Business Decision Using Data Analytics

By Prof. Nimesh Patel

Last update: 28/02/2022

Project III (CT805-N) by:

Lukhi Yashkumar Rajeshbhai (18BECE30561)

Dodiya Mohit Naranbhai (18BECE30106)

Saumay Joshi (18BECE30555)

Table of content

Problem Statement

We're a marketing analyst and we've been told by the Senior Marketing Manager that recent marketing campaigns have not been as effective as they were expected to be. We need to analyze the data set in order to understand this problem and propose data-driven solutions. We are required to solve the following questions to generate a report for your management

Section 01: Exploratory Data Analysis

- Are there any null values or outliers? How will you wrangle/handle them?
- Are there any variables that warrant transformations?
- Are there any useful variables that you can engineer with the given data?
- Do you notice any patterns or anomalies in the data? Can you plot them?

Section 02: Statistical Analysis

What factors are significantly related to the number of store purchases?

Does US fare significantly better than the Rest of the World in terms of total purchases?

Your supervisor insists that people who buy gold are more conservative. Therefore, people who spent an above average amount on gold in the last 2 years would have more in store purchases. Justify or refute this statement using an appropriate statistical test

Fish has Omega 3 fatty acids which are good for the brain. Accordingly, do "Married PhD candidates" have a significant relation with amount spent on fish? What other factors are significantly related to amount spent on fish? (Hint: use your knowledge of interaction variables/effects)

Is there a significant relationship between geographical regional and success of a campaign?

Dataset

- Source: Kaggle open source datasets
- License: CC0: Public Domain
- Usability: 9.4
- Size: 817kb
- Dimensions: 2206 x 39
- File type: CSV

Patterns or anomalies

- Almost 50% of clients' education level is graduate, and few customers have an primary level of education.
- The number of married clients is more than widow and divorce.
- There is a remarkably high percentage of customers in Spain while the percentage of clients in the United States and Montenegro is very small.
- There is a very high percentage of clients between 39 to 59 years old compared to other age groups

Correlation (From Notebook's heatmap)

We can see the correlation of numerical columns/decorations on 'NumStorePurchases'. The columns that have clear correlation (high positive or high negative) are important for the prediction model, but few of those with small (about zero) correlation will not have much effect on the 'SalePrice', therefore, we can still drop few of them.

Regression Models Analysis & Prediction (Answers)

From the graph, we can clearly see that the US is on second last ranked in terms of total purchases

From boxplot we can see that people who spent an below average amount on gold have less in store purchases.So it means that the given statement is correct.+4

From the graph we can see that Married PhD customers does not have a significant relationship with the amount spent on fish.

From the analysis we can see that the success of a campaign has no relation to country.


Thank You!