Customer Segmentation: A Key to Unlocking Business Growth and Success

 [1]

In today's highly competitive market, understanding your customers and tailoring your marketing strategies to meet their specific needs and preferences is critical for business success. Customer segmentation, the process of dividing a large customer base into smaller groups based on shared characteristics, can help businesses gain valuable insights into their customers and develop targeted marketing strategies that maximize customer engagement and loyalty.

For this assignment, you will be provided with two datasets containing information on customer demographics, spending habits, purchasing behavior and historical transitions of the customer. Your task is to perform customer segmentation and identify distinct groups of customers based on their shared characteristics. The two datasets are named - *customer_info* and *customer_basket.*

[1] Cluster icons created by Flat Icons - Flaticon

The data

Here's a description of the datasets and columns - starting with *customer_info:*

Description:

- **Contains information about each customer, namely demographical and spend behavior in the past 2 years of data. This data will help you characterize the customer.**

Columns:

- **customer_id:** Identifier of the customer.
- **customer_name:** Name of the customer (contains degree level).
- **customer_birth_date:** Birth date of the customer.
- **kids_home:** Number of kids at home.
- **teen_home:** Number of teens at home.
- **number_complaints:** Number of complaints formally done by the customer.
- **location_latitude:** Approximate location of the customer's home (Latitude).
- **location_longitude:** Approximate location of the customer's home (Longitude).
- **distinct_stores_visited:** Number of distinct stores visited by the customer.
- **loyalty_card_number:** Number of the customer loyalty card.
- **lifetime_spent_groceries:** Total value spent by the customer on groceries.
- **lifetime_spent_electronics:** Total value spent by the customer on electronics.
- **lifetime_spent_vegetables:** Total value spent by the customer on vegetables.
- **lifetime_spend_nonalcohol_drinks:** Total value spent by the customer on non-alcoholic drinks.
- **lifetime_spend_alcohol_drinks:** Total value spent by the customer on alcoholic drinks.
- **lifetime_spent_meat:** Total value spent by the customer on meat.
- **lifetime_spent_fish:** Total value spent by the customer on fish.
- **lifetime_spent_hygiene:** Total value spent by the customer on hygiene.
- **lifetime_spent_videogames:** Total value spent by the customer on video games.
- **lifetime_total_distinct_products:** Number of distinct products bought by the customer (lifetime)**.**
- **year_first_transaction:** Year of the first transaction of the customer.
- **percentage_of_products_bought_promotion:** Percentage of products that were bought with some promotion.
- **typical_hours:** Typical hour when the customer visits the store.
- **latitude:** Latitude of the Customer's Address

- **longitude:** Latitude of the Customer's Address

Regarding *customer_basket,* that file contains 80.000 thousand random baskets from our customers:

Description:

- Contains information about customers' different baskets bought at the shop.
- Each line consists of the customer_id, transaction_id and products in list format.
- The transactions are related to the last 6 months of data.

Columns:

- **customer_id:** Identifier of the customer. Can be used to connect this dataset with the *customer_info* one.
- **invoice_id:** Identifier of the transaction.
- **list_of_goods:** Products bought in the transaction in list format.

Additionally, you can also access *product_mapping.xlsx,* an excel file that contains a mapping between the product name and "category".

The project

In this project, you will be asked to:

- Identify relevant customer segments: Using statistical and machine learning techniques, you will need to identify meaningful segments within the customer base that share similar characteristics. These clusters can then be used to develop targeted marketing strategies

- Analyze customer behavior: Once you have identified customer segments, you will need to analyze their behavior to gain insights into their motivations, preferences, and needs. This may involve analyzing their purchasing patterns, loyalty card usage, complaint history, and other relevant data.

- Develop targeted marketing strategies: Using the customer_basket dataset, you will need to develop targeted marketing strategies that appeal to each customer segment's unique needs and preferences. This may involve developing personalized promotions, creating targeted advertising campaigns, or tailoring product offerings to meet specific customer demands.

Overall, customer segmentation is a powerful tool that can help businesses gain a competitive edge by understanding their customers better and developing targeted marketing strategies that maximize customer engagement and loyalty. By completing this assignment, you will gain hands-on experience with unsupervised learning and develop valuable skills that can help you succeed in a wide range of data science roles.

Deliverables

You will be asked to deliver a notebook (or more, if you want) with code and explanation of the solution. Clean, "DRY" (Don't repeat yourself) and clear code will earn extra points in the submission. **You can choose to abstract code outside of notebooks but that is not mandatory.**

In addition, you will have to submit an executive report with the detail of each segment and suggested campaigns that you would do for each segment - **be creative on the campaign suggestion** - here are some examples:

- Get an extra 20% on Fish if you buy meat.
- 50% promotion on the entire basket just for you!
- Buy 1 Video Game, get 1 free.

To come up with the promotions, you can play around with association rules on the *customer_basket* dataset and relate those baskets with the customers on the *customer_info* one.

A **suggested** index of the report is the following:

- 1) Executive Summary
- 2) Exploratory Data Analysis
- 3) Customer Segmentation
- 4) Targeted Promotion
- 5) Conclusion of the project

Here is what you **can** include in each section (not extensive):

1) **Executive Summary:**

The executive summary is a brief overview of the project. It should provide a high-level description of the problem you addressed, the approach you took, and the results you achieved. Specifically, you should include the following:

- A brief overview of the problem you addressed
- A summary of your methodology and approach
- A summary of your key findings and results

- A recommendation based on your findings

2) Exploratory Data Analysis:

The exploratory data analysis section should provide a detailed overview of the data you used in the project. Specifically, you can include the following:

- A description of the data sources you used
- A summary of any data cleaning or preprocessing you did
- A visualization of the data (e.g., histograms, scatter plots, etc.)
- A summary of any patterns or trends you identified in the data
- An explanation of any outliers or anomalies in the data
- A summary of any correlations or relationships you identified between variables

3) Customer Segmentation:

The customer segmentation section should describe how you identified and segmented your customers. Specifically, you should include the following:

- A description of the segmentation approach you used
- Any heuristics you used before the segmentation
- A summary of the customer segments you identified
- A description of the characteristics of each segment (e.g., demographics, behavior, needs, etc.)
- A comparison of the segments in terms of size or other relevant metrics

4) Targeted Promotion:

The targeted promotion section should provide a detailed plan of the campaigns for your target customers.

Specifically, you should include the following:

- A description of the promotions approach for each segment
- What mechanics you will use in the promotion (buy one-get one, slash price, combination of products).

5) Conclusion of the project:

The conclusion section should provide a summary of the project and its findings. Specifically, you should include the following:

- A summary of the problem you addressed
- A summary of the methodology and approach you used
- A summary of the key findings and results
- A brief reflection on the overall success of the project

No code should be presented during the report - code should be reserved for the Notebook(s).

Evaluation

The project will be evaluated based on code quality, cleanness and overall quality of the segmentation solution. There is an “ideal solution”, but If you don’t deviate too much from it and you can justify the segments you have discovered, the solution will be considered feasible.

In terms of evaluation parameters, the following will be considered:

- Technical quality of the Segmentation: **8 Points**
- Quality of the Report: **4 Points**
- Visuals and Explanation: **3 Points**
- Interpretation of the Clusters: **2 Points**
- Feasibility of Promotions vs. Cluster Interpretation: **3 Points**
- Extra: Code quality: **+2**