**DATA ANALYST INTERNSHIP** Exploratory Data Analysis (Task- 5)

# Objective:

The objective of this analysis is to explore the Titanic dataset to identify patterns, relationships, and key factors that influenced passenger survival. This includes examining demographic details, class distribution, fare ranges, and survival rates across various groups.

# Dataset Summary:

The dataset contains passenger information from the Titanic disaster, including attributes such as Passenger Class, Name, Sex, Age, Number of Siblings/Spouses (SibSp), Number of Parents/Children (Parch), Ticket Number, Fare, Cabin, and Port of Embarkation. It also includes the Survival status, which is the target variable for analysis.

# Data Cleaning Steps:

* Checked for missing values in columns such as Age, Cabin, and Embarked.
* Filled missing Age values with the median age.
* Dropped the Cabin column from analysis due to a large number of missing values.
* Filled missing Embarked values with the most common port.
* Ensured data types were appropriate for each column (e.g., Age as numeric, Embarked as categorical).
* Removed any duplicate rows if found.

# Observations:

**1.** **Passenger Class Distribution:** Most passengers traveled in 3rd Class, followed by 1st Class, then 2nd Class. Higher class generally meant wealthier passengers.

**2. Survival Rate by Gender**: Females had a much higher survival rate than males. Many males did not survive, likely due to the 'women and children first' policy.

**3. Age Distribution:** Most passengers were between 20–40 years old, with some infants and elderly passengers onboard.

**4. Survival Rate by Age Group:** Children under 10 years had higher survival rates. Middle-aged passengers had lower chances.

**5. Survival Rate by Passenger Class:** 1st Class passengers had the highest survival rate, while 3rd Class passengers had the lowest.

**6. Fare Distribution:** Most passengers paid fares under 50 units, with some very high fares from 1st Class passengers.

# Summary:

The analysis shows that higher-class passengers had better access to lifeboats and safety measures, leading to higher survival rates. Females had significantly better chances of survival, as evacuation protocols prioritized women and children. Younger children were more likely to survive due to this priority, while middle-aged passengers faced lower chances. Third-class passengers were at the greatest risk, possibly because of their location on the ship and limited access to lifeboats. Passengers who paid higher fares, typically in first class, generally had better survival outcomes. The Cabin column contained many missing values, making it less useful for detailed analysis. Overall, social class, gender, and age emerged as the most important factors influencing survival in the Titanic disaster.