Exploratory Data Analysis - Titanic Dataset

Objective:-

To explore the Titanic dataset and extract meaningful insights using statistical and visual techniques such as histograms, heatmaps, boxplots, and pairplots.

Tools Used:-

Python

Pandas

Seaborn

Matplotlib

Google Colab

Steps Performed:-

1. Loaded the dataset and checked its structure using .info() and .head().

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
actions	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

2. Summarized statistical properties using .describe() and checked for null values.

```
Dataset Shape: (891, 12)

Data Types:
 PassengerId        int64
Survived           int64
Pclass             int64
Name              object
Sex               object
Age              float64
SibSp              int64
Parch              int64
Ticket            object
Fare             float64
Cabin             object
Embarked          object
dtype: object
```

3. Performed univariate analysis on 'Age', 'Sex', 'Pclass', and 'Fare'.

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	891	2	NaN	NaN	NaN	681	NaN	147	3
top	NaN	NaN	NaN	Dooley, Mr. Patrick	male	NaN	NaN	NaN	347082	NaN	G6	S
freq	NaN	NaN	NaN	1	577	NaN	NaN	NaN	7	NaN	4	644
mean	446.000000	0.383838	2.308642	NaN	NaN	29.699118	0.523008	0.381594	NaN	32.204208	NaN	NaN
std	257.353842	0.486592	0.836071	NaN	NaN	14.526497	1.102743	0.806057	NaN	49.693429	NaN	NaN

4. Visualized relationships using histograms, boxplots, countplots, and pairplots.

Pairplot of Key Features
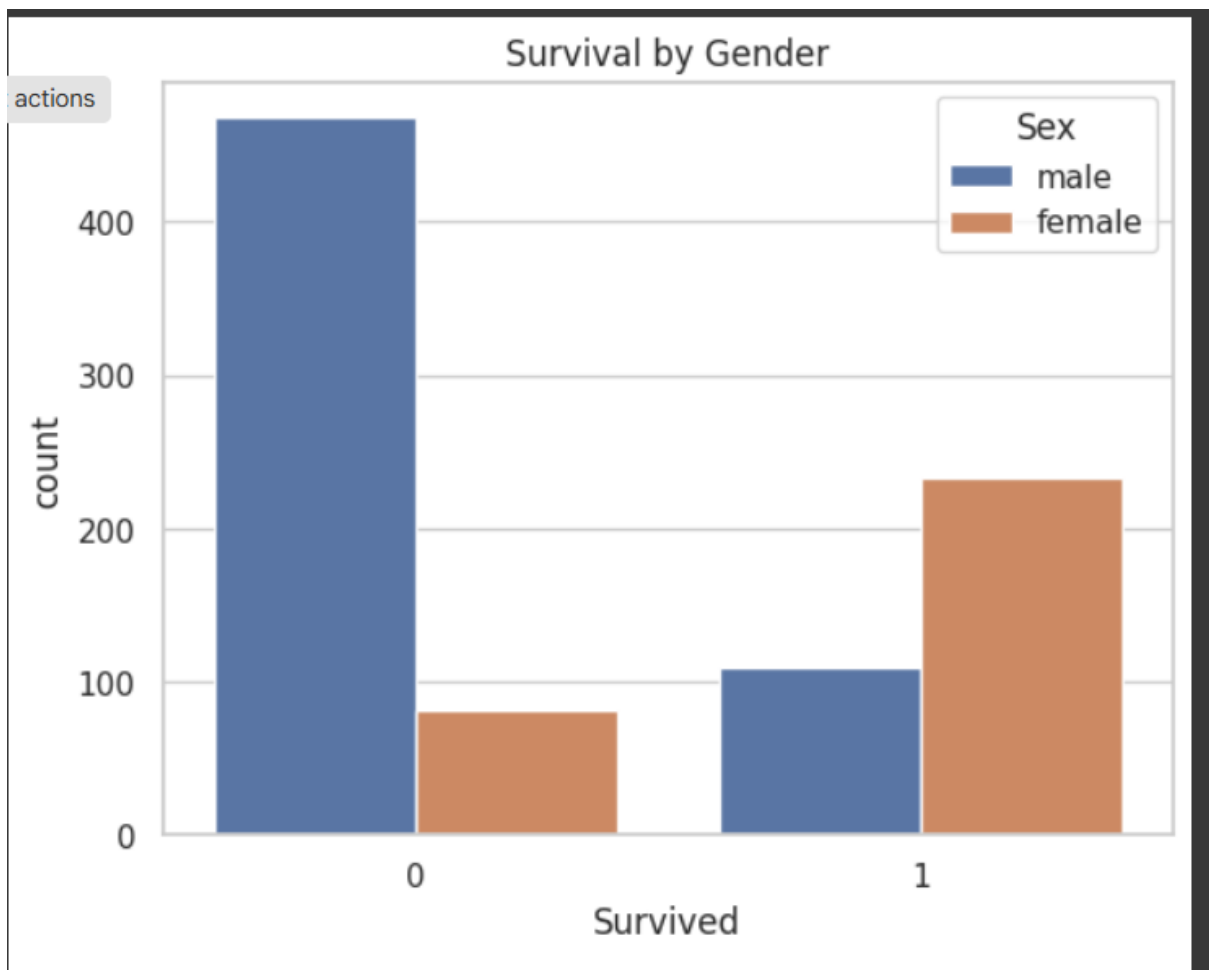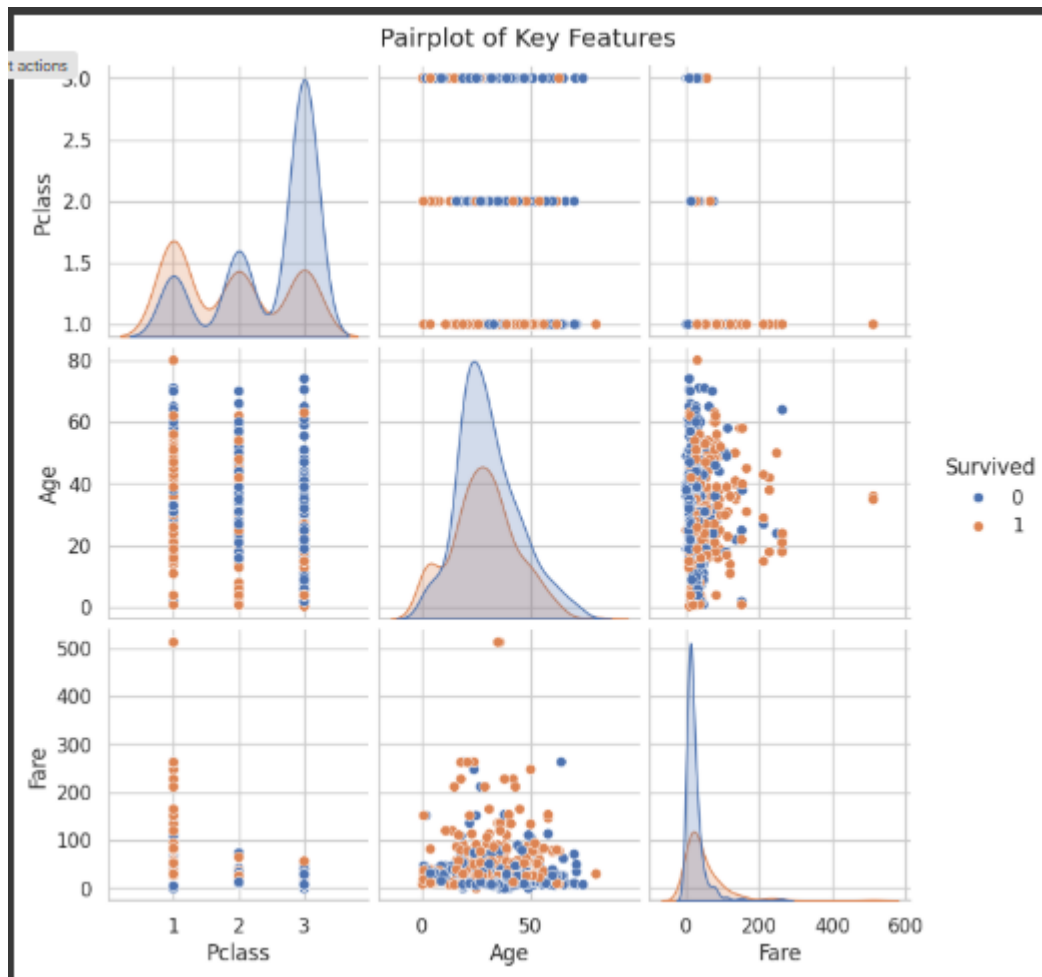
5. Created a correlation heatmap to explore linear relationships.

Correlation Heatmap
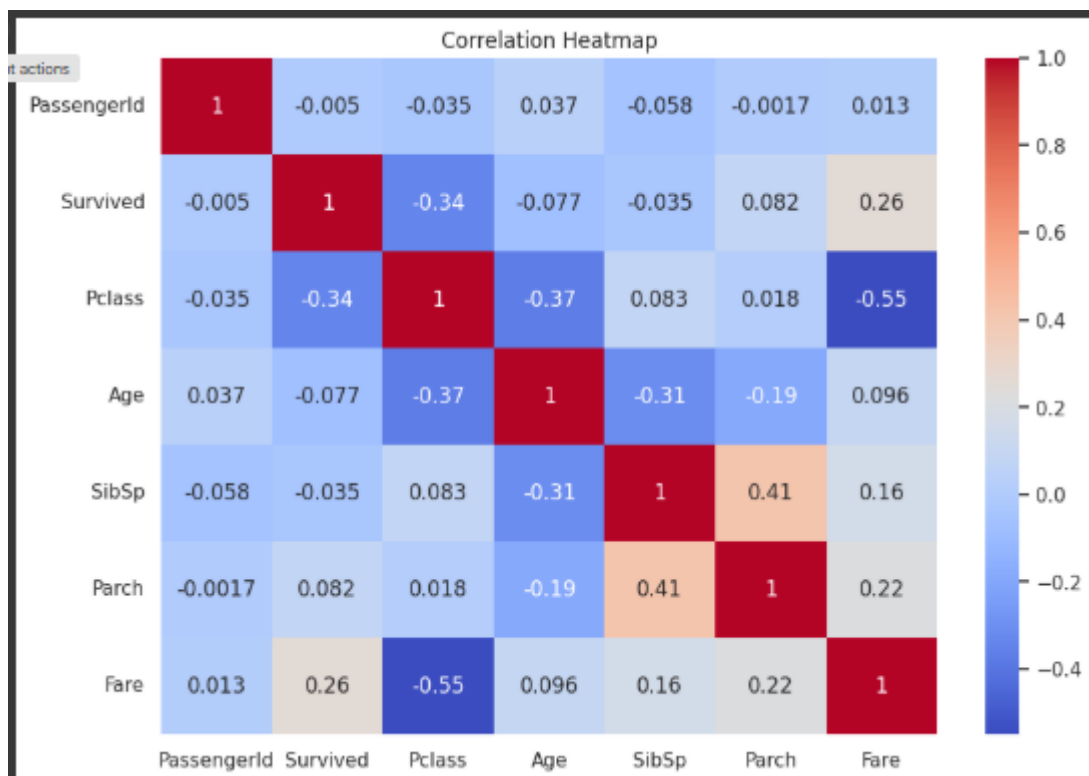
6. Wrote observations under each visualization and summarized insights.

```python
# Step 10: Summary of Findings

print("🔍 Observations:")
print("- More males than females, but more females survived.")
print("- Higher-class passengers had a better survival rate.")
print("- Younger passengers had slightly higher survival chances.")
print("- Fare and Pclass are moderately correlated.")
print("- Strong visual trends show gender and class influenced survival.")
```

Key Insights: –

Females had a higher survival rate than males.

Passengers in higher classes had better chances of survival.

Younger passengers were more likely to survive.

Fare and Pclass showed moderate correlation with survival.

Strong patterns observed between survival and gender/class.