***Rhombix Technologies***

***Internship Month-2***

***Syed Abrar Ul Haq***

***Task02***

**Steps for Feature Engineering on Titanic Dataset**

1. **Time-Based Features**   
   Although the Titanic dataset doesn't have a time column, you can simulate time-based features:
   * **Passenger Age Groups**: Group ages into bins (e.g., child, young adult, senior).
   * **Embarkation Duration**: Use Embarked data and estimate approximate time between embarkation points.
2. **Rolling Statistics**  
   For a dataset like Titanic, rolling statistics aren't directly applicable. Instead, we can apply statistical summaries to groups:
   * Compute the mean survival rate for each Pclass or Embarked category.
   * Calculate median age and fare per Pclass.
3. **Additional Transformations**
   * **Family Size**: Create a feature by adding SibSp (siblings/spouse aboard) and Parch (parents/children aboard) + 1 (the individual passenger).
   * **IsAlone**: A binary feature indicating whether the passenger traveled alone (FamilySize == 1).
   * **Fare per Person**: Divide Fare by FamilySize to get fare allocation per individual.
   * **Title Extraction**: Extract titles (Mr., Mrs., Miss, etc.) from passenger names (if available).

**Explanation of Features Created**

1. **FamilySize**: Reflects how many family members traveled together.
2. **IsAlone**: Indicates solitary travelers.
3. **FarePerPerson**: Helps adjust for large families.
4. **AgeGroup**: Groups passengers by age, capturing different survival probabilities.
5. **Title**: Adds socio-economic context (e.g., Mr., Miss, Dr.).
6. **Pclass\_Survival\_Rate**: Captures survival likelihood based on class.