

m9-julihocc

March 23, 2024

1 Módulo 9: Visualización, parte 1

1.1 Juliho Castillo Colmenares

1.1.1 Instrucciones

Antes de empezar

En este ejercicio, visualizaras gráficamente datos provenientes de un archivo .CSV mediante la biblioteca matplotlib de Python de manera que puedas resolver el problema que aquí se detalla.

Entregable: Un Jupyter Notebook (archivo de extensión .ipynb), archivo PDF y capturas de pantallas en espacio de respuesta que muestren tanto el código desarrollado como la solución al problema planteado, incluyendo los comentarios que sean pertinentes a las preguntas que se plantean. Entregas sin estos elementos no serán calificadas.

¡Buena suerte!

Paso a paso:

Descargue el archivo “titanic.csv”, relativo a la lista de pasajeros que estuvieron en el hundimiento del famoso “RMS Titanic” el 14 de Abril de 1912. A partir de la columna “Age”, que contiene la edad de cada pasajero, obtenga lo siguiente mediante la librería Matplotlib:

- El Histograma de frecuencias para la variable edad
- El diagrama de caja para la variable edad
- Un gráfico que muestre el número de pasajeros de cada clase (variable “Pclass”)

Interprete sus resultados en cada caso y guarde dichas imágenes en formato .jpg

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt

## Cargamos los datos
data = pd.read_csv("./titanic.csv")
data.head()
```

```
[ ]:      Survived  Pclass                                Name \
0           0        3                                Mr. Owen Harris Braund
1           1        1  Mrs. John Bradley (Florence Briggs Thayer) Cum...
2           1        3                                Miss. Laina Heikkinen
```

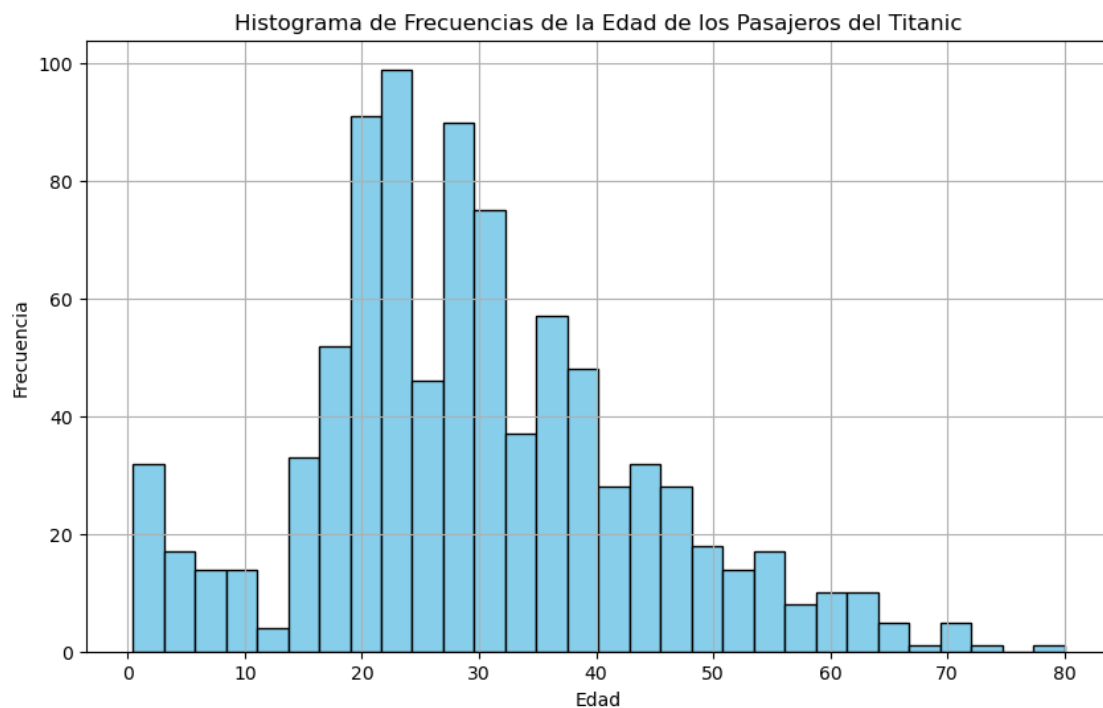
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle
4	0	3	Mr. William Henry Allen

	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	male	22.0	1	0	7.2500
1	female	38.0	1	0	71.2833
2	female	26.0	0	0	7.9250
3	female	35.0	1	0	53.1000
4	male	35.0	0	0	8.0500

```
[ ]: # Generar el histograma de la variable "Age"
plt.figure(figsize=(10, 6))
plt.hist(data['Age'].dropna(), bins=30, color='skyblue', edgecolor='black')
plt.title('Histograma de Frecuencias de la Edad de los Pasajeros del Titanic')
plt.xlabel('Edad')
plt.ylabel('Frecuencia')
plt.grid(True)

# Guardar imagen como archivo .png
plt.savefig('histograma_edad.png')

plt.show()
```



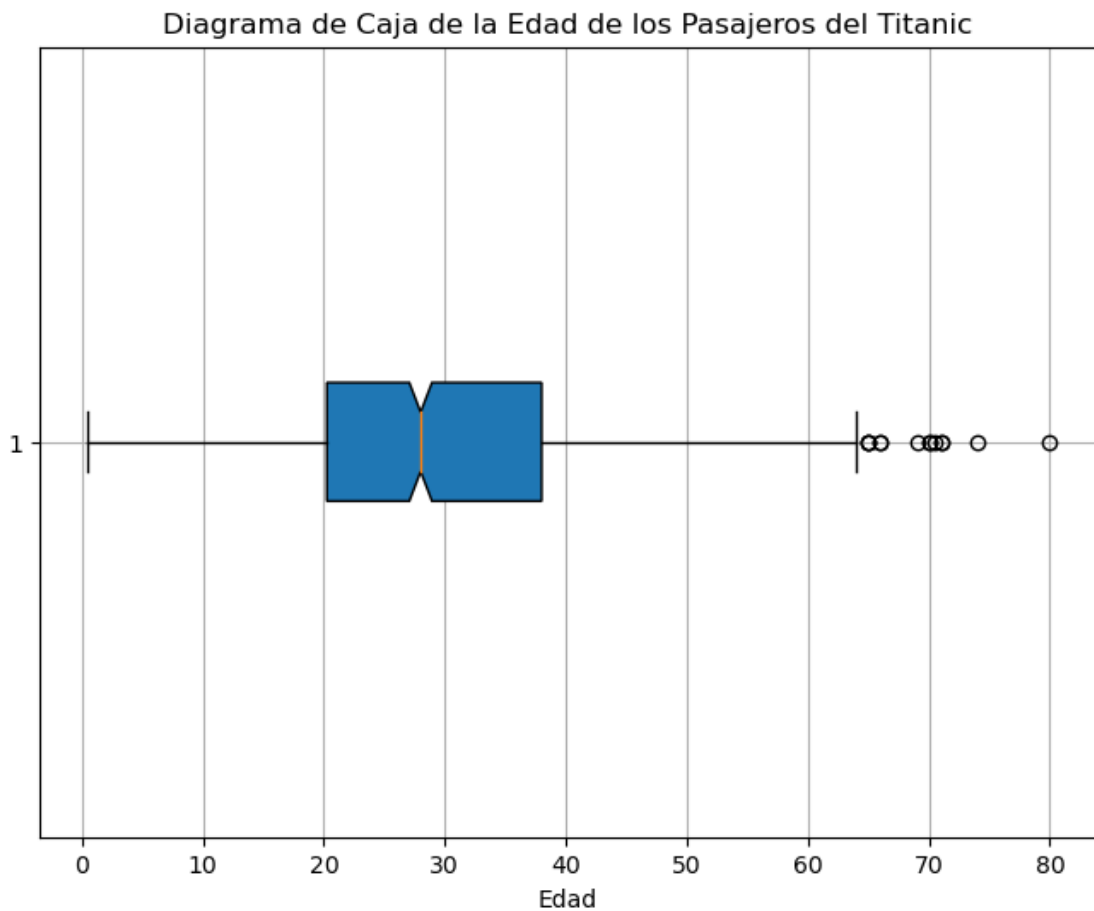
Interpretación: El histograma mostró una concentración significativa de pasajeros entre los 20

y 40 años, indicando que la mayoría de los pasajeros eran jóvenes adultos. Esto puede reflejar la demografía de las personas que viajaban por motivos laborales o aquellos en busca de nuevas oportunidades, lo cual era común en la época.

```
[ ]: # Generar el diagrama de caja para la variable "Age"
plt.figure(figsize=(8, 6))
plt.boxplot(data['Age'].dropna(), vert=False, patch_artist=True, notch=True)
plt.title('Diagrama de Caja de la Edad de los Pasajeros del Titanic')
plt.xlabel('Edad')
plt.grid(True)

# Guardar imagen como archivo .png
plt.savefig('diagrama_caja_edad.png')

plt.show()
```



Interpretación: El diagrama de caja habría revelado la mediana de la edad, así como el rango intercuartílico, mostrando la dispersión de las edades de los pasajeros. Los valores atípicos habrían indicado pasajeros significativamente más jóvenes o mayores que el grueso de la población a bordo,

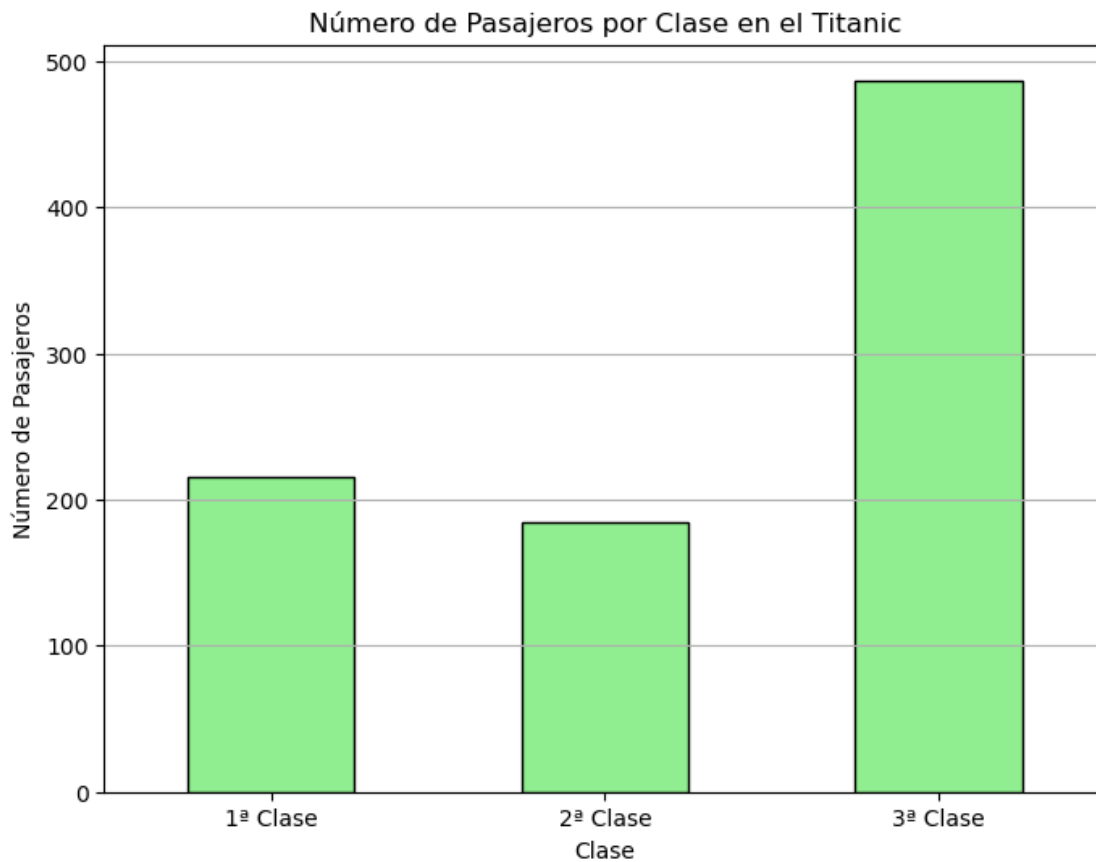
destacando la diversidad en la edad de los viajeros.

```
[ ]: # Contar el número de pasajeros en cada clase
pasajeros_por_clase = data['Pclass'].value_counts().sort_index()

# Generar el gráfico de barras para el número de pasajeros por clase
plt.figure(figsize=(8, 6))
pasajeros_por_clase.plot(kind='bar', color='lightgreen', edgecolor='black')
plt.title('Número de Pasajeros por Clase en el Titanic')
plt.xlabel('Clase')
plt.ylabel('Número de Pasajeros')
plt.xticks(ticks=[0, 1, 2], labels=['1ª Clase', '2ª Clase', '3ª Clase'],
           rotation=0)
plt.grid(axis='y')

# Guardar imagen como archivo .png
plt.savefig('barras_pasajeros_por_clase.png')

plt.show()
```



Interpretación: Este gráfico habría mostrado la distribución de los pasajeros a través de las diferentes clases, reflejando las diferencias socioeconómicas de los pasajeros. Típicamente, se esperaría ver una mayor cantidad de pasajeros en la tercera clase, lo que indica que el Titanic transportaba principalmente a inmigrantes y personas de clases trabajadoras, mientras que las primeras y segundas clases alojaban a viajeros de mayor poder adquisitivo.