

Predictive Analytics(ANZ Virtual Experience)

- Scatter plots created to check for any correlation.
- Based on the unique values for the “txn_description” column new dataframes were created.
- Inter bank and phone bank transactions had very less data so they were not considered for model creation.
- 2 new features were created based on the average payment made by customers, and average amount at the point of sale.

Supervised Model

- Columns selected for the model: "age", "balance", "avg_p", "pos".

```
In [70]: print('Coefficient of determination: ', r2_score(y_test, y_predict))
```

```
Coefficient of determination: -0.11569976344132793
```

- Due to lack of sufficient data `r2_score` is very low. But the predicted values are realistic.

```
In [71]: # Sample prediction value
```

```
print("The predicted model gives a salary of:")  
print(lr.predict([[30,2000,55,55]])[0])
```

```
The predicted model gives a salary of:  
2327.3461972901523
```

Decision Tree Regressor

```
In [75]: dt.score(X_test, y_test)
```

```
Out[75]: -1.4951107833508677
```

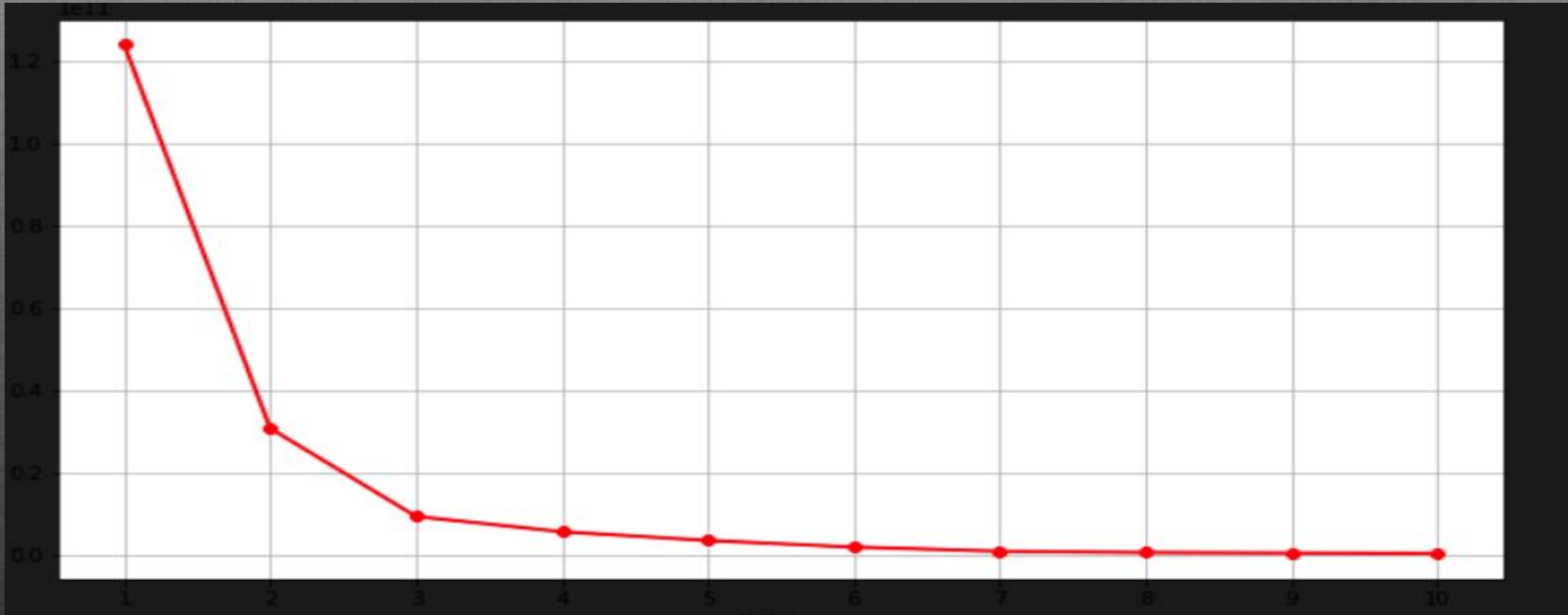
```
In [76]: # Sample prediction values

print("The model predicts salary will be=")
print(dt.predict([[30,2000,55,55]])[0])
```

```
The model predicts salary will be=
1408.08
```

- The score is also low, but the values are more realistic as compared to Linear Regression.

K-Means Clustering(k=3)



- Using the elbow method value of k was determined to be 3. Hence the customers were divided into three segments.