Customer Value Analysis

WITH PYTHON

Pei Yii Ng (Heather)

1 December, 2023

BACKGROUND

Customer Value Analysis is vital for understanding customer perceptions and improving offerings to meet their needs by observing behaviours and spending patterns. It informs decisions on product development, pricing, and marketing.

Companies utilise data analytics to acquire a more precise insight into customer segments and trends in response to the rapid growth of available data.

DATA SOURCE

Kaggle's 'Company's Ideal Customers | Marketing Strategy' dataset by Aman Chauhan, sourced from a superstore's previous marketing campaign, uncovering trends to optimise business outcomes.

Kaggle Source

Data consist of 4 sections:

1. Customer's personal information
2. amount customers spend on each category of products in the last 2 years.
3. information about customer activities and purchases
4. Overview of the success rate of each marketing campaign conducted by the company

ANALYSIS OBJECTIVES

Data Exploration and Quality Checking

Iimprove data quality through a series of data cleansing techniques that address outliers, missing values, duplication and data format standardisation

Customer Segmentation and Behavior Analysis

In-depth analysis of customers' personal information, behaviours, historical transactions and promotional order data, as well as clustering analysis and categorisation of customers into different groups.

Predict Customer Churn and Identify at-risk customers

Develop a model to precisely anticipate churn and identify high-risk consumers

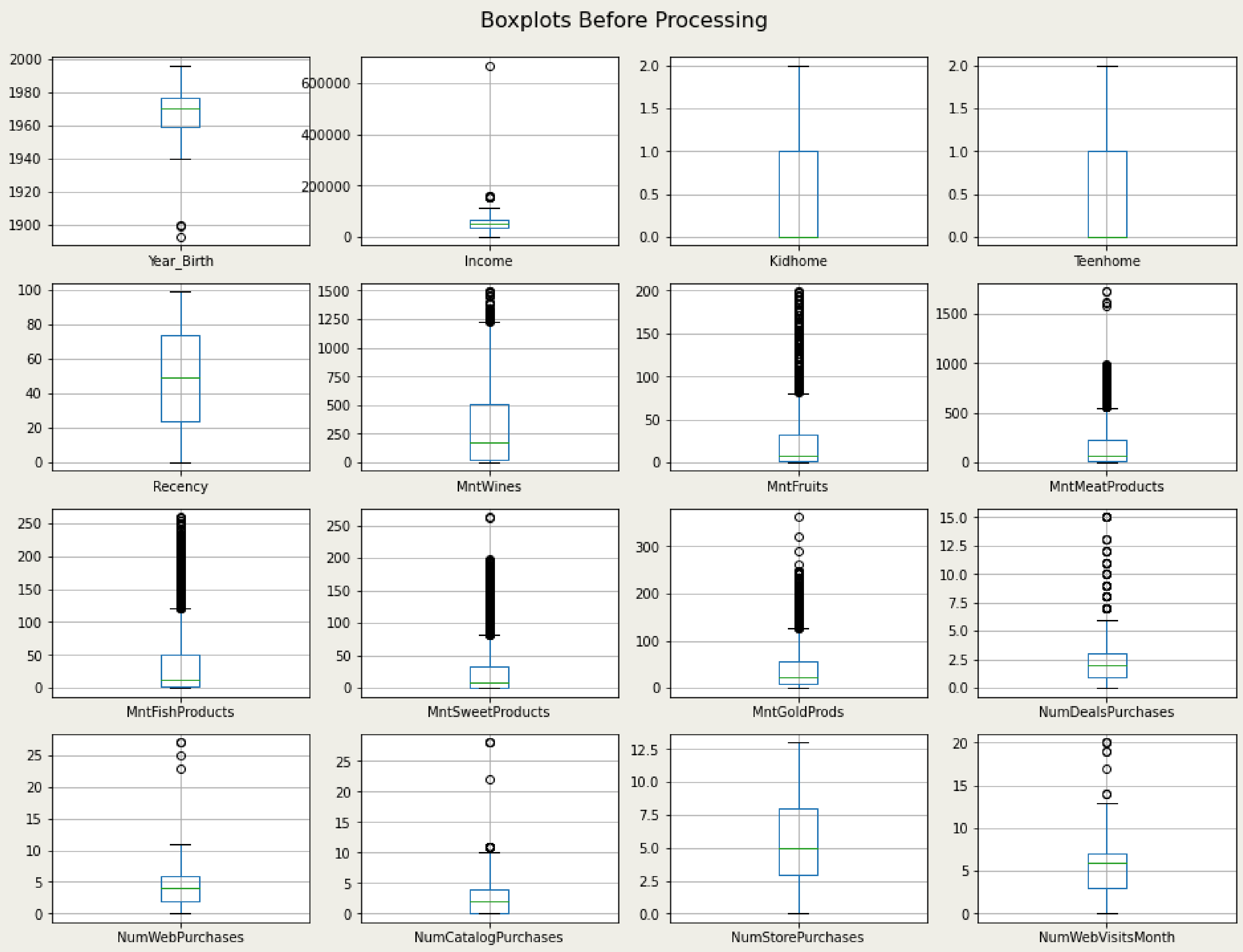
Personalised Marketing Algorithm

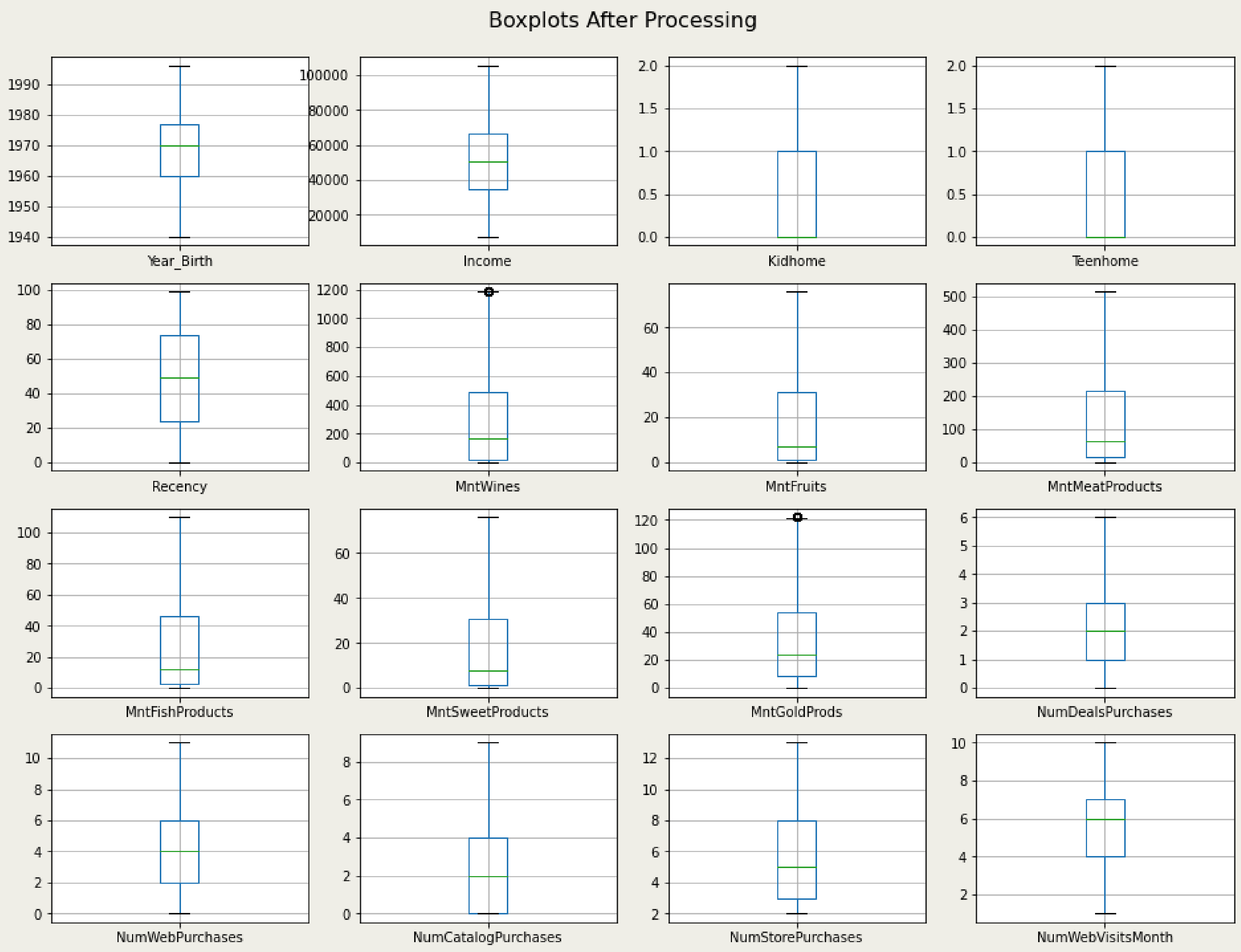
Tailor marketing strategies based on individual customer preferences, optimizing the relevance and effectiveness of promotional efforts

DATA EXPLORATION AND QUALITY CHECKING

- Handling Missing Values: Imputed missing values in the 'Income' column using mean income based on education level.
- Detecting and Treating Outliers: Used boxplots to identify outliers in numerical columns like 'Year_Birth' and 'Income'. Replaced outliers with median values or removed them.
- Data Transformation: Converted 'Year_Birth' and 'Dt_Customer' to datetime format, calculated 'Age_Customer' and 'Days_Customer', combined 'Kidhome' and 'Teenhome' into 'Childhome'.

DATA EXPLORATION AND QUALITY CHECKING



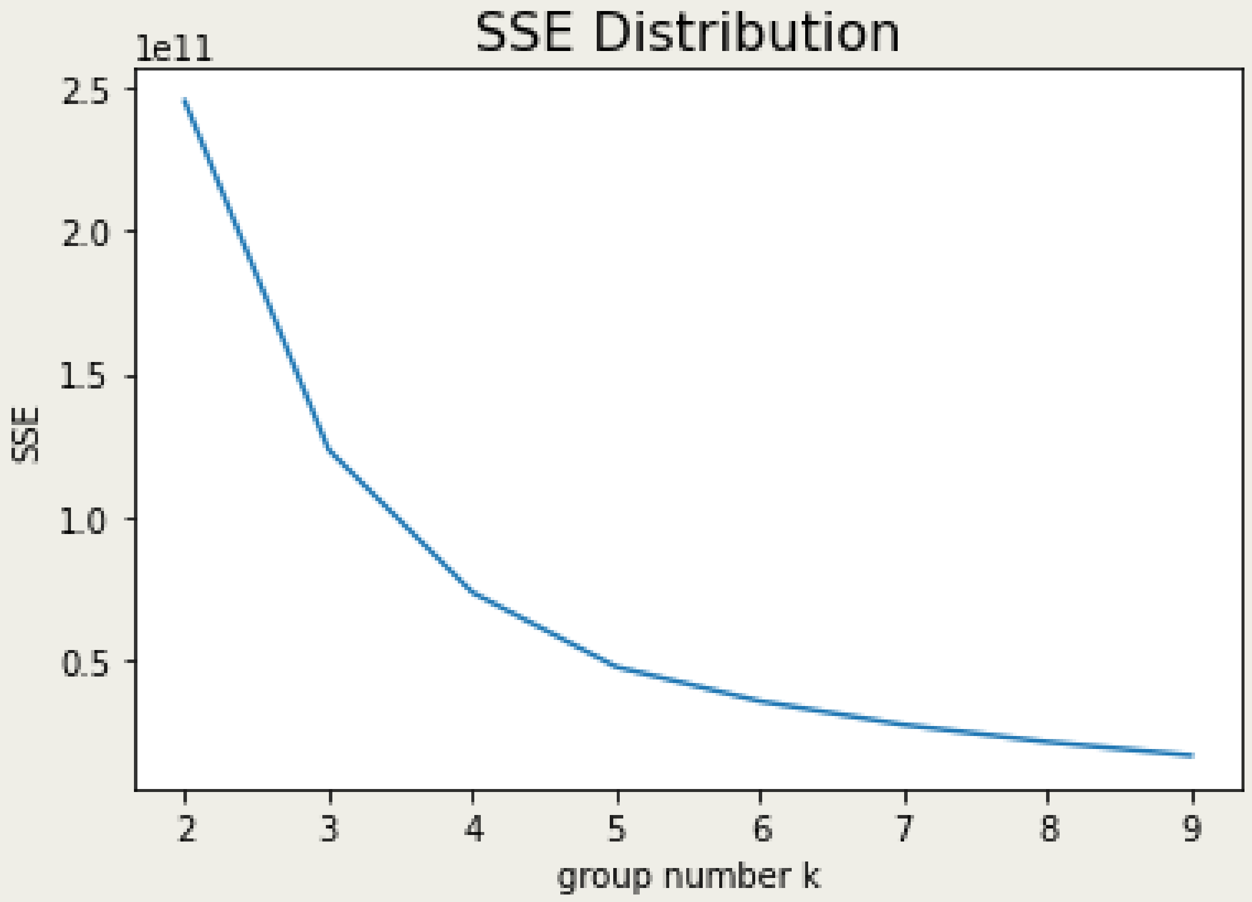


```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2142 entries, 1 to 2239
Data columns (total 30 columns):
 #   Column               Non-Null Count  Dtype         
---  ------               --------------  -----         
 0   ID                   2142 non-null   int64         
 1   Year_Birth           2142 non-null   datetime64[ns]
 2   Education            2142 non-null   object        
 3   Marital_Status       2142 non-null   object        
 4   Income               2142 non-null   float64       
 5   Kidhome              2142 non-null   int64         
 6   Teenhome             2142 non-null   int64         
 7   Dt_Customer          2142 non-null   datetime64[ns]
 8   Recency              2142 non-null   int64         
 9   MntWines             2142 non-null   int64         
 10  MntFruits            2142 non-null   int64         
 11  MntMeatProducts      2142 non-null   int64         
 12  MntFishProducts      2142 non-null   int64         
 13  MntSweetProducts     2142 non-null   int64         
 14  MntGoldProds         2142 non-null   int64         
 15  NumDealsPurchases    2142 non-null   int64         
 16  NumWebPurchases      2142 non-null   int64         
 17  NumCatalogPurchases  2142 non-null   int64         
 18  NumStorePurchases    2142 non-null   int64         
 19  NumWebVisitsMonth    2142 non-null   int64         
 20  AcceptedCmp3         2142 non-null   object        
 21  AcceptedCmp4         2142 non-null   object        
 22  AcceptedCmp5         2142 non-null   object        
 23  AcceptedCmp1         2142 non-null   object        
 24  AcceptedCmp2         2142 non-null   object        
 25  Complain             2142 non-null   object        
 26  Response             2142 non-null   object        
 27  Age_Customer         2142 non-null   int64         
 28  Days_Customer        2142 non-null   int16         
 29  Childhome            2142 non-null   int64         
dtypes: datetime64[ns](2), float64(1), int16(1), int64(17), object(9)
memory usage: 506.2+ KB
```

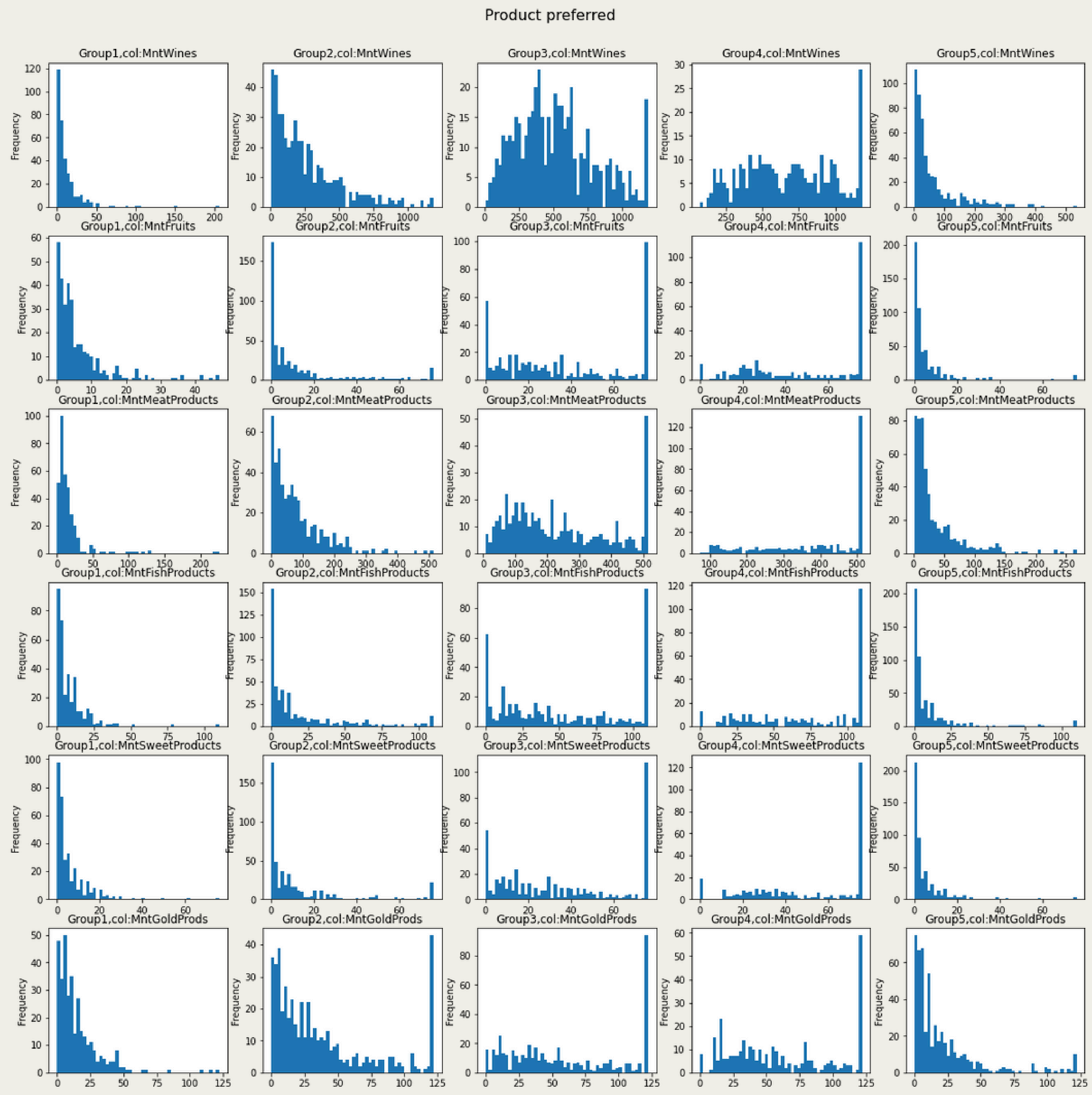
The final cleaned data set included 30 variables with a total of 2142 records.

CUSTOMER SEGMENTATION & BEHAVIOR ANALYSIS

- Clustering Approach: Used K-Means clustering to group customers with similar characteristics and behaviors.
- Determining Optimal Number of Clusters: Calculated Sum of Squared Errors (SSE) for different values of k (2-10). Used the Elbow Method to determine the optimal k=5.



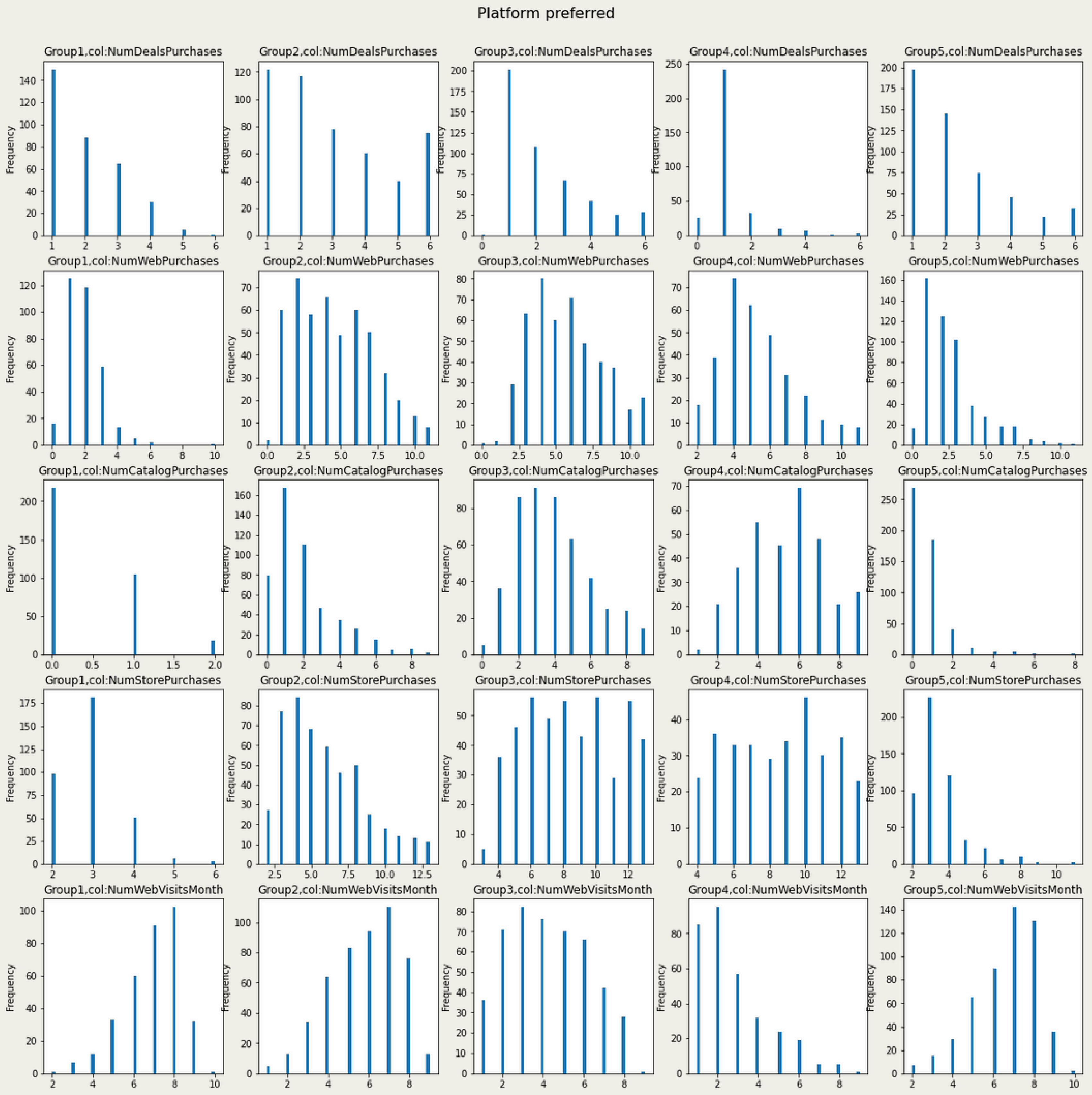
CUSTOMER SEGMENTATION & BEHAVIOR ANALYSIS



Looking at this figure in the horizontal dimension, Group 1 to 5 shows a gradual rightwards deviation in the purchase volume of each product, which indicates an increase. This is highly possible to be determined by income.

CUSTOMER SEGMENTATION & BEHAVIOR ANALYSIS

The peaks in this figure are more obvious, which we guess indicates the preference of each group in terms of purchasing platform. So combining the two plots, we summarised the characteristics of each group of members:



CUSTOMER SEGMENTATION & BEHAVIOR ANALYSIS

Findings:

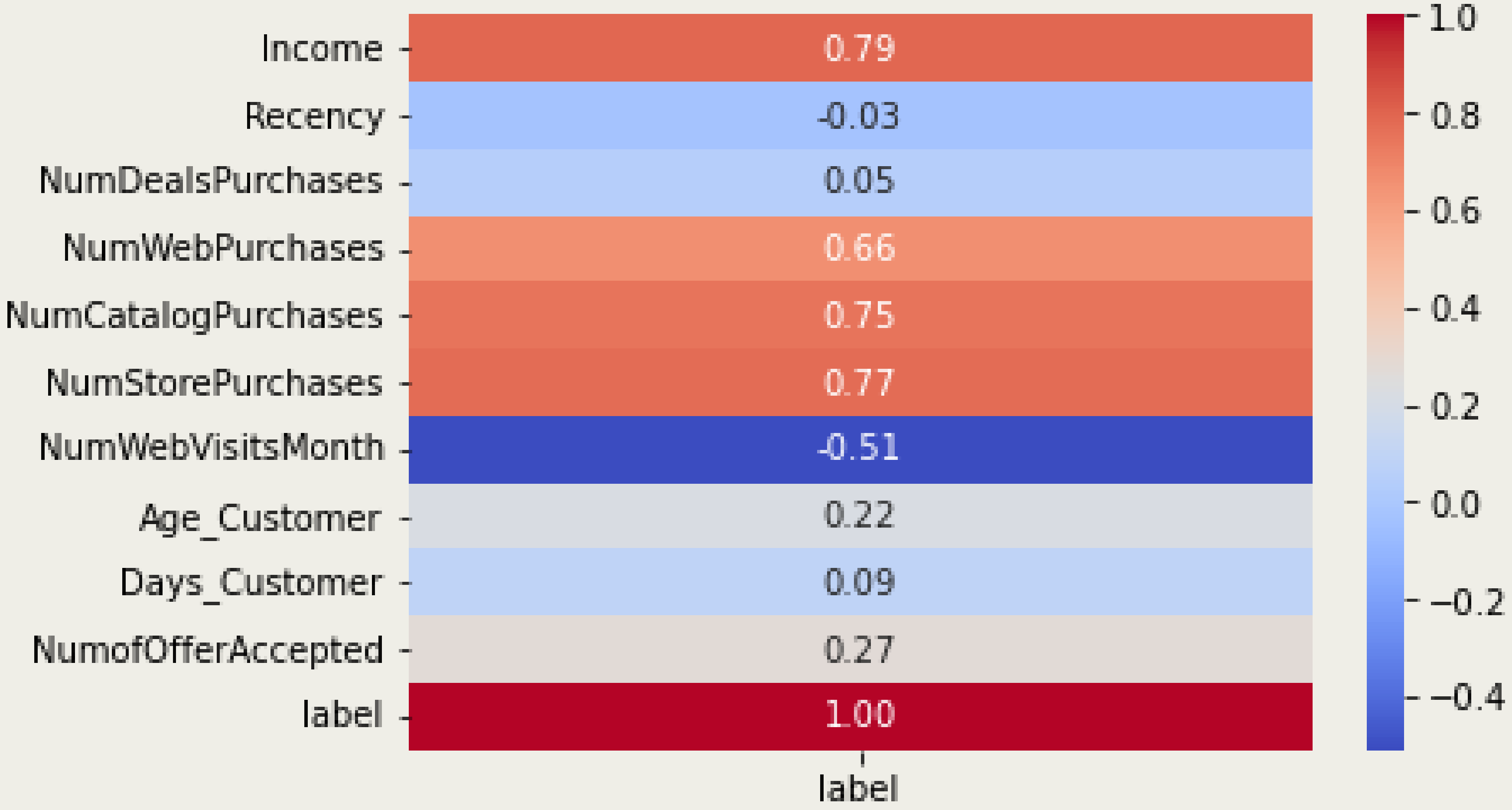
- Group 1: Low-income, many children, low conversion rate
- Group 2: Medium-low income, many children, deal seekers
- Group 3: Medium income, many children, high deal purchases
- Group 4: Medium-high income, online shoppers, prefer gold products
- Group 5: High-income, offline shoppers, wine lovers

CHURN PREDICTION MODEL DEVELOPMENT

- Unsupervised Learning Approach: Used K-Medoids clustering to predict customer churn and identify at-risk customers.
- Feature Selection and Distance Metric: Selected features like 'Income', 'Recency', purchase counts. Used Gower distance to handle mixed data types.

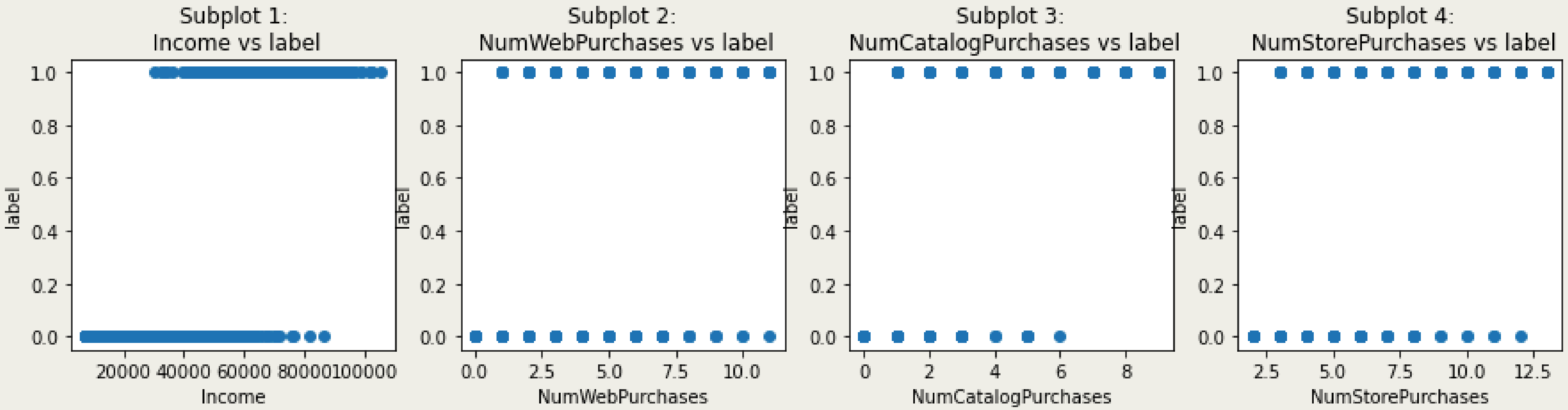
CHURN PREDICTION MODEL DEVELOPMENT

Figure 3.1: Customer Churn Correlation Heatmap

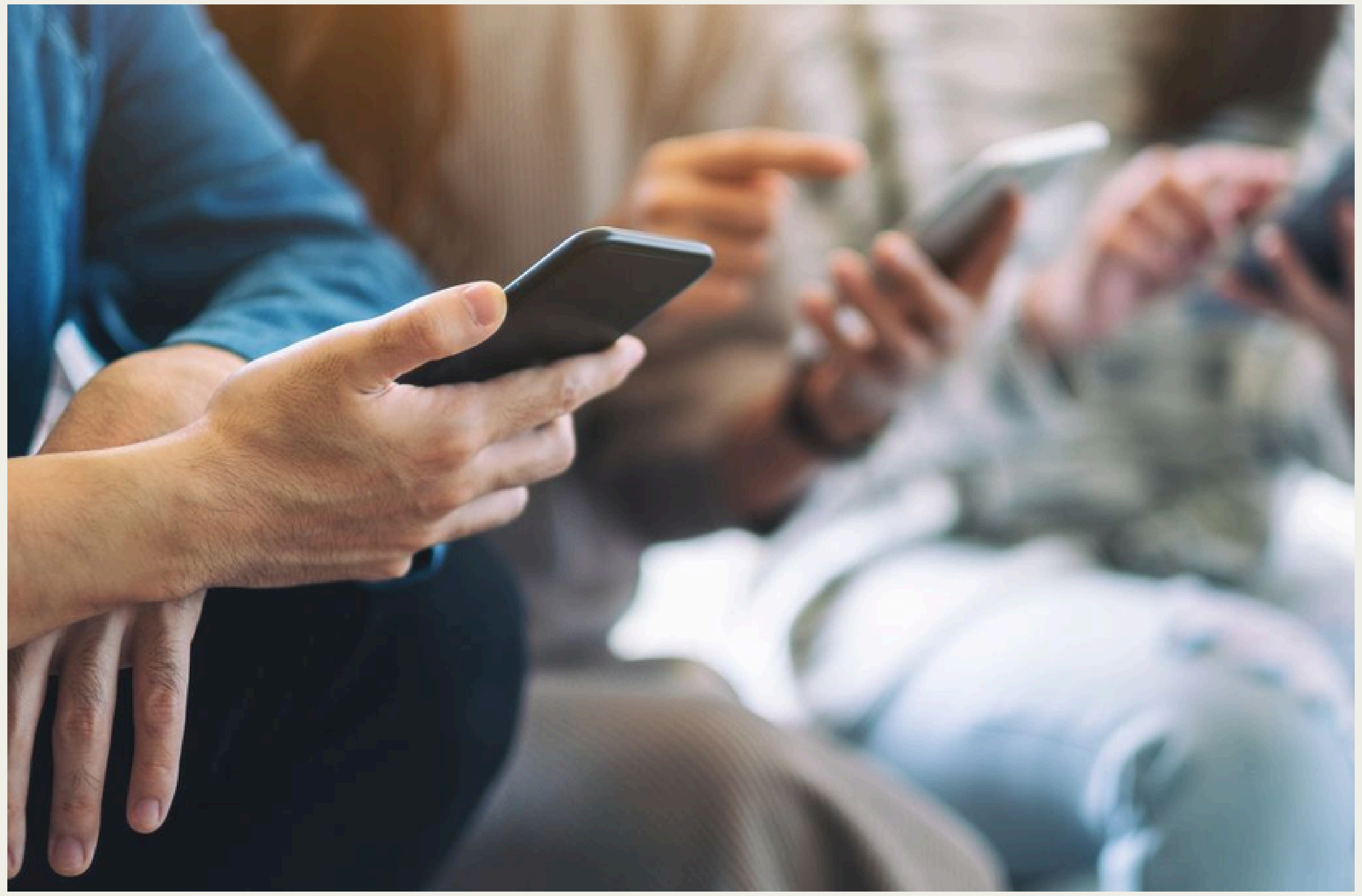


Visualises customer churn patterns by showing the churn rates based on customer behaviours

Figure 3.2



Subplots that plot the variable that has a correlation of at least 0% against the churn label



DATA SOURCE #2

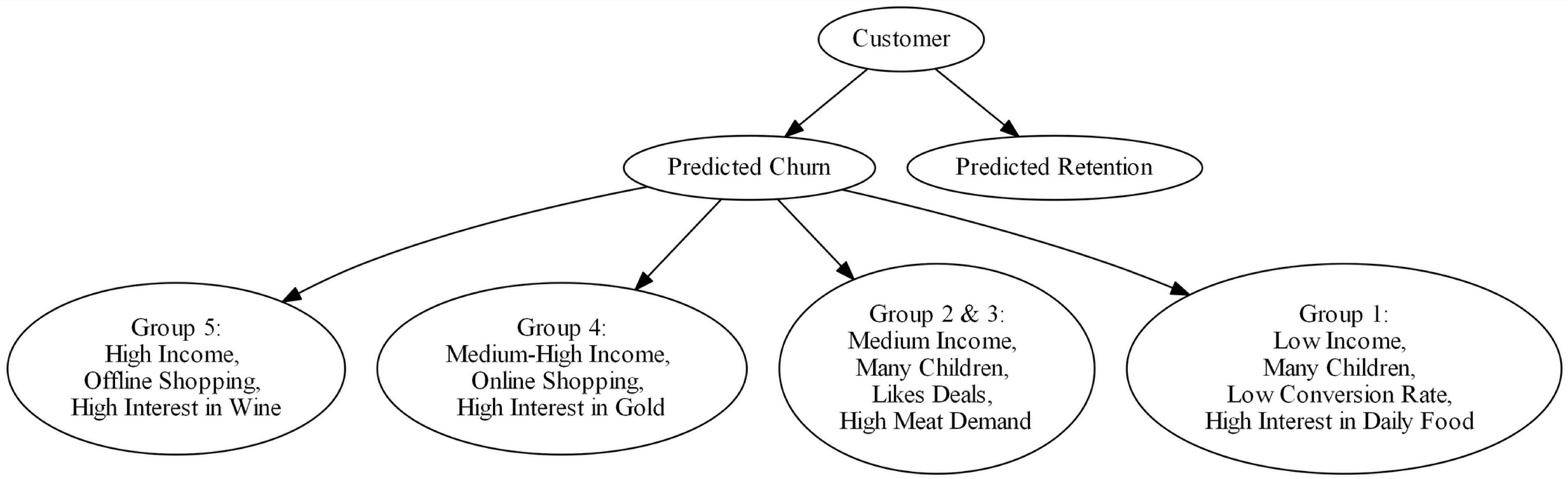
Smartphone Usage Dataset Source

Quarterly percentage of American adults reporting owning and using a smartphone, from 2011 to 2019.

Month_Year	Crashes_Per_100k	Season	Smartphone_Survey_Date	Smartphone_usage
Apr-12	**133.213685**	Spring	4/3/12	46
Apr-15	**150.077792**	Spring	4/12/15	67
Apr-16	**172.401948**	Spring	4/4/16	72
Aug-12	**145.403147**	Summer	8/5/12	44
Dec-12	**169.160811**	Winter	12/9/12	45

First 5 rows of the dataset

PERSONALISED MARKETING ALGORITHM



PERSONALISED MARKETING ALGORITHM

General Strategies:

- Send discounts, coupons via SMS, email, push notifications
- Customer feedback platform (WhatsApp) for service & complaints

Segment-Specific Churn Prevention:

1. Group 5 (High-Income, Offline, Wine Lover)
   a. VIP in-store shopping days, exclusive discounts
   b. "Wine Lover" online section, new products, tasting events
2. Group 4 (Medium-High Income, Online Shoppers)
   a. Promote gold products, additional points/discounts
   b. Premium online experience, exclusive launches, recommendations
3. Group 2 & 3 (Medium-Low Income, Many Children, Deal Seekers)
   a. Weekly meat product discount emails, loyalty coupons
   b. Family packages, discounts for budget-friendly shopping
4. Group 1 (Low-Income, Many Children, Low Conversion)
   a. Daily discounted items, flash sales on online platform
   b. Affordable "Too Good to Go" surprise bags for unsold items

CONCLUSION

Limitations:

- Churn prediction model accuracy can be further improved
- Customer segmentation may require additional refinement for more granular groups
- Limited by the features available in the dataset

Future Work:

- Enhance churn prediction model using advanced algorithms and feature engineering
- Continually update and fine-tune customer segmentation as behaviours evolve
- Integrate decision tree models for direct churn prediction and segment identification
- Gather additional customer-centric data to enrich analysis
- Implement dynamic strategies that adapt to changing customer preferences

Optimising customer value and preventing churn through data-driven insights and targeted marketing efforts. Continuous improvement and adaptation will be key to sustaining success.

Thank you!