*Using the data from the previous session OR from the ORACLE database, select and implement the most adequate type of visualization to answer the following questions from the EDA analysis that you had carried out. Remember to justify your choice:*

# • How much profit do the 10 best clients generate compared to the rest?

Podemos intentar usar esta tabla para ayudarnos a seleccionar que gráfico usar.

Tabla

Descripción generada automáticamente

Ilustración 1

Al ser una comparación, con más de dos dimensiones, baja cardinalidad podríamos usar un grafo de barras.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Ilustración 2

Creamos un *dataframe* con los grupos y los *profits* totales calculados que ya teníamos calculado en de la tarea 4. Posteriormente creamos el gráfico de barras.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Ilustración 3

Como vemos el resto de los clientes es un porcentaje mucho mayor que el del resto de clientes, pero para para la gran cantidad de clientes, se podría decir que los 10 primeros aportan bastante *Profit*. En este caso en especial al ser una composición sería más apropiado usar un grafo de barras apiladas (***stacked bar****).*

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 4

Gráfico, Histograma

Descripción generada automáticamente

Ilustración 5

Este gráfico es útil porque puedes ver el profit total y entender fácilmente la magnitud que representa cada uno.

También podríamos crear un gráfico de tarta (pie):  
Imagen que contiene Texto

Descripción generada automáticamente

Ilustración 6

Gráfico, Gráfico circular

Descripción generada automáticamente

Ilustración 7

Aun así, creo que es el gráfico mas apropiado en este caso es el gráfico de barras apiladas o **stack bar**.

# • How much is the average time and profit of packages according to their category?

Una visualización muy interesante sería ver si el tiempo de preparación y el *profit* en conjunto. Para ello usamos un gráfico de burbujas para ver si encontramos algún patrón.

Texto, Carta

Descripción generada automáticamente

Ilustración 8

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Ilustración 9 Gráfico de puntos

Como podemos ver el tiempo de preparación y el *profit* no parece que tengan una correlación. Usamos el color para representar la dimensión de la categoría. Su lectura es un poco difícil debido al gran volumen de datos, podíamos visualizarlo mejor viendo para cada tiempo de preparación la media del *profit* y tiempo de preparación por cada categoría.

Texto

Descripción generada automáticamente

Ilustración 10

Gráfico

Descripción generada automáticamente

Ilustración 11

Con esta visualización podemos ver como los suministros de oficina son los mas se piden pero los que menos margen de *profit* dejan, lo cuál tiene sentido ya que obtienen mas *profit* por la cantidad de pedidos a un menor precio que por ejemplo los pedidos tecnológicos que son mas caros, se venden menos pero cada uno deja producto deja más *profit*.

Si quisiéramos ver además como se comparan el *profit* y tiempo de preparación entre las categorías entre sí, podríamos usar gráficos de barras

*Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente*

Ilustración 12 Obtenemos los datos necesarios para hacer las vistas.

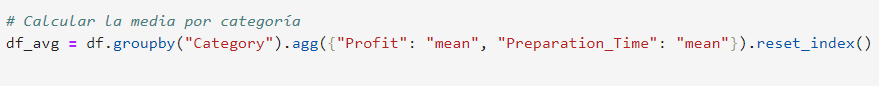


Ilustración 13 Cálculo de media de profit

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 14 Bar plot que compara el profit medio para cada categoría.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 15 Gráfico que compara el tiempo de preparación medio (días) para cada categoría

Aunque no lo piden en el enunciado podríamos ponernos en el hipotético caso que quisiéramos estudiar la distribución del *profit* y del tiempo de preparación de cada categoría, podríamos obtener la mediana y los cuartiles y ver con facilidad los valores anómalos usando un gráfico de cajas y bigotes.

*Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente*

Ilustración 16 Gráfico de cajas y bigotes del profit respecto a las categorías.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Ilustración 17 Gráfico de cajas y bigotes del tiempo de preparación en días respecto a las categorías.

*Focus now on the market that provides the greatest benefit:*

# • How has profit evolved over time in this market?

Los gráficos de líneas son ideales para representar la evolución de un valor cardinal a lo largo del tiempo.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Ilustración 18

El *dataframe* es el *profit* para Europa agrupado por mes que se había obtenido en la tarea anterior. Se combina la librería *seaborn* con *matplotlib* para rotar las etiquetas del eje x y que se visualice mejor el *profit* en cada mes. El resultado es el siguiente:  
Gráfico, Gráfico de líneas

Descripción generada automáticamente

# • What is the average quantity ordered in each region in this market according to the order category?

Esta visualización ha de ser una comparación entre regiones y categorías (**que asumimos que se refiere al modo de envío y no a la categoría del producto**). Como los tipos independientes son región y categoría el *stacked columna chart* es aplicable. La visualización con el gráfico *pie* no es adecuada por comparaciones con datos n-dimensionales y el gráfico de burbujas es menos apropiado debido a la dimensionalidad y tipos dependientes.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 19

Nos ayudamos de la consulta para obtener el número de pedidos agrupados por modo de envío y región para el mercado de Europa.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Ilustración 20

Como vemos hay una gran diferencia de pedidos en cuando a Westen Europe para los envíos de clase estándar.

*Once you have finished these questions, look at the* ***multidimensional schema*** *of the repository.* ***Propose and answer another analytical question*** *using visualizations* ***that involve at least 2 dimensions and one measure.***

Yo propongo responder a la pregunta analítica sobre visualizar el número de productos devueltos a lo largo del tiempo.

Texto

Descripción generada automáticamente

Ilustración 21

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración 22

Parece ser que cada vez el número de devoluciones tiende a crecer, lo cuál tiene sentido porque el número de pedidos también tendía a crecer y probablemente estén mas o menos correlacionados.