DSC PROJECT
GROUP 12

Name1 :Nehal Singh Parmar

Roll No:187141

Name2:Shanu kumar

Roll No:187159

Name3:Bhanuprakash kasibisi

Roll No:187128

Name4:Hitender Singh Shekhawat

Roll No. :187122

Credit card customer clustering

→ The number one challenge faced by marketers is to understand who they are selling to.

→ When you know your buyers' personas you can tailor your targeting and offerings to increase their satisfaction and your revenue as a result.

→ When you already have a pool of customers and enough data on them, it can be very useful to segment them.

→Here, we are going to see how we can use clustering to segment some credit card customers.

->The problem described in this dataset requires us to extract segments of customers depending on their behaviour patterns provided in the dataset, to focus marketing strategy of the company on a particular segment.

->DATA:-The credit card data has 17 attributres for each customer which include the balance (credit owed by the customer), cash advance (when a customer withdraws cash using the credit card), the customer’s credit limit, minimum payment, percentage of full payments and tenure.

METRICS

$Accuracy$ = $Samples\ correctly\ classified/Total\ number\ of\ samples$

F_β=(l+β^2)*$precision$*$recall$/((β^2 * $precision$)+$recall$)

- $Accuracy$($precision$)= 0.5554
- F_β=0.8620

Note: β=2,recall=1

DATA PREPROCESSING

- DATA ANALYSIS
- NORMALISATION

IMPLEMENTATION

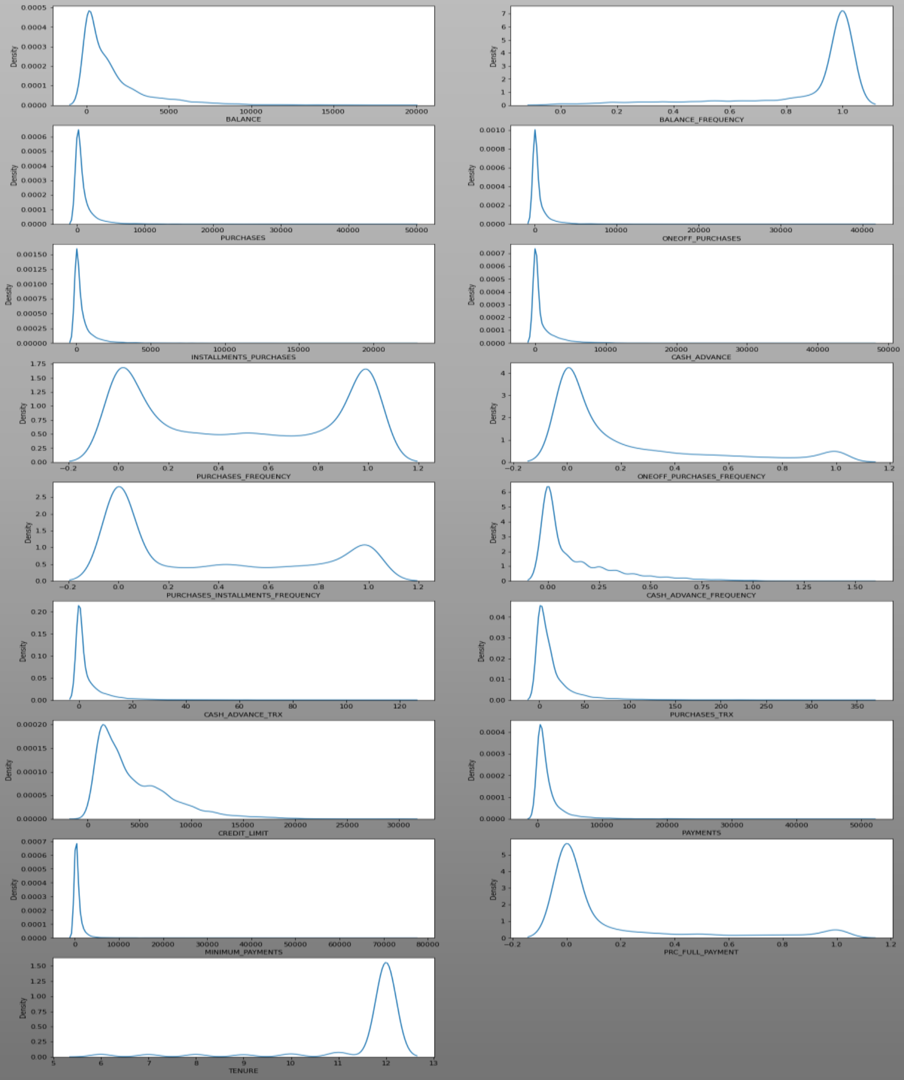
- SPLITTING DATA
- CREATING TRAINING & PREDICTION PIPELINE
- REFINEMENT
- FEATURE IMPORTANCE

STEP WISE APPROACH

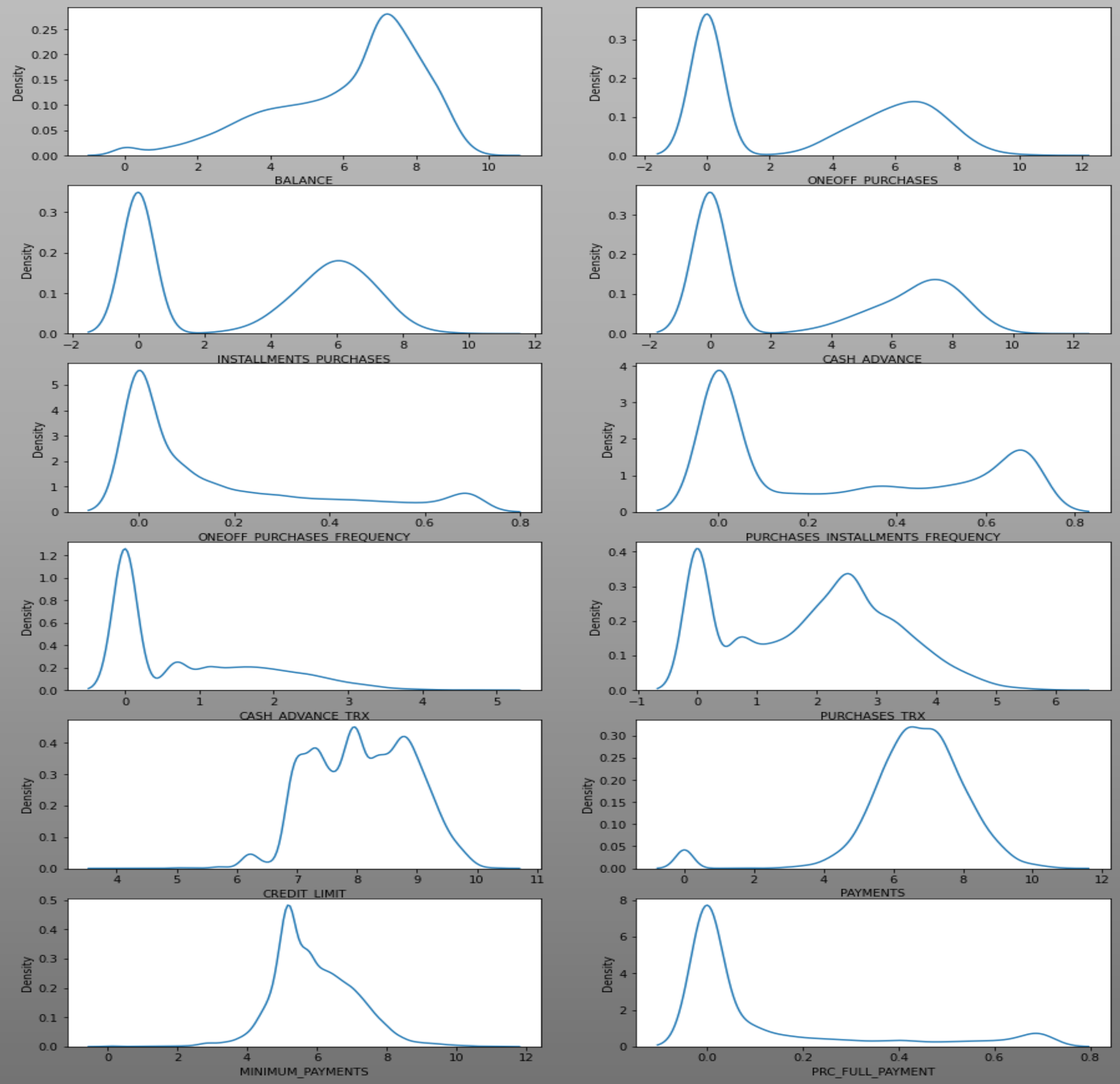
Data taken

```
Data columns (total 18 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   CUST_ID                           8950 non-null   object
 1   BALANCE                           8950 non-null   float64
 2   BALANCE_FREQUENCY                 8950 non-null   float64
 3   PURCHASES                         8950 non-null   float64
 4   ONEOFF_PURCHASES                  8950 non-null   float64
 5   INSTALLMENTS_PURCHASES            8950 non-null   float64
 6   CASH_ADVANCE                      8950 non-null   float64
 7   PURCHASES_FREQUENCY               8950 non-null   float64
 8   ONEOFF_PURCHASES_FREQUENCY        8950 non-null   float64
 9   PURCHASES_INSTALLMENTS_FREQUENCY  8950 non-null   float64
 10  CASH_ADVANCE_FREQUENCY            8950 non-null   float64
 11  CASH_ADVANCE_TRX                  8950 non-null   int64
 12  PURCHASES_TRX                     8950 non-null   int64
```

DATA PROCESSING



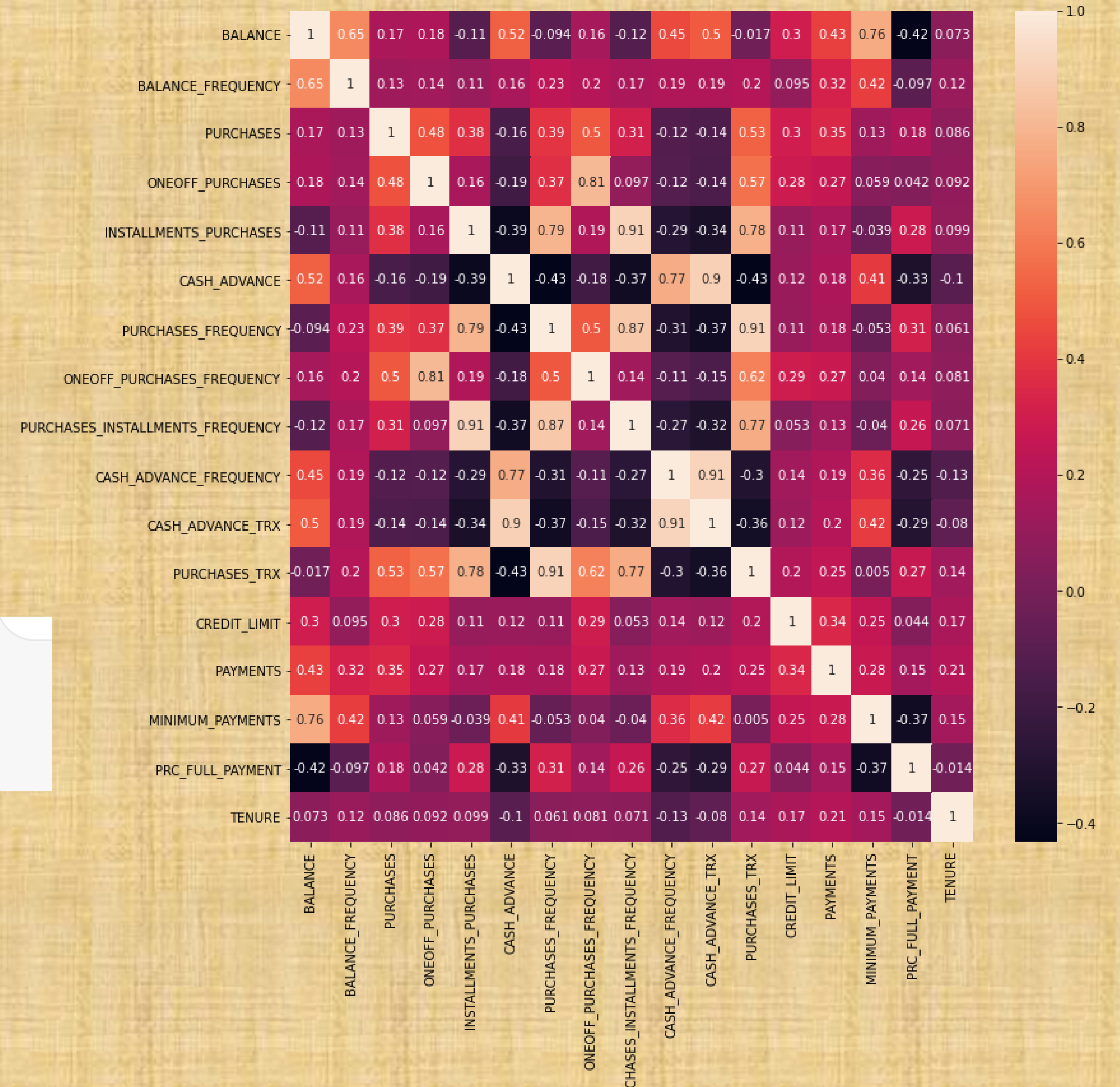
DATASET
SKEWNESS
REDUCTION



CORRELATION IDENTIFICATION AMONG FEATURES

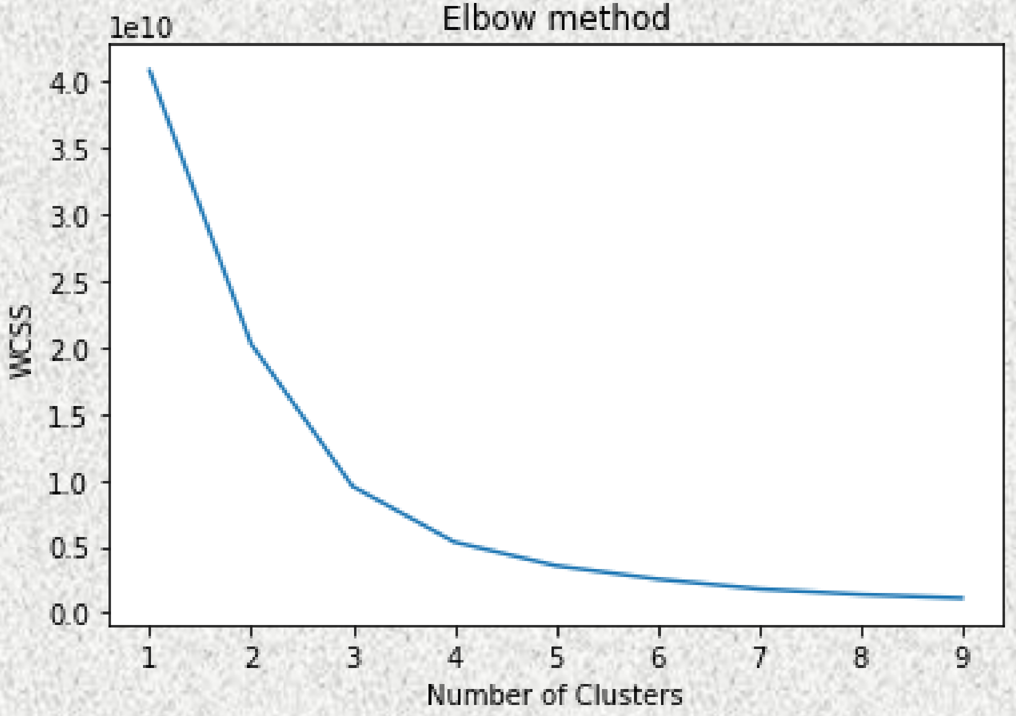
USING PCA FOR DIMENSIONALITY REDUCTION FOR INCREASING NUMBER OF CORRELATED FEATURES

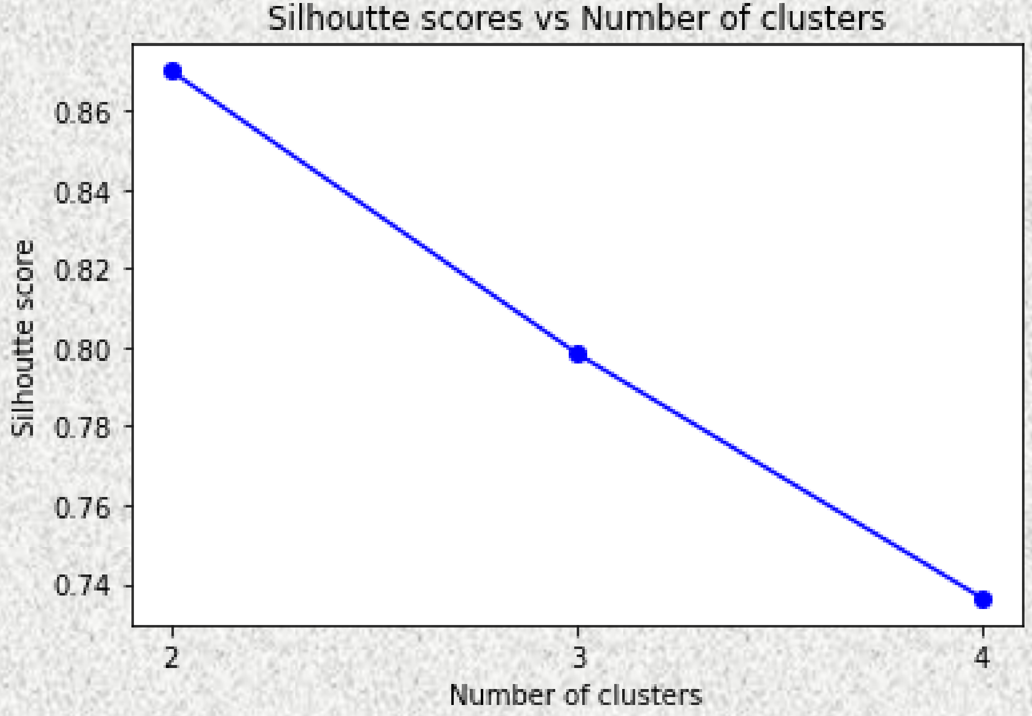
```python
from sklearn.decomposition import PCA

pca = PCA(n_components=0.95)
X_red = pca.fit_transform(df)
```



MODEL TRAINING USING K-MEANS CLUSTERING

DECIDING NO.OF CLUSTERS USING ELBOW METHOD AND SILHOUTTE SCORE





Elbow is around 2,3,4 so using silhouette score for more accurately deciding on no.of clusters

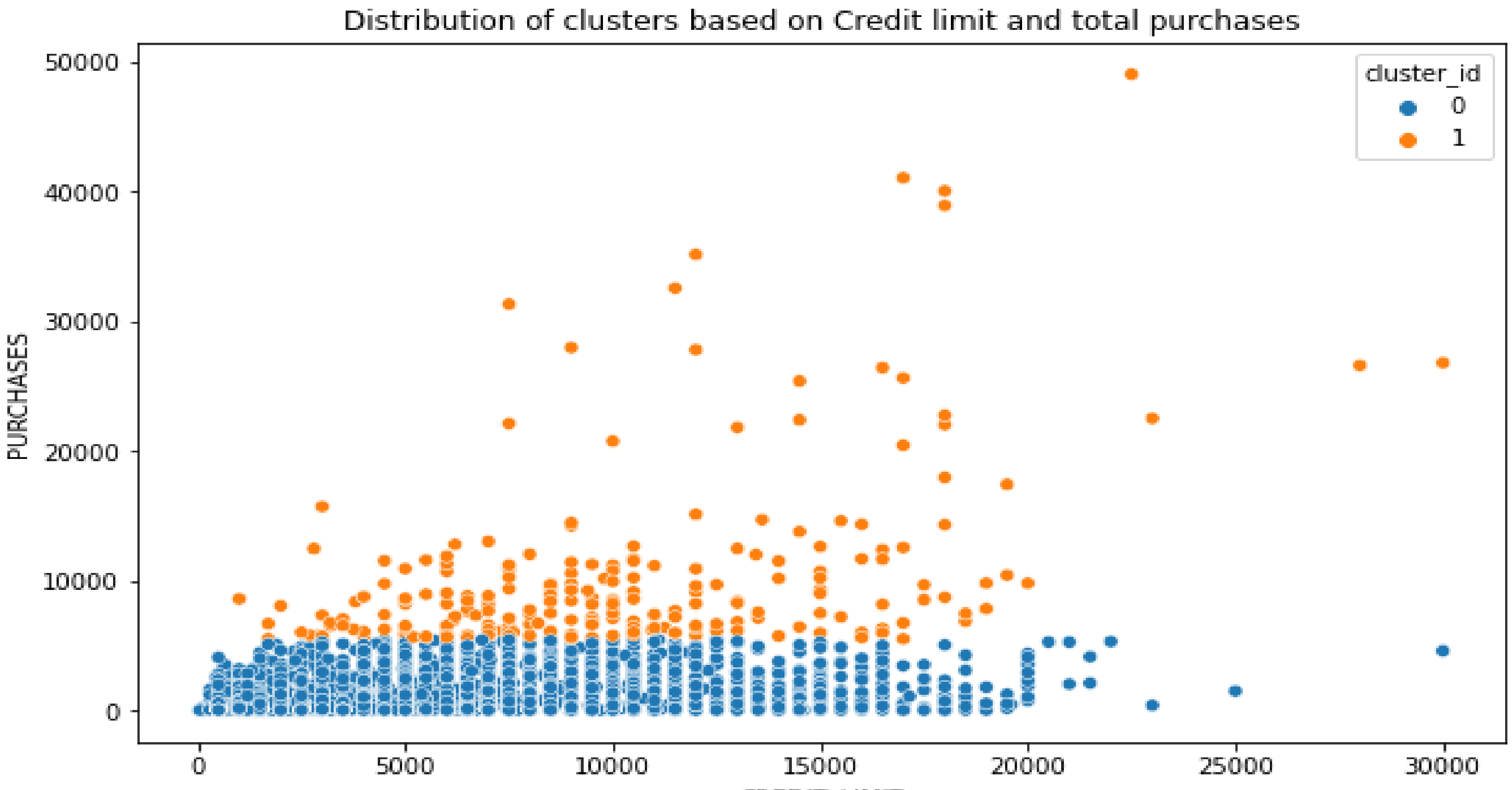
N=2 has high silhouette score than other 2 alternatives

Model Evaluation and Inference.

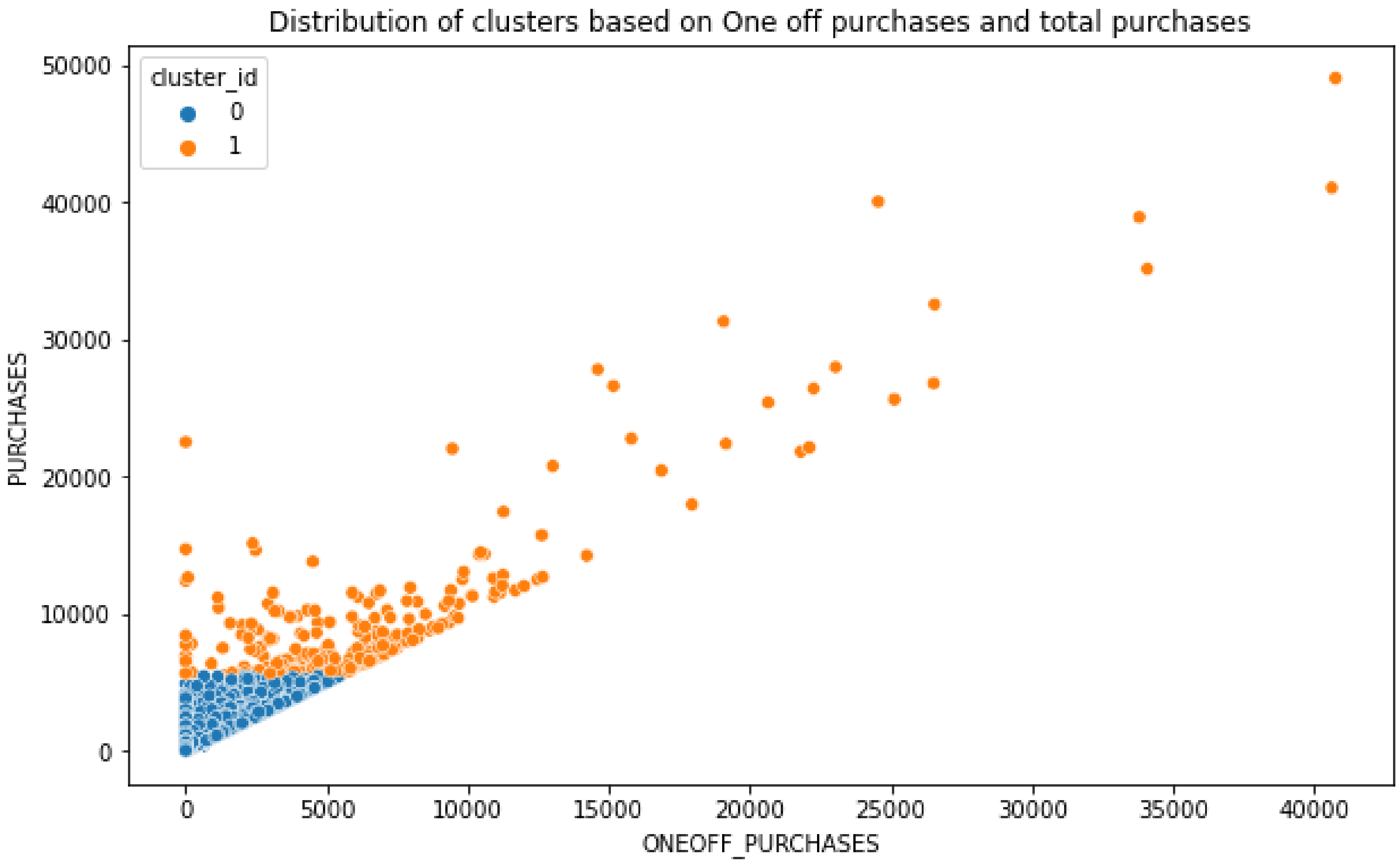
```
from sklearn.metrics import silhouette_score

kmeans = KMeans(n_clusters=2, random_state=23)
kmeans.fit(X_red)

print('Silhoutte score of our model is ' + str(silhouette_scc
d, kmeans.labels_)))
```

```
Silhoutte score of our model is 0.8700455999561516
```

RESULTS



->Looking at the 2 plots it can be concluded that our model has clustered customers with (low usage of credit card in one cluster)and customers with (higher usage) of clusters in other.



CONCLUSIONS

our clustering model will be able to segment our credit card users into distinctive groups. Some of these will be fairly classical such as the prime segment, revolvers and transactors . Understanding the behaviour of customers at this level of granularity is key to tailoring offers which improve customer retention and drive revenues