Report 1

Exploratory Data Analysis

- **Kartik Jadav(kartikadav288@gmail.com)**

The Data The dataset contains 12043 transactions for 100 customers. Transaction period is from 01/08/2018 - 31/10/2018 (92 days duration). For each record/row, information is complete for majority of columns. Some columns contain missing data (blank or NaN cells), which is likely due to the nature of transaction. It is also noticed that there is only 91 unique dates in the dataset, suggesting the transaction records for one day are missing (turned out to be 2018-08-16).

The columns are Taken into account:- ['account', 'txn_description', 'merchant code', 'balance', 'gender', 'age', 'merchant state', 'amount', 'customer_id', 'movement']

Key Findings :

• It is better for the any machine learning algorithm if the data has numeric features , but they are few. Useful numeric features include the customer's id, age, transaction amount and bank balance. And some categorical feature is converted into numeric by using get dummies column.

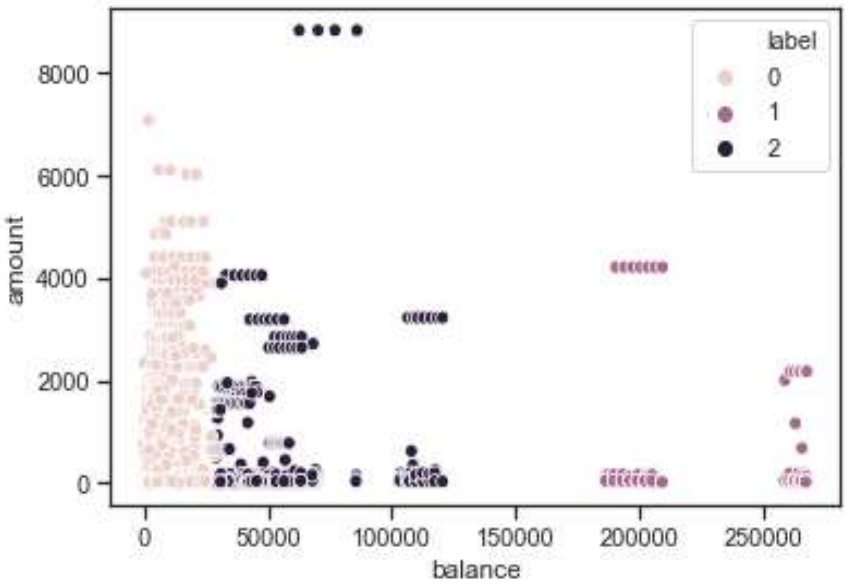
• There is indeed a relation between these features, but the relation is not clear.

• Filtering the type of transaction in the transaction description column, we try to get the amount the person has spent on an average in POS(Point of Sale- Retail, Business, Purchases etc.), Payment(Bills etc.), Interbank(Bank to Bank transfer) and Phone Bank(probably Internet Banking).

• I see phone bank and bank transfer do not have adequate data for all customers. Hence avoid that data.

• The main shortcoming is the lack of data. Perhaps with large amounts of data, more relations can be found.

- Customer Segmentation using KMeans clustering. I shall work with an unsupervised learning algorithm to KMeans.
- The elbow curve of KMeans shows that the elbow is formed at K=3. So we can make 3 clusters. KMeans- The customer segments have been made.
- The segments are based on customer age, account balance, Status, txn_description, movement and merchant state.



- As shown in above figure based on the balance and amount I classified 3 types of customer.

- Please refer to the python notebook for more details. There are correlations between the features. But they are not exactly clear. With more data, we can make more clear. It was a really good experience for me to

understand Exploratory Data Analytics and predictive modeling for customer segmentation in the banking sector.