Point of Sale Merchant Segmentation in the Payments Industry

submitted in partial fulfillment for the degree of master of science

S J Bulder
12309125

master information studies
information systems
faculty of science
university of amsterdam

2019-07-12

	Internal Supervisor	External Supervisor
Title, Name	Dr Adam Belloum	Kees Loeff
Affiliation	UvA	Adyen
Email	A.S.Z.Belloum@uva.nl	kees.loeff@adyen.com

Universiteit van Amsterdam

ABSTRACT

Payment service providers (PSPs) provide businesses (i.e. merchants) with services that allow them to accept electronic payments from customers (i.e. shoppers). Point of Sale (POS) refers to a physical location (ea. check out areas in stores) where an electronic payment occurs. The current categorization of POS merchants is inadequate and does not allow PSPs and merchants to completely utilize all potential insights. Using payment data aggregated on store level, this research aims to develop a clustering model that categorizes POS merchants, which enables merchants to be benchmarked. This will provide insights for both PSPs and merchants.

1 INTRODUCTION

Businesses are engaged in the activity of buying and selling goods to make money. A retailer does this by selling products to consumers. Products can be consumer goods or services. The sale of products occurs through sales channels, which is a way of selling to consumers. Examples of sales channels are e-commerce, which is sales through digital channels and retail, which is sales through physical locations, such as a shop. A sale ensues by exchanging products for monetary value. A payment system allows businesses to settle financial transactions by transferring monetary value. A payment system uses cash-substitutes to transfer this value. Traditional cash-substitutes are for example cheques, but with the rise computers, electronic payment systems appeared. Electronic payments system include credit and debit cards, electronic funds transfers and internet banking. A Payment Service Provider (PSP) offers businesses services that allow them to accept electronic payments. Within the payment-sector retailers are referred to as 'merchants' and consumers are referred to as 'shoppers'. PSPs have merchants as their direct customers, whereas shoppers are direct customers to the merchants. PSPs typically offer services to merchants that allows them to accept e-commerce, mobile, and point-of-sale payments. E-commerce payments comprise payments on the sale of products through online services or over the Internet. Mobile payments are payments for a product through a portable device, such as a smartphone. A Point of Sale (POS) generally refers to a location where a sale occurs. This is, contrary to e-commerce and mobile, a physical location. The POS for a store commonly means the checkout area where the transaction is completed. A POS payment, therefore, is a payment that occurs at a physical location through a checkout system using, for instance, a debit card.

Payment Service Providers often serve a variety of POS merchants, they can differ in many different aspects: Merchants are active in different industries, different parts of the world, solely sell through e-commerce, etc. To categorise a merchant the payment industry uses a Merchant Category Code (MCC). This is a four-digit number. The MCC is used to classify a merchant based upon the sort of products it provides. MCCs are either allocated based on merchant type (e.g. railways or hotels) or based on merchant name (e.g. 3504 for HILTON HOTELS[1]) [28]. Using MCC classification, Walmart Inc. and Aldi are both MCC '5411 - Supermarkets' and McDonald's and Burger King are MCC '5814 - Fast Food Restaurants'. The major card schemes, of which Visa and MasterCard are the largest, maintain MCC lists. Currently, there are nearly 1000 MCCs. The MCC categorisation enables benchmarking of merchants. Benchmarking is the comparison of a merchant to industry peers. McDonald's, for example, can be compared to other merchants within the MCC '5814 - Fast Food Restaurants'. Benchmarking is beneficial for both PSPs and their merchants; To PSPs, it provides insights on how their different clients are performing. From a merchant's perspective, it allows them to compare their performance to competitors. Performance can be measured in

[1]All companies mentioned in this research are purely used for example purposes and may or may not be related to Adyen.

various ways, depending on its context. A common way to measure merchants' performance in the payment industry is to measure the authorisation rate. This rate is the number of successful electronic transactions of a merchant divided by the total number of electronic transactions.

The problem with MCC categorisation is that it restricts merchants to one MCC, whereas modern day businesses are often involved in many more. Amazon, Inc., for example, started as a book store, which corresponds to MCC '5192 - Books, Periodicals, and Newspapers', but today its core businesses are cloud computing, e-commerce, Artificial Intelligence and computer hardware, and it can be related to over a dozen MCCs. Also, the MCC classification is created from a retail perspective, i.e. merchants are grouped based one the same sort of products. From a payments perspective, many more aspects can be taken into consideration when comparing merchants. The average transaction value (ATV) can be calculated by setting the total value of all transactions as the numerator and the total number of transactions as the denominator. The ATV could be used as an interesting benchmarking performance metric, by comparing merchants with similar ATVs. Another interesting factor could be 'shopper country', the country from where the shopper originates. If, for instance, a certain merchant from the Netherlands has a lot of shoppers from China, it could be interesting to benchmark their performance to other Dutch merchants who have a lot of Chinese shoppers, but do not necessarily sell similar products. From a more technical, POS payments perspective, an interesting dimension is the way shoppers interact with a terminal, this can, for example, be a "contactless" or a "chip" payment.

This research focuses on creating a new segmentation system that will enable POS merchants to be benchmarked against newly created clusters. To accomplish this POS payment data from a payment service provider will be analysed. A relevant algorithm will be constructed that is able to cluster stores based on POS related payment features. One of these features is, for instance, the average transaction value mentioned earlier. The clustering model will generate clusters that will ultimately serve as categories for stores. These categories can be used to benchmark stores. Their performance can be benchmarked against certain peers in their respective clusters. This method will generate valuable insights for payment service providers as well as merchants. This research can be summarised using the following research question:

Research Question Is it possible to develop a Point of Sale merchant business market segmentation using a machine learning algorithm?

The remainder of this paper is structured as follows: the next section 2 discusses related work, which is divided into Market Segmentation 2.1 and Clustering 2.2. After that, the Methodology 3 is being discussed. The Methodology section is divided into Data Description 3.1, Data Preprocessing 3.2, Clustering **??**, Determining k 3.4, Cluster Validation 3.5 and Principal Component Analysis 3.6. In section 4 the results, an evaluation and future work are discussed. Finally, the paper is concluded in section 5.

2 RELATED WORK

2.1 Market Segmentation

The classification individuals is an ancient method, its roots date back to Hippocrates (400 BC), who categorised people on the basis of physical attributes. Market segmentation divides a market into a number of smaller markets based on shared properties [35]. Since the introduction of the concept as a potential method to solve marketing problems, it has gotten significant (scientific) attention [5] [4]. Data-driven market segmentation makes use of statistical methods and can be split into two broad approaches [24]: A priori and a posteriori segmentation. A priori segmentation is conducted when a theoretical framework is generated before the segmentation occurs, contrary to a posteriori segmentation, which is not based on a predefined framework. A posteriori segmentation requires access to rich data sets, having a large number of variables and often uses complex algorithms [10] [36]. Most market segmentation is focused on business to consumer markets, whereas business market segmentation is less used [26]. Segmentation in the B2B context is regarded as equally important to the B2C context [31]. Businesses can be segmented using various metrics, like business size, turnover, location, and other relevant metrics [34]. The most used segmentation metrics in B2B markets are geographics and firmographics [37]. Business market segmentation is useful for guiding strategic and tactical decision-making , especially for implementation, set-up, sales and marketing.

2.2 Clustering

Clustering is an unsupervised learning method, with the aim to classify subsets of objects with each other based on some sort of relationship [19]. Where in supervised learning the model responds to feedback, unsupervised cluster analysis identifies common points in the data and reacts depending on the absence or presence of these common points in every new data part [17]. Along with the increase in information and intersection of subjects the number of tools to perform cluster analysis has increased. Every clustering algorithm has strengths and weaknesses because of the intricacy of modern information [39]. Each method adheres different sets of rules that define a "cluster" and whether data points are similar [29] [9]. Clustering can be divided into Soft clustering, where each data point belongs to a cluster to a certain degree and Hard clustering, where each data point does or does not belong to a cluster. Clustering approaches can be divided into two groups [14] [18]: Hierarchical and partitioning clustering. Hierarchical clustering methods cluster by forming iteratively divided patterns using a top-down or bottom-up approach, it aims to create a hierarchy of clusters [29] [32]. In partitional clustering methods, data points are assigned into k clusters based on optimising some criterion function and without any hierarchical structure [22]. The criterion can be a distance function, used for quantitative data or a similarity function, used for qualitative data [38]. Modern, commonly used clustering algorithms are, among others [39]: K-Means [23], BIRCH [40], CLARA [21], CLARANS [25], Fuzzy C-Means [6], DBSCAN [13], etc. Which algorithm is best is ambiguous and depends on its application [39].

3 METHODOLOGY

3.1 Data Description

The data used for this study consists of Point of Sale payment data from Adyen. The data is a partial set from Adyen's entire payment data: Certain merchants have deliberately been left out of the data set and a random sample has been taken from the remaining data set. All values within the data that can be directly linked to Adyen merchants have been tokenised. These measures ensure no (business) insights can be obtained from Adyen or its merchants through this research. All companies mentioned in this research are purely used for exemplary purposes and may or may not be related to Adyen.

The data set is created from 200 million individual POS payments and contains data from more than 10,000 stores. The features (e.g. countrycode, payment method, etc.) are mixed-type, consisting of both numerical and categorical features. The full feature list can be obtained from A.1. The data set is limited to POS payments only and contains data from 2018. One of the features is a string that identifies the physical store at which the payment transpired. All individual payments have been aggregated to store level. See A3 for an example of three entries in the data set.

3.1.1 Data Enrichment. The main data set has been enriched with a United Nations (UN) data set. The UN data set comprises urban agglomerations in the world with 300,000 inhabitants or more in 2018. The UN defines an urban agglomeration as "the population contained within the contours of a contiguous territory inhabited at urban density levels without regard to administrative boundaries. It usually incorporates the population in a city or town plus that in the suburban areas lying outside of, but being adjacent to, the city boundaries."[12] The data set contains data on 1860 cities in 154 countries. A column containing dummy variables is added to the main data set. If a store from the main data set is located in one of the cities from the UN data set the store receives a 1, if not a 0. The 'city' feature is only used for this purpose and not used as a clustering feature. This allows stores that are present in large urban cities to be identified and will add a useful feature to the data set. Stores in large urban cities are thought considered to have different properties compared to stores that are not by Adyen domain experts. For example, they have more foreign shoppers (e.a. tourists).

3.2 Data Preprocessing

3.2.1 Missing values. Missing values have been identified in the numerical data, these values have been replaced with mean values of their respective features. Missing values have been identified for the categorical data, these have been removed from the data-set.

3.2.2 Categorical Data. The K-Means++ algorithm is only able to process numerical data. Because of this, all categorical data has been transformed into numerical data. The categorical features are: pos_entry_mode, payment_method, fundingsource, mccdescription and countrycode. These features are encoded to numerical values using One Hot encoding. One Hot Encoding splits a categorical data column into n columns, where n is equal to the number of unique values in the column, each new column is labelled as one of the values from the original column. The values are replaced with 1s and 0s. Take for example a column containing four country names: China, France, Germany and Brazil. This column would be split up into four columns. Rows that have the value China will get a 1 in the China column and 0s in the other columns. Similarly, if a row contains the value France, it will get a 1 the France column and 0s in the others, etc.

For each of the newly created columns for pos_entry_mode, payment_method and fundingsource, percentages of the total are then calculated by using the 'transaction_count'. For example, 20% of store X's transactions have payment method 'visa'.

3.2.3 Dimensionality. The dimensionality of the data set has been highly increased because of One Hot Encoding. The number of features has increased with 34 for countrycode, 8 for pos_entry_mode, 42 for payment_method, 6 for funding source and 122 for mccdescription, for a total of 212 extra features. Principal Component Analysis, see also 3.6, has been used to reduce the dimensionality of each of the encoded features, reducing the number of features to one feature for each of the original categorical features. See A4, A5 and A6 for an example of the above described data transformation, using 'payment_method'.

3.2.4 Split in Training and Test. The data-set has been randomly split into a training a test set based on a 80:20 ratio.

3.2.5 Feature Scaling. Feature scaling has been applied to the final features so they can be compared on common grounds and no features dominate others. The features have been normalised using:

$$x' = \frac{x - \bar{x}}{\sigma} \tag{1}$$

where x is the original feature value, $\bar{x}$ is the mean of that feature and σ is its standard deviation. This method has been widely adopted for feature scaling in many machine learning applications [15].

3.3 Clustering

3.3.1 K-Means++. K-Means is the algorithm of choice to perform the clustering. K-Means is able to divide a N-dimensional populations into k clusters efficiently [23]. The K-means algorithm, using a set of numeric datapoints $\boldsymbol{X}$ and an integer number k ($\leq n$), identifies a distribution of $\boldsymbol{X}$ into k clusters that minimises the within group sum of squared errors (WGSS). This method can be expressed using the following mathematical notation P, or the cost function [33] [7]:

$$\text{Minimise} \qquad P(W,Q) = \sum_{l=1}^{k}\sum_{i=1}^{n} w_{i,l}\, d(X_i, Q_1) \tag{2}$$

$$\begin{aligned} \text{subject to} \qquad & \sum_{l=1}^{k} w_{i,l} = 1, \quad 1 \leq i \leq n \\ & w_{i,l} \in 0,1, \quad 1 \leq i \leq n, 1 \leq 1 \leq k \end{aligned} \tag{3}$$

where W is an $n \times k$ partition matrix, $\boldsymbol{Q} = \{Q_1, Q_2, ..., Q_k\}$ is a set of objects in the same object domain, and $d(\cdot,\cdot)$ is the squared Euclidean distance between two objects. The Euclidean distance is a dissimilarity metric that is used to calculate the "ordinary" straight-line distance between two numeric data points. The Euclidean distance the most commonly used criterion function [19]. It is used to find the minimum distance between a data point and a cluster. The (squared) Euclidean distance is given by:

$$d_1(X,Y) = \sum_{j=1}^{p}(x_j - y_j)^2 \tag{4}$$

K-Means has been selected as the algorithm of choice because it has a low time complexity; its linear time complexity makes it computationally attractive [19]. This also means the algorithm scales well and works properly on large scale data [39]. Other advantages are that K-Means has been successfully and thoroughly tested on (large) data sets, it is one of the most popular clustering methods, with reasonably straightforward implementation and many validation methods are existent [19]. A major drawback of K-Means is the dependence on the initial seed selection of K-Means; even though K-means always converges, it may not be a local minimum, because of the (random) initial seed choice. To solve this the K-Means++ initialisation scheme is used. This scheme initialises centroids that are generally distant from the other centroids, which leads to better accuracy and results [3].

3.4 Determining *k*

Determining the ideal number of clusters k is an important part of cluster analysis, as this parameter is predefined when executing the K-Means algorithm. Choosing the right k is oftentimes ambiguous [27]. There are several methods that aid in making this decision.

3.4.1 Elbow Method. The elbow method is a visual method that is used to determine the optimal number of clusters. This method plots values of the cost function 2 produced for different k values. When k increases, the average distortion decreases, each cluster has fewer constituent data points and the data points are closer to the clusters' centroids. However, the decrease in the average distortion will decline with each higher value of k and at some point, the marginal gain drops. This point is the "elbow" and shows as an angle in the graph. This point marks the optimal number of clusters.

3.4.2 Silhouette Analysis. See Silhouette Coefficient 3.5.1.

3.5 Cluster Validation

Validating the clustering results is an important aspect of clustering. Clustering analysis always yields results (e.a. the data is always divided into clusters), therefore validating whether these results are good is vital. Clusters can be validated based on two dimensions [16]: Compactness and separation. Compactness means that the data points in each cluster must be as close to each other as possible and separation means that the clusters must be separated as widely as possible. Cluster validation knows three fundamental concepts: External, internal and relative criteria [16]. External evaluation compares results to an existing "ground truth", for example, external benchmarks and known class labels. Internal evaluation evaluates the results of the data that was clustered itself, these methods assign a single quality score to the results. Relative evaluation compares clusters by using different parameter values of the algorithm. As there is no ground truth is available for the data set, internal evaluation methods will be used. Various validation methods have been researched, however, not one single method outperforms the others [2]. The Silhouette Coefficient and The Davies-Bouldin Index are two commonly used internal validation methods and will be used in this research.

3.5.1 Silhouette Coefficient. The Silhouette Coefficient is an internal evaluation score first proposed in 1987 [30]. This score is calculated by determining the proximity of data points in clusters to its centroid [1]. The Silhouette Coefficient s can be calculated using:

$$s_i = \frac{b_i - a_i}{max(a_i, b_i)} \tag{5}$$

where a is the mean distance between a sample and all other data points in the same cluster and b is the mean distance between a sample and all other data points in the next nearest cluster. the resulting score is confined between -1 and $+1$. A score of -1 indicates incorrect clustering, $+1$ indicates highly dense clustering and scores nearby zero indicate overlapping clusters. A higher score means that clusters are compact and well separated. Running the algorithm for various values of k and subsequently determining the Silhouette score aids in determining the optimal value of k.

3.5.2 The Davies–Bouldin Index. The Davies–Bouldin index (DBI) is an internal evaluation method introduced in 1979 [11]. The DBI is defined as the average similarity between each cluster C_i for $i = 1, ..., k$ and its most similar one C_j. The Davies–Bouldin index DB can be calculated using:

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} R_{ij} \tag{6}$$

$$\text{where} \qquad R_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{7}$$

and where s_i is the average distance between each points of cluster i and the clusters' centroid, d_{ij} is the distance between the centroids of clusters i and j. Zero is the DBI's lowest score possible, scores close to zero indicate a good cluster partition.

3.5.3 The Calinski-Harabasz Index. The Calinski-Harabasz Index (CHI) is a criterion that can be used to evaluate a clustering model. For k clusters the CH score is given as the ration of the between-cluster distribution mean and the within-cluster distribution [8]:

$$s(k) = \frac{Tr(B_k)}{Tr(W_k)} \times \frac{N-k}{k-1} \tag{8}$$

where B_k is the between group dispersion matrix, W_k is the within-cluster dispersion and N is the number of data points. A higher Calinski-Harabasz Index indicates better-defined clusters within the model.

3.6 Principal Component Analysis

Principal component analysis (PCA) is used to reduce the dimensionality of data sets that contain a large number of features while preserving the initial variation of the data set as best as possible [20]. The features of the data set are transformed into a new set of features, the principal components (PCs). The PCs are uncorrelated and are ordered in a way to preserve as much variation as possible. PCA transforms the data set to a new coordinate system in a way such that the highest variance comes to lie on the first PC and the second highest variance on the second PC, etc. A useful application of PCA is data visualisation. More than three dimensions (e.a. features) are visually impossible to represent. Reducing the number of dimensions to two allows for the resulting clusters to be visualised in a 2D plot.

Consider a data matrix $\boldsymbol{X}$, with n rows and p columns. Mathematically the PCA transformation is specified by a set of p-dimensional vectors of weights $w_k = (w_1, \dots, w_p)_k$ that map each row vector x_i to a new vector of PC scores $t_i = (t_1, \dots, t_l)_i$ given by: $t_{k(i)} = x_i \times w_k$ for $i = 1, \dots, n$ and $k = 1, \dots, l$ so that that each variable $t_1, \dots, t_l$ successively inherits the highest variance possible from x and with every coefficient vector w constrained into being a unit vector. The full PC transformation of $\boldsymbol{X}$ can be defined as:

$$T = \boldsymbol{X}W \tag{9}$$

where W is a $p \times p$ matrix of weights of which the columns are the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$.

4 EVALUATION

4.1 Number of clusters

4.1.1 Elbow Method. The Elbow method has been used as an aid in determining the ideal number of clusters. Figure 1 shows the Elbow method plot for the data set. The line shows a first 'elbow' around k = 3, this decrease in the slope continues to about k = 6. The slope is not significantly decreasing when k is higher than 6. This indicates the ideal number of clusters is between 3 and 6.

4.1.2 Silhouette Analysis. A Silhouette analysis has been used to determine the ideal number of clusters. The Silhouette score for 2 to 8 clusters are presented in 1. The highest number indicates the ideal number of clusters. From 1 it can be seen that the ideal number of clusters is 4, which has a corresponding Silhouette score of 0.6967. This is confirmed when analysing the Silhouette graphs. See A7 until A13 in the Appendix A. The graphs show the silhouette coefficient values on the x-axis and the cluster labels on the y-axis. The dotted line in each graph represents the average Silhouette width (e.a. Silhouette score). The coloured 'blobs' in the graphs each represent a cluster, with A7 having 2 clusters and A13 having 8 clusters. The closer these blobs are to 1, the better the structure. From the graphs, it can also be seen that 4 is the ideal number of clusters.

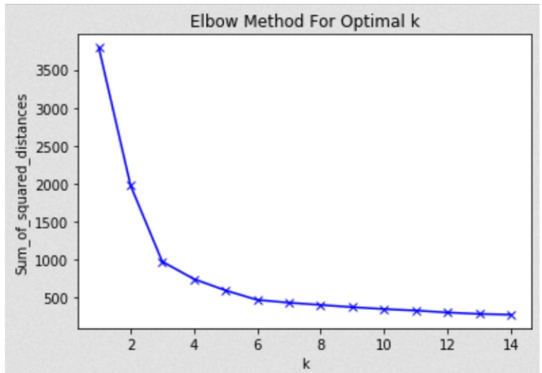

Fig. 1. Elbow method plot.

Number of clusters	Silhouette score
2	0.6326208684288791
3	0.6740801015367901
4	0.6966603450217758
5	0.6796448352763816
6	0.6909733851714268
7	0.624340328008347
8	0.6242220815997035

Table 1. Silhouette scores.

4.2 Clusters

Adyen POS merchants are analysed using the K-Means++ algorithm in order to cluster them. As demonstrated earlier in 4.1, the ideal number of clusters is 4. Therefore, the algorithm distributes all POS merchants (N=13200) into 4 clusters. Table 2 displays these clusters. Cluster 3 is the largest cluster and contains 54.0% of the merchants. The second largest cluster is cluster 1 (27.6%), followed by cluster 2 (11.9%) and the smallest cluster is 4, containing 6.5% of POS merchants.

Cluster	N	%
1	3640	27.6
2	1574	11.9
3	7127	54.0
4	859	6.5
Total	13200	100

Table 2. Cluster distribution.

4.3 Principal Component Analysis

To visually display the clusters of the multi-dimensional data, the number of dimensions has been reduced to 2, using Principal Component Analysis. The results have been plotted in a Principal

Component plot, see figure 2. The plot shows the four clusters in red, blue, yellow and green. The clusters are nicely separated and very dense. Only a few points are outside the main clusters.

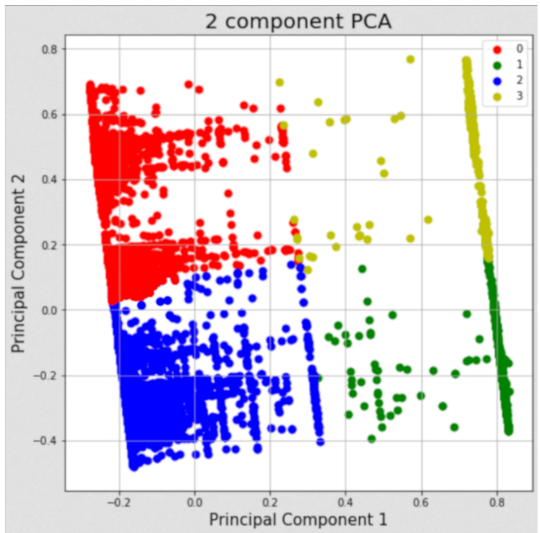

Fig. 2. Principal Component Plot

4.4 Cluster Evaluation

To evaluate the clustering performance of the algorithm a Silhouette Score, Davies-Bouldin Index and Calinski-Harabasz Index have been calculated. These scores are presented in table 3.

Evaluation Name	Score
Silhouette Score	0.6966603450217758
Davies-Bouldin Index	0.4220733667404933
Calinski-Harabasz Index	19099.218572513302

Table 3. Evaluation scores.

4.4.1 Silhouette Score. The Silhouette score varies from -1 to 1, where -1 indicates incorrect clustering and 1 implies near perfect cohesion. A high Silhouette score of 0.6967 indicates that the clusters are very dense and well separated.

4.4.2 Davies-Bouldin Index. The Davies-Bouldin Index is the ratio between the between-cluster distances and the within-cluster distances, and then calculating the average over all the clusters. The score is bounded between 0 and 1, where a lower score indicates better clustering. The DBI for the clusters in this research is 0.4220. This is a low score, that is closer to 0 than to 1, indicating good clustering results.

4.4.3 Calinski-Harabasz Index. The Calinski-Harabasz Index compares the variance within each cluster to the variance between clusters. The CHI score is not bounded and can, therefore, take practically any value, where a high value indicates better cluster separation. The CHI score for the clusters is 19099.22. This is a high score, indicating good cluster separation.

5 CONCLUSIONS

This presents a clustering method for POS merchants. The resulting clusters represent merchant segments and are very dense and well separated. These segments allow PSPs to benchmark merchants, compare integrations and quickly determine set-up settings for new merchants. This research is, to the best of my knowledge, the first academic research that proposes a segmentation method for (Point of Sale) merchants.

REFERENCES

[1] R. Amorim and C. Hennig. 2015. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences* 324 (12 2015), 126–145. https://doi.org/10.1016/j.ins.2015.06.039

[2] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Perez, and I. Perona. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition* 46 (2013), 243–256.

[3] D. Arthur and S. Vassilvitskii. 2007. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1027–1035.

[4] L. Barnett, Norman. 1969. Beyond market segmentation. *Harvard Business Review* 47, 1 (1969), 152–166.

[5] T.P. Beane and D.M. Ennis. 1987. Market segmentation: a review. *European journal of marketing* 21, 5 (1987), 20–42.

[6] J.C. Bezdek, R. Ehrlich, and W. Full. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences* 10, 2-3 (1984), 191–203.

[7] L. Bobrowski and J.C. Bezdek. 1991. c-means clustering with the l/sub l/ and l/sub infinity / norms. *IEEE Transactions on Systems, Man, and Cybernetics* 21, 3 (May 1991), 545–554. https://doi.org/10.1109/21.97475

[8] T. Calinski and J. Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3, 1 (1974), 1–27.

[9] V.E. Castro and J. Yang. 2000. A Fast and Robust General Purpose Clustering Algorithm. *International Conference on Artificial Intelligence*.

[10] C. Constantin. 2012. Post-hoc Segmentation using Marketing Research. 12, 3 (2012), 39–48.

[11] D.L. Davies and D.W. Bouldin. 1979. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1, 2 (April 1979), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

[12] United Nations. DESA / Population Division. 2018. How do we define "urban agglomeration"? United Nations. https://population.un.org/wup/General/FAQs.aspx

[13] M. Ester, H.P. Kriegel, JŠander, and X. Xu. 1996. A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*. AAAI Press, 226–231. http://dl.acm.org/citation.cfm?id=3001460.3001507

[14] C. Fraley and A.E. Raftery. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal* 41, 8 (1998), 578–588.

[15] J. Grus. 2015. *Data Science from Scratch*. 99–100 pages.

[16] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. 2001. On Clustering Validation Techniques. *J. Intell. Inf. Syst.* 17, 2-3 (Dec. 2001), 107–145. https://doi.org/10.1023/A:1012801612483

[17] G.E. Hinton, T.J. Sejnowski, and T.A. Poggio. 1999. *Unsupervised Learning: Foundations of Neural Computation*. MCGRAW HILL BOOK Company. https://books.google.nl/books?id=yj04Y0lje4cC

[18] A.K. Jain, R.C. Dubes, et al. [n. d.]. *Algorithms for clustering data*. Prentice-Hall, Inc.

[19] A.K. Jain, M.N. Murty, and P.J. Flynn. 1999. Data Clustering: A Review. *ACM Comput. Surv.* 31, 3 (Sept. 1999), 264–323. https://doi.org/10.1145/331499.331504

[20] I. Jolliffe. 1986. *Principal Component Analysis and Factor Analysis*. 115–128. https://doi.org/10.1007/978-1-4757-1904-8_7

[21] L. Kaufman and P. Rousseeuw. 1990. *Finding Groups in Data: An Introduction To Cluster Analysis*. https://doi.org/10.2307/2532178

[22] D. Lam and D.C. Wunsch. 2014. Clustering. In *Academic Press Library in Signal Processing*. Vol. 1. Elsevier, 1115–1149.

[23] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, Berkeley, Calif., 281–297. https://projecteuclid.org/euclid.bsmsp/1200512992

[24] J.H. Myers. 1996. *Segmentation and Positioning for Strategic Marketing Decisions*. American Marketing Association.

[25] R.T. Ng and J. Han. 2002. CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge & Data Engineering* 5 (2002), 1003–1016.

[26] A. Parvaneh, H. Abbasimehr, and M.J. Tarokh. 2012. Integrating AHP and data mining for effective retailer segmentation based on retailer lifetime value. *Journal of Optimization in Industrial Engineering* 5, 11 (2012), 25–31.

[27] D. Pelleg and A. Moore. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *In Proceedings of the 17th International Conf. on Machine Learning*. Morgan Kaufmann, 727–734.
[28] Visa Public. 2018. Visa Merchant Data Standards Manual. In *Visa Core Rules and Visa Product and Service Rules*. VISA.
[29] L. Rokach. [n. d.]. *Data Mining and Knowledge Discovery Handbook*. 331–352 pages.
[30] P.J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (1987), 53 – 65. https://doi.org/10.1016/0377-0427(87)90125-7
[31] C. Rygielski, J.C. Wang, and D.C. Yen. 2002. Data mining techniques for customer relationship management. *Technology in society* 24, 4 (2002), 483–502.
[32] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O.P. Patel, A. Tiwari, M.J. Er, W. Ding, and C.T. Lin. 2017. A review of clustering techniques and developments. *Neurocomputing* 267 (2017), 664–681.
[33] S.Z. Selim and M.A. Ismail. 1984. K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6, 1 (Jan 1984), 81–87. https://doi.org/10.1109/TPAMI.1984.4767478
[34] B.P. Shapiro and Bonoma T.V. 1984. How to Segment Industrial Markets," Harvard Business Review. (05 1984).
[35] W.R. Smith. 1956. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing* 21, 1 (1956), 3–8.
[36] M. Wedel and W.A. Kamakure. 2000. *Market Segmentation*. Springer US.
[37] A. Weinstein. 2004. *Handbook of Market Segmentation: Strategic Targeting for Business and Technology Firms*. Haworth Press.
[38] R. Xu and I.I.D. Wunsch. 2005. Survey of Clustering Algorithms. *Trans. Neur. Netw.* 16, 3 (May 2005), 645–678. https://doi.org/10.1109/TNN.2005.845141
[39] T. Xu, D. Yingjie. 2015. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science* 2, 2 (01 Jun 2015), 165–193. https://doi.org/10.1007/s40745-015-0040-1
[40] T Zhang, R. Ramakrishnan, and M. Livny. 1996. BIRCH: an efficient data clustering method for very large databases. In *ACM Sigmod Record*, Vol. 25. ACM, 103–114.

A APPENDIX

A.1 Features

- companyaccountid: Unique (tokenised) string representing Adyen merchants.
- merchantaccountid: Unique (tokenised) string representing Adyen sub-accounts.
- storeid: Unique (tokenised) string representing stores.
- pos_entry_mode: The way the shopper interacts with the termina (e.g. Chip or contactless).
- payment_method: The payment method used for the payment (e.g. VISA)
- fundingsource: The funding source of the payment (e.g. DEBIT).
- mccdescription: The MCC description that applies to a store (e.g. Shoe Store).
- countrycode: Country where the store is located (e.g. US)
- city: City where the store is located.
- transaction_count: The number of transactions.
- terminal_count: The number of active terminals in a store.
- shopper_count: The number of individual shoppers that performed a payment.
- internationalshopper_count: The number of international (e.a. foreign) shoppers.
- auth_rate: The authorisation rate (e.a. how much payments succeeded).
- atv: The average transaction value.
- creation_day_distinctcount: The number of active days of a store during the sample period.

companyaccountid	merchantaccountid	storeid	pos_entry_mode	payment_method	fundingsource	mccdescription	countrycode	...
540504	540510	1423	Chip	maestro	DEBIT	Motion Picture Theaters	BE	
584914	600023	4608	MSR	visa	CREDIT	Fast Food Restaurants	US	
540504	540508	1596	CTLS Chip	amex	DEBIT	Family Clothing Stores	FR	

city	transaction_count	terminal_count	creation_day_distinctcount	shopper_count	internationalshopper_count	auth_rate	atv
Brussels	231	5	300	221	18	0,9891	20,3729
Chicago	54	2	352	53	21	0,9437	115,5475
Cholet	7	3	326	40	0	0,8936	447,8079

Fig. A3. Example entries.

Companyaccountid	merchantaccountid	storeid	payment_method_maestro	payment_method_visa	payment_method_amex	...
540504	540510	1423	1	0	0	
584914	600023	4608	0	1	0	
540504	540508	1596	0	0	1	

Fig. A4. One Hot Encoding example for payment_method

companyaccountid	merchantaccountid	storeid	payment_method_maestro	payment_method_visa	payment_method_amex	...
540504	540510	1423	20%	0	0	
584914	600023	4608	0	30%	0	
540504	540508	1596	0	0	10%	

Fig. A5. Transforming to percentages

companyaccountid	merchantaccountid	storeid	payment_method_PCA
540504	540510	1423	0,3469
584914	600023	4608	-0,1062
540504	540508	1596	0,2123

Fig. A6. Principal Component Analysis

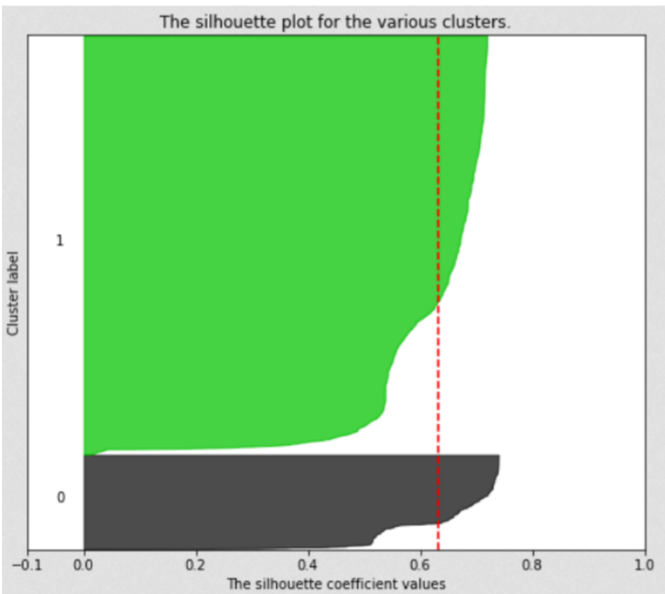

Fig. A7. Silhouette graph for k=2.

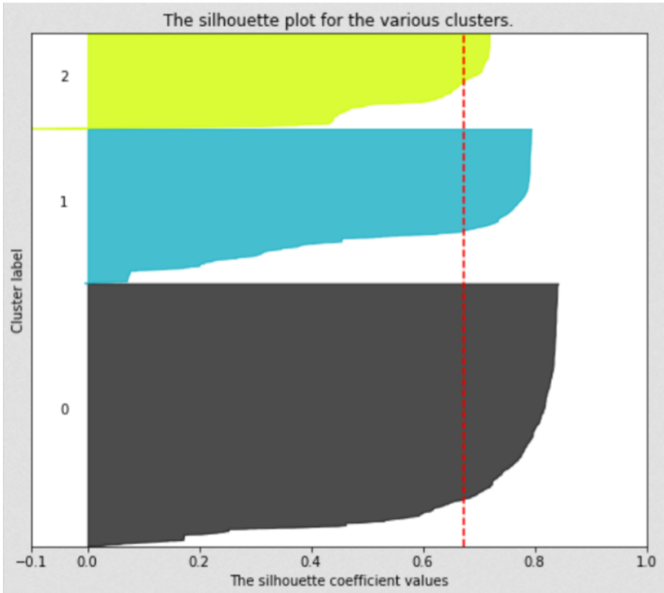

Fig. A8. Silhouette graph for k=3.

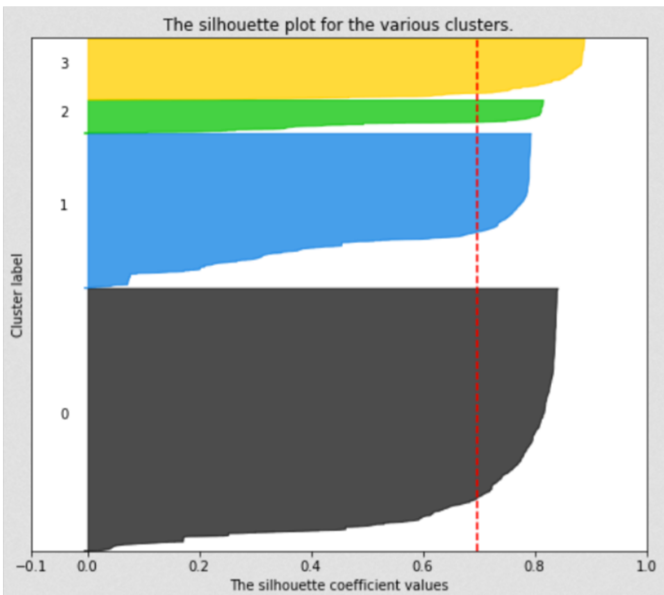

Fig. A9. Silhouette graph for k=4.

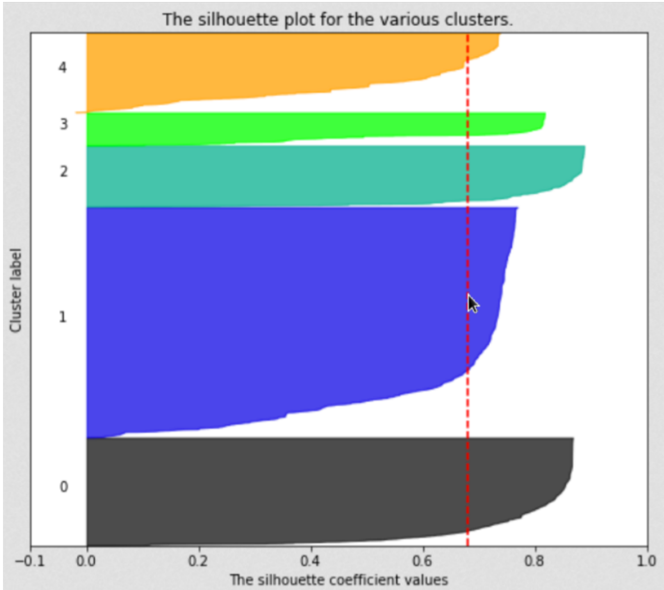

Fig. A10. Silhouette graph for k=5.

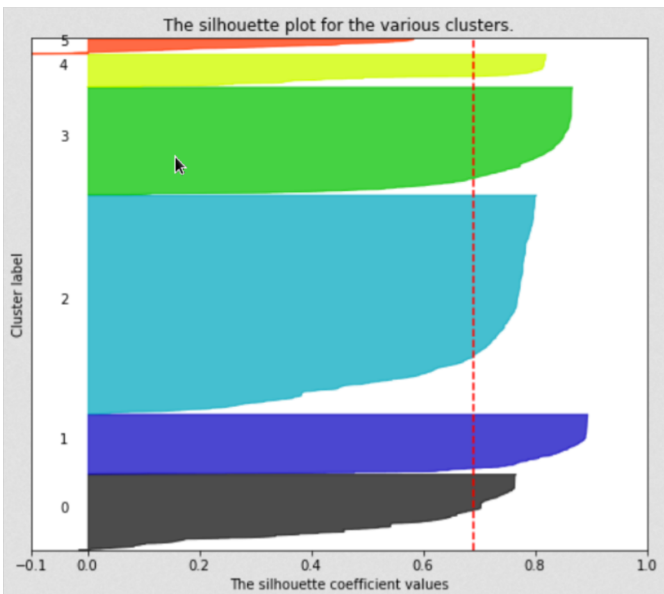

Fig. A11. Silhouette graph for k=6.

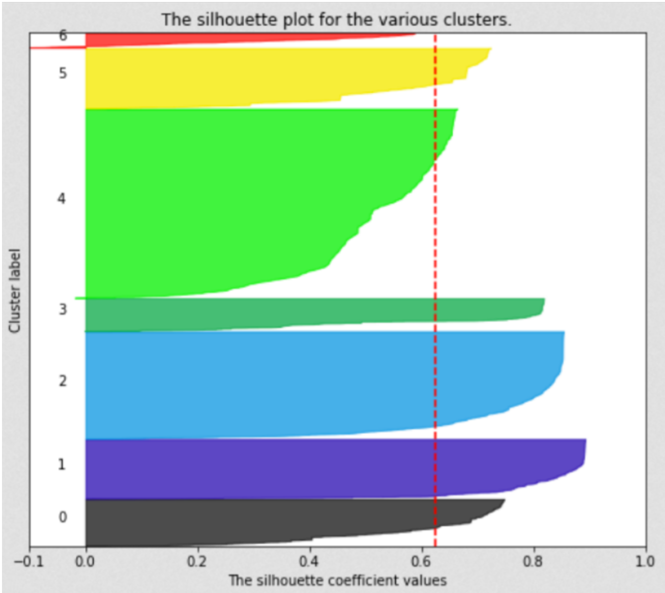

Fig. A12. Silhouette graph for k=7.

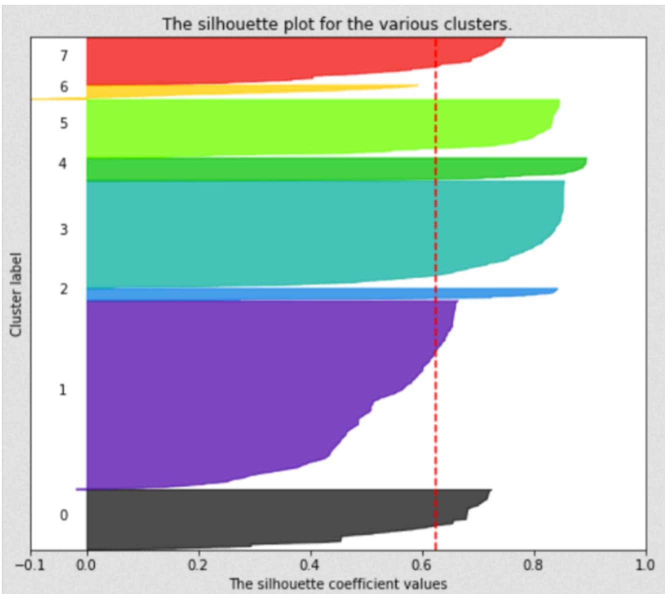

Fig. A13. Silhouette graph for k=8.