Aim:

The goal of this assignment is to:

- Preprocess the dataset and prepare it for machine learning using **numpy** and **pandas**.
- Implement linear regression from scratch using **numpy** and **pandas**.
- Evaluate the performance of the models and interpret the results using metrics such as mean squared error (MSE) and R-squared (coefficient of determination) on the test set.

Preprocessing / Data Loading

The few datasets require some preprocessing before it can be used for machine learning. Specifically, you will need to:

- Load the dataset into a panda DataFrame.
- Handle any missing values in the dataset.
- Remove any outliers from the dataset.
- Normalize (Regularize) the data.
- Split the dataset into a training set and a test set.

Linear Regression

Once you have preprocessed the dataset, you will implement linear regression from scratch using **numpy** and **pandas**. Specifically, you will:

- Implement the least-squares regression line using **numpy** and **pandas**.
- Train the model on the training set of the raw and the preprocessed dataset.
- Evaluate the performance of the model using metrics such as mean squared error (MSE) and R-squared on the test set.

Problem 1. Pizza Franchise (Dataset 1)

In the following data

- X = annual franchise fee ($1000)
- Y = start-up cost ($1000) for a pizza franchise

Problem 2. National Unemployment Male Vs. Female (Dataset 2)

In the following data pairs

- X = national unemployment rate for adult males
- Y = national unemployment rate for adult females

Problem 3. Fire and Theft in Chicago (Dataset 3)

In the following data pairs

- X = fires per 1000 housing units
- Y = thefts per 1000 population within the same Zip code in the Chicago metro area

Note: You should not use external libraries such as scikit-learn for the implementation of linear regression