# Marketing Campaign Analysis

## Understanding Factors Impacting a Campaign Strategy for Increased Campaign Effectiveness

**Group Members**

- Abhinav Chadha
- Puneet Maini
- Rebecca Bland
- Sanjeet Walia
- Siddharth Sharma

# Marketing Campaign Analysis

## Q1 What is the problem you are seeking to answer?

We want to identify potential opportunities to **drive up the customer campaign engagement** in the given dataset. Only about 27% of the target audience accepted any of the Maven Analytics Campaigns (campaign performed 6 times). By identifying how the other 73% of the target audience can better be reached, new advertising methods or messages can be utilized to achieve higher campaign engagement. In the given dataset we found that customers engaging with the campaign have an average sales ticket size twice that of the customers who are not engaging with the campaign.

We have undertaken the following steps to identify potential opportunities to increase customer campaign engagement:

1. Understand the existing campaign engagement (along with average sales ticket) for the target audience (all ~2200 customers)
2. Find customers who did not engage with any of the 6 campaigns and perform a demographic analysis on them to create customer clusters
    1. Create sub-groups of non-engaged customers to identify low, medium, and high spenders.
    2. Understand demographics within each sub-group to better create a campaign specifically for each of the three sub-groups of people who did not accept campaigns.
    3. Perform regression analysis to identify if real campaign improvement could be observed in all three spending sub-groups from the new campaigns.
3. Suggest new marketing strategies (campaigns) for the ~1600 people who did not engage with the campaign initially in order to drive up campaign engagement within this cohort.
4. Perform regression analysis to understand the factors driving the success of the campaigns for the cohort that did accept the campaign(s)

## Q2 Why would a data-driven approach be useful to obtain insights regarding this problem?

Without utilizing a data driven approach to obtain insights regarding this problem, we would be unable to identify specific campaigns for our target audience.

About 73% of the customers did not accept any of the campaigns. Without understanding their demographics and spending patterns, it would be difficult for us to create an effective campaign for each grouping of customers. We would only be able to produce a 7th campaign and hope that it works -- without any real understanding of who we are targeting and how they can best be reached. It is crucial to utilize a data driven approach when answering problems, not only like this, but in general.

**Data and the informational insights provides guided decision making and can save companies a lot of money by creating more effective campaigns.**

## Q3 What is the ideal experiment to examine the issue?

At present, the effectiveness of existing marketing campaigns stands at 27.14%, meaning 27% of the people responded to them. (605 out of a total of 2229)

Having analyzed the data, we recommend certain marketing campaign strategies to target the customers who haven't responded to any of the campaigns so far. (1624 customers). The ideal experiment would be to test these new campaigns on the entire 2229 customer database to see if the marketing effectiveness has increased.

## Q4 Data (1/2)

### A. Where did you obtain your data?

Maven Analytics: https://www.mavenanalytics.io/data-playground?search=marketing

### B. Are you using multiple data sets? If yes, how are you combining the data?

No, we are using a single data set.

### C. What is the aggregation level of your data?

There is a low/medium aggregation level in the data. Each row in the data represents a single customer within data. The data is partially aggregated to the customer level, as the total sales per customer was not aggregated and provided. The only aggregation that is performed in the data is the total sales (in $) for each product category over the two-year period, the number/sum of purchases made by each of the three campaign channels (In store, Web, Catalog), and the sum of website visits over the last month.

### D. Are you using multiple data sets? If yes, how are you combining the data?

Each row in the data represents a single customer that was presented with 6 campaigns over the course of 2 years.

## E. What is the outcome variable(s) that you are interested in?

**Aggregated column**: Total Sales (sum of the following)

1. MntWines: $ amount wine purchases

2. MntFruits: $ amount fruit purchases

3. MntMeatProducts: $ amount meat purchases

4. MntFishProducts: $ amount fish purchases

5. MntSweetsProducts: $ amount sweets purchases

6. MntGoldProds: $ amount gold purchases

**Q5 Present descriptive statistics describing your data. Outline data discrepancies that are important. These could include outliers, missing data, erroneous data, etc. Describe how you handled these issues.**

The data consisted of 2240 customers and their purchases over the past two years. Along with this, we also have their deals purchases, web purchases, store purchases and Catalog purchases over the time period. We also have information about their campaign acceptance during this time period.

**Deals purchases, Web purchases, Store purchases and Catalog purchases**

|  | NumDealsPurchases | NumWebPurchases | NumCatalogPurchases | NumStorePurchases | NumWebVisitsMonth |
|---|---|---|---|---|---|
| count | 2240.000000 | 2240.000000 | 2240.000000 | 2240.000000 | 2240.000000 |
| mean | 2.325000 | 4.084821 | 2.662054 | 5.790179 | 5.316518 |
| std | 1.932238 | 2.778714 | 2.923101 | 3.250958 | 2.426645 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 1.000000 | 2.000000 | 0.000000 | 3.000000 | 3.000000 |
| 50% | 2.000000 | 4.000000 | 2.000000 | 5.000000 | 6.000000 |
| 75% | 3.000000 | 6.000000 | 4.000000 | 8.000000 | 7.000000 |
| max | 15.000000 | 27.000000 | 28.000000 | 13.000000 | 20.000000 |

**Q5 Present descriptive statistics describing your data. Outline data discrepancies that are important. These could include outliers, missing data, erroneous data, etc. Describe how you handled these issues.**

**Campaign Acceptance**

| AcceptedCmp1 | AcceptedCmp1 | AcceptedCmp2 | AcceptedCmp3 | AcceptedCmp4 | AcceptedCmp5 | Response |
|---|---|---|---|---|---|---|
| 0 | 2096 | 2096 | 2096 | 2096 | 2096 | 2096 |
| 1 | 144 | 144 | 144 | 144 | 144 | 144 |

**Customer Demographics include Education, Income, Age, Marital Status, Kids and Teens at home, and their Country.**

**Education level**

| Education | Education |
|---|---|
| 2n Cycle | 203 |
| Basic | 54 |
| Graduation | 1127 |
| Master | 370 |
| PhD | 486 |

**Income and Age**

| | Income | Age |
|---|---|---|
| count | 2216.000000 | 2240.000000 |
| mean | 52247.251354 | 52.194196 |
| std | 25173.076661 | 11.984069 |
| min | 1730.000000 | 25.000000 |
| 25% | 35303.000000 | 44.000000 |
| 50% | 51381.500000 | 51.000000 |
| 75% | 68522.000000 | 62.000000 |
| max | 666666.000000 | 128.000000 |

**Marital Status**

| Marital_Status | Marital_Status |
|---|---|
| Absurd | 2 |
| Alone | 3 |
| Divorced | 232 |
| Married | 864 |
| Single | 480 |
| Together | 580 |
| Widow | 77 |
| YOLO | 2 |

**Kids & Teens**

| Kidhome | Kidhome |
|---|---|
| 0 | 1293 |
| 1 | 899 |
| 2 | 48 |

| Teenhome | Teenhome |
|---|---|
| 0 | 1158 |
| 1 | 1030 |
| 2 | 52 |

**Country**

| Country | Country |
|---|---|
| Australia | 160 |
| Canada | 268 |
| Germany | 120 |
| India | 148 |
| Mexico | 3 |
| Saudi Arabia | 337 |
| Spain | 1095 |
| USA | 109 |

## Q5 Present descriptive statistics describing your data. Outline data discrepancies that are important. These could include outliers, missing data, erroneous data, etc. Describe how you handled these issues.

### Outliers:

- Marital Status: Remove Absurd, Alone, YOLO (7 Customer IDs)

- Age: Exclude all age above 100 (3 outliers)

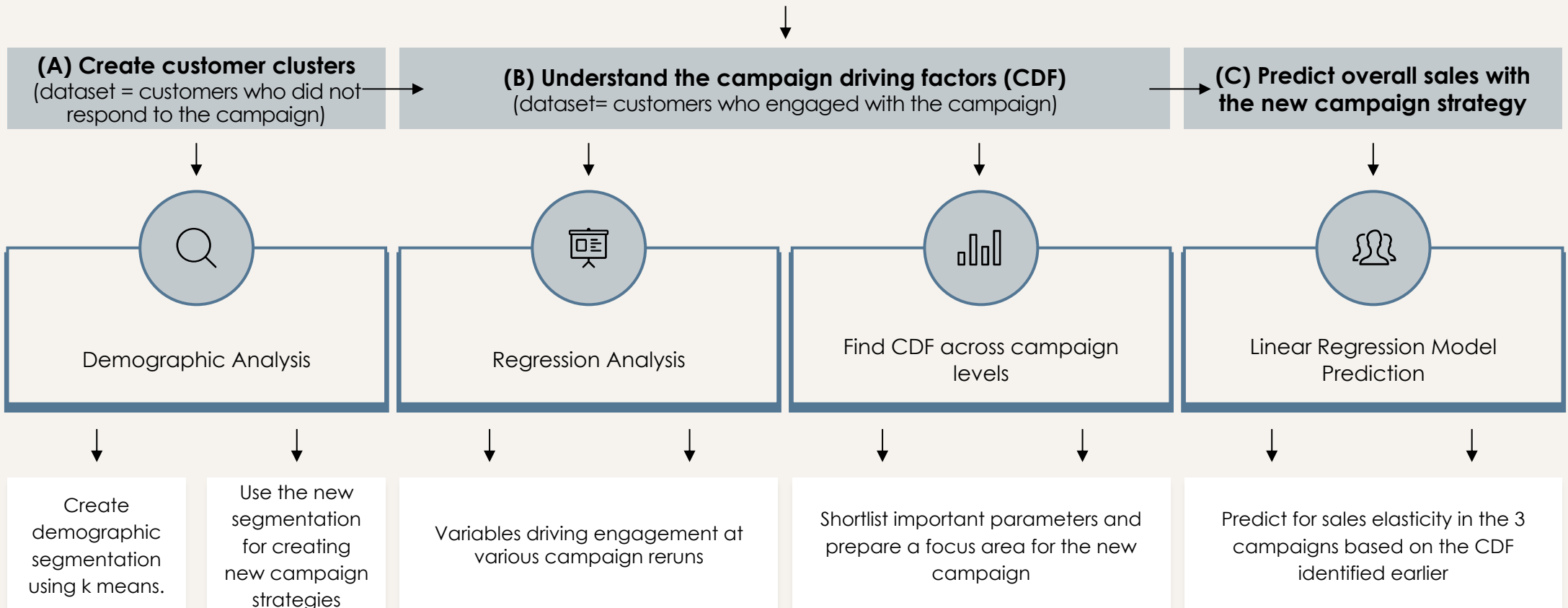- Income: Exclude Customer ID with Income 6666666 (1 outlier)

### Missing Data

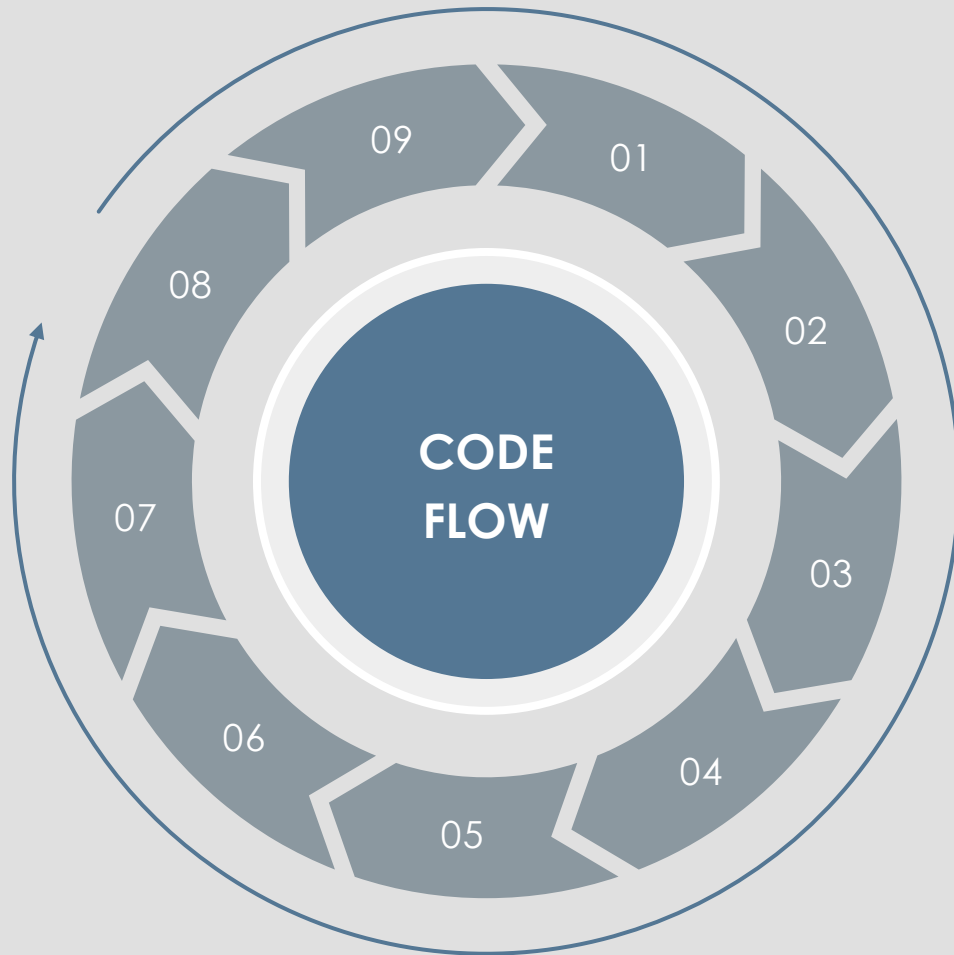For Customer IDs with no income level, we fill the missing values with the median income based on their Country's Median Income

# PROJECT MIND-MAP

## CUSTOMERS DATA SET

*(study the engagement of the campaign and suggest revised campaign targeting strategies to drive up the campaign engagement)*

**(A) Create customer clusters**
(dataset = customers who did not respond to the campaign)

**(B) Understand the campaign driving factors (CDF)**
(dataset= customers who engaged with the campaign)

**(C) Predict overall sales with the new campaign strategy**

Demographic Analysis

Regression Analysis

Find CDF across campaign levels

Linear Regression Model Prediction

Create demographic segmentation using k means.

Use the new segmentation for creating new campaign strategies

Variables driving engagement at various campaign reruns

Shortlist important parameters and prepare a focus area for the new campaign

Predict for sales elasticity in the 3 campaigns based on the CDF identified earlier

# CODE FLOW

**01.** **Data Segregation**: Find customers who accepted the campaign at any level vis-à-vis customers who didn't

**02.** **Create Demographics Variables:** (Education levels, Marital Status, Income levels, Age levels).

**03.** **Demographic clustering using k-mode**

**04.** **Customer segmentation based on k-mode clustering**

**05.** **Channel usage by customer data using linear regression**

**06.** **Linear regression to find prevalent campaign driving factors**

**07.** **Linear regression analysis to determine prevalent channel usage by customers**

**08.** **Linear regression analysis to create a prediction model for various campaigns**