

Applied Data Science Project - Marketing Analytics

Saymon Mameza, Alisha Aery, Pranaykumar Laxman Pagdhare Jr

Olga Scrivner, PhD

FA21: APPLIED DATA SCIENCE: 22077

10 December 2021

Links

App: <https://datasciencemarketing.shinyapps.io/ADSMarketingDashboard/>

Github: https://github.com/aeryali/ADS_project

Research Objective

Our research objective is to utilize advanced analytics to transform data into useful information in the field of marketing. Our five-step approach is to define the problem, clean the data, explore the data, build a model, and deploy the model. After careful consideration of the problem, we came to a consensus regarding the issue of increasing the amount of recurring web visits to the site.

Project Outline

Data Description

Our data originates from a Kaggle competition located at this link:

<https://www.kaggle.com/jackdaoud/marketing-data>. The dataset contains 2,216 rows and 26 columns. Of the 26 columns, 2 columns are categorical, 10 columns are nominal, and 14 columns are continuous variables. The data consists of information on customers like the year they were born, their education, marital status, income, country, and how many kids/teens they have at home. The data also has information on the amount the customer spent on products like wine, fruits, meat, fish, and sweets. In addition, we are given data on where the items were purchased (in store, catalog, or online) and their use of discounts. The data was collected by Jack Daoud who is a Data Integration Specialist at Ellevation Education in Boston, Massachusetts. The information on how the data was collected is unknown.

Purpose

The main purpose of our web app is to help a user understand how many web page visits the company can expect given a customers' income and how much they have spent on wine.

However, there are tabs included to help the user better understand the data behind the predicted web page visits. For example, we have included tabs where the user can explore the data and understand the correlation between certain variables.

App Layout and Design

In the original plan for our project, we were planning to use Flask. However, due to complexity of the application we decided to use Shiny instead. The 5 data science tasks we performed were data cleaning, exploratory data analysis, visualization, correlation (statistical) analysis, and regression.

The application has been divided into 4 tabs as follows :

1. **About** : This tab provides information about the application name along with its name
2. **Exploratory Data analysis and Visualization** : This tab shows a summary of EDA performed on the data by showing summary of Column selected in the lookup. For the selected column, the application also shows its histogram as a visual aid.
3. **Correlation Analysis** : This tab shows correlation between different products available for marketing vs income of the customer. Thus we can check if the product has any correlation with earnings of the customer. The page is divided into a side panel and 2 tabs. The 1st tab will show a correlation plot for each product vs income based on your selected product in the side panel. The 2nd tab has been created to showcase raw and provide summary correlation in each column.
4. **Predicted web page visits** : This page will give prediction insights on number of customer visits based on 'Amount Income' and another product (wine in this case).

The layout is based on Shiny using R. The data for this application is sourced from the GitHub location used during this project.

Challenges

While working on this dashboard, we encountered several challenges that we ultimately overcame. These challenges included our team's minimal knowledge of R and shiny, combining the code for each section we worked on, working with people across different time zones, and funnily enough, figuring out how many parentheses and brackets were needed to run the code. Though our team did not have much experience with R, we were able to look back at previous datacamp assignments to refresh our memory.

During our zoom meetings, we would try combining the code but had a bit of trouble running it without errors. As a team, we were able to work through these issues and successfully run our code. Since our team was located in different states, we had trouble finding the right time for our meetings because we had to work around our job schedules. However, we were able to overcome the challenge through communicating and setting designated time for the project in advance. Completing this dashboard was not as easy as we would have hoped. However, the challenges we encountered along the way helped us better understand front end development and project execution.