Data_A2_Solution

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

1. Summary of Findings

- **Passenger Class (Pclass): There is a significant association between Pclass and Survival. Passengers in higher classes (1st and 2nd) had significantly higher survival rates than those in the 3rd class.**
- **Gender (Sex): Women had higher survival rates than men. Females had a much higher chance of survival, which aligns with the historical narrative of prioritizing women and children during evacuation.**
- **Family Size: Passengers traveling with family members (larger family size) had higher survival rates. Traveling alone reduced survival chances.**
- **Fare: There was a weak positive correlation between Fare and Survival, suggesting that passengers who paid higher fares were slightly more likely to survive.**

2. Key Insights

- **Passenger Class: The survival rate is much higher for 1st class passengers compared to 3rd class passengers. This indicates that class was a critical determinant in the chances of survival, likely due to proximity to lifeboats and prioritization.**
- **Gender: Women had significantly higher survival rates. This reflects the priority given to women and children during the evacuation process, with a higher percentage of women saved compared to men.**
- **Family Size: Traveling alone was associated with a lower probability of survival, as passengers traveling with family members were more likely to be prioritized and helped to the lifeboats. Larger families had a better chance of survival.**
- **Fare: Although weak, the positive correlation between Fare and Survival suggests that passengers who could afford higher fares (often associated with higher-class cabins) were more likely to survive.**

3. Methodology Explanation

1. **Data Preprocessing:**
    - **Missing values were handled by filling categorical variables with the most frequent category and numerical values with the mean.**
    - **Categorical variables like Sex and Pclass were encoded using Label Encoding to convert them into numerical form for analysis.**

2. **Statistical Analysis:**
    - **Pearson Correlation was used to measure the relationship between Fare and Survival.**
    - **Chi-Square Test for independence was applied between Pclass and Survival to determine if the passenger class influenced survival.**
3. **Visualization:**
    - **Several plots were created to visually explore the relationships between survival and variables like Pclass, Gender, Fare, Family Size, and Age.**
    - **The main visualizations included:**
        - **Bar plots for survival rates by Pclass and Gender.**
        - **Age distribution of survivors vs non-survivors.**
        - **Family size impact on survival.**
        - **Fare distribution analysis to assess if wealthier passengers had better survival rates.**

4. Visualizations

1. **Survival Distribution by Passenger Class and Gender:**
    - **Bar Plot comparing survival rates across different classes and gender.**
2. **Age Distribution of Survivors vs Non-Survivors:**
    - **Histogram showing the distribution of ages for survivors vs non-survivors, revealing that younger passengers had higher survival rates.**
3. **Family Size Impact on Survival:**
    - **Bar Plot showing the survival rate of passengers based on family size (whether they were traveling alone or with family).**
4. **Fare Distribution:**
    - **Box Plot showing the fare distribution among survivors and non-survivors, indicating that higher fares were associated with higher survival rates.**

5. Conclusion

The analysis of the Titanic dataset has uncovered several key insights into the factors that influenced survival:

- **Pclass was a strong determinant, with passengers in higher classes having better survival chances.**
- **Gender played a significant role, with women having higher survival rates.**

- **Family size impacted survival, with passengers traveling alone having lower survival chances.**
- **Fare had a weak but positive correlation with survival, suggesting wealthier passengers had slightly higher survival chances.**