

Customer Purchasing Pattern Study Based on Historical Data

Team 1

Qianrong Wen, Ting-Ann Lu, Risheng Guo,
Luke Hong, Kexi Pi, Yesol Lee (Sally)

Content

1. Project overview
2. Data cleaning
3. Insights from data
4. Customer persona and machine learning
5. Conclusion

Project overview

Data set: Marketing Analytics (<https://www.kaggle.com/jackdaoud/marketing-data>)

Data source: Dr. Romero-Hernandez (a professor at UC Berkeley and Hult International Business School)

Business problem: to find relationships between customers' characteristics and the amount they spend at a retail store.

- Average customer persona
- Customers' purchasing patterns
- Evaluate the channel efficiency

Data Cleaning

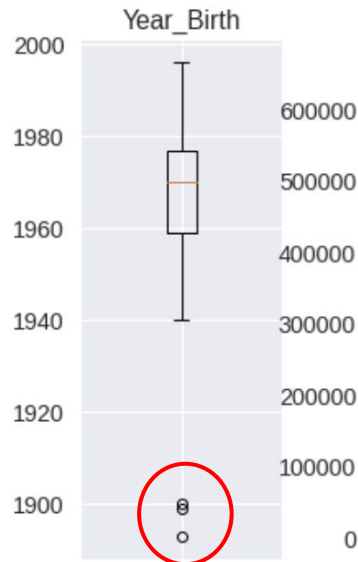
Column Name

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis sit amet odio vel purus bibendum luctus. Morbi iaculis dapibus tristique. In hac habitasse platea dictumst. Mauris convallis quam at. Morbi iaculis dapibus tristique.

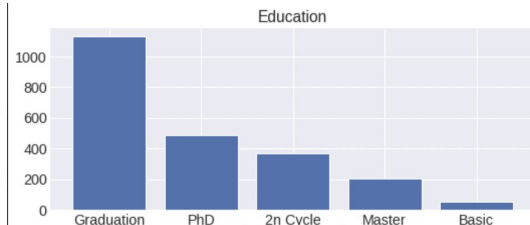
Missing values

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis sit amet odio vel purus bibendum luctus. Morbi iaculis dapibus tristique. In hac habitasse platea dictumst. Mauris convallis quam at. Morbi iaculis dapibus tristique.

Impossible Data

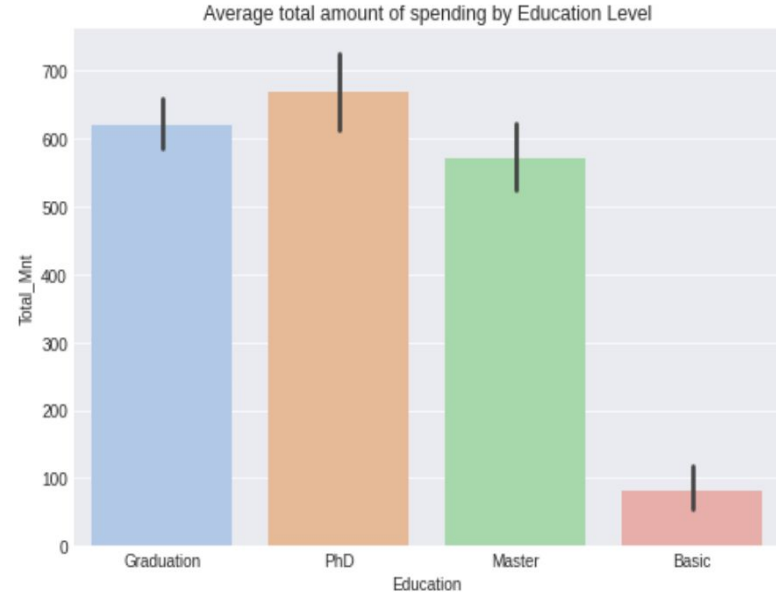
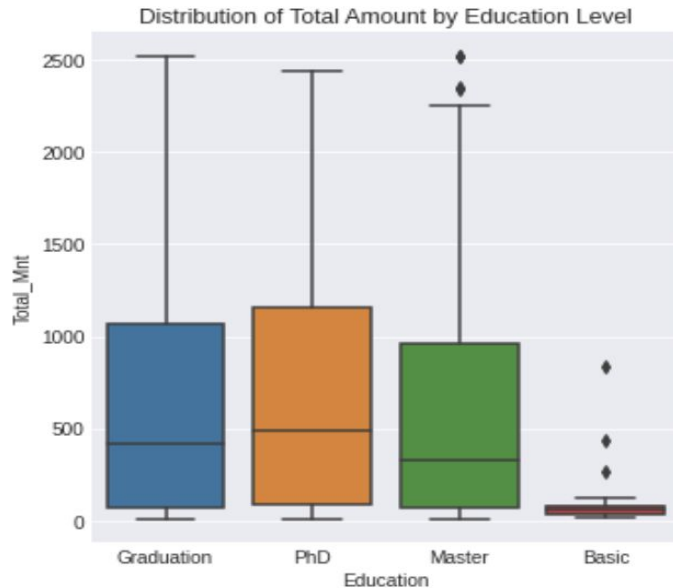


Duplicate categories



Education v. Total Amount

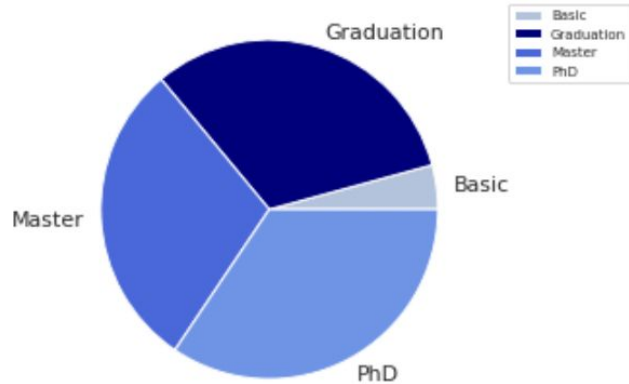
- Median of graduation, PhD and Master are similar
- Basic spent the least in average than other education levels



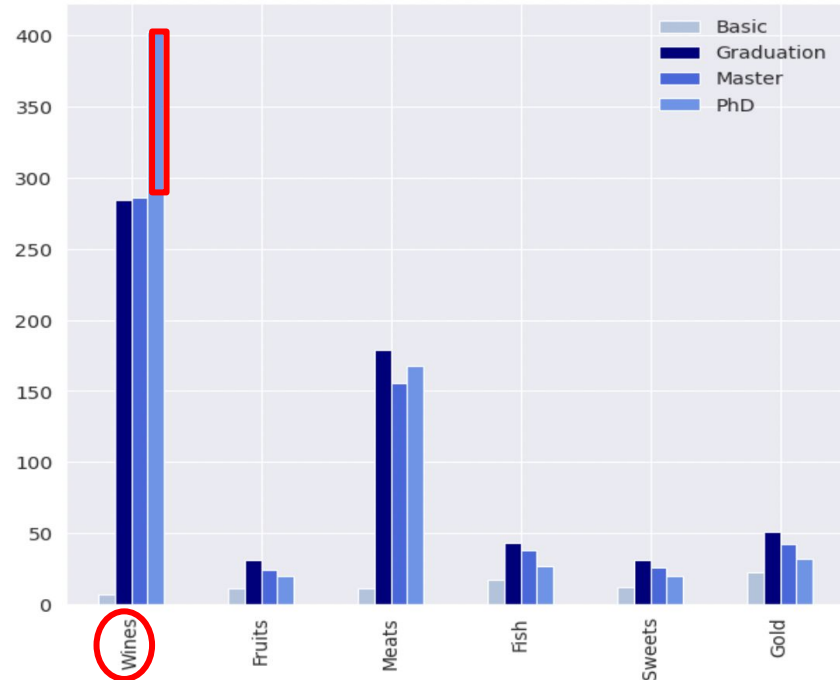
Education Level Spent Amount by Products

- PhD spent significantly more than other education levels on wines

The proportion of education levels in total amount

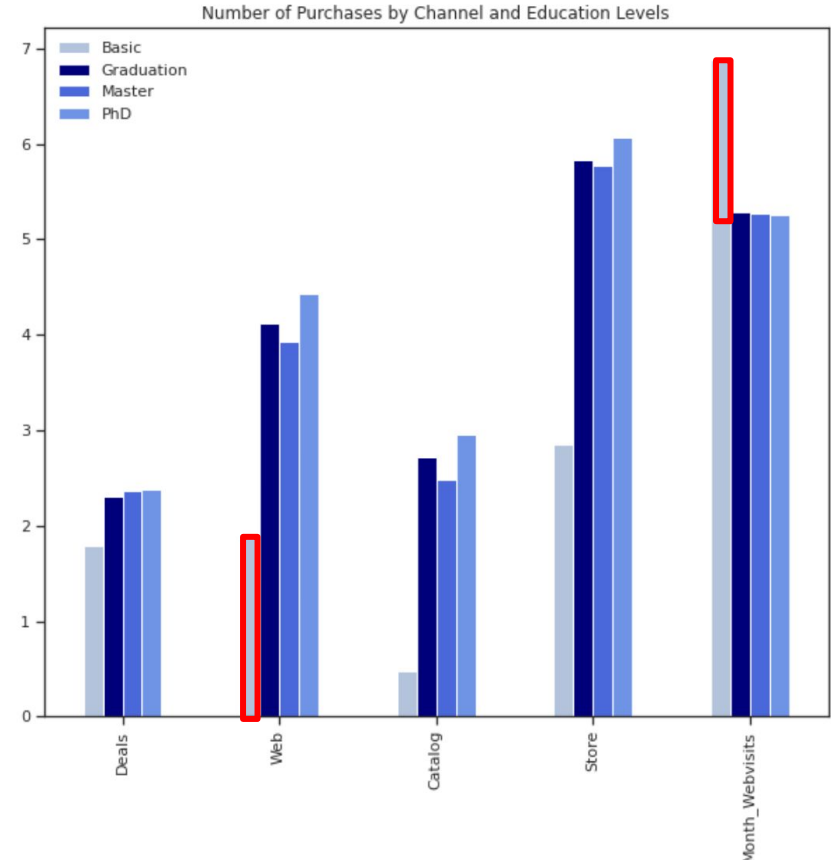


Amount Per Person by Product Category



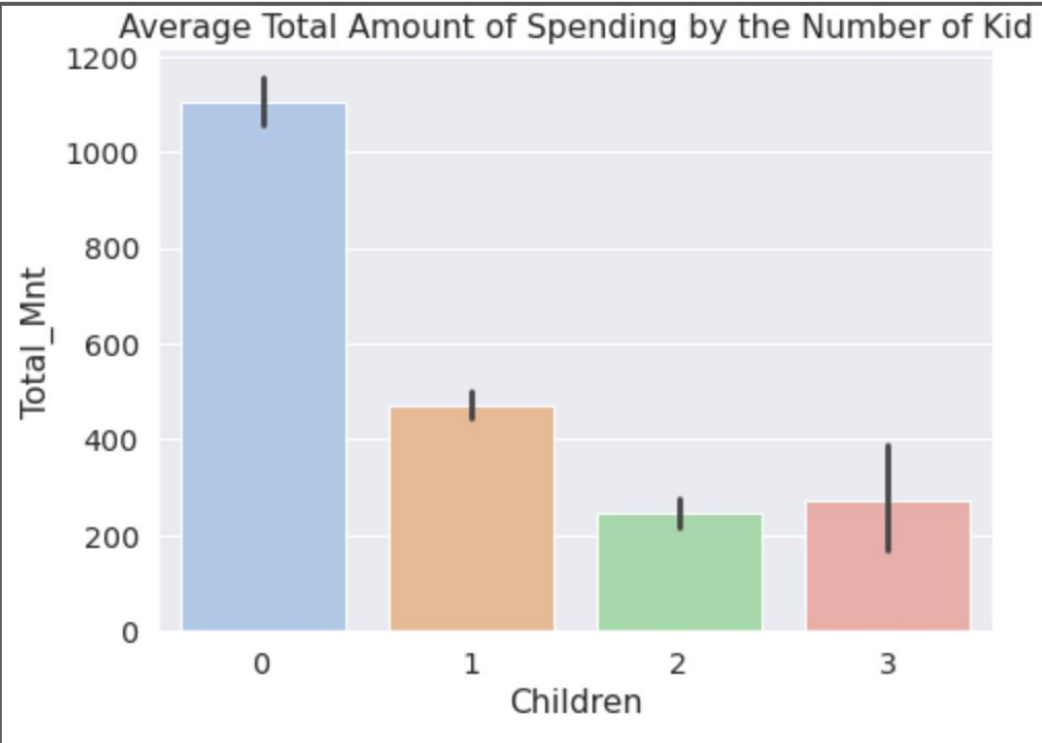
Education Level Spent Amount by Channels

- PhD, who may has higher economic ability, tends to purchase in store
- Basic, who shows the lowest purchasing power in each channel, has highest number on web visits



Suggestions

Number of Children with the Amount of Money Spent



- Assumption: positive relationship
- Result: roughly negative relationship
- Reason:

Different components of purchase

0 kid: wine+gold

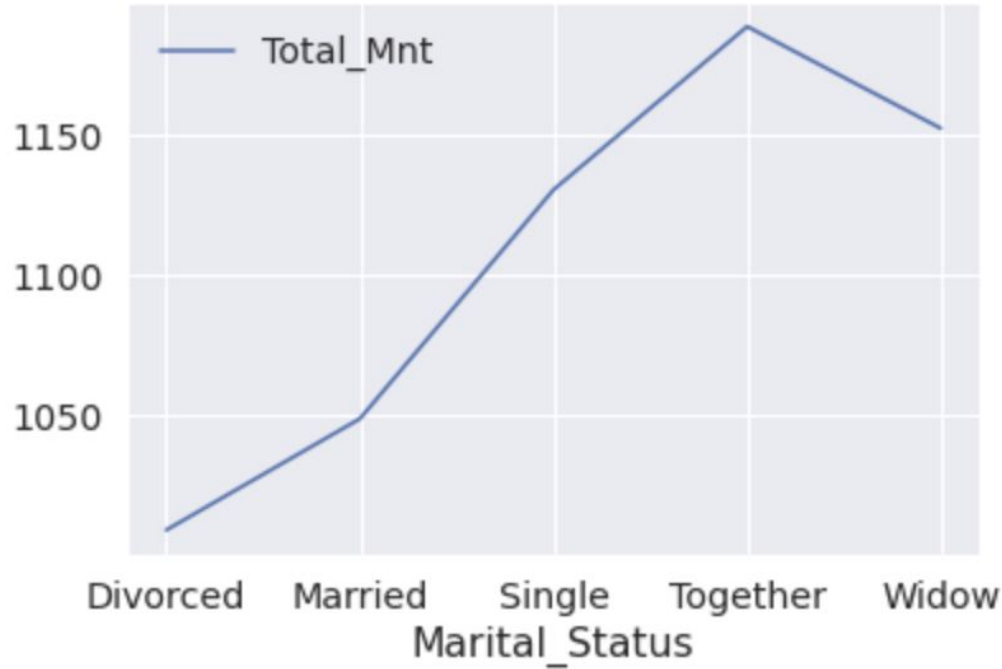
0+kids: utility

Number of Children with the Amount of Money Spent

	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	Total_Mnt
Children							
0	486.990566	52.295597	372.187107	76.338050	53.026730	63.778302	1104.616352
1	266.806050	19.409253	98.643238	26.684164	20.312278	40.824733	472.679715
2	141.591449	7.878860	51.299287	11.387173	8.370546	25.420428	245.947743
3	171.377358	6.905660	64.018868	7.075472	6.622642	18.603774	274.603774

- Data: mean spending on each of the categories
- No kids: wines + meat + gold
- With kids: less gold purchase

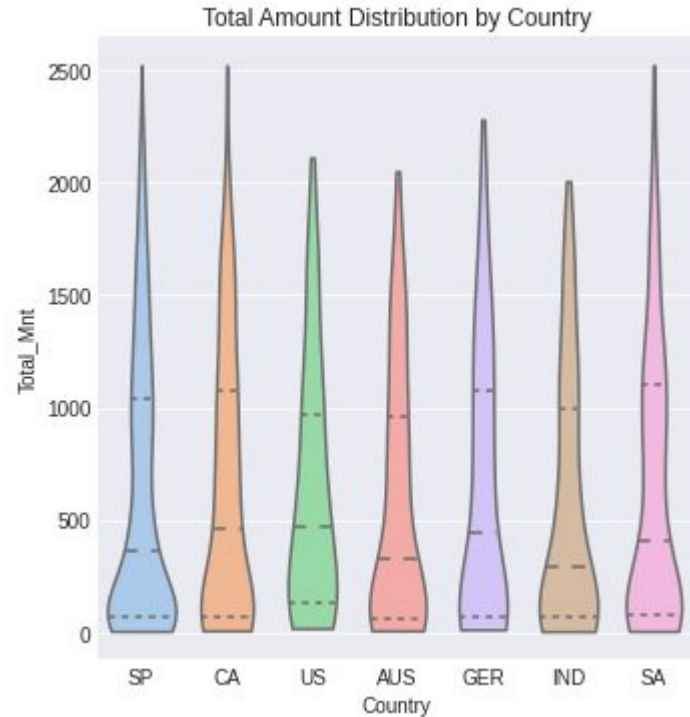
Marital Status with the Amount of Money Spent



- Relationship of spending with marital status (0 kid)
- Peak: Together
 - Stable income
 - Romantic spending
- Trough: Divorced
 - Lack of stable income source

Total Amount vs Country

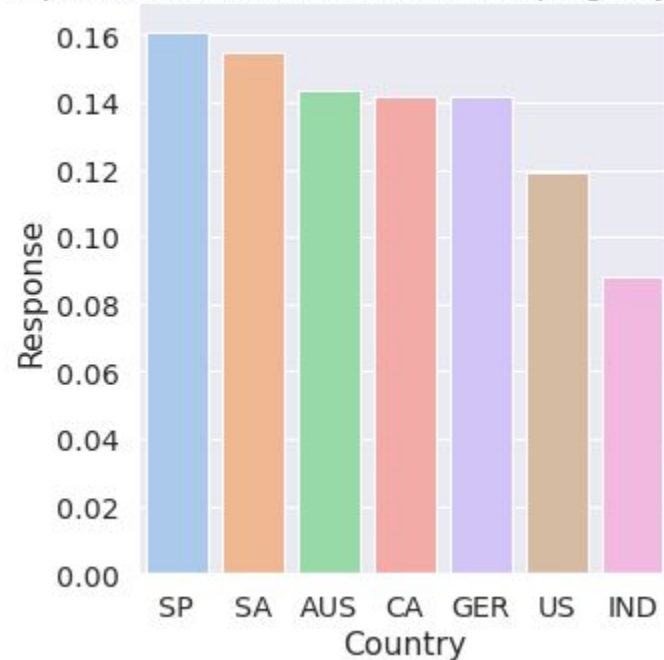
- US has highest median spending
- India has lowest median spending
- Spain has highest frugal spending



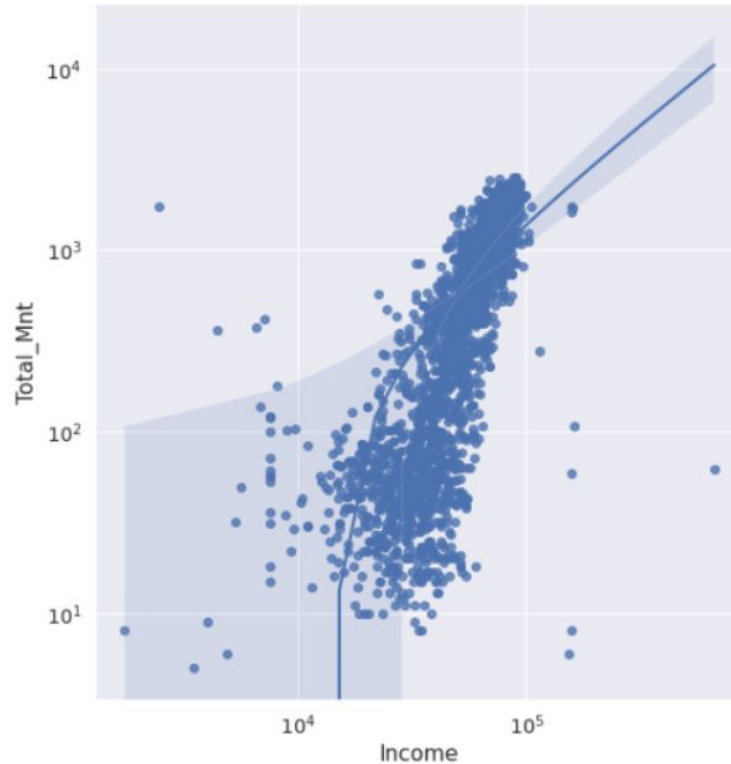
Response Rate vs Country

- Understanding of baseline response rates per country
- Different countries will have different sentiments towards campaigns

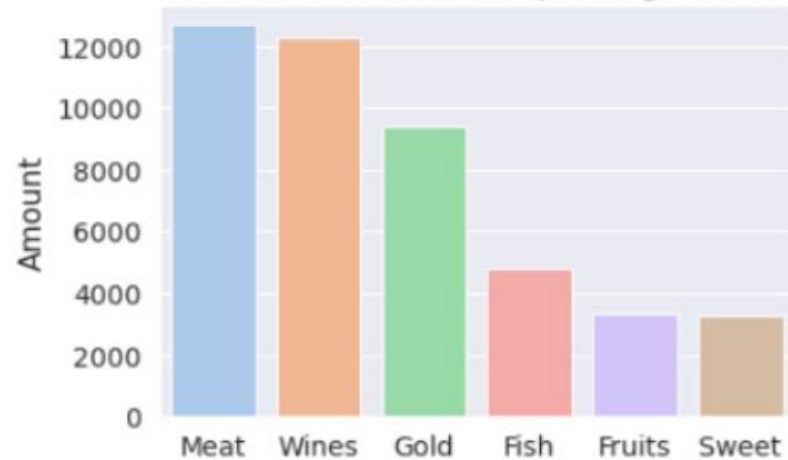
Reponses Rate for the Latest Campaign by Country



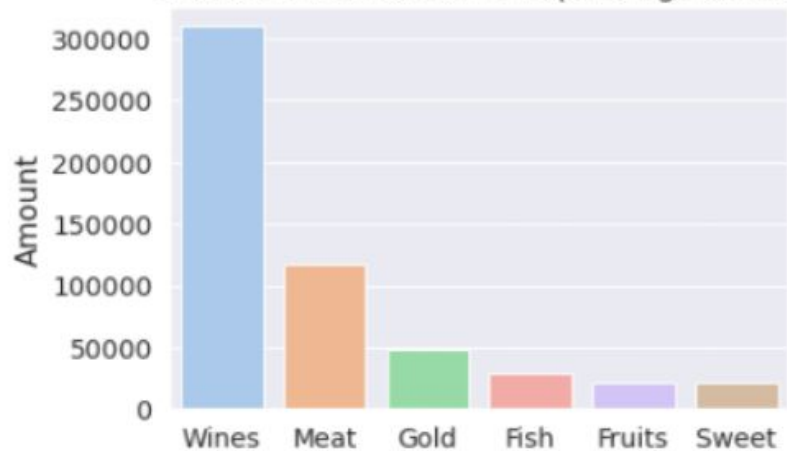
Relationship Between Income and Total Amount of Spending



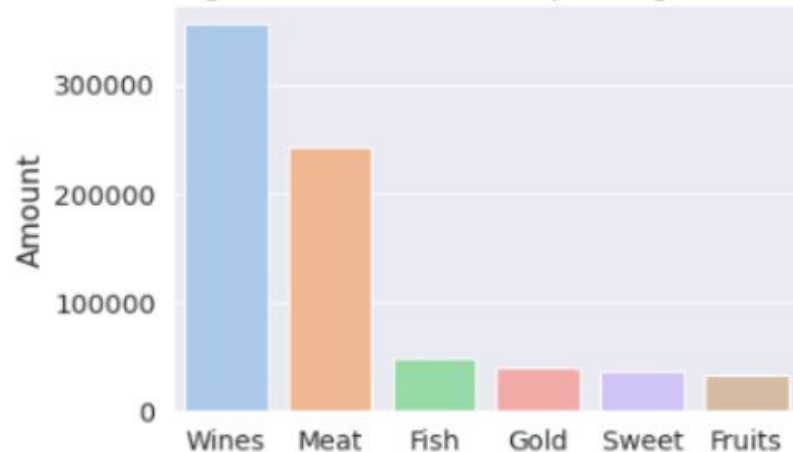
Low Income Customers Spending Patterns



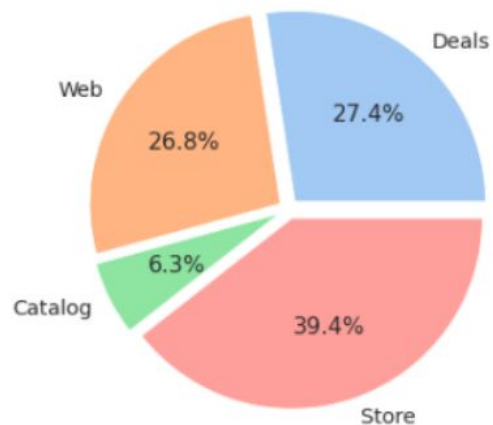
Middle Income Customers Spending Patterns



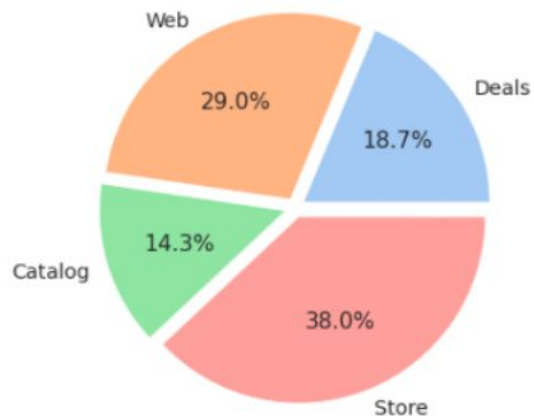
High Income Customers Spending Patterns



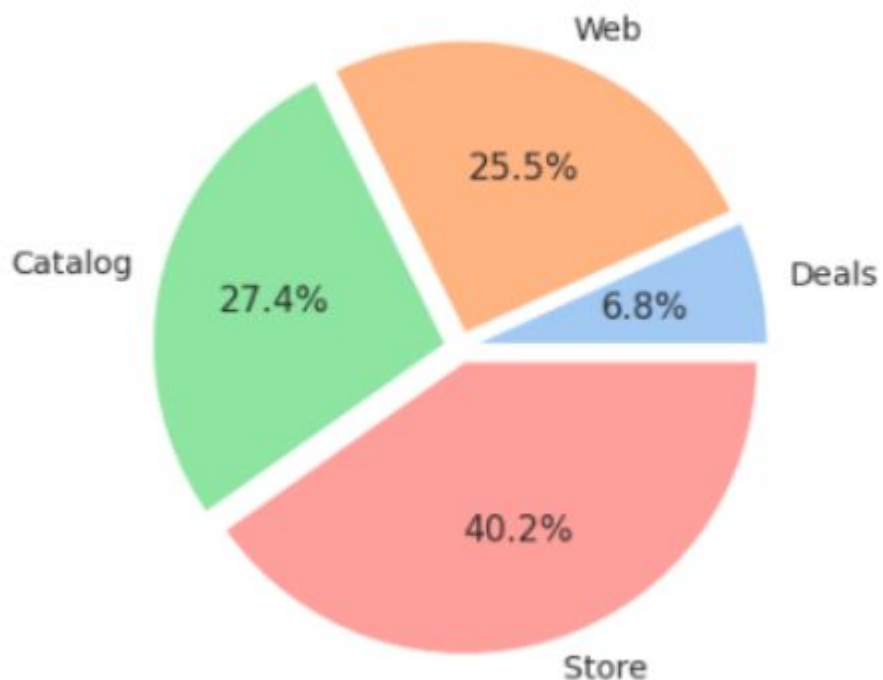
Low Income Customers Purchase Method Preferences



Middle Income Customers Purchase Method Preferences



High Income Customers Purchase Method Preferences



Suggestions

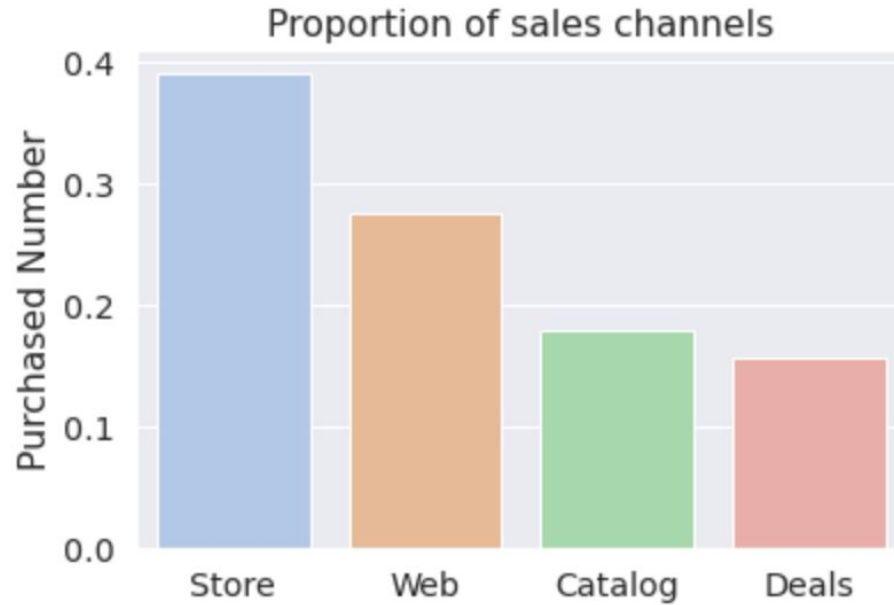
Expand the alcohol area at store

Stock more alcohol brand that covers from low-end all the way to high-end

Put alcohol area to the most conspicuous place at store if possible

Sign long term contract with alcohol suppliers

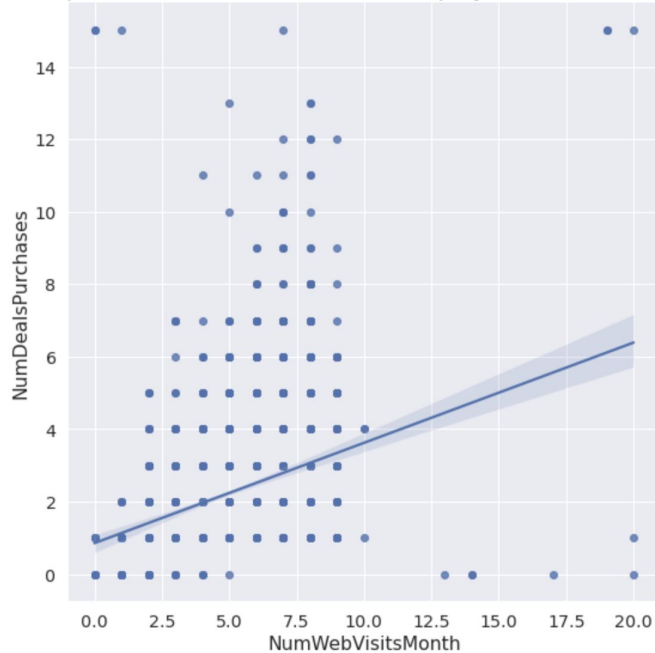
Sale channels



- Shopping preference is similar across different education groups and countries
-

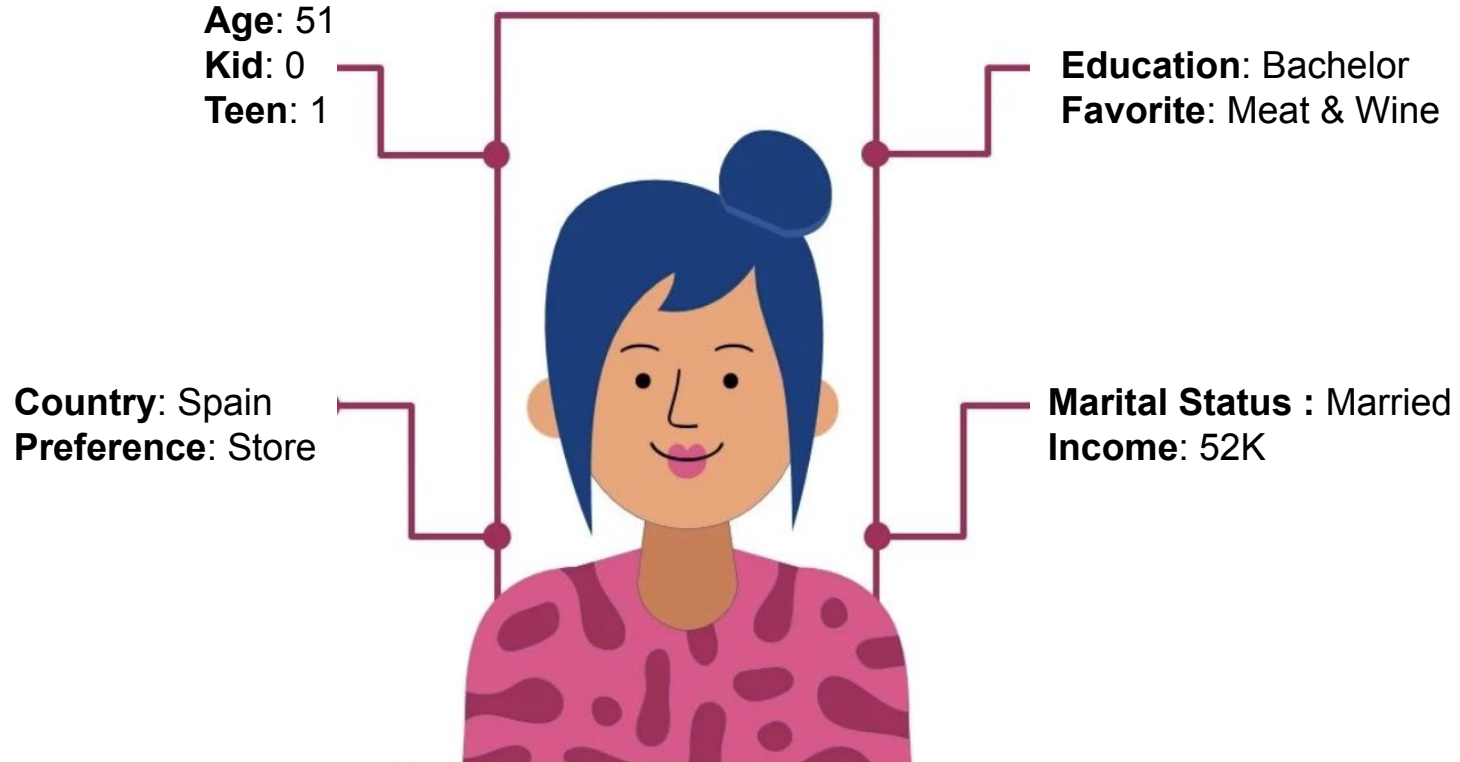
Suggestion on sale channel

Relationship Between the Number of Visits to Company Web Site and Deal Purchases



- Positive relationship between WebVisit and DealPurchase
- Website is a great way to promote deals

Customer Persona



Machine Learning model

- Linear Regression to infer
- Random Forest to predict

- Linear Regression

- R^2 : 0.587626
 - RMSE: 389.59

\$591

- Random Forest

- R^2 : 0.709599
 - RMSE: 326.93

