Customer Segmentation Project

Fall Semester 2025-2026

In today's 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.

Unsupervised machine learning plays an important role in customer segmentation by uncovering hidden patterns in data without requiring labeled examples. Unlike traditional marketing approaches that rely on predefined customer categories, unsupervised learning allows businesses to discover natural groupings within their customer base based on purchasing behavior, demographics, browsing history, or other relevant data points.

For this assignment, you will be provided with a dataset containing information about 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 dataset is named *customer_info.*

The data

Here's a description of the datasets and columns*:*

Description:

- **Contains information about each customer, namely demographical and spend behavior for each customer.**

Columns:

- **customer_id:** Identifier of the customer.
- **customer_name:** Name of the customer (contains degree level).
- **customer_gender:** Gender of the customer.
- **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.
- **distinct_stores_visited:** Number of distinct stores visited by the customer.
- **lifetime_spend_groceries:** Total value spent by the customer on groceries.
- **lifetime_spend_electronics:** Total value spent by the customer on electronics.
- **typical_hour:** Typical hour when the customer visits the store
- **lifetime_spend_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_spend_meat:** Total value spent by the customer on meat.
- **lifetime_spend_fish:** Total value spent by the customer on fish.
- **lifetime_spend_hygiene:** Total value spent by the customer on hygiene.
- **lifetime_spend_videogames:** Total value spent by the customer on video games.
- **lifetime_spend_petfood:** Total value spent by the customer on pet food.
- **lifetime_total_distinct_products:** Number of distinct products bought by the customer (lifetime)**.**
- **percentage_of_products_bought_promotion:** Percentage of products that were bought with some promotion.
- **year_first_transaction:** Year of the first transaction of the customer.
- **loyalty_card_number:** Number of the customer loyalty card.
- **location_latitude:** Approximate location (<1km range) of the customer's home (Latitude).

- **location_longitude:** Approximate location (<1km range) of the customer's home (Longitude).

To access this data, please go to Moodle and find the dataset in the project folder in the header of the course.

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 will involve profiling these clusters and understanding why they are different from each other.

- **Develop targeted marketing strategies:** 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. Use the information you found on profiling to build and recommend your marketing strategy.

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.

Project Deliverables

1. **Written Report** describing:
    - the results and conclusions you gained from the analysis you performed;
    - the process by which you arrived at these insights, justifying the decisions you made;
    - the different clusters you found.

No need to discuss how the algorithms work (unless you used an algorithm not discussed in the practical sessions).

The report should be submitted as a pdf file and named using this format where “99” should be replaced with your group number:

DM1_2526_Group_99_Report.pdf

Make sure that the report contains the **names** and **student numbers** of all the members of the group.

Minimum 5 pages, maximum 15 pages of content (excluding cover page, index and appendices).

2. **Colab Notebook(s)** (*ipynb* file format) containing all the code used to develop your project. Instructions will be provided for how to download your notebook from Google Colab.

The file should be named using this format, where “99” should be replaced with your group number:

DM1_2526_Group_99_Notebook.ipynb

Make sure that the notebook contains a Markdown cell (text cell) with the **names** and **student numbers** of all the members of the group.

3. **Deadline**. 07 January 2026 23:59. A penalty of 10% will be given for each day of delay.

4. **Discussion**. After submitting the projects the students will be called to discuss the project with one of the instructors.