Intuition:

This is a classification problem without ground labels (unsupervised). So the approach was to identify merchants which are similar to each other (using a clustering algorithm – here K means). Ideally, the card transaction price offered to these merchants should be similar. After identifying clusters of similar merchants, if the price offered to them is within 75%-25% of the prices offered to all the merchants in the same group, they will be assigned a label of 'normal' price. On the other hand, if their price is >75 or <25% of the prices offered to the merchants in their group, it will be classified as 'non-competitive' and 'competitive' respectively.

Limitations of this study:

1. Since 'Accepts card payments' was used for clustering, the column 'Current Provider' was not utilized. In a future iteration of this exercise, it as well.
2. Since the MCC code column has a high granularity and the data does not mention what those codes mean (such as whether a particular code refers to the supermarkets and so on) , it did not make sense to use them to interpret the clusters.
3. The column "Annual Card Turnover" is ambiguous because merchants which currently do not accept card payments have a non-zero value for this column. Is it because they used to accept card payments in the past ?

Implementation details

1. Converting "Annual Card Turnover" and "Average Transaction Amount" into log scale

The distribution of these metrics was heavily skewed which was impacting clustering performance (even after scaling) and resulting in clusters which were not meaningful.

This was solved by converting them to logarithmic scale.

2. Aligning both parts of fees

All the 'fee' columns had the following structure – "0.55% + 2p" where the first part is the fee applied as a proportion to each transaction and the second part is the flat fee applied to each transaction.

The flat fee was converted to a percentage:

Flat_fee_percentage = ((flat_fee / 100) / Average_transaction_amount) * 100

And summed together

3. Creating a single fee metric

A single fee metric "Fees (%)" was created from the weighted average of all the aligned fee columns based on the following assumptions:

- 40% of card payments are from Mastercard cards, and 60% are from Visa cards

- 90% of card payments are made using Debit cards, 8% are made using credit cards, and 2% are made using business debit cards

Fees (%) =
(PMD · mastercard_prob · debit_prob)+(PVD · visa_prob · debit_prob)+(PMC · mastercard_prob · credit_prob)+(PVC · visa_prob · credit_prob)+(PMBD · mastercard_prob · business_debit_prob)+(PVBD · visa_prob · business_debit_prob)

Where:

- PMDP_{MD}PMD: Price for "Mastercard Debit"
- PVDP_{VD}PVD: Price for "Visa Debit"
- PMCP_{MC}PMC: Price for "Mastercard Credit"
- PVCP_{VC}PVC: Price for "Visa Credit"
- PMBDP_{MBD}PMBD: Price for "Mastercard Business Debit"
- PVBDP_{VBD}PVBD: Price for "Visa Business Debit"

4. Clustering using K Means

K means clustering was performed on the training dataset (80% samples) with K=7.

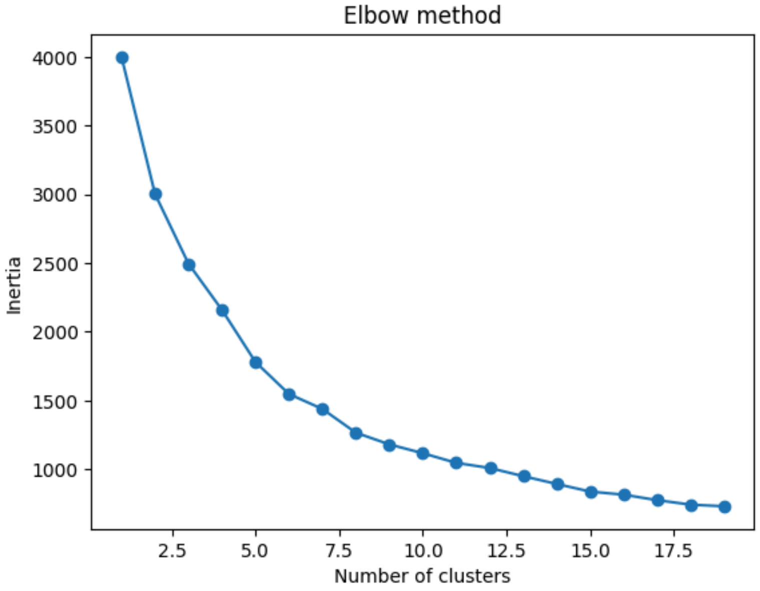

Figure 1 Inertia vs k

The clustering performance was evaluated using Silhouette Score and Davies-Bouldin Index.

A Silhouette score of +1 indicates samples within the same cluster are closer to each other and far away from those in other clusters and a score of -1 represents the opposite end of the spectrum. A score of 0 indicates that some samples lie on the decision boundary between two clusters. The score observed here was 0.33 which shows moderate clustering performance.

The DBI metric measures the ratio of the similarity of each cluster with that of its most similar cluster. A lower score represents a good clustering performance. The score observed for us was 1.18.

The hyperparameter (k=7) was selected based on the elbow-method, if it achieved a reasonable Silhouette Score and DB index score.

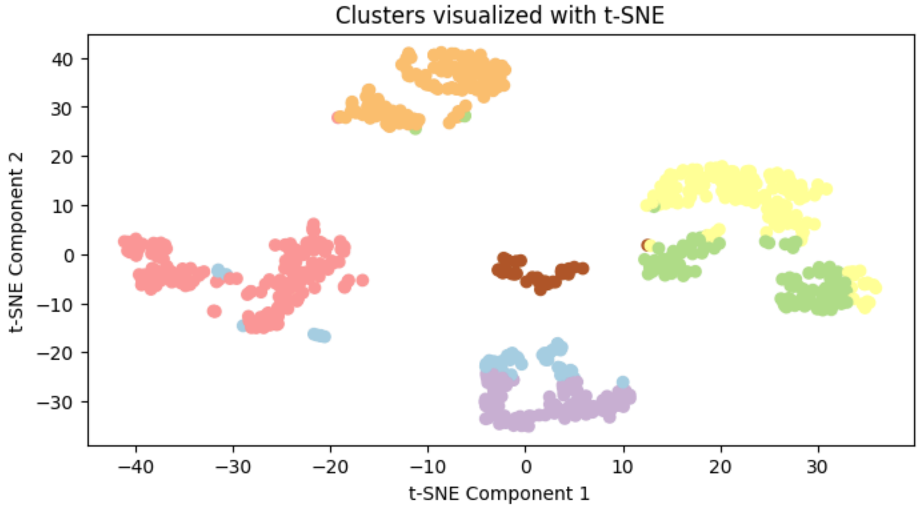

Figure 2 Visualising the clusters using t-SNE

Interpreting Clusters

- Cluster 0
  - Does not accept Card payments
  - Is mostly registered
  - Very high average transaction amount (>£ 1k)
  - –
  - **This is an interesting group because the average transaction amount for this group of merchants is very high, but they do not accept card payments. They are probably high value goods (automobiles) or similar merchants.**
- Cluster 1
  - Accepts card payments
  - Is registered
  - Moderate average transaction amount (> £ 100)
  - Average Annual card turnover (for those merchants which currently accept card payments) < £ 100k (or over 50k)
  - **Companies which are registered and currently accept card payments. The average transaction size is >£ 100 so not groceries but the average annual card turnover is around £ 100k**
- Cluster 2
  - Does not accept card payments
  - Is not registered

    - Low average transaction amount (<£ 100)
    - –
    - **These are companies which are not registered but currently accept card payments and have low ticket size and annual card turnover. So, this group probably represents newly established small stores. It would be interesting to understand why they do not accept card payments yet.**

- Cluster 3
    - Accepts card payments
    - Is not registered
    - Low average transaction amount (<£ 100)
    - Average Annual card turnover (for those merchants which currently accept card payments) < £ 100k
    - **These are companies which are not registered but currently accept card payments and have low ticket size and annual card turnover. So, this group probably represents newly established small stores which have not been making much revenue yet.**

- Cluster 4
    - Does not accept card payments (currently)
    - Is registered
    - Low average transaction amount (<£ 100)
    - –
    - **These merchants are registered, have a moderate/low ticket size and do not currently accept card payments. It might be because the costs of accepting card payments outweigh the additional revenue, they bring in but we have to look deeper into this. We do not have geographical location about these merchants to understand where they are located.**

- Cluster 5
    - Accepts card payments (currently)
    - Is registered
    - Low average transaction amount (<£ 100)
    - Average Annual card turnover (for those merchants which currently accept card payments) over £ 100k
    - **Companies which are registered and currently accept card payments (similar to cluster 1). The average transaction size is <£ 100 so regular stores but the average annual card turnover is over £ 100k. These are the kind of shops where the value of the items are not a lot but people shop more often.**

- Cluster 6
  - May or may not be registered
  - May or may not accept card payments
  - Moderate average transaction amount (> £ 100)
  - Average Annual card turnover (for those merchants which currently accept card payments) < £ 100k
  - **Interesting group because it has a mix of merchants who accept and do not accept card payments and may or may not be registered. Otherwise, like cluster 1.**