Consumer Profiles: EDA

Gabe Anoia, Harrison Hubbard,
Levi Sessions

Customer Personality Analysis Dataset

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

- 2,450 samples over 28 features
- Customer demographic, past spending, responses to marketing
- Data collected from undisclosed European retail company
- Collected from sales data and customer survey responses

1. Overview of Dataset
2. **Overview of Features**
3. Analysis of Numerical Variables
4. Analysis of Categorical Variables
5. What's Next

Demographic Features

Year_Birth	Customer's Birth Year	Categorical
Education	Customer's education level	Categorical
Marital_Status	Customer's marital status	Categorical
Income	Customer's yearly household income	Numerical
Kidhome	Number of children in customer's household	Numerical
Teenhome	Number of teenagers in customer's household	Numerical
Dt_Customer	Date of customer's enrollment with the company	Categorical
Recency	Number of days since customer's last purchase	Numerical

Behavioral Features

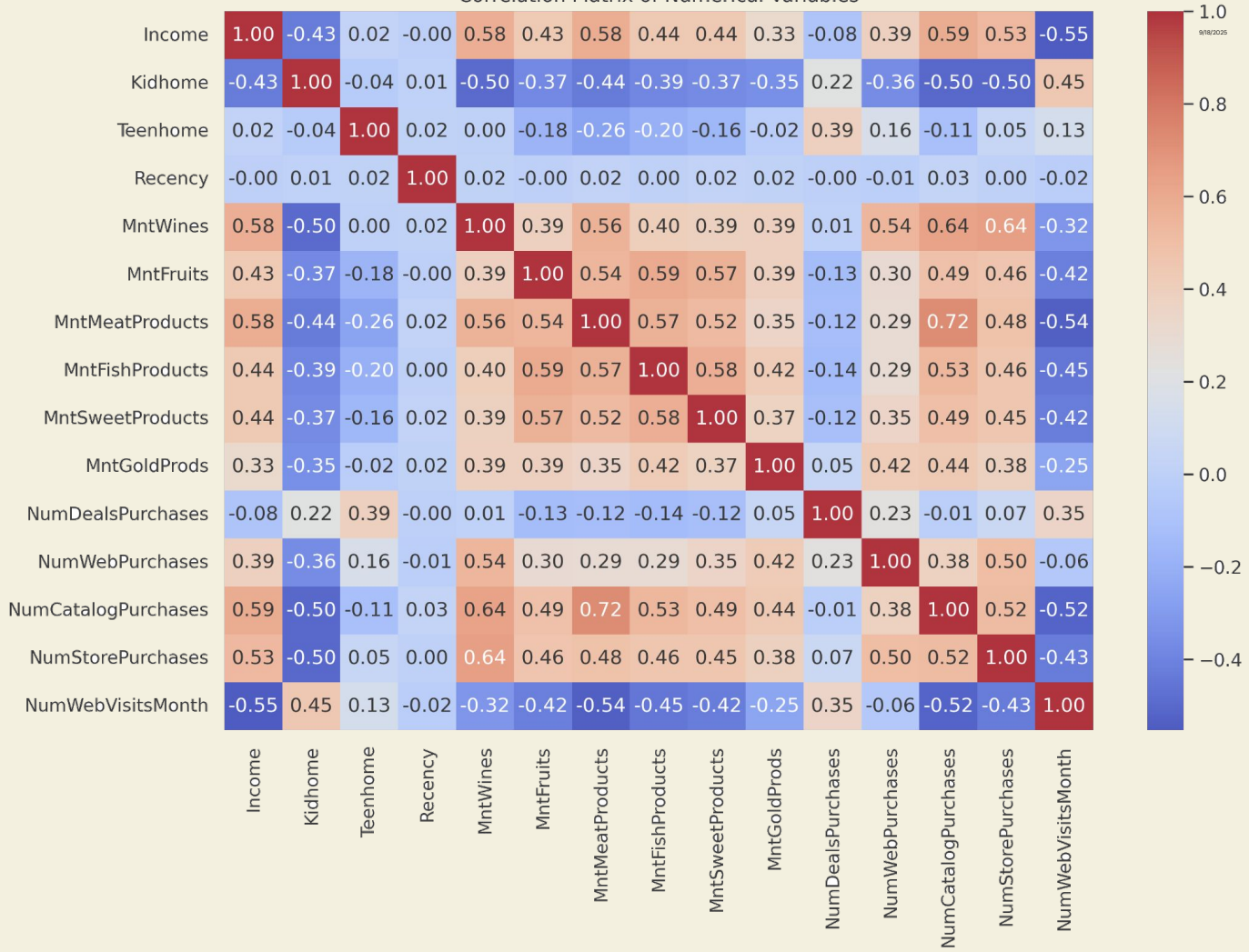
MntWines	Amount spent on wine in last 2 years	Numerical
MntFruits	Amount spent on fruits in last 2 years	Numerical
MntMeatProducts	Amount spent on meat in last 2 years	Numerical
MntFishProducts	Amount spent on fish in last 2 years	Numerical
MntSweetProducts	Amount spent on sweets in last 2 years	Numerical
MntGoldProds	Amount spent on gold in last 2 years	Numerical
NumWebPurchases	Number of purchases made through the company's website	Numerical
NumCatalogPurchases	Number of purchases made using a catalogue	Numerical
NumStorePurchases	Number of purchases made directly in stores	Numerical
NumWebVisitsMonth	Number of visits to company's website in the last month	Numerical

Engagement Features

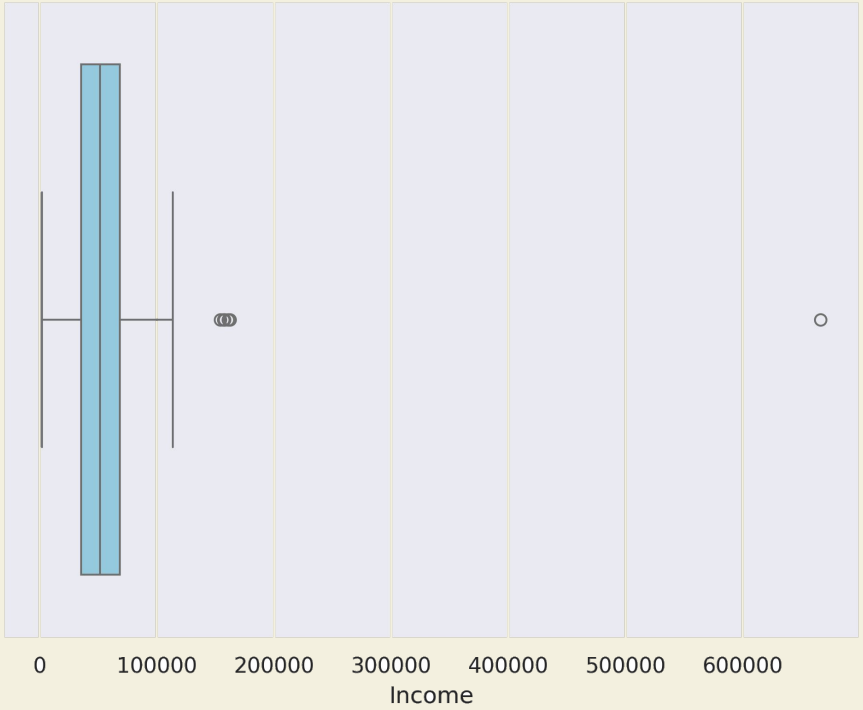
NumDealsPurchases	Number of purchases made with a discount	Categorical
AcceptedCmp1	If customer accepted the offer in the 1st campaign	Categorical
AcceptedCmp2	If customer accepted the offer in the 2nd campaign	Categorical
AcceptedCmp3	If customer accepted the offer in the 3rd campaign	Categorical
AcceptedCmp4	If customer accepted the offer in the 4th campaign	Categorical
AcceptedCmp5	If customer accepted the offer in the 5th campaign	Categorical
Response	If customer accepted the offer in the last campaign	Categorical

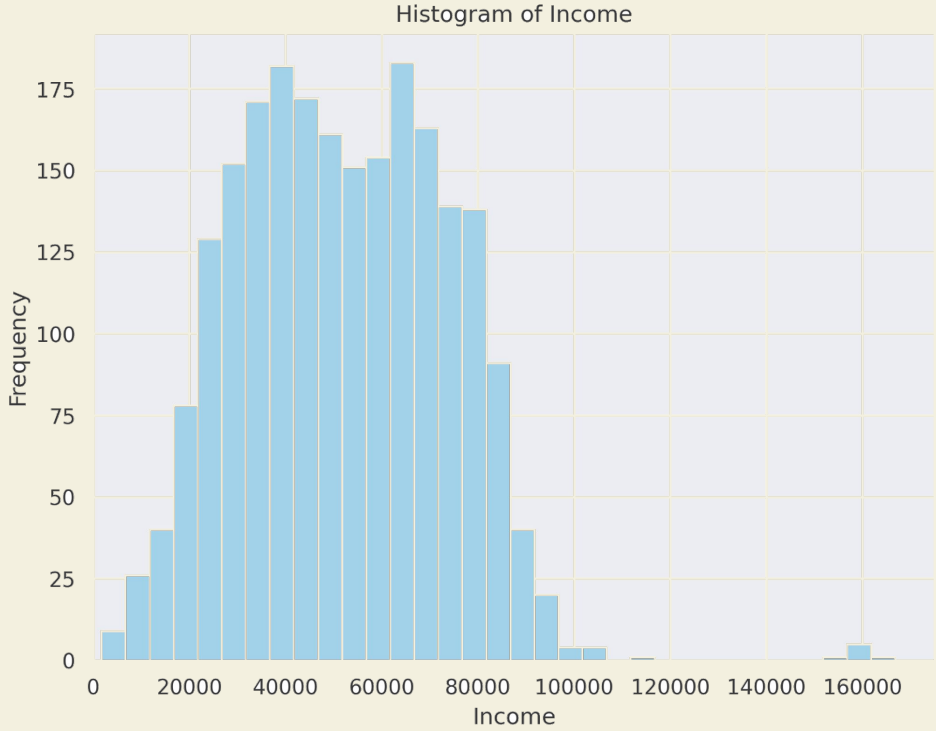
Our Goals

1. Overview of Dataset
2. Overview of Features
3. **Analysis of Numerical Variables**
4. Analysis of Categorical Variables
5. What's Next

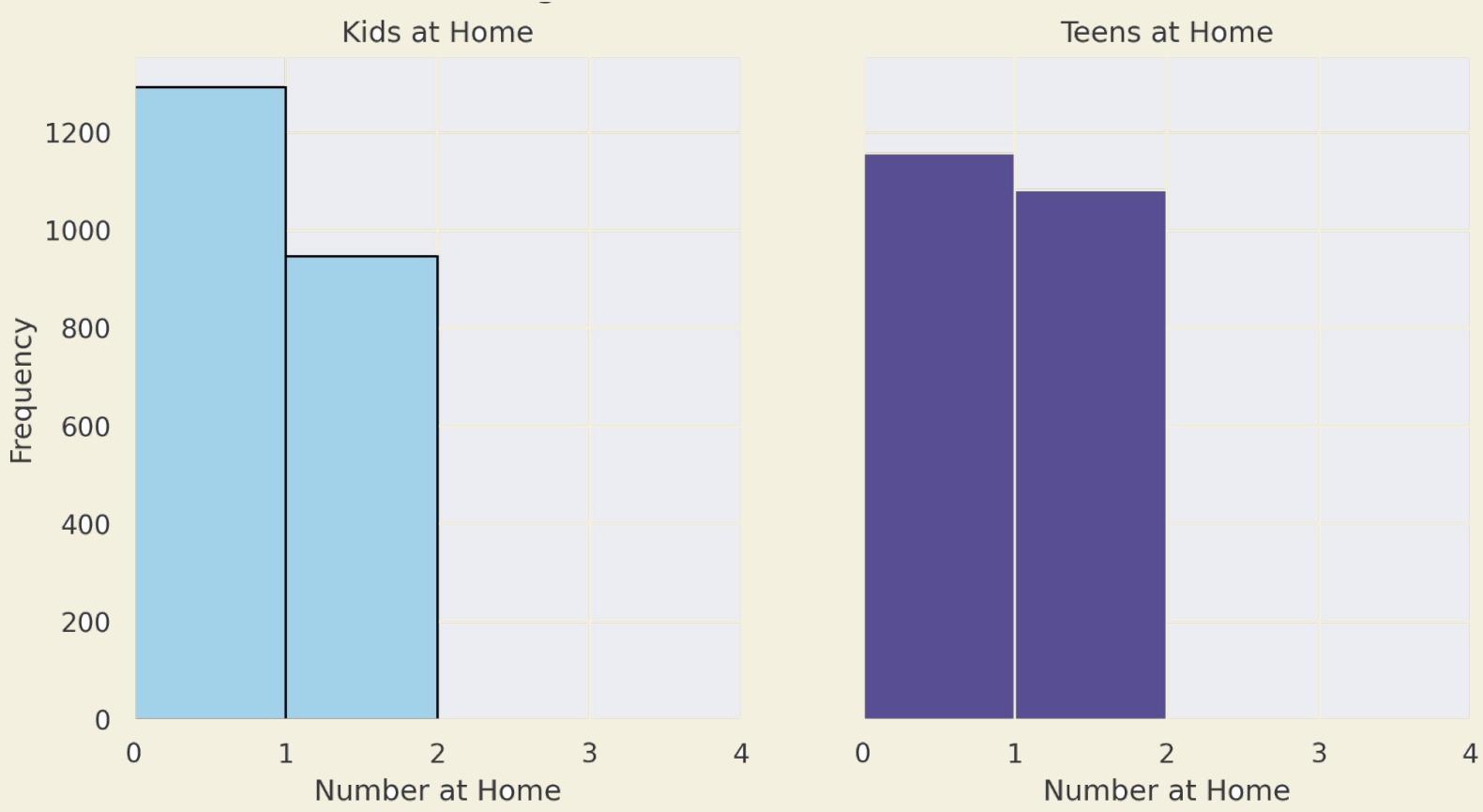

Correlation Matrix of Numerical Variables
Income 1.00 -0.43 0.02 -0.00 0.58 0.43 0.58 0.44 0.44 0.33 -0.08 0.39 0.59 0.53 -0.55
Kidhome -0.43 1.00 -0.04 0.01 -0.50 -0.37 -0.44 -0.39 -0.37 -0.35 0.22 -0.36 -0.50 -0.50 0.45
Teenhome 0.02 -0.04 1.00 0.02 0.00 -0.18 -0.26 -0.20 -0.16 -0.02 0.39 0.16 -0.11 0.05 0.13
Recency -0.00 0.01 0.02 1.00 0.02 -0.00 0.02 0.00 0.02 0.02 -0.00 -0.01 0.03 0.00 -0.02
MntWines 0.58 -0.50 0.00 0.02 1.00 0.39 0.56 0.40 0.39 0.39 0.01 0.54 0.64 0.64 -0.32
MntFruits 0.43 -0.37 -0.18 -0.00 0.39 1.00 0.54 0.59 0.57 0.39 -0.13 0.30 0.49 0.46 -0.42
MntMeatProducts 0.58 -0.44 -0.26 0.02 0.56 0.54 1.00 0.57 0.52 0.35 -0.12 0.29 0.72 0.48 -0.54
MntFishProducts 0.44 -0.39 -0.20 0.00 0.40 0.59 0.57 1.00 0.58 0.42 -0.14 0.29 0.53 0.46 -0.45
MntSweetProducts 0.44 -0.37 -0.16 0.02 0.39 0.57 0.52 0.58 1.00 0.37 -0.12 0.35 0.49 0.45 -0.42
MntGoldProds 0.33 -0.35 -0.02 0.02 0.39 0.39 0.35 0.42 0.37 1.00 0.05 0.42 0.44 0.38 -0.25
NumDealsPurchases -0.08 0.22 0.39 -0.00 0.01 -0.13 -0.12 -0.14 -0.12 0.05 1.00 0.23 -0.01 0.07 0.35
NumWebPurchases 0.39 -0.36 0.16 -0.01 0.54 0.30 0.29 0.29 0.35 0.42 0.23 1.00 0.38 0.50 -0.06
NumCatalogPurchases 0.59 -0.50 -0.11 0.03 0.64 0.49 0.72 0.53 0.49 0.44 -0.01 0.38 1.00 0.52 -0.52
NumStorePurchases 0.53 -0.50 0.05 0.00 0.64 0.46 0.48 0.46 0.45 0.38 0.07 0.50 0.52 1.00 -0.43
NumWebVisitsMonth -0.55 0.45 0.13 -0.02 -0.32 -0.42 -0.54 -0.45 -0.42 -0.25 0.35 -0.06 -0.52 -0.43 1.00
Income
Kidhome
Teenhome
Recency
MntWines
MntFruits
MntMeatProducts
MntFishProducts
MntSweetProducts
MntGoldProds
NumDealsPurchases
NumWebPurchases
NumCatalogPurchases
NumStorePurchases
NumWebVisitsMonth
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

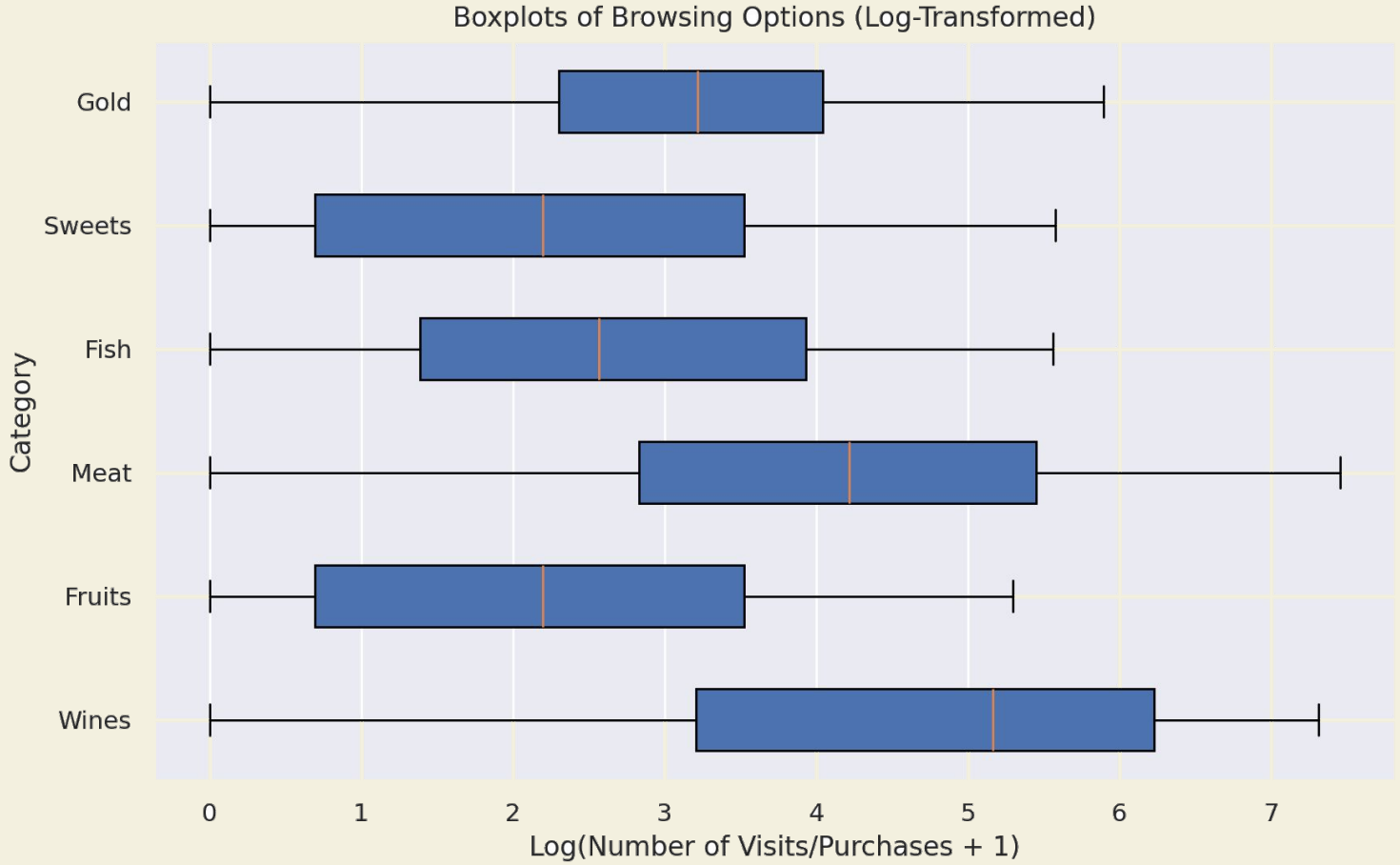
Income


0
100000
200000
300000
400000
500000
600000
Income


Histogram of Income
Frequency
0
25
50
75
100
125
150
175
0
20000
40000
60000
80000
100000
120000
140000
160000
Income

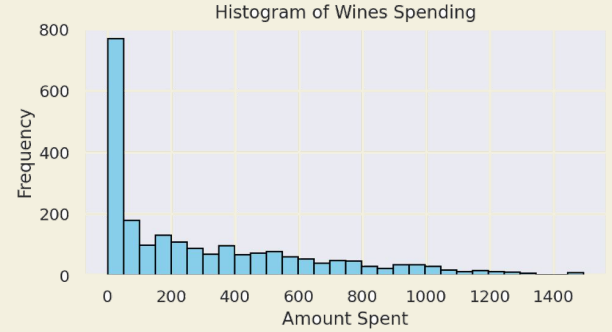
Kids and Teens in Home

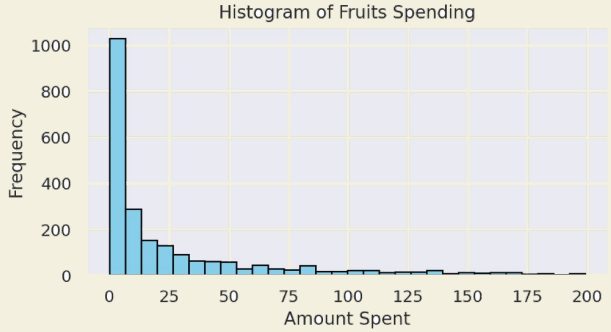


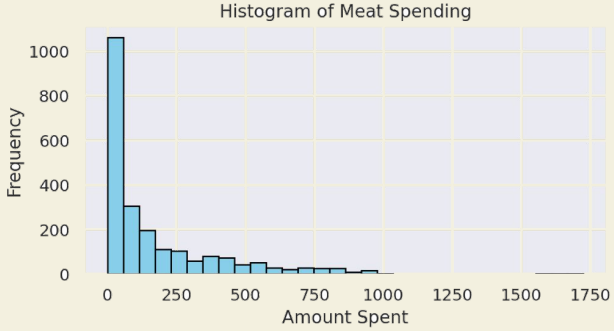

Boxplots of Browsing Options (Log-Transformed)
Gold
Sweets
Fish
Meat
Fruits
Wines
Category
0
1
2
3
4
5
6
7
Log(Number of Visits/Purchases + 1)

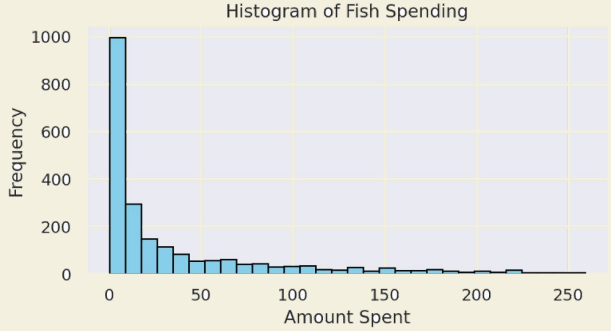
Notable Conclusions:

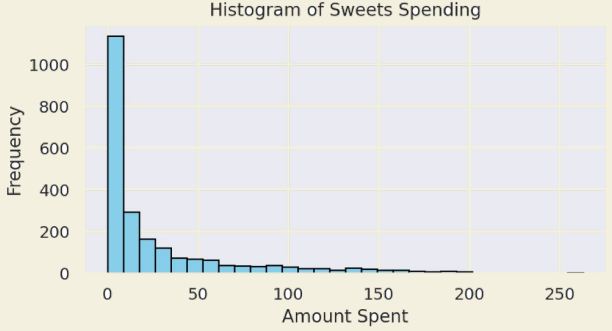
- Hard right skew in all distributions of the spending types
- Wine has the largest spread of the items, while also having the longest right tail of them all
- Gold and meat have a large cluster of data within the first $200 of spending

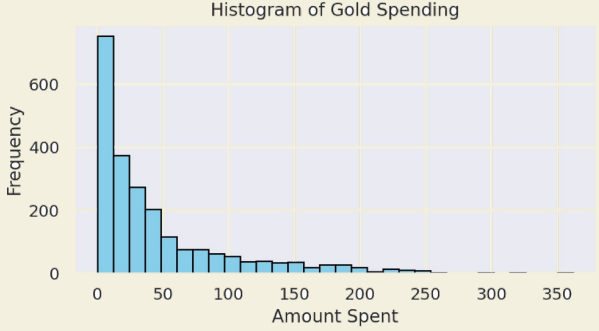


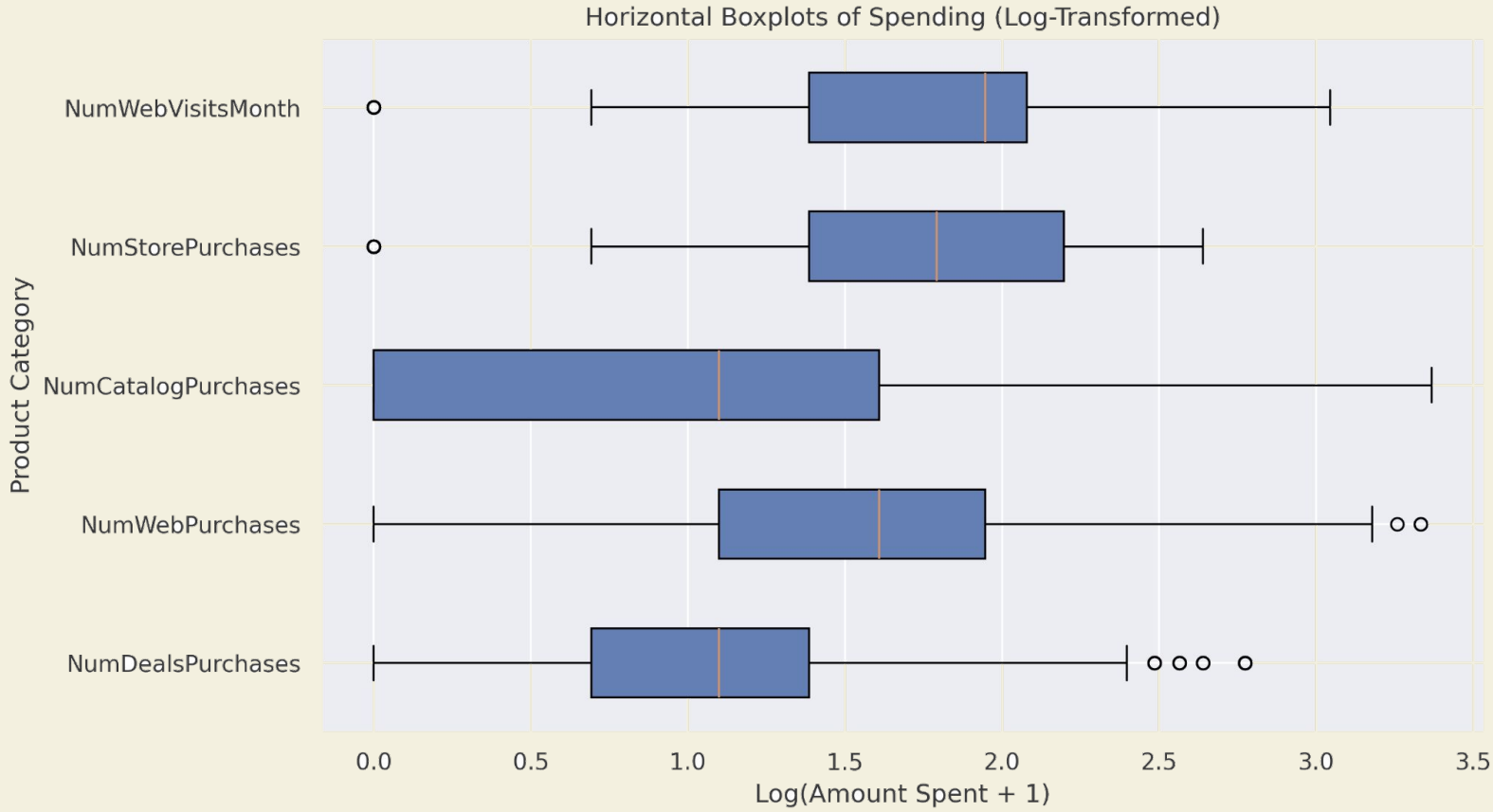


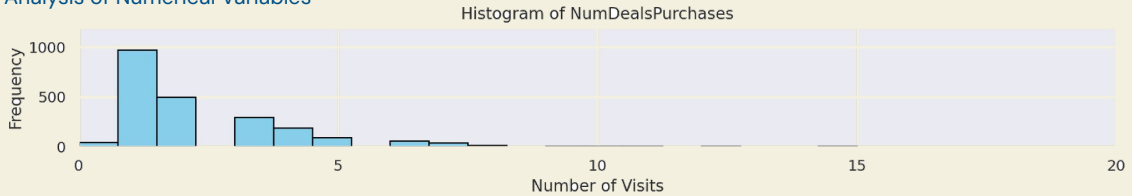


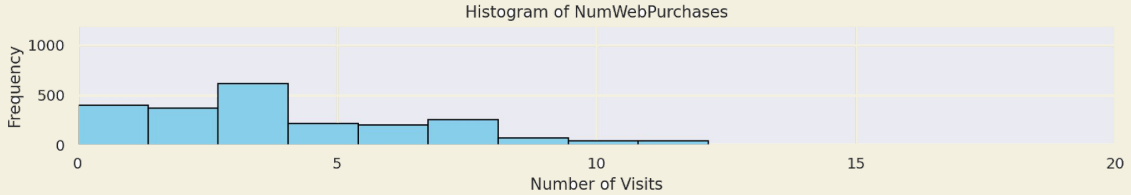


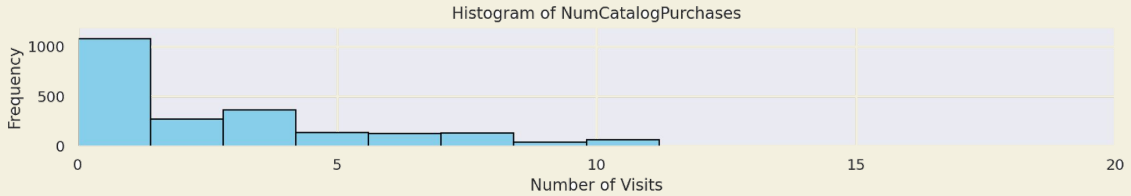


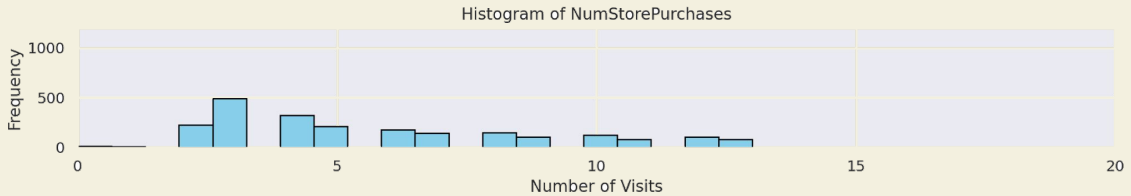


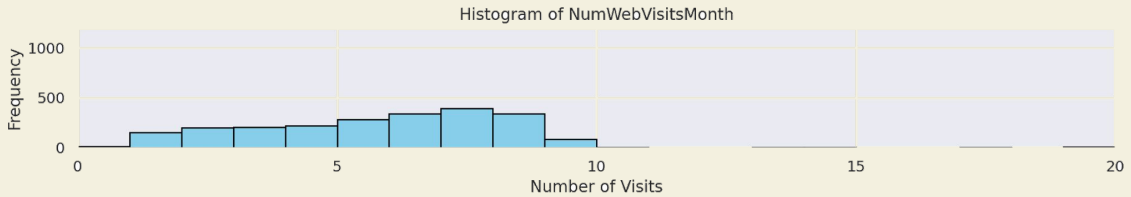

Horizontal Boxplots of Spending (Log-Transformed)
NumWebVisitsMonth
NumStorePurchases
NumCatalogPurchases
NumWebPurchases
NumDealsPurchases
Product Category
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Log(Amount Spent + 1)







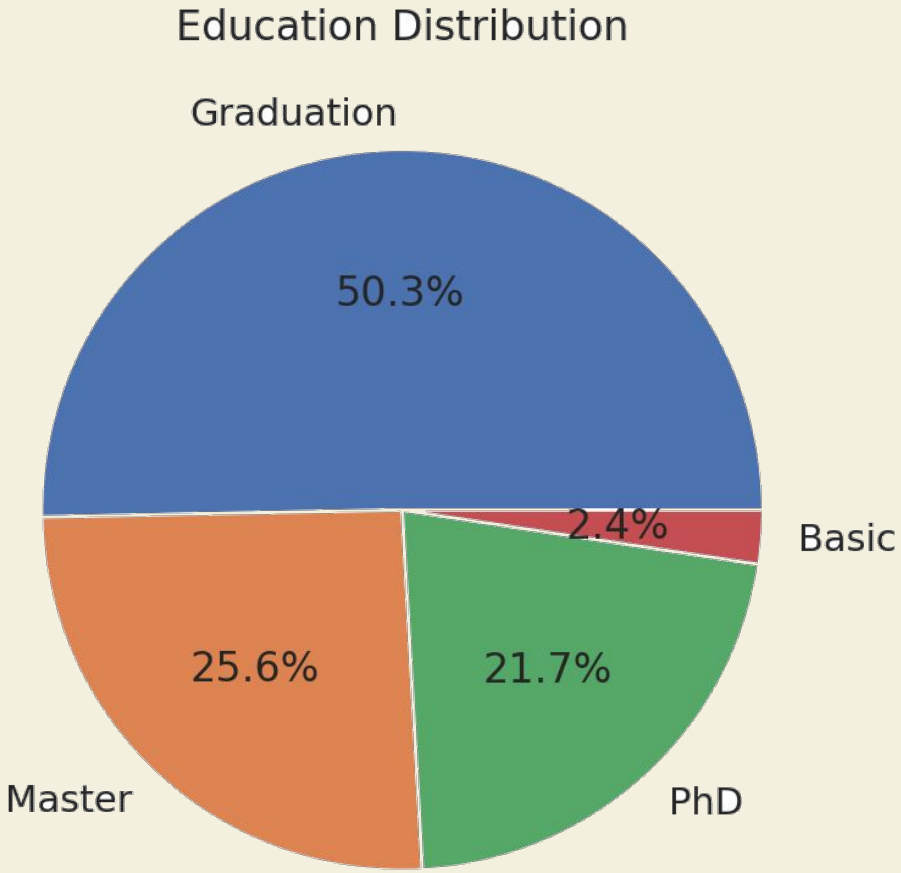




Notable Conclusions:

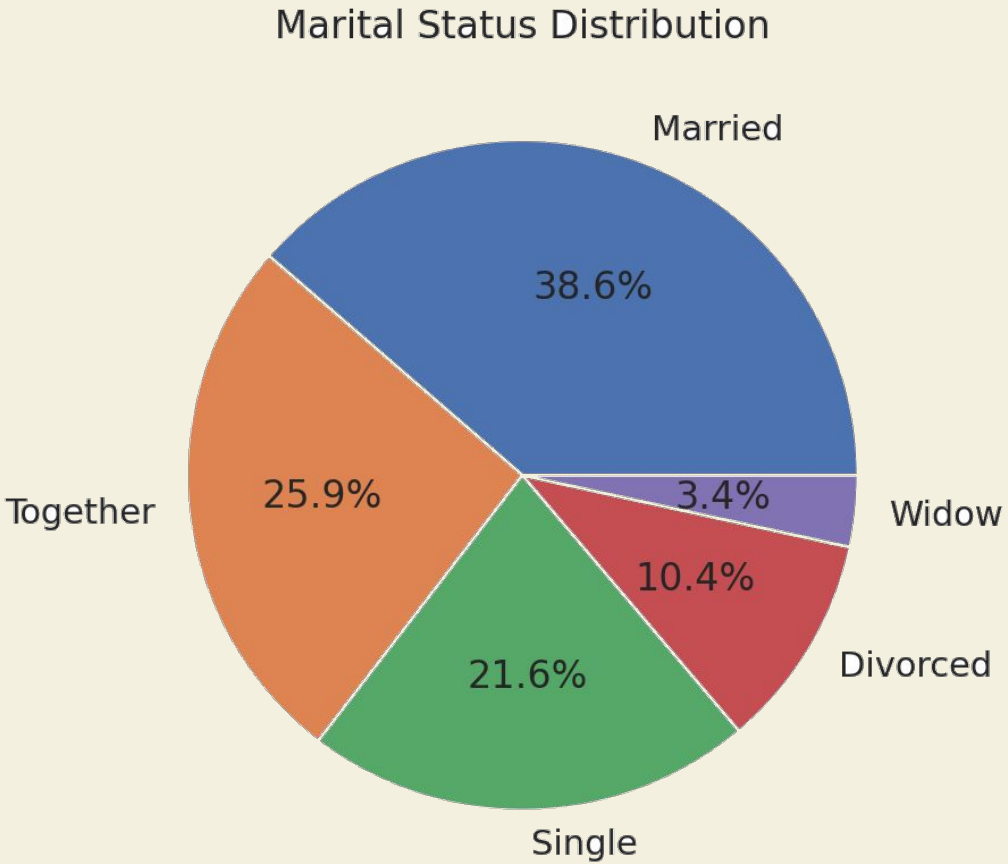
- The distributions of purchases with deals, from catalogs, and from the store are all slightly skewed to the right
- The distribution of the web purchases made is roughly symmetrical, while the number of web visits per month has a slight left skew

1. Overview of Dataset
2. Overview of Features
3. Analysis of Numerical Variables
4. **Analysis of Categorical Variables**
5. What's Next



Notable Conclusions:

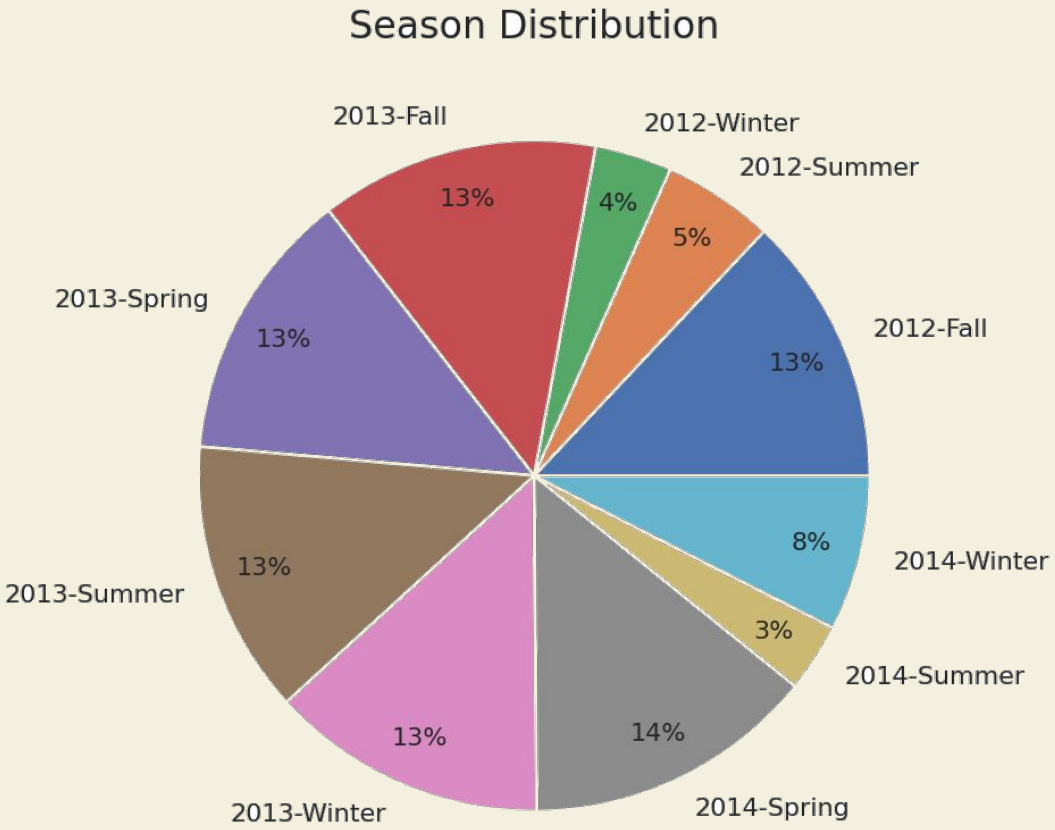
- Majority of customers are highly educated (Graduation, Master's, PhD), with only a small fraction reporting basic education.
- Higher education levels suggest a customer base with greater income potential and stronger spending in premium categories like wine and meat.



Notable Conclusions:

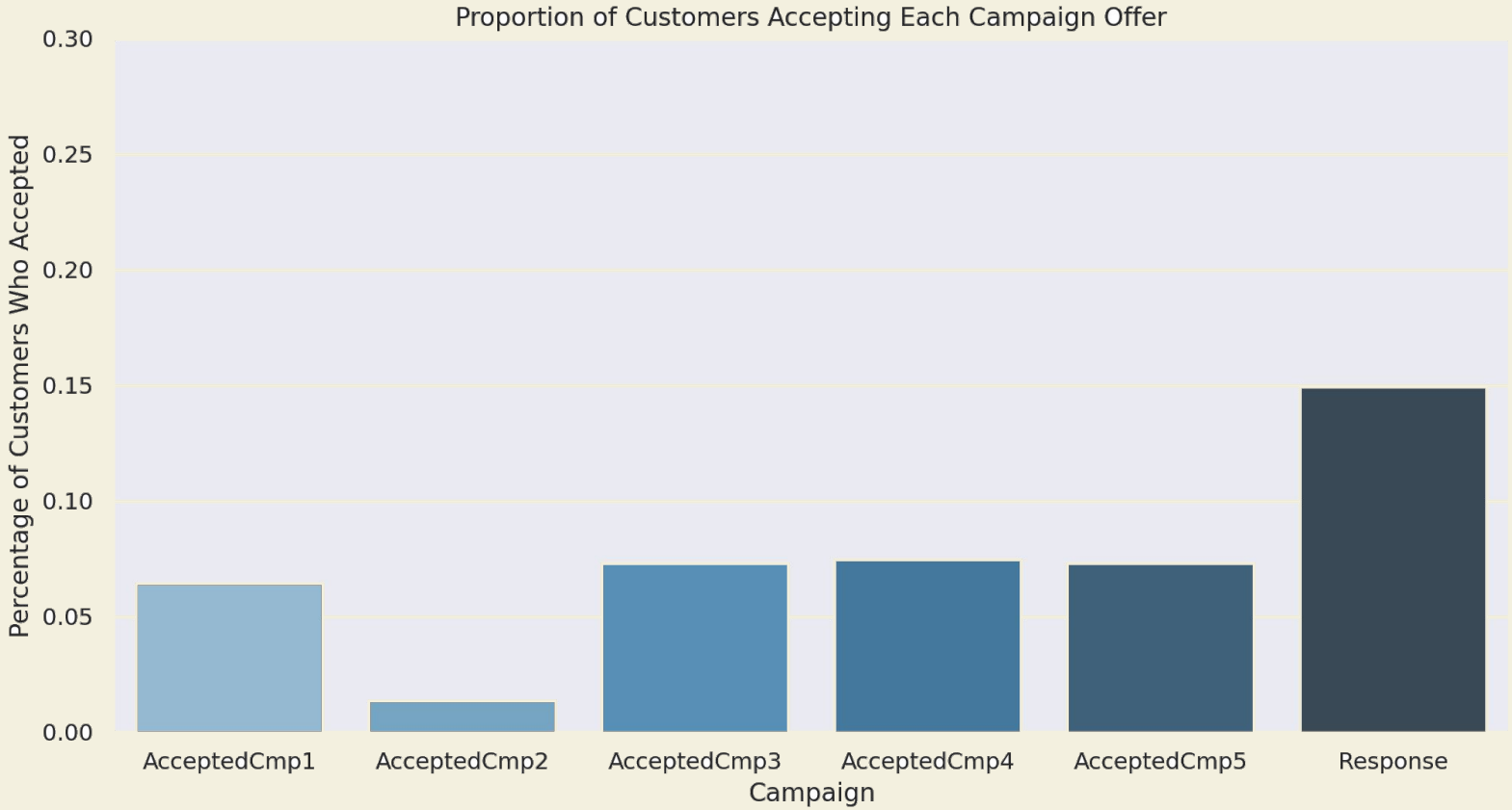
- Multiple categories representing similar structures => Combine into

- Household structure likely influences spending priorities => Families with children showed lower non essential spending, while singles may focus more on time saving purchases.



Notable Conclusions:

- Customer enrollments are evenly distributed across 2013, with smaller representation in 2012 and 2014.
- No single period dominates, reducing temporal bias in the dataset.
- Enrollment season does not strongly correlate with recency or spending behavior.


Proportion of Customers Accepting Each Campaign Offer
Percentage of Customers Who Accepted
0.30
0.25
0.20
0.15
0.10
0.05
0.00
AcceptedCmp1
AcceptedCmp2
AcceptedCmp3
AcceptedCmp4
AcceptedCmp5
Response
Campaign

1. Overview of Dataset
2. Overview of Features
3. Analysis of Numerical Variables
4. Analysis of Categorical Variables
5. **What's Next**