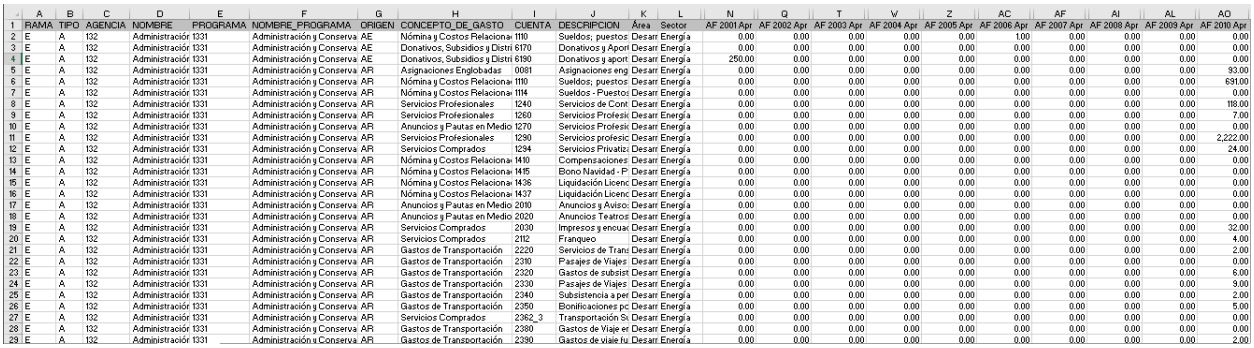


*Un **vector** es una **secuencia de datos de un mismo tipo**. Los miembros en un vector son llamados componentes.*

Para **programas de manejo de datos y visualización** como **Excel**, **Power BI**, **Tableau**, etcétera, puede trabajarse de manera más eficiente con datos vectorizados.

Podemos trabajar un ejemplo que puede ser una tabla presupuestaria, la cual tiene múltiples dimensiones y una sola métrica divida en años u otras características de ese y otros años.

Esta es una hoja derivada de una consulta (*SQL Query*) a la base de datos presupuestarios de la Oficina de Gerencia y Presupuesto. El resultado del query vuelve el presupuesto aprobado para los años 2001 al 2016 de cada agencia, sus características, los conceptos de gasto y origen de fondos. La métrica viene siendo la misma: ***fondos***, divididos por año.



Para un mejor uso de las funciones PivotTable, vamos a combinar las columnas presupuestarias separadas (que es una sola métrica dividida por años), consolidándolas en dos columnas: una contendrá los presupuestos (*Value*) y la otra contendrá el año fiscal (*Attribute*). En Excel prepararemos el archivo recibido y traspondremos dichas columnas.

En lugar de abrir directamente el archivo de hoja de cálculo presupuestario, usaremos la interfaz Query Editor de Excel para preparar los datos antes de traer la hoja Excel con los datos. No haremos cambios al archivo original.

Trasponer usando Unpivot Columns

Query Editor: Unpivot columns

- Combinar múltiples columnas en una.
  - Ejemplo: columnas 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017.
    - Combinarlas en una sola
- CUIDADO: Esto multiplicará el número de filas (rows) multiplicándolas por cada columna
- Importar datos a Excel usando Data | Get Data | From File | From Workbook

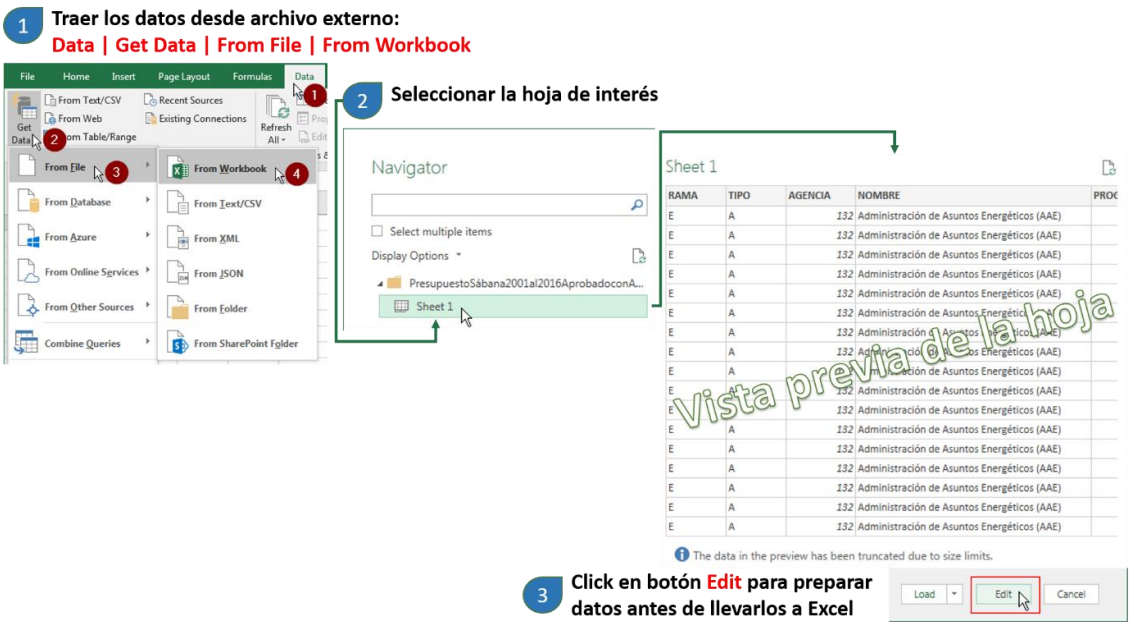
Los pasos para traer los datos y trasponer (*unpivot*) las columnas son los siguientes:

Hay dos procesos principales:

1. **Importar el archivo Excel** original usando **Data | Get Data… From Workbook**
2. Usar la interfaz **Query Editor** de Excel para **escoger las columnas a consolidar y aplicar la herramienta tab Transform | Unpivot Column**.

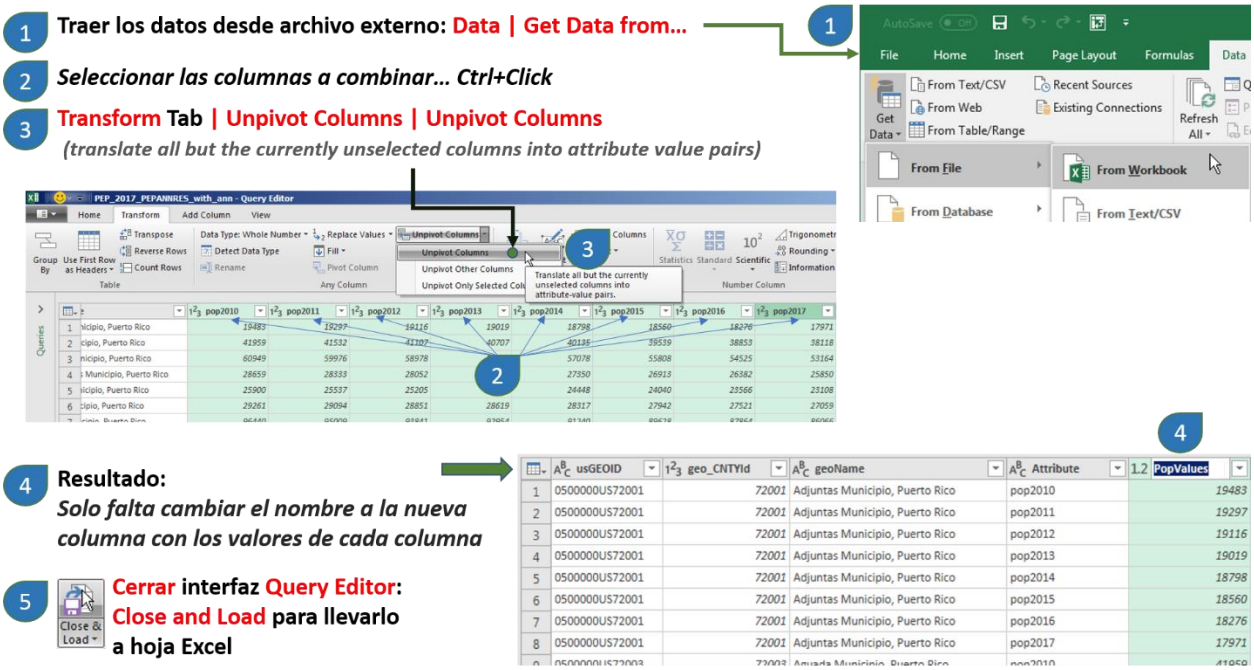
Proceso 1: Importar datos:

Extraer datos desde otro archivo. En este caso, un workbook Excel.



Proceso 2: Usar Query Editor

Importar datos (Extract, Transform, Load: ETL) y traerlos a Excel.

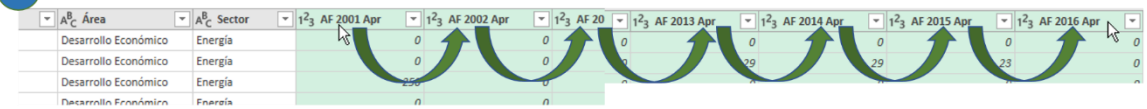


Luego aceptamos los cambios y cerramos la interfaz Query Editor para ver el resultado. Este será una Tabla Excel.

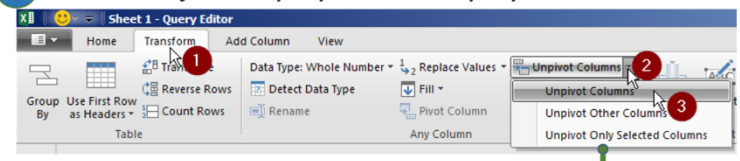
Abrirá entonces la interfaz **Query Editor** de Excel. En esta interfaz haremos la trasposición de columnas. Aquí se puede realizar cambios a los datos, eliminar columnas, filas, cambiar el tipo de dato en columnas, realizar cómputos… etcétera. Solo usaremos la función **Unpivot Columns**.

TRANSFORM | UNPIVOT COLUMNS

A ***Ctrl+Click en cada cabecera de columna AF 2001 al 2016 Apr***



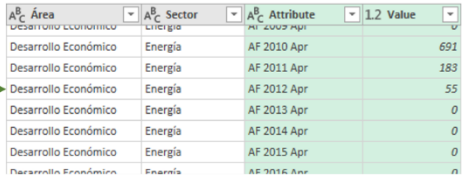
B ***Click en Transform tab | Unpivot Columns | Unpivot Columns***



Nota: Aunque vamos a *trasponer columnas*, NO vamos a usar el botón Transpose.
Sólo vamos a trasponer 16 de las columnas presupuestarias (métricas) y dejaremos las dimensiones (las otras columnas descriptivas) como están. Por esto usamos el botón **Unpivot Columns.**

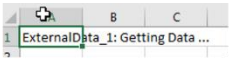
D ***Cambie** los **nombres** de las **columnas**:*
Attribute** → **Año
Value** → **Presupuesto

C ***Resultados...***
Las 16 columnas fueron reducidas a dos: Attribute y Value



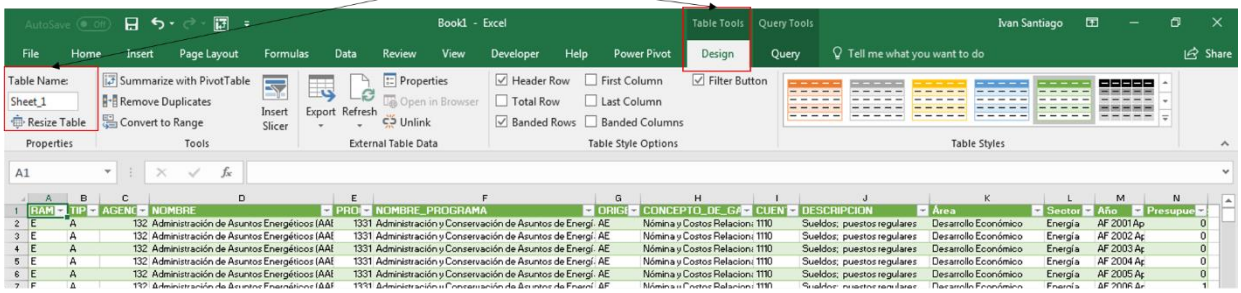
D ***Para ejecutar cambios, Haga click en: Home tab | Close & Load***

Al cerrar la interfaz **Query Editor**, el **resultado** será una **tabla Excel**. Esta tendrá sus filas (rows) multiplicadas por cada columna que consolidamos.



Espere el procesamiento de los datos

La Tabla resultante consiste de 753,536 filas y 14 columnas



Al hacer la trasposición usando Unpivot Columns, consolidamos los años 2001 al 2016 en una sola columna, repitiendo filas por cada año.

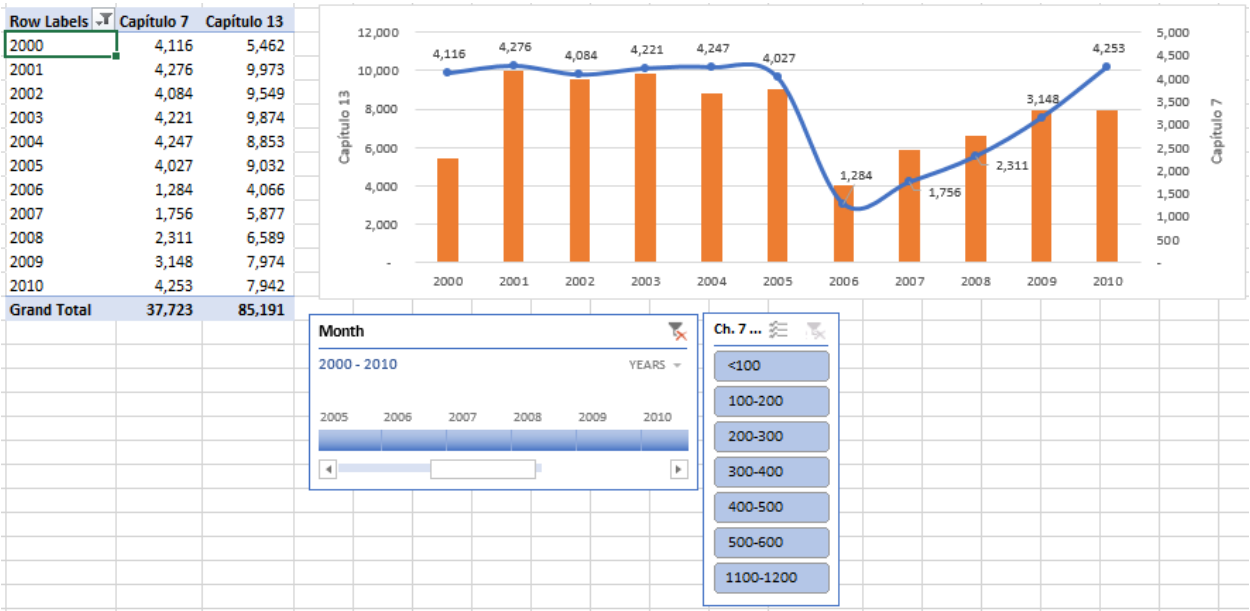
El archivo original contiene **47,096** filas y **58** columnas...

Esta tabla ya está lista para usar las funciones de PivotTable. Usaremos este archivo para uno de los ejercicios.

PIVOTCHARTS

Una gráfica Pivot es igual a cualquier otra gráfica Excel, salvo que está directamente atada a una tabla Pivot. La gráfica Pivot se irá ajustando conforme se le apliquen filtros y campos a la tabla Pivot a la cual esté enlazada.

En este ejemplo arriba, mostramos en una gráfica Pivot las **solicitudes de quiebras ante el tribunal federal, distrito** de **San Juan** por los **capítulos 7 y 13**. Se aplicaron filtros seleccionando solamente los grupos <100, 100-200 y 200-300, y para los años 2000 al 2006.



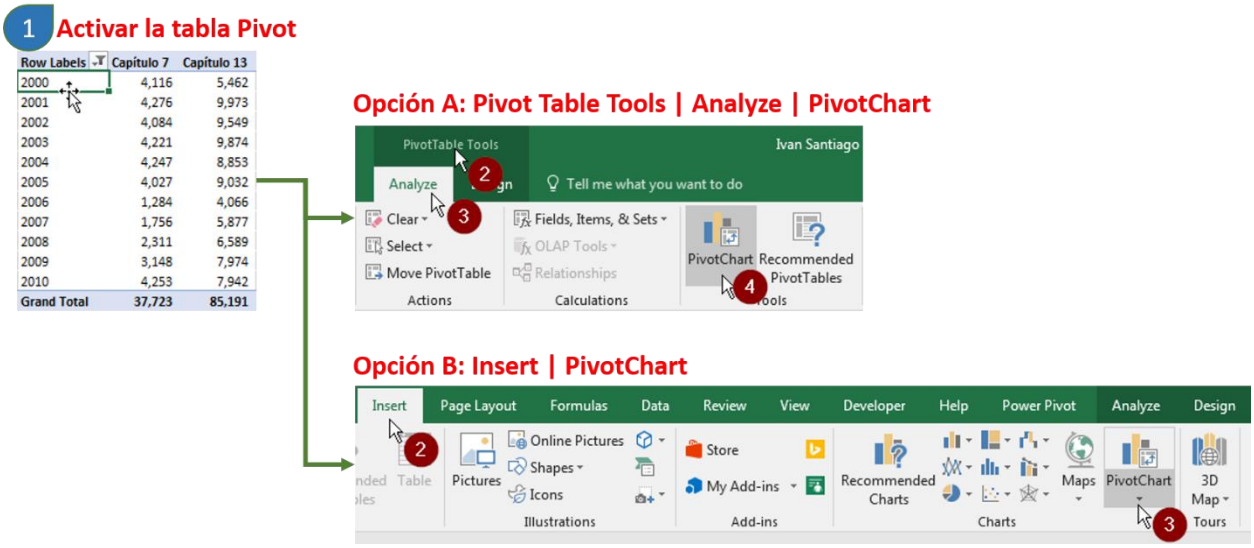
La misma gráfica puede mostrar ahora los casos sometidos para todas las clases desde <100 a 1,200, además de incluir los años 2000 a 2010.

CÓMO AÑADIR UNA GRÁFICA PIVOT

Excel provee dos opciones para añadir una gráfica Pivot. Luego de haber activado la tabla Pivot, puede añadir una gráfica Pivot desde el menú principal:

A: PivotTable Tools | Analyze | PivotChart

B: Insert | PivotChart



Opciones disponibles para formatear gráficas Pivot

Cuando una gráfica Pivot es activada, se activan a su vez una serie de pestañas (tabs) y botones con diferentes funciones. Estas funciones van desde las relacionadas a:

- **Analysis tab**: aplicar filtros, slicers, timelines, conexiones a tablas Pivot, y eliminar de la gráfica los botones de filtros, entre otras.
- **Design tab**: Cambiar cómo luce la gráfica Pivot. Tal como se hace con las gráficas regulares en Excel.
- **Format tab**: Ir más al detalle de los elementos de la gráfica para hacerle cambios individuales a diferentes elementos que la componen.

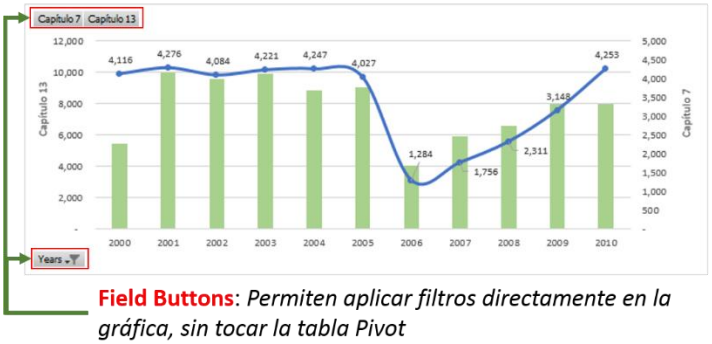
Analyze tab

Design tab

Format tab

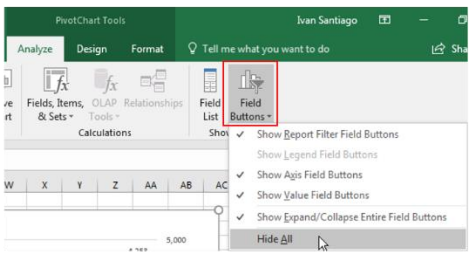
Field Buttons

Por defecto, aparecerán botones correspondientes a las dimensiones que aparecen en la gráfica Pivot. Estos botones son para filtrar datos o quitar un campo para moverlo al área de filtros.



Field Buttons: *Permiten aplicar filtros directamente en la gráfica, sin tocar la tabla Pivot*

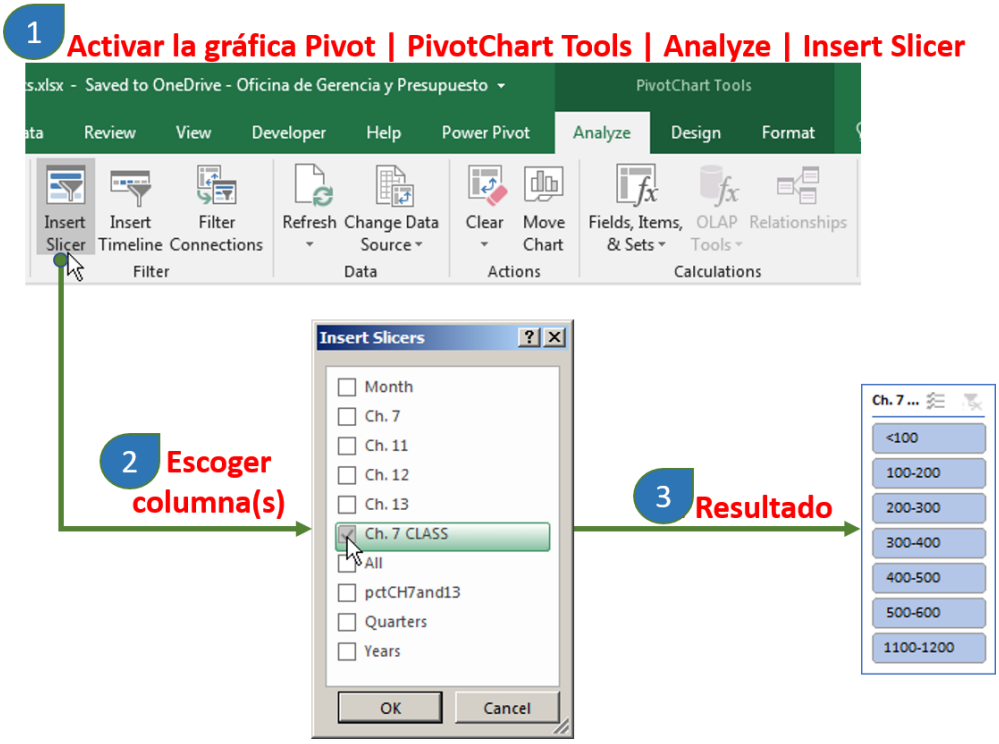
Field Buttons: *Pueden ser apagados o encendidos desde el grupo Analyze | Botón Field Buttons*



Los botones se pueden apagar o mostrar en la gráfica desde el menú **Analyze | Field Buttons**.

SLICERS

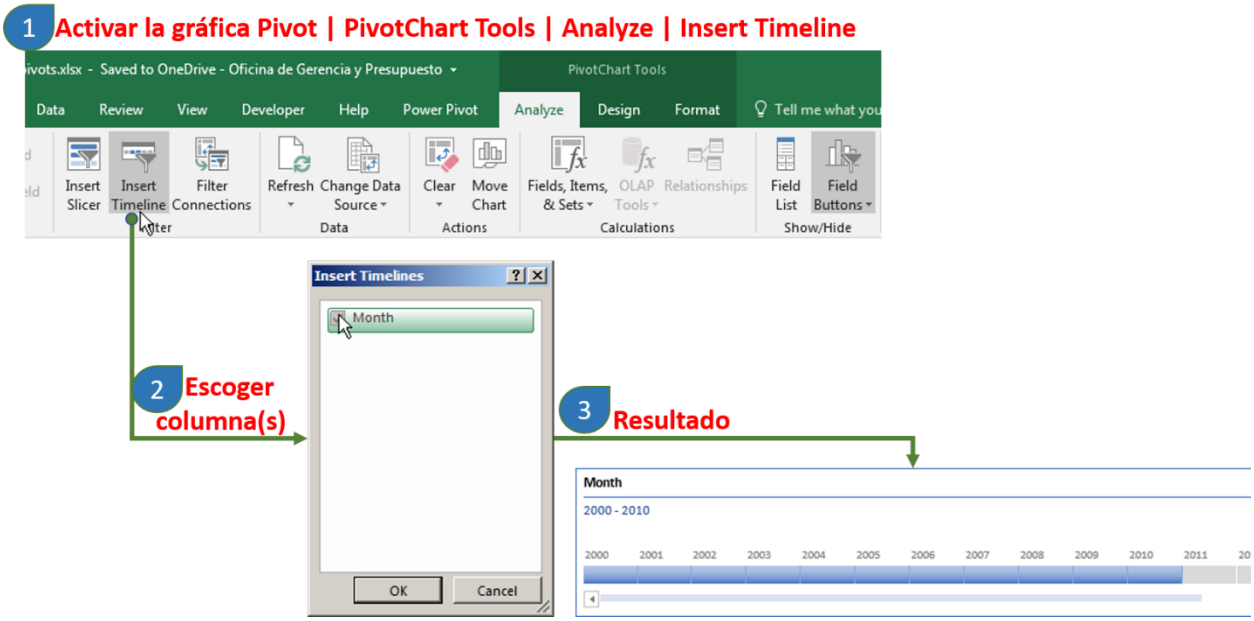
Los slicers son filtros interactivos. Contienen botones que corresponden a las dimensiones que deseamos aplicar como filtros. Los slicers se añaden desde **PivotChart Tools |** tab **Analyze |** botón **Insert Slicer**.



Es más práctico crear slicers para datos *dimensionales* (categorías). Si va a usar mediciones, recomendamos dividirlos en clases (usar funciones LOOKUP) para evitar tener slicers con listas demasiado largas haciéndolos inmanejables.

TIMELINES

Los timelines son slicers para usarlos con datos temporales (fechas, tiempo). Para añadir un timeline, active la gráfica Pivot, y haga **click** en el **botón Insert Timeline**, localizado en el **grupo Filter**, dentro del **tab Analyze** del menú **PivotChart Tools**.



PRÁCTICAS: CREAR PIVOTTABLES Y PIVOTCHARTS

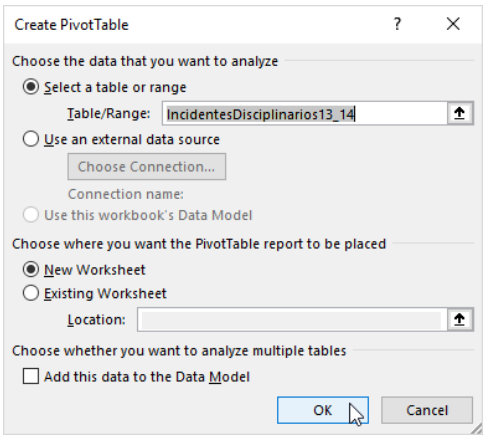
En esta parte vamos a practicar con algunos conjuntos de datos que se han presentado anteriormente como ejemplos de PivotTables y PivotCharts.

EJERCICIO 1: CREAR PIVOTTABLE

Objetivo: generar tabla Pivot utilizando datos crudos en formato Excel, previamente descargados desde el portal http://data.pr.gov

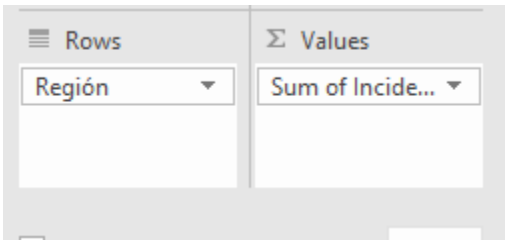
Abra el workbook **IncidentesDisciplinariosEscuelasDE2013-14.xlsx**, localizado en **C:\Tutorial_Excel\DataAnalysis\Modulo_3**.

Active la **tabla** haciendo **click** y vaya al **menú principal, Insert | PivotTable**. En la forma **Create PivotTable**, en el apartado **Choose where you want the PivotTable report to be placed**, escoja la opción **New Worksheet**.



A la **izquierda** aparecerá la tabla vacía **PivotTable1**. A la **derecha** aparecerá la forma/interfaz **PivotTable Fields**.

En la forma **PivotTable Fields**, **arrastre** el campo **Región** en el cuadro **Rows**. **Arrastre** luego el campo **Incidentes** al cuadro **Values**.

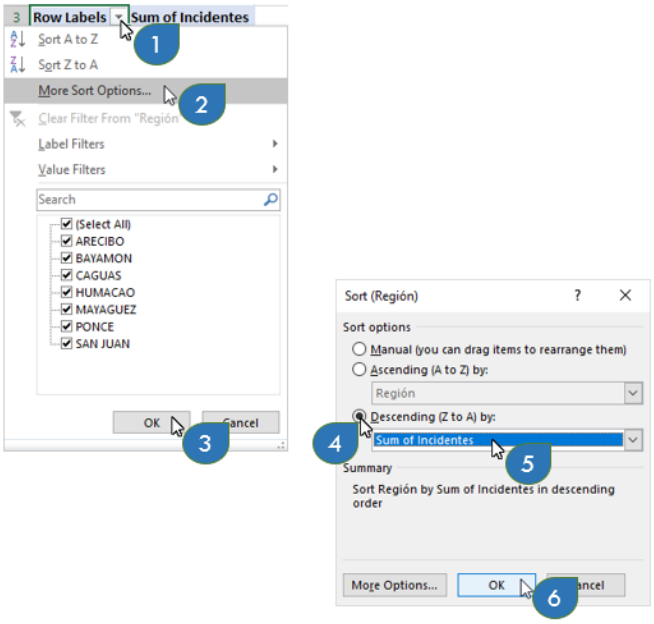


La tabla Pivot deberá aparecer así. Estas son las sumas de distintos incidentes disciplinarios sumados por región.

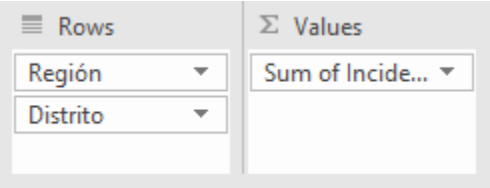
Row Labels	Sum of Incidentes
ARECIBO	15
BAYAMON	18
CAGUAS	132
HUMACAO	87
MAYAGUEZ	12
PONCE	70
SAN JUAN	100
Grand Total	**434**

¿Cuál es la región con más incidentes disciplinarios registrados? _____

(Puede **ordenar** los datos en la columna **Sum of Incidentes** para saber cuál región tiene más incidentes) Siga la siguiente secuencia:



Ahora, **arrastre** la columna **Distrito *debajo* de** la columna **Región** en la caja **Rows**.



La tabla pivot deberá verse así:

Row Labels	Sum of Incidentes
CAGUAS	**132**
BARRANQUITAS	15
CIDRA	13
GUAYAMA	42
GURABO	62
SAN JUAN	**100**
CAROLINA	27
GUAYNABO	32
SAN JUAN I	26
SAN JUAN II	15
HUMACAO	**87**
CANOVANAS	25
FAJARDO	8
LAS PIEDRAS	29
YABUCOA	25
PONCE	**70**
PONCE	46
SANTA ISABEL	22
UTUADO	1
YAUCO	1
BAYAMON	**18**

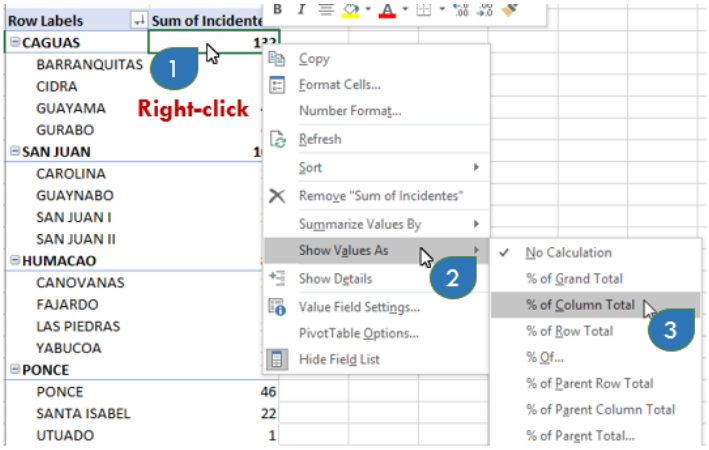
La tabla muestra ahora más detalle. Esta vez desglosa los incidentes por ***región educativa y distrito***.

* Un *distrito escolar* puede estar compuesto de uno o más municipios.

Mostrar valores como porcentajes (Column Total)

Haga **right click** en **cualquier celda** debajo de la columna **Sum of Incidentes**.

En el menú de contexto, escoja la opción **Show Values As** y luego **click** en la opción **% of Column Total**.



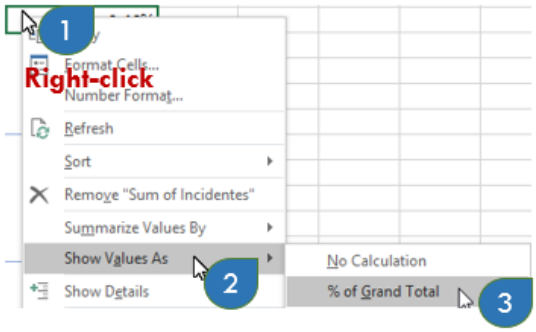
La tabla Pivot debe verse así:

Row Labels	Sum of Incidentes
CAGUAS	**30.41%**
BARRANQUITAS	3.46%
CIDRA	3.00%
GUAYAMA	9.68%
GURABO	14.29%
SAN JUAN	**23.04%**
CAROLINA	6.22%
GUAYNABO	7.37%
SAN JUAN I	5.99%
SAN JUAN II	3.46%
HUMACAO	**20.05%**
CANOVANAS	5.76%
FAJARDO	1.84%
LAS PIEDRAS	6.68%
YABUCOA	5.76%
PONCE	**16.13%**

Fíjese en lo que muestra la tabla. Aquí aparecen las **proporciones en por cientos por el total en cada columna**. Cada **región educativa** y **distrito** muestran su **proporción con relación al valor total de esta columna (Cantidad de Incidentes)**.

En este conjunto de datos **no habrá diferencia si usamos** las opciones **% of Grand Total** y **% of Column Total**.

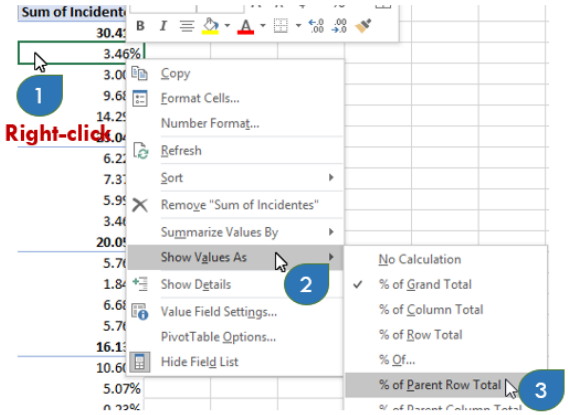
Haga la prueba escogiendo la opción **% of Grand Total** en la columna **Sum of Incidentes**.

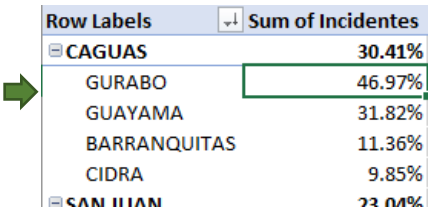


Ver participación de cada distrito dentro de la región

Hasta el momento, hemos visto la participación de cada distrito dentro del total de la columna. Ahora, vamos a mostrar la participación de cada distrito *dentro* de cada región.

Escoja ahora la opción % of Parent Row Total para que pueda ver la participación de cada distrito.





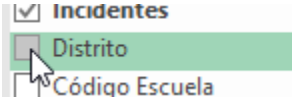
Row Labels	Sum of Incidentes
CAGUAS	30.41%
GURABO	46.97%
GUAYAMA	31.82%
BARRANQUITAS	11.36%
CIDRA	9.85%

Note que dentro de la región educativa de de Caguas, los incidentes se concentran en el distrito de Gurabo.

Ranks (posición)

Se puede preguntar, **cuál será el orden** de cantidad de incidentes por regiones. Esta pregunta puede contestarse usando la opción **Rank**

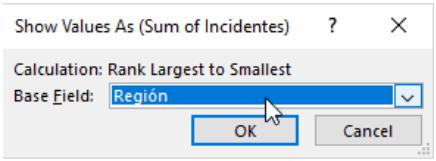
Primero, **remueva** los **distritos** de la caja Rows haciendo **uncheck** en el campo **Distrito**



Haga **right click** encima de cualquier celda de la columna **Sum of Incidentes**.

Escoja la opción **Show Values | Rank Largest to Smallest…**

Escoja el campo **Región** en la forma **Show values as (Sum of Incidentes)**



Como era de esperarse, así aparecen las posiciones.

Row Labels	Sum of Incidentes
CAGUAS	1
SAN JUAN	2
HUMACAO	3
PONCE	4
BAYAMON	5
ARECIBO	6
MAYAGUEZ	7

Cierre este workbook. Guarde los cambios usando el mismo nombre del archivo original.

En el próximo ejercicio, haremos el proceso de integrar varias columnas y compactarlas en dos columnas mediante vectorización (unpivot), usando datos presupuestarios.

Opcional: Contestar:

¿Cuál es el distrito escolar con mayor cantidad de incidentes disciplinarios? _____

¿Cuáles son las 10 escuelas con mayor cantidad de incidentes disciplinarios?

Ayuda: use el campo **Escuela_UniqueID** ☑ Escuela_UniqueID y aplique filtro **Value Filter | Top 10**

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

¿Cuáles son los 10 tipos de incidentes más frecuentes?

EJERCICIO 2: VECTORIZACIÓN O CONSOLIDACIÓN (UNPIVOT) DE MÚLTIPLES COLUMNAS

Como mencionamos anteriormente, cuando los datos están divididos en múltiples columnas, la preparación de cómputos para las opciones *Show Values As* son más tediosas o puedan dar resultados inconsistentes.

Para esta parte, usaremos un workbook con ***datos presupuestarios de la Oficina de Gerencia y Presupuesto*** para los años ***2001 al 2016***.

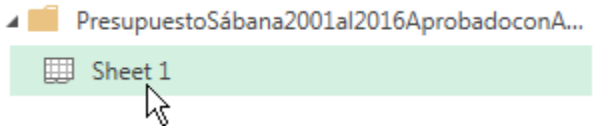
Abra un nuevo archivo workbook de Excel. New | Blank workbook.

Si está usando Office 365, vaya al **menú principal** y escoja **Data | Get data | From File | From Workbook**

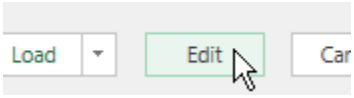
Si está usando Office 2016, deberá ir al **menú principal** y escoja **Data | New Query | Get data | From File | From Workbook**

Traiga (***Import***) el archivo **PresupuestoSábana2001al2016AprobadoconAreaySector.xlsx**, localizado en **C:\Tutorial_Excel\DataAnalysis\Modulo_3**.

En la forma **Navigator**, escoja (click) en la hoja **Sheet1**



Como vamos a hacer modificaciones al archivo resultante, haga **click** en el botón **Edit**.



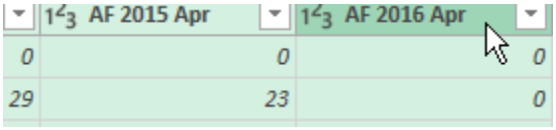
Aparecerá la interfaz **Query Editor** que mostrará el contenido del workbook a importar.

Navegue hacia la derecha para que pueda ver las columnas del ***presupuesto aprobado*** (**Apr**) para los años 2001 al 2016.

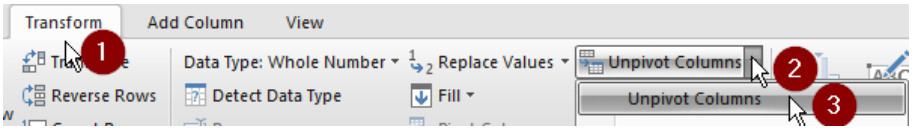
Vectorizar (unpivot) múltiples columnas

Como hemos mencionado antes, este proceso trata sobre integrar los valores guardados en varias columnas (16 en estos datos) y guardándolas en dos columnas. Una columna tiene la dimensión de los años 2001 al 2016 y la otra columna resultante tendrá los presupuestos aprobados que corresponden a cada año fiscal.

Haga **Shift+Click** en las **cabeceras** *headers* de las **columnas AF 2001 Apr** a la **columna AF 2016 Apr**. Aparecerán en azul indicando que están escogidas.



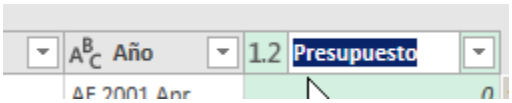
Vaya al **menú principal** de la Interfaz **Query Editor** y escoja **Transform | Unpivot Columns | Unpivot Columns**



Cambiar nombres de las columnas

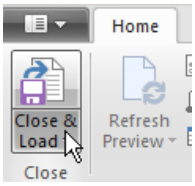
Haga **doble click** en el nombre de la columna **Attribute** y escriba **Año**.

Haga **doble click** en el nombre de la columna **Value** y escriba **Presupuesto**.



Aceptar cambios y cerrar la interfaz Query Editor

Para aceptar los cambios y cerrar la interfaz Query Editor, vaya al **menú principal** y escoja **Home** y haga **click** en el botón **Close and Load**



Esto iniciará el proceso de conversión de los datos para traerlos a una hoja de Excel.

El proceso tardará debido a que se están multiplicando las filas (rows) para cada año fiscal.

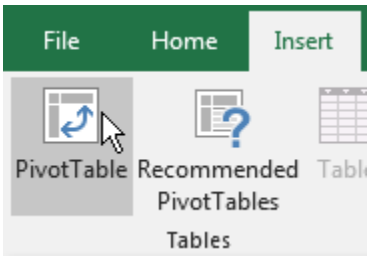
Inspeccione la hoja resultante:

Sector	Año	Presupuesto
Energía	AF 2002 Apr	0
Energía	AF 2003 Apr	0
Energía	AF 2004 Apr	0
Energía	AF 2005 Apr	0
Energía	AF 2006 Apr	0
Energía	AF 2007 Apr	0
Energía	AF 2008 Apr	0
Energía	AF 2009 Apr	0
Energía	AF 2010 Apr	0
Energía	AF 2011 Apr	0
Energía	AF 2012 Apr	0
Energía	AF 2013 Apr	29
Energía	AF 2014 Apr	29
Energía	AF 2015 Apr	23
Energía	AF 2016 Apr	0
Energía	AF 2001 Apr	250

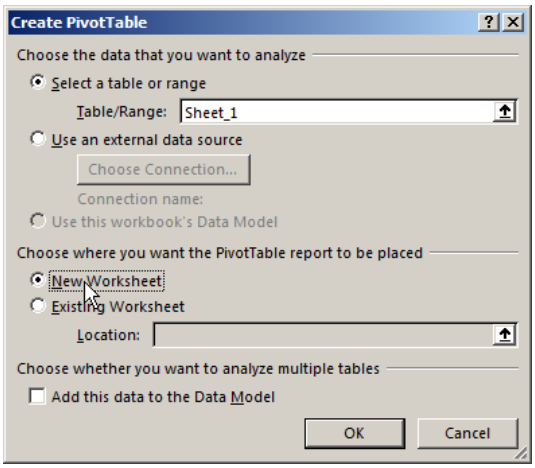
Notará además que **el resultado es** una ***tabla Excel***.

Generar PivotTable

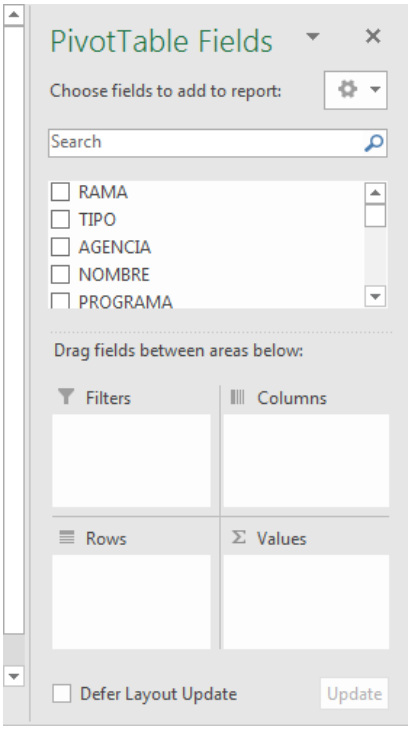
Haga **click** en **cualquier celda dentro de la tabla Excel**. Vaya al **menú principal** (ribbon) y haga **click** en el tab **Insert** y **click** en el botón **Pivot Table**, del **grupo Tables**.



En la forma **Create PivotTable**, bajo el apartado **Choose where you want the PivotTable report to be placed**, escoja la opción **New Worksheet**.



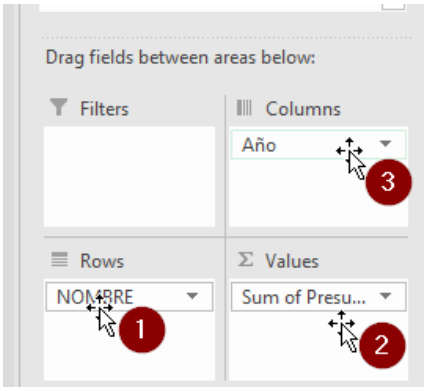
Aparecerán en la nueva hoja la **tabla pivot vacía** a la izquierda y la interfaz **PivotTable Fields**



GENERAR REPORTE PRESUPUESTOS POR AGENCIA

Para generar el primer reporte simple por agencia gubernamental, arrastre los campos (columnas)…

1. **Nombre:** ……. a la caja **Rows**
2. **Presupuesto:** … a la caja **Values**
3. **Año:** …… a la caja **Columns**



La tabla Pivot deberá verse así:

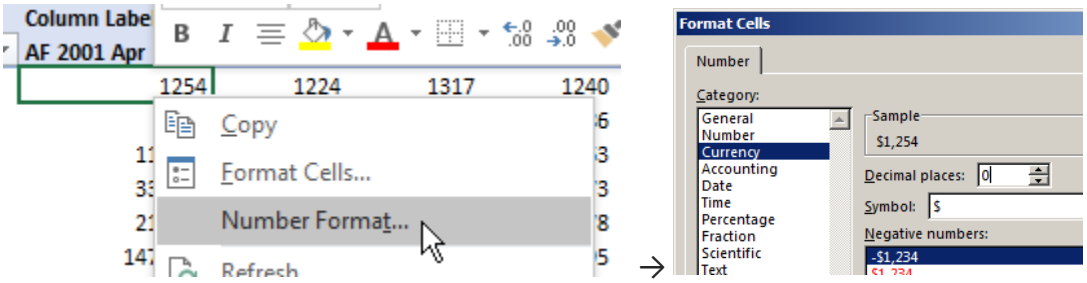
Sum of Presupuesto	Column Labels								
Row Labels	AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr	AF 2006 Apr	AF 2007 Apr	AF 2008 Apr	AF 2009 Apr
Administración de Asuntos Energéticos (AAE)	1254	1224	1317	1240	1696	1628	1627	1063	1405
Administración de Asuntos Federales de Puerto Rico	9350	9232	9245	9986	10257	7296	6255	6255	6044
Administración de Compensaciones por Accidentes de Automóviles	114629	85663	92890	103463	102780	102002	103586	102616	97758
Administración de Corrección	332724	315500	343525	383773	380794	478804	439558	403090	406069
Administración de Desarrollo Laboral	218099	262087	224984	94878	107560	95347	81476	107998	145707
Administración de Desarrollo Socioeconómico de la Familia	1475056	1505138	1575573	1597895	1686526	1704348	1743746	1813684	2049123
Administración de Desarrollo y Mejoras de Viviendas	15469	18541	19604	16717	13051	11568	0	0	0
Administración de Familias y Niños	317613	327545	361462	173327	234404	206573	217316	223591	225079
Administración de Fomento Comercial	13559	13201	12882	9676	0	0	0	0	0
Administración de Instituciones Juveniles	89972	82551	101580	83947	82358	91507	90578	80630	79175
Administración de la Industria y el Deporte Hípico	4009	4362	4701	4716	5073	3820	4656	5078	5483
Administración de Recursos Naturales	72858	74056	98524	68182	80008	121192	89339	69641	77629
Administración de Reglamentos y Permisos	21800	22588	22157	19514	19997	21362	21982	22930	23965

Cambiar formato a Currency

Haga **right click** en cualquiera de las celdas bajo la columna **AF 2001 Apr**.

Escoja la opción **Number Format…**. En la forma **Format Cells**, bajo la lista **Category**, escoja **Currency.**

Escriba **0** en **Decimal places** y haga **click** en el botón **OK** para aceptar los cambios.



Notará que todas las columnas de años fiscales en la tabla Pivot, cambiarán a formato **Currency**:

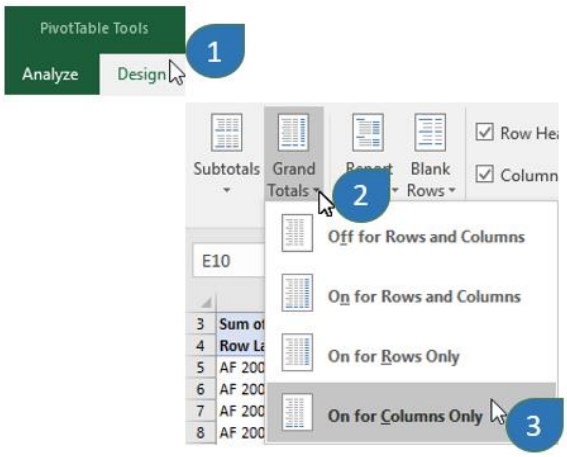
mn Labels															
2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr	AF 2006 Apr	AF 2007 Apr	AF 2008 Apr	AF 2009 Apr	AF 2010 Apr	AF 2011 Apr	AF 2012 Apr	AF 2013 Apr	AF 2014 Apr	AF 2015 Apr	AF 2
$1,254	$1,224	$1,317	$1,240	$1,696	$1,628	$1,627	$1,063	$1,405	$7,561	$27,843	$104,194	$39,384	$36,115	$24,575	
$9,350	$9,232	$9,245	$9,986	$10,257	$7,296	$6,255	$6,255	$6,044	$4,443	$4,568	$4,641	$4,535	$4,468	$4,288	
$114,629	$85,663	$92,890	$103,463	$102,780	$102,002	$103,586	$102,616	$97,758	$94,796	$82,162	$97,516	$97,398	$98,432	$113,266	$
$332,724	$315,500	$343,525	$383,773	$380,794	$478,804	$439,558	$403,090	$406,069	$0	$0	$0	$0	$0	$0	

Remover totales para las filas

Para propósitos de este ejercicio, no usaremos los valores de Grand Total de las filas.

Apr	AF 2016 Apr	Grand Total
,575	$0	$252,126
,288	$3,840	$104,703
,266	$113,266	$1,602,223
$0	$0	$3,483,837

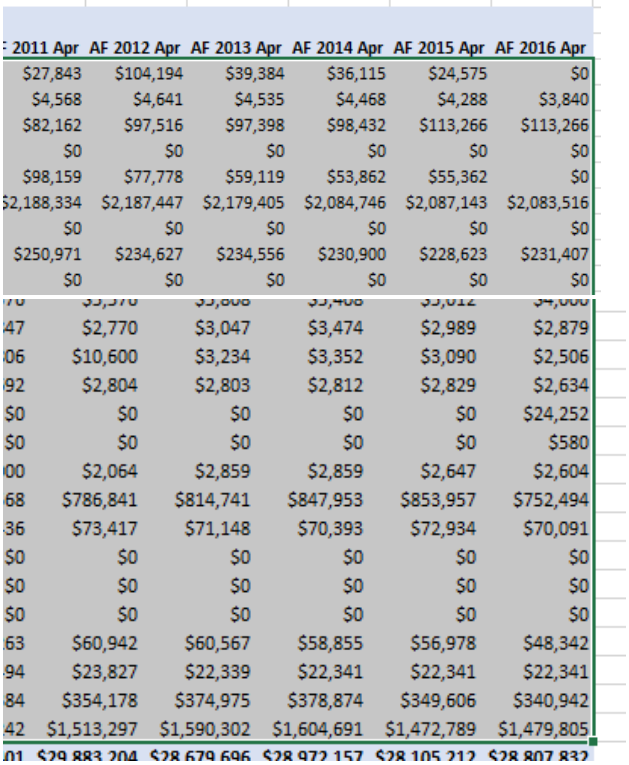
Estas se pueden **remover** yendo al **menú principal** en **PivotTable Tools | Design tab | Grand Totals | On for Columns Only**



Usar Conditional Formatting para resaltar valores extremos

Ya que estamos mostrando muchos números, usaremos una gradación de colores para resaltar valores más altos y más bajos.

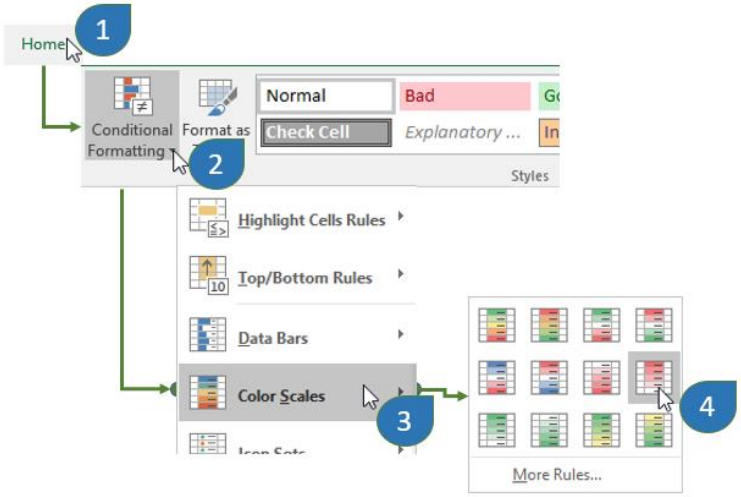
Sombree las celdas bajo las columnas 2001 al 2016 **sin incluir la fila de totales**



2011 Apr	AF 2012 Apr	AF 2013 Apr	AF 2014 Apr	AF 2015 Apr	AF 2016 Apr
$27,843	$104,194	$39,384	$36,115	$24,575	$0
$4,568	$4,641	$4,535	$4,468	$4,288	$3,840
$82,162	$97,516	$97,398	$98,432	$113,266	$113,266
$0	$0	$0	$0	$0	$0
$98,159	$77,778	$59,119	$53,862	$55,362	$0
$2,188,334	$2,187,447	$2,179,405	$2,084,746	$2,087,143	$2,083,516
$0	$0	$0	$0	$0	$0
$250,971	$234,627	$234,556	$230,900	$228,623	$231,407
$0	$0	$0	$0	$0	$0
47	$2,770	$3,047	$3,474	$2,989	$2,879
06	$10,600	$3,234	$3,352	$3,090	$2,506
92	$2,804	$2,803	$2,812	$2,829	$2,634
$0	$0	$0	$0	$0	$24,252
$0	$0	$0	$0	$0	$580
00	$2,064	$2,859	$2,859	$2,647	$2,604
68	$786,841	$814,741	$847,953	$853,957	$752,494
36	$73,417	$71,148	$70,393	$72,934	$70,091
$0	$0	$0	$0	$0	$0
$0	$0	$0	$0	$0	$0
$0	$0	$0	$0	$0	$0
63	$60,942	$60,567	$58,855	$56,978	$48,342
94	$23,827	$22,339	$22,341	$22,341	$22,341
84	$354,178	$374,975	$378,874	$349,606	$340,942
42	$1,513,297	$1,590,302	$1,604,691	$1,472,789	$1,479,805
01	$29,883,204	$28,679,696	$28,972,157	$28,105,212	$28,807,832

Mantenga las celdas sombreadas.

Vaya al menú principal **Home | Conditional Formatting | Color Scales** y **escoja** la opción **Red – White** en la **segunda fila**, **cuarta columna**.



El resultado deberá verse como este. Note cómo resaltan los valores más altos en rojo. Así podrá identificar más fácilmente los valores extremos.

Column Labels																
AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr	AF 2006 Apr	AF 2007 Apr	AF 2008 Apr	AF 2009 Apr	AF 2010 Apr	AF 2011 Apr	AF 2012 Apr	AF 2013 Apr	AF 2014 Apr	AF 2015 Apr	AF 2016 Apr
$1,254	$1,224	$1,317	$1,240	$1,696	$1,628	$1,627	$1,063	$1,405	$7,561	$27,843	$104,194	$39,384	$36,115	$24,575	$0
$9,350	$9,232	$9,245	$9,986	$10,257	$7,296	$6,255	$6,255	$6,044	$4,443	$4,568	$4,641	$4,535	$4,468	$4,288	$3,840
$114,629	$85,663	$92,890	$103,463	$102,780	$102,002	$103,586	$102,616	$97,758	$94,796	$82,162	$97,516	$97,398	$98,432	$113,266	$113,266
$332,724	$315,500	$343,525	$383,773	$380,794	$478,804	$439,558	$403,090	$406,069	$0	$0	$0	$0	$0	$0	$0
$218,099	$262,087	$224,984	$94,878	$107,560	$95,347	$81,476	$107,998	$145,707	$157,491	$98,159	$77,778	$59,119	$53,862	$55,362	$0
$1,475,056	$1,505,138	$1,575,573	$1,597,895	$1,686,526	$1,704,348	$1,743,746	$1,813,684	$2,049,123	$2,374,104	$2,188,334	$2,187,447	$2,179,405	$2,084,746	$2,087,143	$2,083,516
$15,469	$18,541	$19,604	$16,717	$13,051	$11,568	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
$317,613	$327,545	$361,462	$173,327	$234,404	$206,573	$217,316	$223,591	$225,079	$265,870	$250,971	$234,627	$234,556	$230,900	$228,623	$231,407
$13,559	$13,201	$12,882	$9,676	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
$89,972	$82,551	$101,580	$83,947	$82,358	$91,507	$90,578	$80,630	$79,175	$0	$0	$0	$0	$0	$0	$0
$4,009	$4,362	$4,701	$4,716	$5,073	$3,820	$4,656	$5,078	$5,483	$4,262	$4,423	$4,036	$4,024	$3,947	$3,268	$2,984
$72,858	$74,056	$98,524	$68,182	$80,008	$121,192	$89,339	$69,641	$77,629	$70,148	$67,623	$73,703	$56,103	$53,868	$54,884	$50,735
$21,800	$22,588	$22,157	$19,514	$19,997	$21,362	$21,982	$22,930	$23,965	$0	$0	$0	$0	$0	$0	$0
$85,700	$67,515	$72,013	$78,572	$79,008	$73,401	$84,392	$73,925	$82,907	$80,837	$99,459	$96,674	$97,864	$93,568	$86,512	$81,452
$1,296,428	$1,271,244	$1,247,901	$1,362,671	$1,359,410	$1,592,551	$1,641,700	$1,723,432	$1,495,025	$1,628,908	$2,017,294	$1,994,698	$2,250,274	$2,226,236	$2,223,908	$2,823,427
$126,350	$117,282	$126,039	$124,838	$144,812	$156,294	$184,026	$188,469	$155,507	$159,140	$136,754	$139,523	$138,699	$139,594	$128,724	$119,999
$24,799	$11,576	$12,340	$11,647	$15,452	$11,870	$11,569	$11,771	$11,975	$21,262	$9,805	$13,463	$13,403	$14,239	$12,667	$17,171
$113,384	$111,568	$116,717	$178,748	$215,179	$166,417	$192,719	$155,256	$195,265	$188,600	$184,690	$248,847	$234,605	$238,120	$247,832	$227,286

Navegue hacia abajo… podrá ver otros valores extremos.

¿Cuáles son las agencias con mayores presupuestos?

Mostrar valores como porcentaje del total

Ya que hemos *vectorizado* los valores por año fiscal, tendremos acceso a las opciones que provee el menú de contexto ***Show Values as***.

No Calculation	Presenta los valores, solo con el cómputo agregado.
% of Grand Total	Porcentaje del total.
% of Column Total	Porcentaje del total de una columna.
% of Row Total	Porcentaje del total de una fila.
% of …	Porcentaje comparando con otra columna.
% of Parent Row Total	Porcentaje de una fila en relación con el total de la fila madre. Mostrará el porcentaje correcto de la fila madre con relación al total global de la columna.
% of Parent Column Total	Porcentaje en relación con el total de la columna
% of Parent Total	Porcentaje en relación con el total de la fila madre. La diferencia es que la fila madre presentará valor de 100%.
Difference From	Resta entre una columna y otra.
% Difference From	Porcentaje de diferencia entre dos columnas.
Running Total In	Total, acumulado, usualmente entre fechas.
% Running Total In	Porcentaje acumulado, usualmente entre fechas.
Rank Smallest to Largest	Posicionamiento ordenado de menor a mayor.
Rank Largest to Smallest	Posicionamiento ordenado de mayor a menor.
Index	Convierte el valor a un índice ponderado como se explicó anteriormente.

No vamos a practicarlos todos aquí, pero haremos algunos.

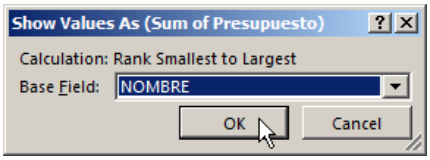
Experimente presentar los valores como:

Percent of Grand Total (¿Qué representan los valores?) _____
_____

Percent of Column Total (¿Qué representan los valores?) _____
_____

Percent of Row Total (¿Qué representan los valores?) _____
_____

Rank: **Largest to Smallest** & **Smallest to Largest**.
Use **NOMBRE** como campo (columna base) para la comparación



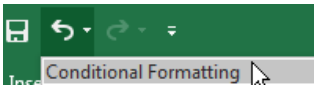
Note cómo algunas agencias mantienen su posición a lo largo de los años. Otras pueden cambiar drásticamente.

Conditional Formatting: Hasta ahora, los valores altos son mostrados usando color rojo oscuro. Esto puede ser engañoso (culturalmente estamos acostumbrados a pintar de rojo los valores negativos) porque los valores más bajos (posición 1,2,3…) son las que tienen presupuestos mayores. ¿Qué otro esquema de color puede usar para representar mejor los números? Recomendado: use gradación roja…

22	52	52	7	12
137	137	138	130	132
137	137	138	130	132
7	6	6	6	6
5	5	7	8	7
129	131	132	126	128
102	102	98	83	85
13	17	21	18	11
1	1	1	1	1
137	137	138	130	132
20	18	20	12	22
62	71	67	77	76
137	137	138	130	132
61	58	49	47	52

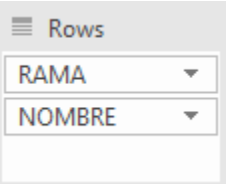
OTRAS OPCIONES AGRUPAR POR JERARQUÍA:
En esta parte usaremos la jerarquía existente para agrupar agencias y presentaremos los valores por grupo o Parent Row (Ramas Gubernamentales) y Agencias como filas.

Pero antes, **regrese** al **Conditional Formatting** previo si lo cambió. Use el comando **Undo** para regresar al estado anterior.



Haga **right click** en cualquiera de las **celdas** con números y escoja **Show Values As | No calculation**

Para agrupar por jerarquía, **arrastre el campo RAMA a la caja Rows**, justo ***encima*** del campo NOMBRE



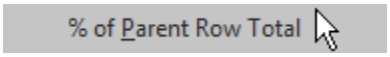
Podrá ver ahora la tabla jerarquizada por rama gubernamental, mostrando además las agencias (o Aportaciones presupuestarias) que las componen.

Sum of Presupuesto / Row Labels	Column Labels / AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr
A	**$1,831,273**	**$2,062,547**	**$2,404,339**	**$2,226,470**	**$2,531,002**
Aportaciones a los Municipios	$356,747	$406,100	$648,444	$439,431	$508,365
Aportaciones a los Partidos Políticos	$12,029	$900	$900	$900	$27,900
Aportaciones al Quehacer Cultural	$360	$360	$360	$500	$500
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro	$43,126	$43,624	$96,732	$146,748	$178,724
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro	$3,284	$6,564	$22,690	$40,410	$60,960
Deuda Pública	$347,354	$432,457	$439,480	$451,569	$490,362
Universidad de Puerto Rico	$1,068,373	$1,172,542	$1,195,733	$1,146,912	$1,264,191
E	**$17,579,732**	**$17,945,070**	**$19,431,459**	**$21,357,847**	**$22,304,250**
Administración de Asuntos Energéticos (AAE)	$1,254	$1,224	$1,317	$1,240	$1,696
Administración de Asuntos Federales de Puerto Rico	$9,350	$9,232	$9,245	$9,986	$10,257
Administración de Compensaciones por Accidentes de Automóviles	$114,629	$85,663	$92,890	$103,463	$102,780
Administración de Corrección	$332,724	$315,500	$343,525	$383,773	$380,794
Administración de Desarrollo Laboral	$218,099	$262,087	$224,984	$94,878	$107,560
Administración de Desarrollo Socioeconómico de la Familia	$1,475,056	$1,505,138	$1,575,573	$1,597,895	$1,686,526
Administración de Desarrollo y Mejoras de Viviendas	$15,469	$18,541	$19,604	$16,717	$13,051

Al organizar por jerarquía, podemos notar otros patrones: qué aportaciones aumentaron o se redujeron en la “Rama A” de Aportaciones Presupuestarias.

Percent of Parent Row Total

Use esta opción para mostrar los valores presupuestarios como porcentajes *dentro* de una jerarquía. Además, mostrará la participación en porcentaje de cada rama con respecto al total de la columna.

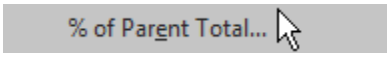


Fíjese cuáles son las proporciones por rama de gobierno y cómo han aumentado o disminuido las aportaciones, por ejemplo.

Sum of Presupuesto / Row Labels	Column Labels / AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr
A	**9.28%**	**10.13%**	**10.82%**	**9.26%**	**10.00%**
Aportaciones a los Municipios	19.48%	19.69%	26.97%	19.74%	20.09%
Aportaciones a los Partidos Políticos	0.66%	0.04%	0.04%	0.04%	1.10%
Aportaciones al Quehacer Cultural	0.02%	0.02%	0.01%	0.02%	0.02%
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro	2.35%	2.12%	4.02%	6.59%	7.06%
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro	0.18%	0.32%	0.94%	1.81%	2.41%
Deuda Pública	18.97%	20.97%	18.28%	20.28%	19.37%
Universidad de Puerto Rico	58.34%	56.85%	49.73%	51.51%	49.95%
E	**89.04%**	**88.09%**	**87.46%**	**88.85%**	**88.16%**
Administración de Asuntos Energéticos (AAE)	0.01%	0.01%	0.01%	0.01%	0.01%

Percent of Parent Total

Utilice la opción **% of Parent Total**. Use el campo **RAMA** como ***Base Field***.



Mostrará las jerarquías como 100% y las proporciones dentro del grupo.

Sum of Presupuesto	Column Labels				
Row Labels	AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr
A	100.00%	100.00%	100.00%	100.00%	100.00%
Aportaciones a los Municipios	19.48%	19.69%	26.97%	19.74%	20.09%
Aportaciones a los Partidos Políticos	0.66%	0.04%	0.04%	0.04%	1.10%
Aportaciones al Quehacer Cultural	0.02%	0.02%	0.01%	0.02%	0.02%
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro de Empleado	2.35%	2.12%	4.02%	6.59%	7.06%
Aportaciones para Pensiones y Seguridad Social - Sistema de Retiro de Maestros	0.18%	0.32%	0.94%	1.81%	2.41%
Deuda Pública	18.97%	20.97%	18.28%	20.28%	19.37%
Universidad de Puerto Rico	58.34%	56.85%	49.73%	51.51%	49.95%
E	100.00%	100.00%	100.00%	100.00%	100.00%
Administración de Asuntos Energéticos (AAE)	0.01%	0.01%	0.01%	0.01%	0.01%

Index

Como se explicó anteriormente, esta opción mostrará cada celda con métricas (measures) ponderando el valor relativo en términos del total de filas y total de columnas.

Primero, **remueva** el campo **RAMA** de la **caja Rows**, **desplazándolo a**l área de las columnas (**Fields**).

Haga **right click** en cualquier celda con números y escoja **Show Values as | Index**

Asignaciones bajo la Custodia de la Oficina de Gerencia y Presupuesto	0.092418256	0.802529807	0.656599339	0.603930918	0.475722881
Asignaciones para Propósitos Especiales	2.512128199	3.283449478	4.305437566	4.100114758	4.362199123
Asignaciones para Propósitos Generales	10.34221877	10.64664276	0	0	0
Autoridad de Acueductos y Alcantarillados	0.872323744	0.901704681	0.776788017	0.79971335	0.730546541
Autoridad de Carreteras y Transportación	1.175230608	0.99239204	0.899482907	1.059916838	0.820408402
Autoridad de Conservación y Desarrollo de Culebra	0.976072734	0.91384617	1.042547919	0.854601479	1.447165435
Autoridad de Desperdicios Sólidos de Puerto Rico	0.60239344	2.502755702	7.573704977	0.818309511	0.493080531
Autoridad de Edificios Públicos	1.64947375	1.899526005	1.597848735	1.398698525	0.986741574
Autoridad de Energía Eléctrica	0.845049249	0.809565679	0.760148021	0.803396425	0.895581913
Autoridad de las Navieras	4.495401424	4.356586718	3.996623313	3.694095294	0.764427943
Autoridad de los Puertos	0.857819601	0.858018411	1.333443635	1.312461011	1.172973631
Autoridad de Tierras de Puerto Rico	1.191661421	0.974134725	0.720057545	0.621711858	0.823364156
Autoridad de Transporte Integrado	0	0	0	0	0
Autoridad de Transporte Marítimo de Puerto Rico y las Islas Municipio	1.026234509	1.035452028	1.044025787	0.943034663	0.783523364
Autoridad del Centro de Convenciones de Puerto Rico	21.32718924	0	0	0	0
Autoridad del Distrito del Centro de Convenciones	0.380514182	0	0.9231733	1.545443178	2.173123039
Autoridad del Puerto de las Américas	0	0	0.253538762	0.234346916	0.824396672

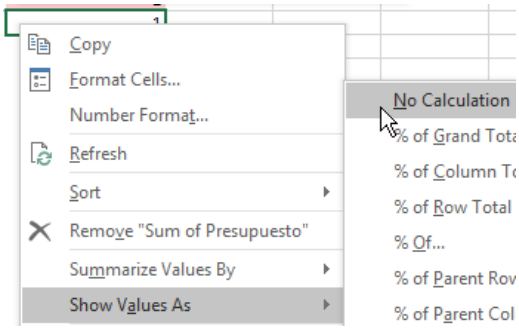
Recuerde que este índice puede verse influido *por grandes diferencias entre columnas*. Por eso algunos índices pueden aparecer muy altos. Pueden tener una asignación significativa en un año, pero en los demás años no aparecerán aportaciones.

Difference from y % Difference from

¿Cuáles son las diferencias entre años, para cada agencia, por ejemplo? Para esta y otras opciones de **Show Values As**, haremos algunos cambios.

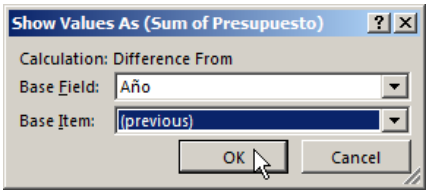
Para ver estas diferencias:

1. Quite los valores como índices. Haga **right click** en cualquiera de los valores de años y escoja **Show Values as | No calculation**



De esta manera, podrá ver los valores originales agregados antes de mostrar las diferencias por año.

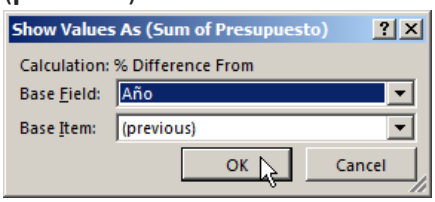
2. Haga **right click** *encima* de cualquiera de las celdas con números.
En el menú de contexto que aparecerá, escoja **Show Values as | Difference from**.
En ***Base Field***, escoja el campo **Año**.
En ***Base Item***, escoja **(previous)**… lo cual representa comparar con el año anterior.



Notará que los valores negativos aparecerán en colores blancos, mientras que los aumentos sustanciales a las asignaciones presupuestarias aparecerán en colores rojos más oscurecidos.

$703	$4,518	-$3,644	-$118	$3,815	-$2,561	-$535	$262	$570	-$1,357
$37,129	$190,181	-$235,725	$7,548	-$8,133	$0	$0	$0	$0	$0
$508,121	-$515,500	-$123,173	-$7,161	-$8,492	$0	$0	$0	$0	$0
$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
$68,934	$29,287	-$115,278	$34,970	-$83,155	$110,995	-$111,882	$39,256	-$10,972	$61,263
$27,000	-$27,297	$0	$597	$29,260	-$30,010	$750	$0	$10,400	-$10,400
$0	-$25	$0	$25	$0	$0	$0	$0	$0	$0
$31,976	$42,481	-$10,165	$17,000	$10,053	-$1,200	$5,732	$7,047	$7,758	-$18,645
$20,550	$0	$0	$7,125	$7,463	$1,185	$15,195	$6,712	$487	-$2,557
-$11,093	$10	-$9,055	$1,297	-$403	$5,500	$15,150	$850	$2,640	$4,175
$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
-$40,845	$157,425	-$121,584	-$195,110	$10,882	$1,065,431	-$613,103	$147,931	-$188,081	$417,071
$31,883	-$297,956	$0	$0	$0	$0	$0	$0	$0	$0
$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
-$32,532	$67,093	$374,068	$315,192	-$272,202	-$39,881	-$65,118	$243,522	-$193,330	$221,953
-$241,265	$441,079	$21,586	$68,351	-$506,964	-$71,936	$22,589	$1,295,988	-$281,569	-$751,688
$417	-$87	-$32	-$6	$92	-$146	-$224	-$28	$21	$122
-$4,412	$12,095	-$13,861	-$38	$378	$4,854	-$681	-$984	-$1,462	$2,554
-$166,483	$113,200	-$104,002	-$88,906	-$155,635	$165,117	$108,130	-$114,229	-$11,245	$130,783
$536,269	$642,376	$31,957	$739,297	-$547,685	$73,457	$262,719	$427,528	-$87,776	-$194,687

Repita el proceso; esta vez, utilice **Show Values as | % Difference From**.
Recuerde usar:
Año como **Base Field**
(**previous**) como **Base Item**



Note que aparecerán celdas con valor #DIV/0!. Esto se debe a que al comparar hay valores cero en el denominador.

Corporación del Centro Regional del ELA de PR	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Corporación del Conservatorio de Música de Puerto Rico	-43.71%	-3.24%	55.04%	57.69%	59.98%	-18.10%	-26.59%	-37.47%	9.11%
Corporación del Fondo del Seguro del Estado	7.84%	-0.21%	11.76%	10.50%	2.66%	4.36%	-0.38%	11.18%	0.02%
Corporación del Proyecto ENLACE del Caño Martín Peña	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	-7.60%
Corporación para el Desarrollo de las Artes, Ciencias e Industria Cinematogr	28.00%	228.19%	-9.88%	0.37%	-15.33%	19.71%	-17.16%	5.82%	-8.61%
Corporación para el Desarrollo de las Exportaciones de Puerto Rico	-2.62%	-9.29%	56.82%	-100.00%	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Corporación para el Desarrollo Rural	-65.95%	13.74%	143.29%	1.09%	-59.29%	35.10%	-47.70%	-51.16%	-100.00%
Corporación Pública para la Supervisión y Seguro de Cooperativas de Puerto I	-1.33%	17.80%	8.01%	-6.52%	-4.79%	10.67%	5.70%	20.67%	24.72%
Cuerpo de Bomberos de Puerto Rico	3.45%	9.06%	21.78%	8.69%	12.75%	1.47%	0.72%	5.98%	-11.96%
Cuerpo de Emergencias Médicas de Puerto Rico	13.97%	-0.04%	32.59%	-19.83%	48.14%	-6.45%	9.01%	2.08%	-3.29%
Departamento de Agricultura	78.75%	-28.44%	-30.75%	24.83%	58.68%	-42.00%	-4.69%	-3.08%	10.68%
Departamento de Asuntos del Consumidor	-3.68%	14.03%	3.24%	13.28%	4.92%	4.18%	-1.06%	7.01%	-23.88%
Departamento de Corrección y Rehabilitación	-17.71%	90.43%	9.65%	15.08%	17.55%	-20.83%	2.13%	-0.81%	9509.97%

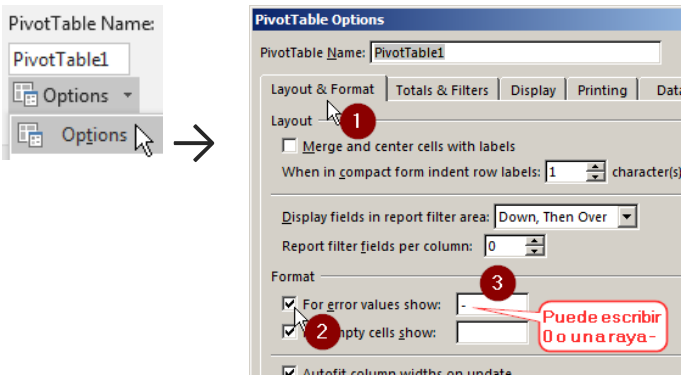
Hay una manera de "esconder" este valor en PivotTables

Esconder valores de error en PivotTable

Para esconder el valor de error #DIV/0!, vaya al **menú principal** bajo **PivotTable Tools | Analyze | Options | Options**

Vaya al tab **Layout & Format** y bajo **Format**, haga **check** en la opción **For error values show**:

En la caja de texto, **escriba** un cero (**0**) **o** una raya (-).



En la fila de totales podrá ver los cambios porcentuales globales.

3.18%	9.07%	8.19%	5.25%	4.99%	0.65%	3.64%	-3.49%	9.35%	-4.38%	6.88%	-4.03%	1.02%	-2.99%	2.50%

Running Total In y % Running Total In

Esta opción le mostrará la acumulación de valores (presupuestos), en este caso, a medida que pasan los años.

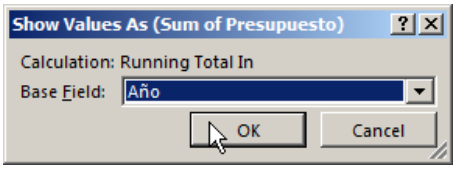
NOTA: En este caso (presupuesto) no es que haya necesariamente acumulaciones; por lo tanto, no hace mucho sentido usarlo porque el presupuesto gubernamental no es para acumular dinero; se asigna y se gasta, por lo general completamente, con pequeños ahorros o deudas…

Si quiere ver los números de todos modos, este sería el proceso:

Haga **right click** encima de cualquier celda que contenga valores numéricos.

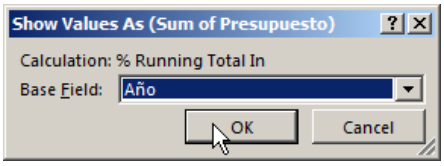
En el menú de contexto escoja **Show Values as | Running Total**

En la caja **Show Values As**, escoja **Año** como ***Base Field***.



$0	$0	$0	$0	$0	$0	$275,000
9,977	$3,272,077	$3,922,108	$4,384,058	$5,263,079	$5,894,904	$6,932,275
6,744	$1,136,744	$1,136,744	$1,136,744	$1,136,744	$1,136,744	$1,136,744
2,567	$72,567	$72,567	$72,567	$72,567	$72,567	$72,567
8,717	$11,430,621	$12,866,047	$14,108,143	$15,572,192	$17,078,993	$18,515,105
7,964	$13,353,460	$15,684,944	$17,734,859	$19,033,086	$20,275,886	$21,518,686
7,274	$7,821	$8,340	$8,880	$9,542	$10,272	$10,925
0,232	$220,629	$230,042	$237,993	$248,498	$254,222	$258,202
8,061	$5,836,124	$6,239,958	$6,632,547	$7,155,919	$7,619,270	$8,098,371
2,823	$42,265,272	$47,525,249	$52,697,450	$57,674,964	$62,559,841	$67,461,590
4,777	$114,777	$114,777	$114,777	$114,777	$114,777	$114,777

Repita el proceso para **Show Values as | % Running Total in**, usando **Año** como **Base Field**



Notará que los porcentajes van aumentando hacia 100% hacia el AF 2016 en las filas.

Recuerde que la aplicación de esta parte es solo demostración. El presupuesto no se acumula.

Otros desgloses

Usted podrá agregar y resumir datos usando las dimensiones que estén disponibles para analizar los datos. Para este conjunto de datos presupuestarios, podemos agregar datos por origen de fondos, concepto de gasto, cuenta, área, sector, etcétera. Este es un ejemplo que puede aplicar, desplazando el campo (field, columna) **NOMBRE_PROGRAMA** a la caja **Rows**

Sum of Presupuesto	Column Labels	
Row Labels	AF 2001 Apr	AF 2002 Apr
Administración de Asuntos Energéticos (AAE)	**$1,254**	**$2,478**
Administración y Conservación de Asuntos de Energía	$1,254	$2,478
Administración de Asuntos Federales de Puerto Rico	**$9,350**	**$18,582**
Asesoramiento en Asuntos Federales	$5,457	$10,712
Colaboradores con Puerto Rico	$27	$54
Coordinación y Orientación de Servicios en Oficinas Regionales	$3,550	$6,950
Gastos de Representación, Comisionado Residente de Puerto Rico	$316	$866
Administración de Compensaciones por Accidentes de Automóviles	**$114,629**	**$200,292**
Apoyo Técnico y Administrativo	$15,712	$28,273
Deuda Pública y Gerencial	$197	$197
Dirección y Administración General	$1,240	$2,697
Educación y Prevención de Accidentes	$1,472	$2,943
Reclamaciones y Compensaciones por Accidentes de Automóviles	$96,008	$166,182
Administración de Corrección	**$332,724**	**$648,224**

Este otro ejemplo muestra el detalle añadiendo el campo **CONCEPTO_DE_GASTO**

Sum of Presupuesto	Column Labels			
Row Labels	AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr
Administración de Asuntos Energéticos (AAE)	**$1,254**	**$2,478**	**$3,795**	**$5,035**
Administración y Conservación de Asuntos de Energía	**$1,254**	**$2,478**	**$3,795**	**$5,035**
Anuncios y Pautas en Medios	$0	$20	$26	$29
Asignación Pareo Fondos Federales	$0	$0	$0	$0
Asignaciones Englobadas	$0	$0	$0	$0
Compra de Equipo	$8	$41	$211	$236
Donativos, Subsidios y Distribuciones	$250	$250	$250	$250
Entidades Gubernamentales	$0	$0	$0	$0
Facilidades y Pagos por Servicios Públicos	$3	$6	$9	$20
Gastos de Transportación	$33	$81	$120	$167
Materiales y Suministros	$0	$21	$38	$48
Nómina y Costos Relacionados	$879	$1,861	$2,809	$3,808
Otros Gastos Operacionales	$75	$135	$178	$222
Servicios Comprados	$6	$40	$90	$146
Servicios Profesionales	$0	$23	$64	$109
Administración de Asuntos Federales de Puerto Rico	**$9,350**	**$18,582**	**$27,827**	**$37,813**

Los ejemplos muestran los valores presupuestarios en sumas agregadas. Usted puede experimentar usando las opciones contenidas en **Show Values as**.

Ejercicio 3: Crear PivotChart

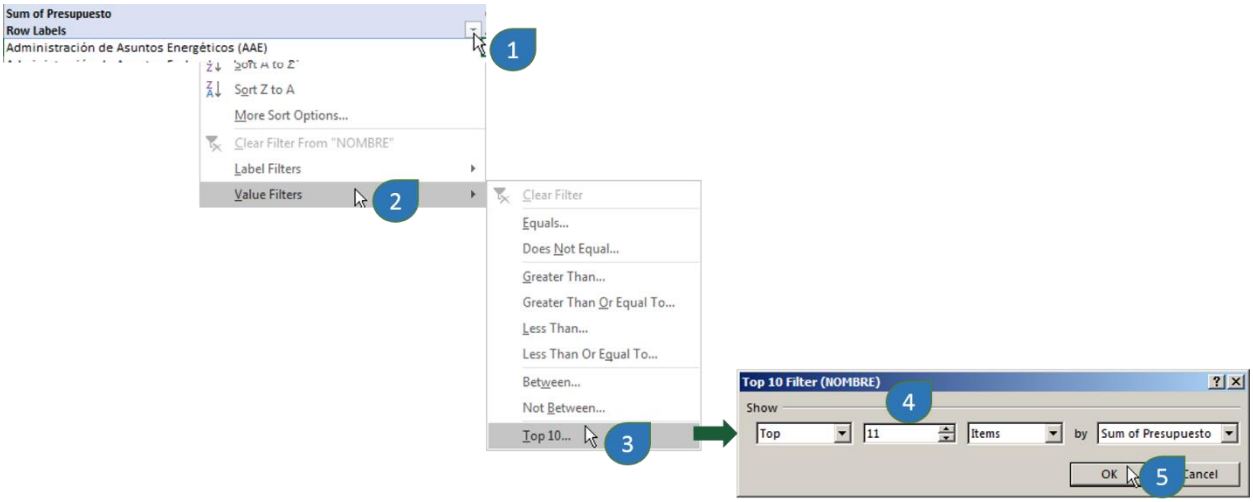
Antes de terminar, haremos un ejemplo de una gráfica pivot usando el mismo conjunto de datos presupuestarios. Para hacer la gráfica Pivot primero debemos preparar los datos un poco:

- Para **simplificar** la gráfica, vamos a **aplicar** un **filtro** (*Top 10*) para seleccionar un grupo de agencias con los presupuestos más grandes.
- Presentaremos los **datos** presupuestarios **agregados** solamente (**No Calculation**)
- Cambiaremos los **Años** a las filas (**Rows**) y las 10 a 15 **agencias** seleccionadas a la caja **Columns**.

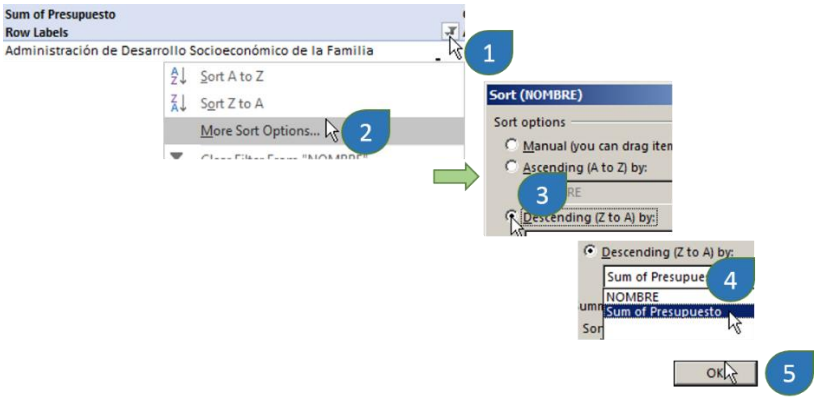
Regrese los valores de las celdas a datos agregados (No Calculation).
Haga **right click** en cualquier celda numérica. En el menú de contexto escoja **Show Values as | No Calculation**.

Aplicar filtro Top 10

Haga **click** en el **botón flecha** al lado derecho de la celda ancha **Row Labels**, **click** en **Value Filters | Top 10**. Aparecerá la forma **Top 10 Filter**. Aplique (escriba) filtro para **11 Items** by **Sum of Presupuesto**



Ordene los datos de forma descendente para que la gráfica aparezca más organizada.
Haga **click** en el **botón flecha**. Escoja **More Sort Options**. En la forma **Sort**, escoja la opción **Descending (Z to A) by**, luego escoja el ítem **Sum of Presupuesto** de la lista en el combo box. Luego **click** en botón **OK** para aceptar las opciones y ejecutar el ordenamiento.

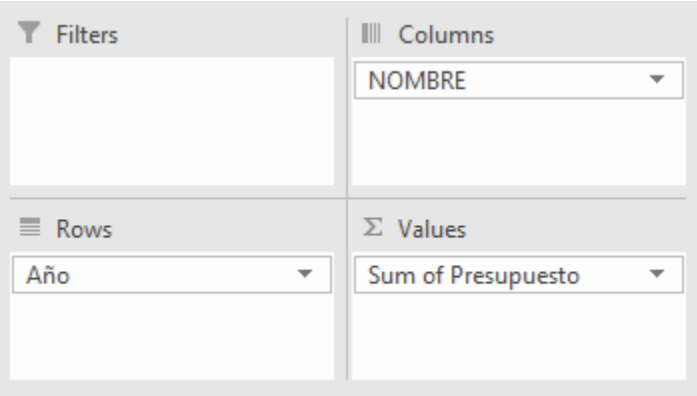


La tabla deberá aparecer así:

Sum of Presupuesto	Column Labels							
Row Labels	AF 2001 Apr	AF 2002 Apr	AF 2003 Apr	AF 2004 Apr	AF 2005 Apr	AF 2006 Apr	AF 2007 Apr	AF 2008
Autoridad de Energía Eléctrica	$2,673,037	$5,315,210	$8,021,110	$11,115,169	$14,745,497	$19,018,201	$23,322,862	$28,36
Departamento de Educación	$2,303,022	$4,572,345	$7,269,797	$9,820,786	$12,509,441	$16,133,656	$19,572,651	$23,11
Administración de Desarrollo Socioeconómico de la Familia	$1,475,056	$2,980,194	$4,555,767	$6,153,662	$7,840,188	$9,544,536	$11,288,282	$13,10
Administración de Seguros de Salud de Puerto Rico	$1,296,428	$2,567,672	$3,815,573	$5,178,244	$6,537,654	$8,130,205	$9,771,905	$11,49
Universidad de Puerto Rico	$1,068,373	$2,240,915	$3,436,648	$4,583,560	$5,847,751	$7,293,463	$8,749,067	$10,22
Autoridad de Carreteras y Transportación	$1,185,783	$2,218,905	$3,240,234	$4,542,290	$5,603,081	$7,104,951	$8,628,407	$10,22
Autoridad de Acueductos y Alcantarillados	$757,304	$1,564,991	$2,323,893	$3,169,177	$3,981,929	$4,861,774	$6,115,687	$7,68
Departamento de Salud	$528,992	$1,122,474	$1,773,165	$2,565,322	$3,307,410	$4,137,248	$4,961,731	$5,78
Policía de Puerto Rico	$590,821	$1,182,365	$1,851,376	$2,649,357	$3,408,545	$4,276,038	$5,138,818	$6,02
Banco Gubernamental de Fomento para Puerto Rico	$363,047	$694,546	$1,094,276	$1,513,040	$2,217,329	$2,600,641	$3,126,108	$3,53
Autoridad de Edificios Públicos	$626,339	$1,370,549	$2,053,344	$2,699,986	$3,180,145	$3,773,504	$4,262,861	$4,66
Deuda Pública	$347,354	$779,811	$1,219,291	$1,670,860	$2,161,222	$2,396,562	$3,028,884	$3,59
Aportaciones a los Municipios	$356,747	$762,847	$1,411,291	$1,850,722	$2,359,087	$2,896,739	$3,319,113	$3,77
Administración de Vivienda Pública	$326,210	$701,441	$1,013,421	$1,380,348	$1,869,143	$2,349,308	$2,798,880	$3,23
Asignaciones bajo la Custodia de la Oficina de Gerencia y Presupuesto	$30,040	$299,187	$539,365	$778,369	$976,528	$1,332,112	$1,566,112	$1,60
Grand Total	$13,928,553	$28,373,452	$43,618,551	$59,670,892	$76,544,950	$95,848,938	$115,651,368	$136,44

Cambiar agencias a columnas y años a Rows

Ahora arrastre el campo (field) **NOMBRE** a la caja **Columns** y el campo **Años** a la caja **Rows**



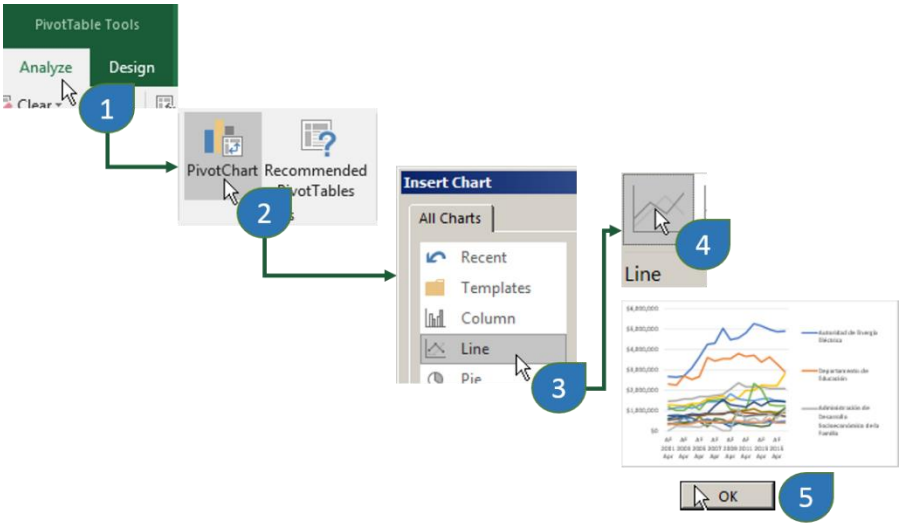
Quite cualquier resumen que parezca a Running Totals. Haga **right click** encima de cualquier celda numérica y escoja **Show Values As | No Calculation**. La tabla debe verse así:

Sum of Presupuesto	Column Labels		
Row Labels	Autoridad de Energía Eléctrica	Departamento de Educación	Administración de Desarrollo Socioeconómico de la Familia
AF 2001 Apr	$2,673,037	$2,303,022	$1,475,056
AF 2002 Apr	$2,642,173	$2,269,323	$1,505,138
AF 2003 Apr	$2,705,900	$2,697,452	$1,575,573
AF 2004 Apr	$3,094,059	$2,550,989	$1,597,895
AF 2005 Apr	$3,630,328	$2,688,655	$1,686,526
AF 2006 Apr	$4,272,704	$3,624,215	$1,704,348
AF 2007 Apr	$4,304,661	$3,438,995	$1,743,746
AF 2008 Apr	$5,043,958	$3,546,651	$1,813,684
AF 2009 Apr	$4,496,273	$3,548,156	$2,049,123
AF 2010 Apr	$4,569,730	$3,822,139	$2,374,104
AF 2011 Apr	$4,832,449	$3,655,058	$2,188,334
AF 2012 Apr	$5,259,977	$3,714,259	$2,187,447
AF 2013 Apr	$5,172,201	$3,374,077	$2,179,405
AF 2014 Apr	$4,977,514	$3,653,007	$2,084,746
AF 2015 Apr	$4,884,877	$3,294,743	$2,087,143
AF 2016 Apr	$4,901,749	$2,904,300	$2,083,516
Grand Total	$67,461,590	$51,085,041	$30,335,784

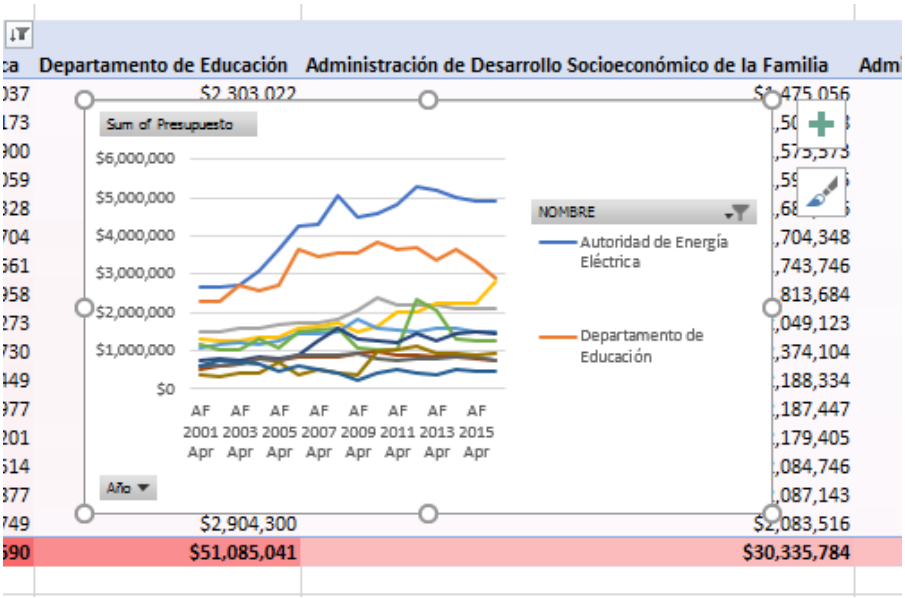
Generar gráfica Pivot

Ahora que esta tabla está lista, pasemos a generar una gráfica Pivot. Primero, asegúrese que la tabla pivot esté activada (click en cualquier celda *dentro* de la tabla pivot).

Vaya al **menú principal**. Asegúrese de ver en el menú **PivotTable Tools**.
Haga **click** en **Analyze | Pivot Chart | Line**. Escoja la opción **Line**. Observe la previsualización.
Haga **click** en botón **OK**

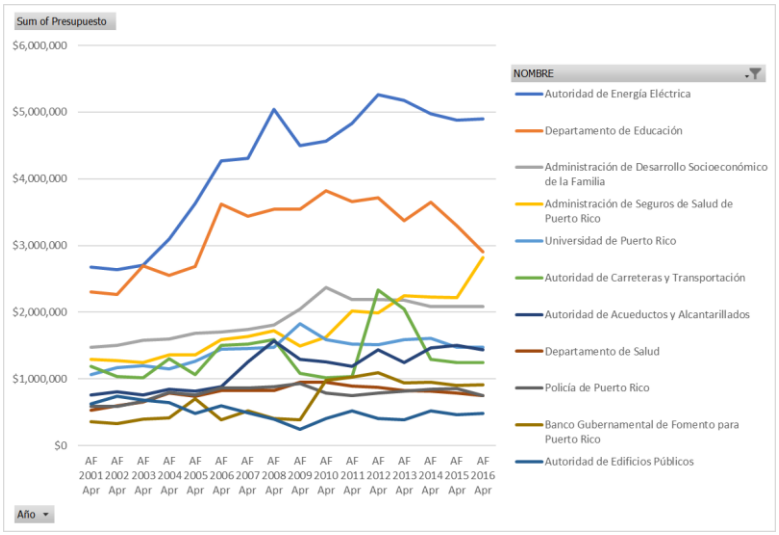


La gráfica Pivot aparecerá en algún lugar dentro o fuera de la tabla Pivot:



Mueva la gráfica debajo de la tabla. Expanda la gráfica, estirando desde los círculos en las esquinas. Estire lo suficiente para que pueda ver en la leyenda los once ítems

La gráfica debe verse así:



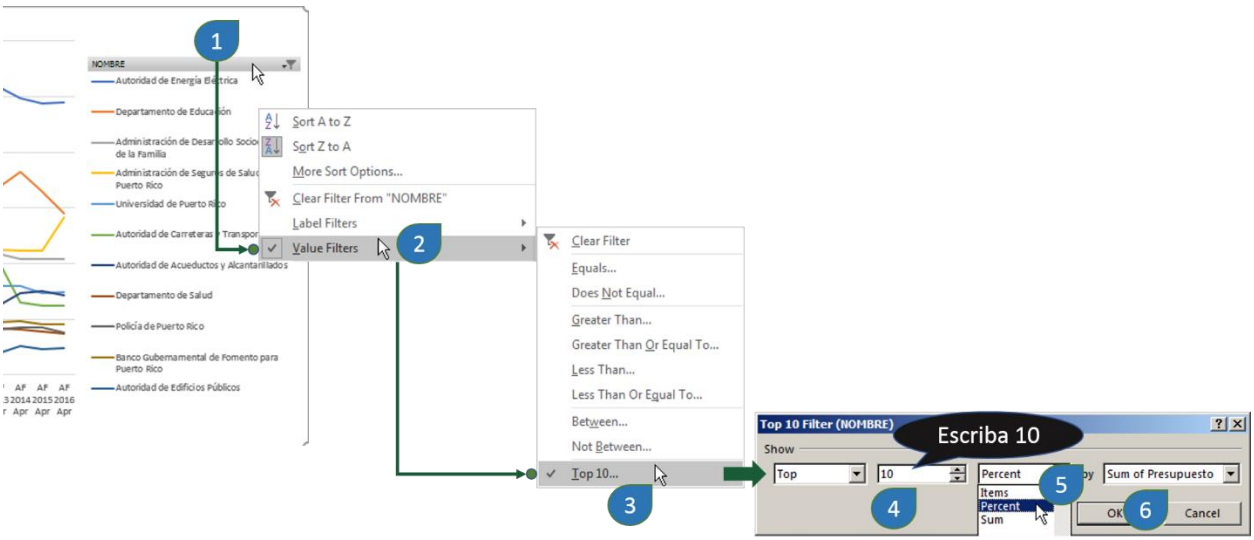
Aplicar filtro en la gráfica Pivot (Top 10, Opción Percent)

Note que hay varios botones en la gráfica. Estos son para seleccionar datos y presentar subconjuntos de estos.

Vamos a experimentar seleccionar usando el filtro ***Top 10*** con la opción *Percent*. Aplicaremos filtro para presentar agencias que estén en el rango "Top 10%" …

Haga **click** en el botón de filtro **NOMBRE | Value Filters | Top 10** …

En la caja **Top 10 Filter** escoja **Top**, **escriba 10** y escoja de la lista a la derecha el ítem **Percent**. **Click** en **OK** para aceptar cambios y ejecutar el filtro.



Para filtro 10%: ¿Cuántas agencias (entidades gubernamentales) aparecen en la gráfica?

_____

Repita el proceso, aumentando el porcentaje a 20, 30 y hasta 50%.

Para filtro 20%: ¿Cuántas agencias (entidades gubernamentales) aparecen en la gráfica?

_____

Para filtro 30%: ¿Cuántas agencias (entidades gubernamentales) aparecen en la gráfica?

_____

Para filtro 50%: ¿Cuántas agencias (entidades gubernamentales) aparecen en la gráfica?

_____

¿Qué tipo de filtro propone utilizar para considerar solo las agencias de gobierno central?

_____

¿Qué filtro propone utilizar para obtener agencias que dependen del Fondo General, Ingresos Propios, Fondos Federales, ¿etc.?

_____

Esto concluye este ejercicio.

Guarde su workbook Excel y nómbrelo **PrácticaPresupuestoPivot.xlsx** en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_3**

Para el próximo ejercicio usaremos datos demográficos del **Registro de Empleados Públicos**, disponible en el portal de datos http://data.pr.gov.

EJERCICIO 4: CAMPOS CALCULADOS EN PIVOTTABLES

En esta parte vamos a practicar generar campos adicionales en PivotTables. En ocasiones, los datos originales no proveen campos calculados, especialmente si vienen de bases de datos. La derivación de datos: resumen y análisis, viene a ser trabajo del *analista de datos*. Las tablas en la base de datos deben ser eficientes y ocupar el menor espacio posible.

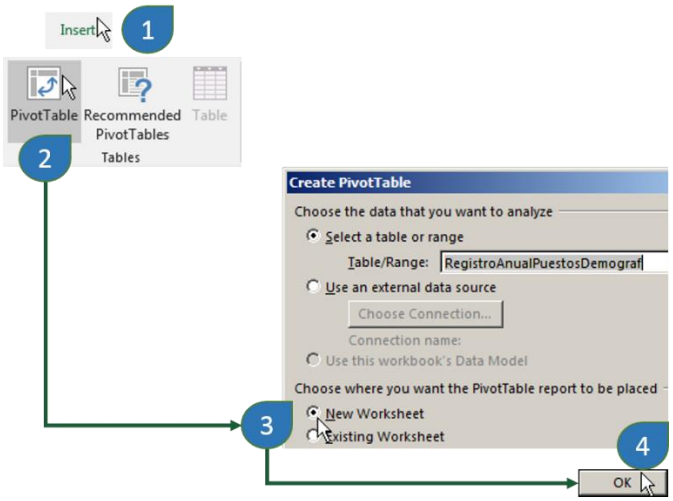
En este ejercicio usaremos el workbook Excel:
Registro Anual de Puestos por Demografía_26abr2017.xlsx, localizado en
C:\Tutorial_Excel\DataAnalysis\Modulo_3

Al abrir este workbook, esta es una muestra de tabla Excel:

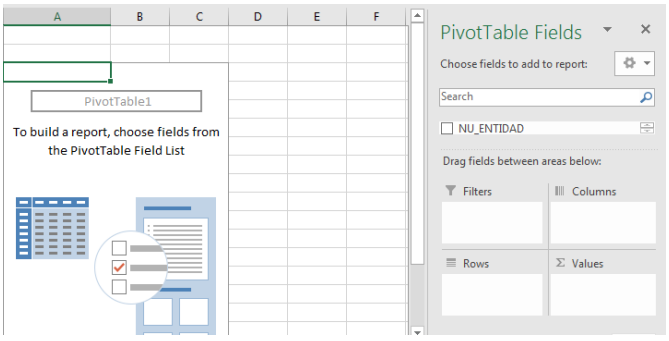
	A	B	C	D	E	F	G	H	I	J	K	L
1	NU_E	Clasificación	AñoFiscal	NB_ENTIDAD	FechaTran	FechaCierre	Estatus	CarreraParcialM	CarreraParcialF	CarreraCompletoM	CarreraCompletoF	CarreraNoGraduadoM
2	400	Rama Legislativa	2013-2014	Oficina del Contralor de Puerto Rico	1/20/2015 0:00	2/5/2015 0:00	1	0	2	214	227	0
3	400	Rama Legislativa	2014-2015	Oficina del Contralor de Puerto Rico	2/8/2016 0:00	2/18/2016 0:00	1	0	4	211	224	0
4	400	Rama Legislativa	2015-2016	Oficina del Contralor de Puerto Rico	3/1/2017 0:00	3/2/2017 0:00	1	0	0	203	220	0
5	400	Rama Legislativa	2012-2013	Oficina del Contralor de Puerto Rico	2/6/2014 0:00	2/7/2014 0:00	1	0	2	224	232	0
6	1000	Departamentos y Agencias	2013-2014	Departamento de Agricultura	2/6/2015 0:00	2/6/2015 0:00	1	0	0	152	125	2
7	1000	Departamentos y Agencias	2014-2015	Departamento de Agricultura	3/1/2016 0:00	3/1/2016 0:00	1	0	0	158	127	2
8	1000	Departamentos y Agencias	2015-2016	Departamento de Agricultura	2/1/2017 0:00	2/1/2017 0:00	1	0	0	159	132	2
9	1000	Departamentos y Agencias	2012-2013	Departamento de Agricultura	1/22/2014 0:00	1/27/2014 0:00	1	1	0	161	137	2
10	1020	Departamentos y Agencias	2014-2015	Departamento de Estado	1/14/2016 0:00	1/14/2016 0:00	1	0	0	36	59	4

AÑADIR PIVOTTABLE

Para añadir una tabla Pivot, haga **click** en cualquier celda de la tabla, vaya al **menú principal** y escoja **Insert | PivotTable**.



Aparecerá la tabla Pivot vacía, junto con la interfaz PivotTable Fields

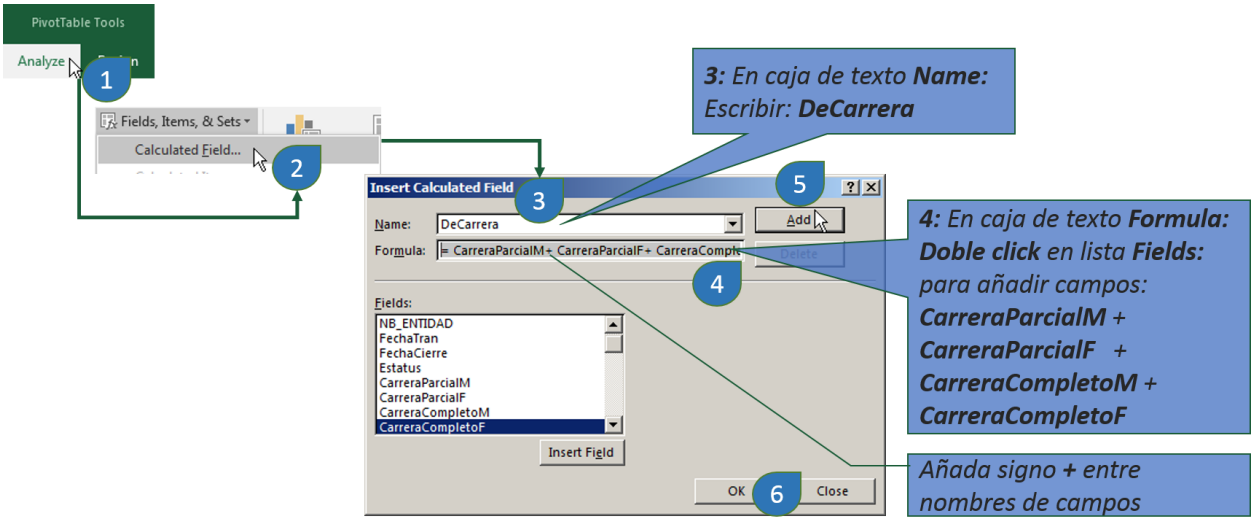


Añadir campo calculado a la tabla Pivot

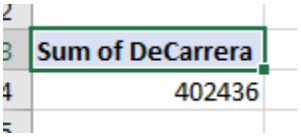
Según aparece en la tabla original de datos, los empleados de carrera (métrica) aparecen divididos en múltiples categorías: por género, jornada completa, jornada parcial, niveles educativos, y otras características. A diferencia de la tabla presupuestaria, estos datos son un poco más complejos por la cantidad de columnas. Por lo tanto, si queremos tener un conteo de empleados públicos en la tabla Pivot, podemos generar un campo calculado.

Con la **tabla pivot** vacía **activada**, vaya al **menú principal**, **click** en **Pivot Table Tools | Analyze | Fields, Items, & Sets**, **click** en **Calculated Field…**

En la forma **Insert Calculated Field**:
En **Name**: escriba **DeCarrera**
En **Formula**: **=CarreraCompletoM+CarreraCompletoF+CarreraParcialM+CarreraParcialF**
Click en botón **Add** para añadir el nuevo campo calculado (en la tabla Pivot)
Click en botón **OK** para aceptar los cambios y salir de esta forma



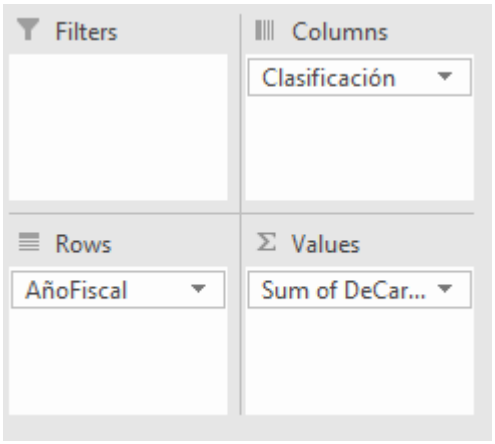
Su tabla Pivot deberá aparecer así:



Aquí aparecerá solamente la sumatoria del nuevo campo calculado **DeCarrera**

Añada las columnas para hacer el desglose por:

Campo	Definición
Clasificación	Clasificación de entidad pública
AñoFiscal	Años fiscales 2012-2013, 2014-2015, 2015-2016



Arrastre el campo **Clasificación** a la caja **Columns**
Arrastre el campo **AñoFiscal** a la caja **Rows**

Su tabla Pivot aparecerá así:

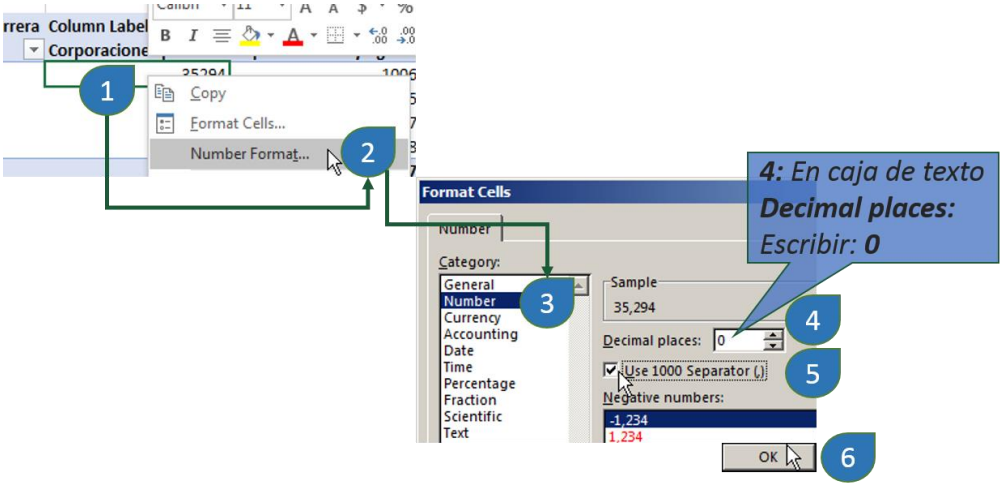
Sum of DeCarrera	Column Labels							
Row Labels	Corporaciones públicas	Departamentos y Agencias	Municipios	Oficina del Gobernador	Otras entidades	Rama Judicial	Rama Legislativa	Grand Total
2012-2013	35294	100668	27872	1069	429	3568	531	169431
2013-2014	18433	40501	14937	1087	225	3663	812	79658
2014-2015	18472	44792	19287	818	149	3581	507	87606
2015-2016	16044	31820	16876	364	147	0	490	65741
Grand Total	**88243**	**217781**	**78972**	**3338**	**950**	**10812**	**2340**	**402436**

Cambiar formato numérico

Cambie el formato numérico para añadir una coma para separar millares.

Haga **right click** *encima* de una de las celdas numéricas. **Click** en opción **Number Format…**

Aparecerá la forma **Format Cells**. Click en **Number**. En **Decimal Places**, escriba **0**.
Check en opción **Use 1000 separator**.
Click en **OK** para aceptar cambios y cerrar esta forma.



La tabla Pivot debe verse así:

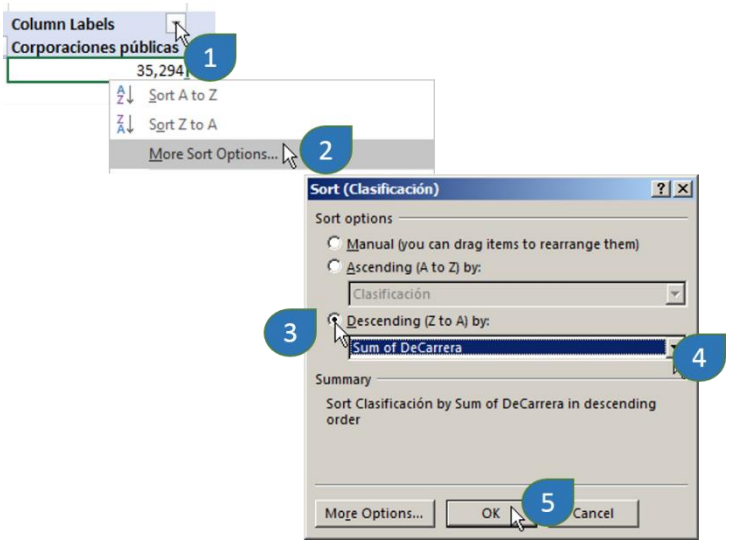
Sum of DeCarrera	Column Labels							
Row Labels	Corporaciones públicas	Departamentos y Agencias	Municipios	Oficina del Gobernador	Otras entidades	Rama Judicial	Rama Legislativa	Grand Total
2012-2013	35,294	100,668	27,872	1,069	429	3,568	531	169,431
2013-2014	18,433	40,501	14,937	1,087	225	3,663	812	79,658
2014-2015	18,472	44,792	19,287	818	149	3,581	507	87,606
2015-2016	16,044	31,820	16,876	364	147	0	490	65,741
Grand Total	**88,243**	**217,781**	**78,972**	**3,338**	**950**	**10,812**	**2,340**	**402,436**

Ordenar datos de forma descendente

Ordene los datos de mayor a menor. Haga **click** en el **botón flecha** al lado derecho de la celda **Row Labels**. Escoja la opción **More Sort Options…**

En la forma **Sort**, escoja la opción **Descending (Z to A) by**. **Escoja** de la lista el campo **Sum of DeCarrera**.

Haga **clic**k en el botón **OK** para aceptar cambios y cerrar esta forma.

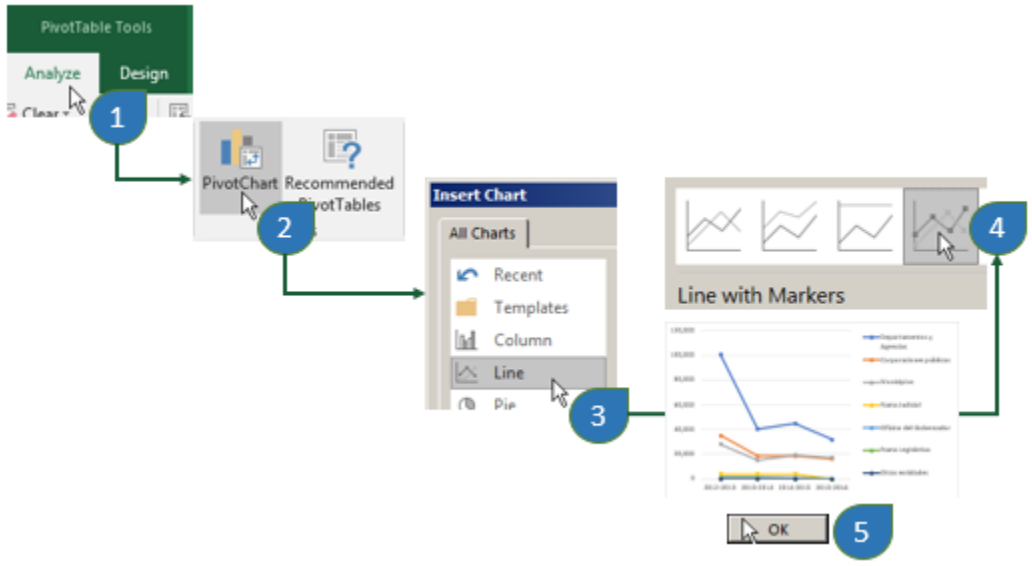


La tabla pivot debe verse así:

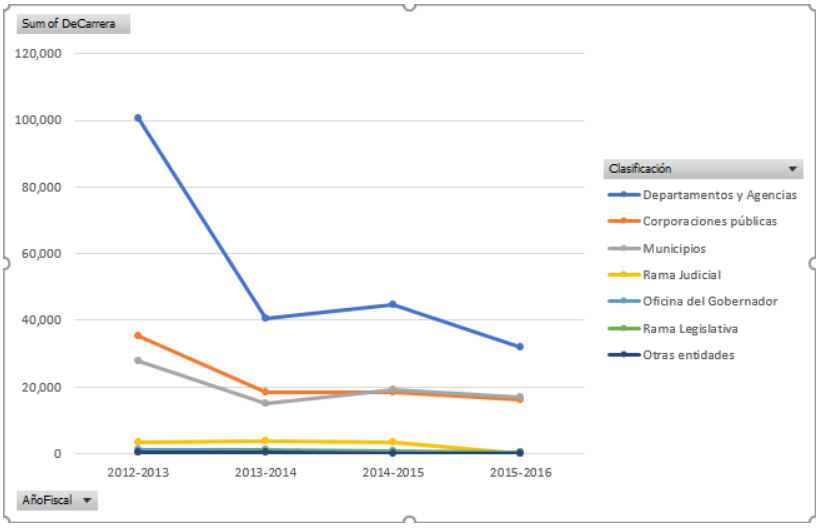
Sum of DeCarrera	Column Labels							
Row Labels	Departamentos y Agencias	Corporaciones públicas	Municipios	Rama Judicial	Oficina del Gobernador	Rama Legislativa	Otras entidades	Grand Total
2012-2013	100,668	35,294	27,872	3,568	1,069	531	429	169,431
2013-2014	40,501	18,433	14,937	3,663	1,087	812	225	79,658
2014-2015	44,792	18,472	19,287	3,581	818	507	149	87,606
2015-2016	31,820	16,044	16,876	0	364	490	147	65,741
Grand Total	**217,781**	**88,243**	**78,972**	**10,812**	**3,338**	**2,340**	**950**	**402,436**

Preparar PivotChart

Para terminar, prepare una gráfica Pivot. Estos son los pasos:

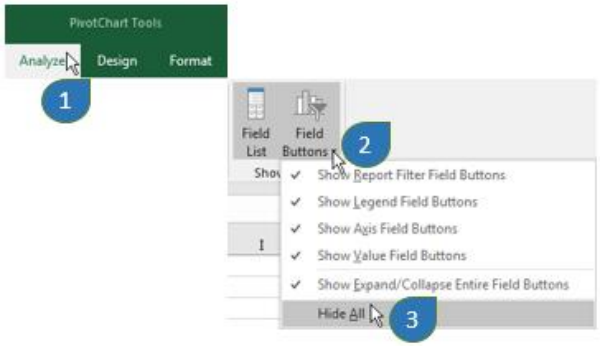


Luego de mover y estirar un poco, la gráfica Pivot debe verse así:

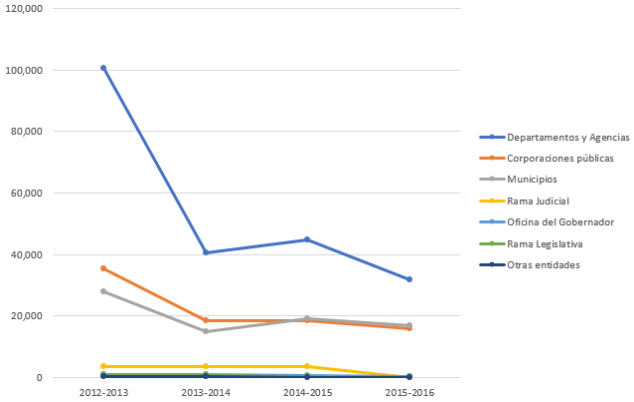


Quitar botones en la gráfica Pivot

Si desea remover los botones para filtrado de la gráfica Pivot. Estos son los pasos:



Así deberá aparecer la gráfica:



Esto concluye los ejercicios.

Guarde este archivo en el folder **C:\Tutorial_Excel\DataAnalysis\Modulo_3**

Fuentes

Funciones útiles, Excel

https://www.analyticsvidhya.com/blog/2016/12/cheatsheet-excel-functions-keyboard-shortcuts/ recuperado en enero 22, 2018

AGGREGATE
https://support.office.com/en-us/article/AGGREGATE-function-43b9278e-6aa7-4f17-92b6-e19993fa26df, recuperado en enero 22, 2018
SUMIFS
https://exceljet.net/excel-functions/excel-sumifs-function, recuperado en enero 22, 2018

Keyboard shortcuts
https://www.analyticsvidhya.com/blog/2016/12/cheatsheet-excel-functions-keyboard-shortcuts/, recuperado en enero 22, 2018.

10 Excel Functions You Need to Know for Data Analysis, https://blog.adeptmarketing.com/10-excel-functions-you-need-to-know-for-data-analysis, recuperado en enero 22, 2018.

LOOKUP
https://support.office.com/en-us/article/LOOKUP-function-446d94af-663b-451d-8251-369d5e3864cb, recuperado en marzo 9, 2018.

VLOOKUP
Excel Vlookup Tutorial Part 3 - Vlookup Example with Exact Match,
http://www.excelfunctions.net/Vlookup-Example-Exact-Match.html, recuperado en enero 24,2018.

Index and Match, named ranges
Excel's INDEX MATCH Alternative To VLOOKUP For Marketers [VIDEO],
https://www.annielytics.com/blog/excel-tips/excels-index-match-alternative-vlookup-marketers-video/, recuperado en febrero 7, 2018.

Referencia a celdas
Relative and Absolute Cell References, https://www.gcflearnfree.org/excel2016/relative-and-absolute-cell-references/1/, recuperado en febrero 7, 2018.

Trasponer celdas:
How to transpose in Excel - convert row to column and viceversa
https://www.ablebits.com/office-addins-blog/2014/03/06/transpose-excel-rows-columns/
recuperado en marzo 26, 2018.

Programas Add-In Utilidades

The XY Chart Labeler, http://www.appspro.com/Utilities/ChartLabeler.htm se usa para tener más opciones para añadir etiquetas en los diagramas de dispersión en Excel. Recuperado en febrero 7, 2018.

Datos temporales

Date and time functions (reference). Documentación Microsoft. https://support.office.com/en-us/article/date-and-time-functions-reference-fd1b5961-c1ae-4677-be58-074152f97b81, recuperado en febrero 12, 2018.

Using DATE function in Excel - formula examples to calculate dates, https://www.ablebits.com/office-addins-blog/2015/06/03/excel-date-formulas/, recuperado en febrero 13, 2018.

Año 1900 aparece como bisiesto en Lotus-1-2-3 y por consiguiente en Excel:
Excel incorrectly assumes that the year 1900 is a leap year, Documentación Microsoft: https://msuppeortn.micrtosofta.com/en-us/help/214326/excel-incorrectly-assumes-that-the-year-1900-is-a-leap-year, recuperado en febrero 13, 2018

Why are Excel weekdays wrong for 1900?, https://superuser.com/questions/481496/why-are-excel-weekdays-wrong-for-1900, recuperado en febrero 13, 2018

Method to determine whether a year is a leap year, Documentación Microsoft, https://support.microsoft.com/en-us/help/214019/method-to-determine-whether-a-year-is-a-leap-year, recuperado en febrero 13, 2018

DATEDIF function

DATEDIF function. Documentación Microsoft, https://support.office.com/en-us/article/datedif-function-25dba1a4-2812-480b-84dd-8b32a451b35c, recuperado en febrero 13, 2018.

Secret Excel DATEDIF Function, https://www.myonlinetraininghub.com/secret-excel-function-datedif, recuperado en febrero 13, 2018. Provee ejemplos más detallados de los errores en esta función y alternativas a esta función.

Cómo representar fechas en inglés a español:
Specifying a Language for the TEXT Function
https://excelribbon.tips.net/T011782_Specifying_a_Language_for_the_TEXT_Function.html
recuperado en 26 abril, 2018
Muestra cómo desplegar una fecha en otro idioma cuando Excel está en otro lenguaje. El código que deberá insertar en la función TEXT para español es **0C0A**. Ejemplo:
=TEXT(DATEVALUE("11/09/1999"),"[$-0C0A]dddd, d ""de"" mmmm, yyyy")

Visualización, gráficas

Recomendaciones:

Data Visualization 101: How to Choose the Right Chart or Graph for Your Data, https://blog.hubspot.com/marketing/data-visualization-choosing-chart, recuperado en enero 25, 2018

Data Visualization – How to Pick the Right Chart Type? https://eazybi.com/blog/data_visualization_and_chart_types/, recuperado en enero 25, 2018

Choosing a good chart, http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html, recuperado en enero 25, 2018

How to Use the Gestalt Principles for Visual Storytelling #PoDV, https://www.fusioncharts.com/blog/how-to-use-the-gestalt-principles-for-visual-storytelling-podv/, recuperado en abril 13, 2018

Pivot tables

Understanding Pivot Table Index Calculations http://datapigtechnologies.com/blog/index.php/understanding-pivot-table-index-calculations/, Recuperado en 11 junio, 2018

Otros temas:

Quiebras

¿Qué es la Quiebra?, http://www.quiebras.com/pages/que_es_la_quiebra/itemid:7, recuperado en febrero 20, 2018.