Titanic Project

Objective

The Main objective of this project is to conduct an Exploratory Data Analysis on Titanic Dataset.

Feature Engineering

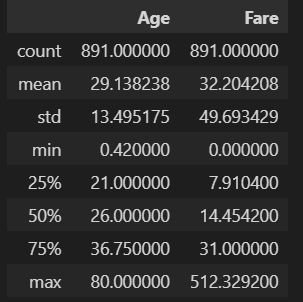
* Family Size
* Title
* Age group
* Alone
* Fare Range

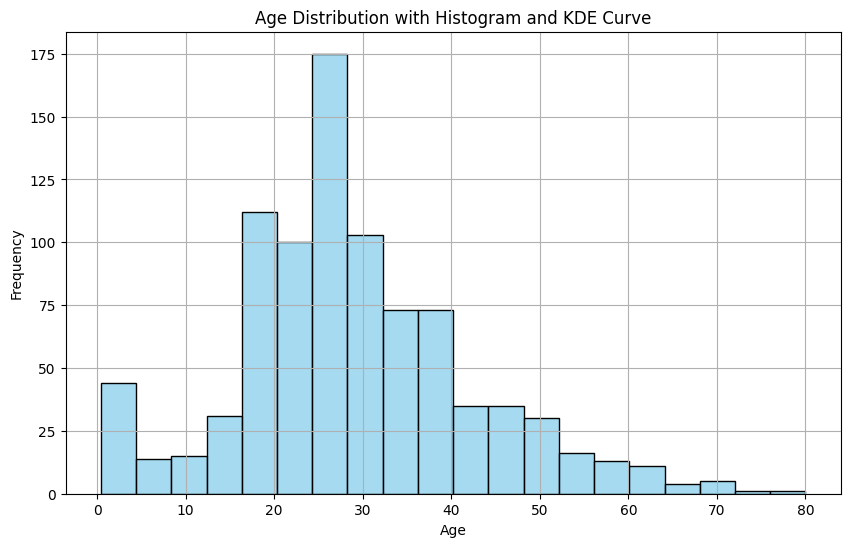
Filling Missing Values

* Mode
* Median
* Filling with Unknown (Cabin)

General/Univariate Analysis

Descriptive Statistics of Age and Fare





**Age**

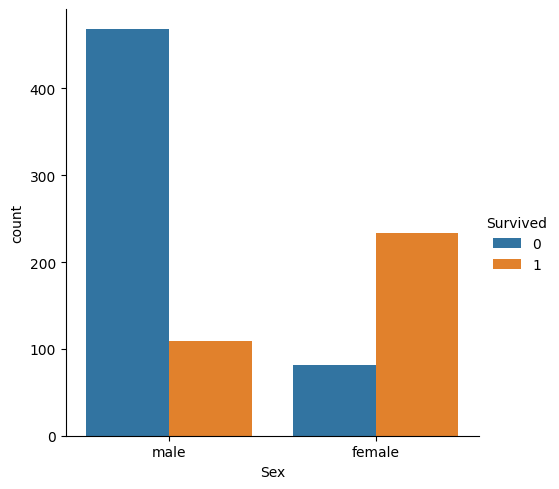
* 25 percentage of passengers are below the age of 21 years.
* 75 percentage of passengers are below the age of 36.75 years.
* Standard Deviation -13 percent denotes that the maximum count of people between 20-40 years of age.

**Fare**

* 25 percent of the fare charged were below 7 dollars.
* 75 percent of the fare charged were below 31 dollars
* As a there is a higher standard deviation, it denotes that there is a huge difference in fare charged

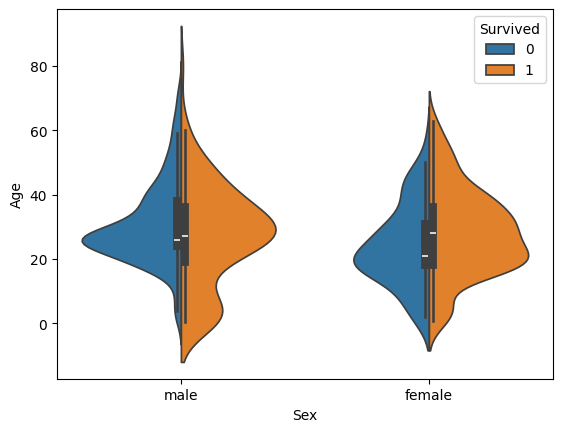
Analysis of Survival of Passengers

(i)Gender



The Above graph shows that survival chance of Female passengers are more the male Passengers.

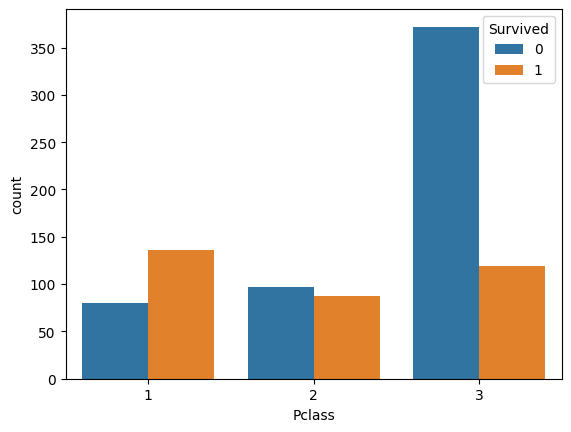
(ii)Gender and Age



**The survival rate is –**

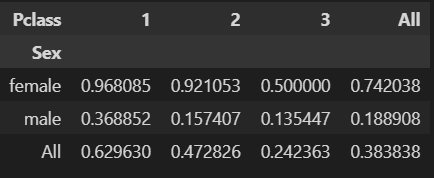
* **Good for children.**
* **High for women in the age range 20-50.**
* **Less for men as the age increases.**

(iii)Passenger Class



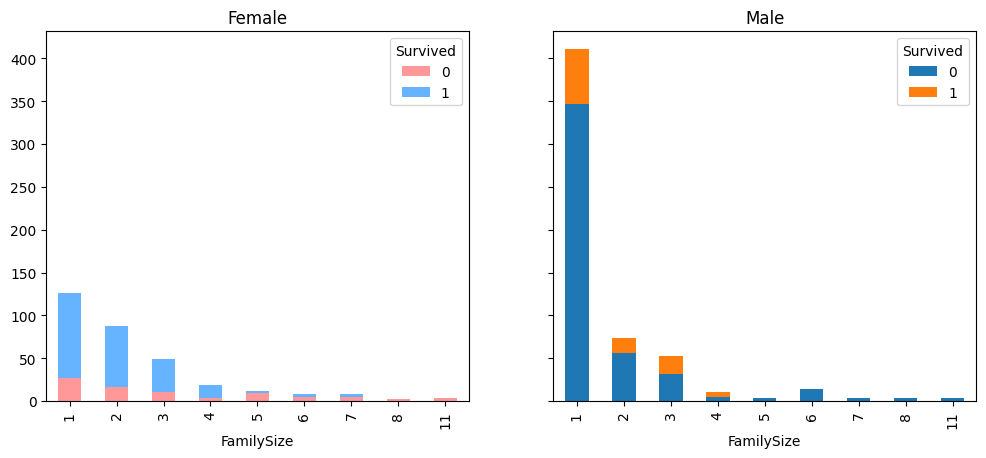
The Above graph shows that economic status plays a crucial role in survival chance of people

(iv)Passenger class and Gender



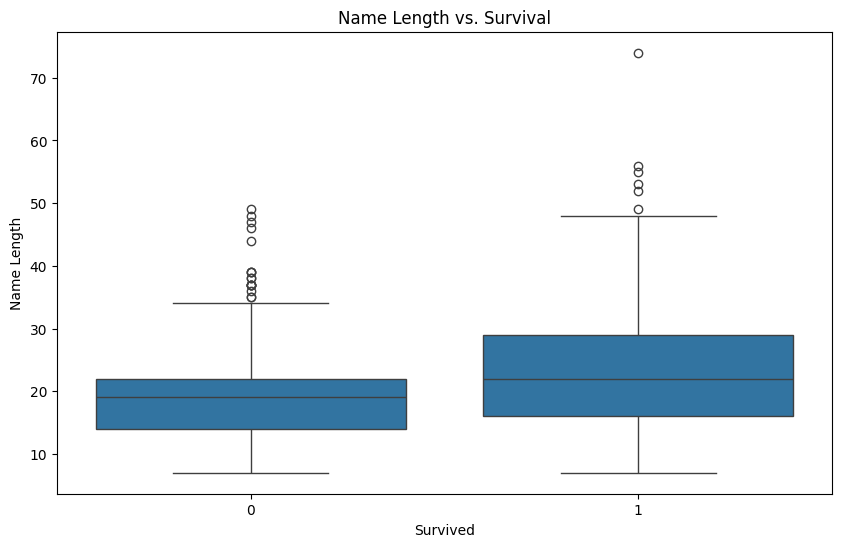
The Survival Rate is good for first Class passengers. Even among first class the survival rate is better for Female compared to Male.

**(v)Family size and Survival**



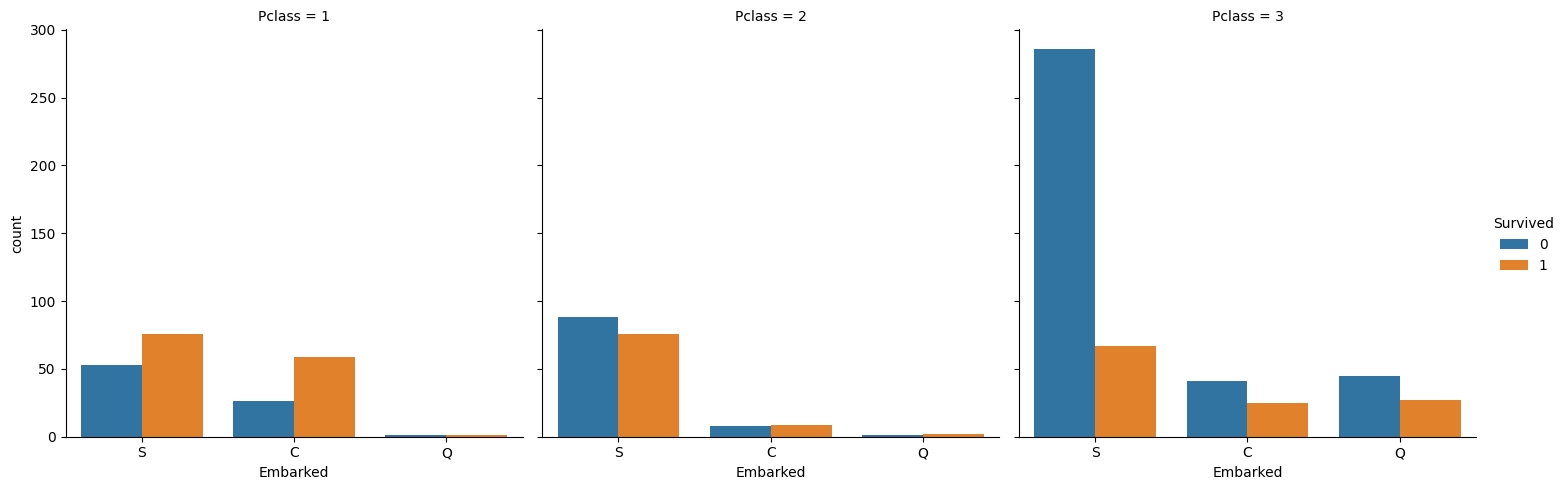
The survival Chance is good for family size of 2 and 3 for women and 1 for male.

**(vi) Name Length and Survival**

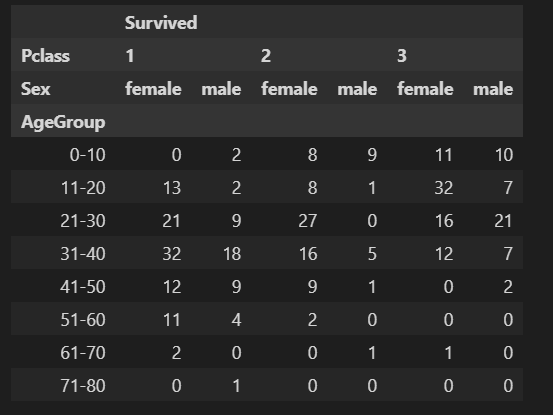


Name Length has some impact on survival Rate as for length of more than 30 has more survival chance.

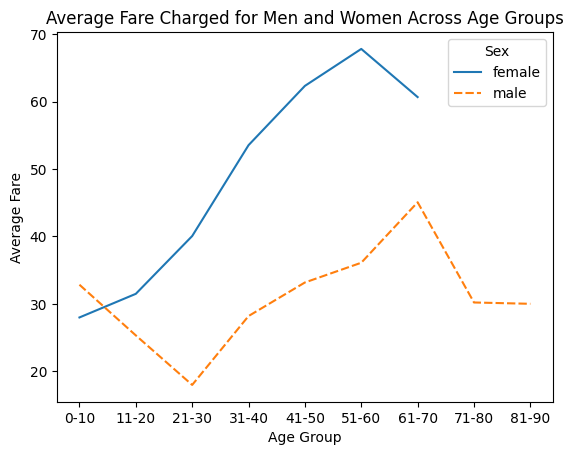
**(vii) Passenger class, port and Survival**

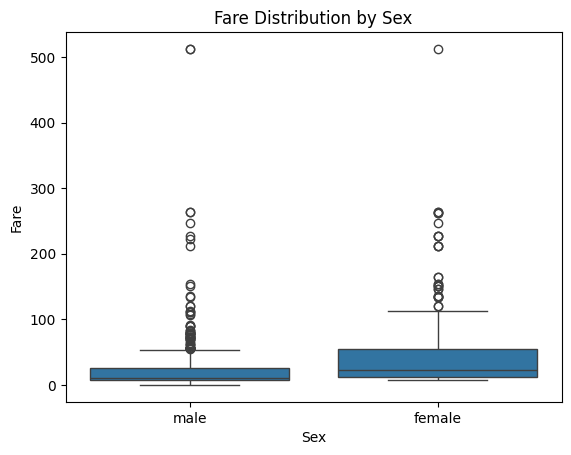


**(viii)Overall Survival Rate**



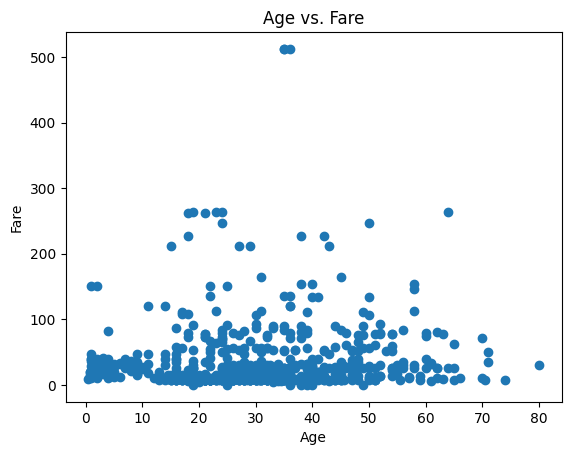
**(x)Fare and Gender**





The above graphs show that average fare charged for women is greater than men. But still both have purchased a highest paid ticket.

**(xi)Age and Fare**



As there is no definite pattern noticed in the scatter plot, we can say that there is no definite relationship between age and fare

(**xii) Title and Fare**

F-statistic: 6.567513364906533 P-value: 5.187405534782054e-06

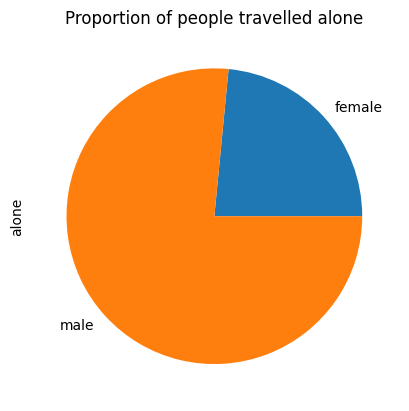
As the p value is less than 0.05, we reject the null hypothesis and conclude that there is significant difference in titles with respect to fares.

**(xiii)Gender and Average Fare**

|  |  |
| --- | --- |
| Sex | **Average Fare** |
| female | 44.47 |
| male | 25.52 |

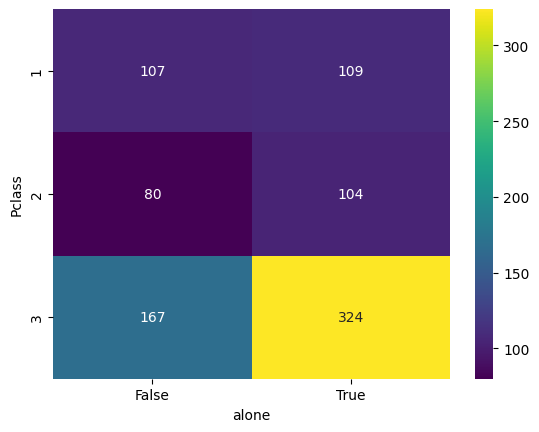
The above table shows that the average fare paid by women is greater than that of male.

**(xiv)Alone Travellers-Proportion**



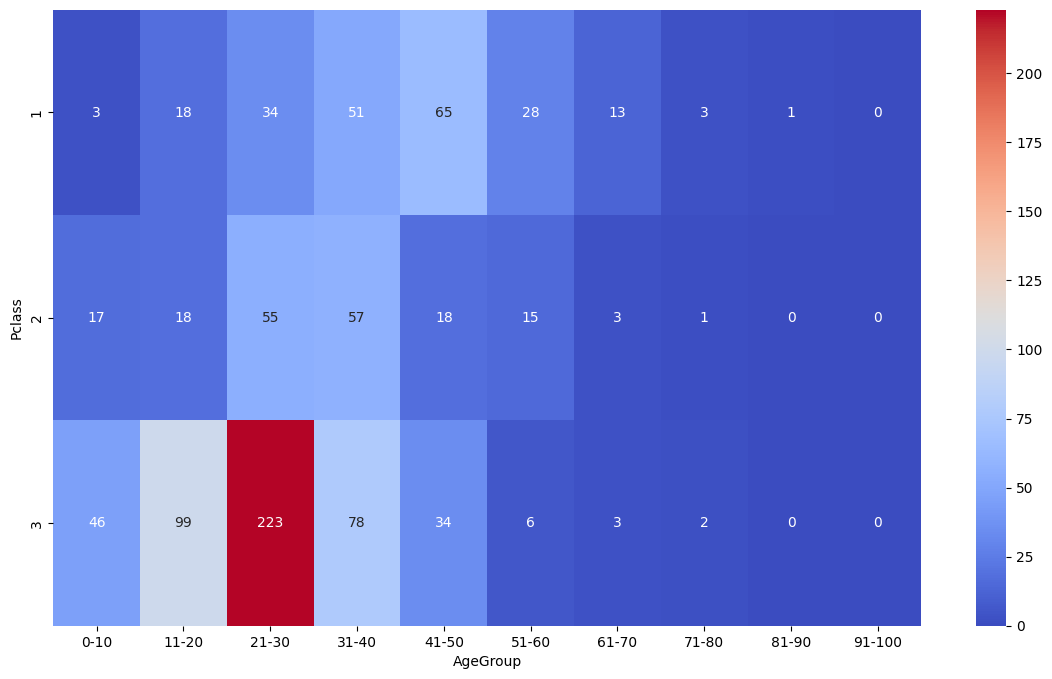
The proportion of Male passengers who travelled alone is greater than female passengers.

**(xv)Distribution of Passengers who travelled alone along the various class**



Mostly they preferred to travel in third class compared to other classes.

**(xvi)Preference of passengers for various passenger classes**



**Other findings**

* The people who travelled with zero fare charged are all Men and mostly belong to Title of Mr. and age group 20-40.
* Family Size of More than 4 mostly preferred Third class.