TITANIC

March 9, 2025

1 Titanic dataset analysis

1.0.1 Esta tarea implica la limpieza y el análisis del conjunto de datos del Titanic.

1.0.2 El dataset contiene información sobre pasajeros del Titanic, incluyendo variables como edad, clase, tarifa pagada y si sobrevivieron o no.

1.0.3 El conjunto de datos está disponible en Kaggle y contiene información sobre los pasajeros del Titanic, como su edad, clase, tarifa, etc.

1.0.4 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	Lastname	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund	Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings	Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen	Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle	Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen	Mr. William Henry	male	35.0	0	0	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
```

1.0.5 Columnas numéricas

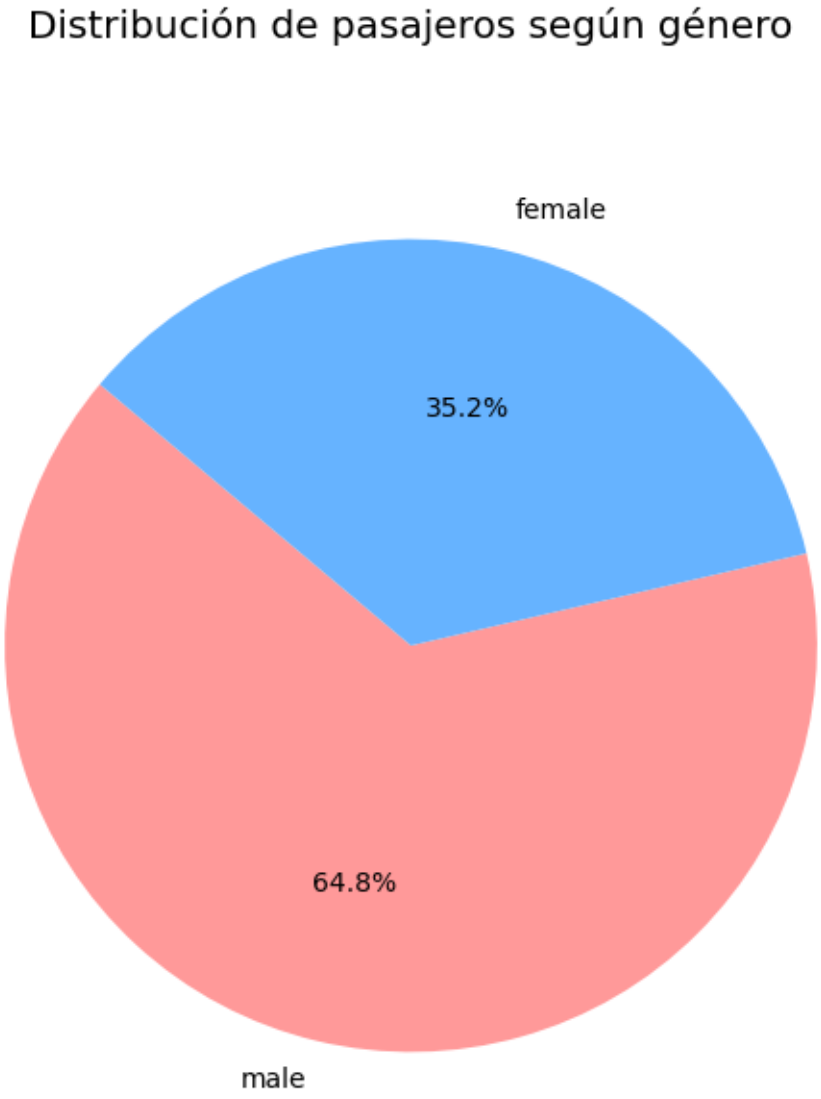
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Survived  891 non-null    int64
 1   Pclass    891 non-null    int64
 2   Age       891 non-null    float64
 3   SibSp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

1.0.6 Resumen de datos

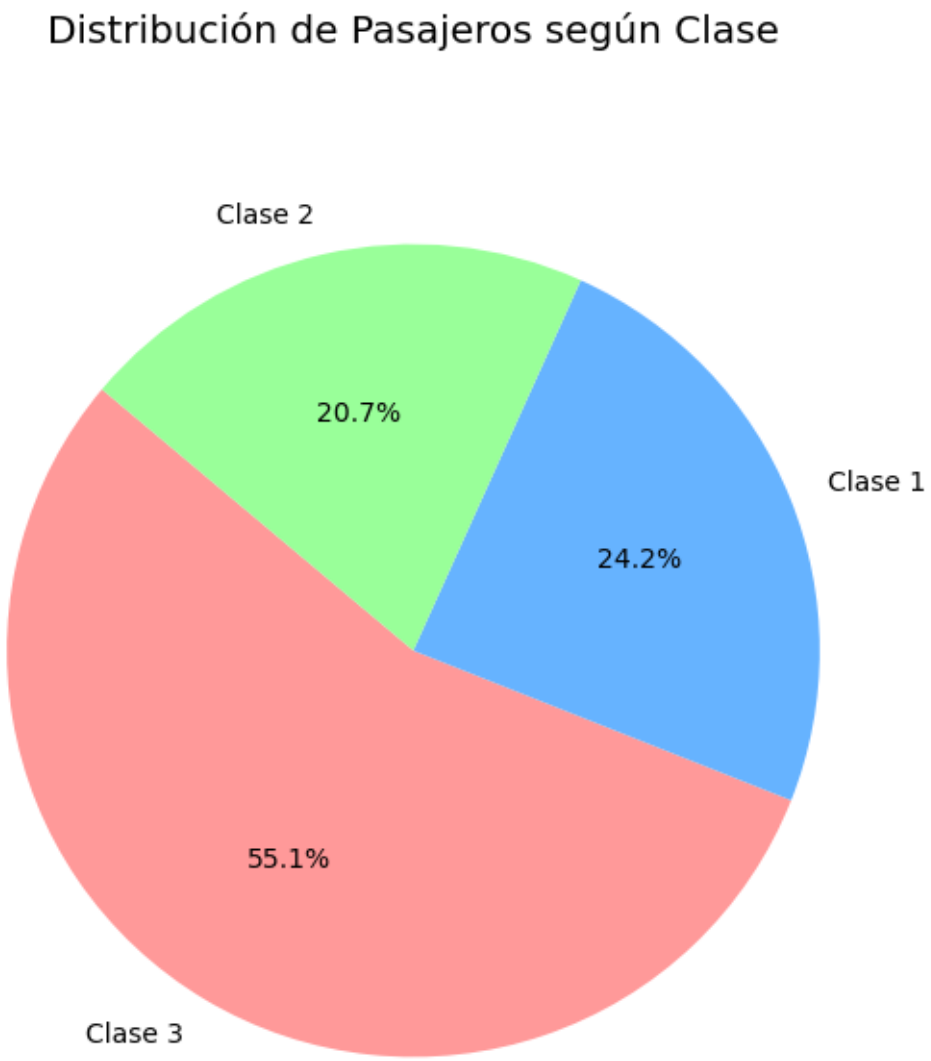
	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.385152	0.523008	0.381594	32.204208
std	0.486592	0.836071	13.259656	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	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

1.0.7 Distribuciones generales

Distribución de pasajeros según género Proporción de hombres y mujeres en el Titanic



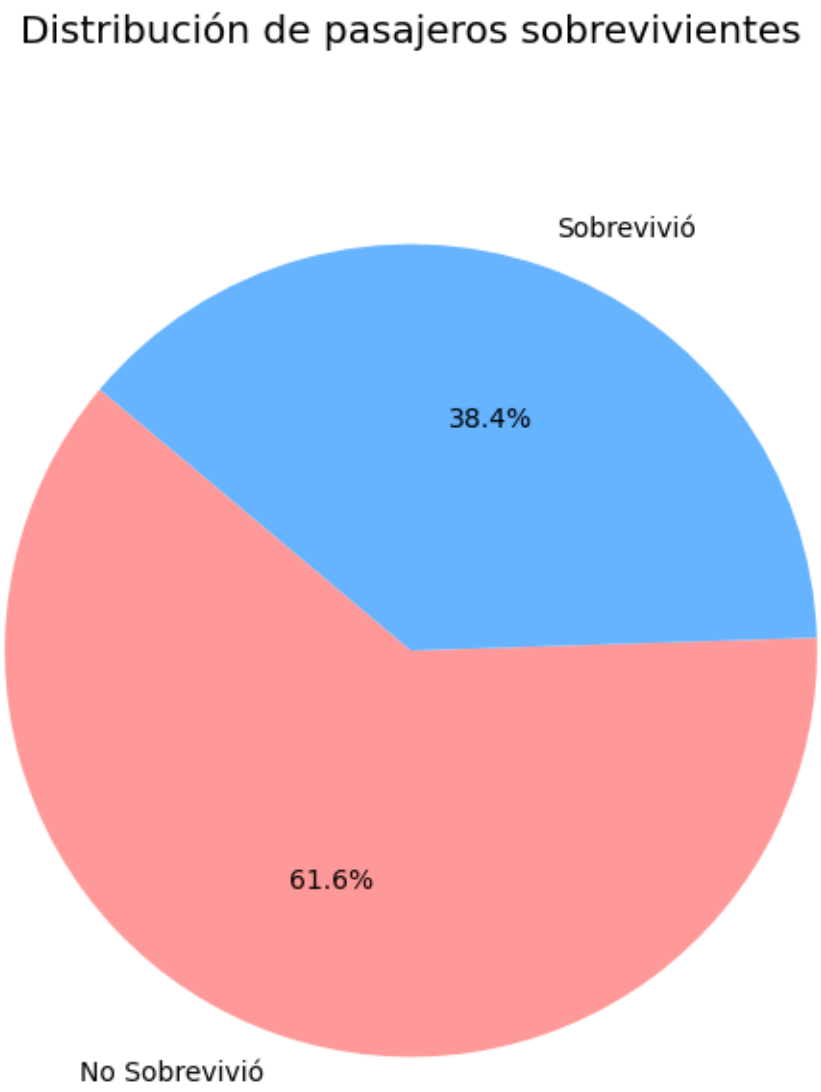
Distribución según clase

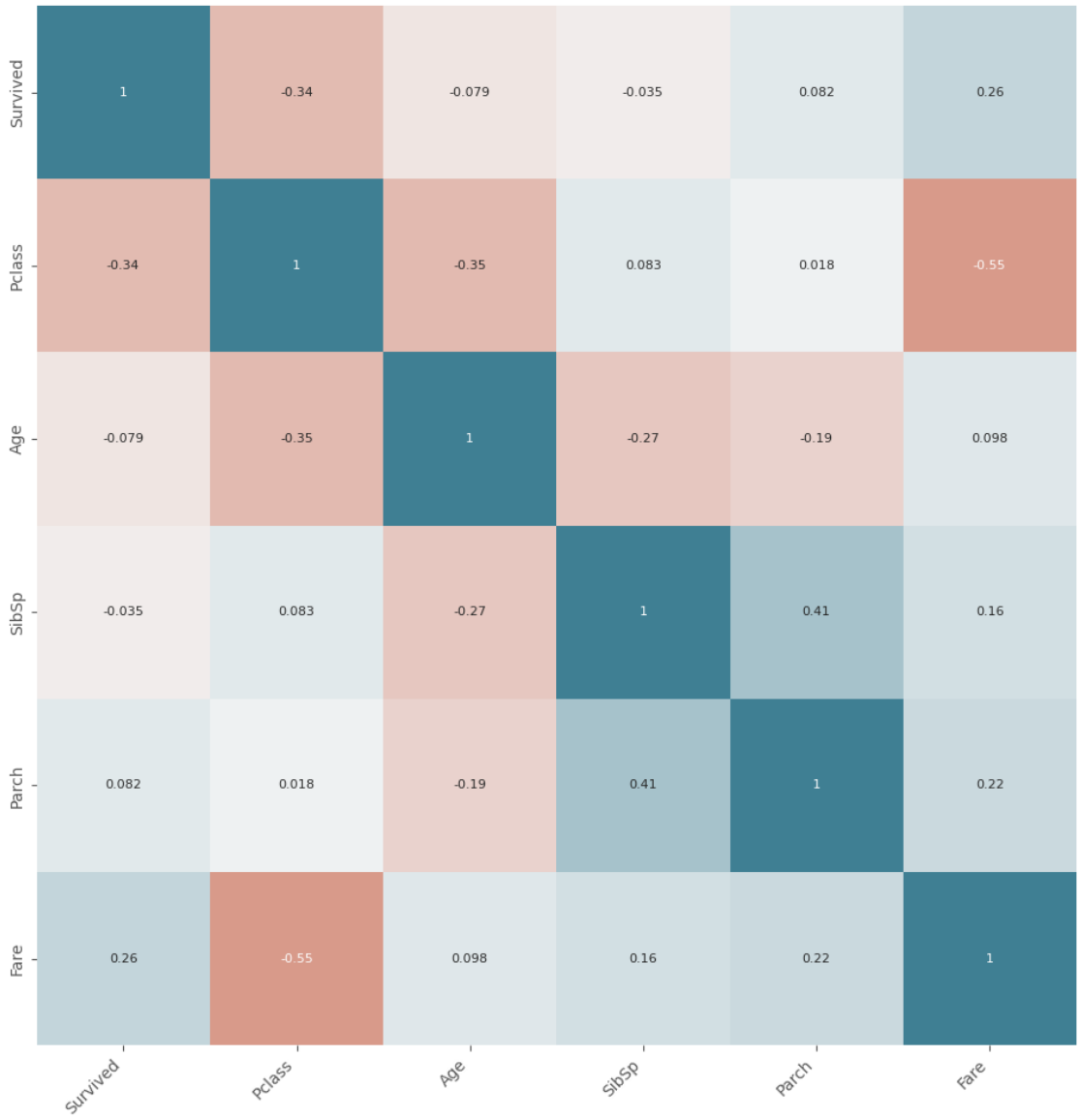


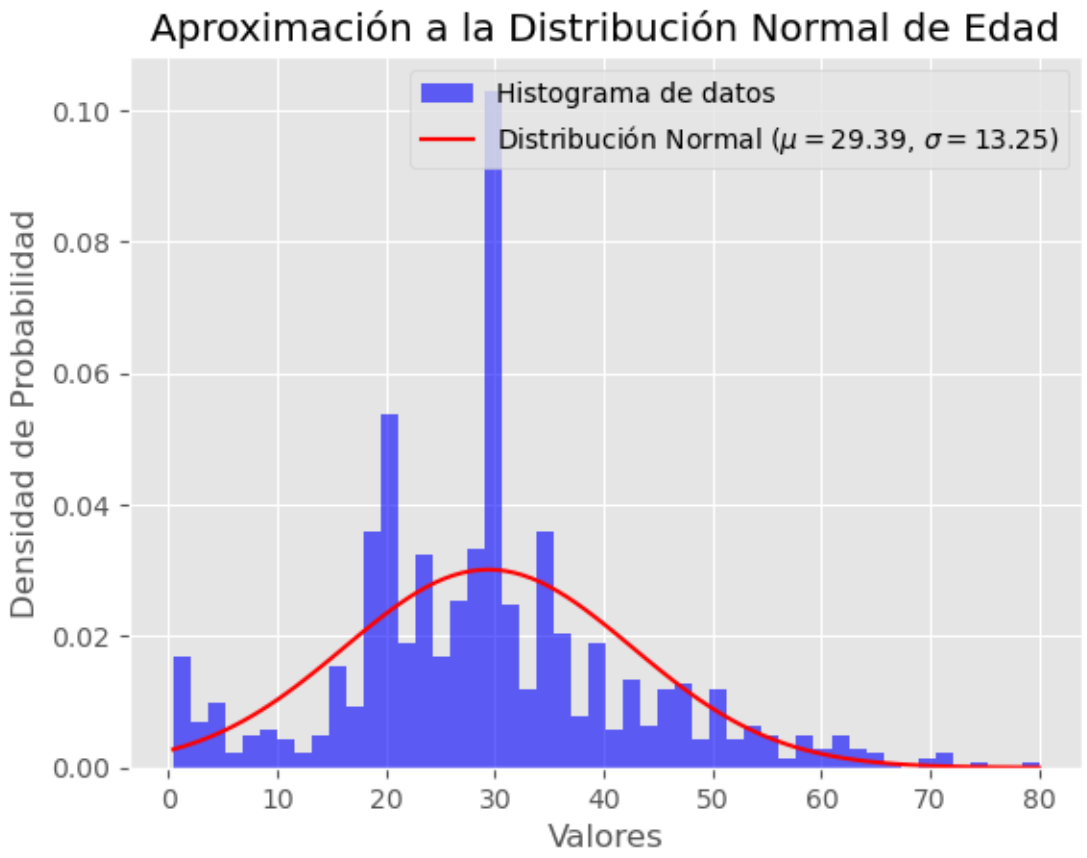
Dado que la clase, a pesar de estar determinada con un valor númerico, se trata de una variable categórica, carece de sentido analizarla por su distribución estadística. Una mejor forma de representar esa información puede ser con gráficos especializados en mostrar variables categóricas.

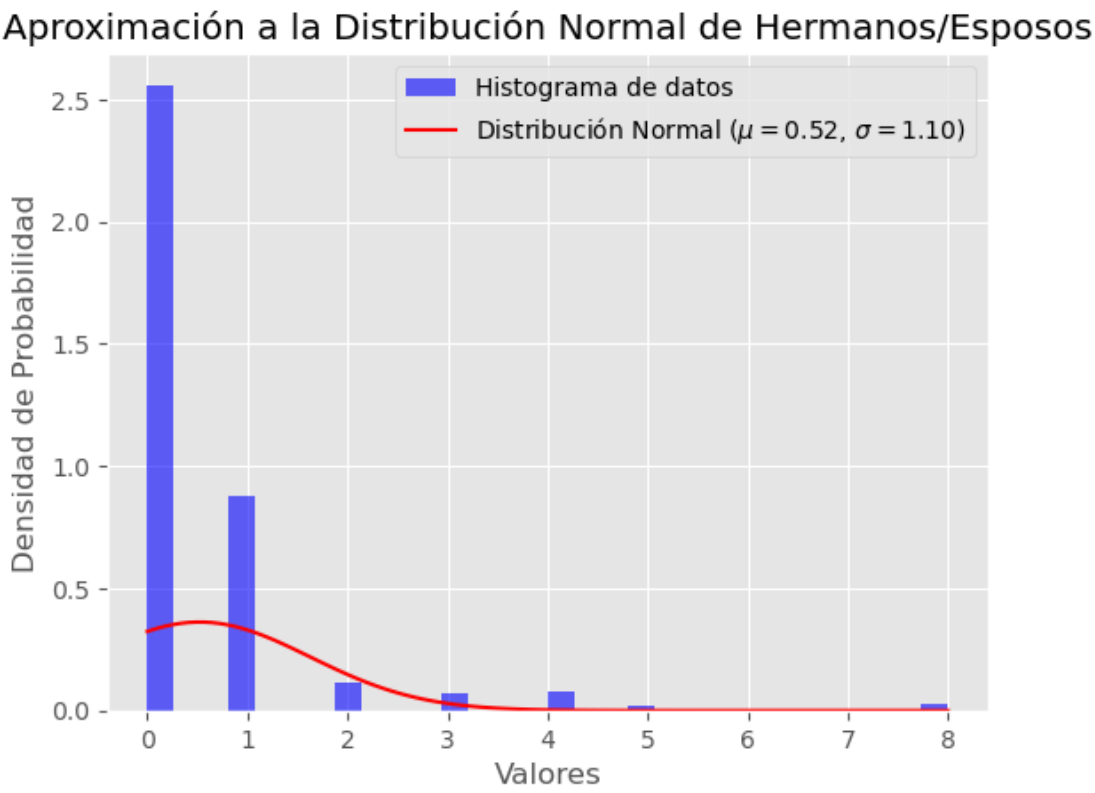
Como podemos observar, la gran mayoría de pasajeros se encontraban en **tercera clase**.

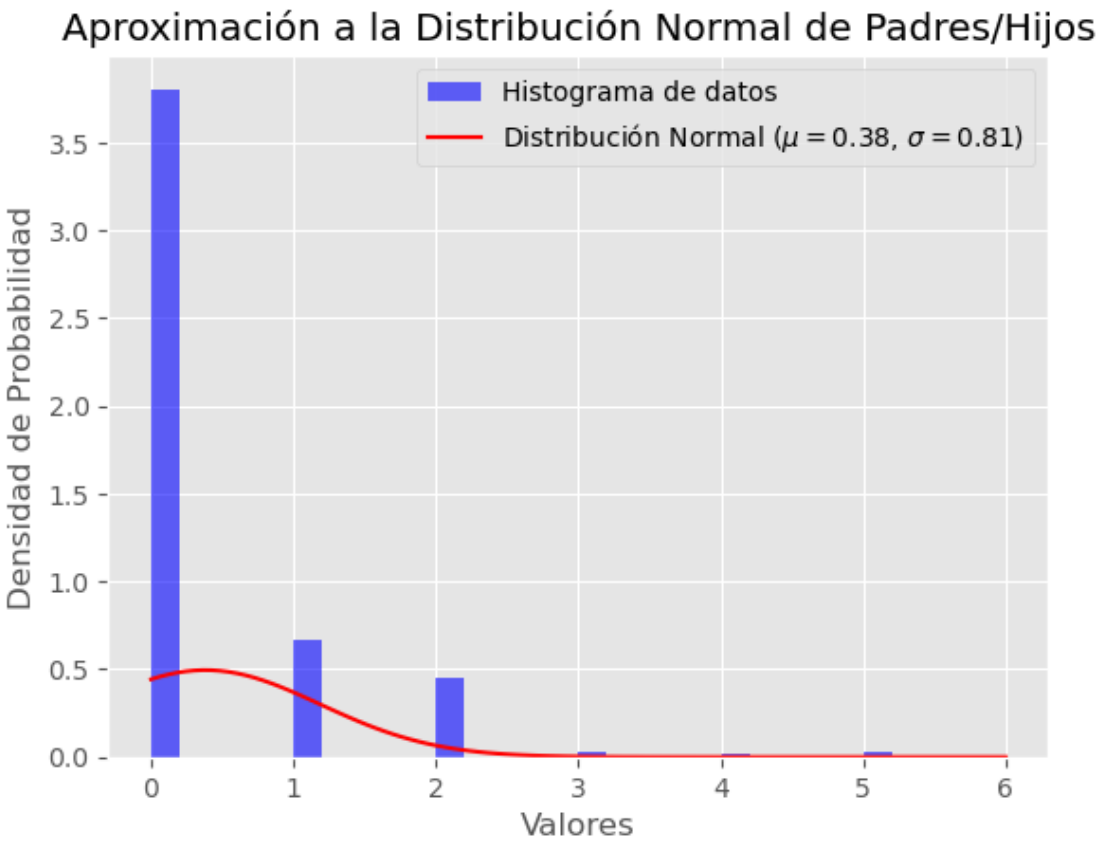
Distribución de sobrevivientes

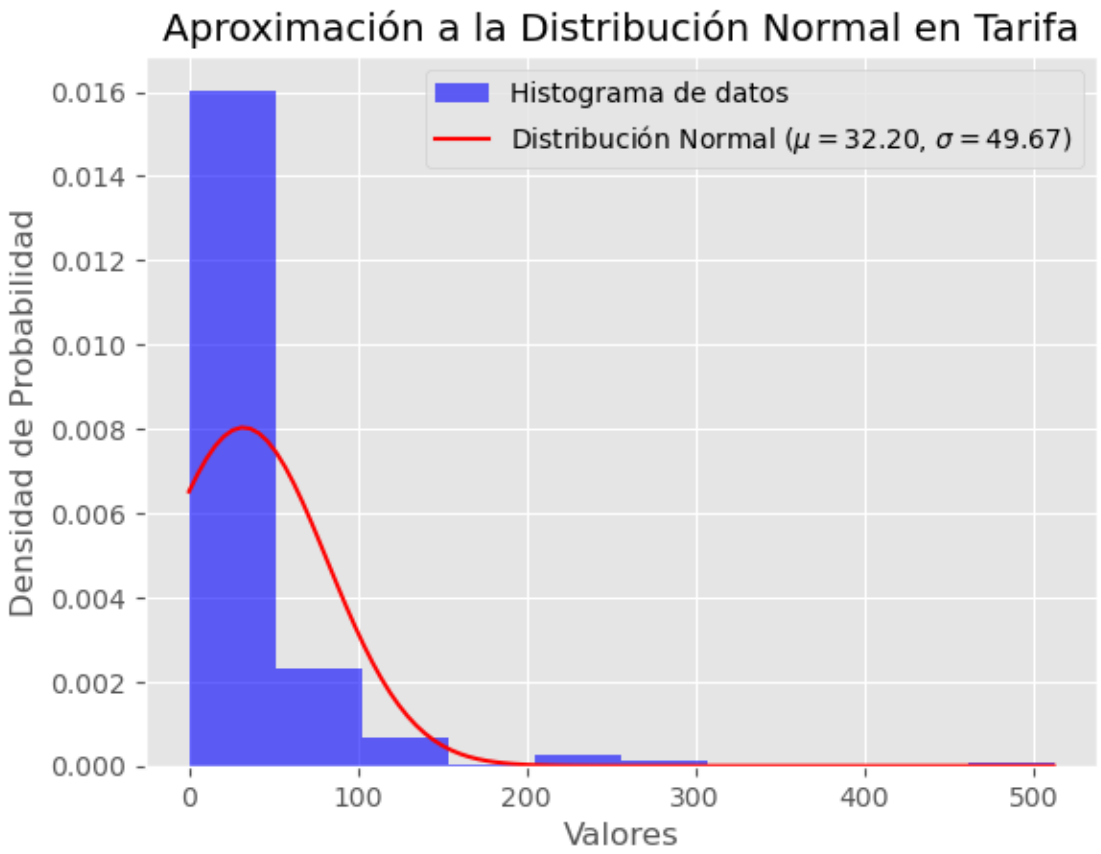

Distribución de pasajeros sobrevivientes
Sobrevivió
38.4%
61.6%
No Sobrevivió

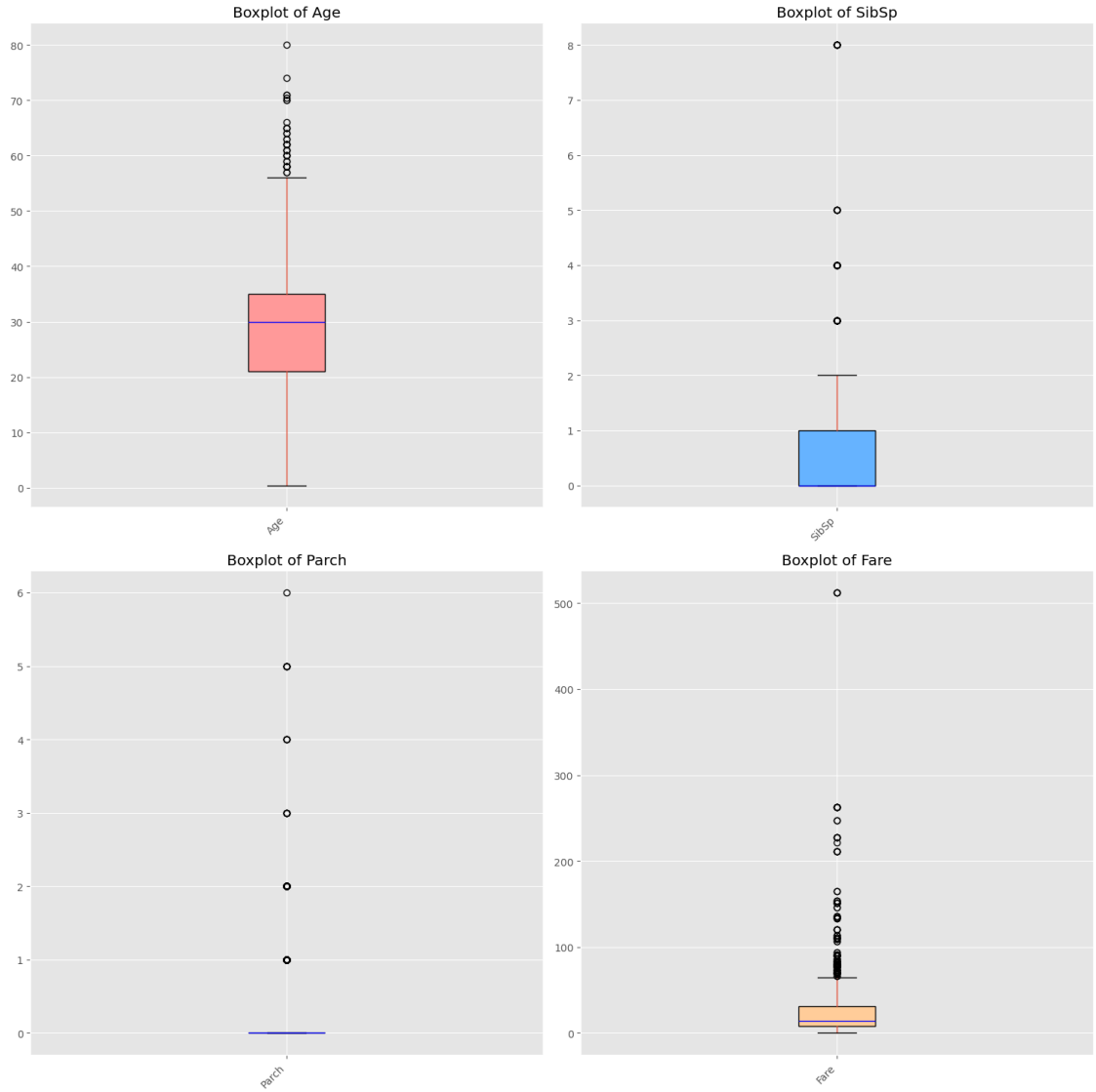

Survived
Pclass
Age
SibSp
Parch
Fare
1 -0.34 -0.079 -0.035 0.082 0.26
-0.34 1 -0.35 0.083 0.018 -0.55
-0.079 -0.35 1 -0.27 -0.19 0.098
-0.035 0.083 -0.27 1 0.41 0.16
0.082 0.018 -0.19 0.41 1 0.22
0.26 -0.55 0.098 0.16 0.22 1
Survived
Pclass
Age
SibSp
Parch
Fare


Aproximación a la Distribución Normal de Edad
Histograma de datos
Distribución Normal ($\mu = 29.39$, $\sigma = 13.25$)
Densidad de Probabilidad
0.00
0.02
0.04
0.06
0.08
0.10
0
10
20
30
40
50
60
70
80
Valores


Aproximación a la Distribución Normal de Hermanos/Esposos
Histograma de datos
Distribución Normal ($\mu = 0.52$, $\sigma = 1.10$)
Densidad de Probabilidad
Valores
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
6
7
8


Aproximación a la Distribución Normal de Padres/Hijos
Histograma de datos
Distribución Normal ($\mu = 0.38$, $\sigma = 0.81$)
Densidad de Probabilidad
Valores
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
1
2
3
4
5
6


Aproximación a la Distribución Normal en Tarifa
Histograma de datos
Distribución Normal ($\mu = 32.20$, $\sigma = 49.67$)
Densidad de Probabilidad
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
Valores



1.1 Equipo:

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

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_priori
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_priori
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 67249 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 354631 bytes to titanic_reports\TITANIC3.pdf
```

```
[NbConvertApp] Converting notebook TITANIC.ipynb to pdf
c:\ProgramData\miniconda3\share\jupyter\nbconvert\templates\latex\display_priori
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_priori
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 29066 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 314087 bytes to titanic_reports\TITANICr.pdf
```