Titanic Dataset – Exploratory Data Analysis (EDA)

Objective

The objective of this project is to explore and analyze the Titanic dataset to understand the factors that influenced passenger survival. This analysis uses visual exploration to identify patterns related to survival, focusing on class and gender. Each visualization is analyzed with a clear aim, observation, conclusion, and reasoning to demonstrate analytical thinking.

Overall Survival Distribution

Aim: Understand overall survival outcomes.

Observation: Deaths significantly outnumber survivors.

Conclusion: Majority of passengers did not survive.

Reasoning: Limited lifeboats, lack of preparedness, and rapid disaster progression.

Note: This graph does not consider class or gender distinctions.

Survival by Passenger Class (Count)

Aim: Examine survival by passenger class using raw counts.

Observation: Third-class deaths are highest; first-class survivors are most prominent.

Conclusion: Passenger class influenced survival outcomes.

Reasoning: Higher classes had better access and priority.

Note: Unequal class sizes make raw counts misleading.

Survival by Passenger Class (Percentage)

Aim: Remove population bias using normalization.

Observation: First-class survival rate is highest.

Conclusion: Class strongly impacts survival probability.

Reasoning: Percentage-based analysis reveals true risk.

Note: This view is more reliable than absolute counts.

Survival by Gender (Count)

Aim: Compare male and female survival using counts.

Observation: Male deaths far exceed female deaths.

Conclusion: Females had a survival advantage.

Reasoning: 'Women first' evacuation protocol.

Note: Counts alone do not indicate survival probability.

Survival by Gender (Percentage)

Aim: Compare survival probability across genders.

Observation: Female survival rate is dramatically higher.

Conclusion: Gender is a major survival factor.

Reasoning: Social norms and enforced evacuation priority.

Note: This plot best represents gender-based survival differences.