**Summary of The EDA on Titian Dataset**

**Summary Statistics**

We began our analysis by computing summary statistics for the dataset. The dataset contains 891 rows and 12 columns. There are missing values in Age, Cabin and Embarked columns.

**Methods**

All the analysis are performed using side-by-side bar charts and chi-square tests.

**Hypothesis and Results**

1. Determine if the survival status is associated to the class of passenger

H0: Survived and Sex are independent among all subjects in the population.  
Ha: Survived and Sex are not independent among all subject in the population.

|  |  |  |
| --- | --- | --- |
| **Table  Description automatically generated** | **Chart, bar chart  Description automatically generated** | **Graphical user interface, text, application  Description automatically generated** |

Since the p\_value is less than 0.05,  
we will reject the null hypothesis and conclude that there are significate association between 'Survived' and 'Sex'.

1. Determine if the survival status is associated to gender

H0: Survived and Sex are independent among all subjects in the population.  
Ha: Survived and Sex are not independent among all subject in the population.

|  |  |  |
| --- | --- | --- |
| **A picture containing table  Description automatically generated** | **Chart, bar chart  Description automatically generated** | **A picture containing text  Description automatically generated** |

Since the p\_value is less than 0.05,  
we will reject the null hypothesis and conclude that there are significate association between 'Survived' and 'Sex'.

1. Determine the survival status is associated to age

H0: Survived and Age\_Group are independent among all subjects in the population.  
Ha: Survived and Age\_Group are not independent among all subject in the population.

|  |  |  |
| --- | --- | --- |
| **Table  Description automatically generated** | **Chart, bar chart  Description automatically generated** | **Table  Description automatically generated with low confidence** |

Since the p\_value is less than 0.05,  
we will reject the null hypothesis and conclude that there are significate association between 'Survived' and 'Age\_Group'.