



BIA | BOSTON[®]
INSTITUTE OF
ANALYTICS

Prediction Product AD Campaign Performance

Presented by : Patoliya Yash

Introduction

- An advertising campaign is a set of advertisements that work together to promote a product or service. An ad campaign is designed around a specific and unique theme to create brand awareness about the company's product or service.
- An advertising campaign can be a series of different individual ads or the same ad across mediums used to create awareness and interest in a product or service.
- This is achieved through different forms of media, including radio, television, print advertising, direct mail, or the internet.

Dataset Information

Here are the key details about the dataset used in this project:

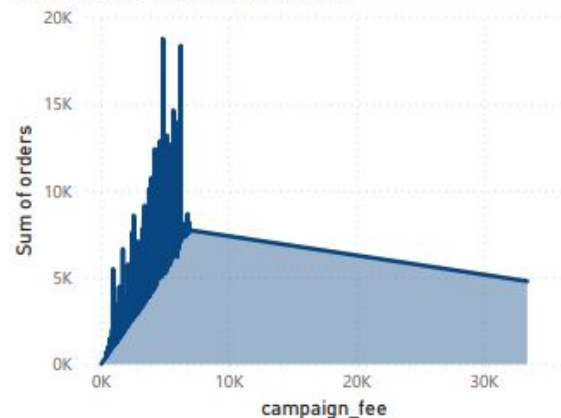
- Our has 731 entries and 11 columns. The columns include
- There are three columns with float data types and eight with integer data types.
- The dataset has 5 categorical variables namely limit_infor, campaign_type, campaign_level, product_level, resource_amount whose values have been represented by single digit numbers. Here, the target variable is **"orders"**

Exploratory Data Analysis (EDA)

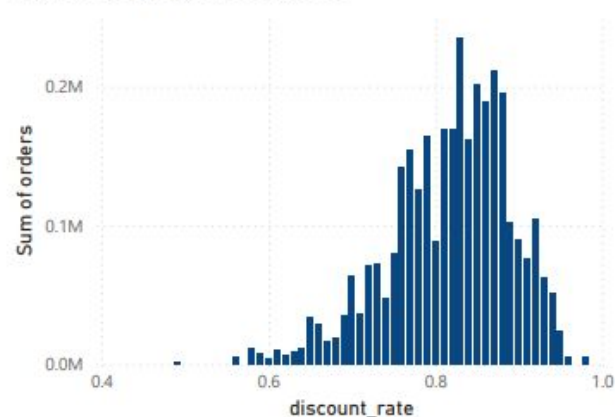
- EDA is used to provides a provides a better understanding of data set variables and the relationships between them.
- The dataset had no duplicates and 2 missing values in "price" which was 0.27% of total data. Hence, the rows with missing values were dropped.
- While observing the relationship of numeric values with "orders", the campaign_fee had one outlier which was removed for a cleaner data.

products campaign sales

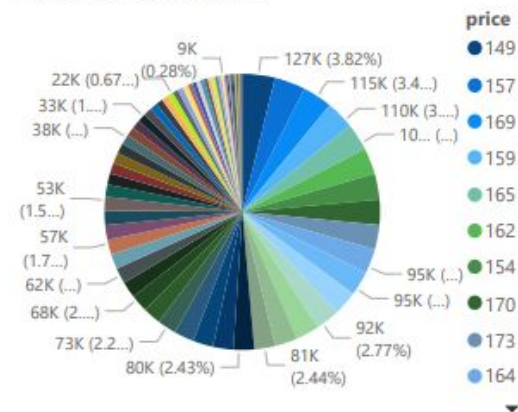
Sum of orders by campaign_fee



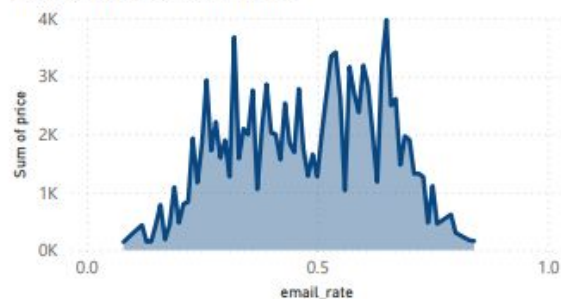
Sum of orders by discount_rate



Sum of orders by price



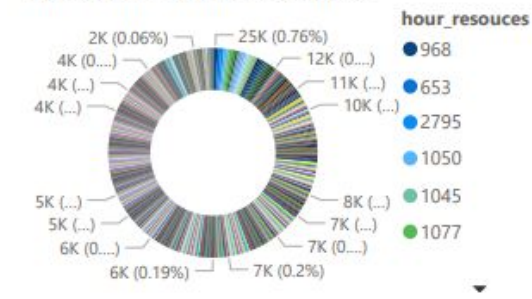
Sum of price by email_rate



Sum of product_level by discount_rate



Sum of orders by hour_resouces



Splitting the data into X and y

- In this step, we divided the dataset into two parts: X and y.
- X contains all the independent variables, which are the features used to make predictions.
- Meanwhile, y represents the dependent variable or target variable, which is the outcome we want to predict.

Train-Test Split

- The dataset was split into training and testing sets.
- An 80:20 ratio was used, with 80% of the data allocated to training and 20% to testing, and the test size set to 0.2.
- A random state of 42 was specified to ensure the reproducibility of results across different runs

Model Selection

The Prediction Product AD Campaign Performance is a regression problem. Hence following models were used:

- **Linear Regression** is best for simple, linear relationships and offers high interpretability.
- **Random Forest** is powerful for complex, non-linear relationships and provides robust performance but is less interpretable and more computationally intensive.

Predictions

Linear Regression		Random Forest Regressor	
RMSE on Train Score:	0.04030	RMSE on Train Score:	0.0539
RMSE on Test Score:	0.04055	RMSE on Test Score:	0.0480
Difference between RMSE on train and test set	0.00025	Difference between RMSE on train and test set	0.00025

Conclusions

- The analysis and predictions provide valuable insights that can significantly enhance the ad campaign performance for the company.
- **Campaign Fee and Hour Resources:** Increasing campaign fees and allocating more resources positively influence the number of orders, suggesting that investments in these areas are likely to yield higher returns.
- **Pricing Strategy:** Higher prices tend to reduce the number of orders. Therefore, maintaining competitive and minimal prices can attract more customers and boost sales.