**Assignment 2**

**Smart bridge externship**

**Applied data science**

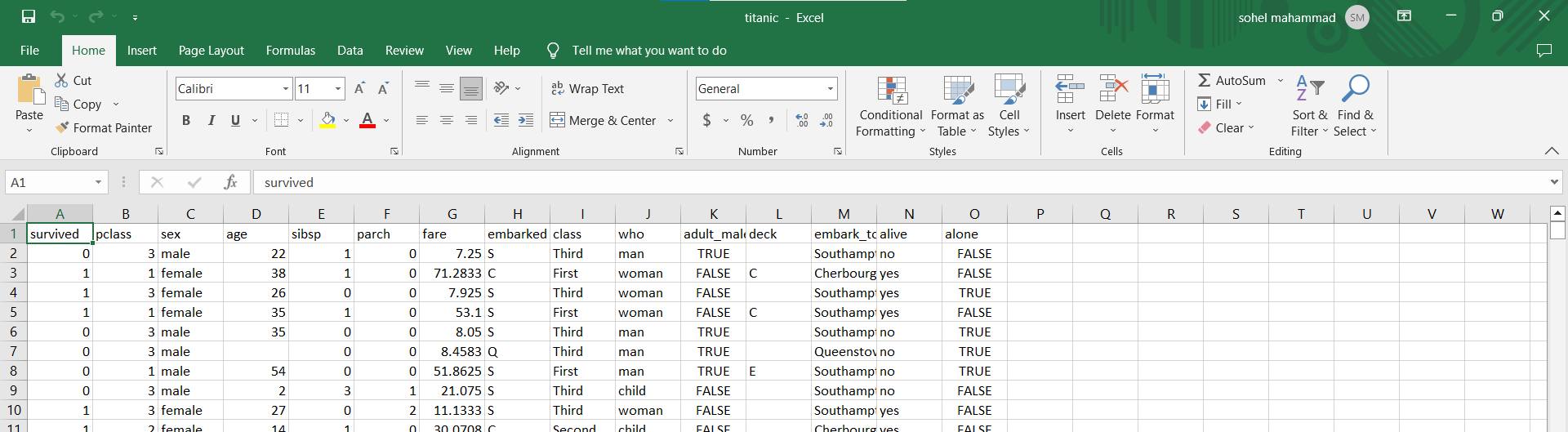
**Submitted by**

**Ch.Chandra Shekhar**

**Vitap**

**20BCB7034**

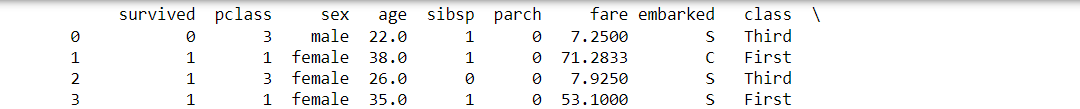
1. Download the dataset: Dataset

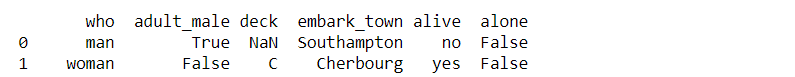


2Load the dataset









1. Perform Below Visualizations. ● Univariate Analysis ● Bi - Variate Analysis ● Multi - Variate Analysis
2. Univariate Analysis: Univariate analysis involves examining individual variables in isolation to understand their distribution, central tendency, and variability. Here are some common visualizations for univariate analysis:

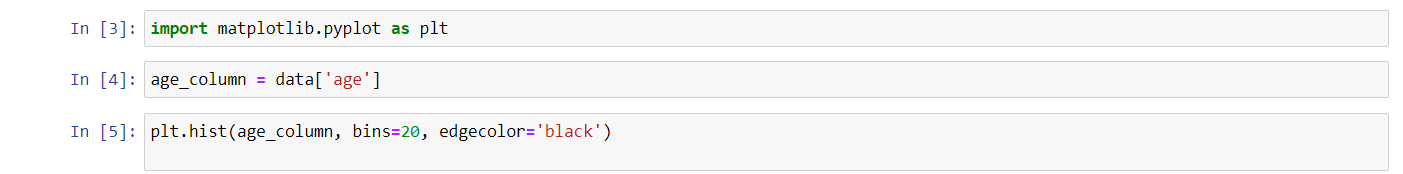
* Histogram: Displays the distribution of a continuous variable by dividing it into bins and showing the frequency or count in each bin.
* Bar Chart: Represents the distribution of a categorical variable using rectangular bars, where the height of each bar corresponds to the frequency or count.
* Box Plot: Illustrates the summary statistics of a numerical variable, such as the median, quartiles, and outliers.
* Kernel Density Plot: Shows the estimated probability density function of a continuous variable.

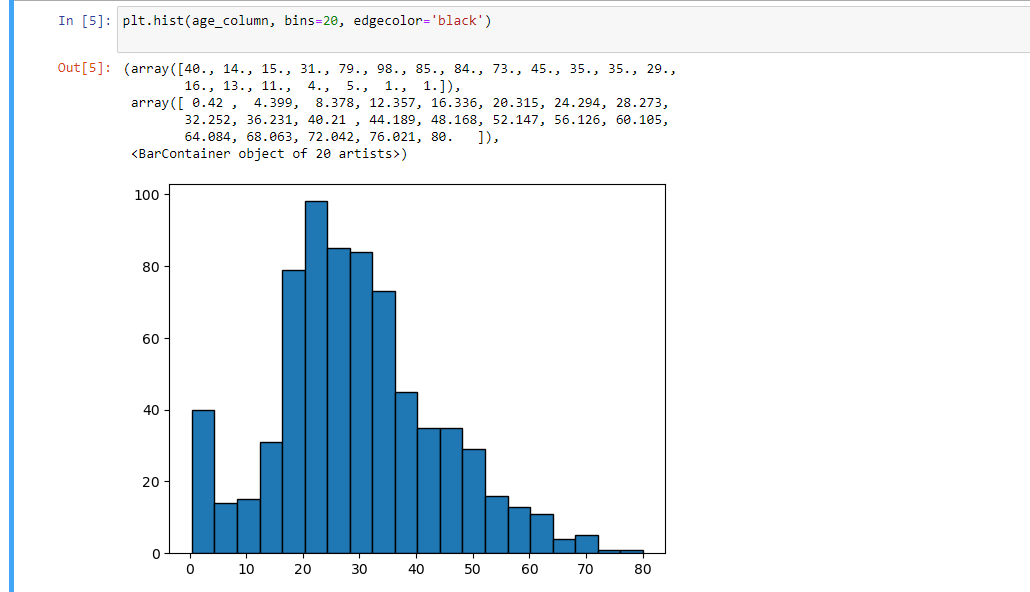
1. Bivariate Analysis: Bivariate analysis involves exploring the relationship between two variables. It helps to understand the correlation, association, or dependency between the variables. Here are some common visualizations for bivariate analysis:

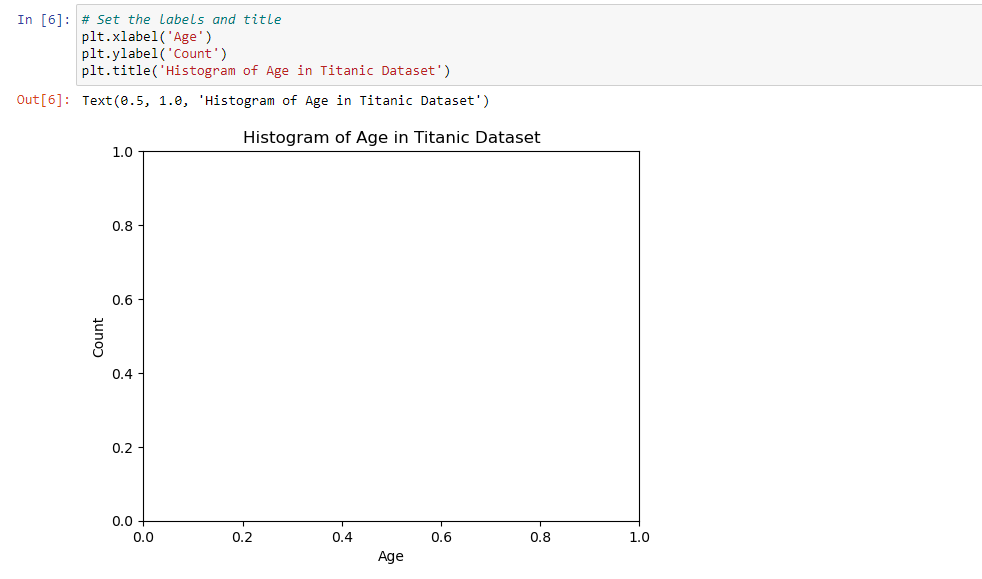
* Scatter Plot: Displays the relationship between two continuous variables by plotting each data point on a two-dimensional plane.
* Line Chart: Shows the relationship between two continuous variables by connecting data points with lines.
* Bar Chart or Grouped Bar Chart: Compares the distribution of a categorical variable across different levels of another categorical variable.
* Heatmap: Represents the correlation or association between two numerical variables using a color-coded grid.

1. Multivariate Analysis: Multivariate analysis involves examining relationships between three or more variables. It helps to understand complex patterns, interactions, and dependencies between multiple variables. Here are some common visualizations for multivariate analysis:

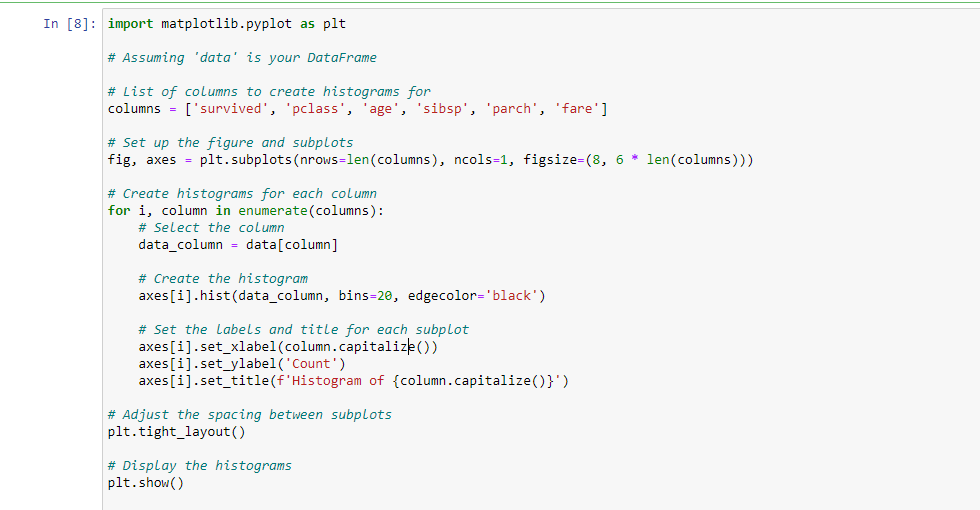
* Scatter Plot Matrix: Displays pairwise scatter plots for multiple variables to visualize their relationships simultaneously.
* Parallel Coordinates Plot: Represents multiple variables as vertical axes and plots lines that connect data points based on their values on each variable, providing insights into patterns and clusters.
* 3D Scatter Plot: Extends the scatter plot to three dimensions, allowing the visualization of relationships between three continuous variables.
* Treemap: Hierarchically displays multiple categorical variables using nested rectangles, with the area of each rectangle representing a variable's proportion.

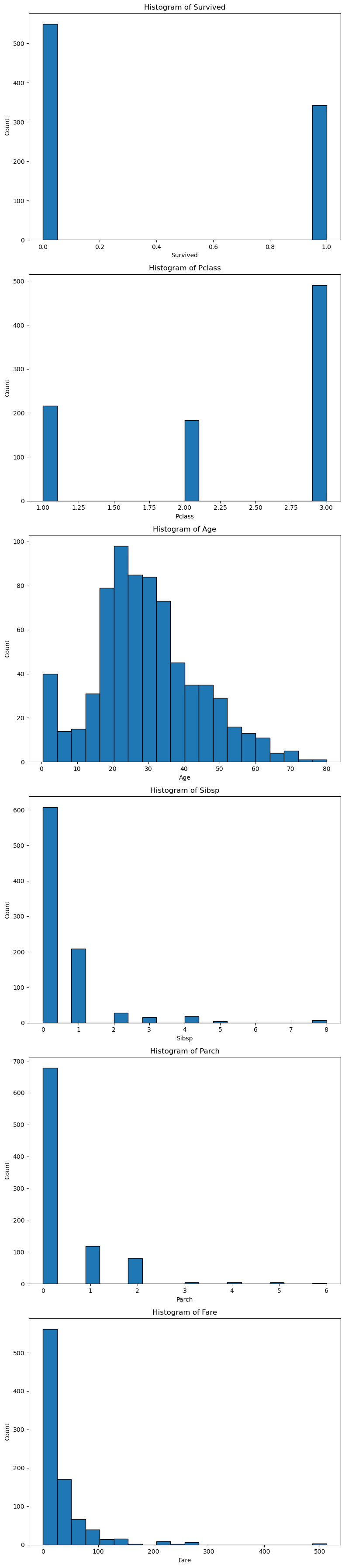




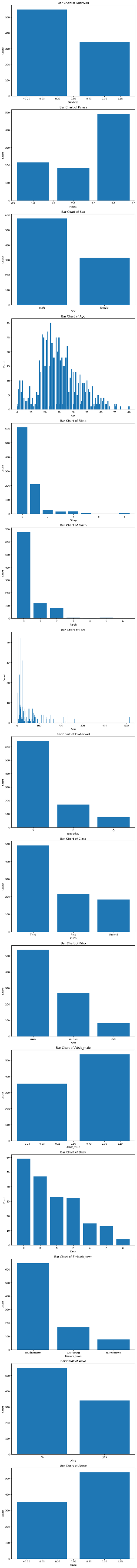


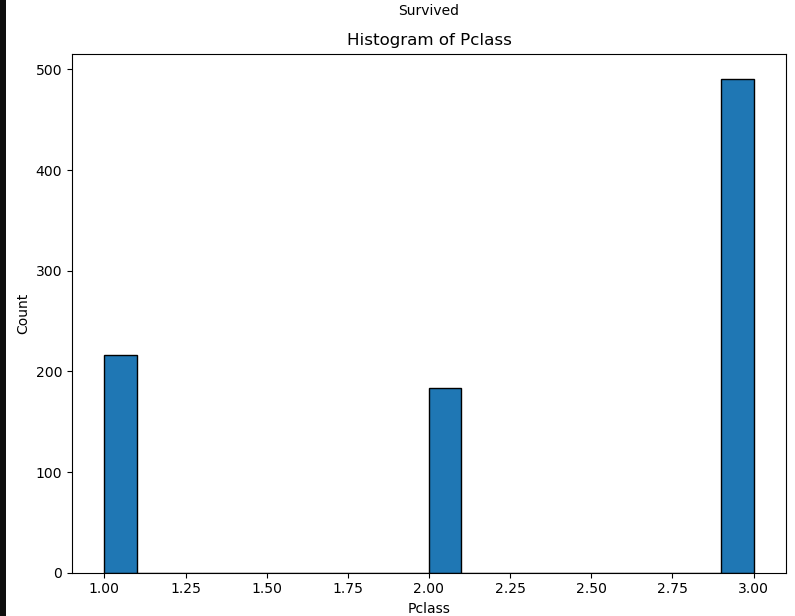
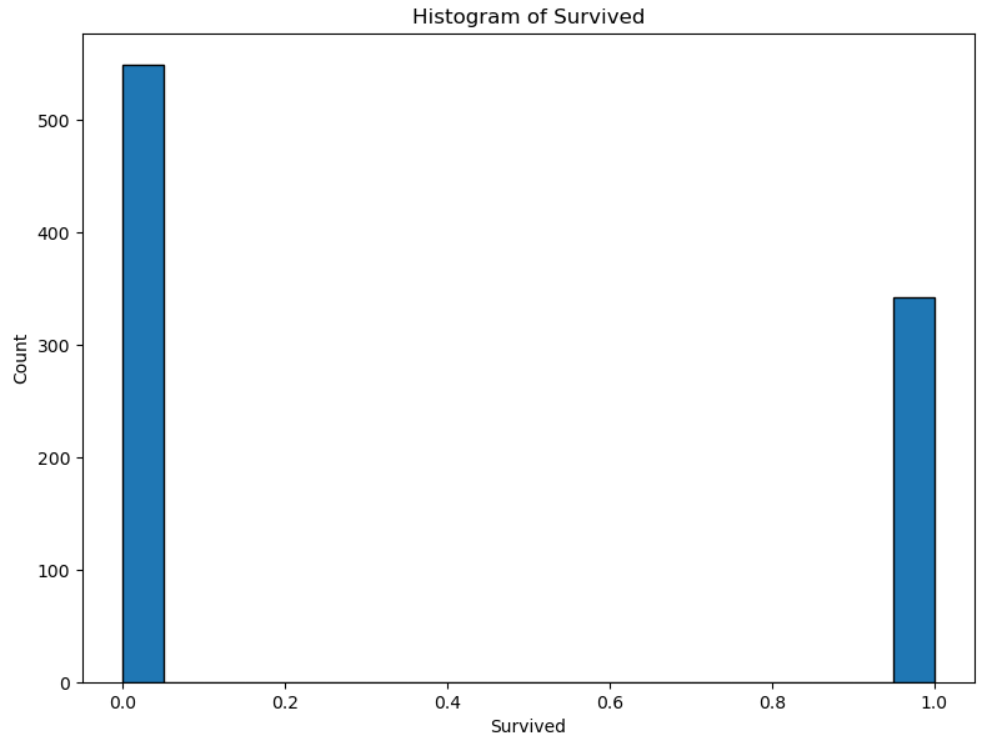


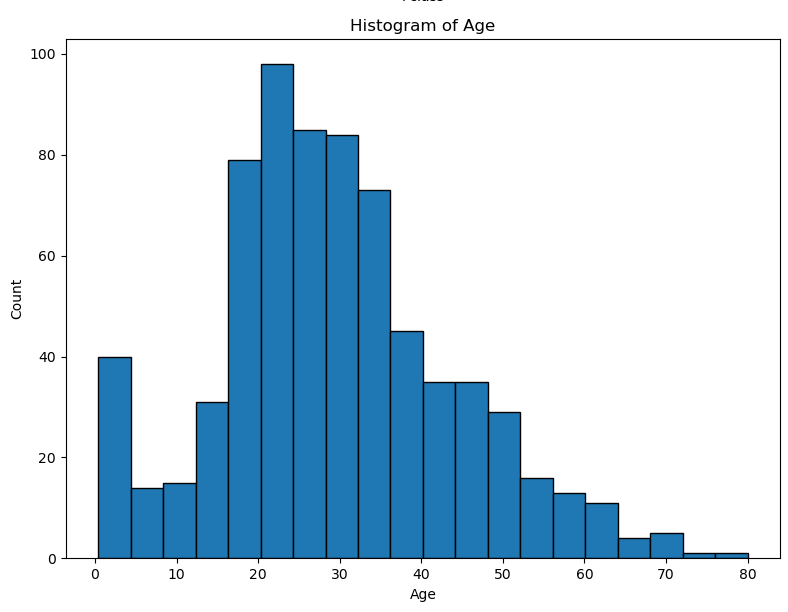


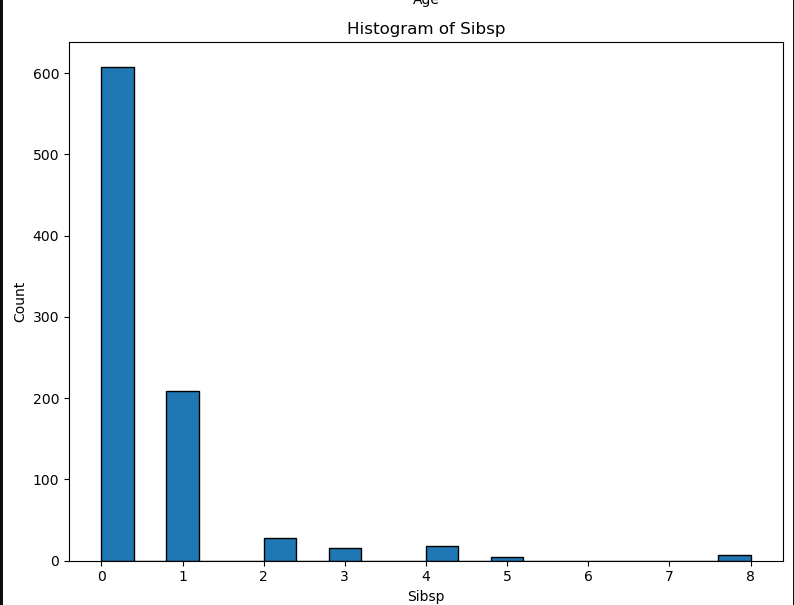


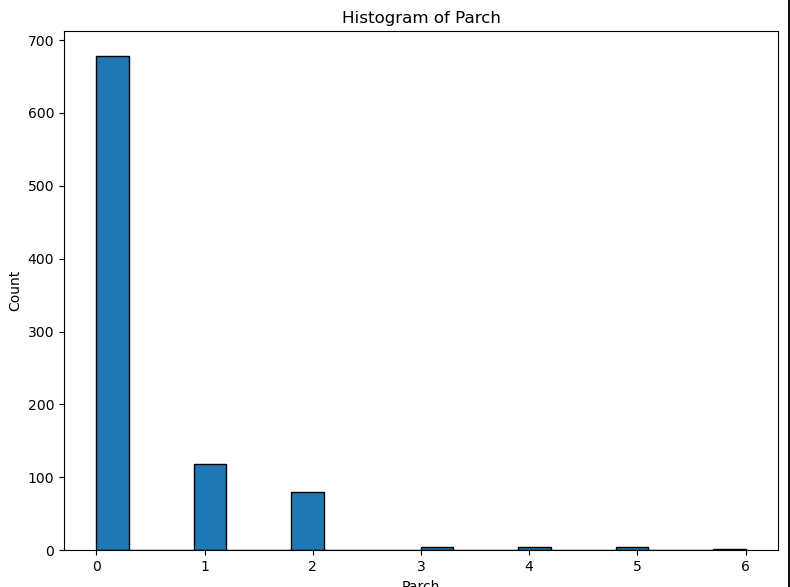


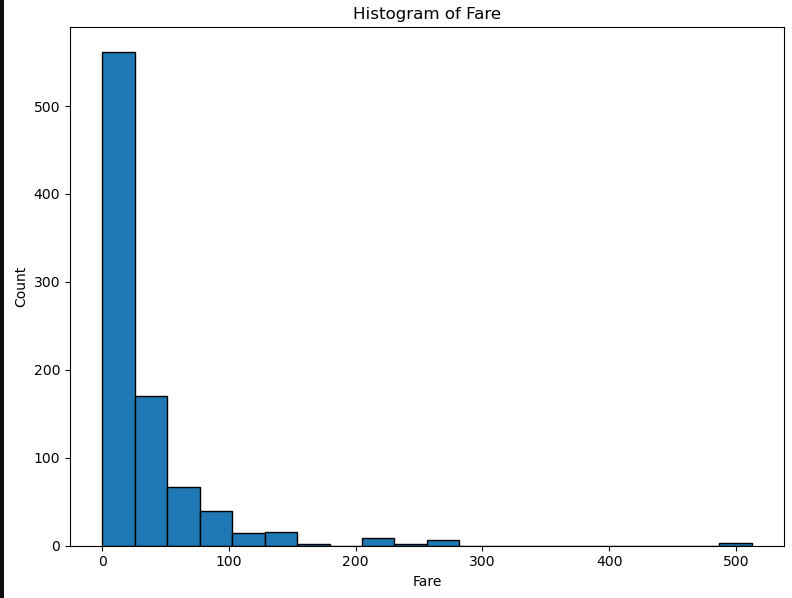


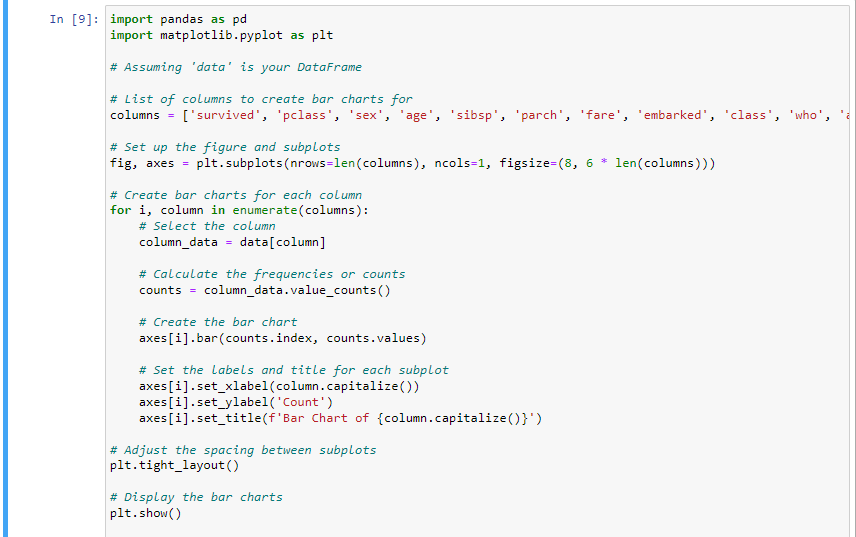


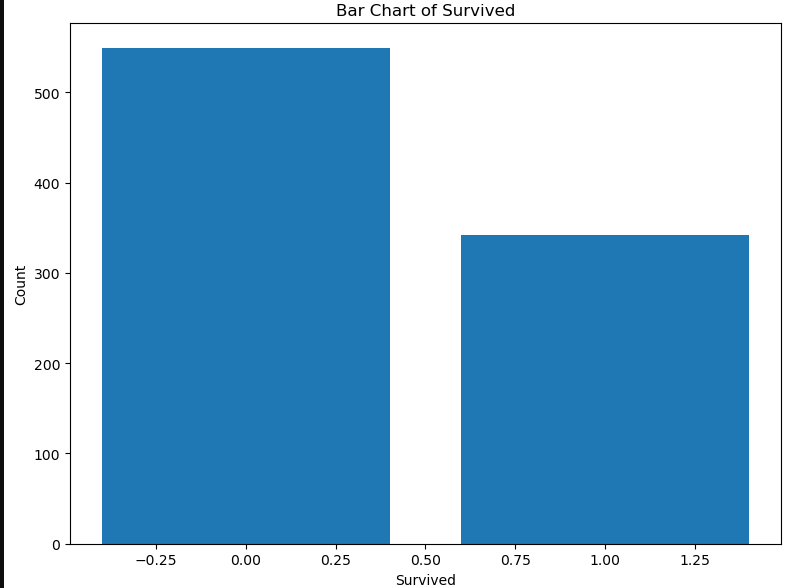


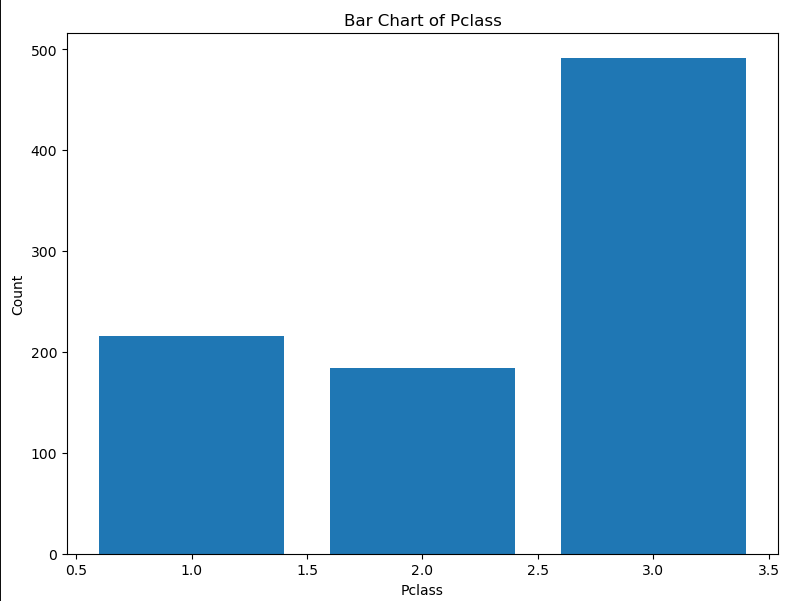


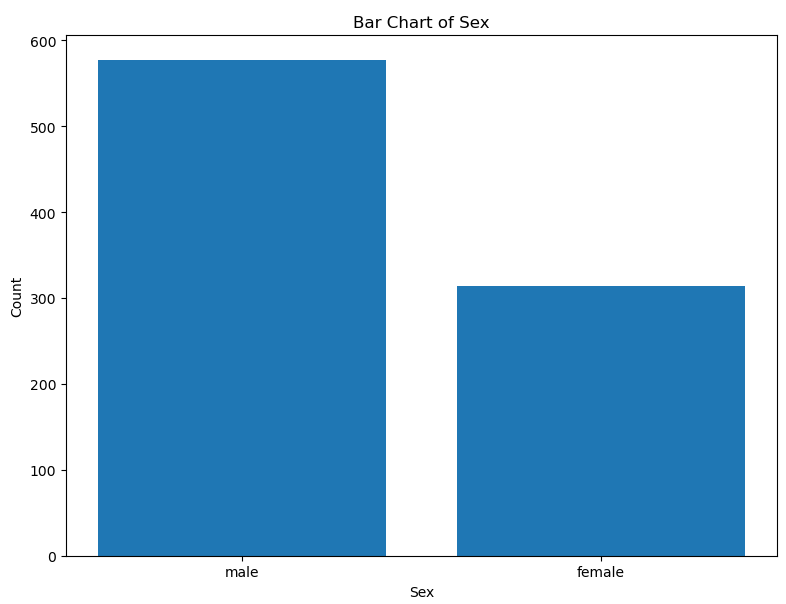


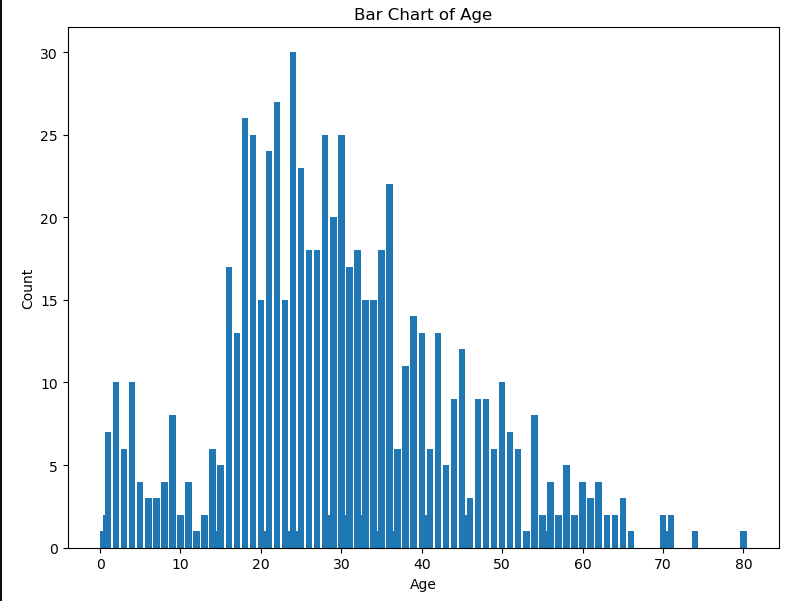


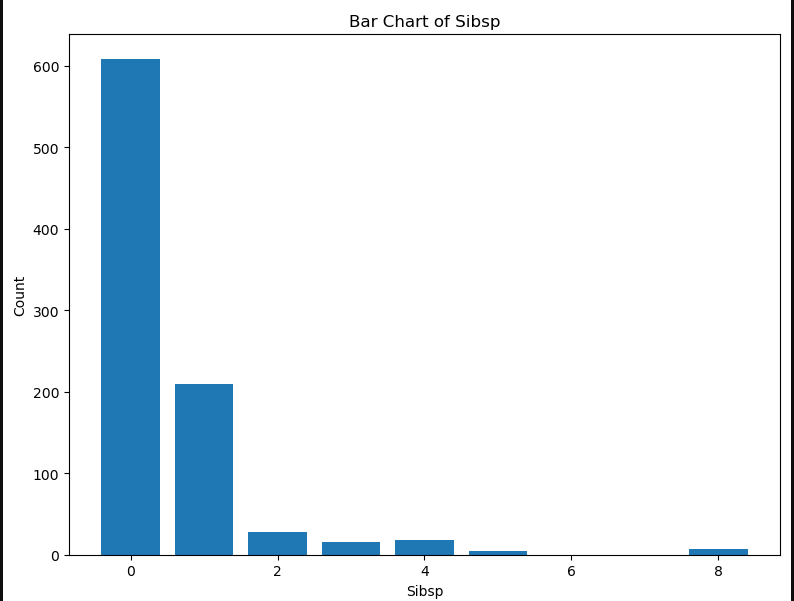


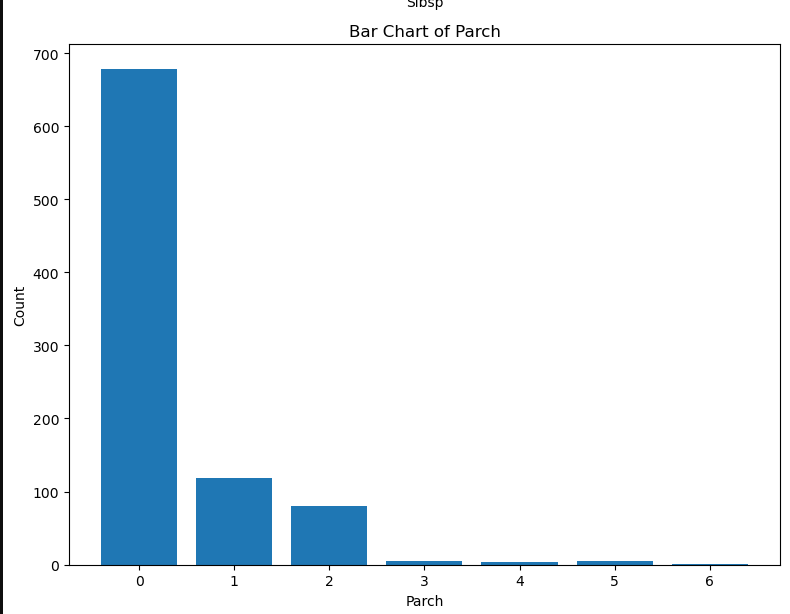


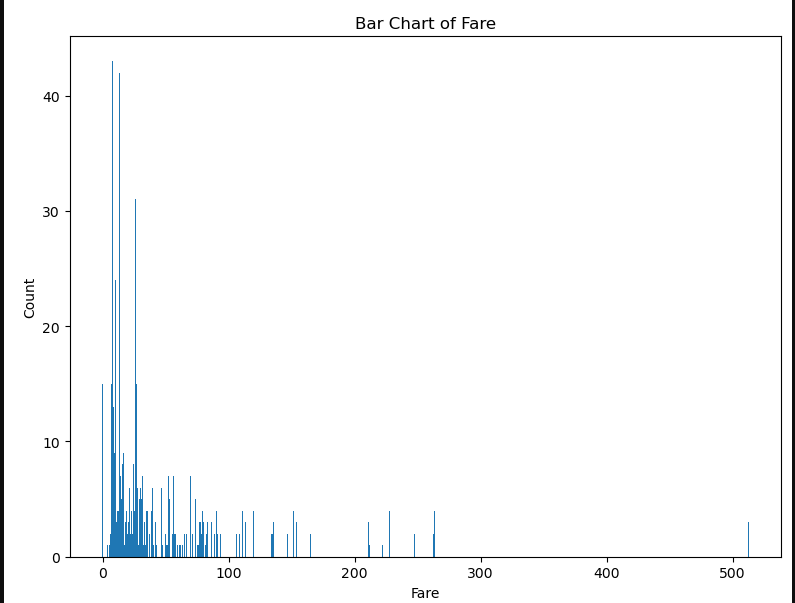


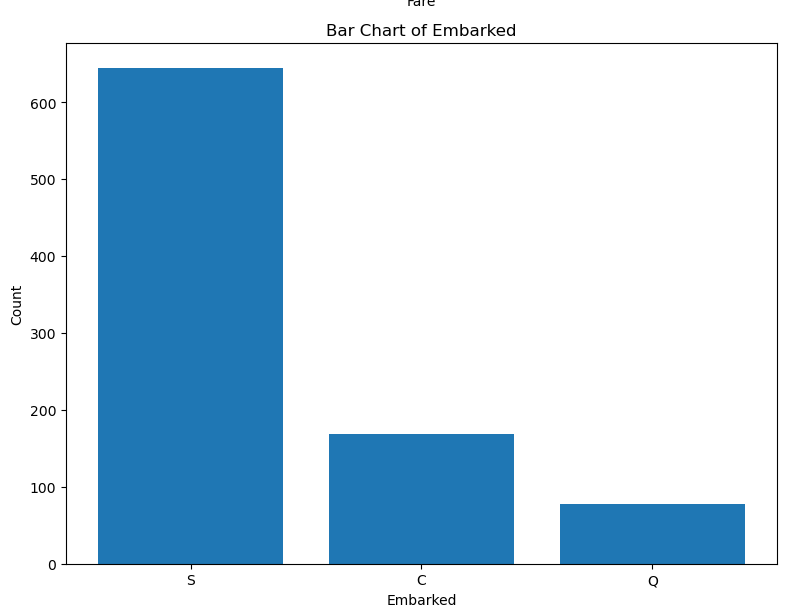


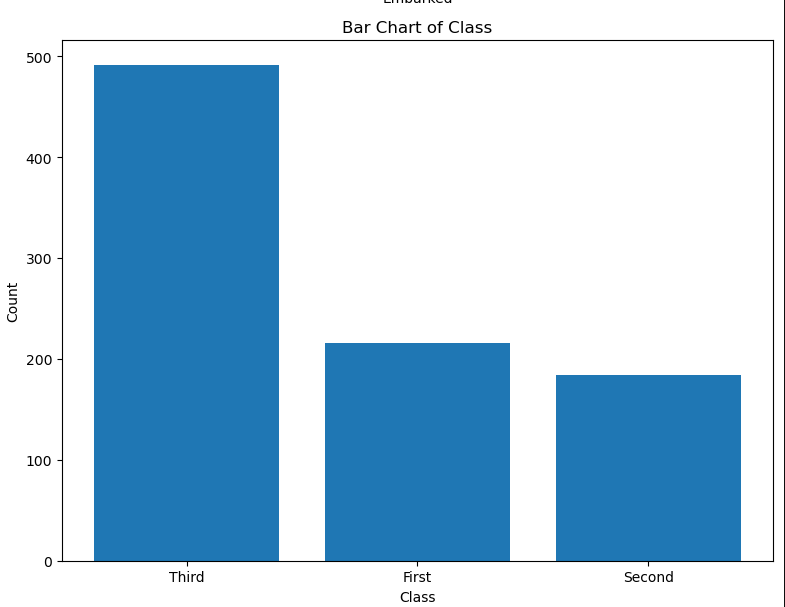


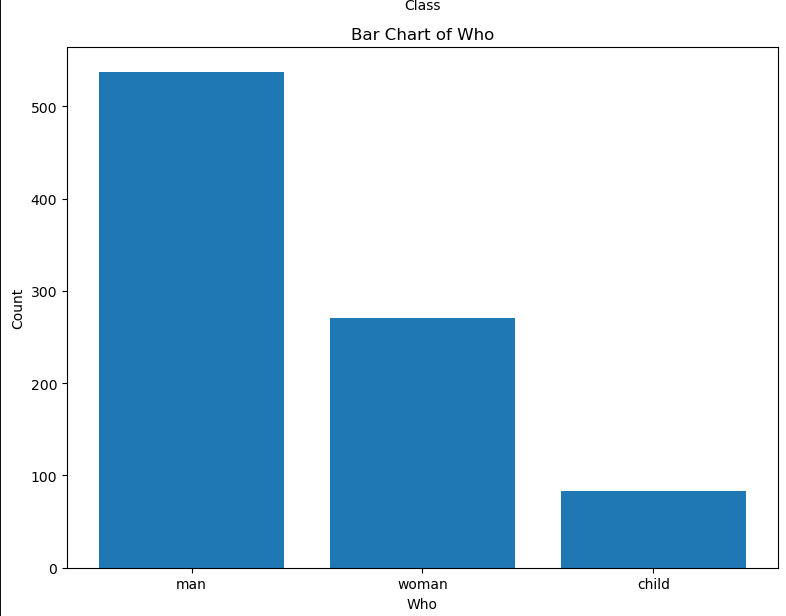


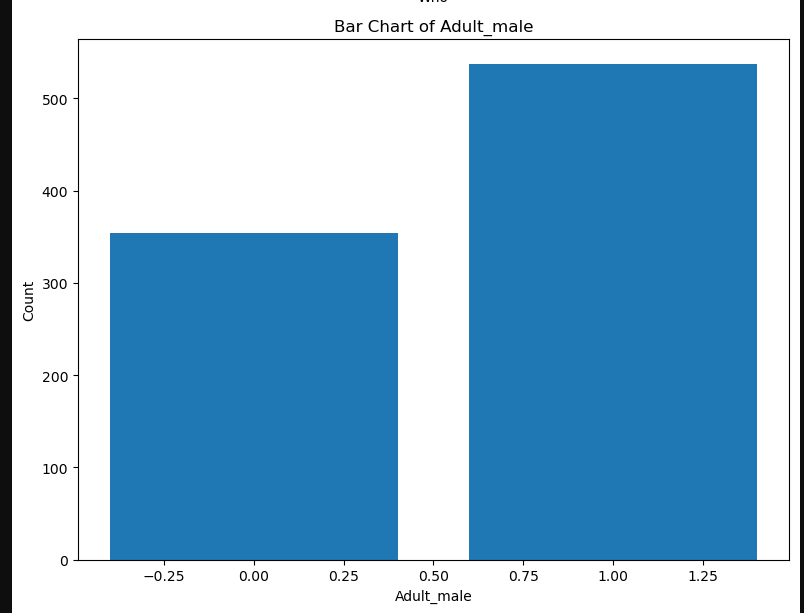


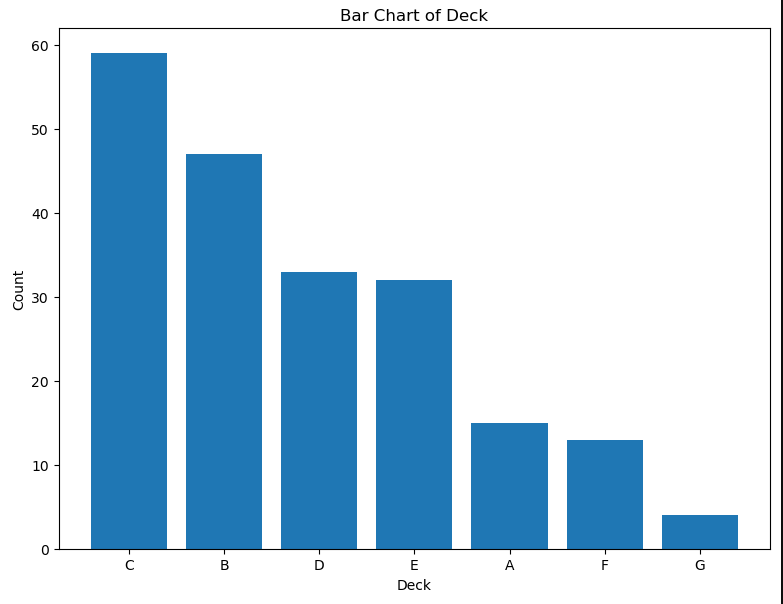


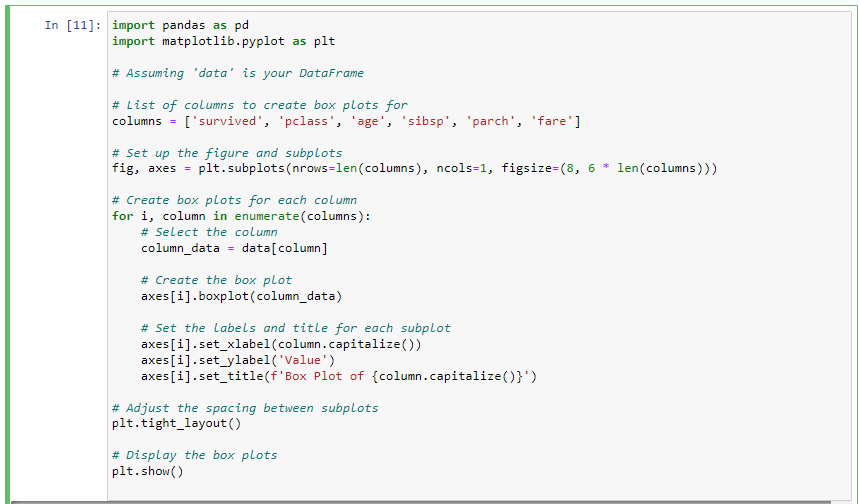


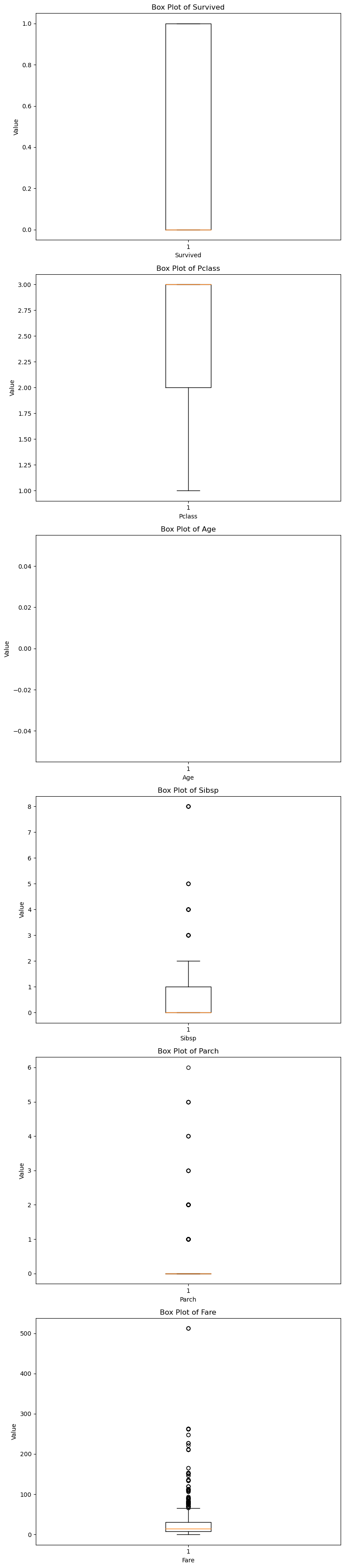


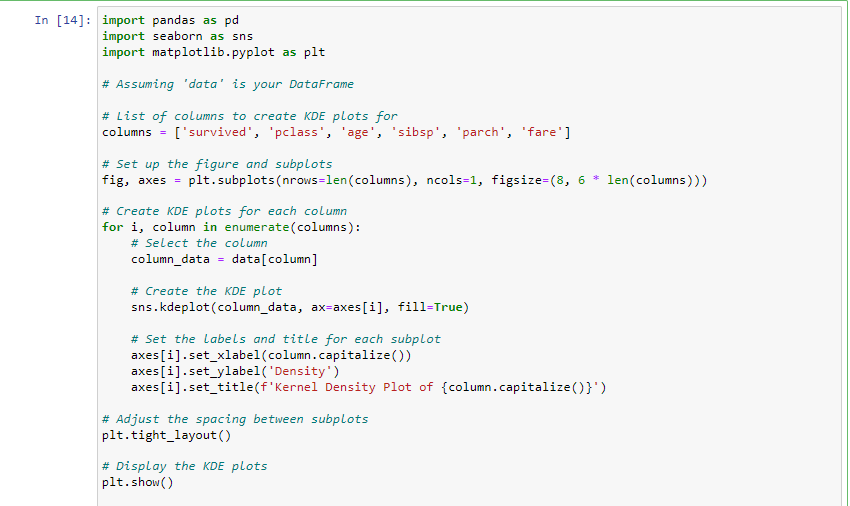


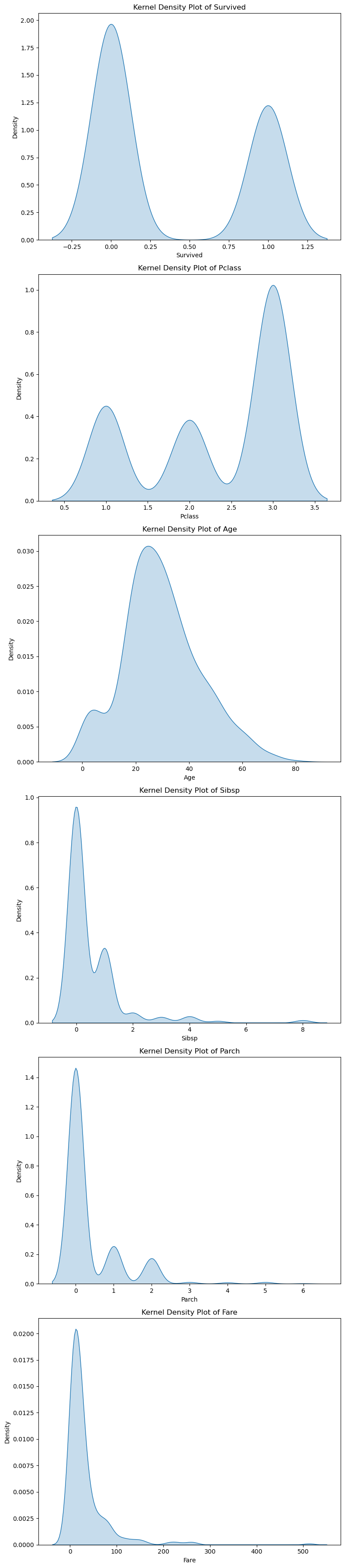


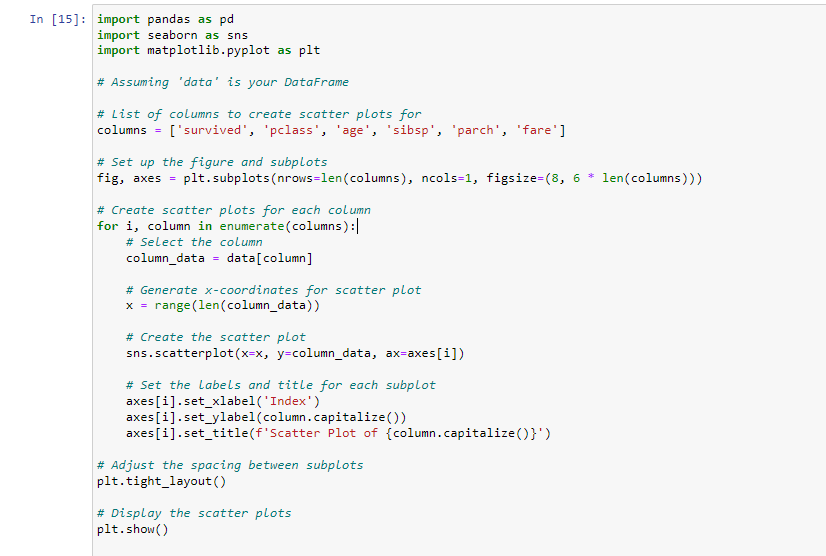


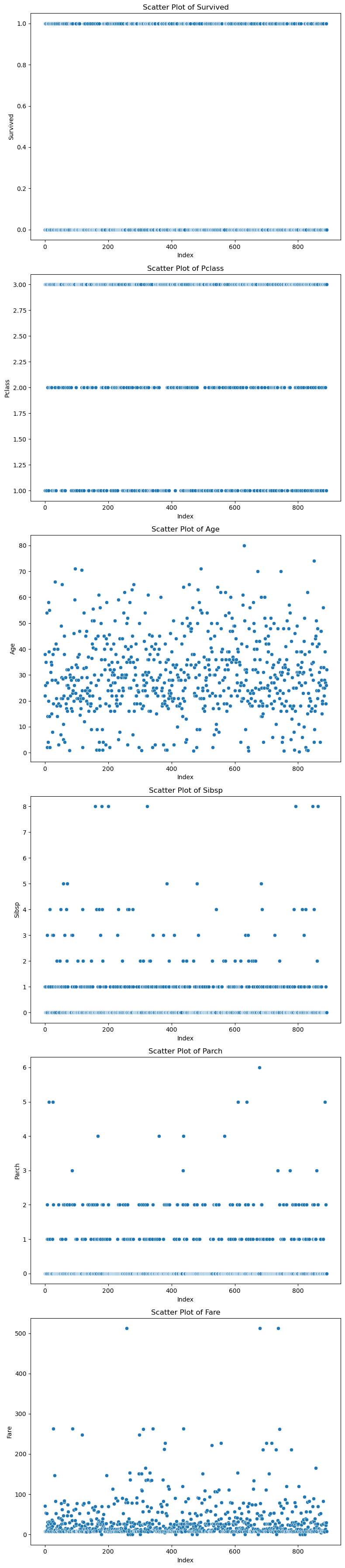


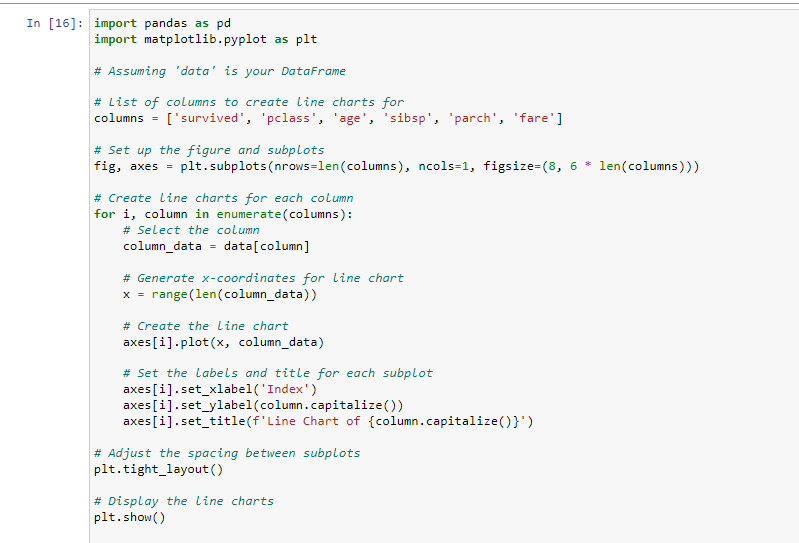


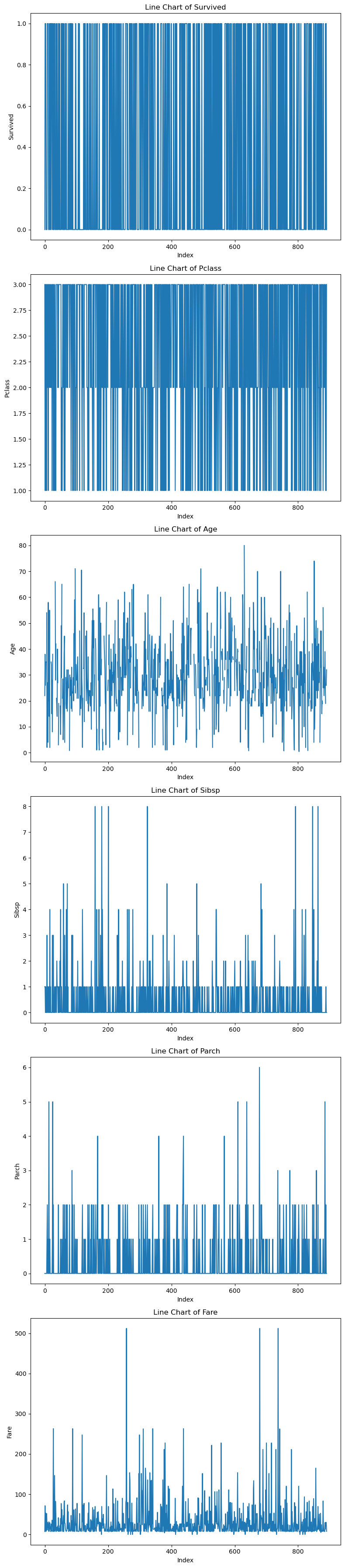


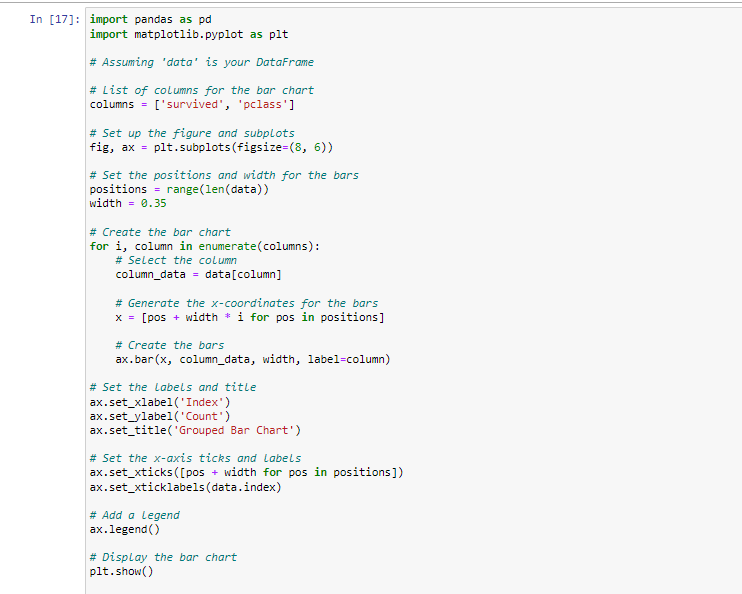


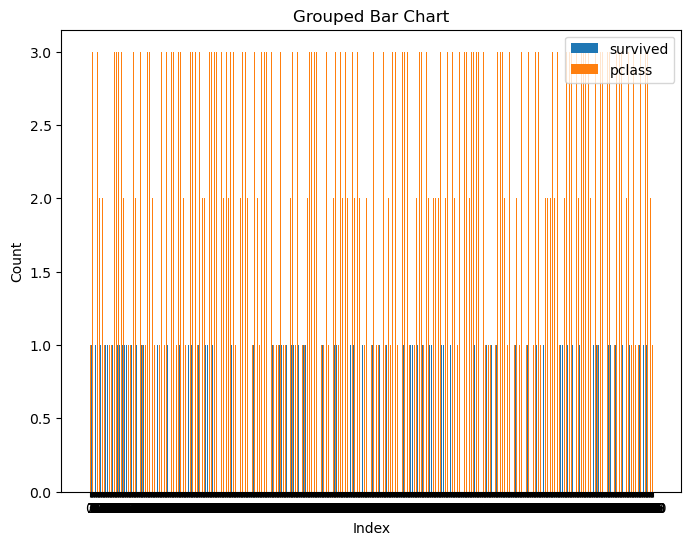




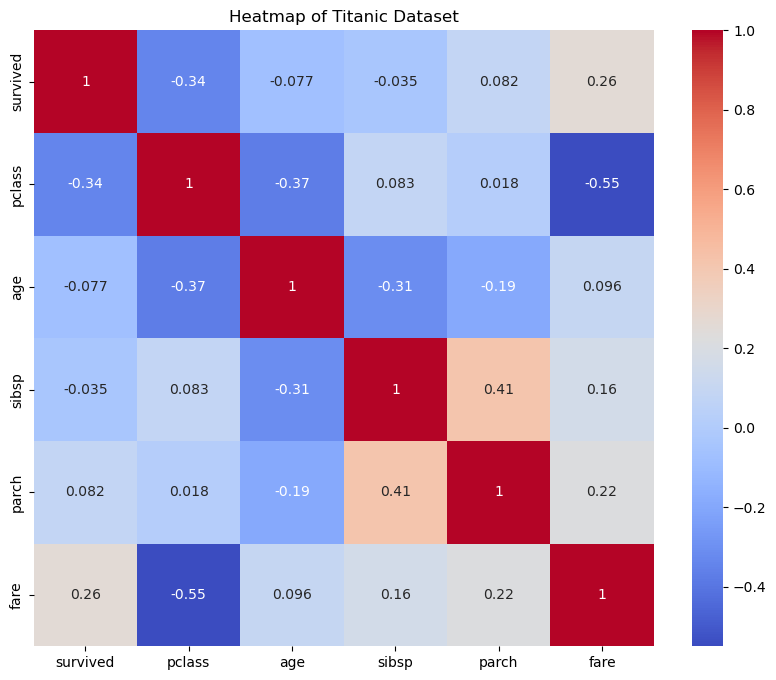


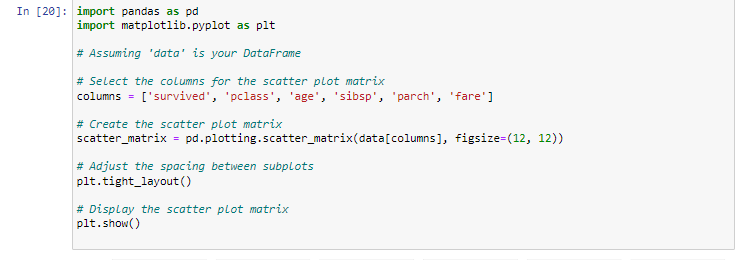


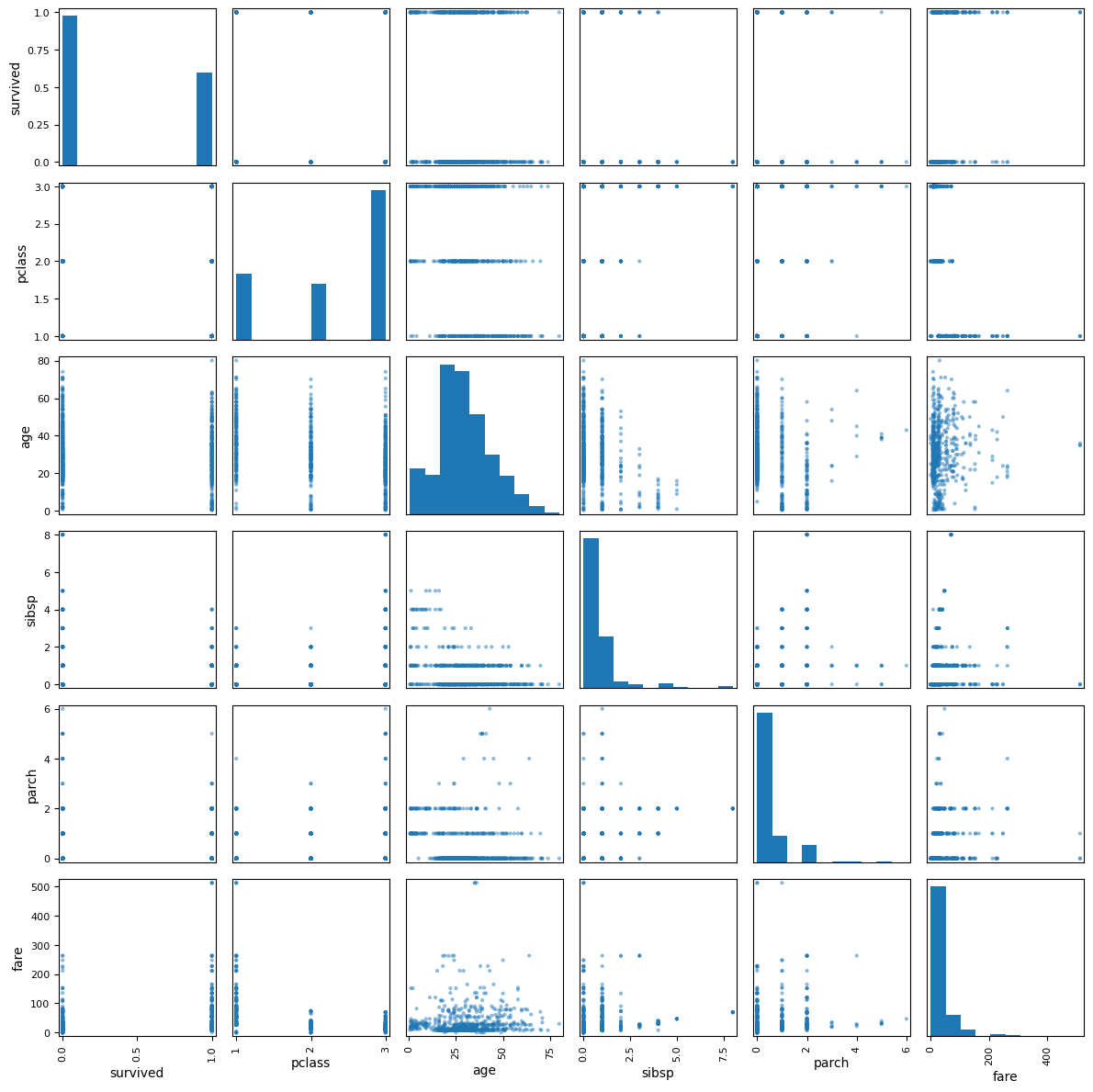


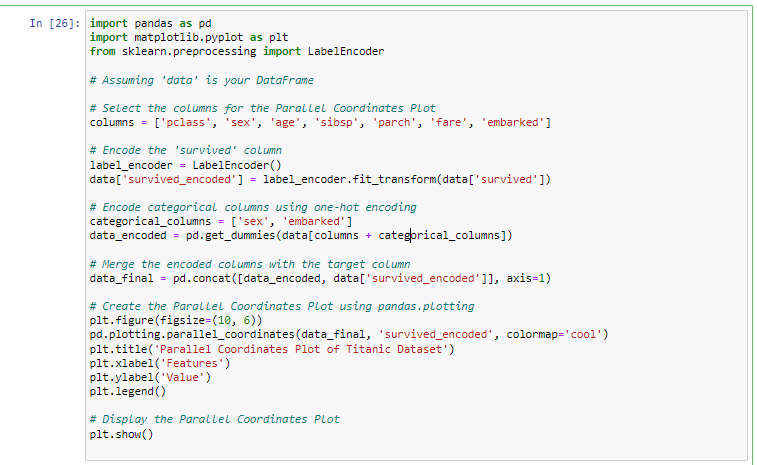


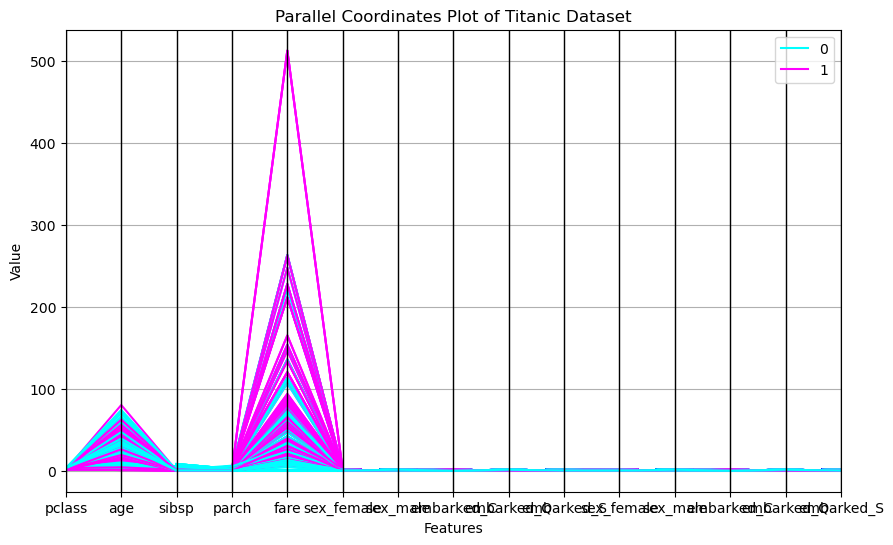


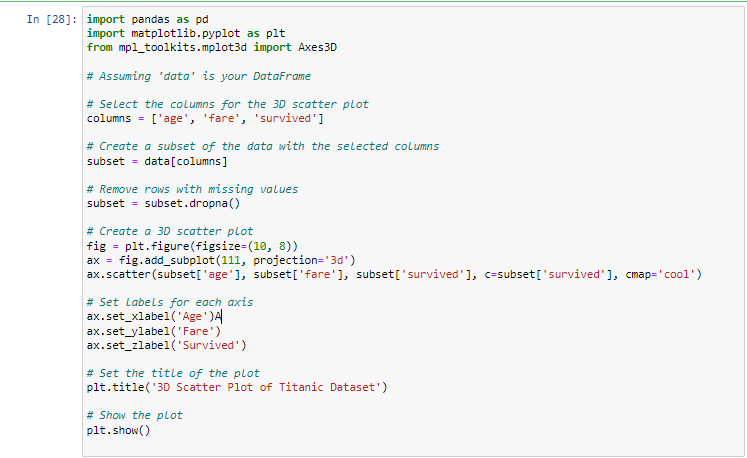


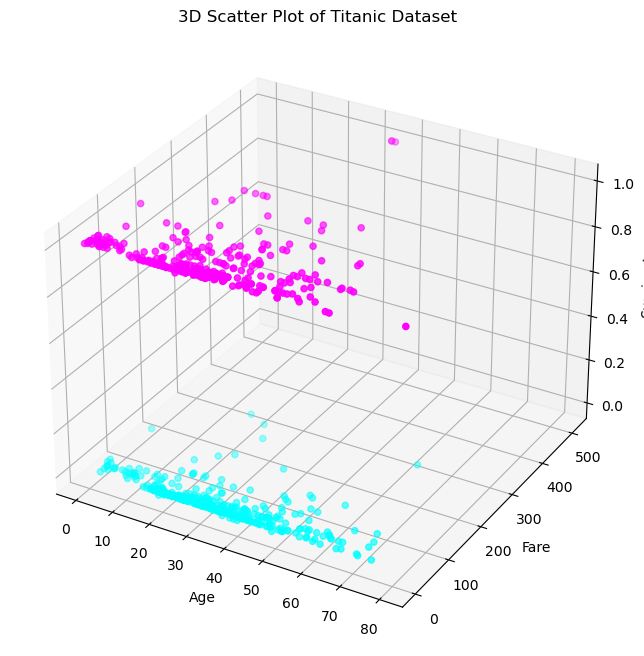


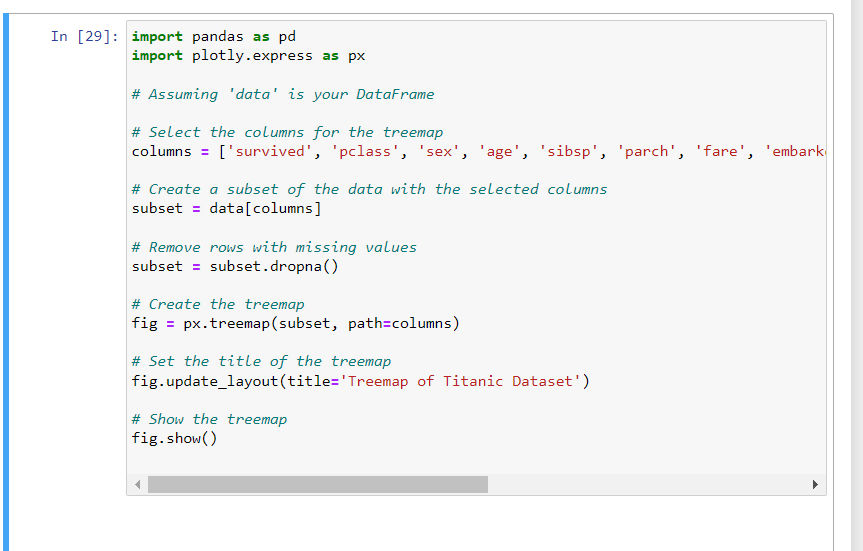




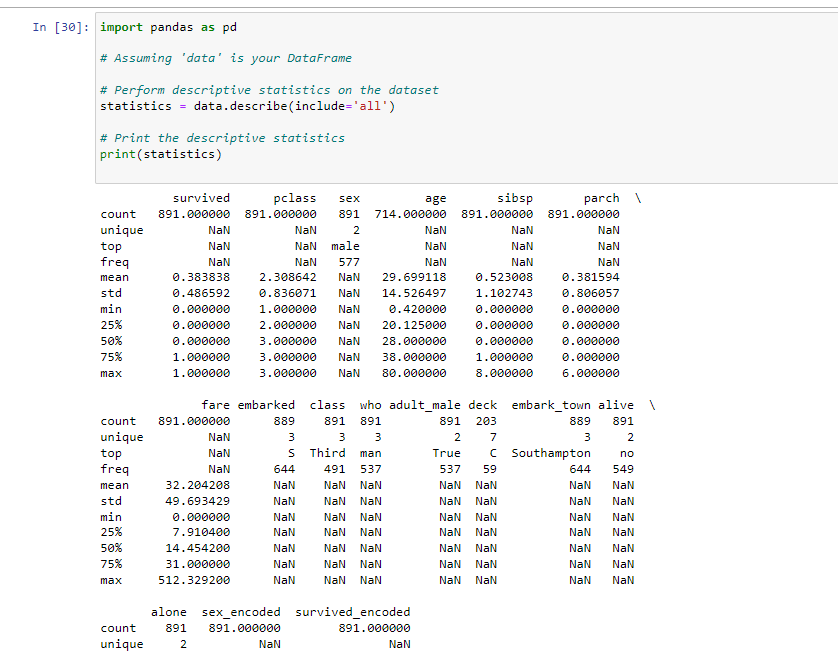




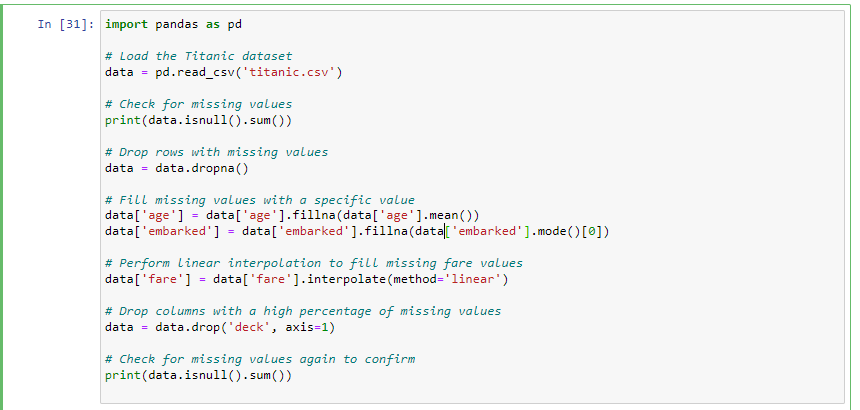


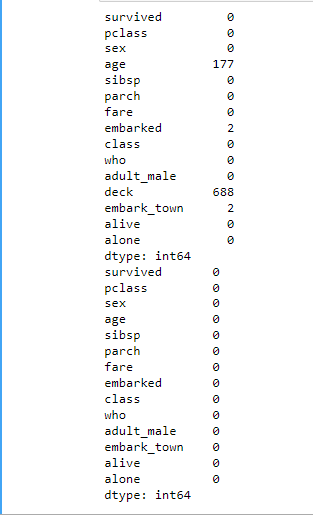


4. Perform descriptive statistics on the dataset

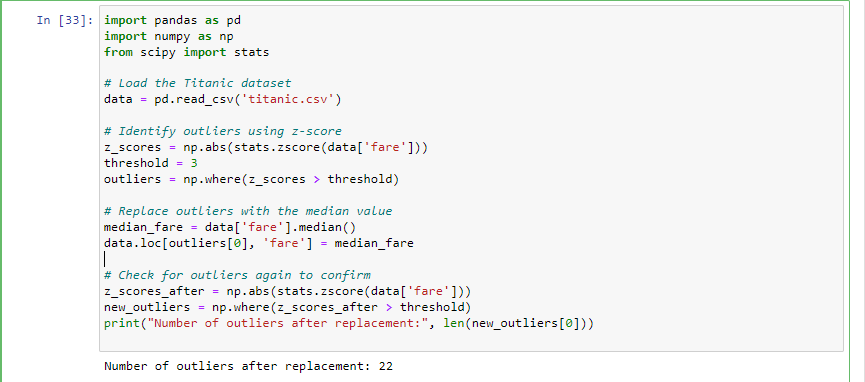


5. Handle the Missing values

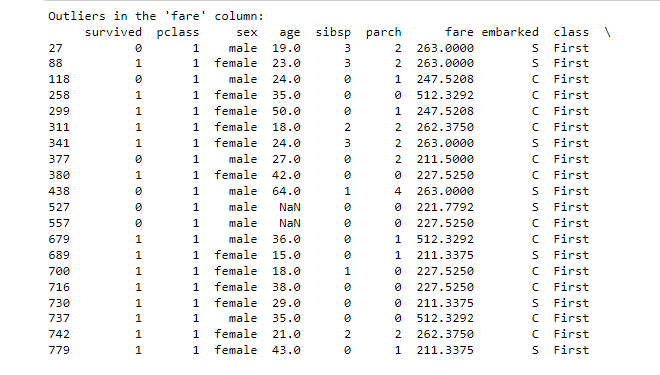


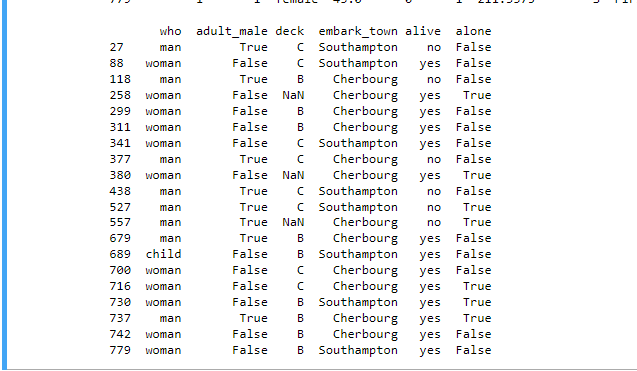


6. Find the outliers and replace the outliers

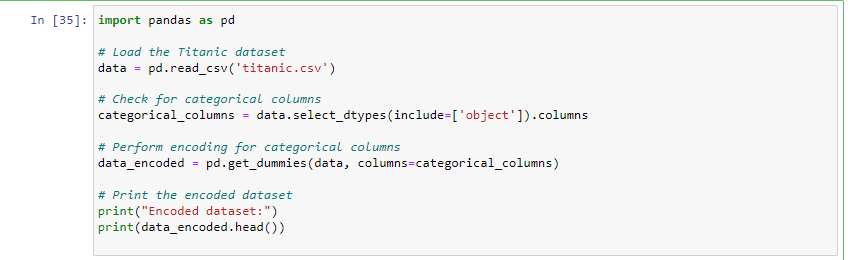


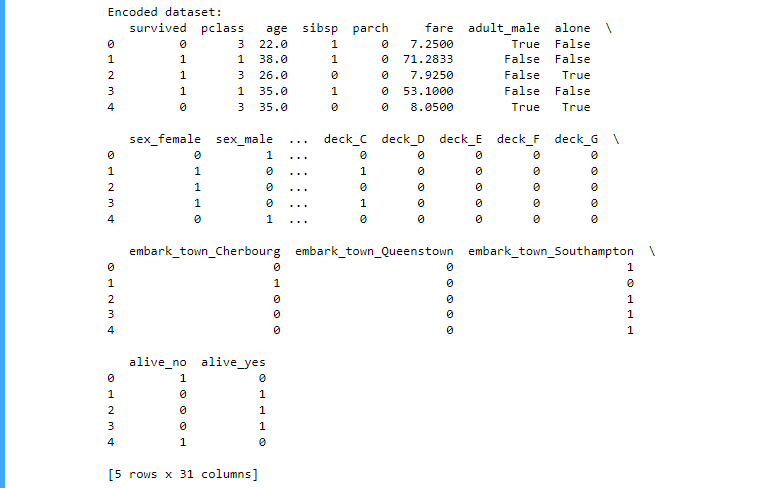




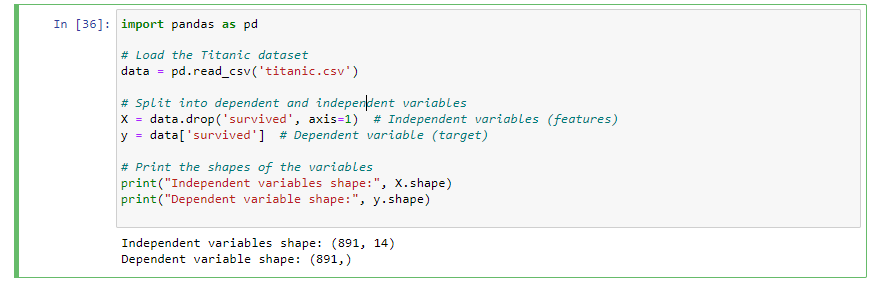


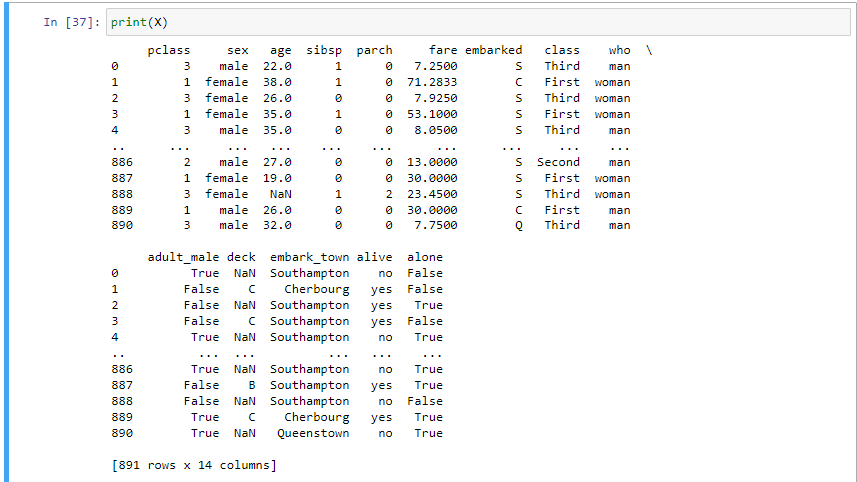
7. Check for Categorical columns and perform encoding

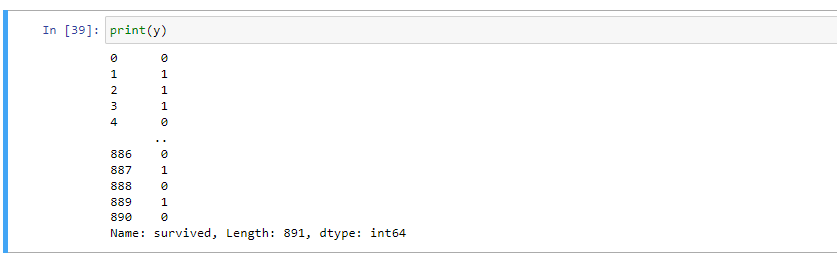




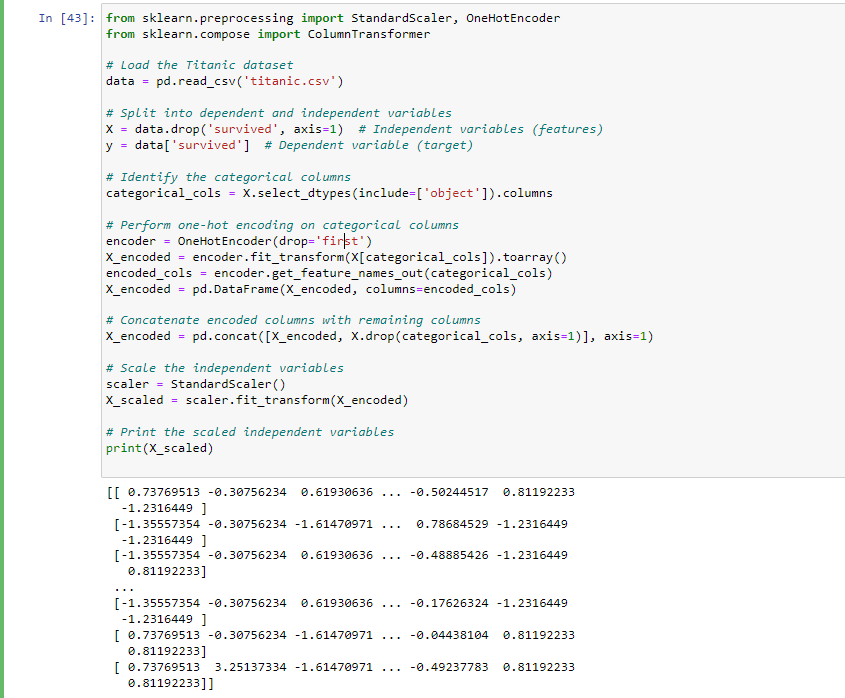
8. Split the data into dependent and independent variables.







9. Scale the independent variables



10. Split the data into training and testing

