Bank Customers Clustering

PCA, K-mean, K-medoids, GMM, Hierarchical Clustering

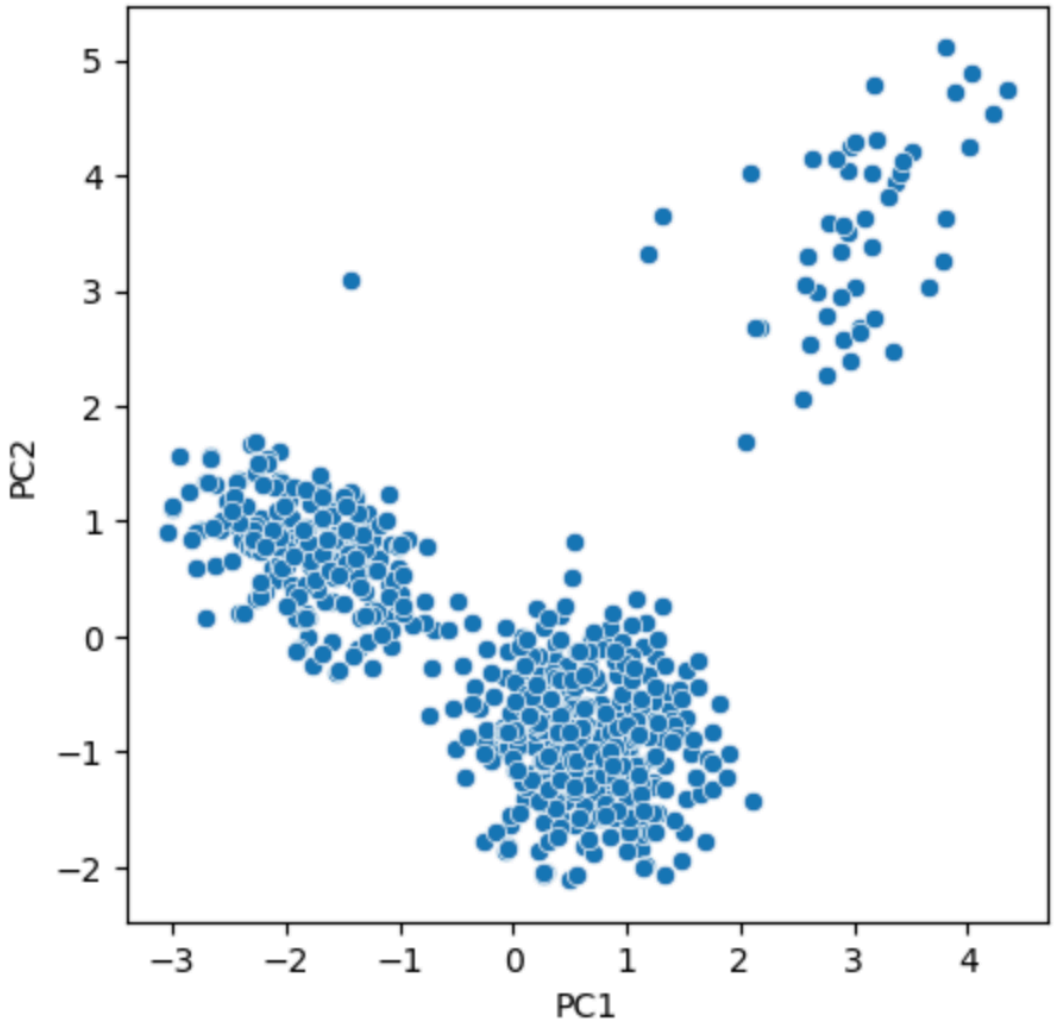
‘Femi Bolarinwa

Data Snapshot

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 660 entries, 0 to 659
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Sl_No                 660 non-null    int64
 1   Customer Key          660 non-null    int64
 2   Avg_Credit_Limit      660 non-null    int64
 3   Total_Credit_Cards    660 non-null    int64
 4   Total_visits_bank     660 non-null    int64
 5   Total_visits_online   660 non-null    int64
 6   Total_calls_made      660 non-null    int64
```

PCA

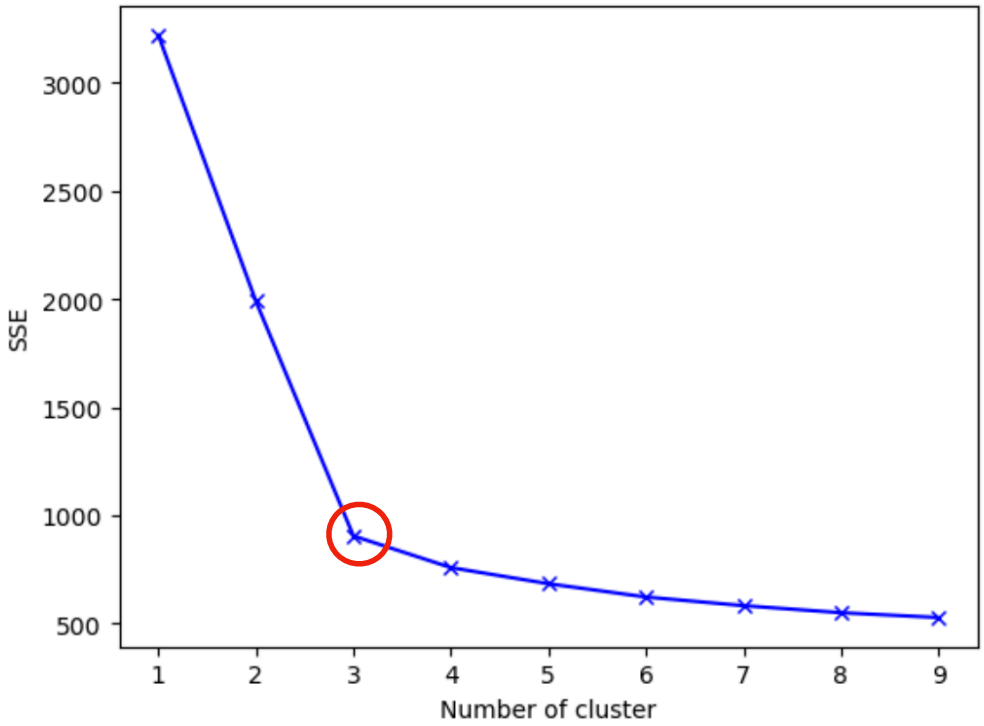
- Principal component analysis helps project the data into a dimension (eigenvectors) that amplifiers or aid clustering.
- Its evident there are 3 clusters of customers.
- But I'll use K-Means' ELBOW method. to automatically detect the optimal number of clusters.



Elbow Method

SSE v N

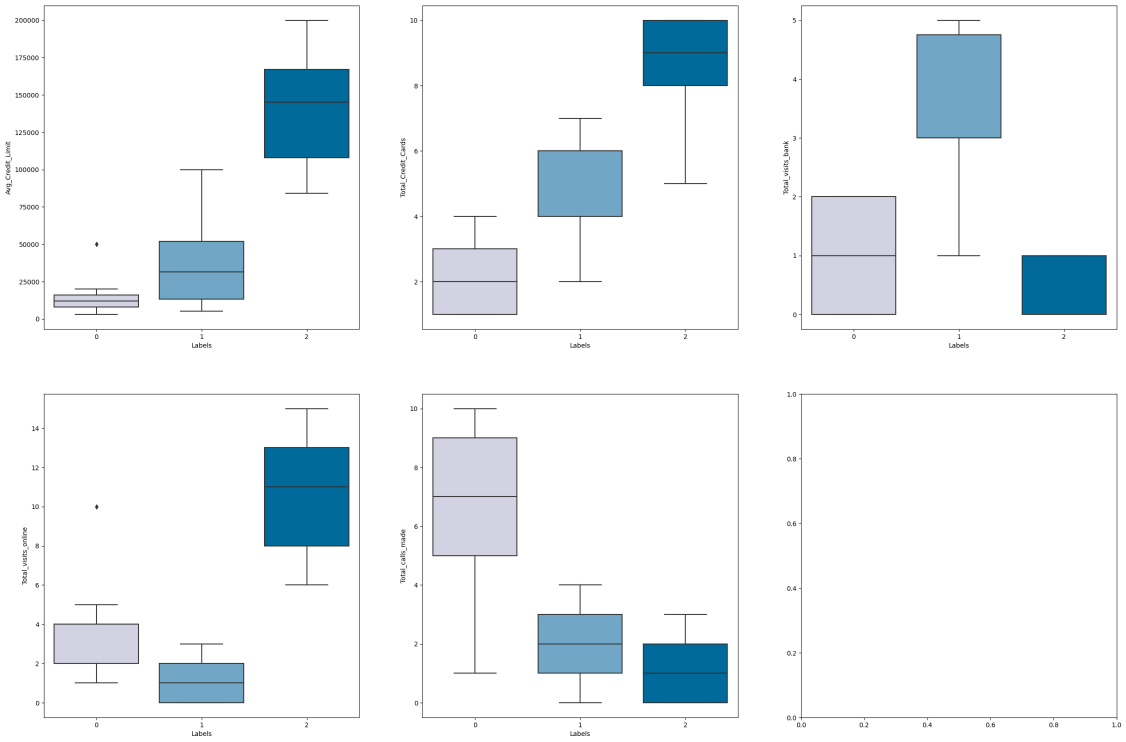
- Plot shows within-cluster sum-of-squares error (SSE) V Number of clusters.
- Optimal number of clusters (3) is obtained at the elbow point of the SSE vs K curve.
- Consistent with visual observation from the principal component analysis above.



K-Means Clustering

Cluster Profiles

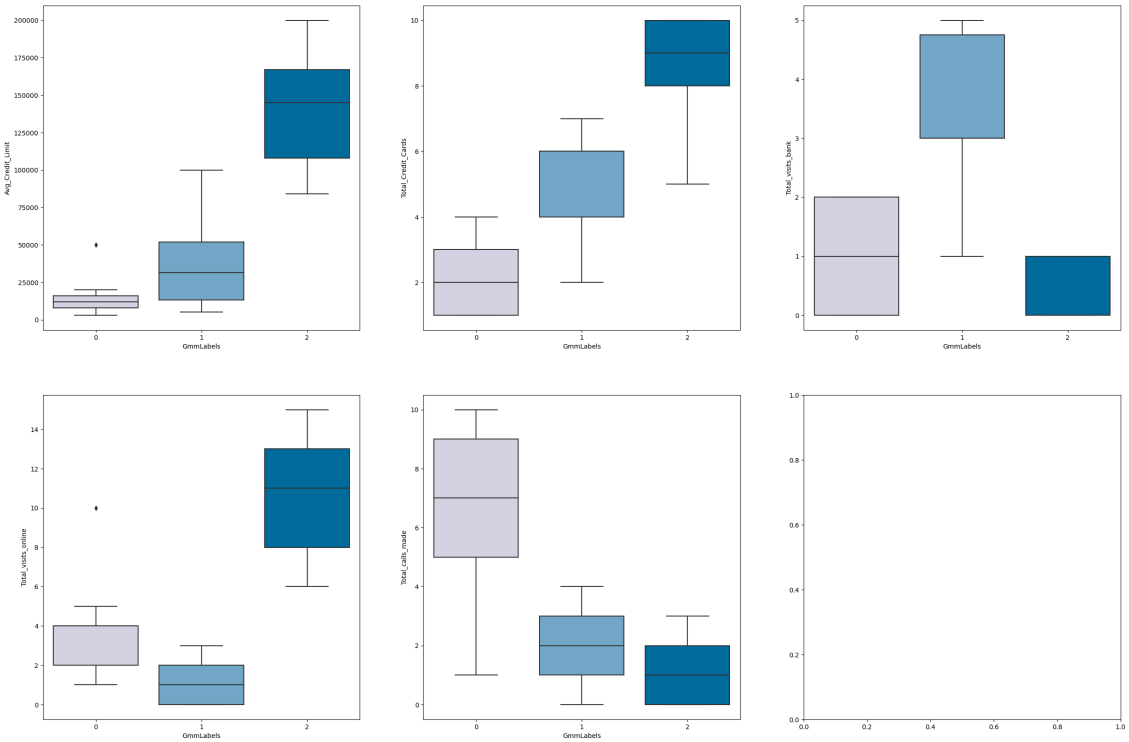
- **Cluster 0**: customers with low credit limit and few credit cards who tend to make a lot of calls and sometimes visit the bank or bank online. This appears to be RECENT personal bankers not so familiar with the bank's tech and are having a lot of issues with there the accounts.
- **Cluster 1**: customers with moderate credit limit and more credit cards who tend to visit the bank a lot more, rarely use online banking and sometimes call the bank. This appears to be SEASONED customers not so tech savvy or enjoy OLD-FASHION banking experience
- **Cluster 2**: customers with high credit limit and lots of credit cards who mostly bank online, rarely visit the bank or call in. This appears to be a group of BUSINESS owners and who prefer ONLINE banking for ease.



GMM Clustering

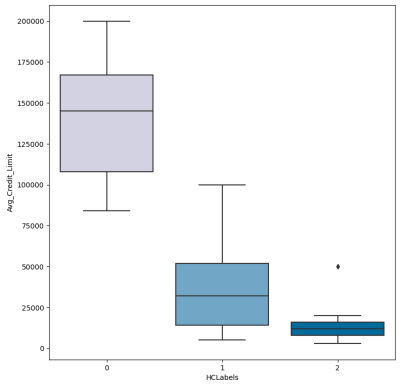
Gaussian Mixture

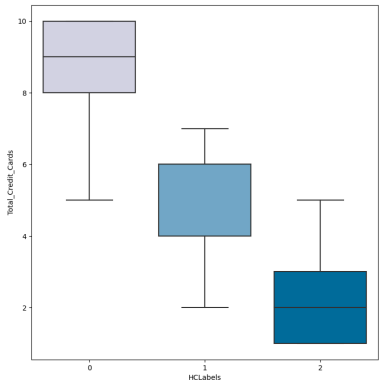
- Identical clustering to K-Means technique

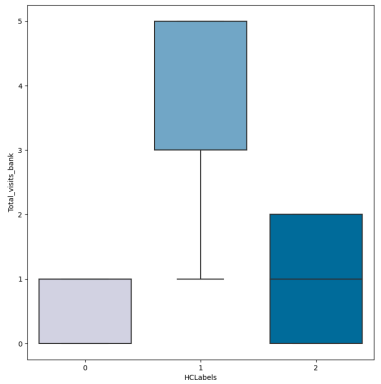


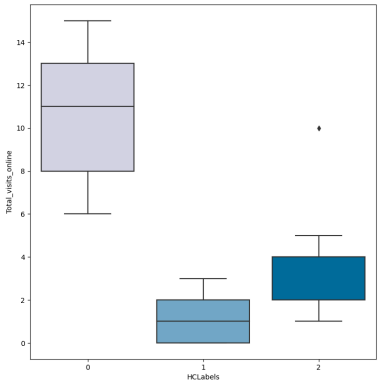
Hierarchical Clustering

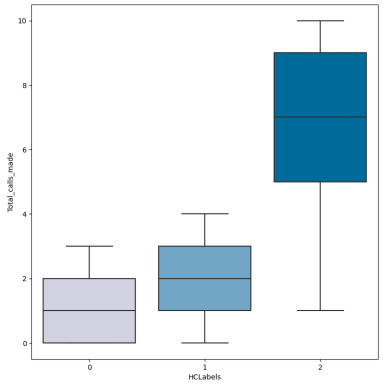
Agglomerative

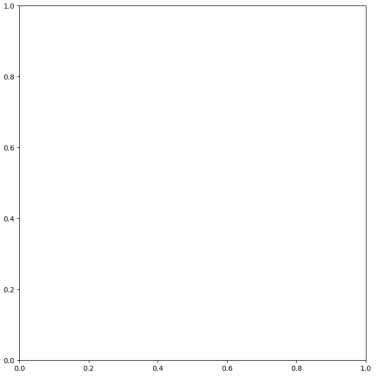












Business Insights and Recommendation

- There are essentially 3 groups of customers in the data. These groups are different in there spending pattern and preferred way of reaching the bank
- There appears to be a group of RECENT customers with low credit limit and fewer cards who appear to be having difficulty using or not inclined to use the online banking. An awareness campaign to show the ease and benefit of online banking is needed. This will help reduce traffic experienced at the call centre.
- A second cluster is a group of SEASONED customers who seem to enjoy the OLD-FASHION banking experience of walking in. The positive experience they have while visiting the bank needs to be sustained. A special campaign to brand online banking could be targeted towards these customers that prefer to walk into the banking hall.
- BUSINESS customers appear to favour online banking experience. This needs to be sustained.