Course-End Project: Marketing Campaigns

Problem Scenario: 'Marketing mix' is a popular concept used in implementing marketing strategies. A marketing mix includes multiple areas of focus as part of a comprehensive marketing plan. This all revolves around the four Ps of marketing - product, price, place, and promotion.

Problem Objective: As a data scientist, you should perform exploratory data analysis and hypothesis testing. The goal is to gain a better understanding of the various factors that contribute to customer acquisition.

Note: Download ***Data Dictionary – Response to marketing* campaigns.xlsx and marketing_data.xlsx dataset** from the course resource section in the LMS.

Data Description:
The variables birth-year, education, income, and so on are related to the first 'P' or 'People' in the tabular data provided to the user. The amount spent on wine, fruits, gold, etc., is related to 'Product'. The information pertinent to sales channels, like websites, stores, etc., is related to 'Place', and the fields which talk about promotions and results of different campaigns are related to 'Promotion'.

Steps to Perform:

- Once data is imported, investigate variables like Dt_Customer and Income, etc., and check if they are imported correctly.

- Income values for a few customers are missing. Perform missing value imputation. Assume that the customers with similar education and marital status make the same yearly income, on average. You may have to clean the data before performing this. For data cleaning, look into the categories of education and marital status.

- Create variables to populate the total number of children, age, and total spending.

  Hint: From the number of purchases through the three channels, people can derive the total purchases.

- Create box plots and histograms to understand the distributions and outliers. Perform outlier treatment.

- Use ordinal encoding and one hot encoding according to different types of categorical variables.

- Create a heatmap to showcase the correlation between different pairs of variables.

- Test the following hypotheses:

- Older people are not as tech-savvy and probably prefer shopping in-store.
- Customers with kids probably have less time to visit a store and would prefer to shop online.
- Other distribution channels may cannibalize sales at the store.
- Does the US fare significantly better than the rest of the world in terms of total purchases?

- Use appropriate visualization to help analyze the following:

- Which products are performing the best, and which are performing the least in terms of revenue?
- Is there any pattern between the age of customers and the last campaign acceptance rate?
- Which Country has the greatest number of customers who accepted the last campaign?
- Do you see any pattern in the no. of children at home and total spend?
- Education background of the customers who complained in the last 2 years.