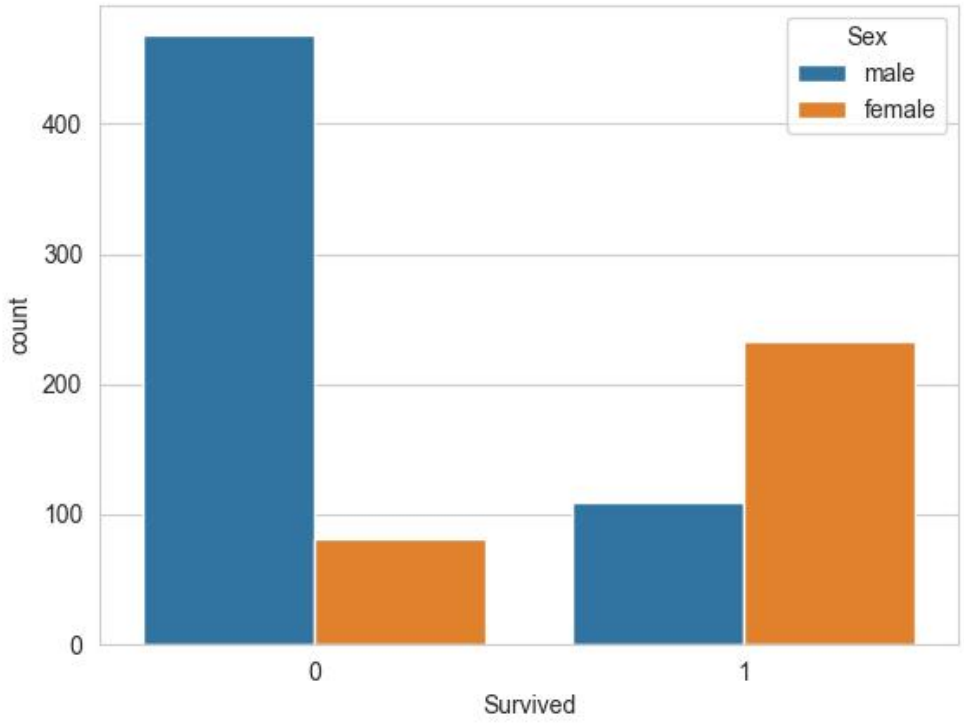
Titanic Data Analysis Report

1. Understanding Data

After importing all the necessary libraries and loading the data set the of the data in it is understood but executing codes like *data.describe() ,data.describe(include="all") , data.info() which gives the brief and the detail summary of the data values .*

2. Visualising Columns ("Survived","Sex")

Plotting a countplot() using "seaborn" package for colums "Survived", and Sex" to understand the relationship within the survived people and their gender.



FINDINGS:

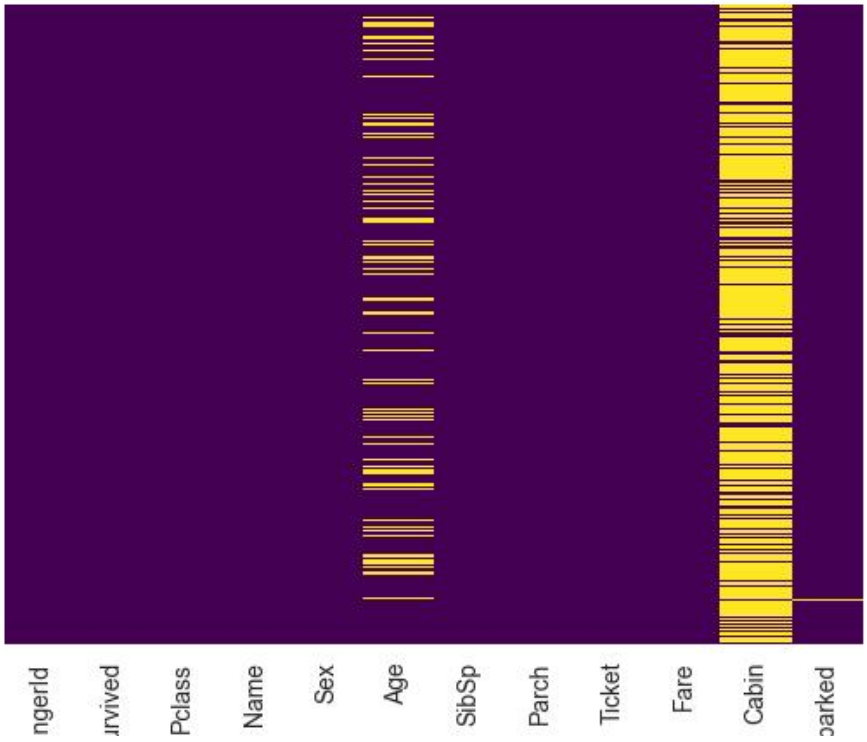
The above countplot we can observe that Male who could'nt survive is more than 450 wherein the Female is below 100.

Similarly in those who survived the count of Female is more than 200 wherein the count of Male is just crossed 100.

3. Visualizing the null values in the dataset

Checking if there is any null values in the data set is done using code *data.isnull().sum()* and is also visualised using a heatmap.

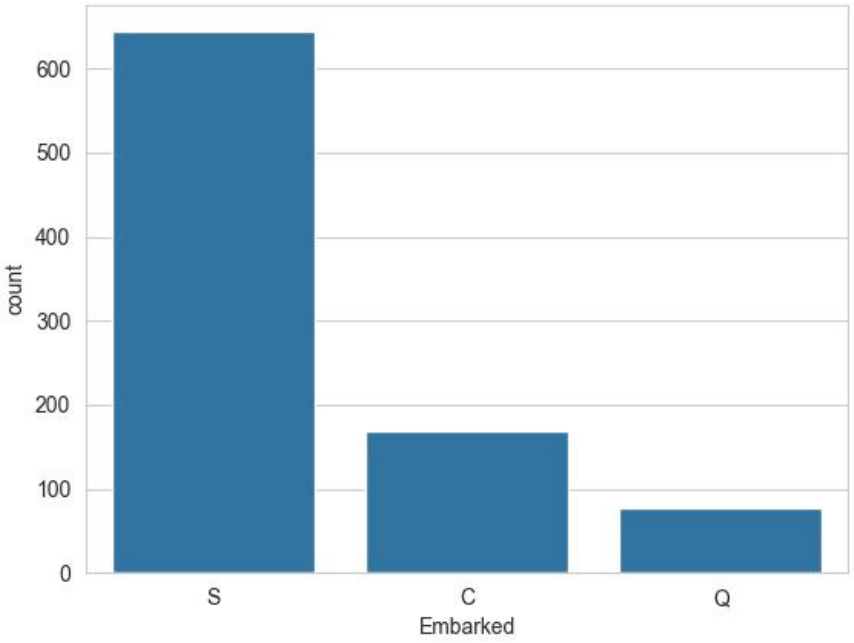
HEATMAP REPRESENTING THE NULL VALUES IN DATASET



FINDINGS :
By the above heatmap we can understand that there null values in the column of dataset : “Age”,”Cabin”,”Embarked”.

4. Visualizing the count of each category in column "Embarked"

This helps to understand count of different categories among the column Embarked so that the most reated is known, it can also be done using code data["Embarked"].value_counts()

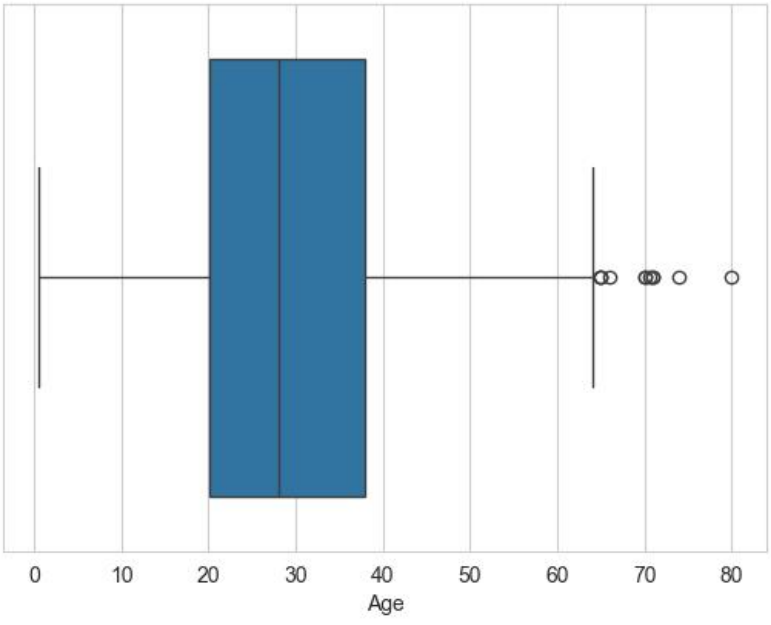


FINDINGS :

we can see that the category "S" is the dominating one .¶

5. Analysing the "Age" column to handel the missing values

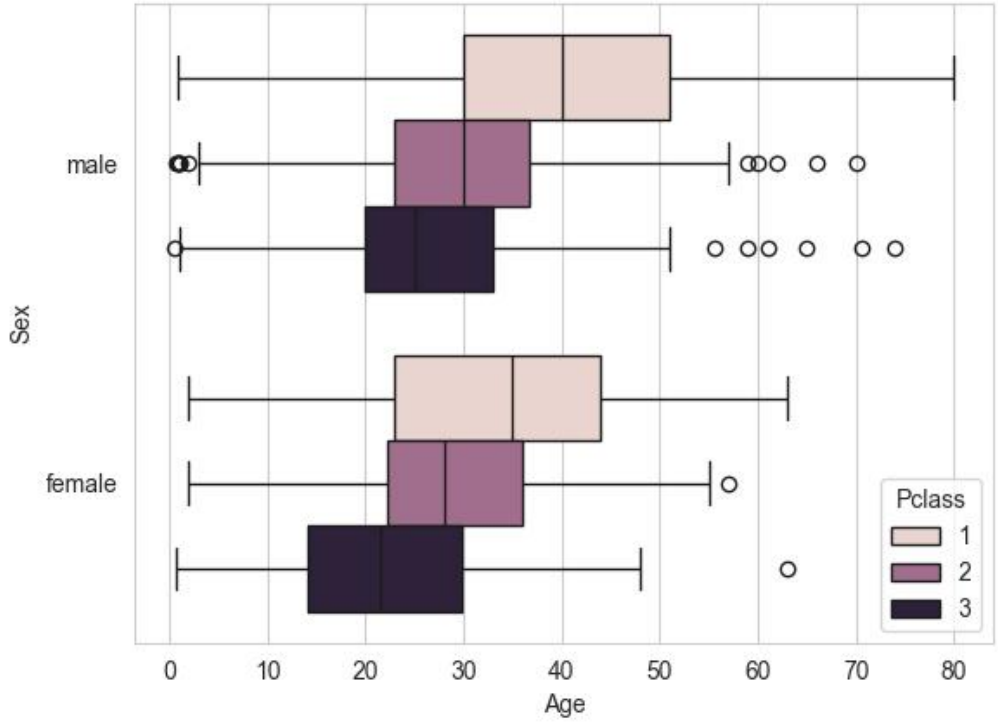
Plotting the boxplot so that we can know the median, q1,q3 of the values in the age .



FINDINGS:
So by this plot we understand that most of the passengers are in the age group of 28 .

6. Analysing the age of passengers based on sex and Pclass

This gives us the more specific information about the age distribution among passengers.And the correct value can be known by code data.groupby(["Sex","Pclass"])["Age"].median()



FINDINGS:

Sex Pclass
female 1 35.0
2 28.0
3 21.5
male 1 40.0
2 30.0
3 25.0
Name: Age, dtype: float64

so by the above outputs we can understand that most of the female of 1st class are of 32 yrs , 2nd and 3rd class are of 28 yrs and the male of 1st class are 36 , 2nd class are 29 and 3rd class are of 28 yrs.

7. HANDELLING MISSING VALUES and Droping other unwanted columns

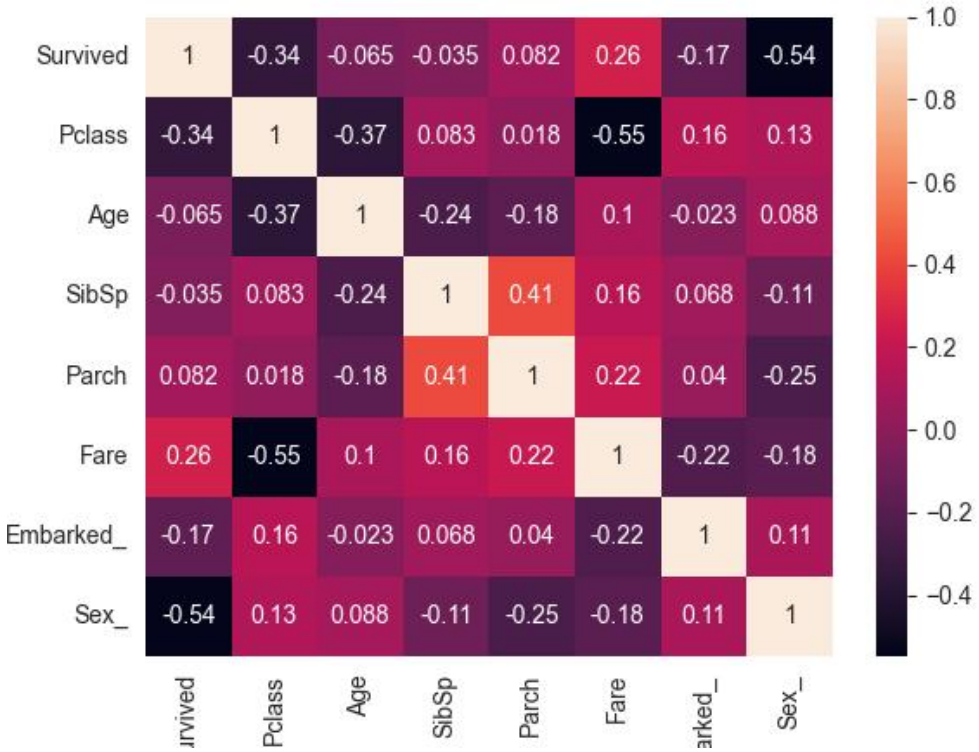
"cabin" has a lot of null values we don't require for analysis . "embarked" is a categorious variable and has only 3 unique caterories we fill the null values with the mode of col vlues = "s". "age" is a continous variable the null values can be filled with the median specificaly w.r.t their gender and pclass.Droping other unwanted columns ["PassengerId" ,"Name","Ticket"]

8. Ploting the HEATMAP of CORRELATION MATRIC for with the numerical variables of dataset

Before creating the correlation matrix we need to encode the categorical variables ic. “Embarked”,”Sex” by code

data['Embarked_'] = data['Embarked'].astype('category').cat.codes
data['Sex_'] = data['Sex'].astype('category').cat.codes

CORRELATION HEATMAP



FINDINGS :

with the above heat map we see that

*** little negative correlation among "Pclass","age","survived","fare".**

*** negative correlatin anomg "sex" ,"survived"**

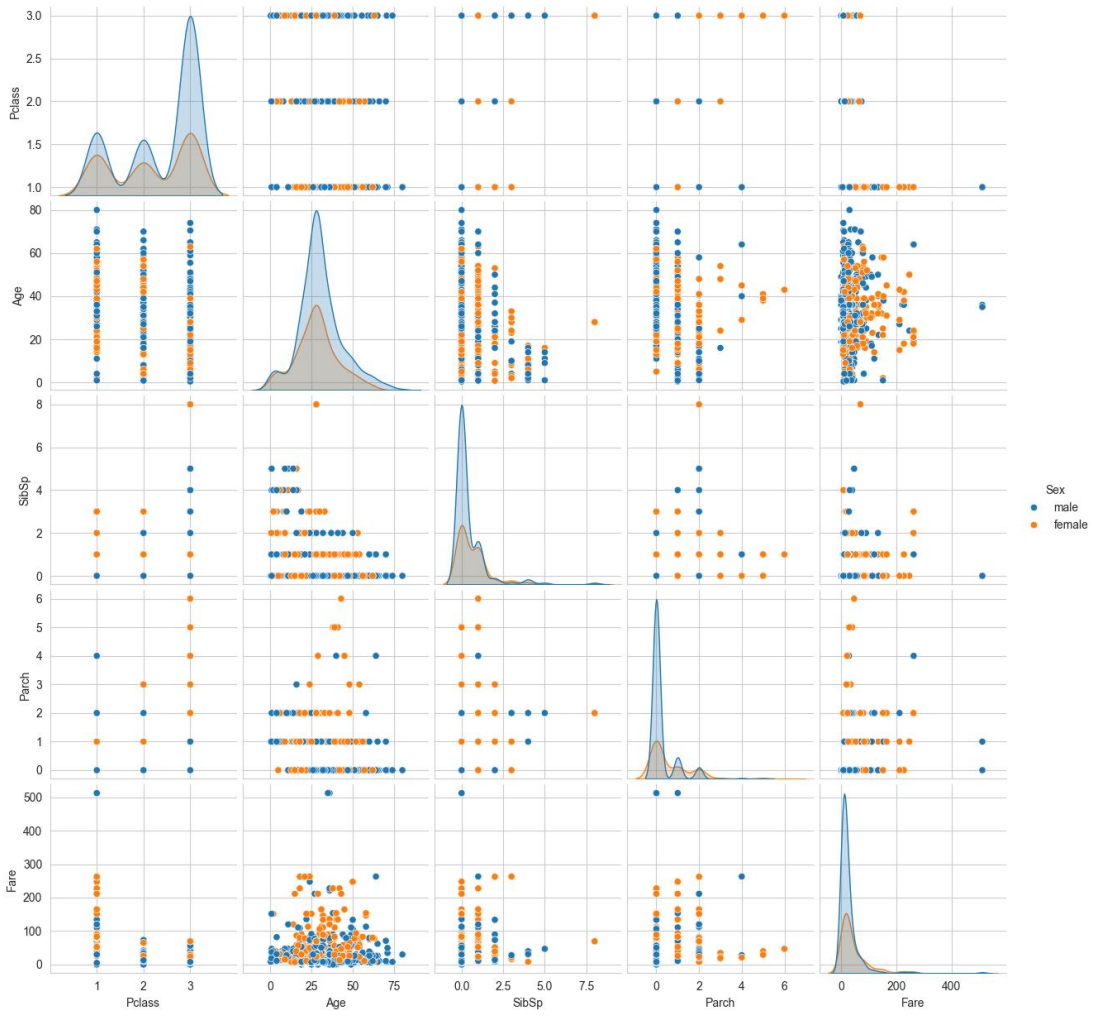
*** little positive correlation among "fare" and "survived"**

*** positive correlation among "parch" and "sibsp"**

But most of the variables are NEUTRAL.

9. PAIR PLOT

With HUE = “Sex”



FINDINGS :

*** Age : Right-skewed . Most passengers are between 20 and 40 years.**

***SibSp (Siblings/Spouses aboard): Strong positive skew. Most people traveled alone or with one family member.**

***Parch (Parents/Children aboard): Strong positive skew .Majority of passengers had no parent/child aboard.**

***Fare : Very high positive skew. Most fares are below 100, but a few outliers paid above 500. First-class passengers typically paid more.**

*** Pclass vs Age : In 1 st class females are around age 20yrs and 50yrs, in 2nd many are above 50 yrs and in 3rd class many females are below 25 yrs.**

***Pclass vs Fare : 1st class fares are higher and fare above 200 are paid only by females in 1st class. Fare in 2nd and 3rd class is almost similar.**

***Age vs Fare : Middle aged females paid more fare and male of all age have mostly paid fare >100.**

***SibSp vs Parch : Most people traveled alone or with small families**