

Summary: Titanic Dataset Exploratory Data Analysis (EDA)

🔍 1. Dataset Overview & Preparation

- The dataset contains **891 rows** and **12 columns**.
 - Major missing values were found in:
 - Age (~19.9% missing)
 - Cabin (over 77% missing)
 - Embarked (0.2% missing)
 - **Missing Age values** were imputed with the mean age (~30 years).
 - The target variable `Survived` was converted from numerical (0/1) to categorical (No/Yes) to improve interpretability in charts.
-

📊 2. Key Findings & Visual Insights

📦 Survival Breakdown

- **61.6% (549 passengers) did not survive**
- **38.4% (342 passengers) survived**

📝 *Insight:* More than 3 out of every 5 passengers died, indicating a high mortality rate.

📦 Gender Distribution & Survival

- **65% of passengers were male, 35% were female**
- **74.2% of women survived, whereas only 18.9% of men survived**

📝 *Insight:* Females were prioritized during evacuation. Women had nearly **4x higher survival rate** than men.

📦 Port of Embarkation

- Most passengers boarded from:
 - **Southampton (S): 72.4%**
 - **Cherbourg (C): 18.9%**
 - **Queenstown (Q): 8.5%**

□ *Insight:* Southampton was the primary embarkation port. Passengers from Cherbourg had a higher survival rate, likely due to more 1st class travellers.

□ Passenger Class (Socioeconomic Status)

- Class distribution:
 - **3rd Class: 55.1%**
 - **1st Class: 24.2%**
 - **2nd Class: 20.6%**
- Survival by class:
 - **1st Class: 62.96% survived**
 - **2nd Class: 47.28% survived**
 - **3rd Class: 24.24% survived**

□ *Insight:* Higher class meant higher survival. **First-class passengers were 2.5x more likely to survive than third-class passengers.**

□ Age Distribution & Survival

- Majority of passengers were in the **20–30 age bracket**, which also had the **highest number of deaths**.
- Visual histograms showed that:
 - Children had **higher survival rates**
 - Young adults (especially in 3rd class) had **lower survival**

□ *Insight:* Younger children and upper-class adults were prioritized; middle-aged adults had the worst odds.

□ Survival by Gender & Class

- Among **1st class females**, survival rate exceeded **95%**
- Among **3rd class males**, survival dropped below **15%**

□ *Insight:* The worst affected group was **3rd class males**, likely due to location (lower decks) and evacuation priorities.

- **Heatmap of survival by ticket class** visually confirmed the high concentration of non-survivors in 3rd class.
- **Combined plots** (hue = "Survived") offered layered insights (e.g., gender + survival, class + survival).

□ Conclusion

This analysis uncovers stark inequalities in survival aboard the Titanic. **Women, children, and upper-class passengers** had significantly higher survival rates. The dataset reflects not only the tragedy of the incident but also the strong influence of social hierarchy and gender roles on survival chances.
