SUMMERY

OBSERVATIONS:

Age Distribution of Passengers

- Most passengers are between 20-40 years old, indicating that the majority are young adults.

- There is a noticeable presence of children below 10, though fewer in number.

- The density curve (KDE) suggests a peak around 30 years, meaning a large fraction of passengers were in this age group.

- Few elderly passengers, with very few aged above 60.

Survival Rate by Passenger Class

- Higher-class passengers (Pclass = 1) had a higher survival rate, possibly due to better access to lifeboats.

- Lower-class passengers (Pclass = 3) had the lowest survival rate, suggesting class influenced survival probability.

- This aligns with historical records showing privileged passengers had better survival odds.

Feature Correlation Heatmap

- Fare and Pclass are strongly negatively correlated (-0.55) → Higher-class passengers paid significantly more.

- Age has weak correlation with survival, meaning age alone didn't determine survival chances.

- SibSp (number of siblings/spouses) and Parch (number of parents/children) have weak correlations but could impact survival when combined.

- Survival correlates positively with Fare, showing higher-paying passengers had better survival odds.

Pairplot (Age, Fare, Pclass, Survived)

- Younger passengers in higher classes had better survival rates.

- Higher Fare values strongly correlate with survival, as seen in earlier analyses.

- Pclass visually separates survivors, with most Class 1 passengers surviving.

- Overall, class and fare played a more decisive role than age alone.

Boxplot of Passenger Fare

- Significant outliers in fare values → Some passengers paid extremely high fares.

- The median fare is relatively low, meaning most passengers paid less than premium prices.

- High fares mostly belong to Pclass 1 passengers, supporting the correlation found earlier.

Summary of Findings from Exploratory Data Analysis (EDA)

Survival Trends:

- Passenger class (Pclass) was the strongest survival factor → Higher-class passengers had significantly better survival chances.

- Women and children had a higher survival rate, supporting the "women and children first" approach.

- 3rd Class passengers had the lowest survival rate, likely due to limited access to lifeboats.

Fare vs. Survival:

- Higher ticket fares correlated with better survival odds → premium-paying passengers (likely in 1st class) were more likely to survive.

- Outliers in fare prices suggest some passengers paid exceptionally high fares.

Age Distribution Insights:

- Most passengers were aged 20-40, with a peak around 30 years.

- Very few elderly passengers (above 60 years).

- Children (under 10 years) were aboard but in small numbers.

Feature Relationships & Correlations:

- Fare and Passenger Class (Pclass) showed strong negative correlation (-0.55) → Lower-class passengers paid much less.

- Survival positively correlated with fare, meaning higher-paying passengers had better chances.

- Age alone did not strongly impact survival, though younger passengers in higher classes survived more.

Outlier Detection:

- Fare had extreme outliers, showing premium ticket holders.

- Some high-ticket fares in Pclass 1 indicate VIP passengers.

Key Takeaways

- Class & Fare were major survival determinants.
- Women & children survived more than men.
- Age didn't significantly influence survival alone.
- Fare data contained extreme values (high-ticket individuals).
- Passengers with small/medium-sized families had better survival odds.
- Cherbourg-embarked passengers had a survival advantage.
- Survivors paid significantly higher fares, supporting class-based favoritism.
- Cabins mattered—certain letters corresponded with survival probability.
- Luxury Score (Fare/Class) was a strong survival indicator.