CUSTOMER SEGMENTATION USING CLUSTERING

Akram Hareem,
3rd-year student, VSU named after P.M. Masherov, Vitebsk, Republic of Belarus;
Scientific supervisor – Yermochensko S.A., candidate of physical
and mathematical sciences, associate professor

In the increasingly competitive retail and e-commerce industries, knowing customer behaviour is critical for designing focused marketing tactics and increasing customer happiness. Customer segmentation, which groups customers according to their purchase habits, is one of the most important strategies for accomplishing this. Using real-time consumer transaction data, this work proposes a clustering-based method [1] for client segmentation. The ability to understand customer purchasing behaviour is significant for businesses, so, as to enhance customer engagement and significantly increase revenue growth. The goal of this work is to use customer segmentation using real-time transaction data from Unbanx [2], which delivers anonymized transactional data for UK consumers. The segmentation will use clustering algorithms, including K-means and hierarchical clustering to find groups of customers with similar purchasing patterns.

Material and methods. The data used in this study is obtained from Unbanx Real-Time UK Consumer Transaction Data, which includes de-identified consumer transactions with attributes such as transaction amount, timestamp, merchant category, and payment method. To ensure meaningful clustering, data pre-processing steps such as cleaning, normalization, and feature selection are performed.

Clustering analysis is performed using the following methods [3]:

– ***K-means Clustering*:** Applied to segment customers based on spending frequency and transaction volume. The ideal number of clusters is determined using the 'elbow method'.

– ***Hierarchical Clustering*:** Is used to compare and choose results obtained from K-means clustering. Dendrograms are made to visualize cluster formations.

Results and Discussion. To evaluate the clustering results, we finally use the *silhouette score* and *Davies-Bouldin* index. The clusters are visualized using scatter plots and dendrograms which provide a clear interpretation of customer groups.

Implementation in Python [4]

The Python code mentioned below is used to process the real-time dataset using clustering and visualising results:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns from sklearn.cluster
import KMeans, AgglomerativeClustering from sklearn.preprocessing
import StandardScaler from sklearn.metrics
import silhouette_score
# Load real-time dataset (example: Unbanx API data)
data = pd.read_csv("unbanx_transaction_data.csv")
# Data preprocessing
data = data.dropna()
data['TransactionAmount'] = data['TransactionAmount'].astype(float)
scaler = StandardScaler()
data[['TransactionAmount', 'TransactionFrequency']] =
                      scaler.fit_transform(data[['TransactionAmount',
                                                 'TransactionFrequency']])
# K-means clustering
kmeans = KMeans(n_clusters=4, random_state=42)
```

```
data['Cluster'] = kmeans.fit_predict(data[['TransactionAmount',
                                           'TransactionFrequency']])
# Hierarchical clustering
hierarchical = AgglomerativeClustering(n_clusters=4)
data['Cluster_Hierarchical'] =
                  hierarchical.fit_predict(data[['TransactionAmount',
                                                 'TransactionFrequency']])
# Visualizing
clusters plt.figure(figsize=(8,6))
sns.scatterplot(data=data, x='TransactionAmount', y='TransactionFrequency',
                hue='Cluster', palette='viridis')
plt.title("Customer Segmentation using K-means")
plt.show()
# Evaluation metrics
silhouette_kmeans = silhouette_score(data[['TransactionAmount',
                                  'TransactionFrequency']], data['Cluster'])
print(f"Silhouette Score for K- means: {silhouette_kmeans}")
```

As a result of clustering analysis, several distinct customer segments are formed and identified. The segments include high-spending (frequent buyers), occasional buyers and low-spending (infrequent buyers). The analysis of these groups reveals a deep insight that can help the marketing strategies positively:

– High-value customers display frequent purchases with higher transaction amounts, which demands a need for loyalty programs.

– Occasional customers can be targeted with promising promotional offers to increase engagement.

– Low-value customers may require a different approach, such as discounts and personalized persuasion.

The comparison of clustering techniques shows that while hierarchical clustering offers improved interpretability of customer relationships, K-means also produces well-defined clusters. The evaluation metrics validate the effectiveness of the chosen approach.

The following table summarizes the cluster characteristics:

Cluster	Avg. Transaction Amount	Purchase Frequency	Marketing Strategy
1	High	High	Loyalty Programs
2	Medium	Medium	Personalized Offers
3	Low	Low	Discounts & Retargeting
4	High	Low	Exclusive Promotions

Conclusion. This work emphasises the use of clustering techniques for customer segmentation based on real-time consumer transaction dataset. The insights taken from the segmentation process can greatly contribute to optimizing marketing growth and marketing strategies, improving customer engagement as well enhancing business decision-making. To further refine segmentation models we can later integrate additional features as well.

1. Ермоченко, С.А. Применение кластерного анализа в построении поисковых информационных систем / Наука – образованию, производству, экономике: материалы XVIII (65) Регион. науч.-практ. конф. преподавателей, научн. сотрудников и аспирантов, Витебск, 13–14 марта 2013 г.: в 2 т. – Витебск: ВГУ имени П.М. Машерова, 2013. – Т. 1. – С. 11–13. Электронная версия: https://rep.vsu.by/handle/123456789/9731

2. Unbanx. (n.d.). Unbanx Real-Time UK Consumer Transaction Data. Retrieved from Datarade: https://datarade.ai/data-categories/consumers-transaction-data

3. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems). – Morgan Kaufmann. – 744 P.

4. Wes McKinney (2022). Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. – O'Reilly Media. – 579 P.