Extracting insights using visual and statistical exploration.

Observations :

1. Use of .info() and .describe()

- Columns like Age, Cabin, and Embarked have missing values.
- Fare and Age columns have a wide range of values.
- Data types are a mix of numbers and text.

2. Use of .value_counts() for Sex and Embarked

- There are more males than females in the dataset.
- Most passengers boarded from the port "S".

3. Use of Age Histogram

- Most passengers were between 20 and 40 years old.
- There are fewer very young and very old passengers.

4. Use of Fare Boxplot

- Most passengers paid lower fares.
- Some passengers paid very high fares, which are seen as outliers.

5. Use of Survival Countplot

- The number of people who did not survive is higher than the number who survived.

6. Use of Survival by Sex

- A larger proportion of females survived compared to males.

7. Use of Survival by Passenger Class

- First-class passengers had a higher survival rate compared to second and third class passengers.

8. Use of Pairplot

- Passengers who paid higher fares generally had better survival chances.
- There is no very clear separation between age and survival outcome.

9. Use of Heatmap

- Fare and Survival are positively related (higher fare, more chance of survival).
- Age and Survival show very weak or almost no relationship.

Summary of Findings:

- Female passengers had a higher chance of survival compared to males.
- First-class passengers survived more than second- and third-class passengers.
- Higher fares were associated with higher chances of survival.
- Most passengers were young adults between 20 and 40 years old.
- The majority of passengers boarded from Southampton port.
- Age, Cabin, and Embarked columns have missing data and should be handled carefully during further analysis.
- A few passengers paid very high fares, indicating outliers.
- Overall, gender, passenger class, and fare amount were significant factors that affected survival.

Conclusion

This analysis of the Titanic dataset provided valuable insights into the factors influencing passenger survival. It was observed that gender, passenger class, and fare were significant indicators of survival chances. Females, first-class passengers, and individuals who paid higher fares had better survival rates. Most passengers were young adults, and the majority boarded from Southampton. Some missing values were also identified, particularly in the Age, Cabin, and Embarked columns, which should be addressed in further modeling. These initial findings offer a strong foundation for building predictive models and for deeper analysis of the Titanic disaster data.