

# TITANIC

March 10, 2025

## 1 Titanic dataset analysis

Esta tarea implica la limpieza y el análisis del conjunto de datos del Titanic. El dataset contiene información sobre pasajeros del Titanic, incluyendo variables como edad, clase, tarifa pagada y si sobrevivieron o no. El conjunto de datos está disponible en Kaggle y contiene información sobre los pasajeros del Titanic, como su edad, clase, tarifa, etc.

### 1.1 Importación y limpieza del dataset

Se cargan los datos y se revisa su estructura para entender qué información contiene el dataset.

- Dataset Head

	PassengerId	Survived	Pclass	Name \
0	1	0	3	Braund
1	2	1	1	Cumings
2	3	1	3	Heikkinen
3	4	1	1	Futrelle
4	5	0	3	Allen

	Lastname	Sex	Age	SibSp	Parch \
0	Mr. Owen Harris	male	22.0	1	0
1	Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0
2	Miss. Laina	female	26.0	0	0
3	Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0
4	Mr. William Henry	male	35.0	0	0

	Ticket	Fare	Embarked
0	A/5 21171	7.2500	S
1	PC 17599	71.2833	C
2	STON/O2. 3101282	7.9250	S
3	113803	53.1000	S
4	373450	8.0500	S

- Dataset Info

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null   int64
1   Survived        891 non-null   int64
2   Pclass          891 non-null   int64
3   Name            891 non-null   object
4   Lastname        891 non-null   object
5   Sex             891 non-null   object
6   Age            891 non-null   float64
7   SibSp           891 non-null   int64
8   Parch           891 non-null   int64
9   Ticket          891 non-null   object
10  Fare            891 non-null   float64
11  Embarked        891 non-null   object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

- Columnas numéricas

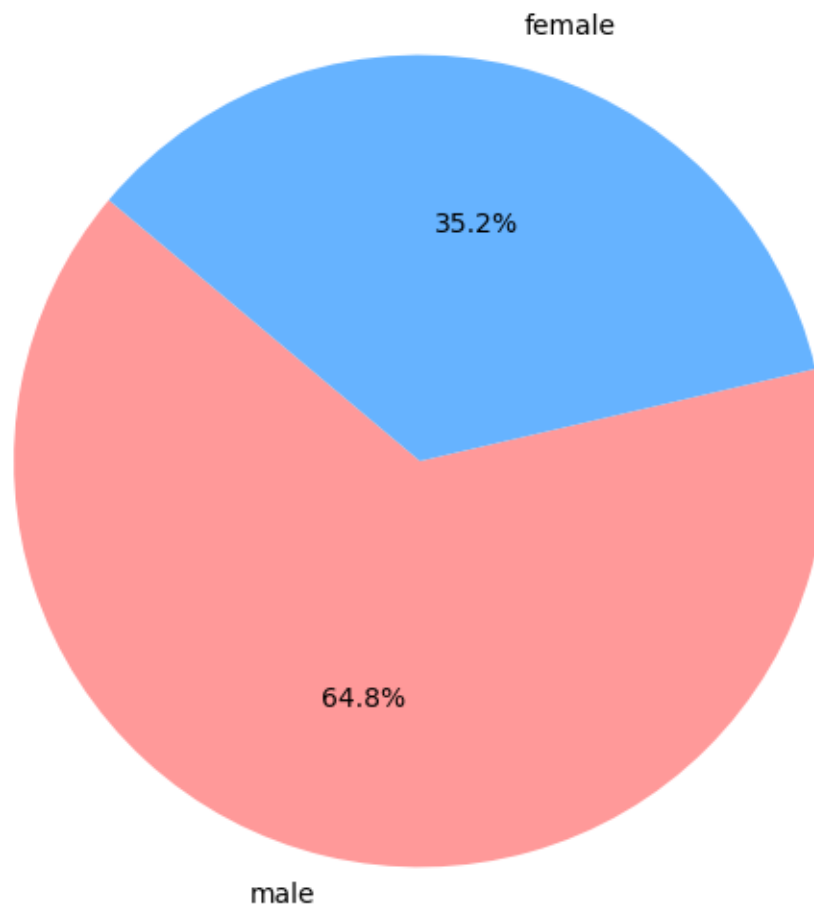
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Survived        891 non-null   int64
1   Pclass          891 non-null   int64
2   Age            891 non-null   float64
3   SibSp           891 non-null   int64
4   Parch           891 non-null   int64
5   Fare            891 non-null   float64
dtypes: float64(2), int64(4)
memory usage: 41.9 KB
```

- Resumen de datos

	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.385152	0.523008	0.381594	32.204208
std	0.486592	0.836071	13.259656	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	21.000000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	30.000000	0.000000	0.000000	14.454200
75%	1.000000	3.000000	35.000000	1.000000	0.000000	31.000000
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

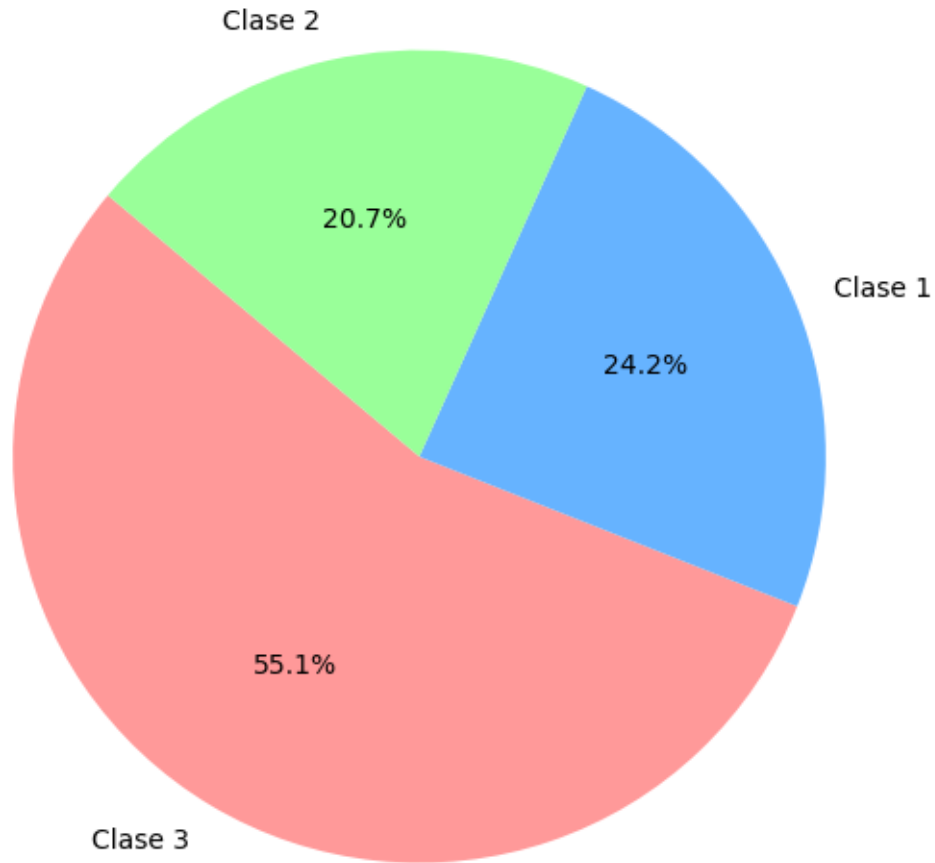
## 1.2 Distribuciones generales

### Distribución de pasajeros según género



Proporción de hombres y mujeres en el Titanic

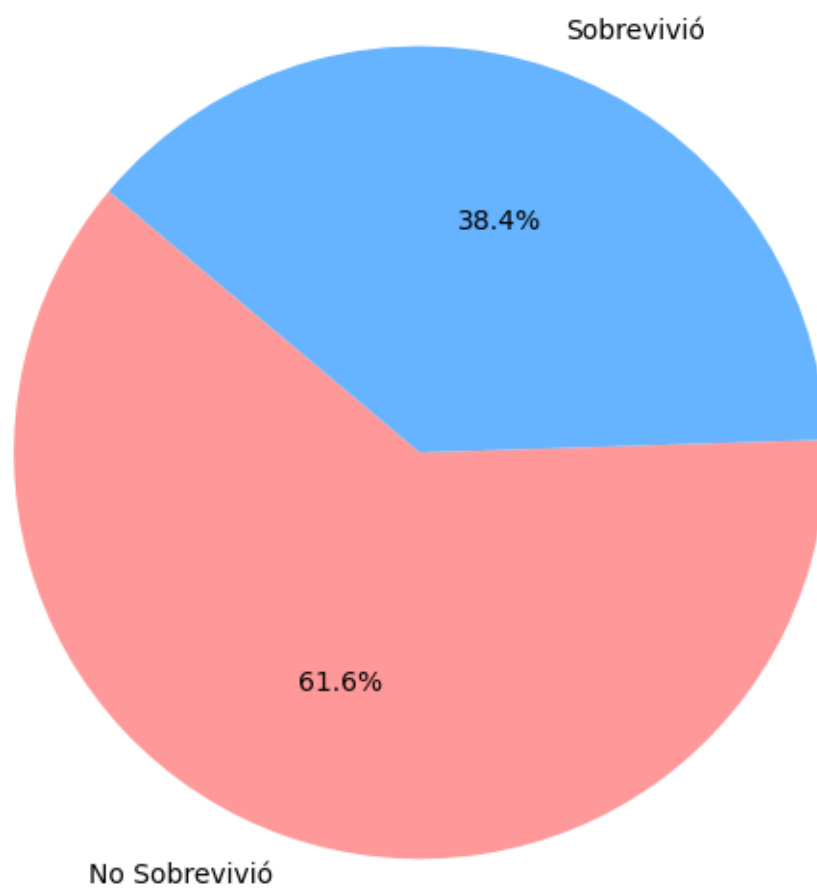
## Distribución de Pasajeros según Clase



Dado que la clase, a pesar de estar determinada con un valor numérico, se trata de una variable categórica, carece de sentido analizarla por su distribución estadística. Una mejor forma de representar esa información puede ser con gráficos especializados en mostrar variables categóricas.

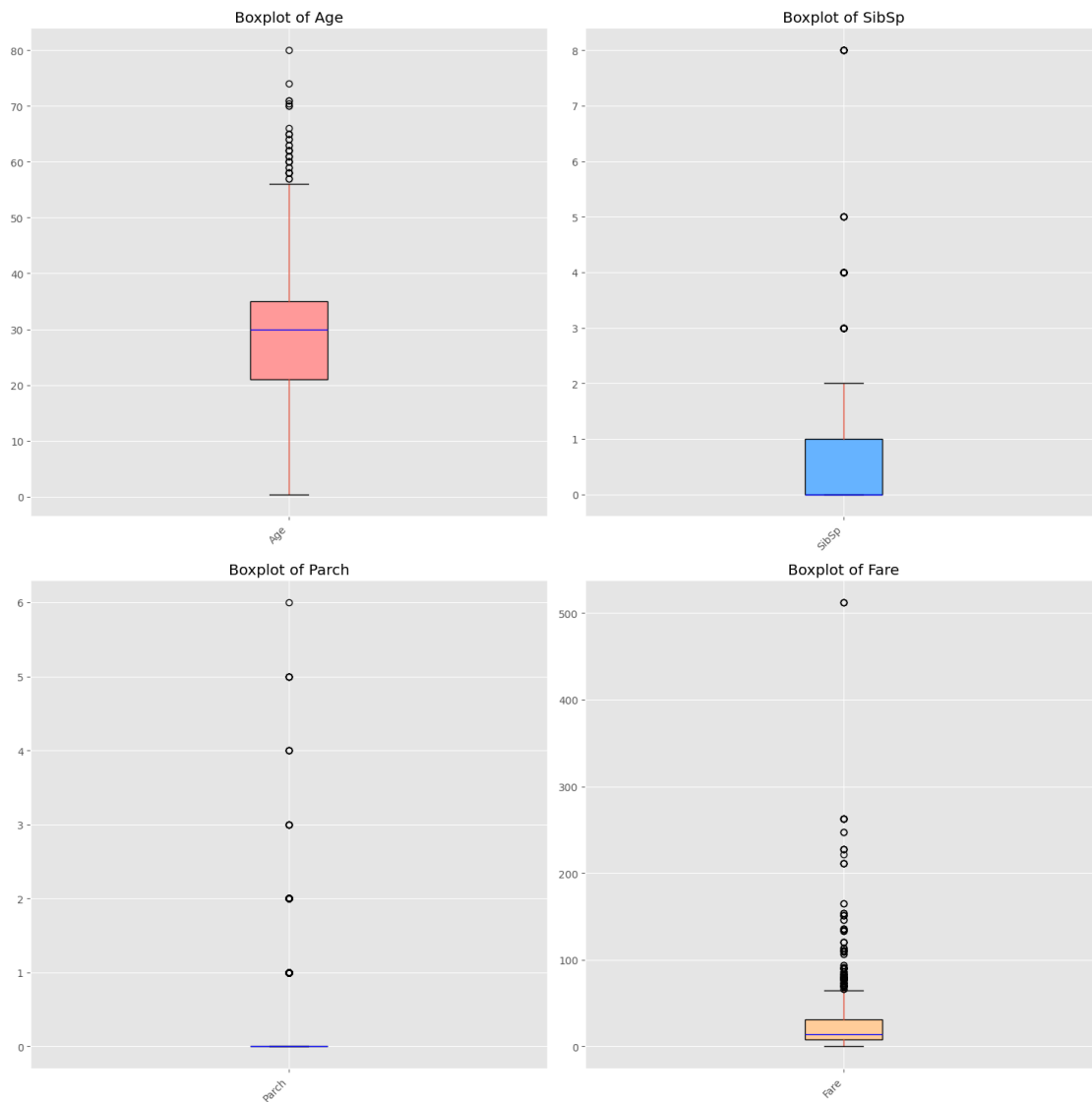
Como podemos observar, la gran mayoría de pasajeros se encontraban en **tercera clase**.

## Distribución de pasajeros sobrevivientes



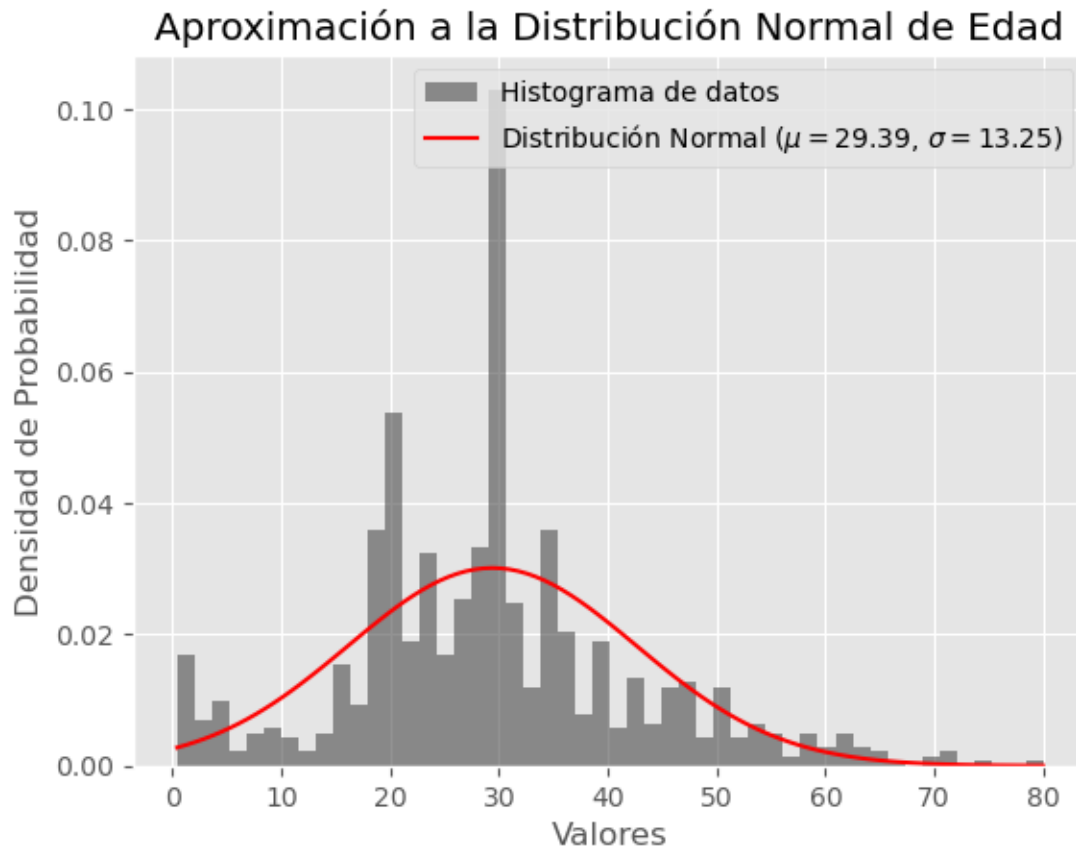
Porcentaje de sobrevivientes y no sobrevivientes.

### 1.3 Boxplots



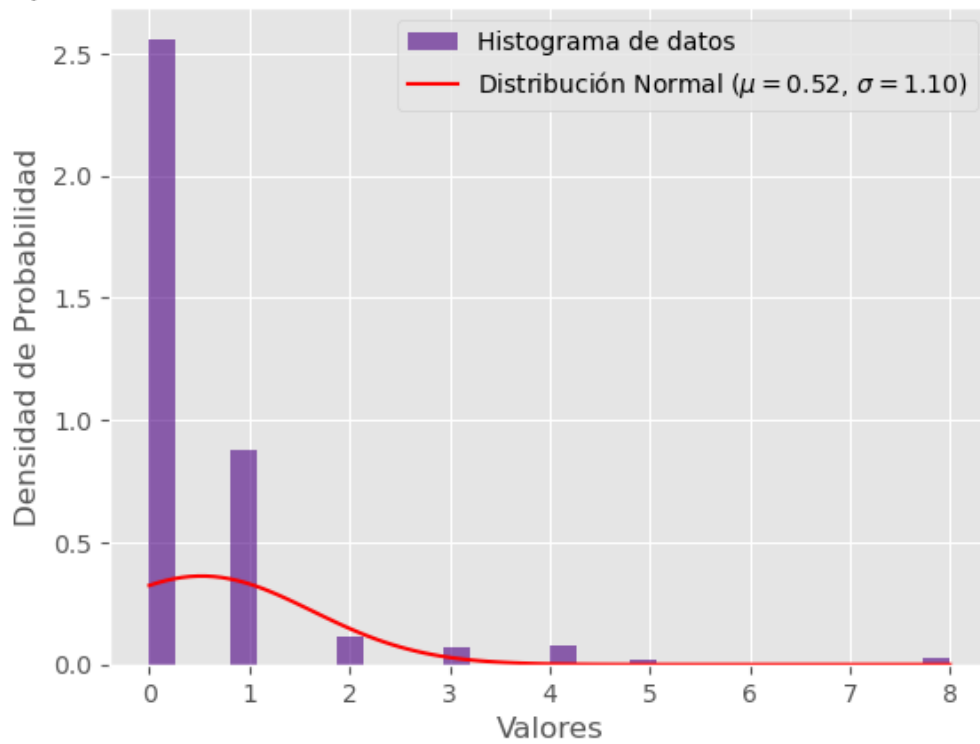
- Age: Los datos se concentran entre los 20 y 35 años. La mediana se encuentra en los 30 años.
- SibSp: La mayoría viajaba sin hermanos o esposos. Hay outliers con valores de 3 a 8.
- Parch: Muchos viajaron sin padres ni hijos. Hay outliers con valores de 1 a 6.
- Fare: La mayoría pagaron tarifas cercanas a la mediana. Existen outliers en tarifas altas.

## 1.4 Histogramas



Distribución de las edades de los pasajeros a bordo del Titanic. La mayoría están entre los 20 y 30 años.

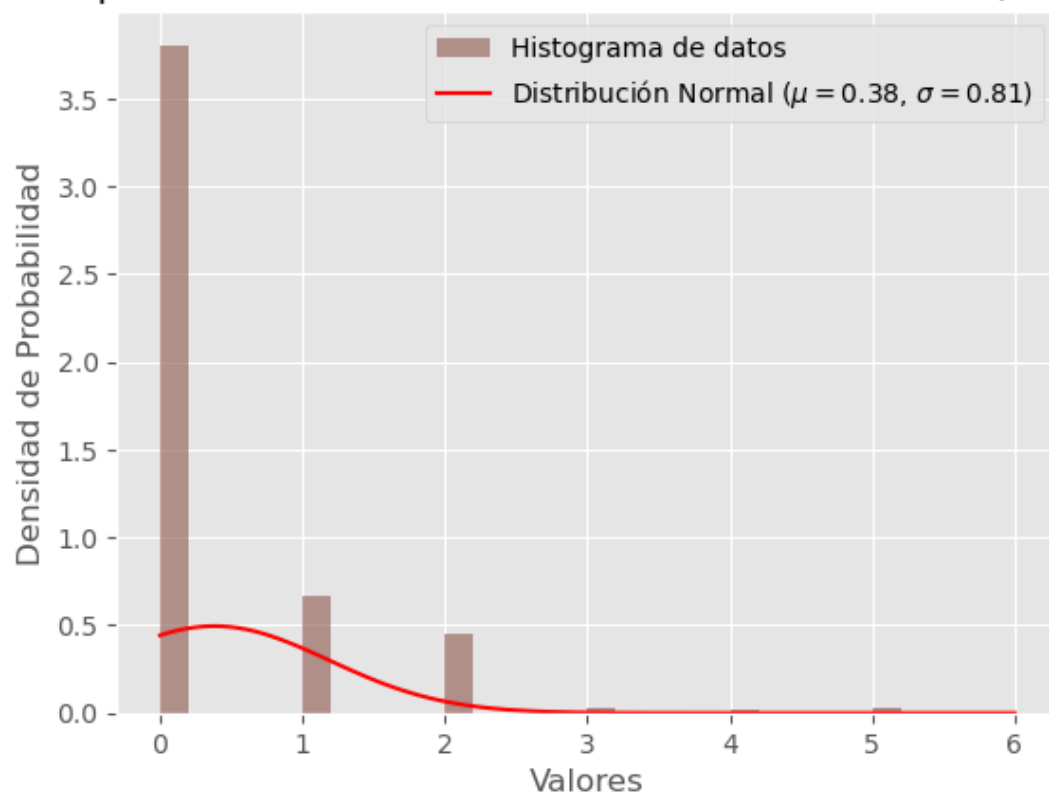
### Aproximación a la Distribución Normal de Hermanos/Esposos



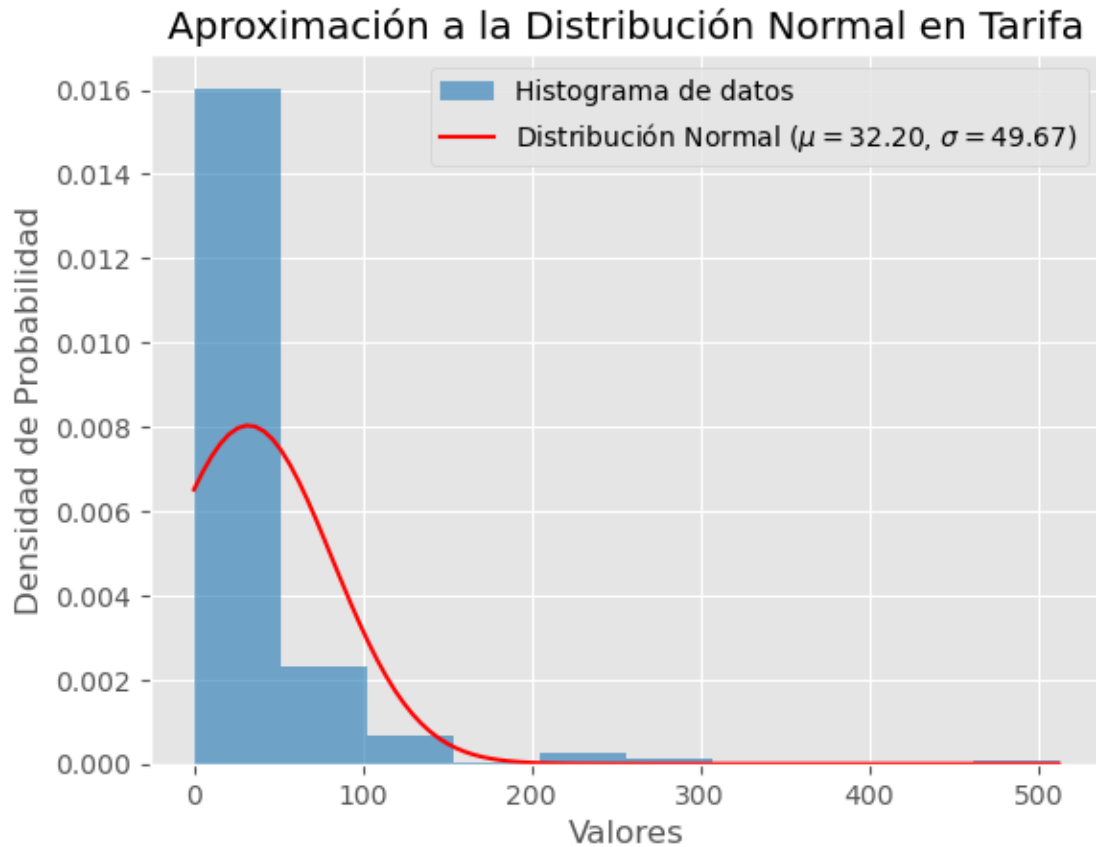
El histograma de la variable SibSp revela que la mayoría de los pasajeros tenían pocos o ningún hermano o esposo a bordo, con un pico en los valores cercanos a cero.



### Aproximación a la Distribución Normal de Padres/Hijos

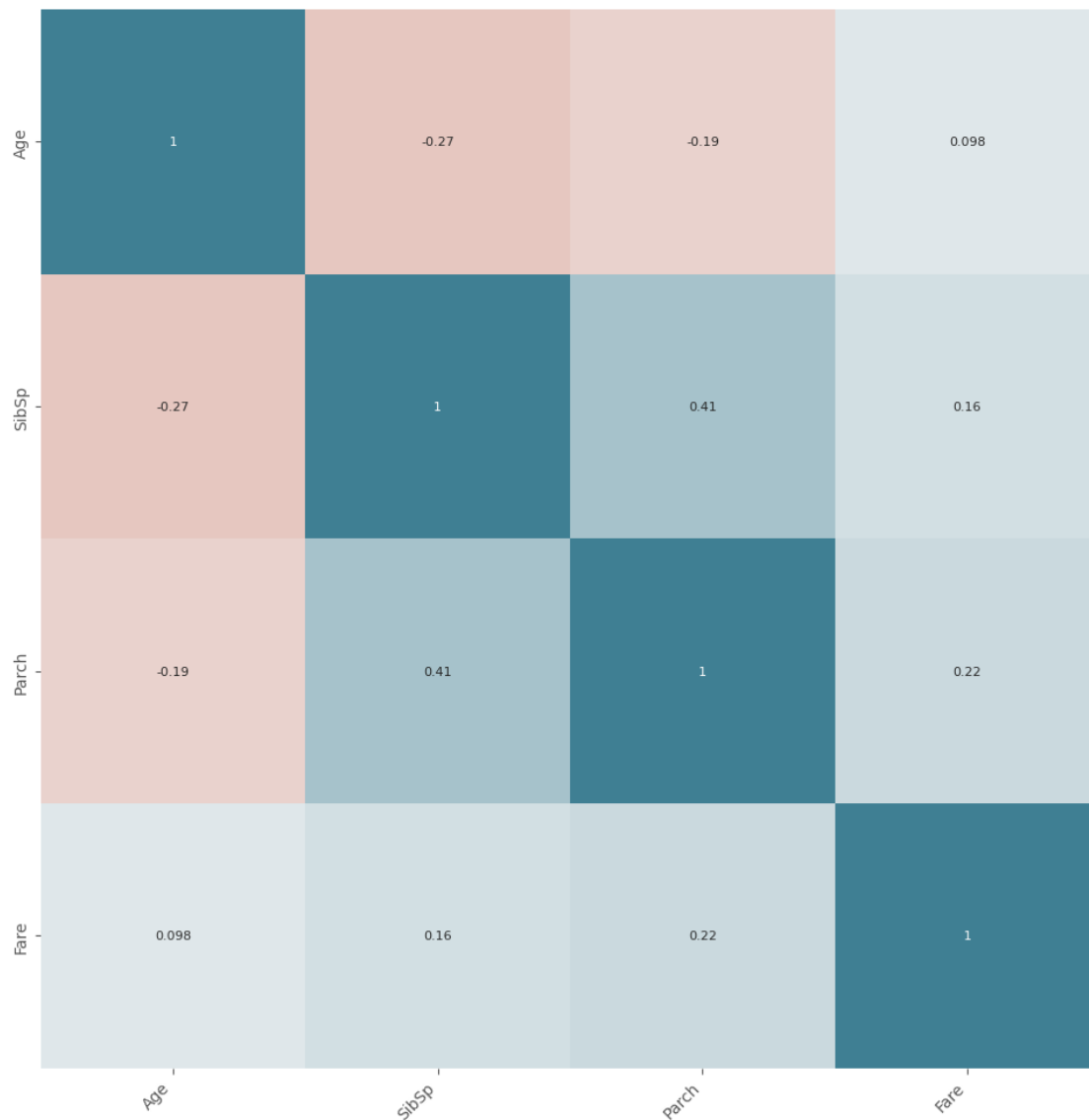


La variable Parch muestra que la mayoría de los pasajeros no viajaban con sus padres ni con hijos, ya que la mayoría tiene un valor de cero.



La mayoría de los pasajeros pagaron tarifas relativamente bajas por el pasaje. Existen algunos pasajeros que pagaron tarifas significativamente más altas, lo que indica que había una clase de pasajeros de alto estatus que viajaba en camarotes más lujosos. La distribución está sesgada hacia la izquierda, con una cola a la derecha que refleja a los pocos pasajeros con tarifas altas.

## 1.5 Matriz de correlación



### Ejemplo de grafico interactivo con plotly

<IPython.core.display.HTML object>

### Comando para generar reporte PDF

Se utilizo nbconvert para guardar los datos en pdf

```
[NbConvertApp] Converting notebook TITANIC.ipynb to pdf
c:\ProgramData\miniconda3\share\jupyter\nbconvert\templates\latex\display_prio
ty.j2:32: UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is
```

```

not able to be represented.
  ((*- endblock -*))
c:\ProgramData\miniconda3\share\jupyter\nbconvert\templates\latex\display_prio
ty.j2:32: UserWarning: Your element with mimetype(s)
dict_keys(['application/vnd.plotly.v1+json', 'text/html']) is not able to be
represented.
  ((*- endblock -*))
[NbConvertApp] Support files will be in titanic_reports/TITANIC3_files\
[NbConvertApp] Making directory .\titanic_reports/TITANIC3_files\titanic_reports
[NbConvertApp] Writing 70128 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | b had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 343733 bytes to titanic_reports\TITANIC3.pdf

[NbConvertApp] Converting notebook TITANIC.ipynb to pdf
c:\ProgramData\miniconda3\share\jupyter\nbconvert\templates\latex\display_prio
ty.j2:32: UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is
not able to be represented.
  ((*- endblock -*))
c:\ProgramData\miniconda3\share\jupyter\nbconvert\templates\latex\display_prio
ty.j2:32: UserWarning: Your element with mimetype(s)
dict_keys(['application/vnd.plotly.v1+json', 'text/html']) is not able to be
represented.
  ((*- endblock -*))
[NbConvertApp] Support files will be in titanic_reports/TITANICr_files\
[NbConvertApp] Making directory .\titanic_reports/TITANICr_files\titanic_reports
[NbConvertApp] Writing 31511 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | b had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 301849 bytes to titanic_reports\TITANICr.pdf

```

## 1.6 Equipo:

- Coconi Dafne
- Cortés López
- Sánchez Erik
- Villegas Getsemaní