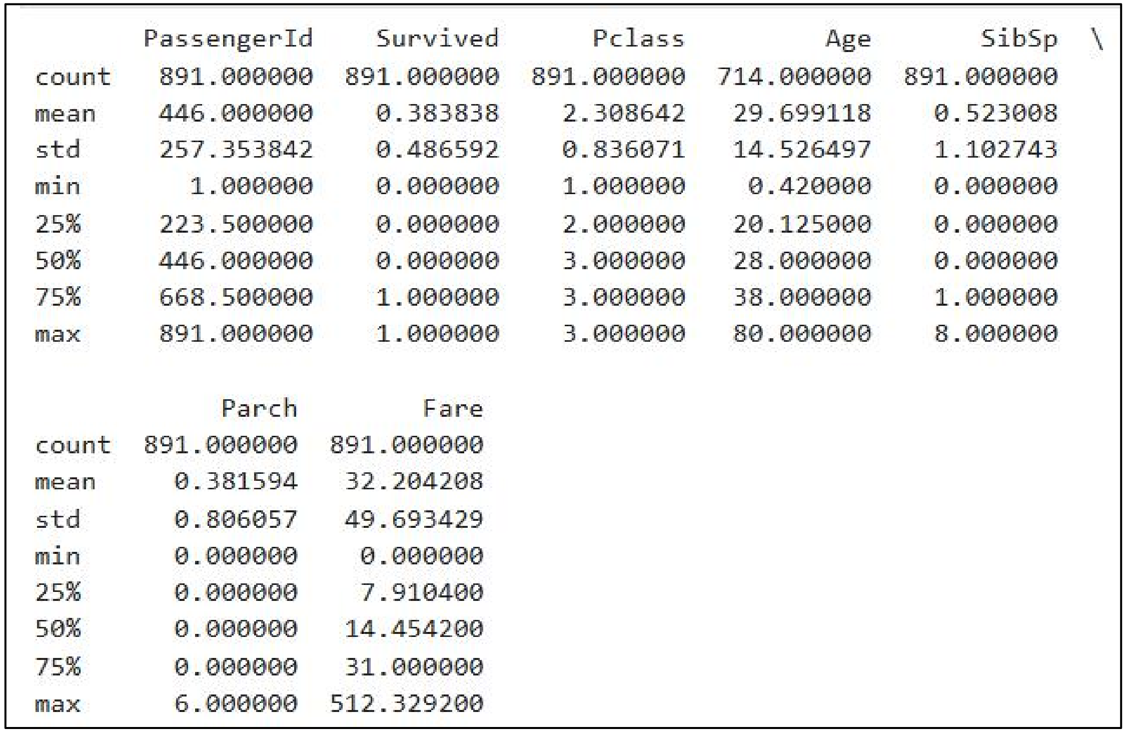
ELEVATE LABS

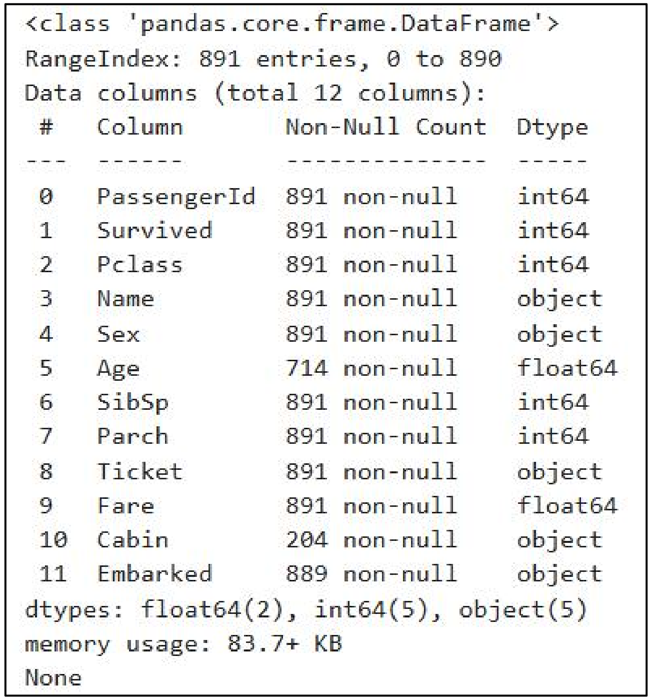
Task - 5

.describe(): generates a descriptive statistic of numerical columns of the dataset



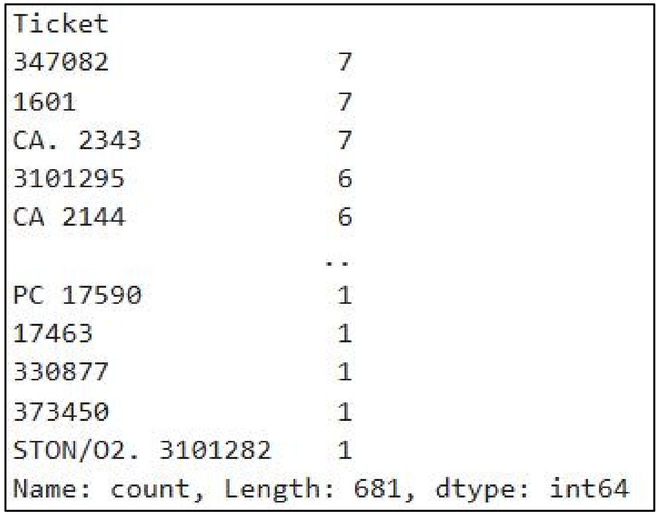
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

.info(): provides a summary of data types, null counts, memory usage of the dataset



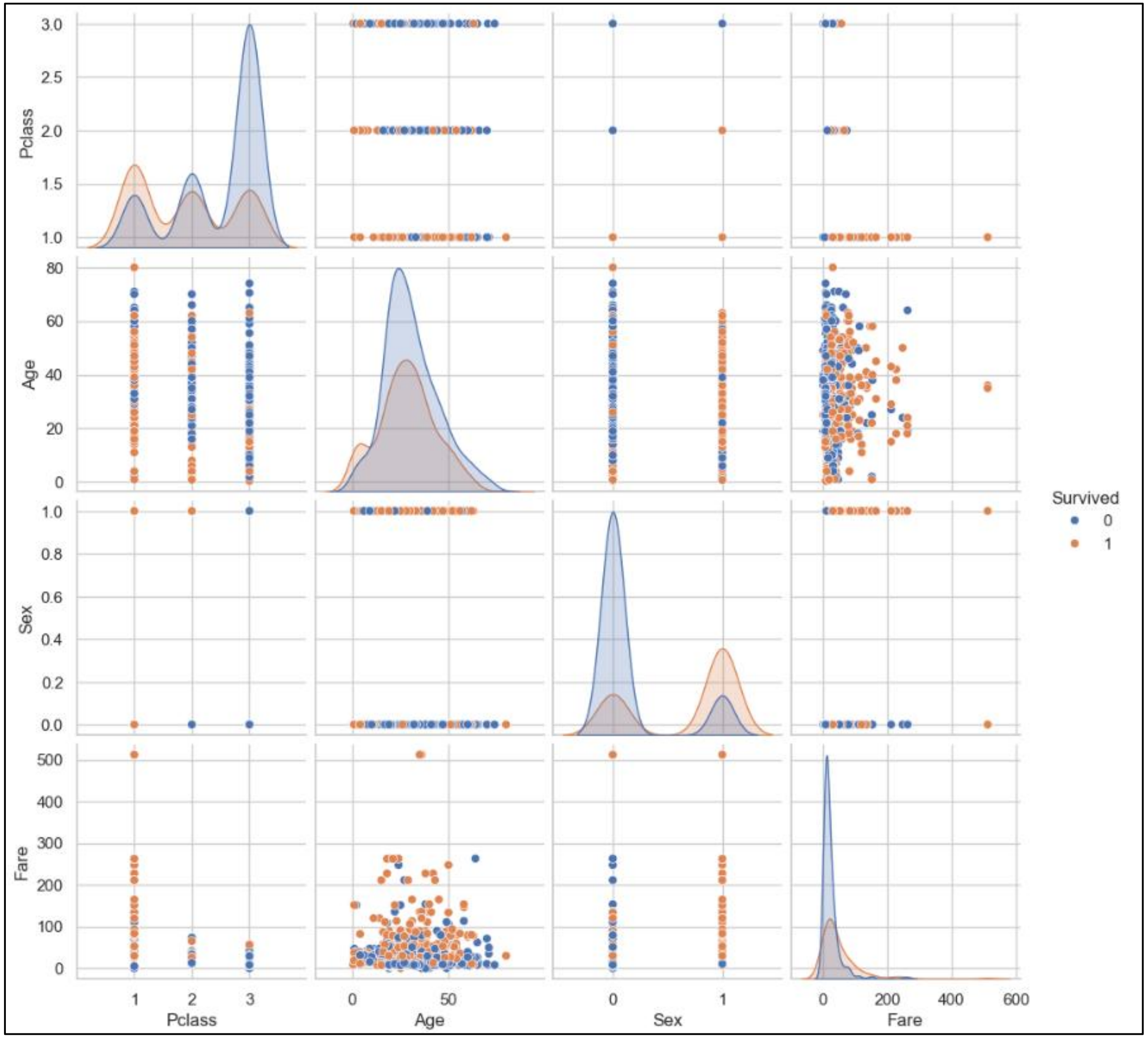
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    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
None
```

.value_counts(): displays how often a unique value appears in the specified column



```
Ticket
347082              7
1601                7
CA. 2343            7
3101295             6
CA 2144             6
                   ..
PC 17590            1
17463               1
330877              1
373450              1
STON/O2. 3101282    1
Name: count, Length: 681, dtype: int64
```

sns.pairplot(): shows scatterplots between every pair of vairables as well as the distribution of each variable. Indicates the relationship and patterns in the data.

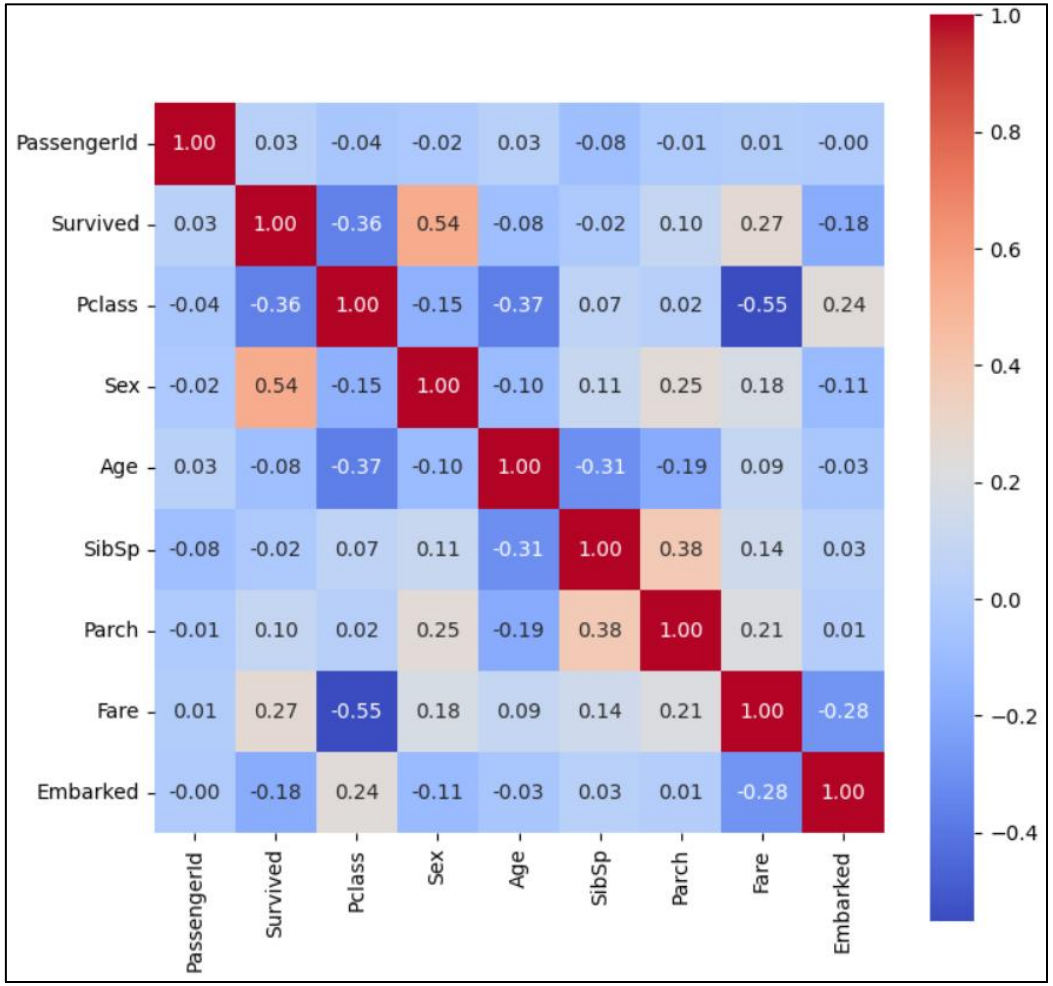


Summary of finding from the pairplot:

- From the given pairplot we can observe that passengers in first class had a higher survival rate (more orange) as compared to the ones in third class who had more deaths (more blue); this indicates that class affected survival.

- Most survivors were females, confirming the “women and children first” rescue priority.
- Survivors paid a higher fare, indicating that wealthier passengers had better chances of survival.
- Children less than 10yrs old had a higher survival rate.
- Males in 3rd class had the lowest survival rate.

sns.heatmap(): shows how strongly variables are connected or related to each other, the colour shows the strength the connection. Each cell in the heatmap indicates the value of correlation (ranging from -1 to +1)



	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
PassengerId	1.00	0.03	-0.04	-0.02	0.03	-0.08	-0.01	0.01	-0.00
Survived	0.03	1.00	-0.36	0.54	-0.08	-0.02	0.10	0.27	-0.18
Pclass	-0.04	-0.36	1.00	-0.15	-0.37	0.07	0.02	-0.55	0.24
Sex	-0.02	0.54	-0.15	1.00	-0.10	0.11	0.25	0.18	-0.11
Age	0.03	-0.08	-0.37	-0.10	1.00	-0.31	-0.19	0.09	-0.03
SibSp	-0.08	-0.02	0.07	0.11	-0.31	1.00	0.38	0.14	0.03
Parch	-0.01	0.10	0.02	0.25	-0.19	0.38	1.00	0.21	0.01
Fare	0.01	0.27	-0.55	0.18	0.09	0.14	0.21	1.00	-0.28
Embarked	-0.00	-0.18	0.24	-0.11	-0.03	0.03	0.01	-0.28	1.00

Summary of the finding from the heatmap:

- Sex -> +0.54: indicates that women were more likely to survive
- Pclass -> -0.36: people in first class had better survival chances
- Fare -> +0.27: people who paid higher ticket prices survived more
- People with more family members onboard might have had a better survival rate
- PassengerId, Embarked and Age have no relation to anything

Breakdown of parameters specified in “sns.heatmap(correlation, annot=True, cmap = 'coolwarm', fmt=".2f", square = True)”

- ‘correlation’: input data for the heatmap
- ‘annot=True’: shows the actual correlation values inside each cell
- ‘cmap=‘coolwarm’’: specifies the color map, blue indicates negative values, white indicates zero and red indicates positive values
- ‘fmt=“.2f”’: controls the formatting of the numbers in the cells
- ‘square=True’: makes each cell in the heatmap squre-shaped rather than rectangular, gives a clean and symmetric look to the matrix

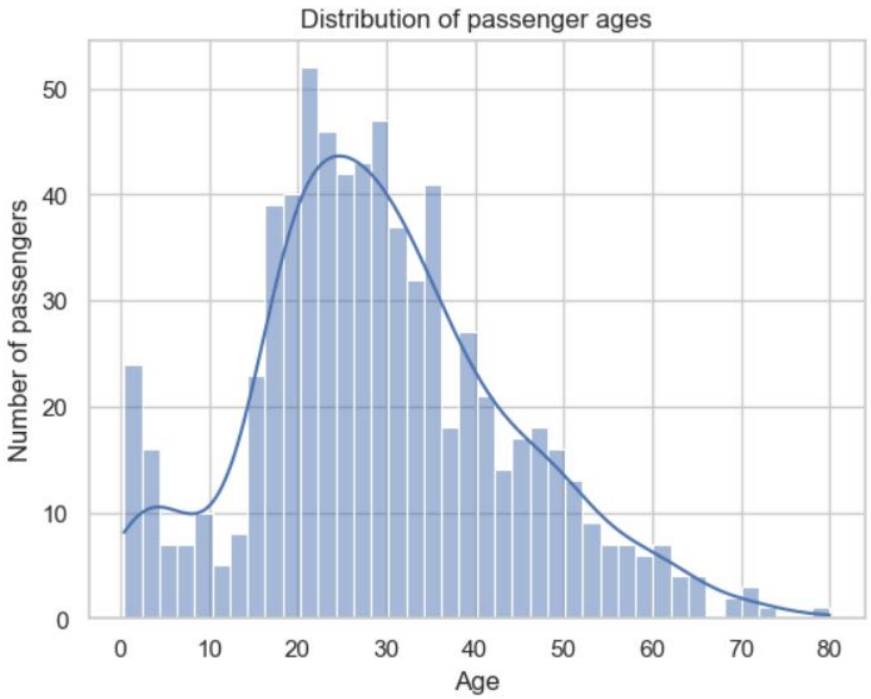
Relationships:

Column 1	**Column 2**	**Type of relationship**	**Meaning/Interpretation**
Sex	Survived	Positive	Females survived more
Pclass	Survived	Negative	3rd class had a lower survival rate
Fare	Survived	Positive	Higher fare -> more murvival
SibSp	Parch	Positive	Families traveled together
Pclass	Fare	Negative	Higher class = higher fare
Parch	Survived	Weak positive	Some family helped survival

Trends:

- Females survived much more than males
- Higher class = better survival chances
- People who paid more had a better chance of survival
- People with family aboard had slightly higher survival
- Age and boarding location didn’t matter much on their own

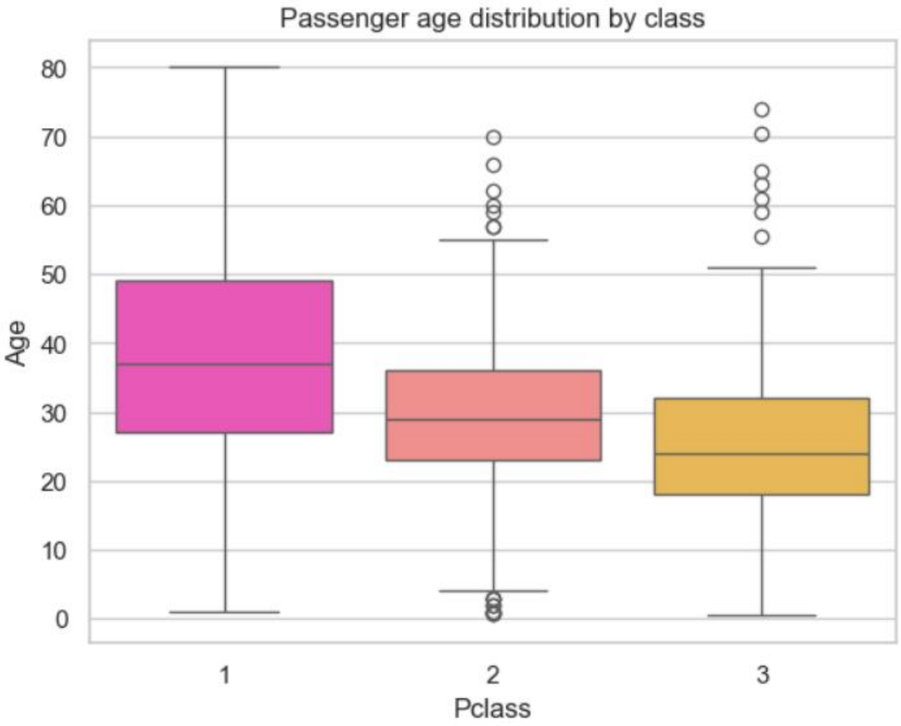
Histogram



Observations:

- From the given histogram, we can observe that the peak is around 20 to 30 year olds, which means there were many young adults on board.
- Some children and few passengers above 60 were present but most of the passengers were middle-aged.
- The curve is right-skewed, therefore we can note that although most passengers were in their 20s or 30s, a few older passengers in their 80s were also on board.
- The smooth KDE curve helps visualize that the majority were between 20 and 40, with a sharp drop after that.

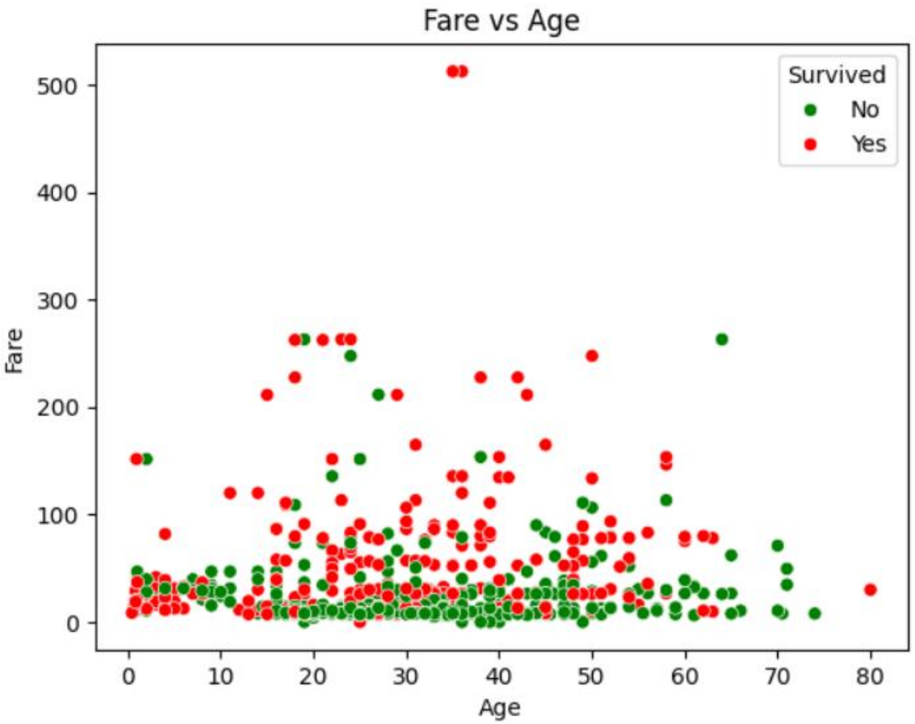
Box plot



Observations:

- The given boxplot indicates that the median age in each class is ~ 38yrs, ~29yrs, ~24yrs, respectively.
- There’s a clear trend in the given dataset, i.e., higher class passengers tended to be older, indicating the socioeconomic patterns of the time- older, wealthier people could afford 1st clss tickets, while younger possibly families and labourers, traveled in 3rd class.

Scatter plot



Observations:

- The given scatter plot indicates that most fares were low,
- younger to middle-aged passengers were more in number (between 20 and 40 years old)
- people who paid a higher fare for the ticket had a higher survival rate, while most of the passengers who paid a lesser fair did not survive.