Exploratory Data Analysis (EDA) - Titanic Dataset

Internship Task 5 - Elevate Labs

1. Dataset Overview
- Dataset: Titanic (train.csv, test.csv, gender_submission.csv)
- Features include Passenger details: Age, Sex, Pclass, Fare, Embarked, Survived, etc.
- Train Data Shape: (891, 12)
- Test Data Shape: (418, 11)
- Gender Submission: Survival outcomes for test dataset passengers.

2. Key Findings
- More passengers did not survive than survived.
- Females had a significantly higher survival rate compared to males.
- First class passengers had higher survival rates than 2nd and 3rd class.
- Most passengers were in 3rd class, but survival was lowest in this group.
- Younger passengers and higher fare-paying passengers had better survival chances.
- Passengers embarked from port 'C' had better survival rates than 'Q' and 'S'.

3. Visual Exploration Techniques Used
- Countplots for categorical features (Survival, Gender, Passenger Class).
- Histograms for age distribution.
- Heatmap for correlation among numerical variables.
- Boxplots and scatterplots for relationships (Age vs Survival, Fare vs Age).

4. Conclusion
The Titanic EDA revealed strong influence of Gender, Class, and Fare on survival rates.
Women and first-class passengers had the best survival chances, while third-class males had the lowest.