import pandas as pd
import numpy as np
import seaborn as sns

import matplotlib.pyplot as plt

from google.colab import files
uploaded = files.upload()

Choose Files No file chosen enable.
Saving train.csv to train.csv

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to

df = pd.read\_csv("train.csv") # Replace with actual filename
df.head()

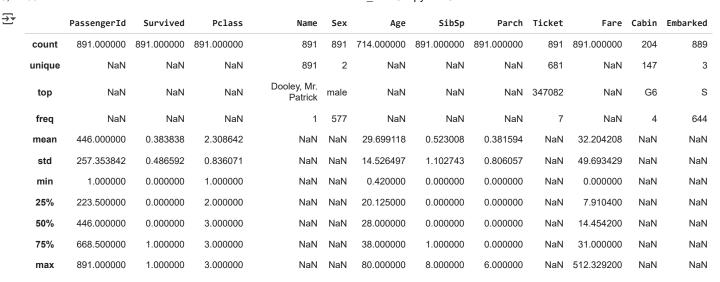
₹		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
	0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C85	С
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S

df.isnull().sum()



dtype: int64

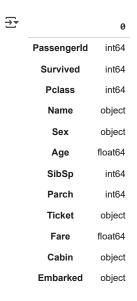
df.describe(include='all')



```
print("Shape of data:", df.shape)
print("Column Names:", df.columns.tolist())
```

```
Shape of data: (891, 12)
Column Names: ['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked']
```

df.dtypes



dtype: object

```
# Convert 'Survived', 'Pclass', 'Embarked' to category
df['Survived'] = df['Survived'].astype('category')
df['Pclass'] = df['Pclass'].astype('category')
df['Embarked'] = df['Embarked'].astype('category')
df = pd.get_dummies(df, columns=['Sex', 'Embarked'], drop_first=True)
# sex - 0 & 1 Embarked_Q, Embarked_S onehotencoding
df.drop(columns=['Cabin', 'Ticket', 'Name'], inplace=True)
df['Age'] = df['Age'].fillna(df['Age'].median())
df.info()
df.head()
```

<class 'pandas.core.frame.DataFrame'>
 RangeIndex: 891 entries, 0 to 890

Data columns (total 10 columns): # Column Non-Null Count

#	Column	Non-Null Count	Dtype
0	PassengerId	891 non-null	int64
1	Survived	891 non-null	category
2	Pclass	891 non-null	category
3	Age	891 non-null	float64
4	SibSp	891 non-null	int64
5	Parch	891 non-null	int64
6	Fare	891 non-null	float64
7	Sex_male	891 non-null	bool
8	Embarked_Q	891 non-null	bool
9	Embarked_S	891 non-null	bool
44	1/21	+(2) Cl-	-+(1/2) :-+(

dtypes: bool(3), category(2), float64(2), int64(3)

memory usage: 39.5 KB

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
0	1	0	3	22.0	1	0	7.2500	True	False	True
1	2	1	1	38.0	1	0	71.2833	False	False	False
2	3	1	3	26.0	0	0	7.9250	False	False	True
3	4	1	1	35.0	1	0	53.1000	False	False	True
4	5	0	3	35.0	0	0	8.0500	True	False	True

df

<del></del>											
<u> </u>		PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Sex_male	Embarked_Q	Embarked_S
	0	1	0	3	22.0	1	0	7.2500	True	False	True
	1	2	1	1	38.0	1	0	71.2833	False	False	False
	2	3	1	3	26.0	0	0	7.9250	False	False	True
	3	4	1	1	35.0	1	0	53.1000	False	False	True
	4	5	0	3	35.0	0	0	8.0500	True	False	True
	886	887	0	2	27.0	0	0	13.0000	True	False	True
	887	888	1	1	19.0	0	0	30.0000	False	False	True
	888	889	0	3	28.0	1	2	23.4500	False	False	True
	889	890	1	1	26.0	0	0	30.0000	True	False	False
	890	891	0	3	32.0	0	0	7.7500	True	True	False

891 rows × 10 columns

df.isnull().sum()



dtype: int64