Data_A1_Solution

NOTE:- تم الأستعانة ب chat gpt في كتابة ال Report و تعديل و لكن تم قراءة و فهمه كله

1. Summary of Findings

- **Passenger Class**: First-class passengers had significantly higher survival rates than those in second and third class.
- **Gender**: Females were more likely to survive than males.
- **Age Groups**: Children (ages 0–15) had the highest survival rates, while middle-aged and senior passengers were less likely to survive.
- **Family Size**: Small families (1–3 members) had better chances of survival compared to those traveling alone or in large groups.
- **Embarkation Point**: Passengers boarding from Cherbourg (C) had slightly better survival odds than those from Southampton (S) or Queenstown (Q).

---

2. Key Insights

- **Survival Priority**: Women and children, especially in higher classes, were given priority during evacuation.
- **Deck Assignments**: Passengers with cabin assignments (decks A–C) were more likely to survive than those with no recorded cabin (lower decks or steerage).
- **Fare Correlation**: Higher fares, indicating wealth and first-class status, were associated with higher survival chances.

---

3. Methodology Explanation

1. **Data Cleaning**:
   - Handled missing values:
     - Filled missing `Age` using median values based on `Pclass`.
     - Categorized missing `Cabin` data as `Unknown`.
     - Filled missing `Embarked` values with the most frequent value (`S`).
2. **Feature Engineering**:
   - Created new features:
     - **FamilyMembers**: Combined `SibSp` and `Parch`.
     - **Title**: Extracted titles (e.g., Mr., Mrs., Miss) from names.
     - **Age Groups**: Categorized passengers into age bins.
     - **FareRange**: Grouped fares into bins based on quartiles.
     - **Deck**: Extracted the first letter of the `Cabin` field.
3. **Visualization**:
   - Used **matplotlib** and **seaborn** to plot relationships between features and survival outcomes.
4. **Analysis**:
   - Explored correlations, survival rates by key features, and overall patterns.

---

4. Visualizations

1. **Survival by Passenger Class**:
    - First-class passengers had the highest survival rates, while third-class passengers had the lowest.
2. **Gender Distribution of Survival**:
    - A significant survival advantage for females.
3. **Age Group Survival Rates**:
    - Children had the highest survival rate among all age groups.
4. **Fare Distribution and Survival**:
    - Passengers who paid higher fares were more likely to survive.
5. **Deck Distribution and Survival**:
    - Passengers on upper decks (A–C) had better survival outcomes.

---

5. Conclusions

The survival during the Titanic disaster was significantly influenced by socioeconomic status, gender, and age. Wealthier passengers in higher classes, women, and children were prioritized during evacuation. Visualizing the dataset emphasized stark differences in survival based on class, gender, and deck location. These findings reflect both the social norms of the early 20th century and the practical challenges faced during the evacuation.

Further analysis, such as using machine learning models, could enhance predictive insights into survival factors.