Project: Sales Prediction

Group Members:

- Arooba Minhas (21K-3094)
- Bushra Khan (21K-3081)

Summary:

Objectives:

- Cleaning Data.
- Performing EDA.
- Linear Regression.
- K-Means Algorithm.

Overview:

We have two data sets in which we have performed data analysis and according to the data sets it will predict the change of quantity in different branches.

Before analysis of data we have performed data cleaning which include as following:

Firstly we have divided our data set in two data sets with 70:30:

1. Train Data
2. Test Data

Cleaning Operations:

- Branch count.
- Product line count.
- Mode of Branch.
- Mode of Product line.
- Replacing the categorical and numeric columns of train and test data.
- Product line counts after replacement in train and test data.

Performing EDA:

We have done data visualization on different attributes which are shown below in different form of graphs (bar graph, pie chart, and box plot) on both test and train data.

(X-axis , Y-axis)

1. (Product line, No of Transactions).
2. (Gender, No of Transactions).
3. (Members, No Customer).
4. (Product Line, Mean Gross Income).
5. (Branch, Mean Gross Income).
6. (Percentage of payment methods (E -Wallet, Cash, Credit-Card)).
7. (Product Line, Female).
8. (Product Line, Male).
9. (Product Line, Ratios) (Composition of all branches).

Linear Regression:

- We are calculating the mean absolute error, mean squared error, linear regression score, and average linear regression score.
- And by using linear regression we are predicting the quantity on the basis of rating in the branches.

K-means Algorithm:

- K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems.
- Here we use the elbow method to find the K in K-means algorithm, and maintain our clustering on given data set.