Executive Summary - Titanic Exploratory Data Analysis (EDA)

This analysis explores the Titanic dataset to uncover key insights into passenger survival rates, demographic distributions, and influencing factors. The study involves data cleaning, handling missing values, and performing visualizations to derive meaningful conclusions.

Key Findings and Insights

1. Survival Distribution

- Out of all passengers, **38.38% survived**, while **61.62% did not**.
- **Female passengers had a significantly higher survival rate** (74.2%) compared to males (18.9%), suggesting gender played a crucial role in survival outcomes.
- Visualizations indicate that a higher proportion of first-class passengers survived compared to lower classes.

2. Passenger Demographics & Age Distribution

- The **age distribution ranged from 0.42 to 80 years**, with most passengers between **20-40 years old (50% of total passengers)**.
- The median age was **28 years**.
- **Children (age < 10) had a higher survival rate (~61%)**, indicating priority was given to younger passengers during evacuation.
- **Elderly passengers (age > 60) had a much lower survival rate (~22%)**.

3. Impact of Passenger Class on Survival

- **First-class passengers had a survival rate of 62.96%**, significantly higher than **second-class (47.28%)** and **third-class (24.24%)**.
- **Over 75% of third-class passengers perished**, indicating that socio-economic status played a key role in survival.

4. Fare Analysis & Its Influence on Survival

- The median fare paid was **£14.45**, but the highest fares were paid by first-class passengers.

- Passengers who survived had a **higher average fare (£48.40)** compared to those who didn't survive (£13.67), indicating that passengers paying higher fares had better access to lifeboats or safer areas of the ship.

5. Missing Data Handling & Cleanup

- **The "deck" column was removed** as it had over **77% missing values**.

- **Age column had ~19.87% missing values**, which were handled using visualization techniques.

- A **heatmap was used to identify missing values** in other columns, ensuring data consistency.

6. Visual Insights from EDA

- **Survival vs. Age:** A box plot showed that younger passengers had a higher survival probability.

- **Fare vs. Survival:** Higher fare-paying passengers had better survival chances, indicating a class-based advantage.

- **Gender-based analysis:** Count plots confirmed that women had a significantly higher survival rate.

- **Distribution of Data:** Histograms and KDE plots highlighted the passenger demographics and survival patterns.

Conclusion

The analysis confirms that **gender, class, and fare price were the most influential factors in survival**. Female passengers, first-class ticket holders, and those who paid higher fares had significantly better survival odds. The data suggests that social hierarchy played a crucial role during the evacuation process.

Future work could involve implementing machine learning models to predict survival chances based on multiple features, further refining our understanding of the dataset.