

# Customer Segmentation: Clustering

Wang Bi



# Project Structure

01

## Introduction

- Project Background
- Data Introduction & Process

02

## Cluster Model

- Four Different Model
- Model Selection

03

## EDA

- Customer Feature & Behavior
- Purchase Channel
- Customer Activation

04

## Profile

- Customer Profile
- Strategy Suggestion



01

# Introduction



# Background of Customer Clustering

## Definition:

- Customer clustering is the grouping of customers based on their similarities to better understand and meet the needs of different groups.

## Significance:

- Personalized Marketing and Service
- Enhanced Customer Satisfaction and Loyalty
- Increased Sales and Profits
- Predictive Insights and Planning



# About Data



## Dataset

This dataset comes from Kaggle. It contains 1000+ data items and 27 variables. The main dimensions are people, products, promotions and channels.

## Data Cleaning

Before starts, I delete NULL value, change the data types, drop outliers, do feature engineering and Scale the feature using the standard scaler.



02

# Clustering



# Clustering Model

## Four Clustering Models

- K-means
- Agglomerative Clustering
- Spectral Clustering
- DBSCAN



## Model Selection

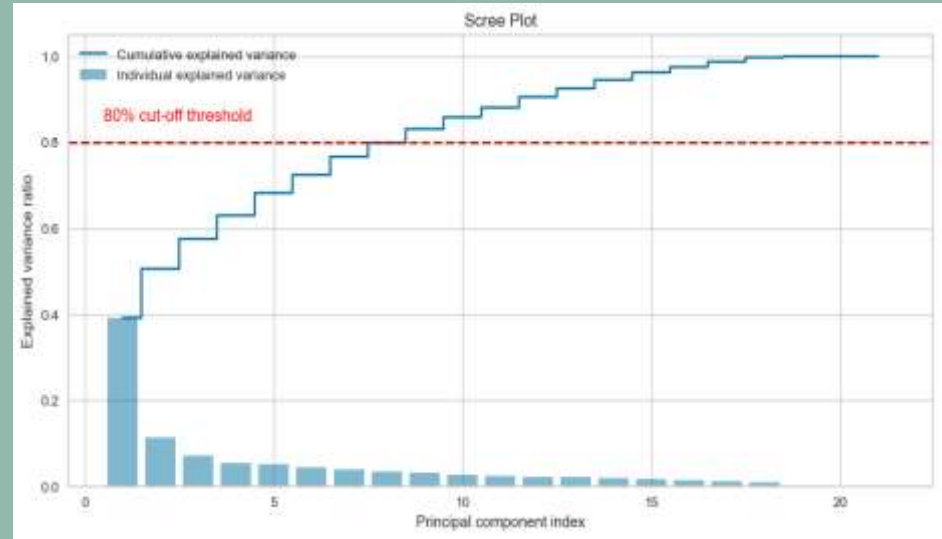
- Calculate the average silhouette score for every model
- Select model with highest score



# Clustering Model

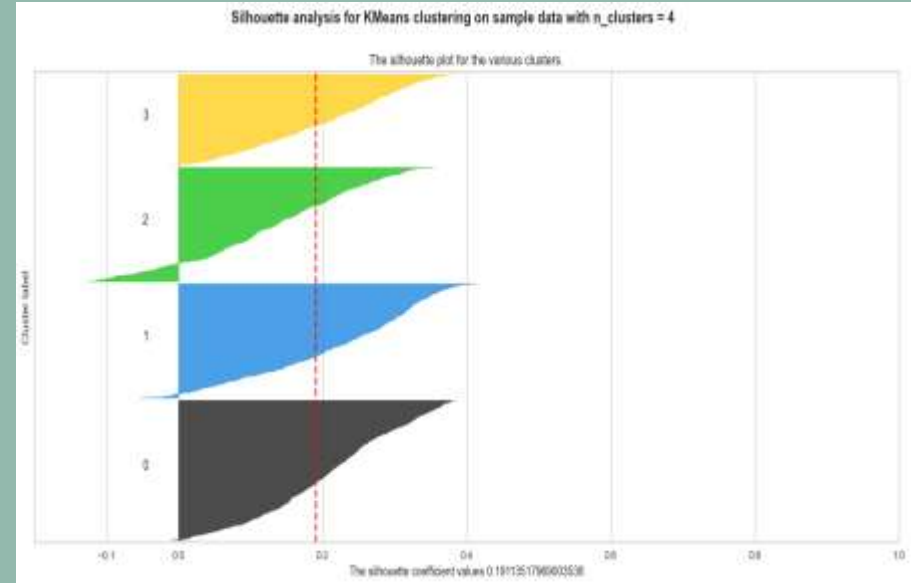
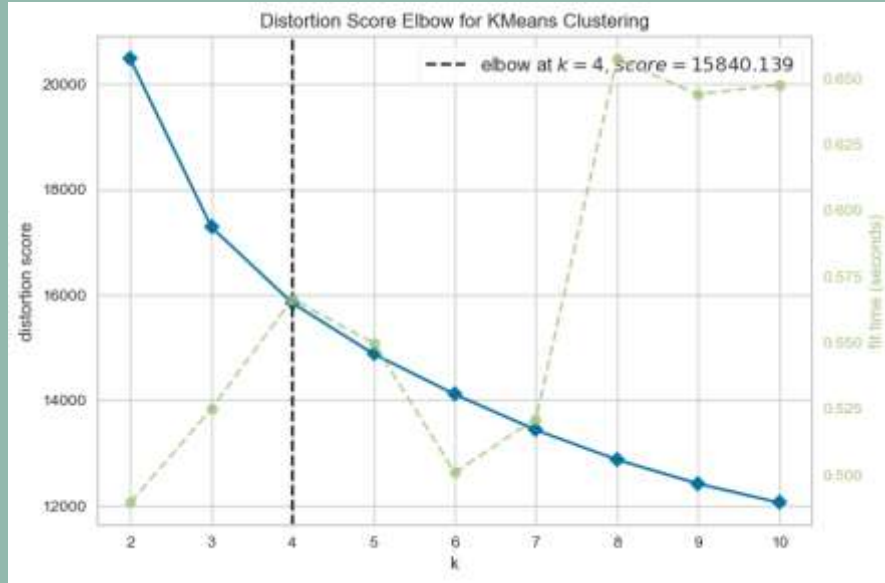
## PCA before Clustering

- 27 features is too much
- I picked eight principal component while maintaining 80% explanatory power for the variance of the variables





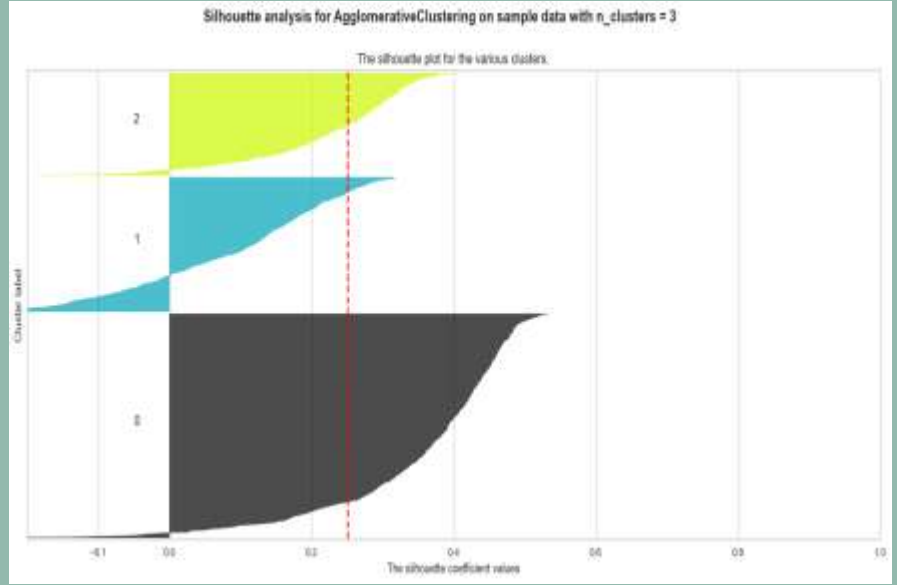
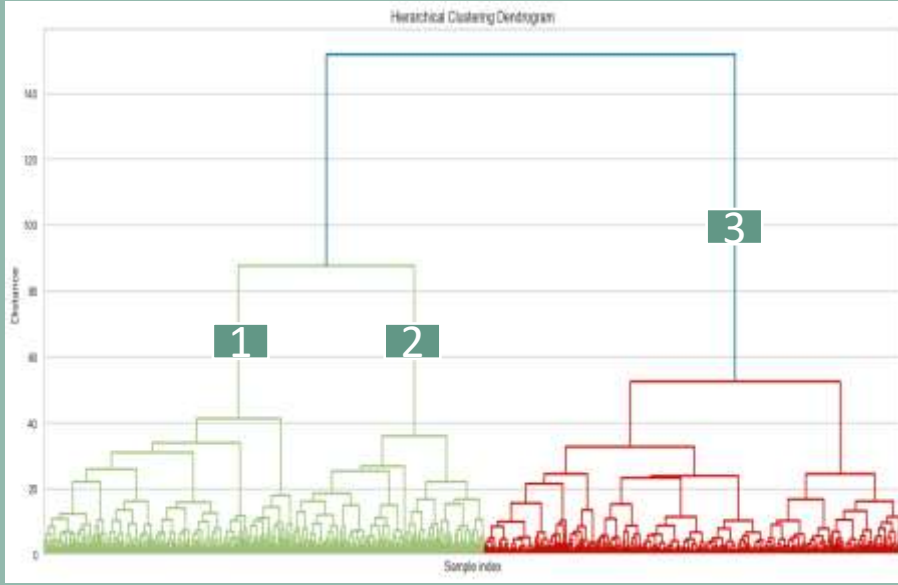
# Model1: K-Means



- Choose cluster number: Calculate distortion score and use Elbow rule to select **clusters = 4**
- Calculate average silhouette score = 0.1911



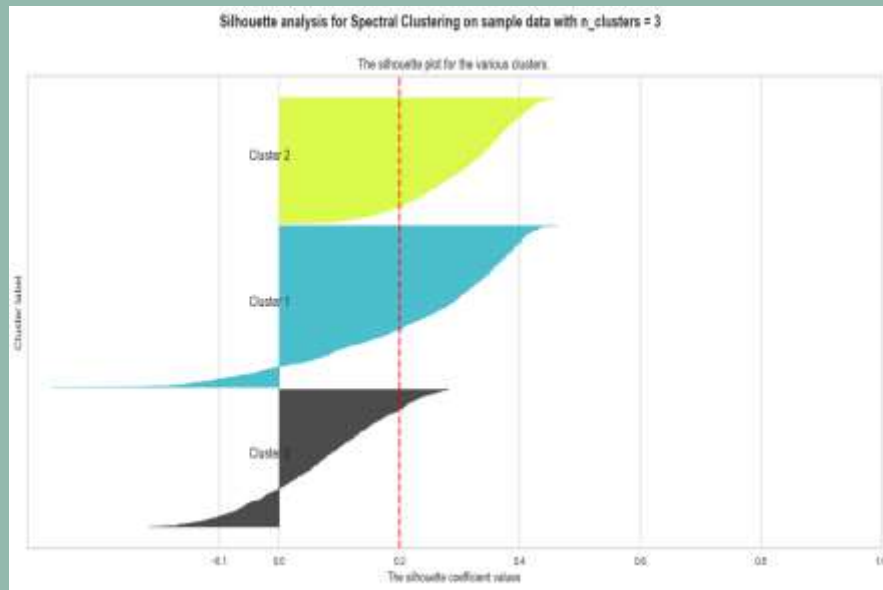
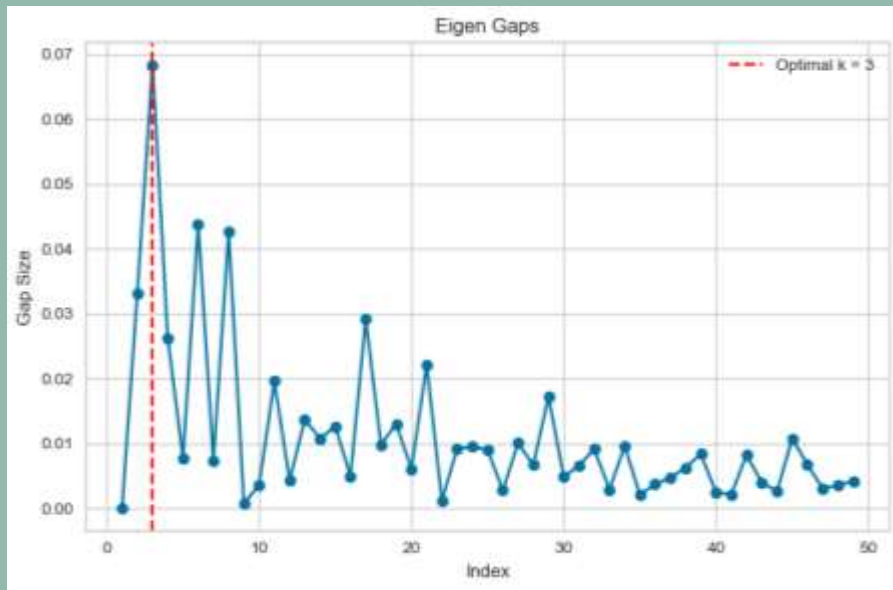
# Model2: Agglomerative Clustering



- Choose cluster number: use Dendrogram to find clusters with the longest merge distance, and finally set clusters = 3.
- Calculate average silhouette score = 0.2507



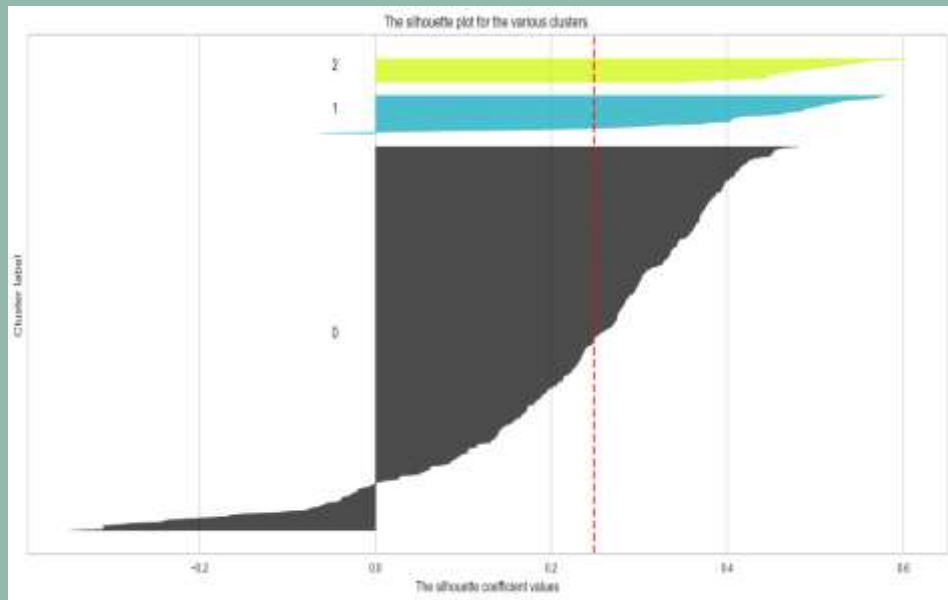
# Model3:Spectral Clustering



- Choose cluster number: calculate eigen gaps and select cluster number with highest gap size, and finally set clusters = 3
- Calculate average silhouette score = 0.1999



# Model4:DBSCAN



```
df['d_cluster'].value_counts()
```

✓ 0.0s

-1 1861

0 309

2 26

1 16

Name: d\_cluster, dtype: int64

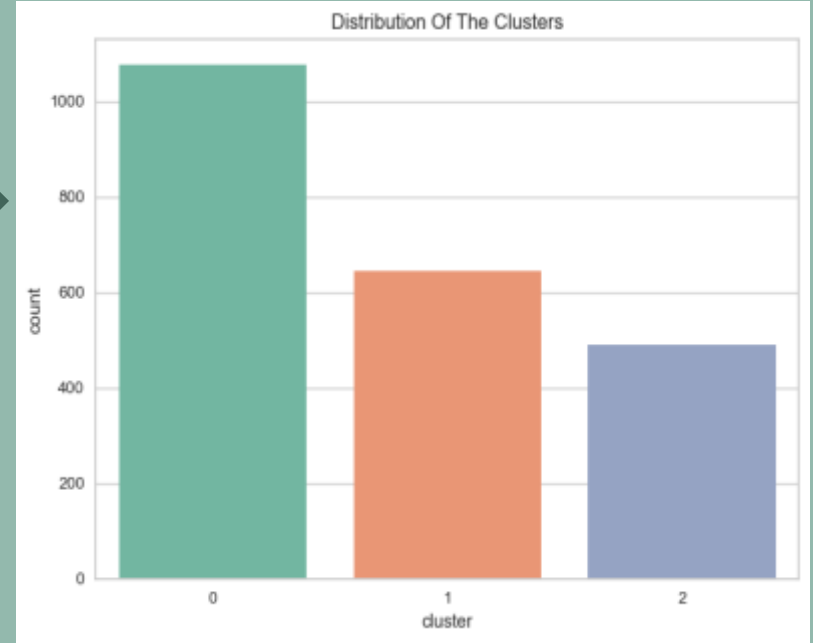
- “**d\_cluster**” is the cluster label through DBSCAN
- “-1” label means noisy point

- No need to set cluster number in advance, calculate average silhouette score = 0.2487
  - Distribution of cluster is markedly uneven
  - Many data points are clustered as noisy point
- } → **Not a suitable model**



# Model Selection

Model	Average Silhouette Score
K-Means	0.19
Agglomerative Clustering	0.25 ★
Spectral Clustering	0.20
DBSCAN	0.24



- Select Agglomerative Clustering finally
- The clusters seem to be fairly distributed

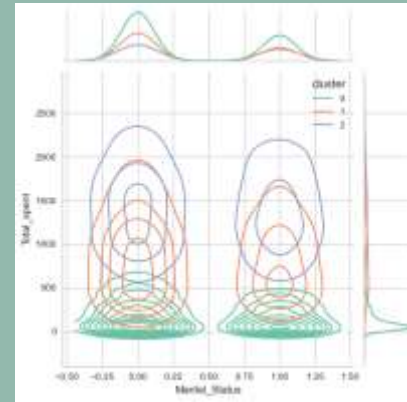
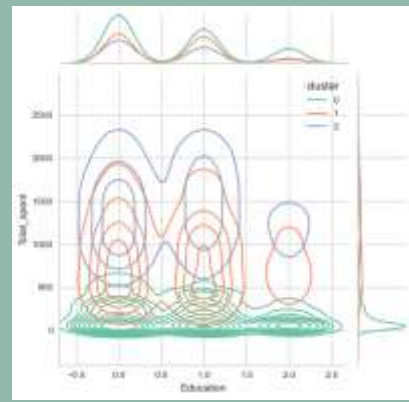
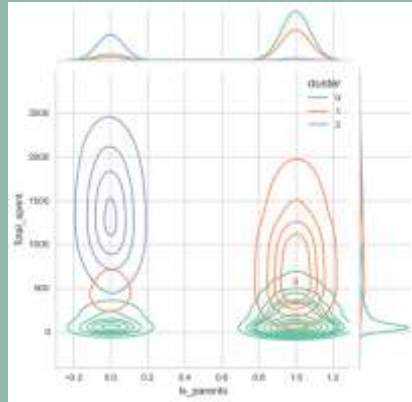
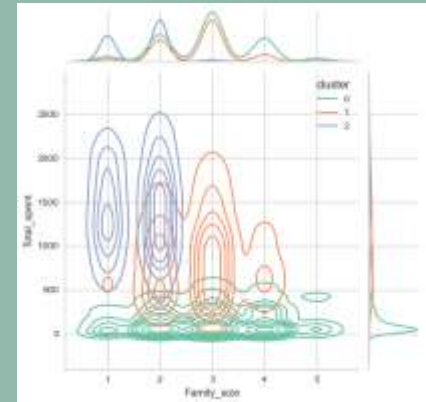
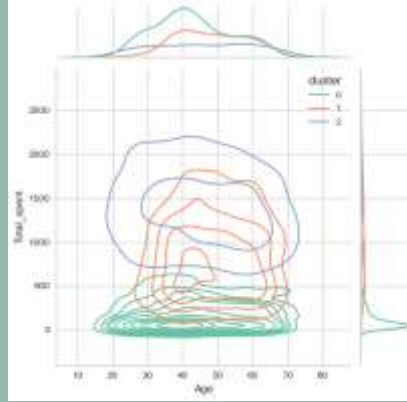
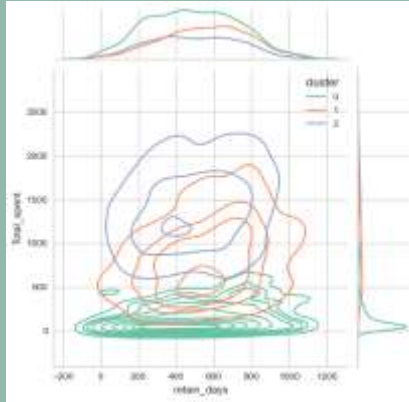
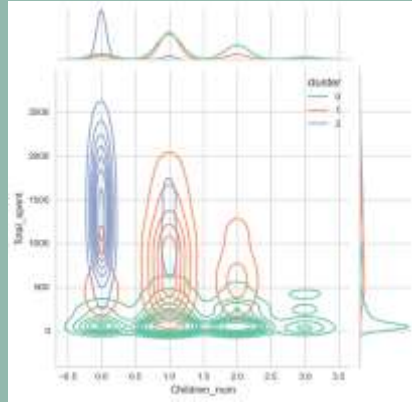


03

# Exploratory Data Analysis



# Customer Features

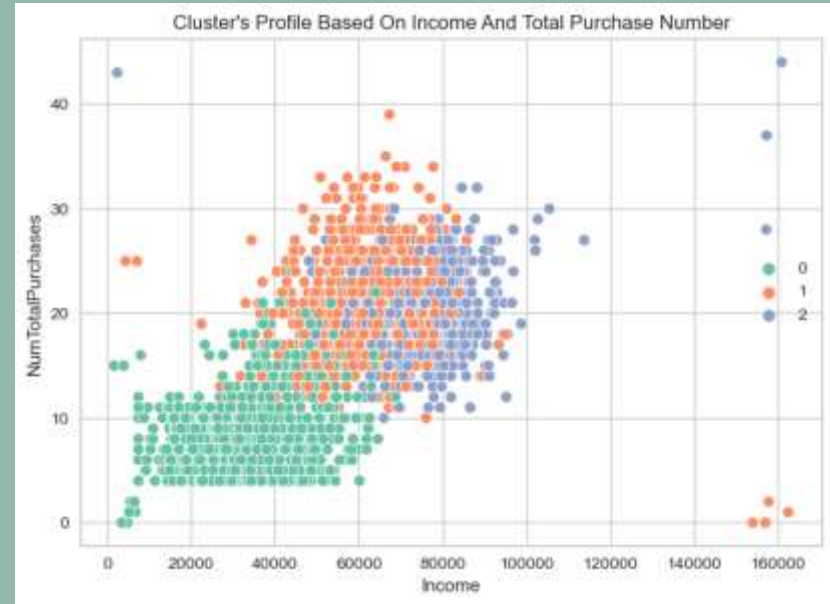


## Kernel Density Estimate plot

- Final analysis shown in profile part



# Customer Features

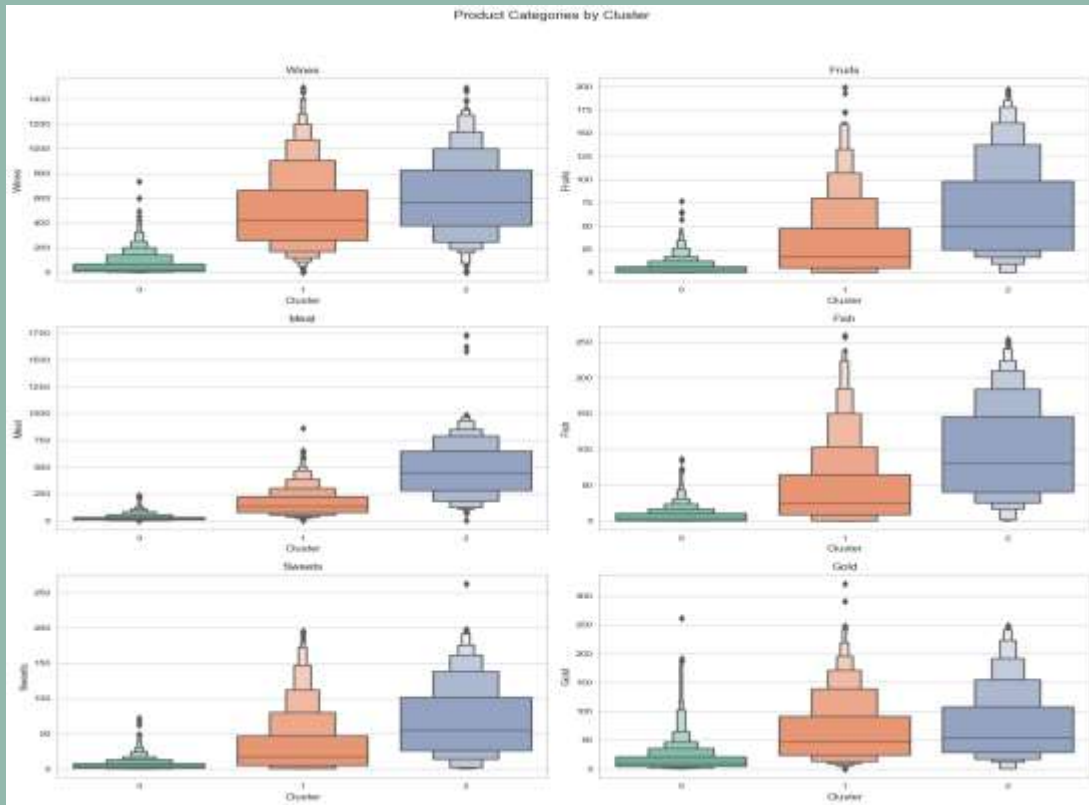


- High-income customers do not necessarily make more frequent purchases, may because they prefer high-value goods
- Total spent: Cluster2 > Cluster 1 > Cluster 0





# Product Preference

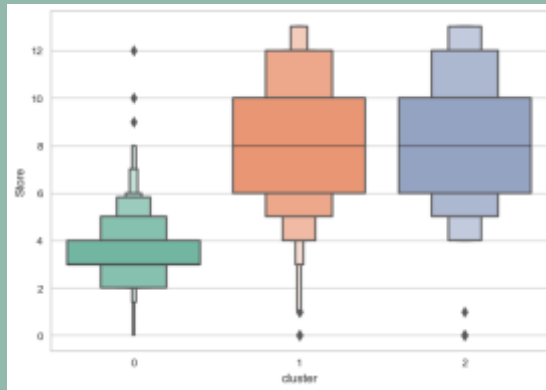
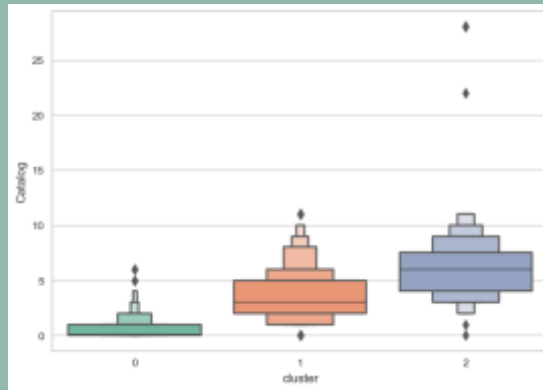
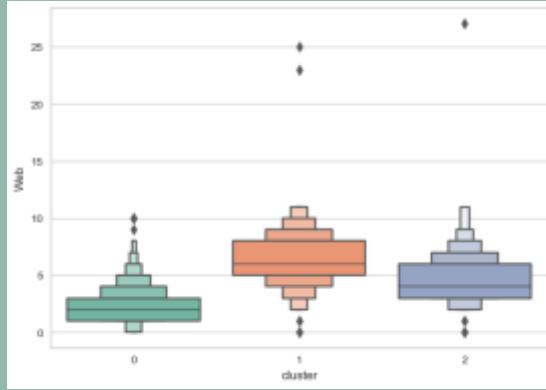


## Boxplot

- Total purchase number:  
Cluster2 > Cluster 1 > Cluster 0
- Cluster 2 spends more than the other two groups on most product categories, especially on “Meat” and “Wine” products.



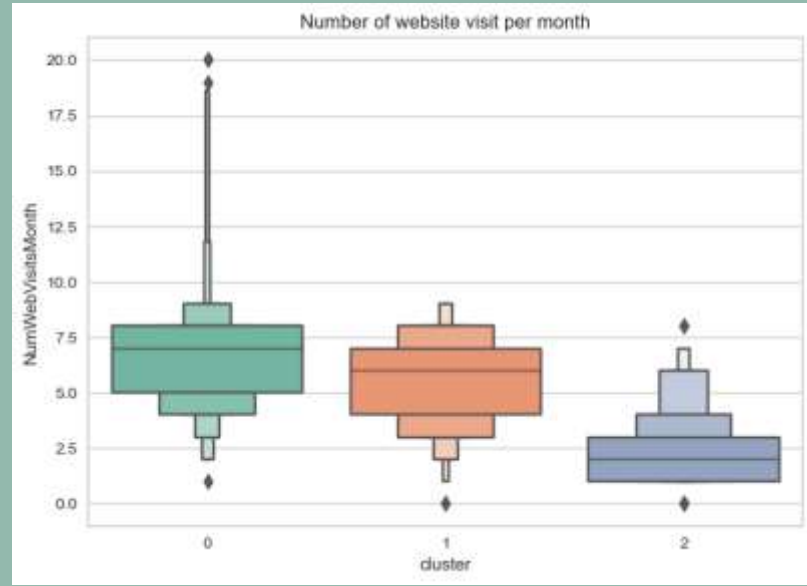
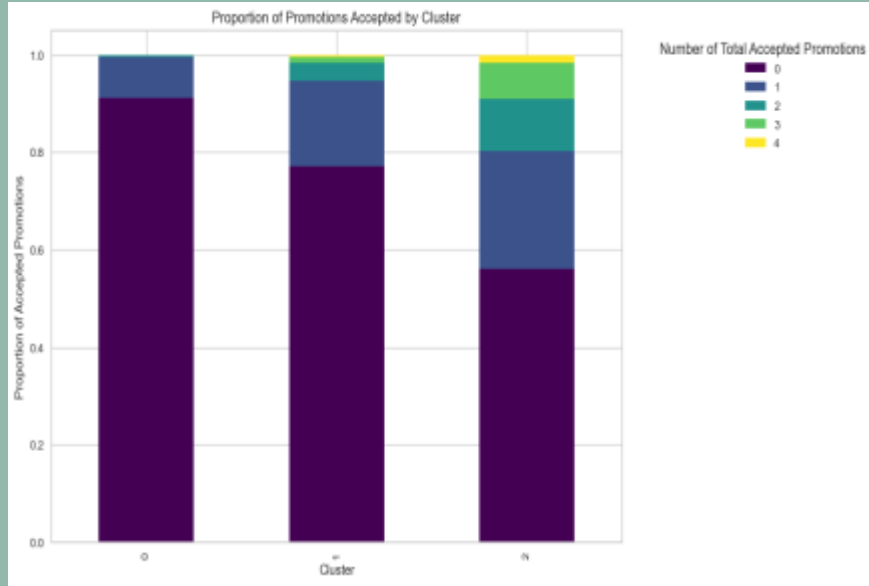
# Purchase Channel



- Cluster 1 and Cluster 2 is **active across all channel**.
- Web shopping: Cluster 1 > Cluster 2, Catalog shopping: Cluster 2 > Cluster 1.
- Cluster 0 makes **relatively few purchases across all channels**, which may indicate that they buy less frequently overall.



# Customer Activation



- The majority of customers tended not to take up any promotions, cluster2 participated in a relatively high number of promotions
- Cluster2 has the lowest website activity, the other two clusters are similar



04

# Customer Profile



0

## Budget-Conscious Family Shoppers

- Lower income, Lower spent
- Most around 40 years old
- Most are parents with 1-3 children
- Many Single parents
- Relatively higher Education

1

## Prudent Established Spenders

- Middle income, Middle spent
- Most 40-60 years old
- Definitely parents with 1 children
- Most are graduation or lower
- Lowest website visits

2

## Affluent Non- Parents

- Higher income, Higher spent
- Span all ages
- Definitely not parents
- Most are graduation or lower
- Relatively the most participation in promotions



# Common Characteristics

0

- Prefer store shopping
- Participate in few promotions
- Wines and meat account for the largest amount of consumption

1

2



# Marketing Strategy

## For Budget-Conscious Family

### Price Sensitive Offer Strategy

- **Price offers and discounts:** These customers have lower incomes and lower spending levels, so we can offer ongoing price discounts, especially on everyday essentials and children's products. For example, a "Family Savings Package" can be set up, which includes a combination of discounts on a variety of common household items.

### Education and Information Provision

- **Consumer Education:** Provide educational resources and workshops on budget management and household savings to help these customers manage their household budgets more effectively.
- **Health and Nutrition Coaching:** Provide free coaching and workshops on child nutrition and family health to increase consumer trust and reliance on the brand.
- **Building single-parent communities:** Build customer communities through social media or offline community events to promote mutual help and support among single-parent families and enhance brand loyalty.

### High-quality content marketing

- **Specialized content:** The customer base is relatively well-educated, so it is possible to provide high-quality, information-rich advertising content, such as in-depth articles, industry analysis reports, and special lectures on financial management, healthy living, technological innovations and other areas they may be interested in.
- **Brand endorsement:** industry experts or celebrities can be invited to write blogs or conduct interviews on a regular basis to provide insights and practical advice to improve brand endorsement.



# Marketing Strategy

## For Prudent Established Spenders

### Value Optimization

- **Low and mid-range product:** mainly launching mid-range products to ensure that the product price is not too high nor too cheap, focusing on value for money.

### Traditional media advertising

- **Utilizing traditional media:** Given the low online activity of this type of customer, advertisements can be placed through TV, radio and print media (e.g. newspapers, magazines), especially in local media, to increase brand visibility.
- **Direct mail marketing:** distributing coupons or promotional information to be delivered directly through the mail, this method is still effective for some older age groups.

### Simplified digital touchpoints

- **Simplify the website interface:** for occasional users, ensure that the interface is intuitive and easy to use, minimising complex navigation and excessive clicking steps.
- **Enhance the mobile experience:** Given that some users may access the site via mobile phones, optimising the access experience for mobile devices is also key, such as increasing page load speeds and improving the friendliness of the mobile interface to attract more pageviews.





# Marketing Strategy

## For Affluent Non-Parents

### High-end and luxury product

- **Promote high-end products and services:** Given their high spending power, promote high-end, luxury or customized services.
- **Limited edition products:** Introduce limited edition or exclusive products to increase the scarcity and uniqueness of the product and attract consumers interested in uniqueness and novelty.
- **Co-operation with other brands:** Co-operate with other high-end brands for cross-border co-branding activities to enhance brand endorsement.

### Enhanced promotions and loyalty programs

- **Innovative promotions:** As this group is more inclined to participate in promotions, innovative and attractive promotions can be designed, such as high discounts, doubling of rebate points, exclusive member days.
- **Premium membership system:** Establish a premium membership system to provide members with privileges such as exclusive discounts, priority purchase rights and VIP customer service.

### Focus on personal interests and lifestyle

- **Promote high-end personal consumer products:** Concentrate on high-end personal consumer products and services such as premium personal care products, beauty and spa services, etc.
- **Health and well-being:** Given that this group is likely to be more focused on personal health, health-related products and services can be promoted, such as organic food and health supplement.



# Thanks!

Does anyone have any questions?

E1162679@u.nus.edu

