

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [3]: train_df=pd.read_csv("C:\\Users\\DELL\\OneDrive\\Desktop\\train.csv")
test_df=pd.read_csv("C:\\Users\\DELL\\OneDrive\\Desktop\\test.csv")

In [4]: train_df.columns

Out[4]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
      Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype=object')

In [6]: train_df.head()

Out[6]:
   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      C
2             3         1       3  Heikkinen, Miss. Laina   female  26.0    0    0  STON/O2 3101282  7.9250   NaN      S
3             4         1       1  Furelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0    1    0  113803  53.1000  C123      S
4             5         0       3    Allen, Mr. William Henry   male  35.0    0    0   373450  8.0500   NaN      S

In [7]: train_df.describe()

Out[7]:
   PassengerId  Survived  Pclass    Age  SibSp  Parch    Fare
count  891.000000  891.000000  891.000000  714.000000  891.000000  891.000000  891.000000
mean    446.000000   0.383838   2.308642  29.699118   0.523008   0.381594  32.204208
std    257.353842   0.486592   0.836071  14.526497   1.102743   0.806057  49.693429
min      1.000000   0.000000   1.000000   0.420000   0.000000   0.000000   0.000000
25%    223.500000   0.000000   2.000000  20.125000   0.000000   0.000000   7.910400
50%    446.000000   0.000000   3.000000  30.000000   0.000000   0.000000  14.454200
75%    668.500000   1.000000   3.000000  38.000000   1.000000   0.000000  31.000000
max    891.000000   1.000000   3.000000  80.000000   8.000000   6.000000  512.329200

In [8]: print(train_df.isnull().sum())

PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age           177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin         687
Embarked        2
dtype: int64

In [12]: train_df.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   Sex          891 non-null     object
5   Age          714 non-null     float64
6   SibSp        891 non-null     int64
7   Parch        891 non-null     int64
8   Ticket       891 non-null     object
9   Fare         891 non-null     float64
10  Cabin        284 non-null     object
11  Embarked     889 non-null     object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB

In [15]: sns.heatmap(train_df.isnull(), cbar=False, cmap="viridis")

Out[15]:
<Axes: >



In [24]: sns.set_style("darkgrid")
sns.countplot(x="Survived", data=train_df)

Out[24]:
<Axes: xlabel='Survived', ylabel='count'>



In [25]: sns.countplot(x="Sex", data=train_df)

Out[25]:
<Axes: xlabel='Sex', ylabel='count'>



In [28]: sns.countplot(x="Survived", hue="Sex", data=train_df, palette="cubehelix")

Out[28]:
<Axes: xlabel='Survived', ylabel='count'>



In [29]: sns.countplot(x="Pclass", data=train_df)

Out[29]:
<Axes: xlabel='Pclass', ylabel='count'>



In [38]: train_df["Age"].hist(bins=80)

Out[38]:
<Axes: >



In [40]: train_df["Age"].hist(bins=40)

Out[40]:
<Axes: >



In [41]: sns.catplot(x="Survived", data=train_df, hue="SibSp", kind="count")

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)
<seaborn.axisgrid.FacetGrid at 0x230cfc3890>

Out[41]:



In [42]: sns.catplot(x="Embarked", data=train_df, hue="Pclass", kind="count")

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)
<seaborn.axisgrid.FacetGrid at 0x230cf57eb90>

Out[42]:



In [43]: train_df["Age"].mean()

Out[43]: 29.69911764795882

In [44]: train_df.groupby(by="Pclass")["Age"].mean()

Out[44]:
Pclass
1    38.233441
2    29.877630
3    25.146620
Name: Age, dtype: float64

In [45]: def m_age(c):
    Age=c[0]
    Pclass=c[1]
    if pd.isnull(Age):
        if Pclass==1:
            return 38
        elif Pclass==2:
            return 29
        else:
            return 25
    else:
        return(Age)

In [46]: train_df["Age"]=train_df[["Age","Pclass"]].apply(m_age, axis=1)

In [47]: sns.heatmap(train_df.isnull(), yticklabels=False, cbar=False, cmap="viridis")

Out[47]:
<Axes: >



In [48]: train_df.drop("Cabin",axis=1, inplace=True)

In [49]: sns.heatmap(train_df.isnull(),cbar=False, cmap="viridis")

Out[49]:
<Axes: >



In [50]: train_df=train_df.dropna()

In [56]: train_df=pd.concat([train_df.sex, embark],axis=1)

In [57]: train_df

Out[57]:
   PassengerId  Survived  Pclass  Age  SibSp  Parch  Fare  male  Q  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  True
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  25.0    1    2   23.4500  False  False  True
889           890         1       1  26.0    0    0   30.0000  True  False  True
890           891         0       3  32.0    0    0    7.7500  True  True  False

889 rows × 10 columns

In [ ]:
```