 # Homework-2

The ABC Employee Salary and Expenditure

Introduction:

This assignment is a part of the Data Science Machine Learning Course projects can be found at the main GitHub repo in tile ABC Employee Salary Expenditure

Project Overview

The ABC Employee Salary and Expenditure dataset contains six hindered rows and six columns. It is out sourced from "Kaggle: Your Machine Learning and Data Science Community Website" , I found the dataset interesting to apply the unsupervised machine learning algorithm particularly K Means algorithm. to study the pattern of the individual salaries and their respective expenditure habits.

Project Intro and Objective:

The goal of the project is to discover the pattern of the Employee Income and Spending habit using unsupervised learning clustering K means algorithm **.**

Methods Used

Based on the study objective, dataset from the Kaggle has been retrieved and preprocessed in order to In order to apply K Means Clustering algorithm.

Technologies

- **Python**
- **Jupiter Notebook /Panda**
- **Scikit-learn library**

In this project K Means was used to study to classify and form clusters . to this effect, Elbow method was used to find out the size of "K" equal to 5 to place similar cluster together and data points in the different clusters are farther apart.

Needs of this project:

- **data exploration/descriptive statistics**
- **data processing/cleaning**
- **statistical modeling**

Conclusion:

I found the K means algorithm more advantageous compare to the K Nearest Neighbors which I uses for classification of data . I found the K means algorithm It is the fastest and most efficient algorithm to categorize data points into groups. It is also simple to implement, scales to large data sets and haven't gotten problems to study the spending habit of the employees . However, its performance is usually not as competitive as those of the other sophisticated clustering techniques because slight variations in the data could lead to high variance**.**