

## Project Description

Our Imaginary Enterprise:

### WalSmart

**WalSmart** is a ten-year-old supermarket chain that seeks to provide customers with a diverse range of high-quality products, including groceries, fresh produce, and unique specialty items sourced from around the world. Its mission is to delight customers with well-curated, affordable, and exciting products that might not be available at other large retail stores.

Through aggressive promotions in food magazines, online ads, and local media, WalSmart has built a strong customer base, currently maintaining a database of over 10,000,000 shoppers. Most customers are regular grocery shoppers, but many also seek out specialty and gourmet items. WalSmart sometimes offers a variety of products, ranging from dairy and meat products to wines.

WalSmart collected customer data in the last year but has so far focused on mass marketing. Now, WalSmart wants to "get smart" about its database, start differentiating its customers, and develop more targeted marketing programs.

The company has provided a random sample of ~120,000 of its active customers. These customers have made a purchase during 2023.

Using this data, you will perform a customer segmentation with the goals of:

- **Engagement/Value Segmentation:** Identify different groups of customers based on how valuable they are to the company (e.g., high spenders, frequent shoppers, etc.). You will need to select relevant variables from the dataset for this analysis.
- **Buying Behavior Segmentation:** Understand different customer groups based on their product preferences (e.g., fresh produce, gourmet items, household goods, etc.). This analysis will focus on the categories of products that customers tend to buy more frequently.

You can decide if you would like to build one or more different segmentations.

Finally, you will consolidate the findings from both segmentations and propose tailored marketing strategies for each identified customer segment.

## Dataset Description

### Variables in the ProjectData table

Field	Description
<b>ID_Client</b>	Customer ID
<b>Frequency_items</b>	Quantity of items bought by customer (2023)
<b>Potential_Score</b>	Customer's potential score. Measure of customer's potential for increased purchases (-100 to 100). % from previous year. Computed by other models.
<b>ZIP_Code</b>	Macro Zip Code of customer's address
<b>Longevity_months</b>	Measured from 31st December 2023 backwards
<b>Latitude</b>	Latitude of customer's address
<b>Longitude</b>	Longitude of customer's address
<b>Credit_factor</b>	Customer's credit worth, (0 to 100). Computed by other models.
<b>Relevance_criteria</b>	Relevance to potential upward tier
<b>Store_state</b>	State where the store is located
<b>Recency_in_weeks</b>	Time from last purchase (as of December 31, 2023)
<b>Returns</b>	Monetary value of returns (2023)
<b>ID_Store_last</b>	Point of Sale of last purchase
<b>Gender</b>	Customer's gender
<b>Flagged</b>	Customer's card flagged as risky (potential cloned card)
<b>Promotional_percentage</b>	Percentage of items bought in promotion (value)
<b>Education</b>	Customer's education
<b>Checked_ok</b>	Flag for customer purchases exported from transactional database
<b>Beer</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Bottled_Water</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Bread</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Meat</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Dairy</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Fresh_Foods</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Frozen_Foods</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Fruit_Beverages</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Pastry</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Sodas</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Toiletries</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax

<b>Veggies</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax
<b>Wines</b>	Revenue from sales in the category (2023), in monetary units, net of sales tax

## Project Deliverables

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