

Statistical Report for Titanic Dataset

I. Overview of Dataset

Objective: To identify major factors affecting the survival of passengers in Titanic

Sample size: 891

Features Used:

Categorical	Numerical
Sex: Gender [male or female]	Survived: 1=Yes, 0=No
Ticket: Ticket number	Pclass: Ticket Class [1 = upper, 2 = middle, 3 = lower]
Embarked: Port embarked on [C = Cherbourg, Q = Queenstown, S = Southampton]	Age: in years [in decimal when <1]
	SibSp: number of siblings or spouse on board
	Parch: number of parents or children on board
	Fare: Passenger's fare

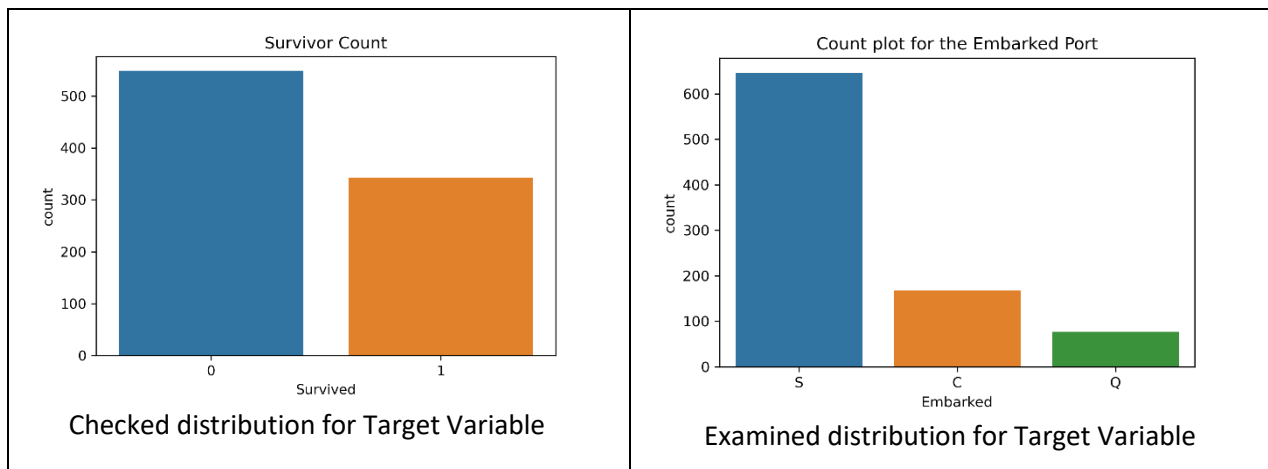
Target Feature: Survived (Yes or No | 1 or 0) *the target variable is moderately right skewed.*

Inspected the dataset, and checked dimensions.

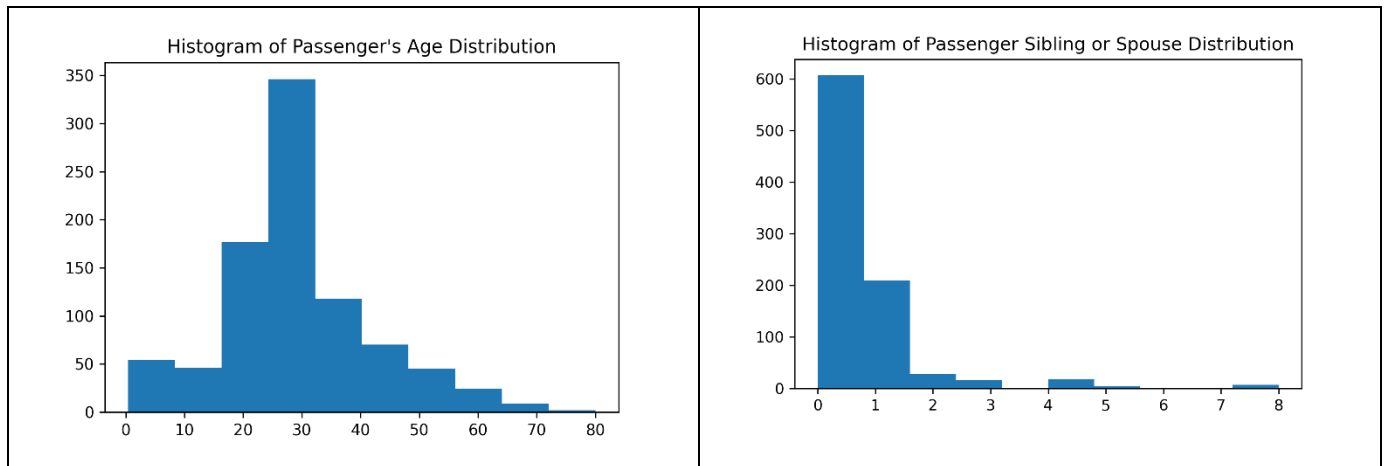
- II. **Data Preparation:** checked for missing values (identified: Age, Cabin and Embarked). Imputed with median and mode for Age and Embarked respectively. Dropped Cabin since it had majority of values missing. Also dropped columns that would not have any impact on survival (Name, PassengerID, Ticket Number). Checked for outliers as well.

III. Univariate and Bivariate Analysis

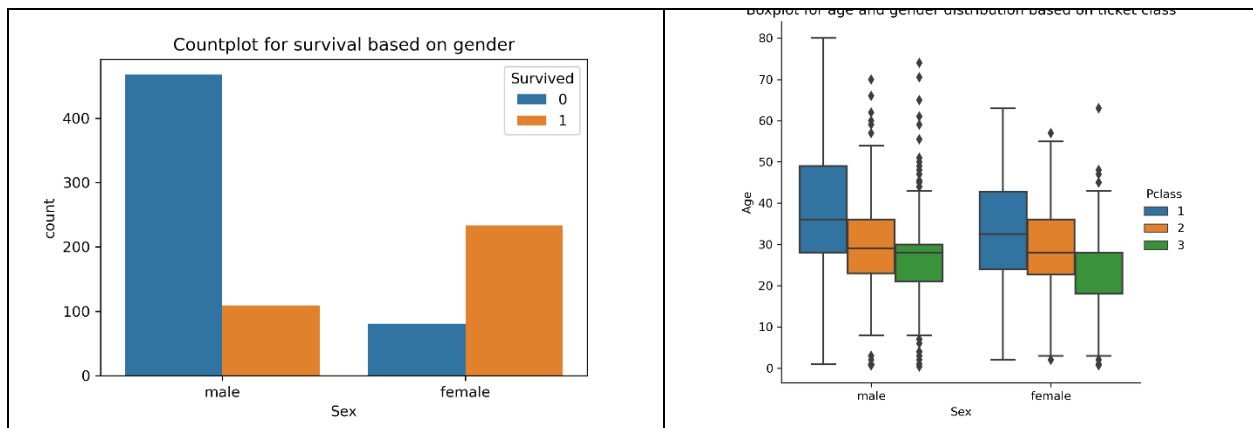
Univariate for categorical variables. Majority of passengers did not survive. Maximum embarked from Southampton and least from Queenstown.



Univariate for numerical variables. Both right skewed.



Bivariate and Multivariate Analysis

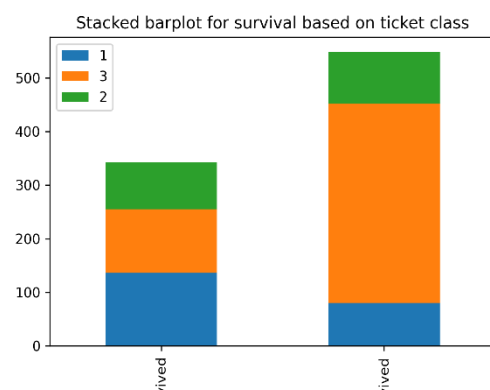


Can infer that most of the older people were traveling in first class. It may be because they were rich. The youngsters who were aged between 25 and 35 were mostly traveling in second and third classes.

IV. Hypothesis Testing

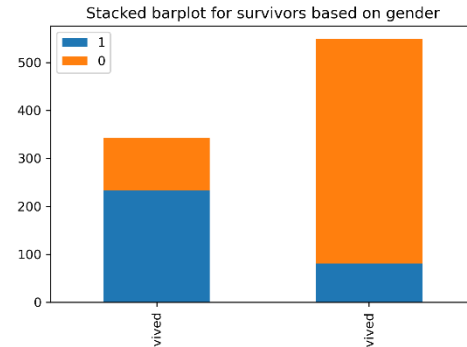
Determine if the survival rate is associated to the class of passenger.

- the survival rate decreases as passenger class decreases.
- From the analysis, it was found that the survival rate is associated with the class of the passenger.



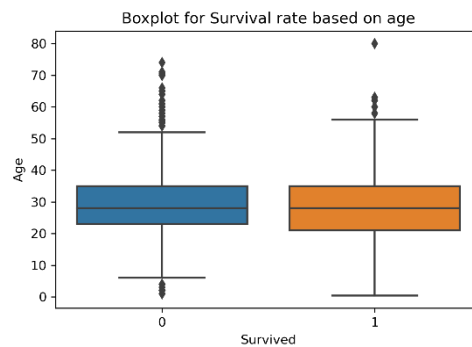
Determine if the survival rate is associated with the gender.

- Survival rate is associated with gender.
- The resulting chart shows that the survival rate for females is much higher than for males.



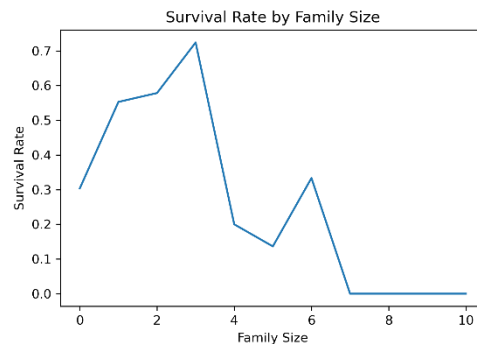
Determine the survival rate is associated with age.

- Passengers who were younger had a higher survival rate compared to those who were older.
- In general, children had a higher survival rate than adults.



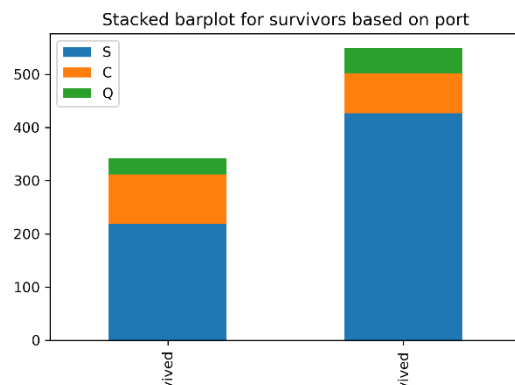
Is Survival rate associated with the number of family members aboard?

- Survival rate is associated with the number of family members aboard.
- Passengers with 1-3 family members aboard had a higher survival rate than those with more than 3 family members aboard.
- However, no significant correlation between the number of family members aboard and survival rate.

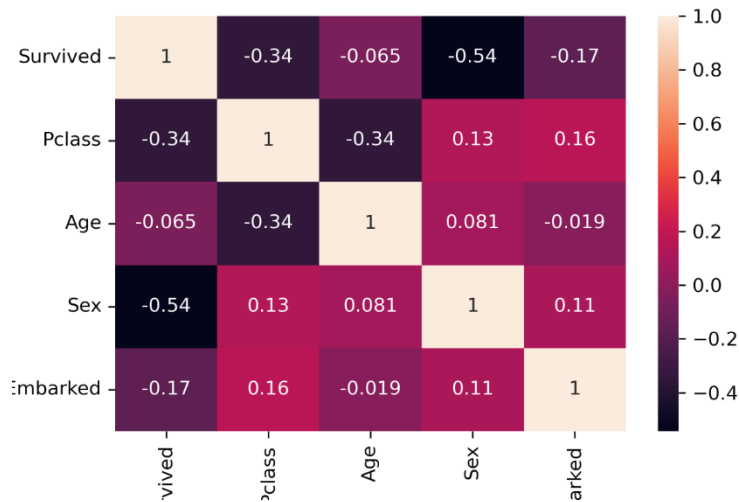


Determine if the survival rate is associated with the port embarked by the passenger

- Survival rate was weakly associated with the port embarked by the passenger.
- Passengers who embarked at Cherbourg had a slightly higher survival rate compared to those who embarked at Queenstown and Southampton.
- no significant correlation between the port embarked and survival rate.



V. Conclusion



Overall, survival on Titanic was affected by several factors. The passenger's class and gender are the strongest predictors of survival rate on the Titanic. Women had a higher chance of surviving than men. Passengers with higher-class tickets were more likely to survive than those who were in third class. Age also played a factor in determining survival as younger passengers were more likely to survive.

Lastly, the survival rate of passengers who boarded at Cherbourg was slightly higher than that of those who boarded at Queenstown and Southampton. However, no substantial association found between the port of embarkation and survival rate. Traveling with a small group of one to three people also improved the odds of survival.