**TASK1-Explanatory data analysis**

**Titanic dataset**

It contains data about passengers on the Titanic, with features like age, sex, class, and whether they survived.

# Columns in the Titanic Dataset:

1. **PassengerId**: A unique identifier for each passenger.
2. **Survived:** Indicates whether the passenger survived.It has 2 variables:1,2
3. **Pclass**(**Passengers Class**):The passenger's class. It has three possible values: 1,2,3
4. **Name**: The name of the passenger.
5. **Sex**: The gender of the passenger.
6. **Age:** The age of the passenger in years.
7. **SibSp**:(**Sibling/Spouse**): The number of siblings or spouses the passenger had aboard the Titanic.
8. **Parch**:(Parent/Child)**:** The number of parents or children the passenger had aboard the Titanic.
9. **Ticket:** The ticket number of the passenger.
10. **Fare**: The fare the passenger paid for the ticket.
11. **Cabin**:The cabin number the passenger was staying in. This column has many missing values.
12. **Embarked**: The port where the passenger embarked. It has three possible values:C,Q,S

# From the dataset above:

1. Read the Csv titanic.csv
2. Check if it is clean or not.
3. Clean the csv file.
   * Clean age column (fill mean of age). Clean cabin column (fill with unknown)
   * Replace the values for the following column eg in survived, replace 1 with the word survived:
4. **Pclass**
5. First class
6. Second Class 3-Third class **Embarked**

C-Cherbourg Q-Queenstorm

S-Southermpton

1. How many people survived? How many are female? How many are male?
2. How many are in firstclass? secondclass? thirdclass?
3. What is the average age?
4. What is the mean fare?
5. Did gender (Sex) have a statistically significant impact on the survival rate (0 means didn’t survive, 1 means survived) of passengers on the Titanic?
6. Did the passenger class (Pclass) have a statistically significant influence on the likelihood of surviving the Titanic disaster?
7. Plot a boxplot to check for outliers in Age and Sex
8. Do a countplot of those who survived based on the gender.

**TASK2-Classification**

**Obesity Dataset**

1. Read obesity classification.csv
2. Perform the Machine Learning pipe cycle (get data, data cleaning, data enriching, and data visualization).
3. Classify someone’s label based on their Height, Weight and BMI

**TASK3-Regression**

**Income.csv.**

Read income.csv

1. Perform Data-Cleaning, Preprocessing, Visualizing.

Predict the amount a income one will get is based on their age and experience.

1. How much will I earn as a 30 year old with 10 years’ experience?

**TASK 4-Clustering using KMeans**

**Customerdata.csv**

A Customer Credit Card Information Dataset which can be used for Identifying Loyal Customers, Customer Segmentation, Targeted Marketing and other such use cases in the Marketing Industry.

1. Perform Data-Cleaning, Preprocessing, Visualizing.
2. Read customerdata.csv
3. Cluster customers with the same similarities
4. View cluster 10.How Many are they? How often do they visit?