

# Individual Company customer Data

## Abstract

Customer data analysis is a detailed analysis of a company's ideal customers. It helps businesses better understand their customers and makes it easier for them to modify products according to the specific needs, behaviors, and interests of different types of customers.

Customer data analysis helps the company to identify the types of target groups in terms of age, gender, residence, country,... etc. and to adjust its products based on its target customers from different types of customer segments. For example, instead of spending money to market a new product to each customer in the company's database, the company can analyze which customer segment is likely to buy the product and then market the product to only that specific segment

## Design

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project.

## Data

- They are the customer of an individual company and information about the customer in terms of family income has been collected to find out more ways to attract the customer and what ages, marital status and educational qualifications are affected by sales and through which statistics are made and predicted the most buying and most researched customers ,
- The dataset is provided in `.csv` ,and dataset includes about 40,000 rows and 15 columns  
Each row corresponds to a customer information, and includes the variables:
  1. flag: Whether the customer has bought the target product or not
  2. gender: Gender of the customer
  3. education: Education background of customer
  4. house\_val: Value of the residence the customer lives in
  5. age: Age of the customer by group
  6. online: Whether the customer had online shopping experience or not
  7. customer\_psy: Variable describing consumer psychology based on the area of residence
  8. marriage: Marriage status of the customer
  9. children: Whether the customer has children or not

10. occupation: Career information of the customer
11. mortgage: Housing Loan Information of customers
12. house\_own: Whether the customer owns a house or not
13. region: Information on the area in which the customer are located
14. car\_prob: The probability that the customer will buy a new car(1 means the maximum possible)
15. fam\_income: Family income Information of the customer(A means the lowest, and L means the highest)

## Algorithms

- When working on analyzing data, you'll likely come across data that is missing (also called null values or NaNs) .Data cleaning by used drop null value is an important part of the data analysis pipeline and making sure that it's all tidy up will make your analysis much stronger.
- rename column names is one of the most often applied data manipulations, rename name [val\_house ]to [value\_house]
- DataFrame will contain columns that are not useful to your analysis. Such columns should be dropped from the DataFrame to make it easier for you to focus on the remaining columns, Drop

## Tools

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting
- Tableau for interactive visualizations
- re for clean data
- nltk for natural language processing