Predictive Analytics

Ankit Kumar Singh,
ankit.kumarsingh20@ifheindia.org

Model building parameters:

- Scatter plots were made for Salary vs Age, Salary vs Account Balance and Salary vs Transaction Amount.

- Filtering the type of transaction in the transaction description column, The amount the person has spent on an average in POS, Interbank, Phone Bank.

- Phone bank and bank transfer do not have adequate data for all customers. Data is not considered for predictive analysis

- Created 2 new features for the data. The average payments made by customers and the average money spent by customers at point of sale.

Supervised Model:

Linear Regression Model:

```
#Defining the X and y of machine learning

X=df[["age","balance","payment","pos"]].values
y=df["salary"].values
```

```
#r2 score
print('Coefficient of determination: ', r2_score(y_test, y_pred))

Coefficient of determination:  0.07666219584715894
```

- Took these columns as X and y. The r2 score was low, main reason being lack of data, hence less training examples.
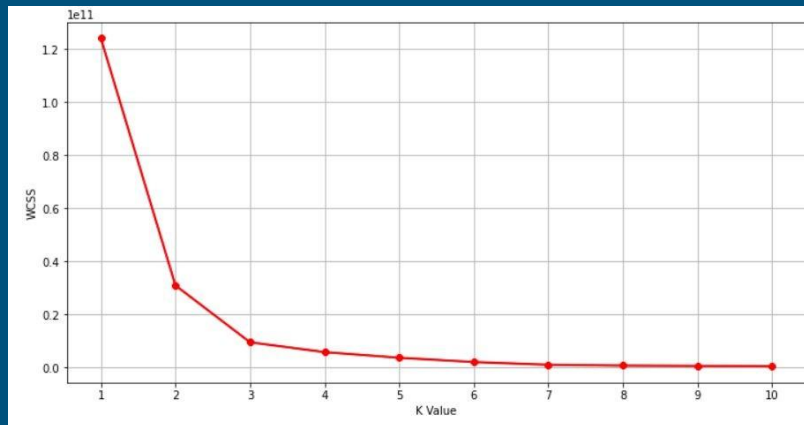
Decision Tree Regressor:

```
#score using test values
dt.score(X_test, y_test)

0.2041445918495219
```

- This model also gave a very low score.
- But the sample prediction it made was also in in realistic values.

Unsupervised Model:

**Customer Segmentation using
K-Means:**

- The elbow curve of Kmeans shows that the elbow is formed at K=3. So we can make 3 clusters.
- The customer segments have been made. The segments are based on customer age, account balance, average payment and POS transactions and salary.