

Ejercicio: Uso de herramientas de Pentaho y R

Has sido asignado como analista de datos en una empresa que mantiene una base de datos con información sobre ventas.

Tu tarea es la siguiente:

1. Utiliza Pentaho para extraer datos de las tablas (ventas y regiones) y combinarlos en un solo conjunto de datos. La relación entre ellas es el id de región. La información relevante incluye:

Fichero	Contenido	Campos
#	Nombre	Tipo
1	mes	String
2	importe	Number
3	region	String

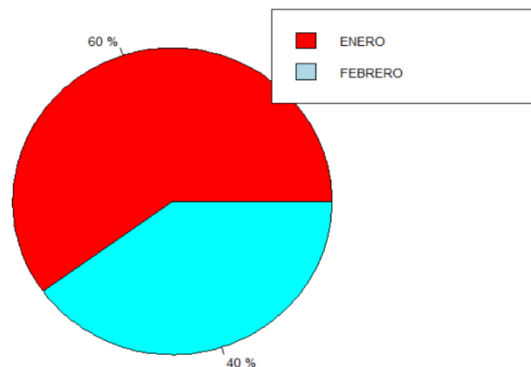
Nota: El campo region de la figura anterior, está en la tabla "Regiones" como "nombre". (Habrá que renombrarlo a "Región").

2. A continuación, utiliza Pentaho para generar un archivo CSV que contenga estos datos combinados.

Region, Mes y Ventas(Importe)

Se filtrará por las Regiones A y B y aquellos comerciales < 3 (idcomercial)
Generará un archivo separado por "," con la información anterior

3. Desde R, usando dataframes, recuperar dicha información y generar un gráfico de "tarta" con la información de ventas agrupadas por mes. Deberá tener una apariencia similar a :



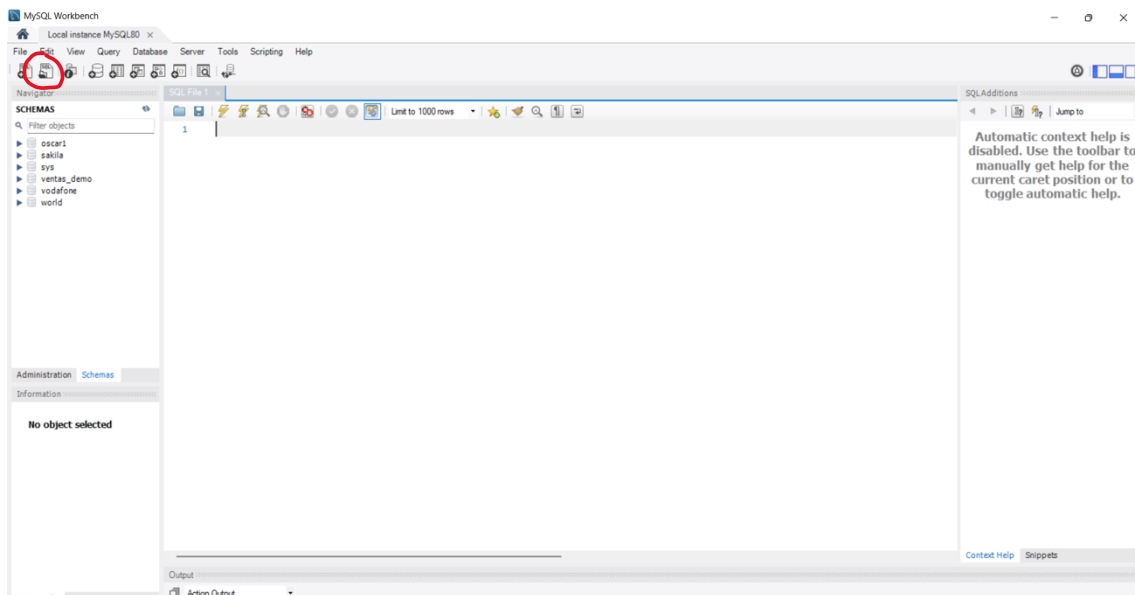
NOTA: Paso previo a la realización del ejercicio:

Volcar en la herramienta MySQLWorkBench el archivo “**ejercicio1.sql**”, con él crearemos la bbdd necesaria para la realización del ejercicio.

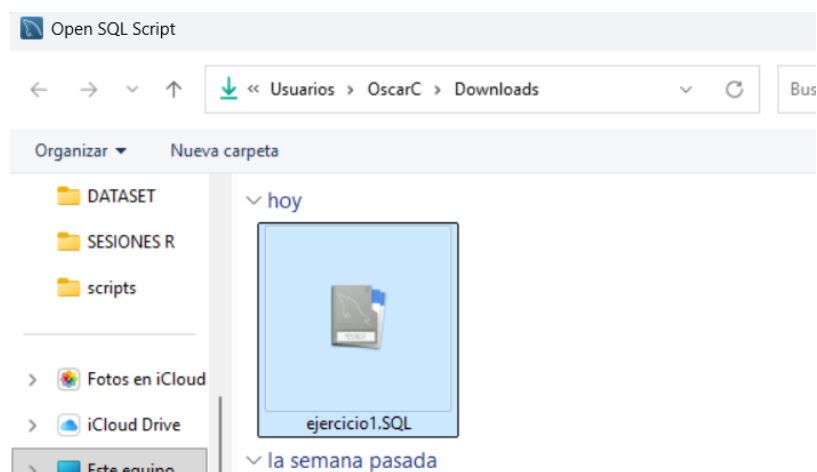
MANUAL SESIÓN

1. SUBIR INFORMACIÓN A LA BASE DE DATOS.

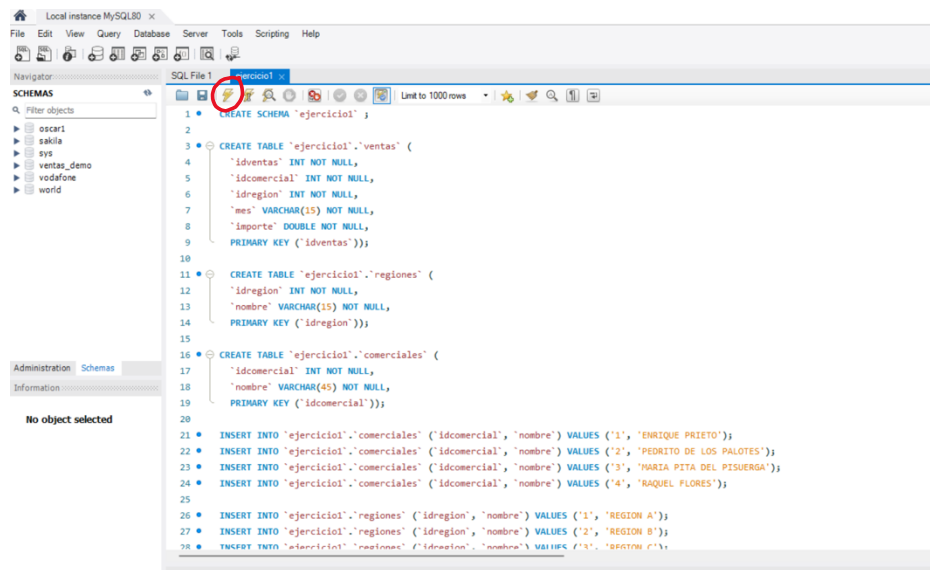
Abrir la herramienta de Workbench y pulsar en el icono de la carpeta que pone SQL.



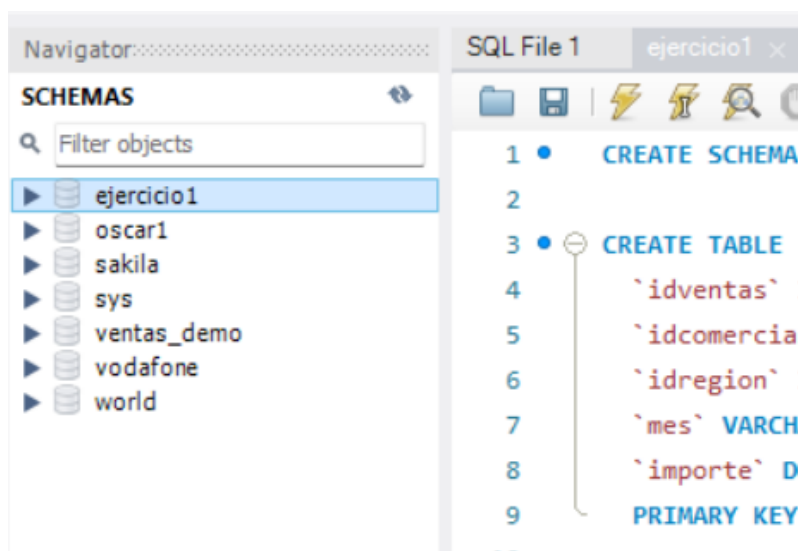
Cargar el fichero ejercicio1.SQL



Cuando se pueda ver el código, ejecutarlo. Para eso, pulsar en el símbolo de ejecutar (El del rayo amarillo).



Puedes actualizar y observar que se ha creado correctamente la base de datos ejercicio1.



2. UNIFICAR LOS DATOS EN PENTHAHO POR REGIONES Y EXTRAER EN CSV.

Usando la herramienta pentaho, sacar la información de las ventas por comerciales y regiones de una base de datos llamada “ejercicio1”. Las tablas son regiones, ventas y comerciales.

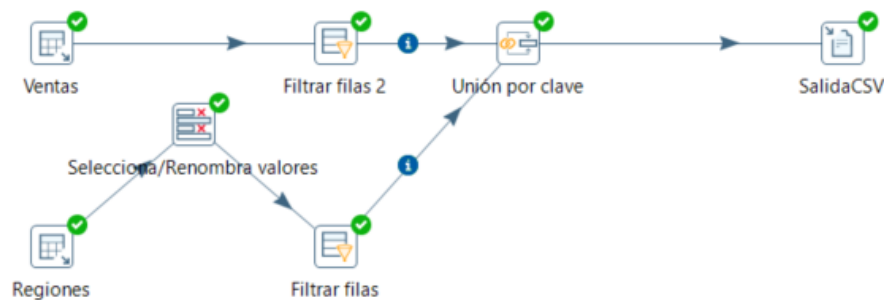
Preparar un csv con las siguientes columnas:

Region, Mes y Ventas

Se filtrará por las Regiones A y B y aquellos comerciales < 3 (idcomercial)

Generará un archivo separado por “,” con la información anterior.

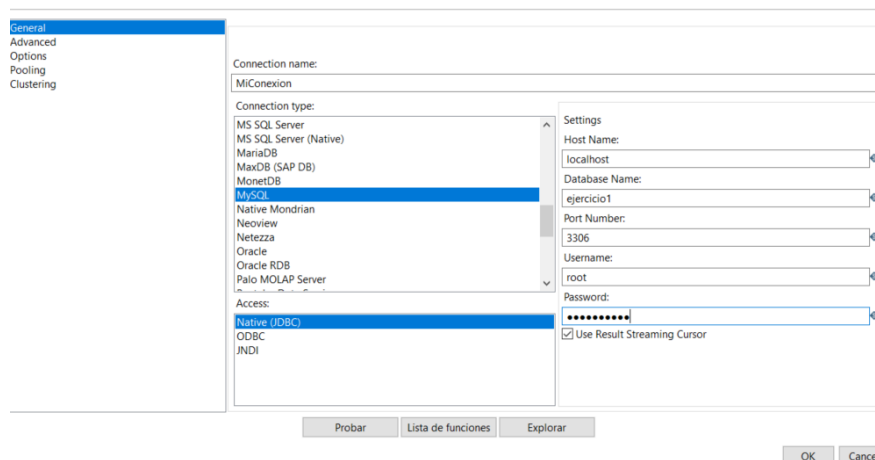
Código:



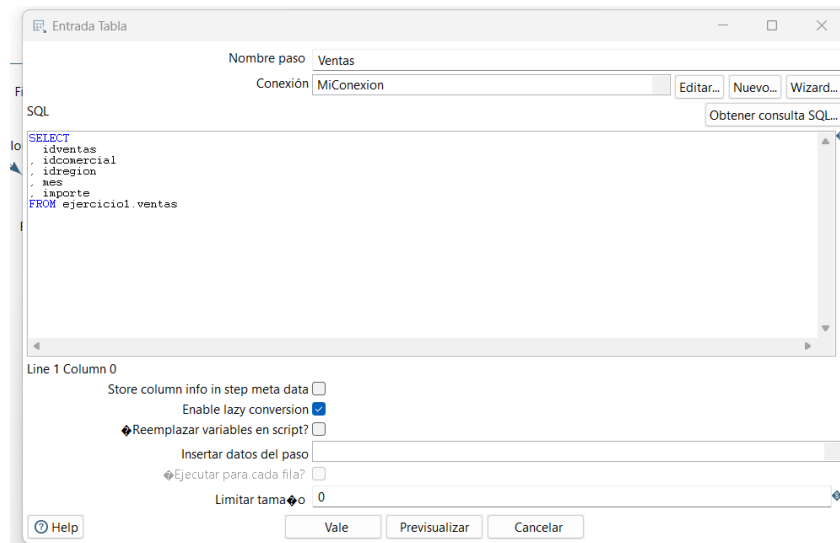
Configuración de cada bloque:

Ventas: Entrada Tabla.

1. Creamos la nueva conexión.

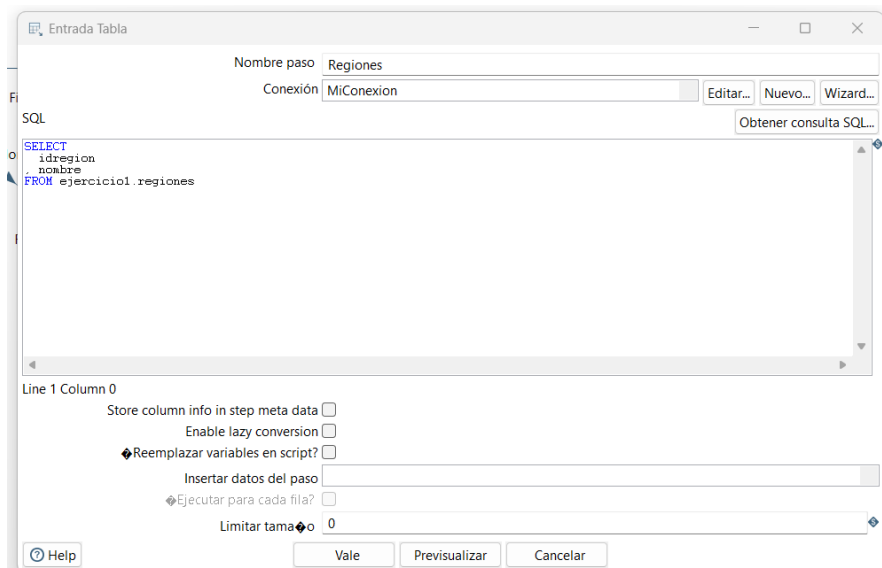


2. Obtener consulta de SQL, seleccionar la tabla y añadir la consulta. Previsualizar y VALE.



Regiones: Entrada Tabla.

Hacemos igual para el bloque de regiones. Como la conexión ya está creada, solo tenemos que hacer la consulta SQL.



Filtrar filas 2: Filtrar Filas.

se c

Nombre de paso: Filtrar filas 2

Enviar 'verdadero' a paso: Unión por clave

Enviar 'falso' a paso:

La condición:

(Integer)

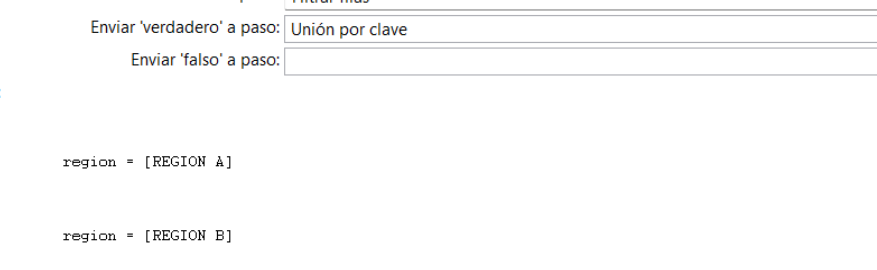
+

Help Vale Cancelar

Selecciona/Renombrar valores:

[illegible]

Filtrar filas:



Filtrar filas

Nombre de paso: Filtrar filas

Enviar 'verdadero' a paso: Unión por clave

Enviar 'falso' a paso:

La condición:

+

region = [REGION A]

OR

region = [REGION B]

Help Vale Cancelar

Unión por clave:

Unión por clave

Nombre de paso: Unión por clave

Primer Paso: Filtrar filas 2

Segundo Paso: Filtrar filas

Tipo Unión: INNER

Claves de primer paso:

#	Campo clave
1	idregion

Claves de segundo paso:

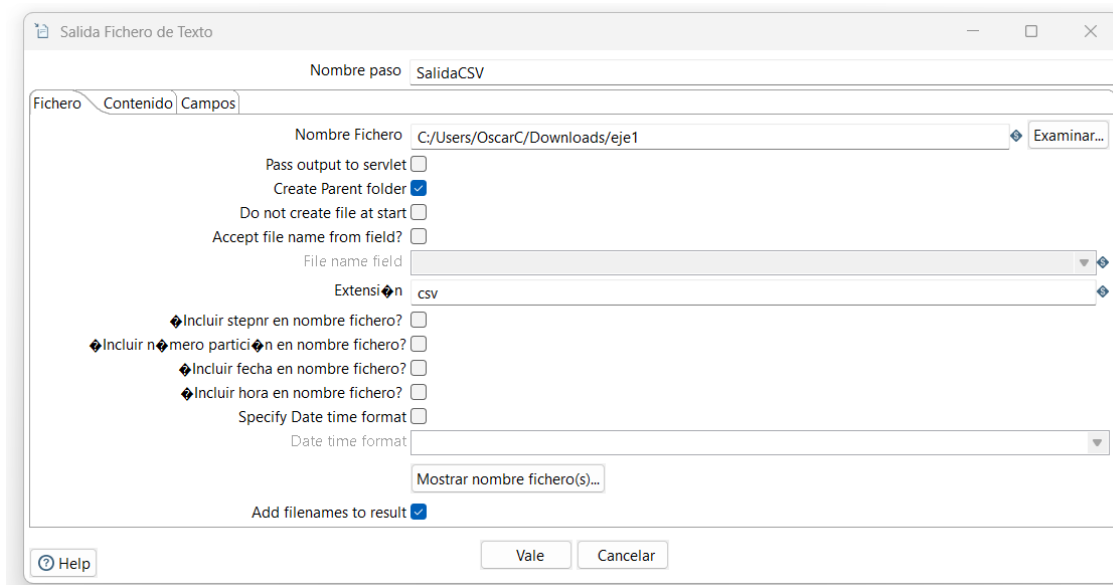
#	Campo clave
1	idregion

Obtener campos clave

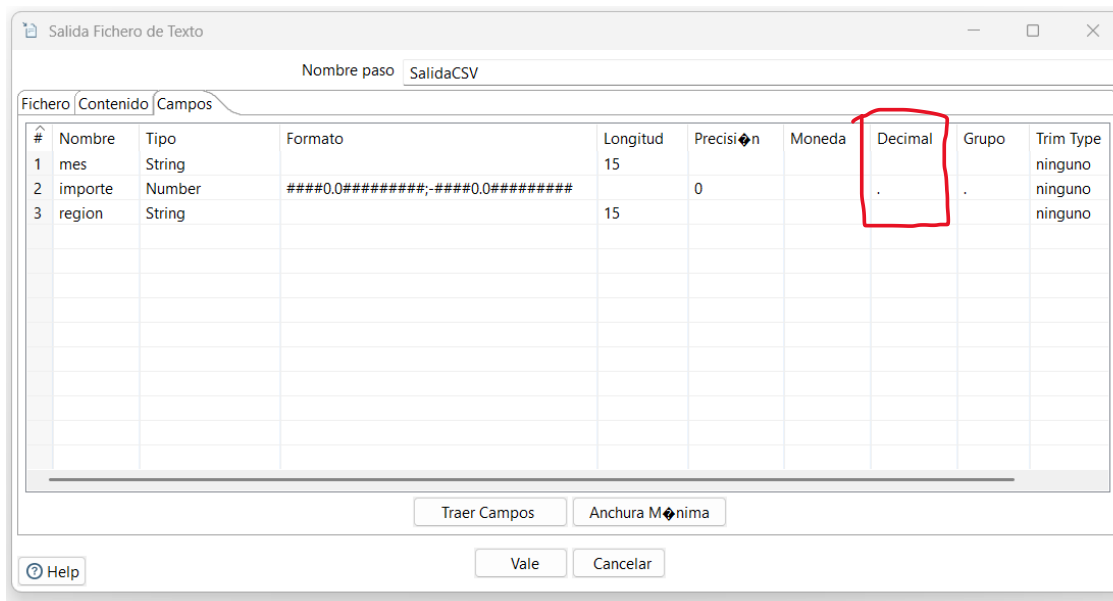
Obtener campos clave

Help Vale Cancelar

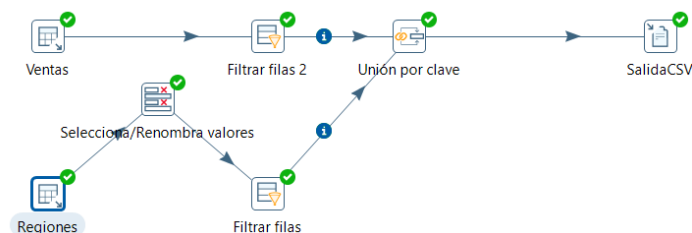
Salida CSV:



VERIFICAR QUE EL DECIMAL ES UN “.”



EJECUTAR EL CÓDIGO:



Si todo ha ido bien, se creará un archivo en descargas así:

Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Automatizar

Pegar

Portapapeles

Fuente

Alineación

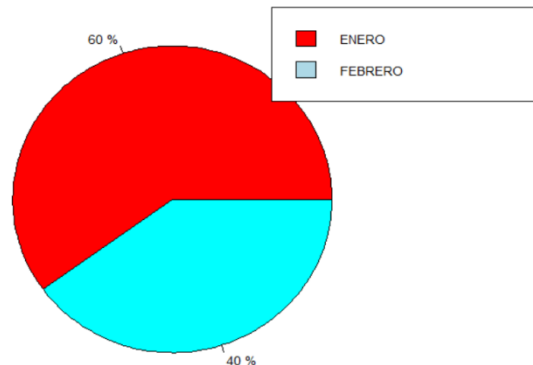
Núm

G7

	A	B	C	D	E	F	G	H	I
1	mes,importe,region								
2	ENERO	,1000.0,REGION A							
3	ENERO	,2000.0,REGION A							
4	ENERO	,1500.0,REGION B							
5	FEBRERO	,5000.0,REGION B							
6	ENERO	,3000.0,REGION B							
7									
8									
9									
10									

3. PROGRAMACIÓN EN R.

Desde R, usando dataframes, recuperar dicha información y generar un gráfico de “quesitos” con la información de ventas agrupadas por mes. Debera tener una apariencia similar a :



CÓDIGO:

```
ruta <- "D:/vodafone/Ejercicios/ejercicio1/ejer1.csv" (PONER LA RUTA DE TU ARCHIVO)
```

```
archivo <- read.csv(ruta)
```

```
# Convertir a dataframe:
```

```
ventas <- data.frame(
```

```
  Region = archivo$region,
```

```
  Mes = archivo$mes,
```

```
  Ventas = archivo$importe)
```

```
# Calcular el total de ventas por región y mes
```

```
total_ventas <- aggregate(Ventas ~ Mes , data = ventas, FUN = sum)
```

```
# Calcular el porcentaje que representa cada entrada
```

```
porcent <- (total_ventas$Ventas/sum(total_ventas$Ventas))*100
```

```
# Crear las ventanas del gráfico.
```

```
x_pos <- 0.5 # Donde va la leyenda en X
```

```
y_pos <- 1 # Donde va la leyenda en Y
```

```
pie(total_ventas$Ventas,labels=paste(round(porcent,2,"%"), col = rainbow(2))
```

```
# Agregar la leyenda en una posición personalizada
```

```
legend(x = x_pos, y = y_pos, legend = total_ventas$Mes,  
      fill = c("red", "lightblue"))
```