Unlocking Customer Value: Grocery Customer Segmentation

Key Takeaways

1. Increase in **income** → decreases the number of **deals** purchased
2. Existing **marketing campaigns** are effective only on **high income** and high expenditure customer segment
3. **Customer Lifetime Value (CLV)** increases with Income
4. Increase in **complains** → decreases the amount of products **sold**
5. Increase in **income** → decreases number of **complains**

Project Objectives	Questions
Clustering the customers into different segments	How to reduce the dimensions? How many clusters are optimal?
Analyzing effect of **income** on existing **deals** and number of **complains**	How does **income level** impact the types of **deals** customers prefer? Does increase in income affect the number of **complaints** received?
Finding the **Customer Lifetime Value (CLV)** to find important customers	How much does a customer **spend** on average? What is the average **purchase frequency**? What is the average **customer lifetime**?
Analyzing the effect of existing **marketing campaigns** on different customer segments	How many customers responded to the campaigns? Which group of customers responded to which campaigns? Which campaigns are more effective on which group(s)?
Analyzing if increase in **complaints** decreases the number of products sold	How many complaints are received per how many customers? How many products are bought by complaining customers?

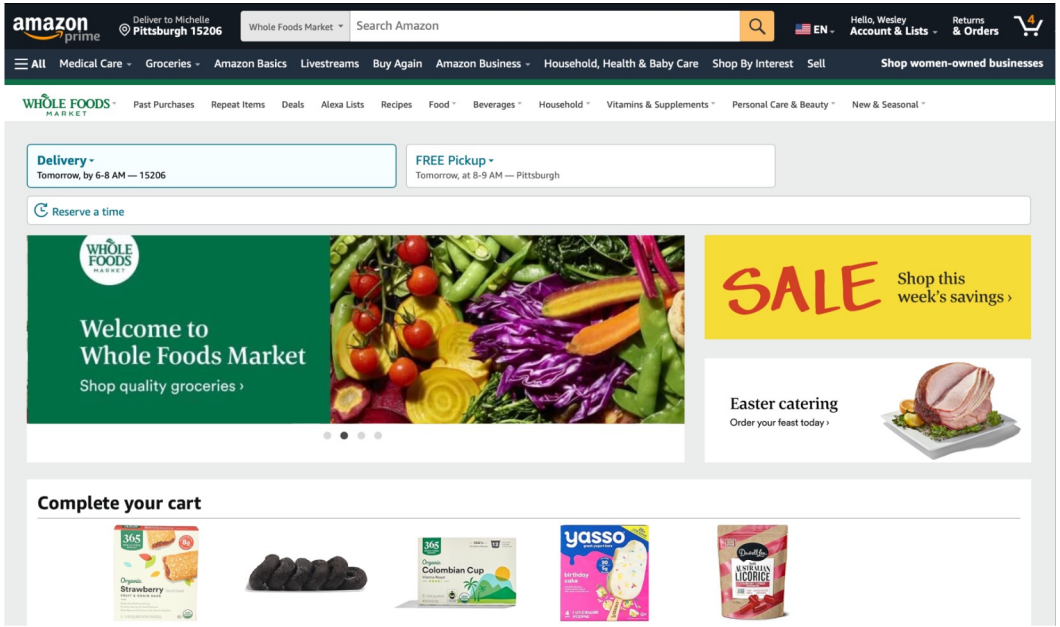
Dataset

[Kaggle dataset link]

2240 rows x 29 columns

Data includes:

- **Customer** information
- **Products** bought
- **Promotion** and Campaigns
- **Channels** of purchase



Tools

Altair, YellowBrick (ML), Matplotlib, Seaborn, Plotly (3D)

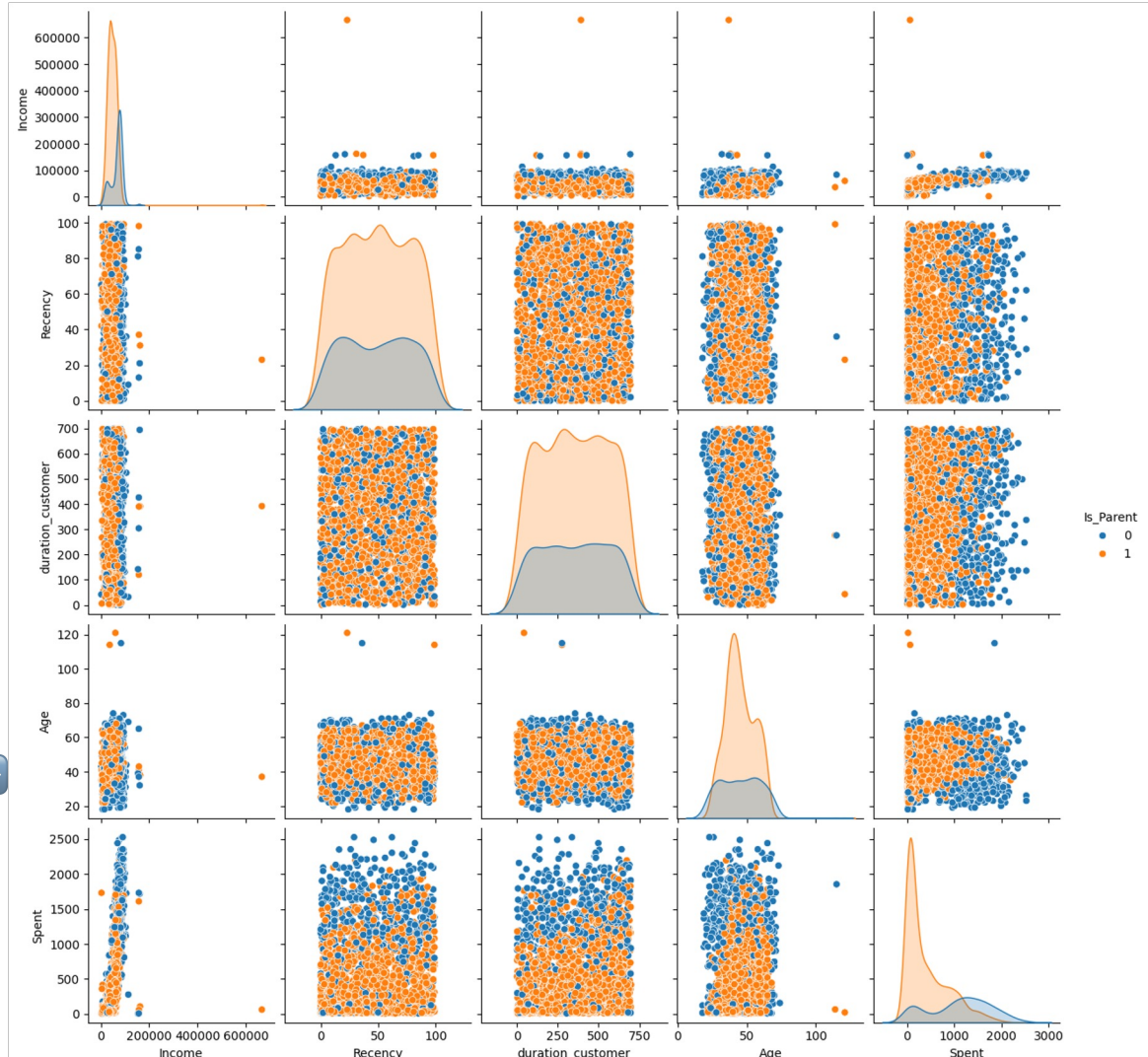
Data Cleaning

1. Drop **Missing** Data
2. **DateTime** Data
3. **Categorical** Data:

   Embedding

1. Dropping **Outliers:**

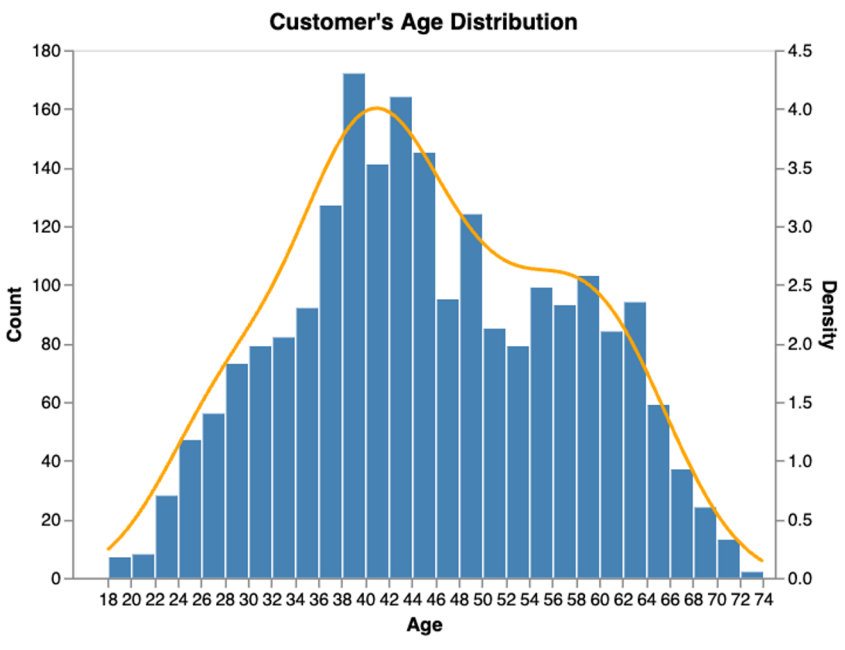
   With the pairplots➡️

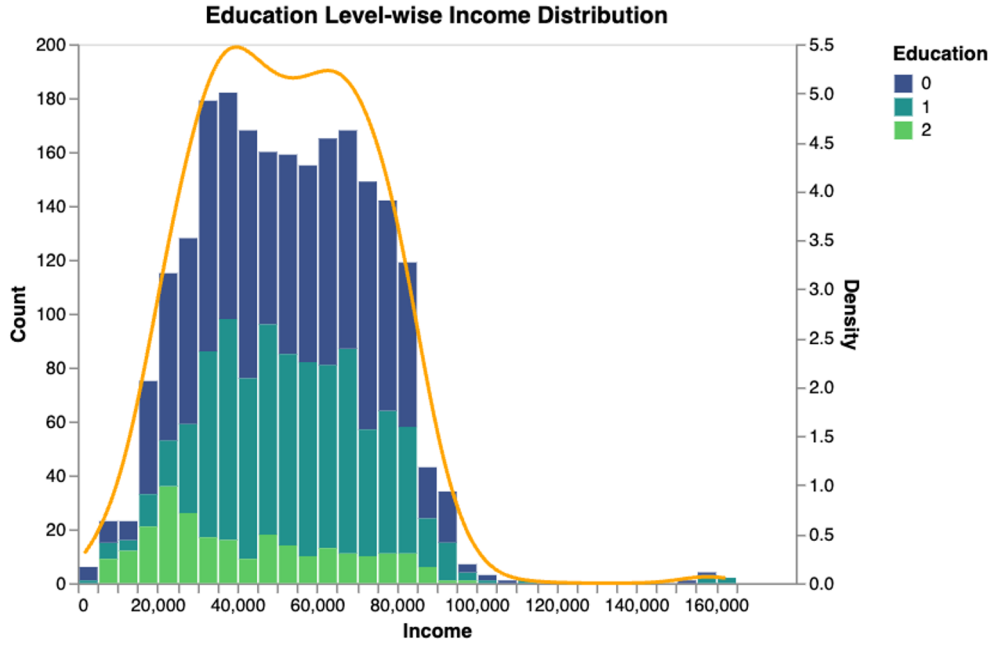


Exploratory Data Analysis

1. Distribution of **Age** - with Density Plot Line
2. **Education** Level-wise Income Distribution

Mostly Normally Distributed → Good for clustering

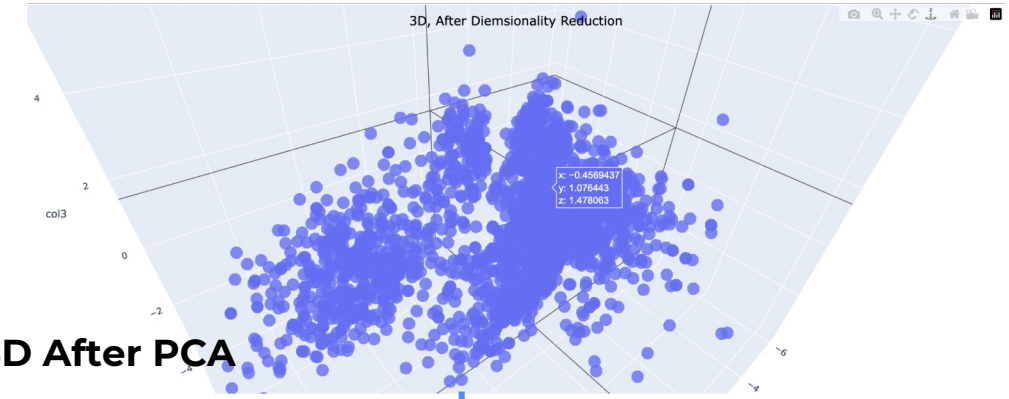




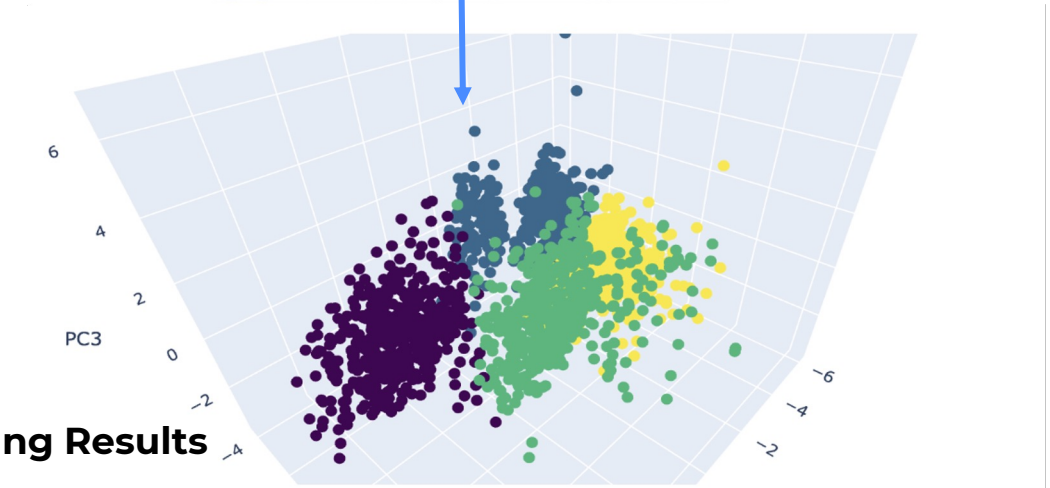
Method: PCA & Clustering

PCA:
Reduce number of features → 3 **P**rincipal **C**oordinates

Clustering:
Segmentation of customers → know the typical characteristics of each group



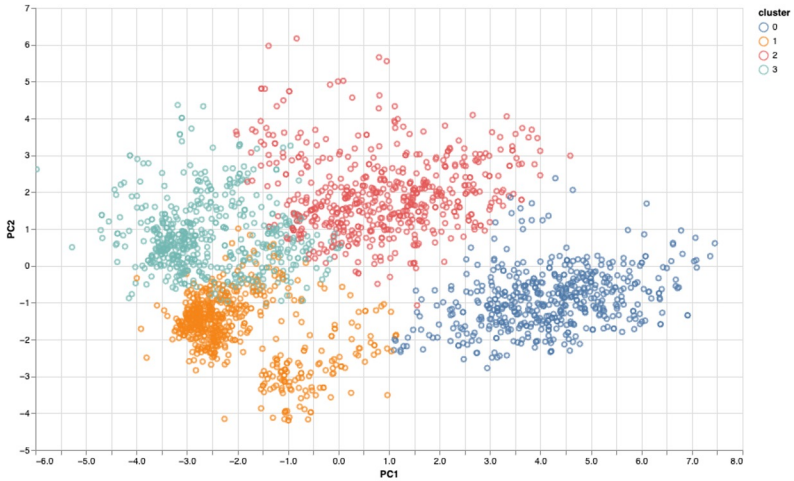
3D After PCA



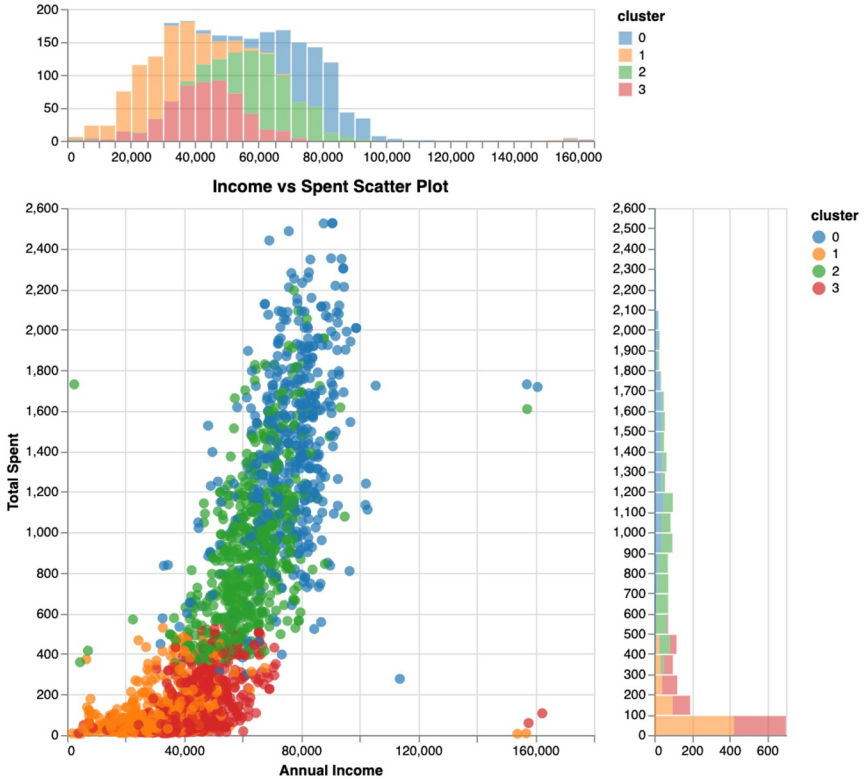
3D Clustering Results

Method: PCA & Clustering

Cluster 0 (blue): high income, high spent
Cluster 1 (orange): Low income, low spent
Cluster 2 (green): medium - high income, medium spent
Cluster 3 (red): medium to low income, low spent



2D Clustering Results



Income VS Spent, segmented by clusters

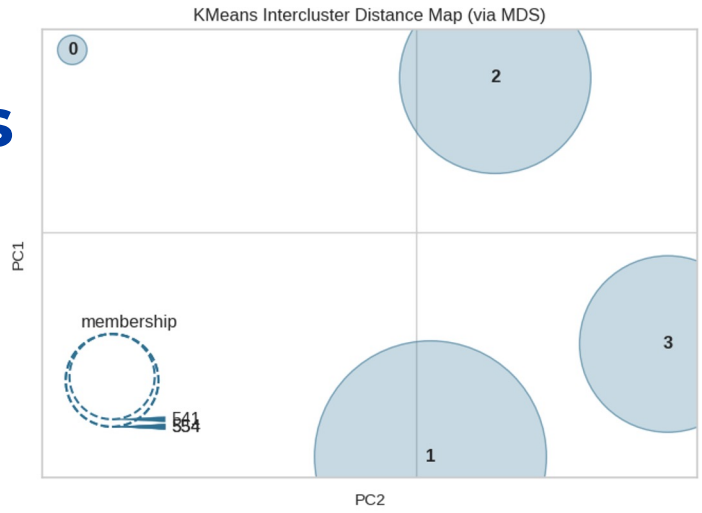
Evaluating Clustering Results

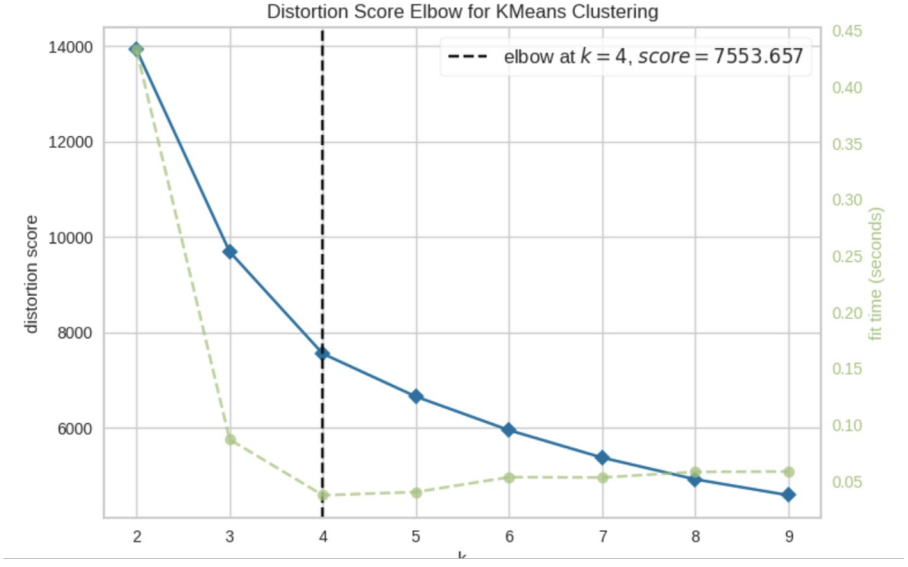
Tool: YellowBrick - ML visualization

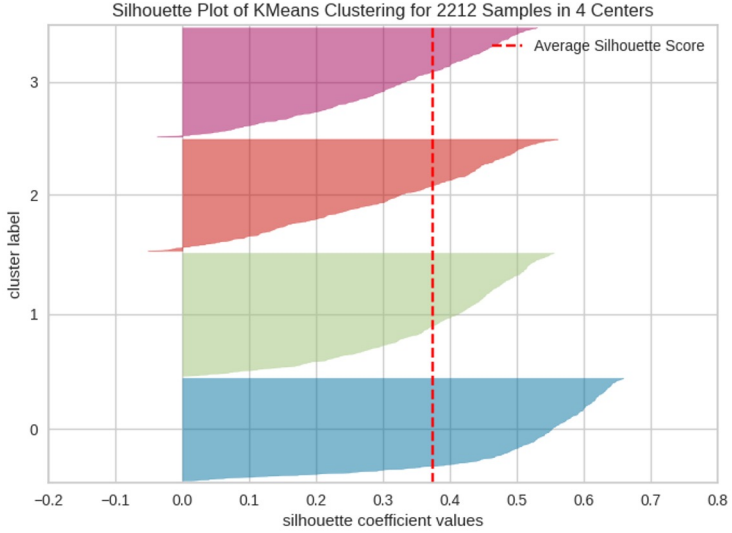
1. Distortion Score Elbow for KMeans
2. KMeans Intercluster Distance Map
3. Silhouette Plot

→ **Best k (number of clusters)**

→ **Clusters are well segmented**

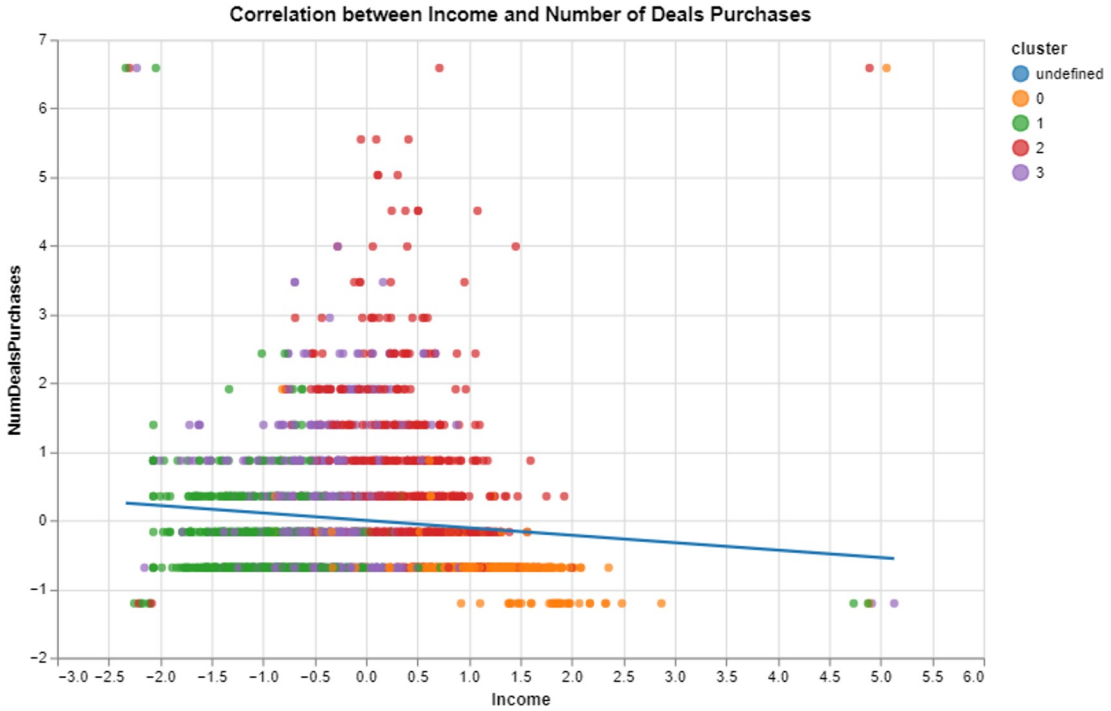






Hypothesis 1

Increase in income decreases the number of deals purchased



Cluster 0 (Orange): high income, high expenditure

Cluster 1 (green): Low income, low expenditure

Cluster 2 (red): medium income, medium expenditure

Cluster 3 (purple): medium income, low expenditure

Hypothesis 2

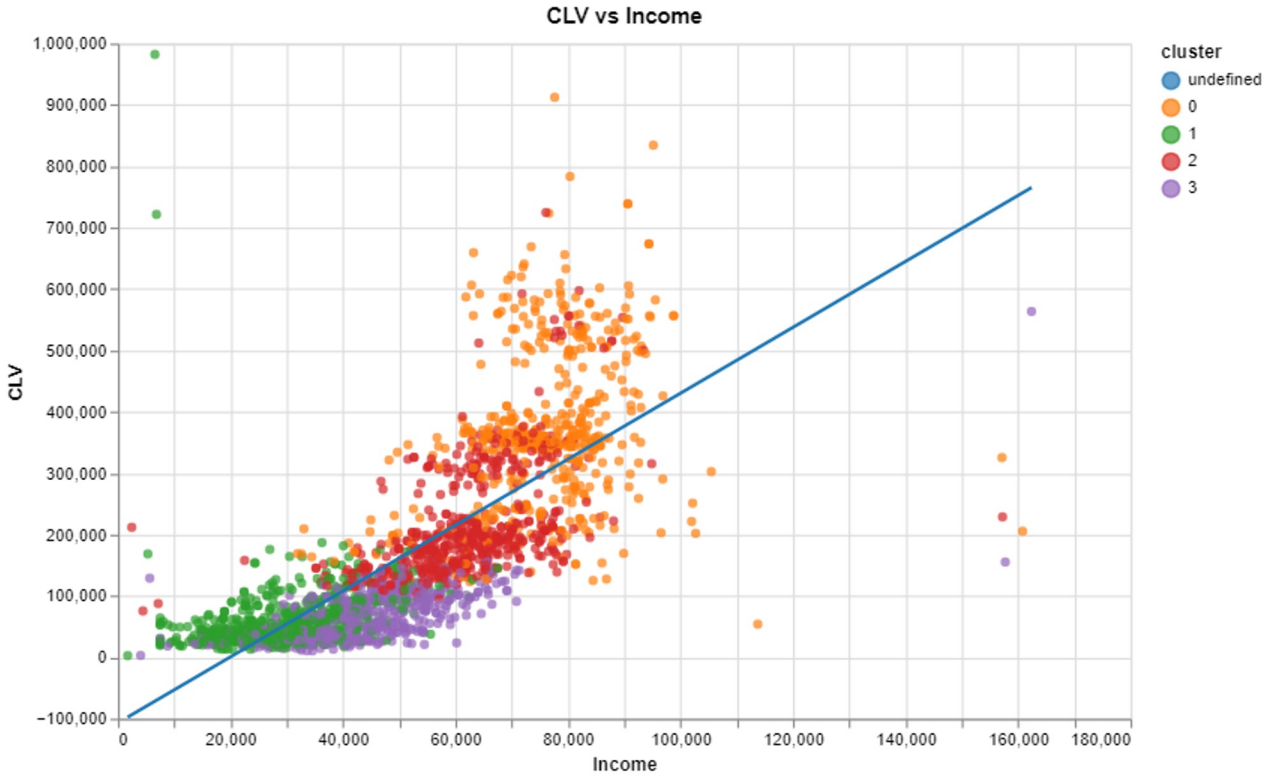
Customer Lifetime Value (CLV) increases with an increase in income.

Cluster 0 (Orange): high income, high expenditure

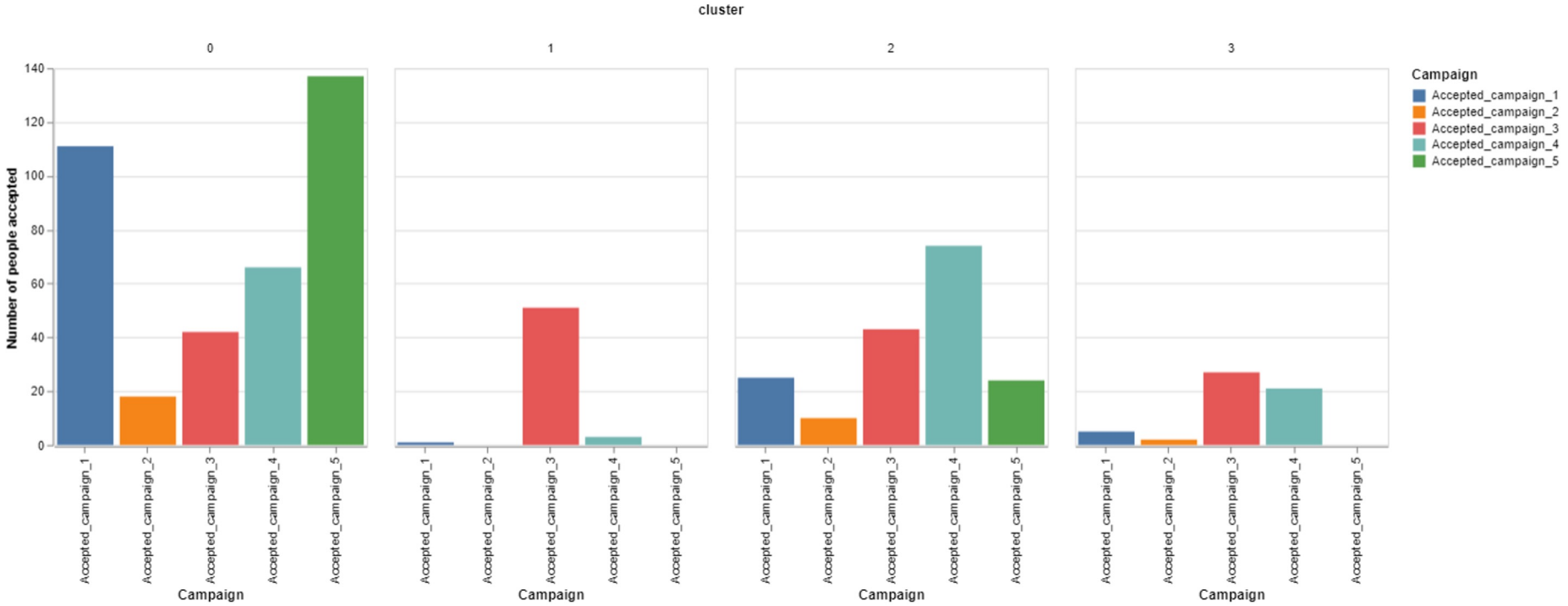
Cluster 1 (green): Low income, lov expenditure

Cluster 2 (red): medium income, medium expenditure

Cluster 3 (purple): medium income, low expenditure



Analysis of marketing campaigns on customer segments



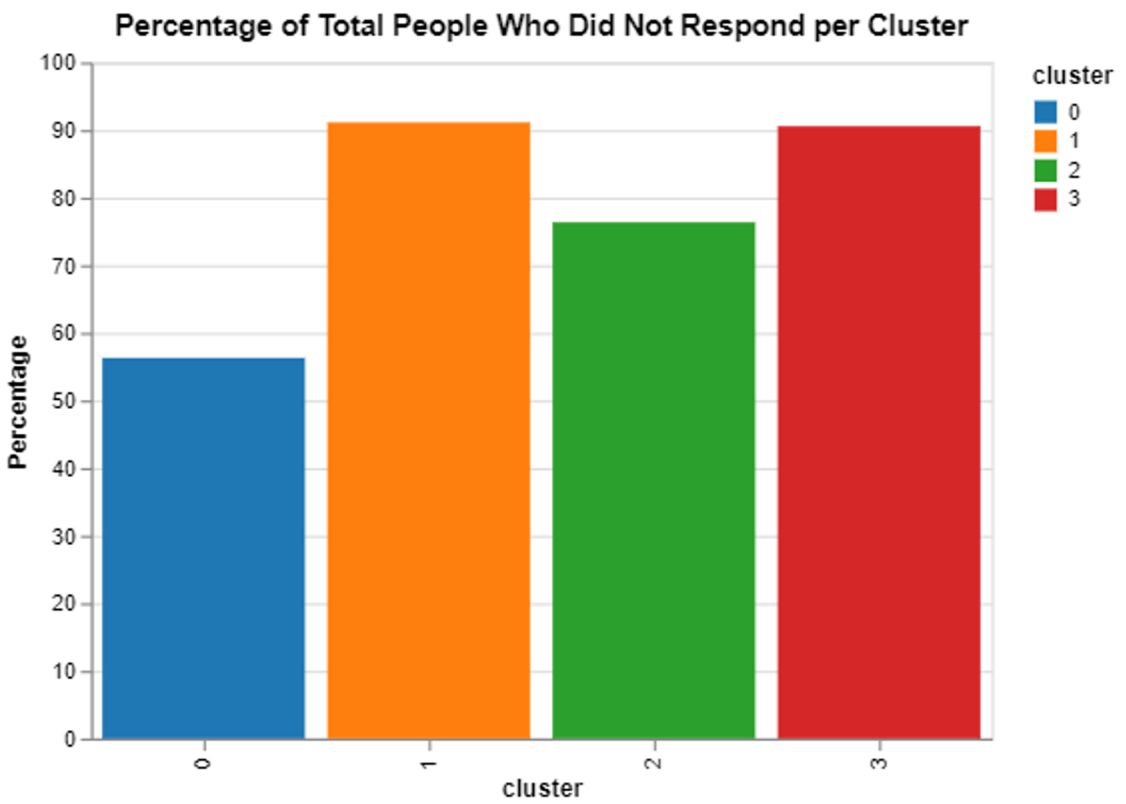
Analysis of marketing campaigns on customer segments

Cluster 0 (blue): high income, high expenditure

Cluster 1 (orange): Low income, low expenditure

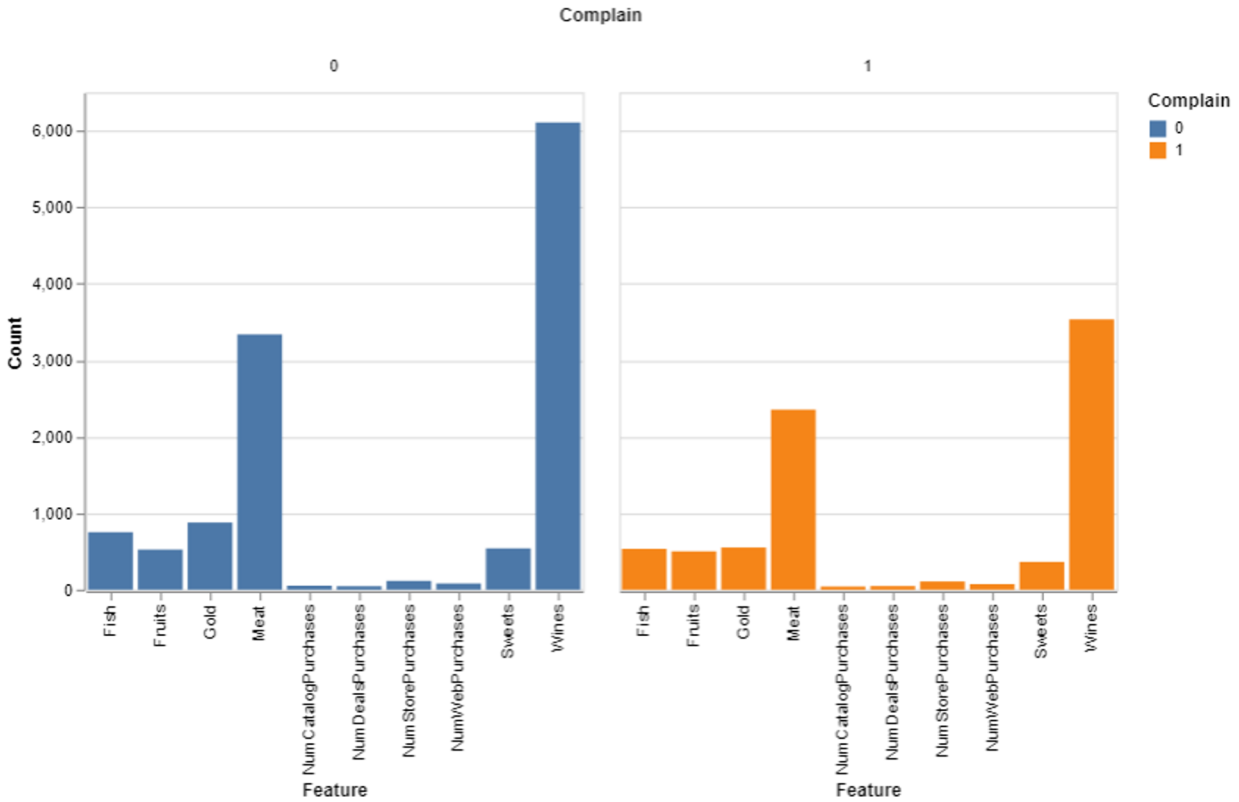
Cluster 2 (green): medium income, medium expenditure

Cluster 3 (red): medium income, low expenditure



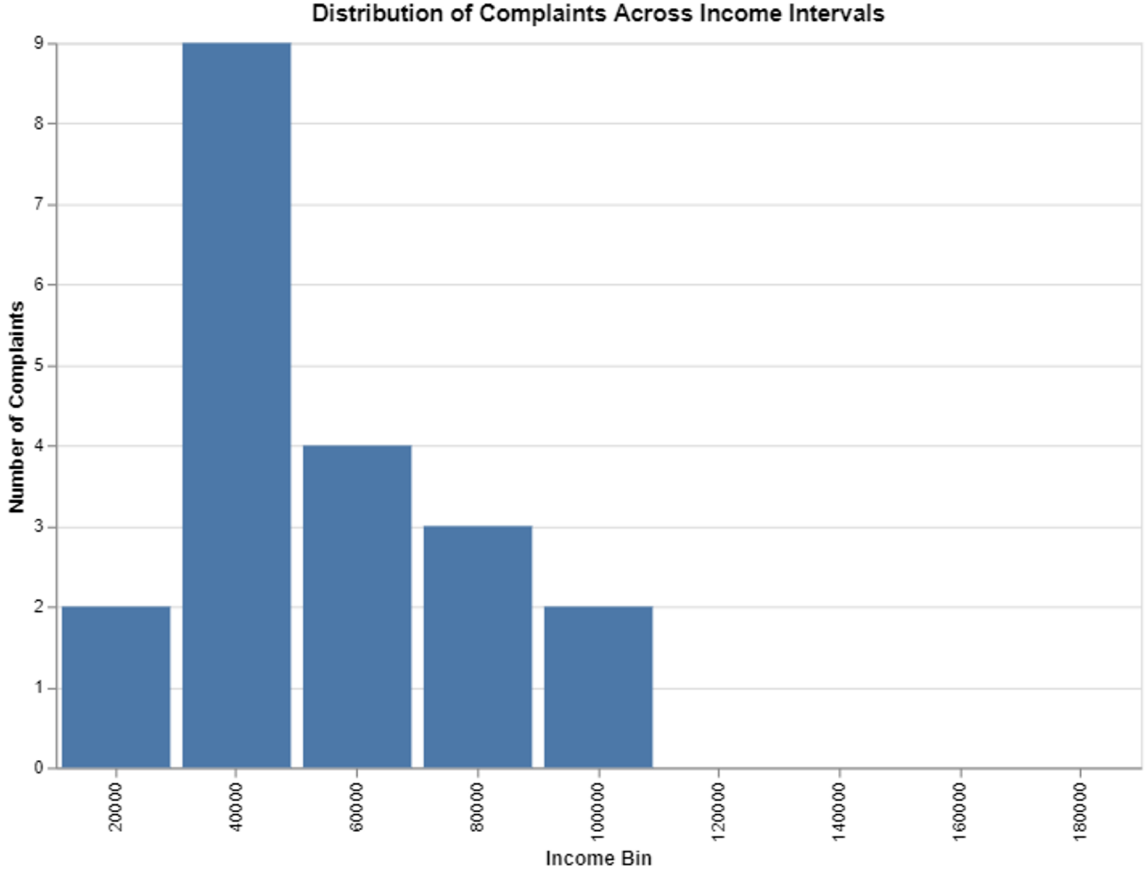
Hypothesis 3

Increase in **complaints** → decreases the number of **products sold**



Hypothesis 4

Number of **complaints** decreases for higher income groups



Future Work

Key Findings	Future Work
Existing **marketing campaigns** are effective only on **high income** and high expenditure customer segment	Design **marketing strategies** targeting low income and low-medium expenditure groups
Customer Lifetime Value (CLV) increases with Income	Optimize community engagement strategy to increase **customer loyalty**
Low **income** → Increase number of **complains** → Decrease **Sales**	Arrange specific agents to solve **after-sale** problems for low income customer group