



# Customer Segmentation

Clustering customers based on consumer behaviour



# Table of Contents

**01**

## **Introduction**

About the project

**03**

## **Clustering**

How we cluster the customers

**02**

## **EDA & Data Cleaning**

How data is checked  
and explored

**04**

## **Profiling & Marketing Strategies**

Profiling the clusters,  
& discuss marketing  
strategies



01

# About the Project

The Context, The Data, The Approach

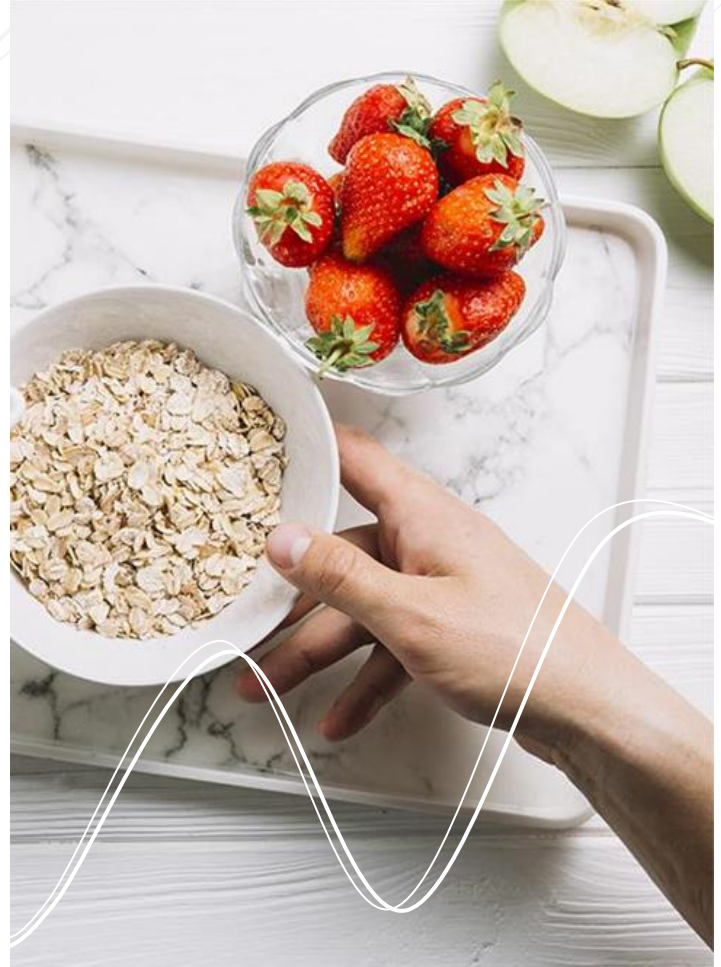
# The Project

## **Aim:**

Segment retail customers into clusters based on their consumer behavior, to strategies marketing efforts to improve sales.

## **How:**

Clustering algorithms from [scikit-learn](#).



# The Approach



## Data Cleaning

### Checked & Addressed:

Missing Data  
Duplicate Data  
Data Type  
Data Values  
Outliers

Rename features,  
Drop features,  
Create features

## EDA

Consumer Behaviour

Consumer  
Demographic

Correlations



## Clustering

Based on  
Consumer Behaviour

StandardScaler

KMeans  
DBSCAN  
AgglomerativeClustering

Silhouette Score

## Profiling & Marketing Strategies

Consumer Behaviour

Consumer  
Demographic



# The Data

Total: 27 features



## People

Demographics

*Age, Income, Education,  
Children, Relationship*

## Consumer Behaviours



## Place

Place of purchase:

*Web, Catalog, Store*



## Promotion

6 Campaigns Offers

*Whether customers  
accepted them or not*



## Purchase

Amount spend of  
product types

*Wine, Meat, Fish, Fruits,  
Sweets, Gold*

# 02

## EDA

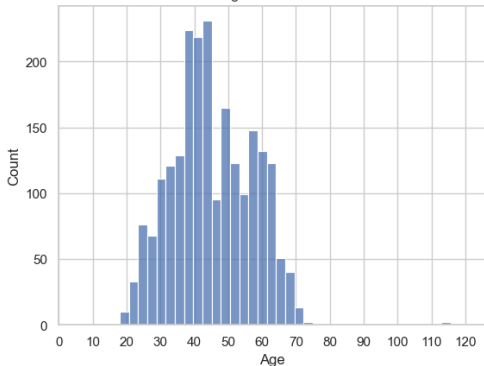
Exploratory Data Analysis





# Demographic

Age Distribution



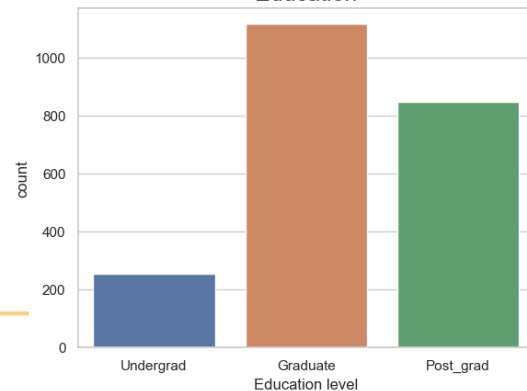
## AGE

Mainly Middle-age Baby Boomers  
IQR 37-55 yo

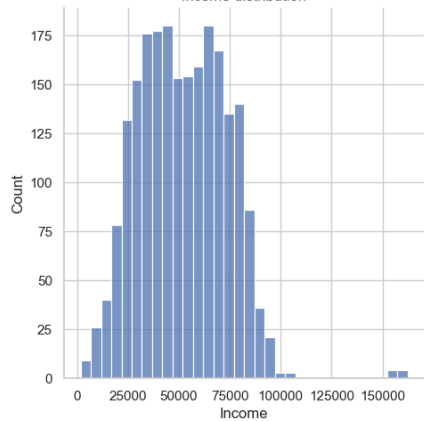
## EDUCATION

Dominantly GRADUATES

Education



Income distribution



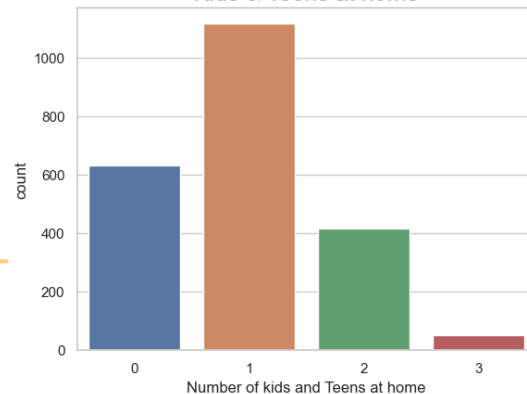
## INCOME

IQR \$35k - \$68k

## HOME KID/TEEN

~ 30% Have no kid/teens

Kids & Teens at home







## Consumer Behaviour



### PRODUCT PURCHASE

Consumers spent most on **WINE** and **MEAT**



### PROMOTION

Campaign acceptance rate very low (typically **7%**)



### PLACE

**IN-STORE** has most purchases

The background of the slide features a top-down view of a wooden cutting board on a light-colored wooden surface. On the board, there is a whole orange in the upper left, a sliced orange half in the center, and a glass of orange juice in the upper right. A thin white line, resembling a sine wave, is drawn across the image, passing through the oranges and the glass. Overlaid on the right side of the image is a large, bold orange number '03'.

# 03

## **Clustering**

Clustering approaches explored

# Ideally, the clusters should:

**0.5**

## **Silhouette Score**

Have minimally a  
Silhouette score of 0.5

*Silhouette Score measures how  
well-separated clusters are from each other*

**10%**

## **Cluster Size**

Have minimally around 10%  
of customers per cluster  
(~200 pax per cluster)

*In order to avoid having overly small clusters  
that are insignificant for marketing efforts*

**3-5**

## **Clusters**

Have 3-5 Clusters to  
keep marketing  
strategies focused



# Clustering Approach Explored

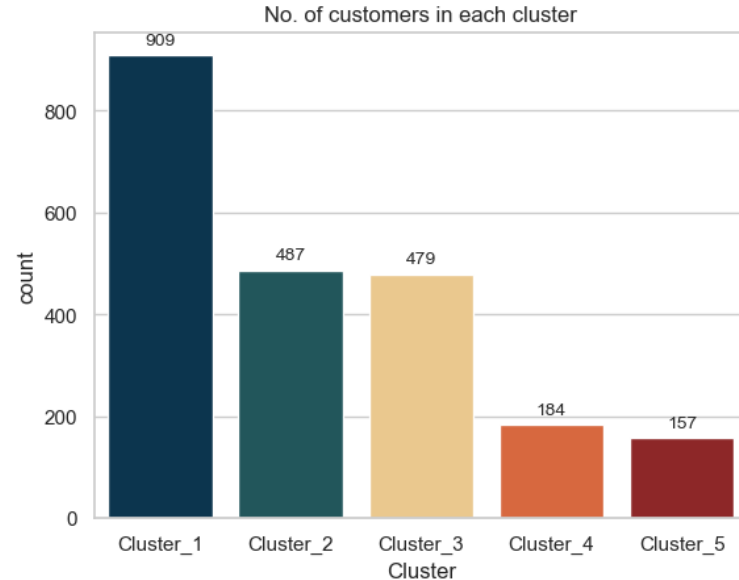
<i>Note: For each approach, the cluster with the closest characteristics to ideal has its score and cluster description displayed here</i>		KMeans	DBSCAN	Agglomerative Clustering
1 <sup>st</sup> Attempt	Silhouette Score	0.341	0.447	0.305
	No. of clusters	3	5	3
	Cluster size	Range from 200-1230	Range from 6 - 2141	Range from 229-1265
2 <sup>nd</sup> Attempt (After applying PCA)	Silhouette Score	0.424	0.460	0.344
	No. of clusters	5	4	4
	Cluster size	Range from 157-913	Range from 5 - 2155	Range from 136 - 838
Selected Clustering Approach				

# Selected Clustering Approach

Description	
Algorithm	KMeans
Features used	1 <sup>st</sup> four PCA features which explains 66% of the variation in the data
Number of Cluster	5
Silhouette Score	0.424

## Observations made when exploring different clustering methods:

- Clustering with higher Silhouette Score have greater range in cluster size.
- It may have scored close to 0.5, but will face significantly imbalanced clusters.
- As such, the clustering approach is chosen despite having a score of 0.4, as the customers appear to be more balanced across the clusters – each clusters makes up 10-40% of the customers.



A top-down view of a workspace. In the upper right, a portion of a silver laptop is visible, showing keys like 'V', 'F', 'D', 'C', 'X', 'Z', 'Command', 'Option', and 'Control'. To the left of the laptop is a silver fork. In the center is a white bowl filled with a fresh salad of lettuce, sliced tomatoes, cucumbers, and basil leaves. In the lower right, there is a glass of red wine. A large green number '04' is overlaid on the left side of the image, and a green banner at the bottom contains the title and subtitle. Thin white lines are drawn over the image, looping around the bowl and extending towards the laptop and wine glass.

# 04

## The Clusters

Profiling the 5 clusters and  
strategising marketing efforts



# Plan:

**Cluster  
Overview**

Income  
VS  
Total Purchases, Total Expenditures, Total Campaigns  
Accepted

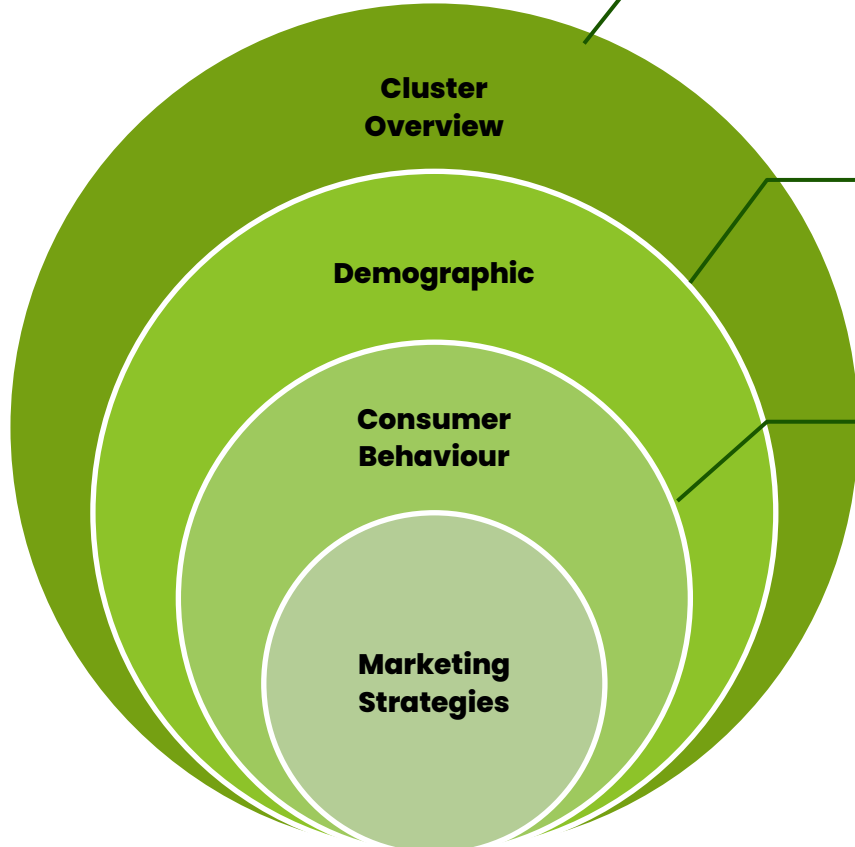
**Demographic**

Age, Education, Kids/Teens at home

**Consumer  
Behaviour**

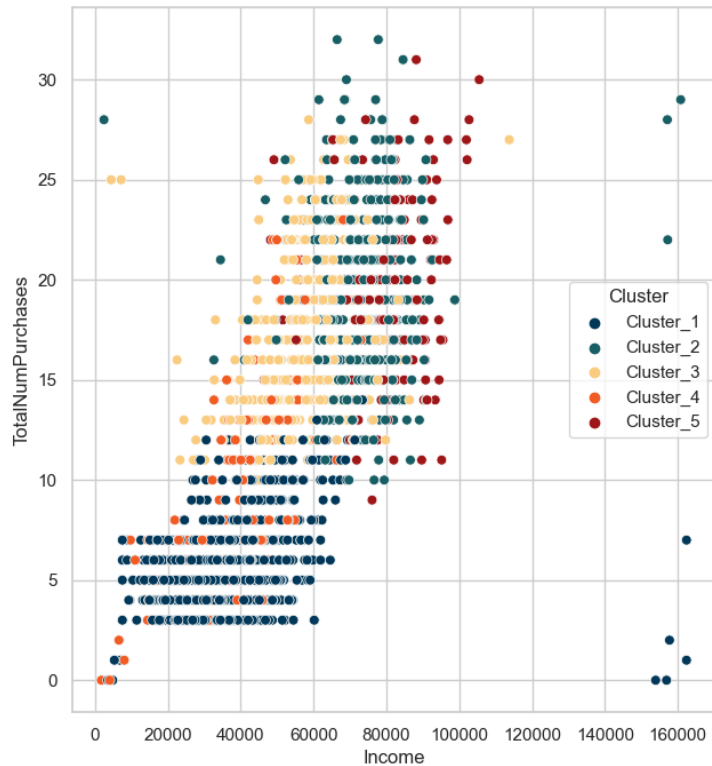
Points of purchase,  
Spending on Product Types  
Acceptance rate of Campaigns offers

**Marketing  
Strategies**

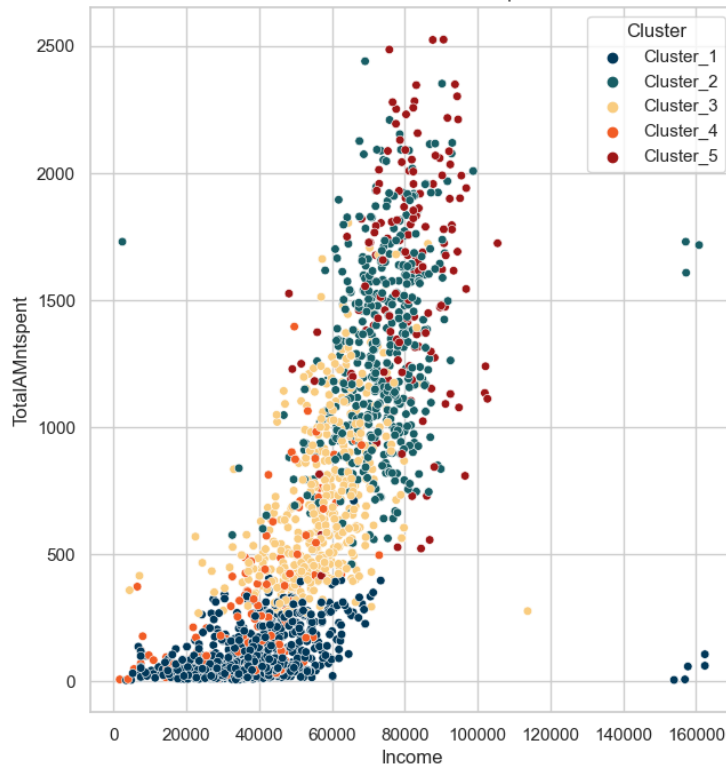


# Cluster Overview

Income VS Number of Purchases



Income VS Total Amount Spent



## Comments

Generally  
Customers with  
Higher income  
Have  
More purchases  
Higher expenditures

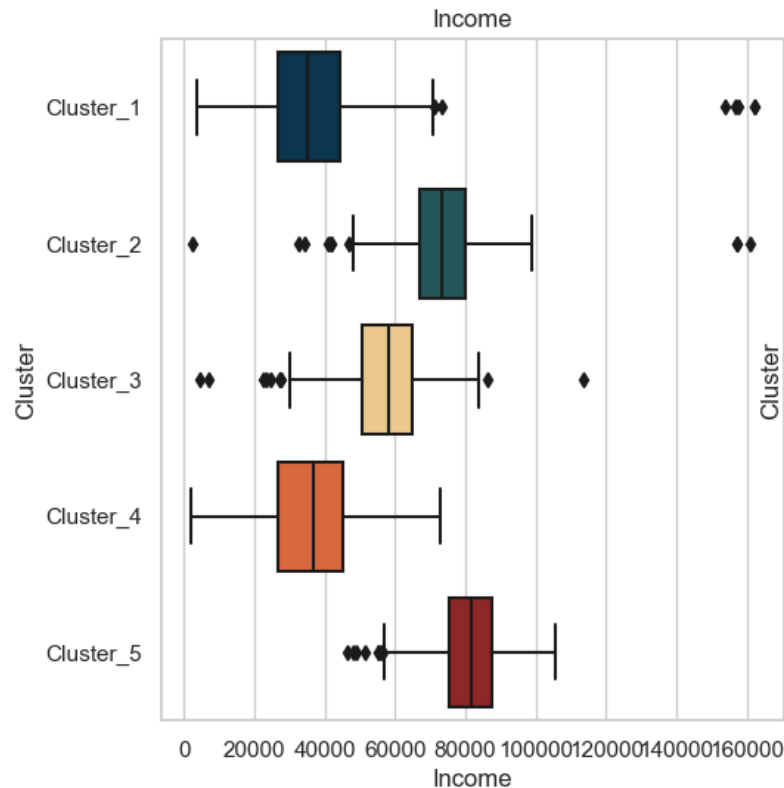
# Cluster Overview

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High

## Some unique individuals:

In **Cluster 1**: several customers have high income >150K but have consumer behavior similar to the lowest income group individuals

In **Cluster 2**: several customers have lowest income <500 but have consumer behavior similar to the higher income group individuals



# Cluster Overview

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High
Purchases & Expenses	Low	High	High	Low	High
Campaigns Accepted per customer	Zero	Mostly Zero	Zero	1 to 2	2 to 4

## Generally:

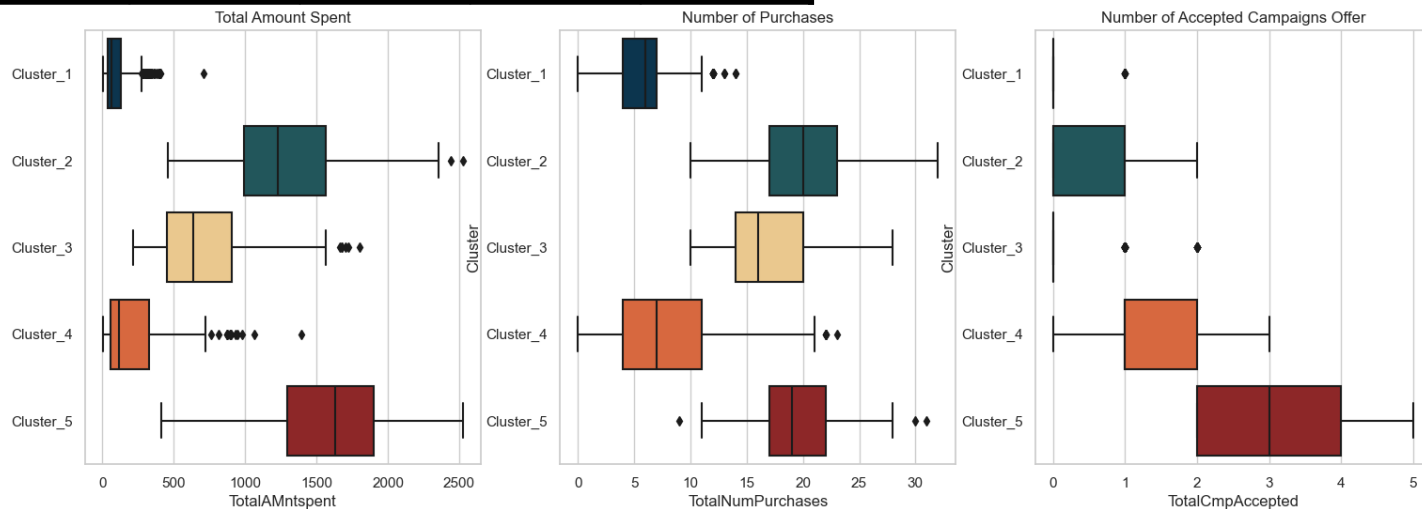
### Observation (1)

Cluster 1 and Cluster 4 who have lower income tend to have lower purchase and expenditure

Cluster 2 and Cluster 5 who have higher income tend to have higher purchase and expenditure

### Observation (2)

Cluster 4 and Cluster 5 are more responsive to campaign offers

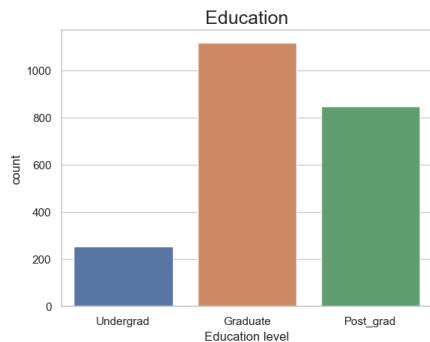


# Demographic

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High
Purchases & Expenses	Low	High	High	Low	High
Campaigns Accepted per customer	Zero	Mostly Zero	Zero	1 to 2	2 to 4



## Education



Cluster Education distribution follows the general education distribution.

No distinct difference across clusters



## Home Kid/Teen

Dominantly have Kids  
~45% have Teens

Cluster 2

Cluster 5

Cluster 3

Dominantly **no** Kids  
Dominantly have 1 Teen

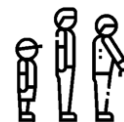
Cluster 1

Cluster 4

Dominantly **no** Kid  
Dominantly **no** Teen

Cluster Age distribution follows the general age distribution.

No distinct difference across clusters



Age

# Consumer Behaviours

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High
Purchases & Expenses	Low	High	High	Low	High
Campaigns Accepted per customer	Zero	Mostly Zero	Zero	1 to 2	2 to 4



## Product Expenses

Similar to general trends,  
Wine and Meat are  
most spent on across clusters



Cluster 3

Cluster 5



Cluster 3 & 5 spent proportionally more on wine (median 55% & 65%)  
compared to other clusters (~40%)

Cluster 4



Cluster 4 spent proportionally the most on Gold (median ~25%)  
compared to other clusters (median <10%)



# Consumer Behaviours

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High
Purchases & Expenses	Low	High	High	Low	High
Campaigns Accepted per customer	Zero	Mostly Zero	Zero	1 to 2	2 to 4



Store is most popular across clusters



VS



Web is equally popular across cluster

Cluster 2 Web & Catalogue are equally popular for  
Cluster 5 Cluster 2 & 5

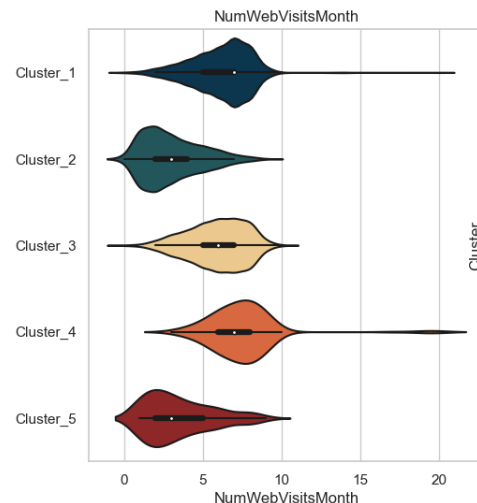
Cluster 1 Catalogue is least popular &  
very rare in Cluster 1 (median 0%)



Points of Purchase



Web Visits Last Month



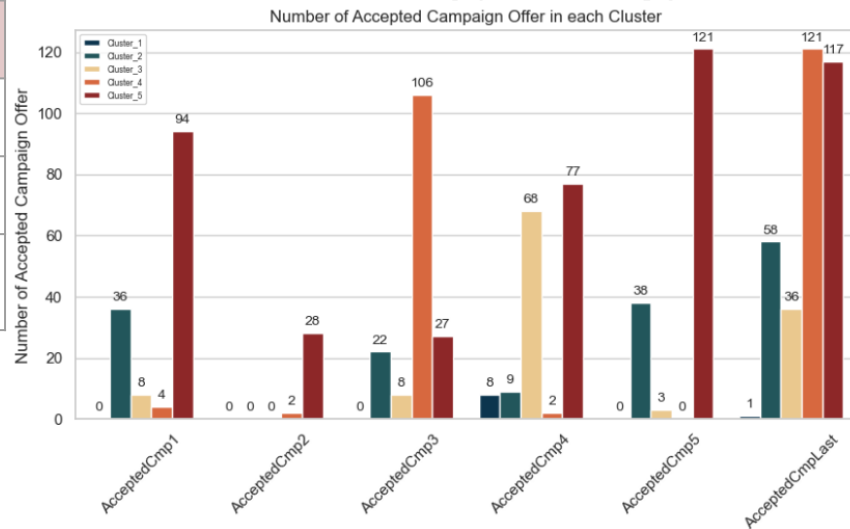
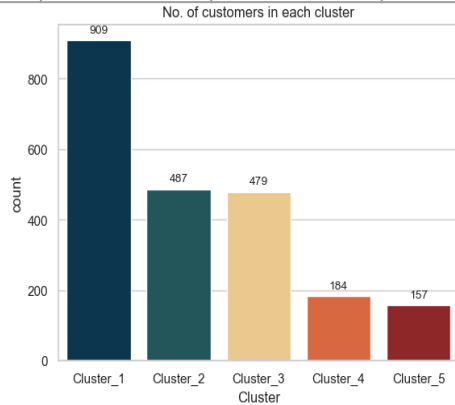
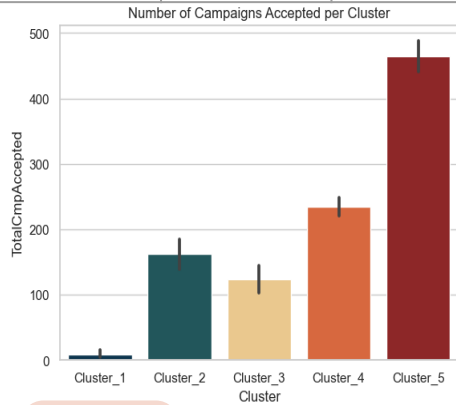
Cluster 2 Visited Web much lesser than  
Cluster 5 other clusters

# Consumer Behaviours



## Campaigns Offer Acceptance

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Income	Low	High	Moderate	Low	High
Purchases & Expenses	Low	High	High	Low	High
Campaigns Accepted per customer	Zero	Mostly Zero	Zero	1 to 2	2 to 4



'Last' Campaign was most popular

Cluster 5 Had most acceptance across the six campaigns

Cluster 4 Had favoured campaign 3 and 'Last' campaign

Cluster 3 Had favoured campaign 4 and 'Last' campaign

Cluster 4  
Cluster 5 have highest campaigns acceptance relative to cluster size

# Cluster Summary

## Cluster 1

~40% of customers  
<10 response to campaigns  
in past 2yr

Lower income (IQR \$25K-45K)  
Low spenders (<\$500 in past 2yrs)

Have Kids and/or Teens at home

## Cluster 4

~10% of customers  
Campaigns enthusiast  
(Especially Cmp 3 & Cmp 'Last')

## Cluster 3

~20% of customers  
Moderate income (IQR \$50K-65K)  
Moderate spenders (IQR \$500-1K in past 2yrs)  
No Kids, but dominantly 1Teen at home  
Some response to campaigns,  
particularly to campaign 4 & Last Campaign

## Cluster 2

~20% of customers  
Some response to campaigns

Higher income (IQR \$65K-80K)  
Higher spenders (\$1K-2K in past 2yrs)

No Kids and Teens at home

## Cluster 5

~10% of customers  
Campaigns enthusiast  
(Especially Cmp 3 & Cmp 'Last')

# Possible Marketing Strategies



**Cluster  
1**

**Cluster  
4**

**Cluster  
2**

**Cluster  
5**

## **‘Daily Essential Promotions’**

Running price promotions on essential groceries items to maintain footfall in stores & increase sales volumes

## **‘Self-Indulgence Promotions’**

Targeted promotions on more luxurious items like wine & meat to increase basket size for high income consumers and drive profitability for the retailer

# Possible Marketing Strategies



Cluster  
1

Cluster  
4

Cluster  
2

Cluster  
5

## **‘Participate in 3 Campaigns & Win’ promotions**

To encourage consumers responses to more campaigns by offering prizes in return for participating in different campaigns over a period of time.

Objective of this promotion is to drive customer loyalty & change consumer behaviour of Cluster 1 & 2 so that they behave more like Cluster 4 & 5 (campaign enthusiasts)



**End**