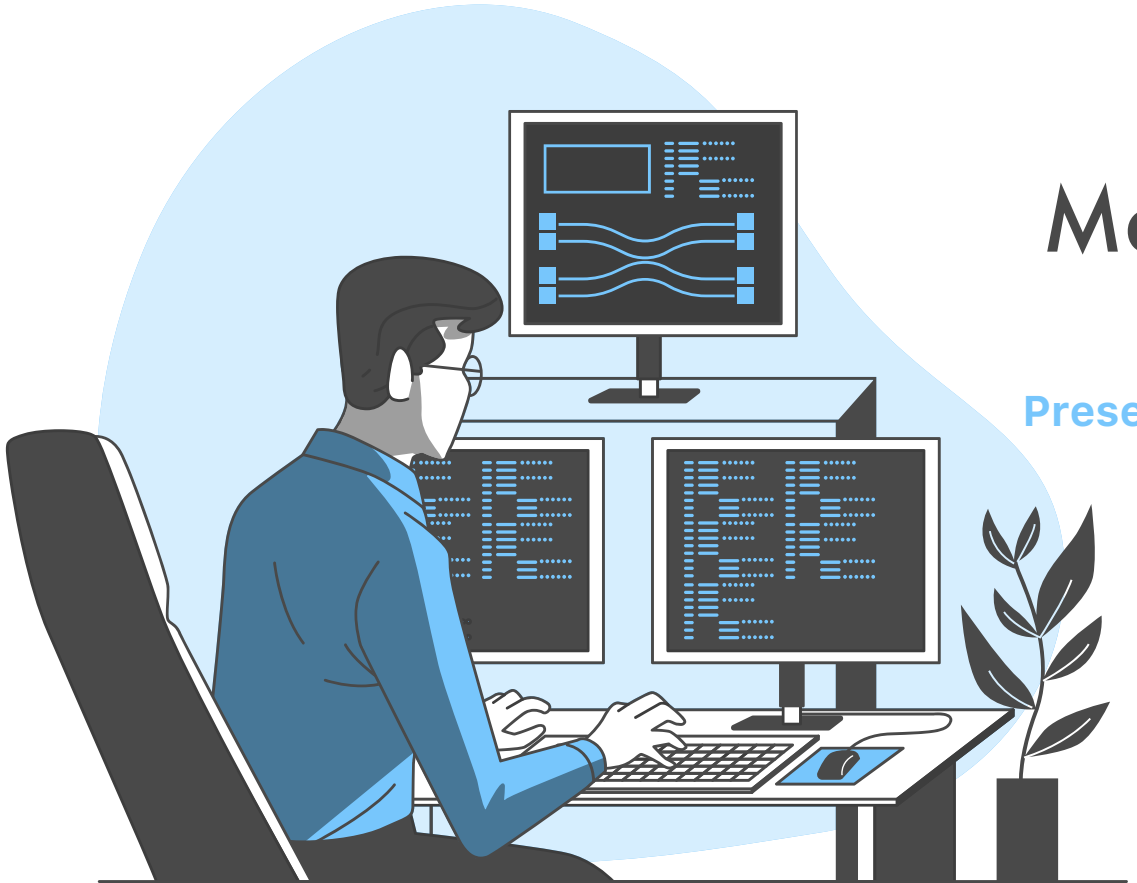


Marketing Campaign Analytics

Presented by Yung-Hui (Stephanie) Pan

Agenda

01

Introduction

- Introduce the goal of the project
- Introduce the dataset

02

Modeling Framework

- Data Preprocessing
- Feature Engineering

03

Performance of Model

- Random Forest : Variable Selection
- K–Means: Clustering

04

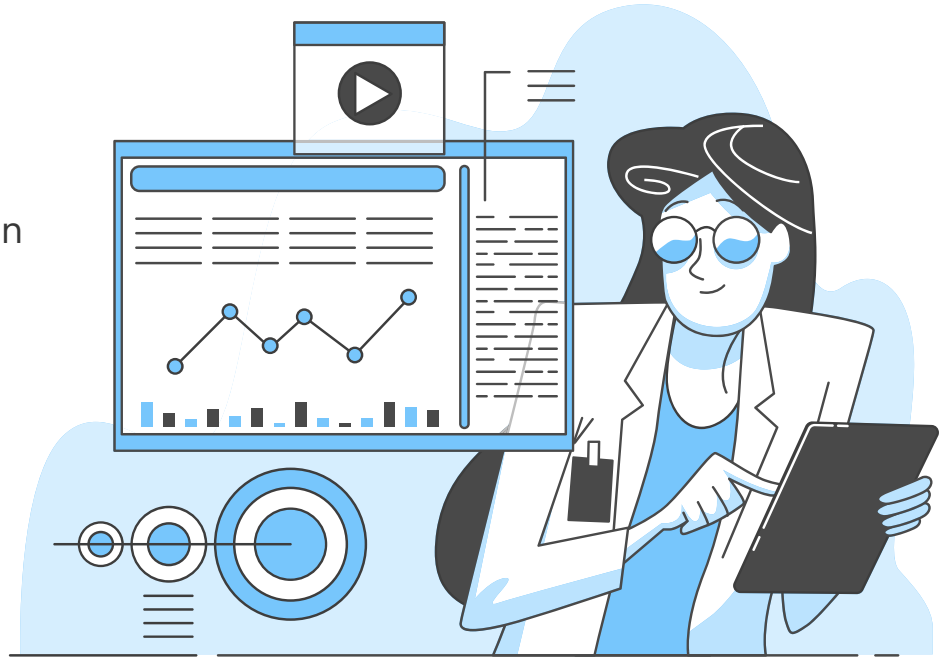
Business Solution

- Marketing Strategy for different groups of customers

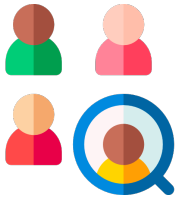
05

Model Interpretation

- Try to explain the result of model



Introduction



Customer Segmentation

Targeted Marketing Strategy

Enhanced Return on marketing campaigns

- Customer segmentation is crucial for retailers because marketers can develop more targeted marketing strategies based on different customer groups, thereby enhancing the return on marketing activities
- To implement customer segmentation, I would apply clustering algorithms such as k-means
- To determine attributes used in clustering algorithms, I will use Random Forest to find the key variables and exclude variables that are considered irrelevant

Dataset Introduction - marketing_campaign.csv

	attribute	type	meaning
Demographic	ID	int	Customer ID
	Year_Birth	int	The Year of a customer's birth
	Education*	chr	The level of education that a customer completed
	Maritial_Status*	chr	Status of Marriage
	Income	chr	Annual Income
	Kidhome	int	# of children under the age of 13 in Customer's household
	Teenhome	chr	# of children between 13-19 in Customer's household
	Dt_Customer	chr	Date of Customer Enrollment
Marketing	AcceptedCmp1	int	1 if customer accepted the offer in the 1st/2nd/3rd/4th/5th campaign, 0 otherwise
	AcceptedCmp2	int	
	AcceptedCmp3	int	
	AcceptedCmp4	int	
	AcceptedCmp5	int	
	Response	int	1 if customer accepted the offer in the last campaign, 0 otherwise

	attribute	type	meaning
Sales	MntWines	int	Dollar amount of Wines/Fruits/Meat/Fish/Sweet/ Gold purchased in last 2 years
	MntFruits	int	
	MntMeatProducts	int	
	MntFishProducts	int	
	MntSweetProducts	int	
	MntGoldProds	int	
	NumDealsPurchases	int	# of purchases made with discount
	NumWebPurchases	int	# of purchases made through the company's website
	NumCatalogPurchases	int	# of purchases made using the catalog
	NumStorePurchases	int	# of purchases made directly in-store
Others	Recency	chr	# of days since last purchase
	NumWebVisitsMonth	int	# of visits made through company's website
	Z_CostContact	int	unknown
	Z_Revenue	int	unknown

* Represents categorical variable

Modeling Framework

Input

Customer
- Demographic data
- Marketing data
- Sales data

Process

Basic Data Processing

- Missing Value Processing
- Outlier Detection
- One-hot Encoding
- Min-Max Scaling

Feature Construction

- CampaignsAccepted: 1 if customer accepted any offer in the 1st/2nd/3rd/4th/5th/last campaign, 0 otherwise
- AmtSpent: Total Amount Spent in last 2 years
- NumPurchases: Total Amount of Purchases in last 2 years

Core Algorithms

Random Forest (Variable Selection)

+

K-means (Clustering Model)

Output

Clustering Model For Customer Segmentation

Random Forest

- Goal : Divide customers into two groups - engaged and non-engaged
- Target attribute : **CampaignsAccepted** (1 if customer accepted any offer in previous campaigns , 0 otherwise)

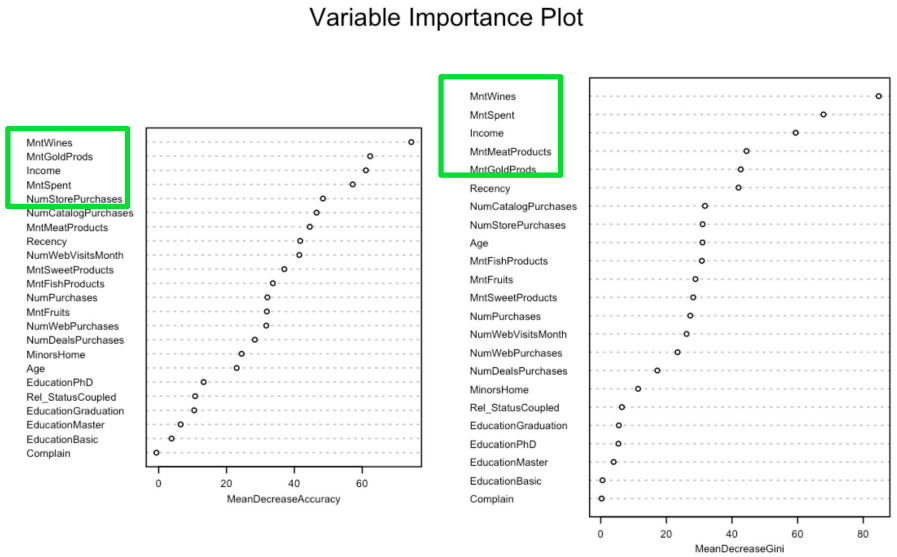
Key Metrics

Mean Accuracy	0.82
Mean Recall	0.53
Mean Precision	0.75
Mean Specificity	0.93

May suffer from imbalanced data, Class 0 : Class 1 = 7 : 3

- Validation method : k-fold cross-validation (k = 5),
- n = 2010

Variable Importance Plot (in k = 1)



Summary

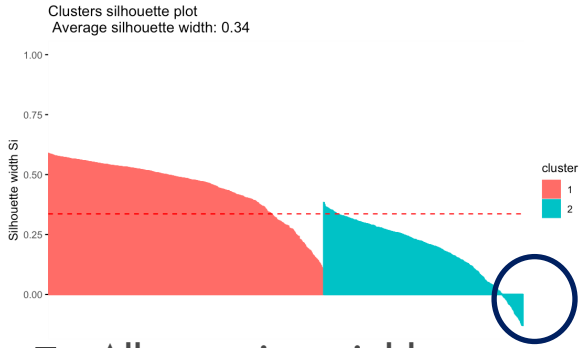
MntWines, MntSpent, MntGoldProds, Income, MntMeatProducts, NumStorePurchases are the variables with highest importance score. (based on the overall performance in each round)
Therefore, those variables would be applied to k-means model, other variables would be excluded.

K-Means Clustering

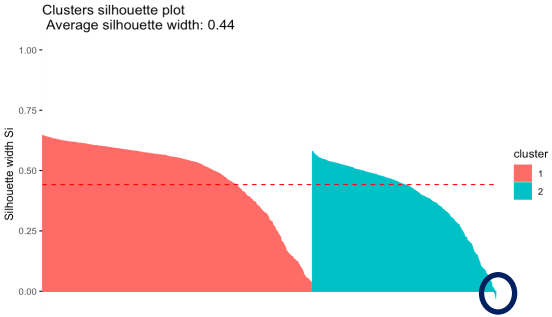
Variables Used	Average Silhouette width
All numeric variables	0.34
PCA (first 5 principal components)	0.44
Variables with highest importance score	0.52

- k= 2
- variables related to ‘Campaigns Accepted’ are excluded to prevent data leakage

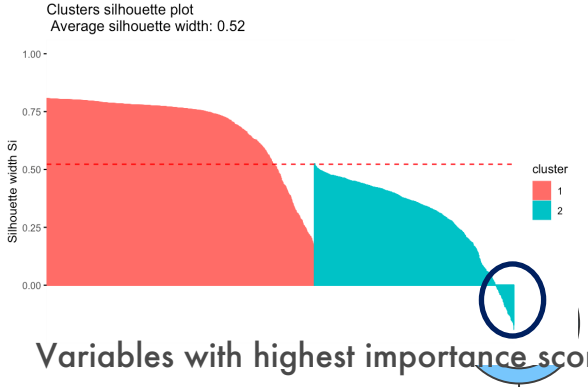
Silhouette plot



All numeric variables



PCA (first 5 principal components)



Variables with highest importance score

Clustering Results

Cluster 1
Cluster 2

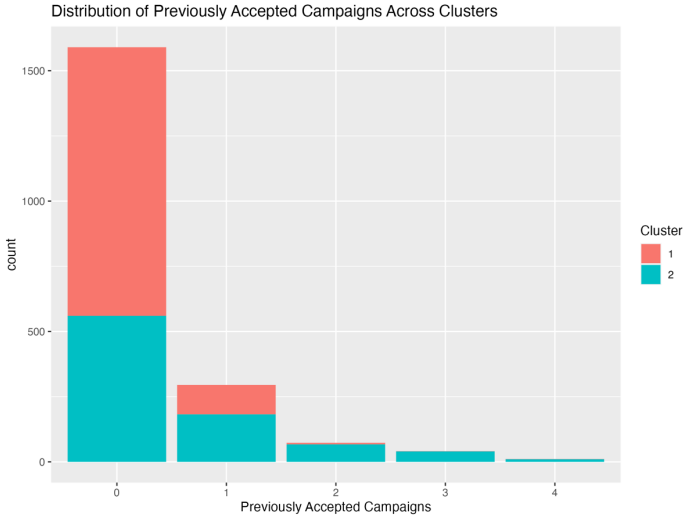
Variable	Cluster 1	Cluster 2
MntSpent	< 500	> 500
MntWines	< 500	> 500
MntMeatProducts	< 300	0～1000
Income	< 80,000	50,000～100,000
MntGoldProds	0～100	100～250
NumStorePurchases	0～8	5～15

The customer in Cluster 2 tends to

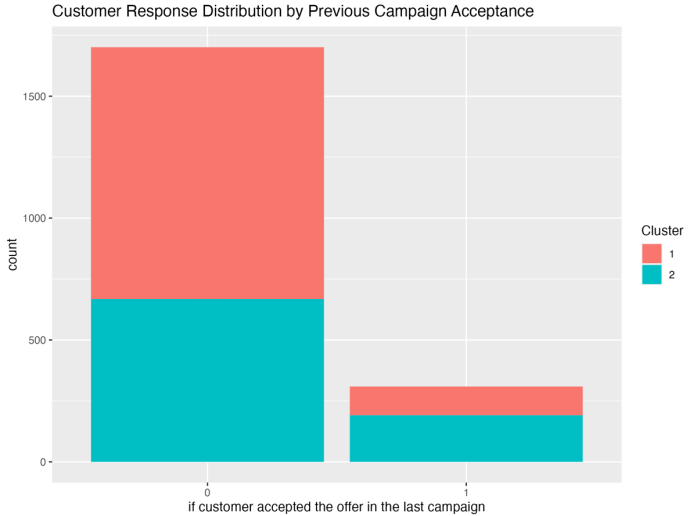
- **spend more**
- **shop more frequently in stores**
- **have higher Income**

Meaning of Clustering Results

Cluster 1
Cluster 2



Distribution of Previously Accepted Campaigns Across Clusters



Customer Response Distribution by Previous Campaign Acceptance

Cluster 2 may represent customers who are
- **more engaged in marketing campaigns**
- **more likely to engage in future campaigns**

Cluster 1 may represent customers who are
- **less engaged in marketing campaigns**
- **less likely to engage in future campaigns**

Marketing Strategy for customers in different clusters

Cluster	Attended any campaigns ?	Possible Meaning	Strategy
1	Yes	Inactive Customer	**Maintain regular communication**: Stay in touch with customers through various channels, such as email newsletters, social media, or personalized messages. Keep them informed about new products, promotions, or relevant updates to maintain engagement
	No	Customers Churn	Conduct **customer satisfaction surveys** : Regularly gather feedback from customers to assess their satisfaction levels and identify areas for improvement. Actively address concerns and make necessary changes to enhance customer experience.
2	Yes	Loyal Customers	Providing **special privileges** such as birthday bonuses, free parking, etc., to enhance customer loyalty.
	No	Potential customers	**Maintain regular communication**: Stay in touch with customers through various channels, such as email newsletters, social media, or personalized messages. Keep them informed about new products, promotions, or relevant updates to maintain engagement

Why those variables are important

Chi-Square Test with CampaignsAccepted

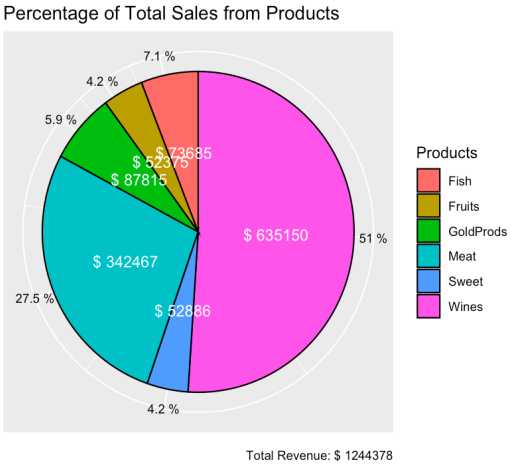
Variable	P-value	Cramer's V	Association	Degree of Freedom
MntWines	< 2.2e-16	0.356	Medium	1
MntSpent	< 2.2e-16	0.309	Medium	1
MntGoldProds	9.08e-07	0.113	Small	1
Income	< 2.2e-16	0.204	Small	1
MntMeatProducts	< 2.2e-16	0.254	Small	1
NumStorePurchases	4.554e-15	0.177	Small	1

*The numerical variables are transformed into categorical variables by binning before applying chi-square test
**Break Points of binning are based on the result of k-means clustering

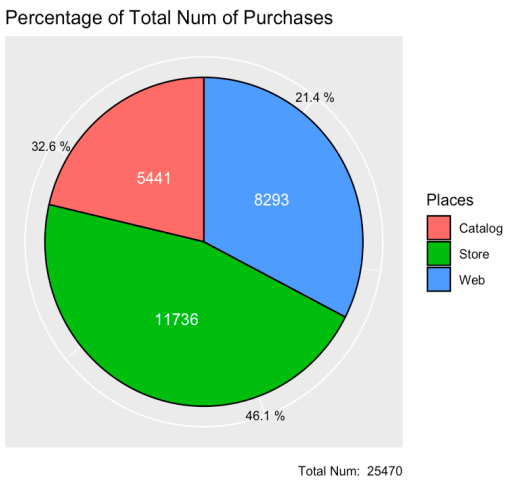
Why those variables are important

Charts

A



B



Findings

A. Wines easily account for a majority of total sales (at 51%), with meat products being a second with nearly half the sales as Wines (at 27.5%). Other products accrue a similar amount of sales

B. Most of sales come from stores, with catalog being a second and Web being a third

It is not surprising that MntWines, MntMeatProducts, MntGoldProds, NumStorePurchases are considered important by Random Forest

Thank You!