

Capstone proposal

A study of sales through consumer behaviors

What is the problem I want to solve?

Based on sales observations in a retail store during Black Friday, make decisions on strategy of targeted advertising. For a specific group of people, predict the categories of products they are most likely to buy and include the categories in the advertising email or other advertising method.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have otherwise?

The client of this analysis would be the owner of the retail store. With the results, the client would be able to determine their advertising strategy to promote sales. They would be able to decide target customers for different categories of products.

What data are you going to use for this? How will you acquire this data?

The dataset for this project comes from <https://www.kaggle.com/mehdidag/black-friday>. The data consists of 538000 observations on Black Friday sales in a retail store.

Variables

|User_ID|

|Product_ID|

|Gender| Sex of user

|Age| Age in bins

|Occupation| Occupation of user

|City_Category| Category of the City (A, B, C)

|Stay_In_Current_City_Years| Number of years stay in current city

|Marital_Status| Marital status of user

|Product_Category_1| Product category

|Product_Category_2| Product may belong to other categories also

|Product_Category_3| Product may belong to other categories also

|Purchase| Purchase amount in dollars

The Approach

1. Data wrangling and cleaning
Handle with any missing values in the dataset
2. Exploratory Data Analysis
Through data visualization, determine potentially significant features that affect categories of products
Are there any relationships among different features?
3. Machine Learning
Use different algorithms to establish classification model correlating customer information with category of product
4. Data Visualization and Report
Summarize the model and results

Deliverables

My deliverables will consist of a report, a slide deck, and python code used in analyzing the data.