

P1-Delivery1

Group-6

September 10, 2025

1 Members

- Font I Cabarrocas, Marc
- Garcia Ayala, Jesus
- Hernández Navarro, Arnau
- Khalipskaya Soboleva, Aglaya
- Mediavilla Jiménez, Alex Michel

2 Data Sources

The dataset used in this project was obtained from **Kaggle**, an online platform for data science competitions and open datasets. Specifically, it comes from the dataset “*Marketing Data*” uploaded by **Jack Daoud**. The dataset contains customer information, including demographic details, purchasing behavior, and responses to marketing campaigns, which makes it suitable for customer segmentation and predictive modeling. Link to the dataset: [Marketing Data on Kaggle](#).

3 Data Meaning

The dataset captures a wide range of customer-related attributes that collectively describe marketing and purchasing patterns within a selected group of individuals. It includes demographic characteristics (such as age, education, and marital status), financial information (such as income), and behavioral variables (such as product spending, purchase channels, and campaign responses). These variables are valuable for identifying customer segments, analyzing trends, and uncovering purchase preferences that can guide targeted marketing strategies.

4 Dataset Structure

The original dataset contains 2240 rows and 29 columns, representing different customer attributes.

4.1 Types of variables

The dataset includes:

- **Numerical (quantitative) variables:**

- Year_Birth
- Income
- Kidhome
- Teenhome
- Recency
- MntWines
- NumWebPurchases

- **Categorical (qualitative) variables:**

- Education
- Marital_Status
- Dt_Customer
- AcceptedCmp1 (0 = No, 1 = Yes)
- AcceptedCmp2
- Complain
- Response

- **Binary variables:**

- AcceptedCmp1
- AcceptedCmp2
- AcceptedCmp3
- AcceptedCmp4
- AcceptedCmp5
- Complain
- Response

4.2 Missing Values

- **Income:** 1.07% missing values
- **All other variables:** 0% missing values
- **Overall missing percentage in dataset:** 0.037%