Titanic Data Exploration

Dataset Overview

1. How many passengers are in the dataset? What columns are available for analysis?
2. Count passengers by **Sex**, **Pclass**, and **Embarked**.
3. Describe the distribution of **Age** (mean, median, min, max, missing values).
4. Describe the distribution of **Fare** (mean, median, min, max, outliers).

Survival Patterns

1. What is the **overall survival rate**?
2. Compare the survival rate of **male vs female passengers**.
3. Compare the survival rate across passenger classes (**Pclass = 1, 2, 3**).
4. Compare the survival rate across embarkation ports (**Embarked = C, Q, S**).
5. Create an **AgeGroup** column with categories:
    a. Child (Age ≤ 14), Adult (15–59), Senior (≥ 60).
    b. Compare survival rates across these groups.
6. Compare survival by the combination of **Pclass × Sex** (e.g., 1st-class women vs 3rd-class men).

Family and Companionship

11. Compare survival rates for passengers with **SibSp > 0** vs **SibSp = 0**.
12. Compare survival rates for passengers with **Parch > 0** vs **Parch = 0**.
13. Create a new column **FamilySize = SibSp + Parch + 1**.
    Compare survival rates for FamilySize categories:
    a. 1 (traveling alone),
    b. 2–4 (small family),
    c. 5+ (large family).

Socio-Economic Factors

14. Compare the **average fare** across classes (**Pclass = 1, 2, 3**).
15. Categorize passengers into **Fare Groups**:

- Low: $0–$10
- Lower-Mid: $10–$30
- Mid: $30–$60
- Upper-Mid: $60–$100
- High: $100–$200
- Very High: $200+

Compare survival rates across these fare groups.

16. Within each class, check whether passengers who paid higher fares had higher survival odds.

Demographics

17. Compute the **average age** of passengers in each class (Pclass). Compare to the overall mean age.
18. Extract **Title** from the Name column and assign passengers into these groups:

- Commoners: Mr, Mrs, Miss, Ms, Mme, Mlle, Master
- Nobility: Lady, Sir, the Countess, Jonkheer, Don
- Professionals: Dr, Rev, Col, Major, Capt

Compare survival rates across these three title groups.

19. Compare the distribution of **Age** across different classes (Pclass = 1, 2, 3).
20. Compare the distribution of **Age** across embarkation ports (Embarked = C, Q, S).