

# Instrucciones Miniproyecto 2

## Implementación de casos

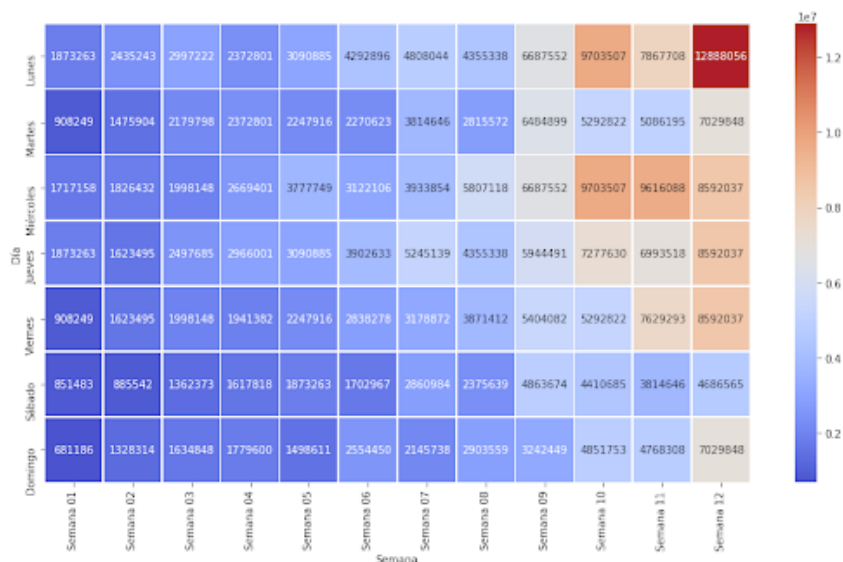
Esta evaluación tiene por propósito poner en práctica la implementación de visualizaciones mediante programación. Se establecen tres casos con *datasets* y codificaciones objetivo para que implementes. Debes entregar un archivo Jupyter Notebook con los resultados generados, utilizando las librerías pandas, matplotlib y seaborn.

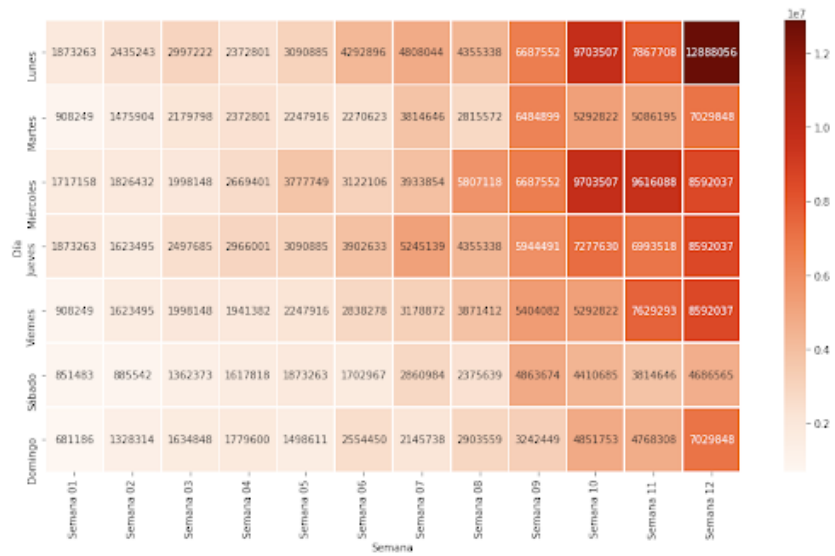
## Caso 1

El año 2020 las plataformas de videoconferencia presentaron un aumento masivo en su uso y flujo de usuarios. Se provee un *dataset* tabular (**videoconferencia.csv -descargar desde plataforma**) que especifica la cantidad de horas de llamadas utilizadas en la plataforma por día y semana del año, hasta la semana 12. Usa este *dataset* y la librería seaborn para visualizar esta información como un mapa de calor, donde las columnas corresponden a las semanas, las filas a los días de la semana, y cada celda se colorea como codificación del valor de horas de llamadas correspondiente. Considera que es necesario preprocesar levemente este *dataset* para utilizar la función heatmap de seaborn. Para ello puedes usar las funciones `read_csv` y `pivot` de pandas.

Además, realiza las siguientes personalizaciones: agrega anotaciones sobre las celdas para leer los valores exactos; prueba una paleta de colores secuencial y otra divergente (**algunas disponibles en el link que se encuentra en plataforma**). Al tener en cuenta la última personalización, responde: ¿cuál paleta es más efectiva para codificar estos datos?

### Resultados posibles:





## Caso 2

El hundimiento del Titanic fue uno de los sucesos más impactantes del siglo XX. De los datos que se han recolectado de los pasajeros, llama especialmente la atención la proporción de pasajeros por clase (First, Second, Third) y cómo se distribuyeron estas clases considerando si sobrevivieron o no al naufragio. Usando el *dataset* del titanic, al cual puede acceder como dataframe usando el comando Python:

```
import seaborn as sns
df_titanic = sns.load_dataset("titanic")
```

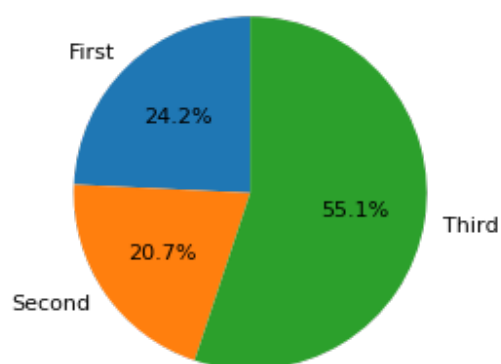
Se le pide que haga lo siguiente:

1. Implemente una visualización con gráfico de tipo torta usando matplotlib o pandas. El gráfico debe mostrar la proporción de pasajeros de cada clase. Para mostrar las proporciones, investigue sobre el argumento `autopct` de `plt.pie`. El título del gráfico de torta debe ser "Pasajeros del Titanic por clase". Cambie la paleta de color por defecto que tiene este tipo de gráfico.
2. Implemente una visualización con dos gráficos de tipo torta usando matplotlib o pandas. Cada gráfico debe mostrar la proporción de pasajeros de cada clase. El primer gráfico debe presentar dicha proporción para el caso de los no sobrevivientes, y el segundo para el caso de los sobrevivientes. El título de este gráfico debe ser: "Pasajeros del Titanic por Supervivencia (0/1) y por Clase (T/F/S)". Para mostrar las proporciones, investigue sobre el argumento `autopct` de `plt.pie`.

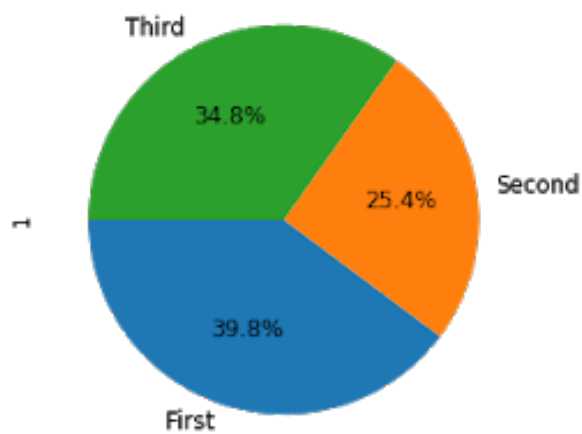
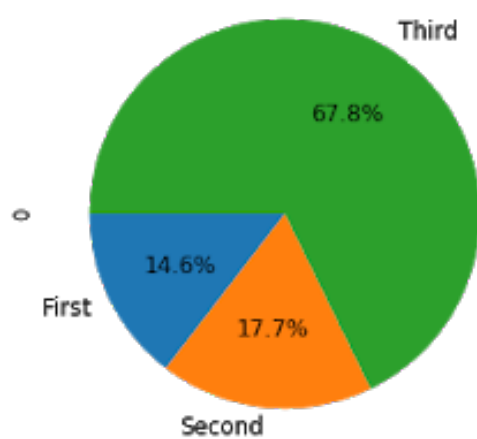
Luego de hacer ambas visualizaciones, indique: a) de los supervivientes, ¿cuál es la clase de pasajero más común?, b) de los no supervivientes ¿cuál es la clase de pasajero más común? c) ¿cree que este gráfico es más efectivo que un gráfico de barras para comparar categorías?

## Resultados posibles:

Pasajeros del Titanic por clase



Pasajeros del Titanic por Supervivencia (0/1) y por Clase (T/F/S)



## Caso 3

Durante los últimos años ha aumentado el interés de las personas por ir al cine a ver películas. Dado lo anterior, durante los años 2011 y 2019 se ha recopilado cuantas películas por género han sido estrenadas para poder determinar si hubo algún género más popular o no. Se provee un *dataset* tabular (**peliculas.csv – descargar desde plataforma**) que especifica la cantidad de películas que han sido estrenada según su género principal y el año de estreno. Usa este *dataset* y las librerías vistas en el curso para visualizar esta información con diferentes *small multiple* (pequeños múltiples). En particular se pide que hagas:

1. Un *small multiple* donde cada gráfico será uno de línea cuya categoría y color será definido según el género. Uno de los ejes será el año y el otro será la cantidad de películas estrenadas en dicho año y con ese género. Personalice la marca para que cada dato del gráfico de línea tenga un símbolo. Recuerde el argumento `marker` que recibe el método `map` permite definir la marca por dato.
2. Un *small multiple* donde cada gráfico será uno de barra horizontal cuya categoría y color será definido según el año. Uno de los ejes será el género y el otro será la cantidad de películas. Recuerde que en el primer argumento del método `map` se indica qué tipo de gráfico se quiere realizar. `plt.barh` define un gráfico de barra horizontal. Además, este gráfico no recibe un argumento `marker`.

Importante: en este gráfico puede que salgan el nombre de los ejes intercambios, es decir, donde están los géneros diga “cantidad” y donde se mapea la cantidad diga “géneros”, esta situación no es necesario arreglar, solo que se vean las barras de forma vertical.

Luego de hacer ambas visualizaciones, responde: ¿cuál de los dos gráficos permite comparar la tendencia de un género entre los diferentes años? y ¿cual de los dos gráficos permite identificar el género de películas más estrenados en 1 año específico?

Resultados posibles:

