Applied Data Science with Python

Course-End Project Problem Statement

Marketing Campaigns

Problem scenario:

Marketing mix stands as a widely utilized concept in the execution of marketing strategies. It encompasses various facets within a comprehensive marketing plan, with a central focus on the four Ps of marketing: product, price, place, and promotion.

Problem objective:

As a data scientist, you must conduct exploratory data analysis and hypothesis testing to enhance your comprehension of the diverse factors influencing customer acquisition.

Data description:

The dataset aligns with the Four Ps of Marketing, categorizing variables to analyze consumer behavior. Product-related variables track spending across categories, while Price factors like income and deal-based purchases indicate affordability. Place covers shopping channels and web visits, reflecting purchase preferences. Promotion measures campaign engagement, complaints, and recency. Additionally, demographics support segmentation for personalized marketing. This structured approach helps businesses optimize products, pricing, distribution, and promotions for better customer engagement and market performance.

Steps to perform:

1. After importing the data, examine variables such as Dt_Customer and Income to verify their accurate importation.

2. There are missing income values for some customers. Conduct missing value imputation, considering that customers with similar education and marital status tend to have comparable yearly incomes, on average. It may be necessary to cleanse the data before proceeding. Specifically, scrutinize the categories of education and marital status for data cleaning.

3. Create variables to represent the total number of children, age, and total spending.

   a. Derive the total purchases from the number of transactions across the three channels.

4. Generate box plots and histograms to gain insights into the distributions and identify outliers. Implement outlier treatment as needed.

5. Apply ordinal and one-hot encoding based on the various types of categorical variables.

6. Generate a heatmap to illustrate the correlation between different pairs of variables.

7. Test the following hypotheses:
   a. Older individuals may not possess the same level of technological proficiency and may, therefore, lean toward traditional in-store shopping preferences.
   b. Customers with children likely experience time constraints, making online shopping a more convenient option.
   c. Sales at physical stores may face the risk of cannibalization by alternative distribution channels.
   d. Does the United States significantly outperform the rest of the world in total purchase volumes?

8. Use appropriate visualization to help analyze the following:
   a. Identify the top-performing products and those with the lowest revenue.
   b. Examine if there is a correlation between customers' age and the acceptance rate of the last campaign.
   c. Determine the country with the highest number of customers who accepted the last campaign.
   d. Investigate if there is a discernible pattern in the number of children at home and the total expenditure.
   e. Analyze the educational background of customers who lodged complaints in the last two years.