

Team B5

Bolded and blue words: Updates since Proposal

Intro: To enhance the strategic planning of marketing campaigns, our group wants to look into the ~~likelihood of consumers to purchase~~ **factors that affect whether or not a consumer purchases** an unspecified product **within the energy sector** utilizing predictive modeling. By analyzing customer responses from the previous campaigns and their demographic backgrounds, we aim to predict the ideal target consumer. With this information, companies can either expand or narrow their target audience when issuing marketing strategies and better utilize capital for marketing success.

1. What kinds of projects does your team want to explore?

"Forecasting the Target Market: Analyzing Factors Influencing New Customer Purchase Decisions"

'We think that this will be helpful to companies who want to increase their sales through targeted marketing campaigns as well as save money by being accurate.

At the end, based on the factors and target audience, we can also make an assumption on what the product is (since the product information is not given in the original dataset)

2. Datasets you have explored:

We found a dataset from Kaggle. The dataset contains rich and varied information about customers' demographics and their reactions to marketing campaigns. Moreover, this dataset offers us exploring how different characteristics of customers influence their response to marketing activities. It would be valuable to know the patterns and trends that can help in tailoring marketing strategies more effectively to meet customer needs and preferences.

Background:

<https://www.kaggle.com/code/khanimar/bi-marketing-campaign-eda-analysis-prediction>

Training Dataset:

<https://www.kaggle.com/code/khanimar/bi-marketing-campaign-eda-analysis-prediction/input?select=train.csv>

3. Any preliminary stories you would like to explore of the chosen dataset?

As today's market continues to progress, incorporating more digitized marketing strategies and international competitors, it has become more important than ever to understand consumer behavior and adjust one's targeting strategies to keep an edge over the growing competition. Our project aims to analyze marketing and customer data to predict factors influencing new customer conversions, including demographics like a customer's age, gender, and education level. By employing advanced data analytics techniques, we seek to uncover patterns, trends, and correlations that may guide future marketing targeting strategies by improving conversion rates, returns on investments (ROI) and gaining better insights on ideal time frames and target audiences.

4. Any preliminary findings/insights of the chosen dataset?

- 31480 observations(rows), 20 features(columns)
- 9 numerical features, 11 categorical features
- Turns out some numerical features are actually categorical features such as ID, Contact ID.

- Need to convert some variables such as target and gender into dummy variables using one hot encoded code.
- Figure out if there are any missing values using info() function or is.na() function.
- drop the rows or columns depending on the number of missing values on the columns in order to obtain valid insight.
- The target column is the most crucial for our analysis because it records the amount of customers who end up purchasing the products or not. We are planning to use stratified sampling of the datasets by taking 3700 success conversions and taking 3700 out of 27.8k failure conversions to make the dataset balanced.
- **It may be effective to use classification or clustering for this data since intersectionality may have a greater effect on consumer decision making than individual demographic factors**
- **Use Exploratory Data Analysis to uncover underlying patterns and insights within the marketing campaign data.**
- **Model could be Naïve Bayes**
- **Look into predictive power of house/credit/education because they might be related**
- **Predicting whether a sale is made**
- **Find background information for products or just hone it on some other predictions that do not need background information. Use subset, hone in one marketing campaign**
- **Bin day and month**
- **Job and day of the month-end of the month is payday and might be more likely to create an account or buy**
- **Drop daysincelastcampiagn because of missing data and that it contradicts, duration could still be ongoing**
- **Address in presentation that we had a lot of 0s which might lead to imbalance**
- **Days since last campaign column (82% of data is empty, but we decided to keep it because the empty ones mean that they are new targets(current campaign only))**
 - **-1 means no contact in the previous campaign (means they are the new targets now)**
- **Renamed the “target” variable to “purchase” variable so it is clearer**
- **Duration is the call time with the potential customers in seconds**
- **EDA: created visualizations showing distribution for target variable “purchase” and found that the target variable is imbalanced**
 - **Created heatmap from numerical variables, no strong correlation is founded**
-

Furthermore, we want to provide the best possible predictive model for the marketing campaign of their new product which shows if a customer buys the new product or not and how much is the possibility of the purchase.

