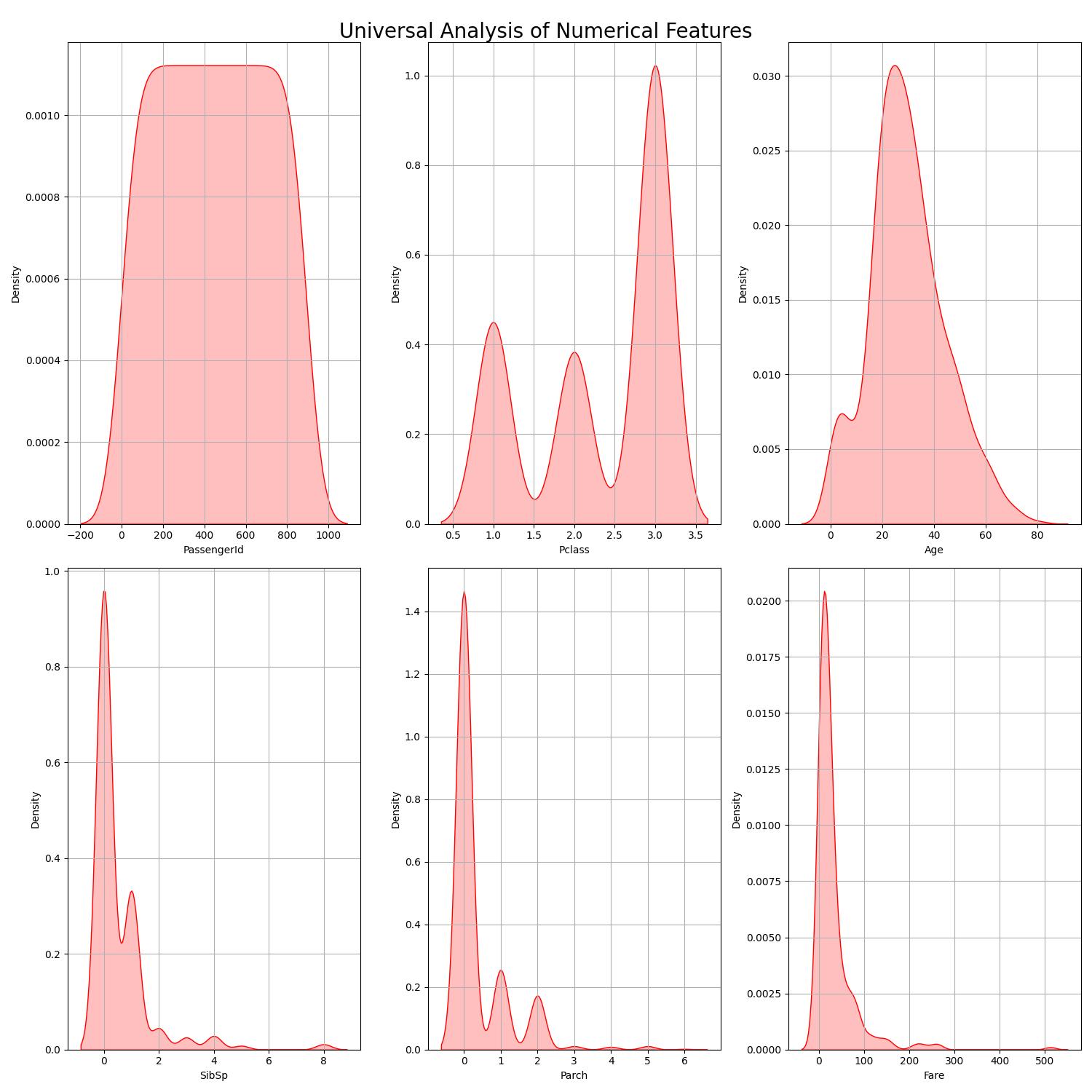
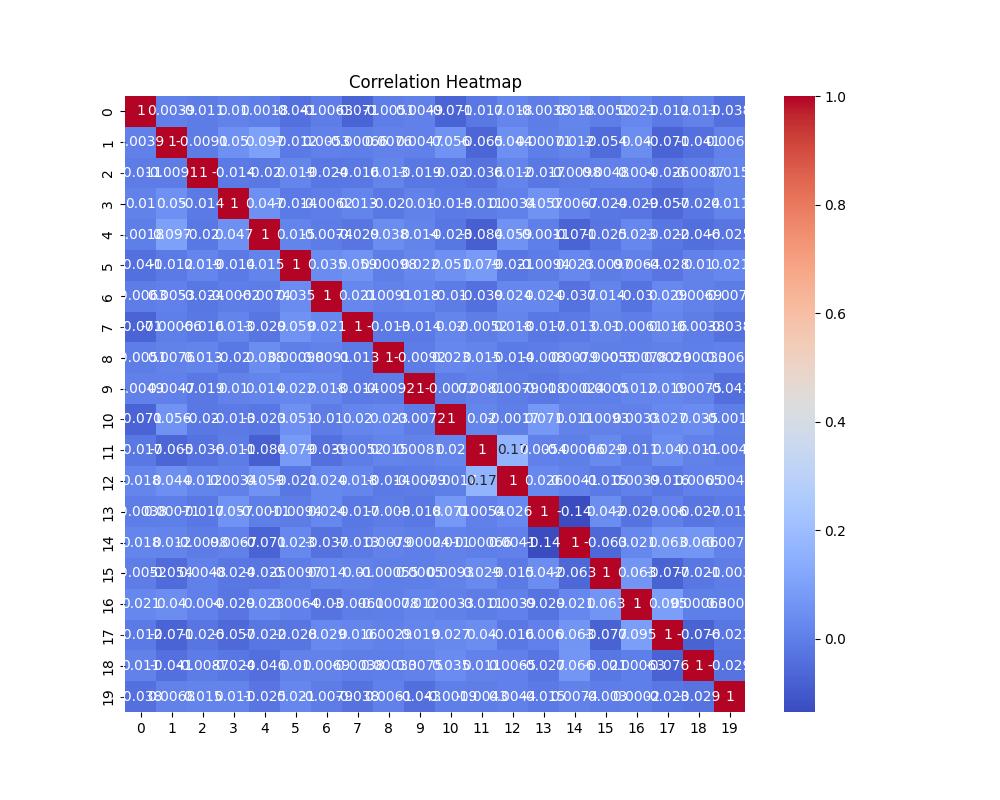
# Data Analysis Report

## Numerical Analysis



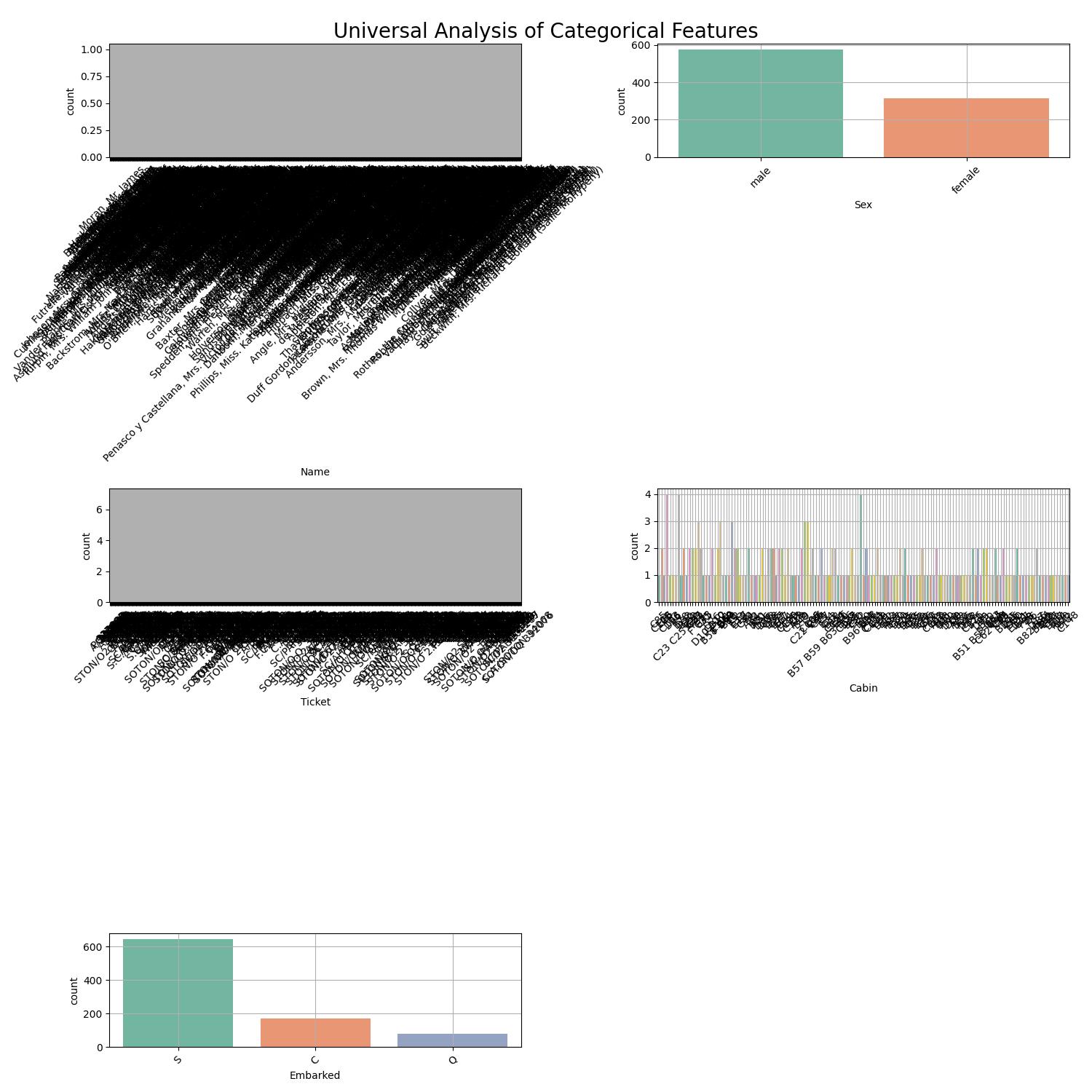
The above plot shows the distribution of numerical features.

## Correlation Heatmap



The above heatmap shows the correlation between features.

## Categorical Analysis



The above plot shows the distribution of categorical features.

## Descriptive Analysis

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Stat | PassengerId | target | Pclass | Age | SibSp | Parch | Fare |
| count | 891.000 | 891.000 | 891.000 | 714.000 | 891.000 | 891.000 | 891.000 |
| mean | 446.000 | 0.384 | 2.309 | 29.699 | 0.523 | 0.382 | 32.204 |
| std | 257.354 | 0.487 | 0.836 | 14.526 | 1.103 | 0.806 | 49.693 |
| min | 1.000 | 0.000 | 1.000 | 0.420 | 0.000 | 0.000 | 0.000 |
| 25% | 223.500 | 0.000 | 2.000 | 20.125 | 0.000 | 0.000 | 7.910 |
| 50% | 446.000 | 0.000 | 3.000 | 28.000 | 0.000 | 0.000 | 14.454 |
| 75% | 668.500 | 1.000 | 3.000 | 38.000 | 1.000 | 0.000 | 31.000 |
| max | 891.000 | 1.000 | 3.000 | 80.000 | 8.000 | 6.000 | 512.329 |

The above table shows the descriptive statistics for numerical features.

The following information summarizes the DataFrame:

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 891 entries, 0 to 890  
Data columns (total 12 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 PassengerId 891 non-null int64   
 1 target 891 non-null int64   
 2 Pclass 891 non-null int64   
 3 Name 891 non-null object   
 4 Sex 891 non-null object   
 5 Age 714 non-null float64  
 6 SibSp 891 non-null int64   
 7 Parch 891 non-null int64   
 8 Ticket 891 non-null object   
 9 Fare 891 non-null float64  
 10 Cabin 204 non-null object   
 11 Embarked 889 non-null object   
dtypes: float64(2), int64(5), object(5)  
memory usage: 83.7+ KB

Best model for provided data is: RandomForestClassifier with accuracy: 0.8507462686567164. This model has been downloaded.