**Personalized Ad- Analytics for small businesses**

Team: - Data Explorers

1. **Problem Statement: -**

With the emergence of giant e-commerce firms such as Amazon or retail corporations such as COSTCO and Walmart, it is extremely crucial for new growing or small businesses to use various marketing methods to promote their business. Marketing campaigns cannot always be tailored to each individual client, but they do have the flexibility to send many adverts combined with imagery to a diverse range of individuals. Family members may be drawn to several types of advertisements, whereas single working professionals may be drawn to commercials that interest them. The issue for marketing teams arises here; they may know who they are attempting to reach, but they want to know what handful of advertising they should create to appeal to a broad range of individuals without splintering their campaign into thousands of tailored ads, which may be prohibitively expensive.

1. **Data Source: -**

We are using the Kaggle data set: - [Kaggle | Marketing Analytics](https://www.kaggle.com/code/jennifercrockett/marketing-analytics-eda-task-final/data)

In this dataset holds the information related to customer and their purchases over 3 years from 2012- 2014. There are 28 features in the dataset and about 2300 tuples as shown in the below image

A picture containing table

Description automatically generated

1. **Methodology:**
   1. **DATA PRE-PROCESSING:**

Data available to us from Kaggle has some additional information which plan to clean, so that we will be able to proceed with analyzing the data and perform initial EDA on the same. For the first instance, the type of cleaning of data that we will need to perform are related to the following:

* + 1. Income – This variable currently consists of string data type value, e.g., $150, here after importing the data we will need to perform string manipulation and conversion operations so that we are able to use this column for numerical analysis
    2. Education – This variable currently consists of string data type value, e.g., Masters or PhD, here after importing the data we will perform data manipulation operations to the column to convert this column to ordinal data, which will refer to these values.
  1. **EDA: -**

We currently have sufficient data of customers, their purchase history and past campaign data. Our initial task will be to try and understand the data through various EDA techniques, also identify if any outliers are available that may help us identify any issues that might help with modelling.

* 1. **Challenges / Risks: -**
     1. Outlier detection for the type data available
     2. Survey information from customer for response to campaigns does not include historical information, can data be interpreted in a way that this relation can be found for tracking historical performance of all campaigns
     3. Pricing suggestions for regular and campaign sales, can the data provide information to interpret whether any product available currently can be priced higher or lower for better profit margins or greater quantity of sales.
     4. Purchase history of products and customers, can we provide analysis for customer and product relation based on purchase history whether there is trend of increasing or decreasing rate of purchase over the period
     5. Income categories, can we find meaningful insights about how income level affects purchasing power of customers and how they are related to various product types.
     6. Product bundling, as per given data can we find what kind of products available for purchase are more likely to be purchased together
     7. Future Campaign, can we provide meaningful insights based on existing campaigns and product purchase history that can lead to predicting what kind of products or offers can increase profitability of future campaigns.
  2. **Modelling: -**

We have customer data, relevant consumer purchase data, and previous campaign data.

One method would be to forecast if a consumer would utilize a specific offer in the forthcoming campaign and identify the key aspects that influence the amount of shop transactions. We may utilize the target variable 'Response,' which is the last campaign outcome, i.e., if the client accessed a certain offer in the previous campaign. To forecast the target variable, we may use a classification model. We may also use the Random Forest model to forecast the number of shop purchases, and then rank the characteristics that are strongly connected to the number of store sales using the feature importance algorithm.

Another option for marketers is to apply cluster analysis on the universe of people they want to target for their campaign. If you know ahead of time who your target's largest group is, you may divide them into clusters and then design campaign variants that appeal to each cluster. Cluster analysis is a type of unsupervised machine learning, as opposed to other types of machine learning issues in which we know what problem we want to answer. We will present the model a person with various features, and we want the model to tell us how likely a person with those traits is to buy our product; however, we do not know how many clusters are optimum for splitting our data or what criteria are appropriate for splitting them up by. We don't know what we want the model to learn from the data, thus we don't have a variable to act as the model's supervisor.(This data contains columns that show which campaigns the consumers responded to, but for the sake of implementing and evaluating clusters, we will remove those columns and only examine the data we would have on customers prior to an advertising campaign.)