

CUSTOMER PERSONALITY ANALYSIS

FOR

ENHANCED BUSINESS STRATEGIES

INSY 695

Jennifer Liu jenniferliu6
Kelly Liu CLIUKELLY
Ivy Lou ItIsvyLou

Yu Lu
Zhiming Zhang
Yvette Zhu

yuluhaley
Frederickzhimingzhang
Yvettez13

Agenda

01

**Challenge &
Hypothesis**

02

Data Landscape

03

**Exploratory data
analysis**

04

Clustering Model

05

**Causal Inference
Model**

06

**Conclusion & Future
Steps**

Enhance Business Strategies with Customer Segmentation



Need For Segmentation

- Diverse customer base
- Inefficient one-size-fits-all promotions
- Opportunity for targeted marketing



Hypothesis

- Segmentation improves promotion effectiveness
- Customized offers increase customer engagement
- Data-driven insights lead to higher sales

Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Prepare the Foundation with Data Preprocessing

- **Size:** 2240 rows x 29 columns
- **Attributes:** People, Products, Promotion, Place
- **Removal of outliers:** **1** age ≥ 100 years **2** income $\geq \$160K$
- **Assumption:** Current year set to 2016
- **Missing Values:** Iterative Imputation



- Age
- Kid & Teen Amount
- Income



- Education Level
- Marital Status



Imputed 1% of missing value from Income

Challenge & Hypothesis

Data Landscape

Exploratory data analysis

Clustering Model

Causal Inference Model

Conclusion & Future Steps

Uncover Hidden Patterns Through Exploratory Data Analysis

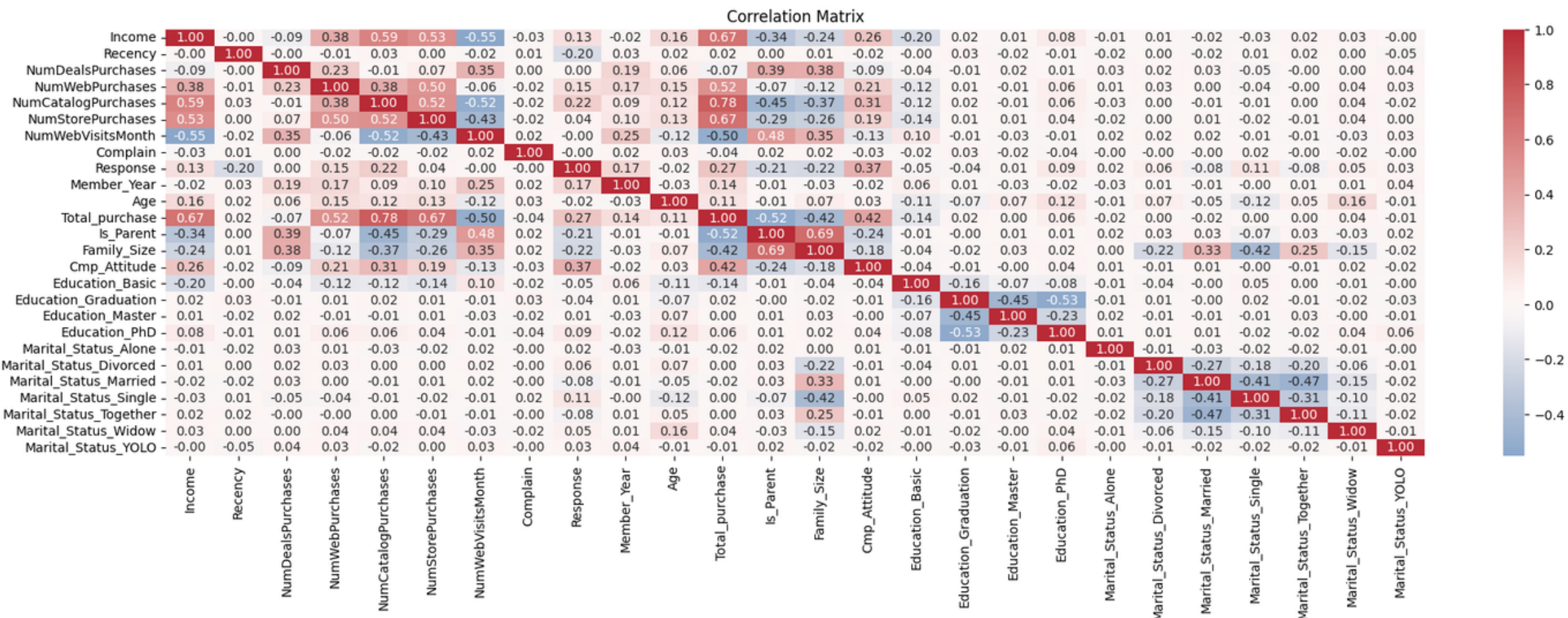
Feature Engineering

No High Correlation

Creation of

- Total Purchase
- Total Children
- Is Parent
- Family Size
- Campaign Attitude

To capture holistic customer profile



Challenge & Hypothesis

Data Landscape

Exploratory data analysis

Clustering Model

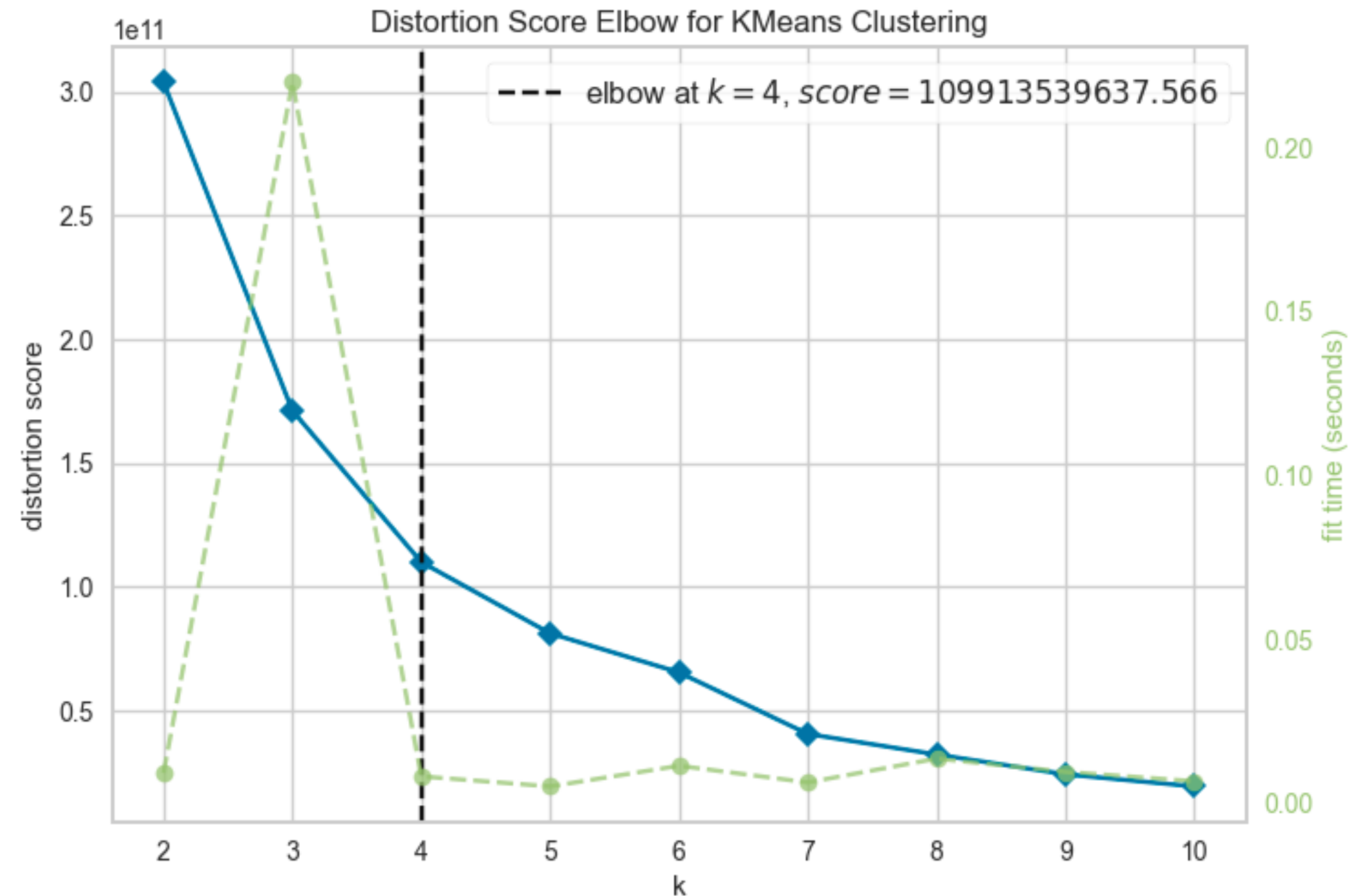
Causal Inference Model

Conclusion & Future Steps

Segmenting the Market: Creating Clustering Model

**Determined cluster number
with Kmeans for
consistency across
methods to be tested**

n= 4



Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

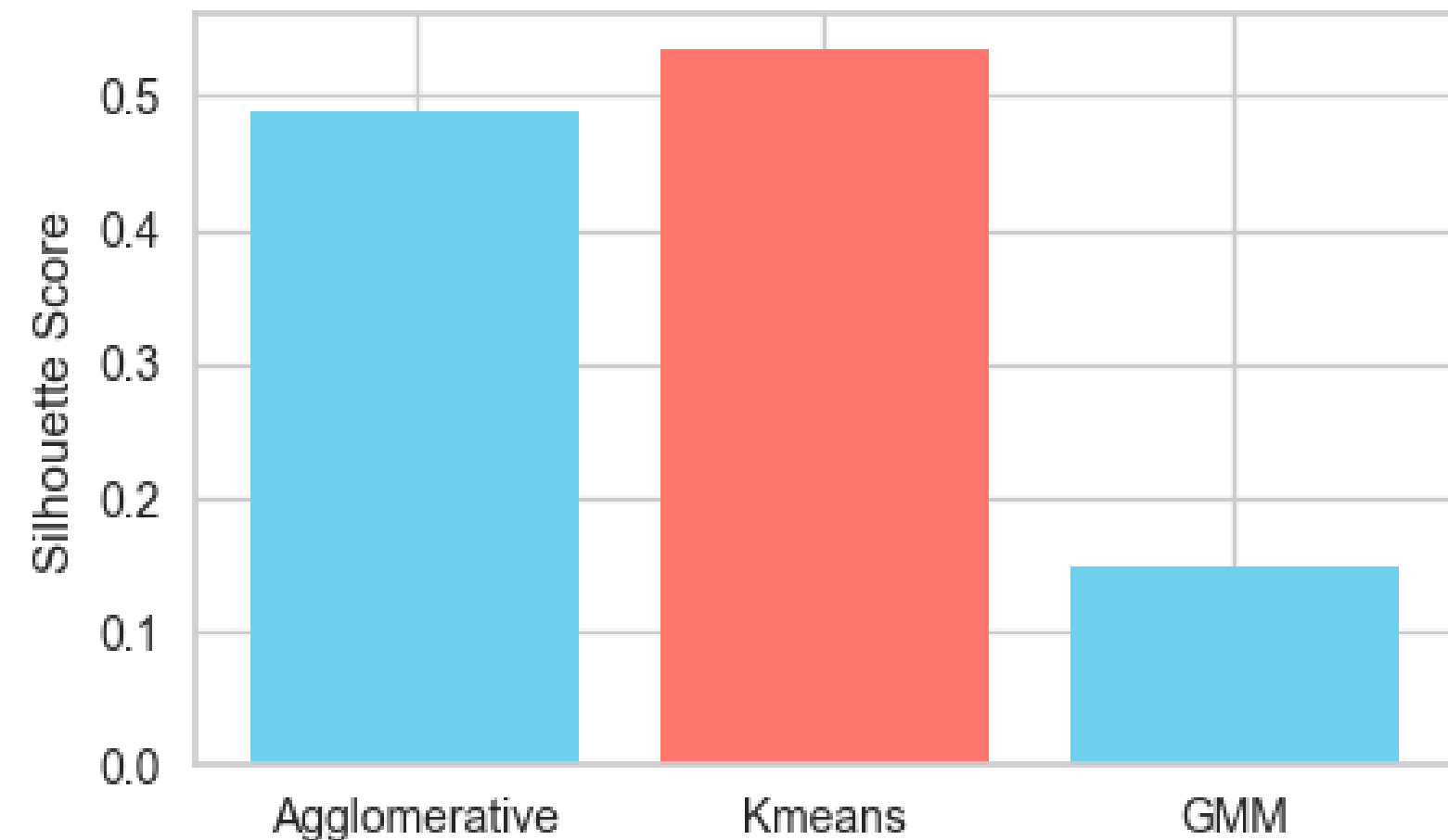
Causal Inference
Model

Conclusion & Future
Steps

Segmenting the Market: Model Selection

- **Tested with Kmeans, Agglomerative, GMM, DBSCAN**
 - Excluding DBSCAN as it yields only a single cluster
- **Chose to use Kmeans for its highest Silhouette Score**

Comparison of Silhouette Scores for Different Clustering Methods



Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

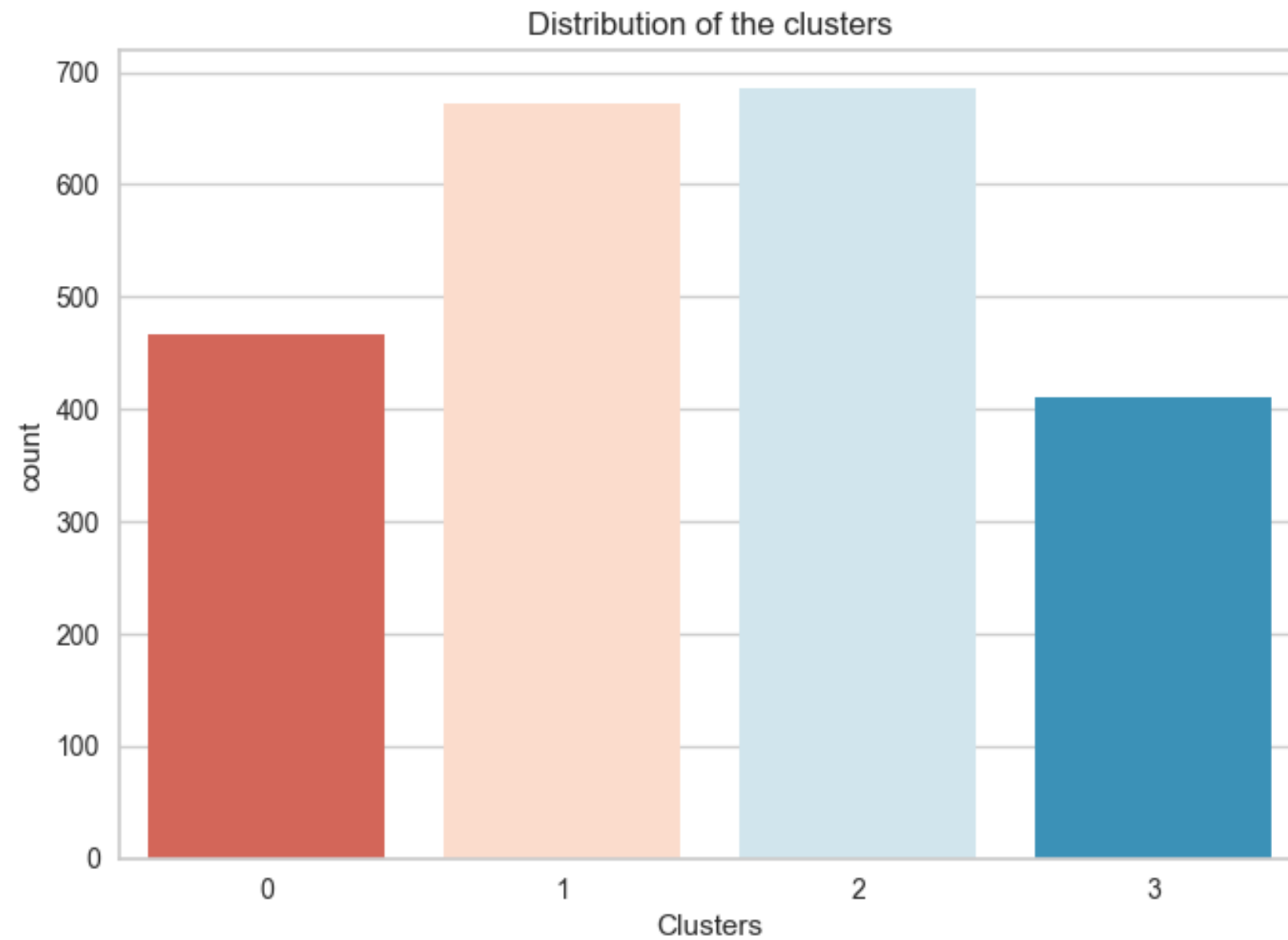
Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Segmenting the Market: Creating Clustering Model

Cluster Distribution



Total Purchases by Income



Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Segmenting the Market: Insights from Clustering Analysis

Segment 0: Affluent Enthusiasts (20.9%)

High income
High total purchase
Small family size
Low likelihood of being parents
Middle-aged

Segment 2: Budget-Conscious Parents (30.7%)

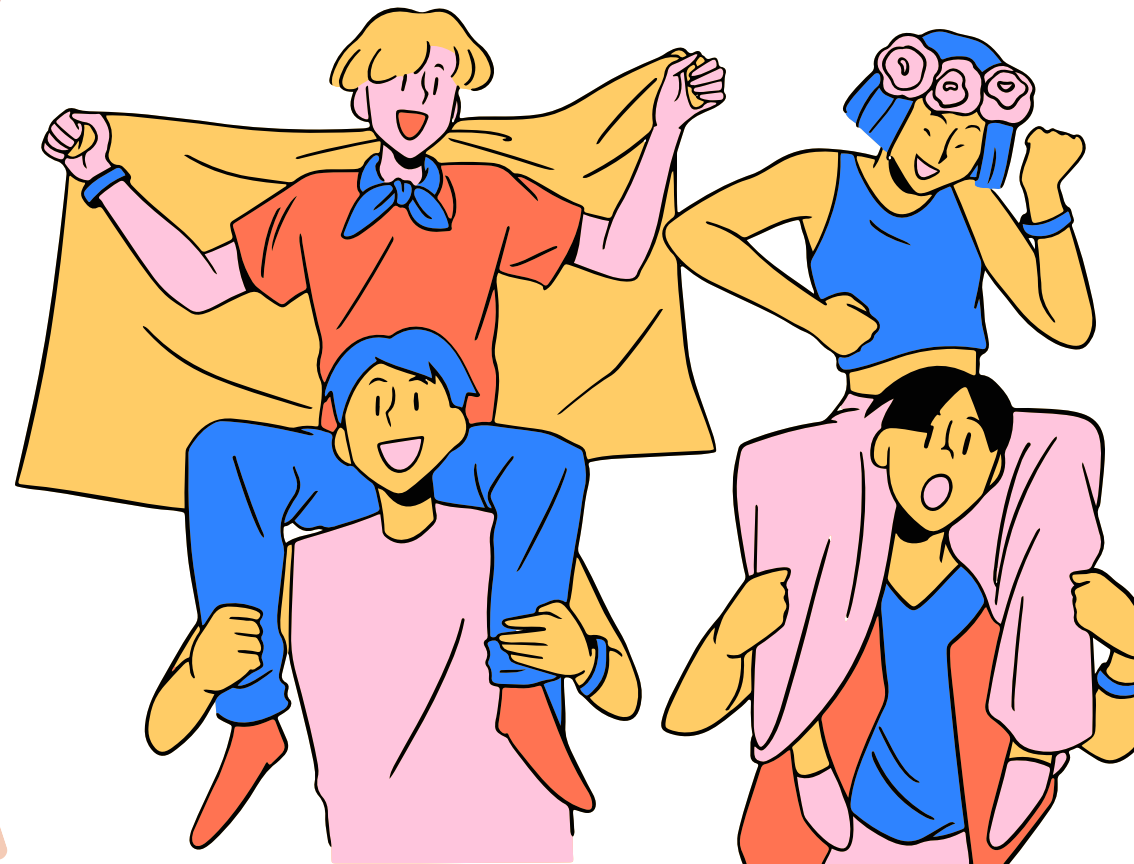
Low income
Very low total purchase
Moderate family size
High likelihood of being parents
Younger age

Segment 1: Established Families (30.1%)

Moderate income
Moderate total purchase
Larger family size
High likelihood of being parents
Middle-aged

Segment 3: Prosperous Mature (18.4%)

Higher income
High purchasing power
Older age
Moderate family size



Challenge &
Hypothesis

Data Landscape

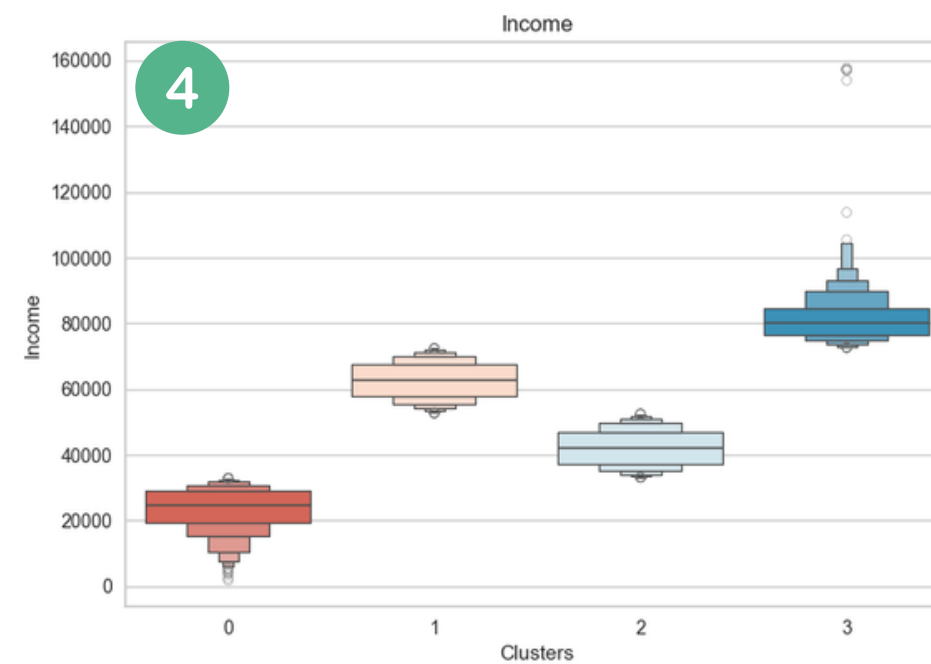
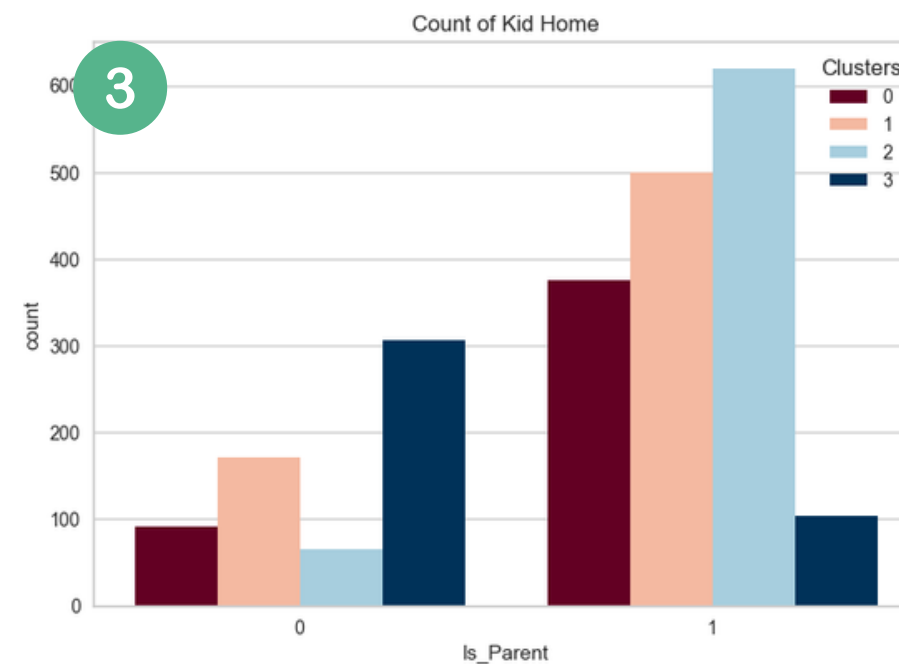
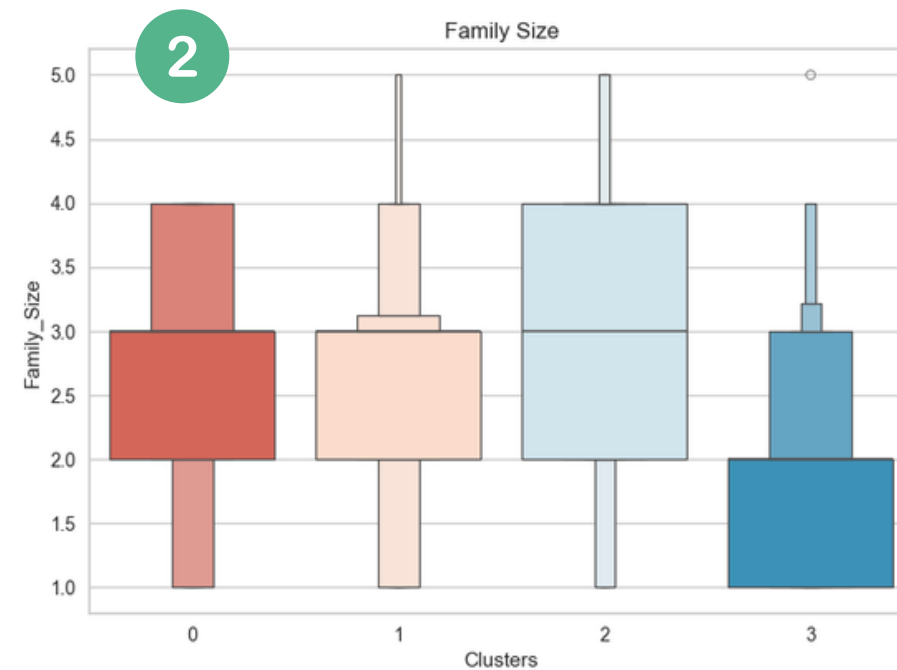
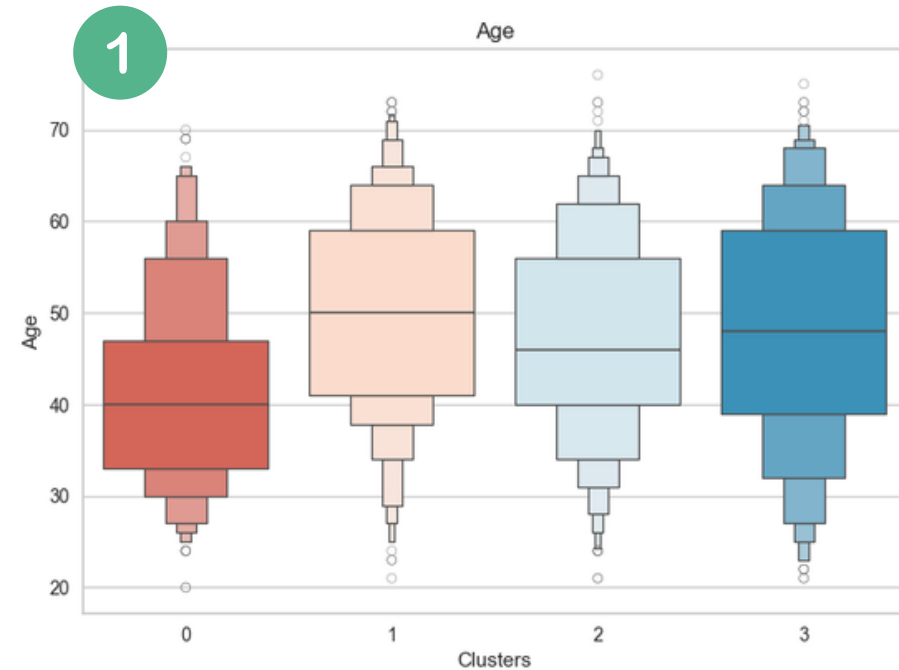
Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Segmenting the Market: Insights from Clustering Analysis



Demographic Insights

1. Affluent Enthusiasts (S0) are younger
2. Prosperous Mature (S3) has the smallest family size
3. Budget-Conscious Parents (S2) have the most children
4. Prosperous Mature (S3) earn the highest income

Affluent Enthusiasts (S0)

Budget-Conscious Parents (S2)

Established Families (S1)

Prosperous Mature (S3)

Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

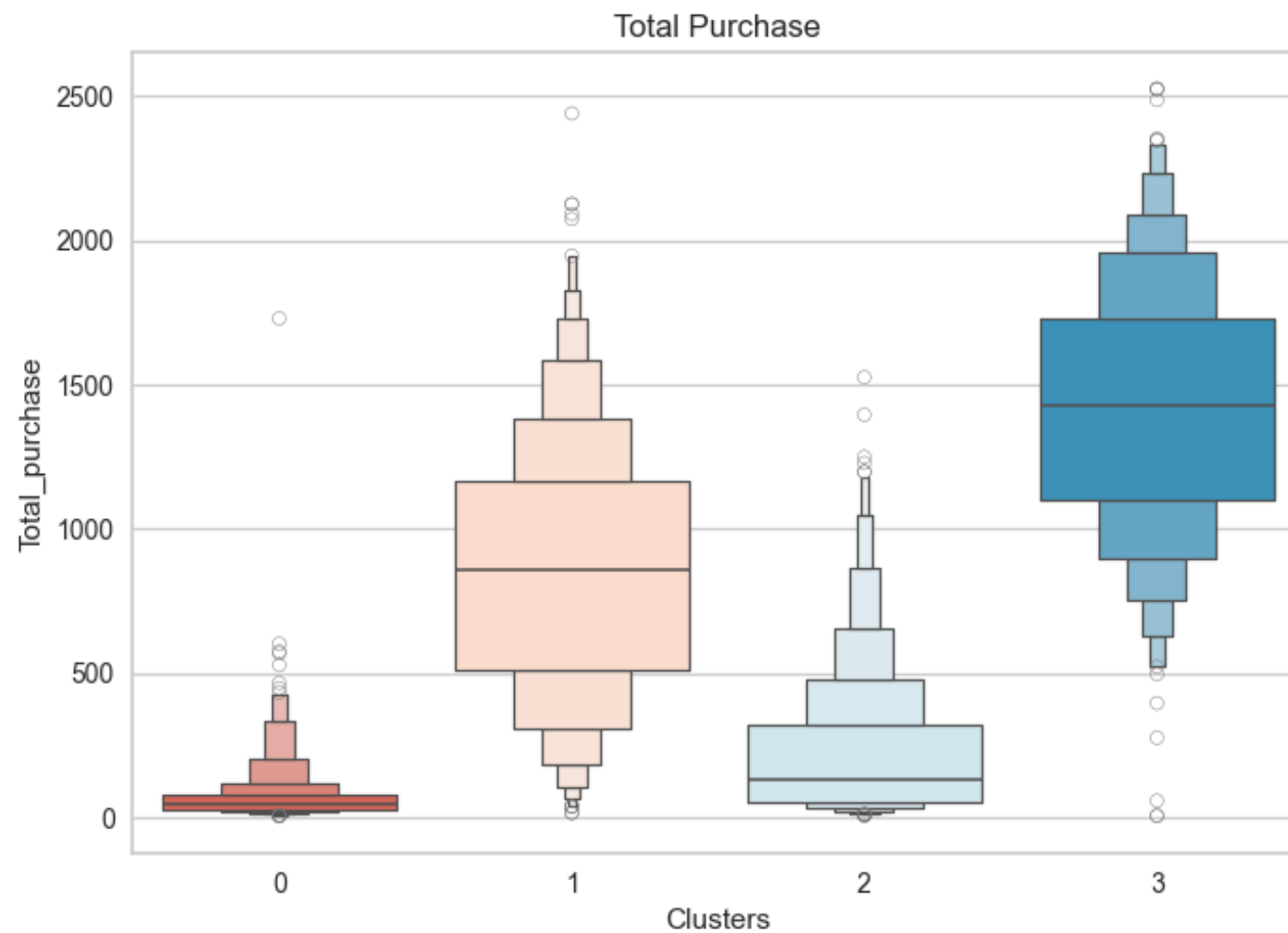
Clustering Model

Causal Inference
Model

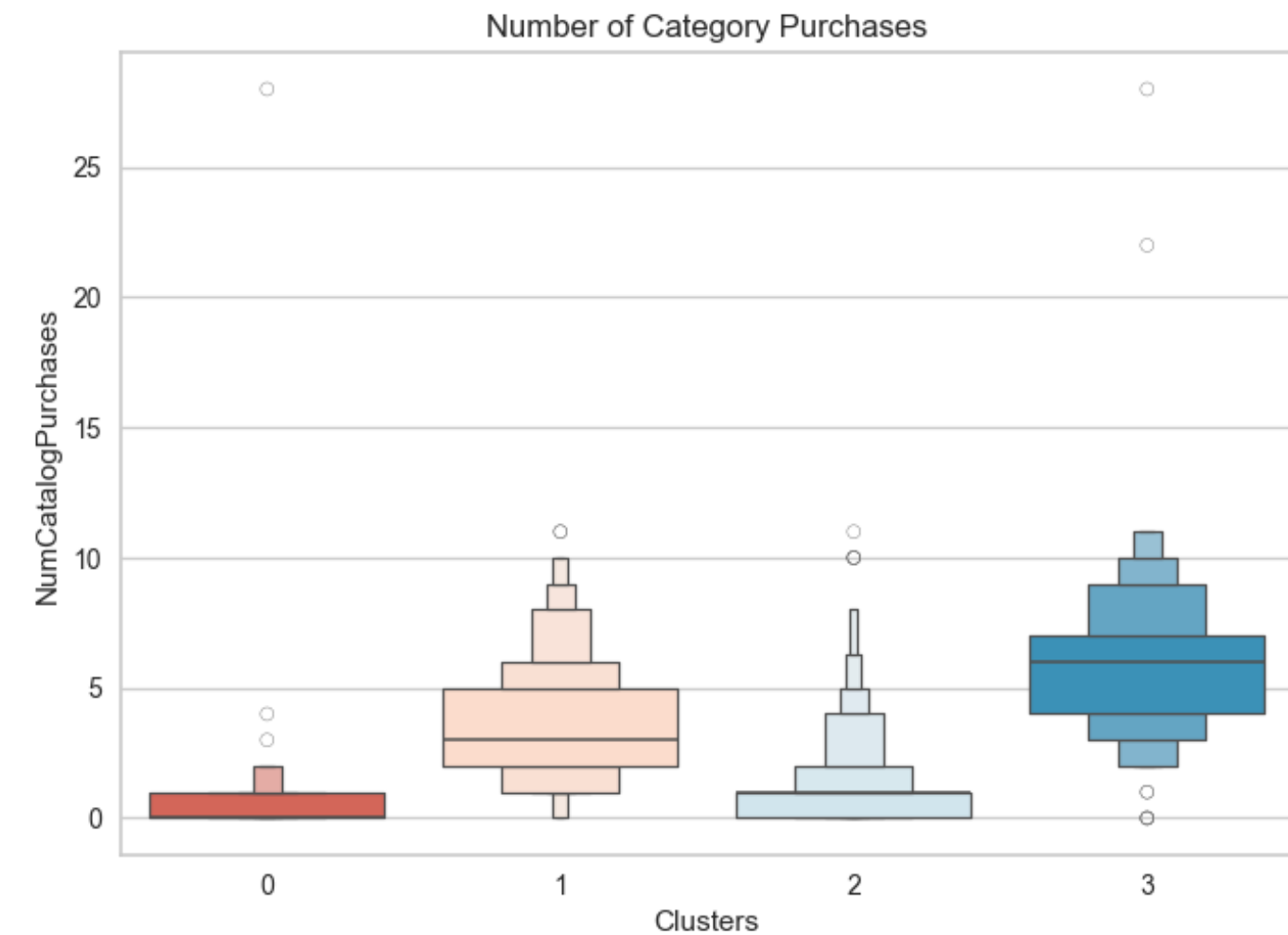
Conclusion & Future
Steps

Segmenting the Market: Insights from Clustering Analysis

Behavioral Insights - Purchase



- Prosperous Mature (S3) and Established Families (S1) have higher purchasing power



- Prosperous Mature (S3) purchased the most categories of products
- Followed by Established Families (S1)

Affluent Enthusiasts (S0)

Budget-Conscious Parents (S2)

Established Families (S1)

Prosperous Mature (S3)

Challenge &
Hypothesis

Data Landscape

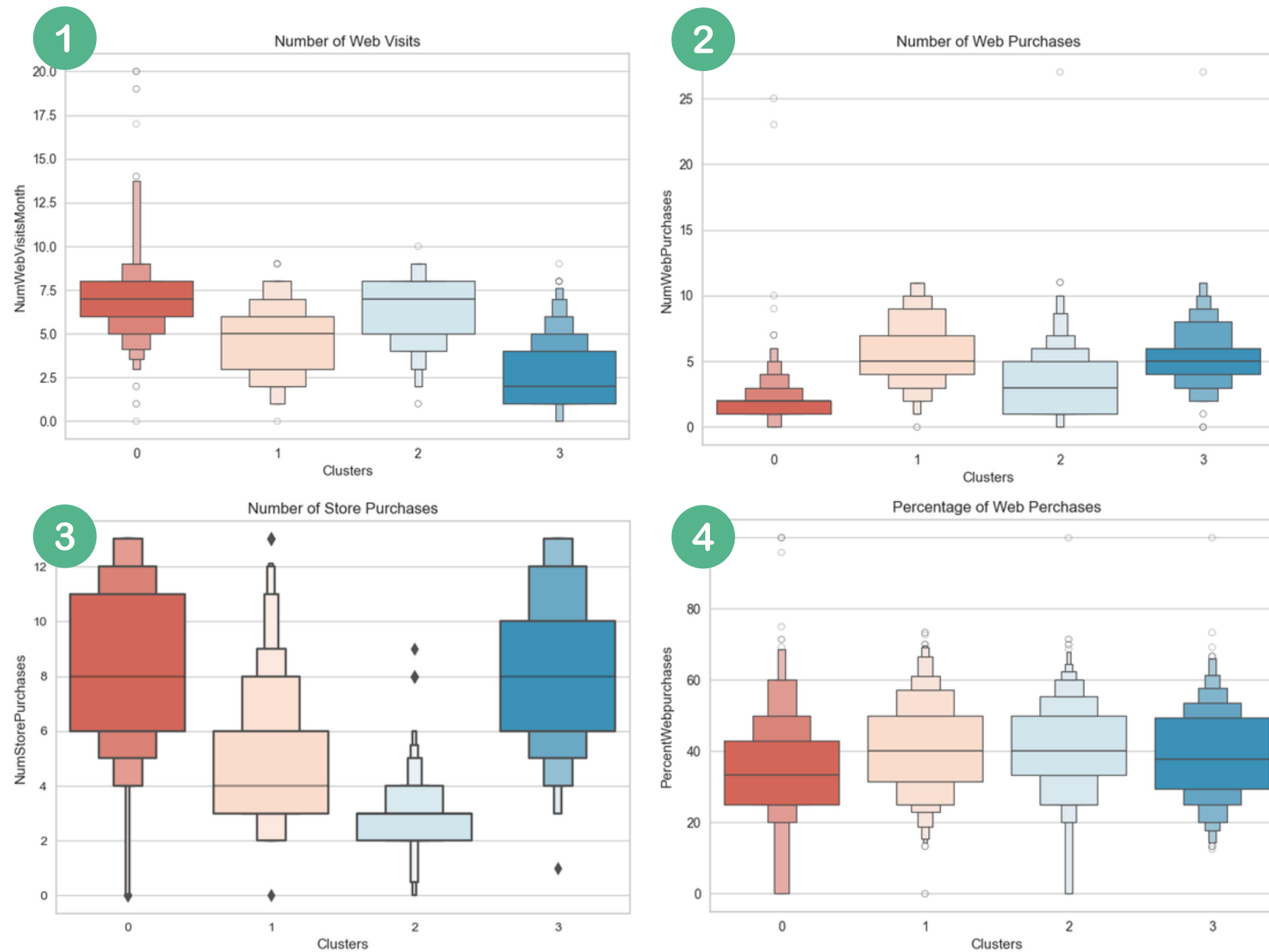
Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Segmenting the Market: Insights from Clustering Analysis



Behavioral Insights - Channel

1. Affluent Enthusiasts (S0) and Budget-Conscious Parents (S2) have the most website visits but the least to website purchase
2. Established Families (S1) and Prosperous Mature (S3) lead in the number of web purchases
3. Affluent Enthusiasts (S0) and Prosperous Mature (S3) made more store purchases
4. % of online shoppers are similar among segments, S1, S2, S3 slightly prefers online shopping than S0

Affluent Enthusiasts (S0)

Budget-Conscious Parents (S2)

Established Families (S1)

Prosperous Mature (S3)

Challenge & Hypothesis

Data Landscape

Exploratory data analysis

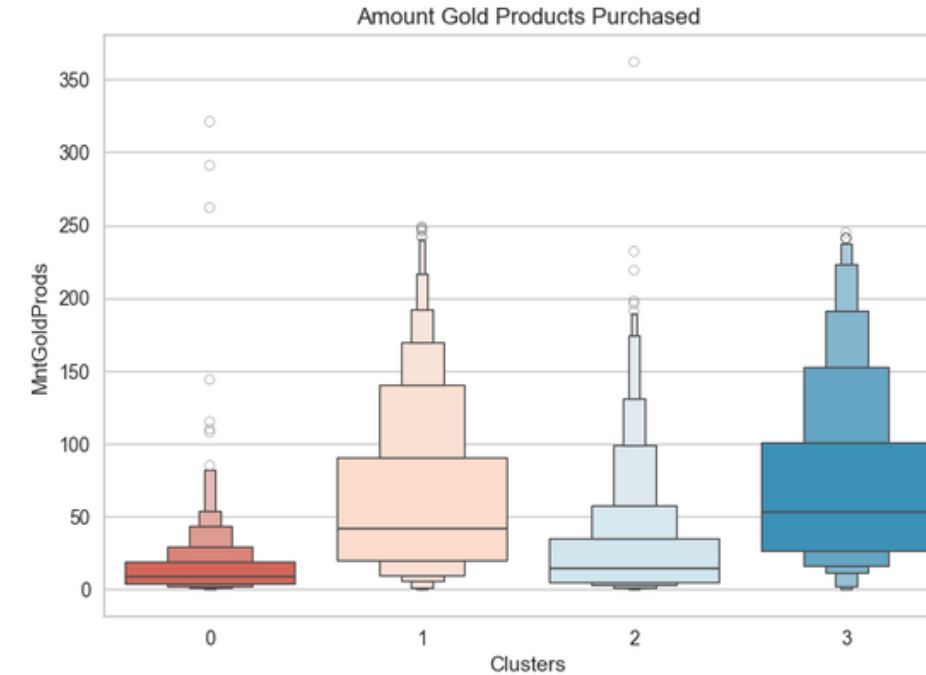
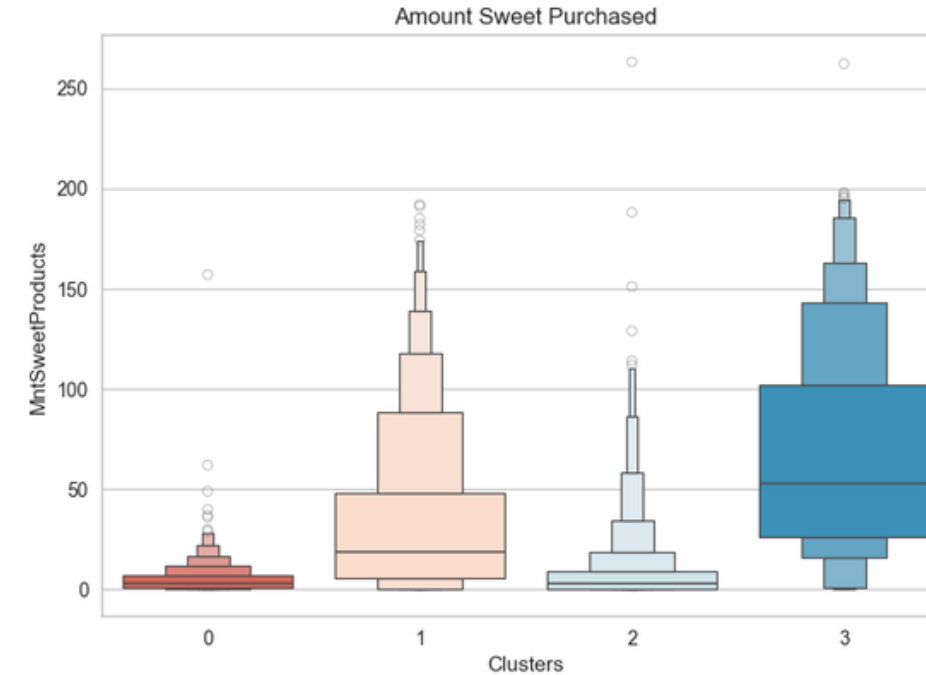
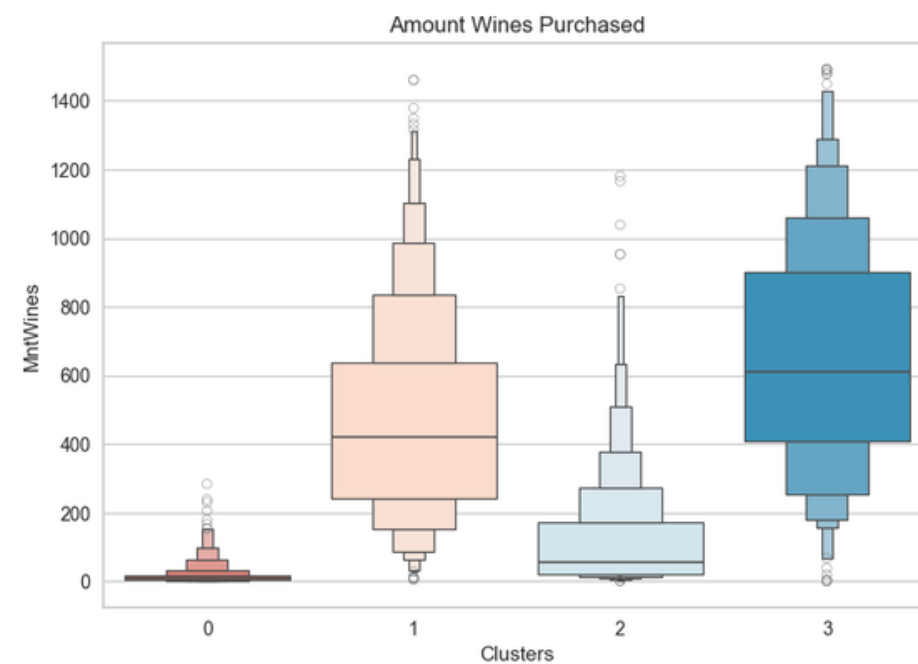
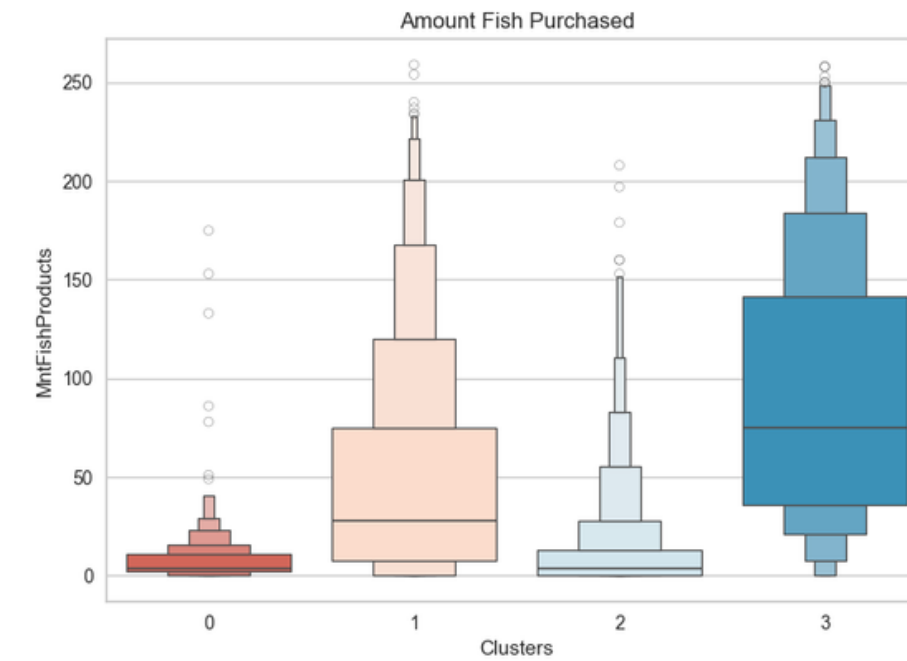
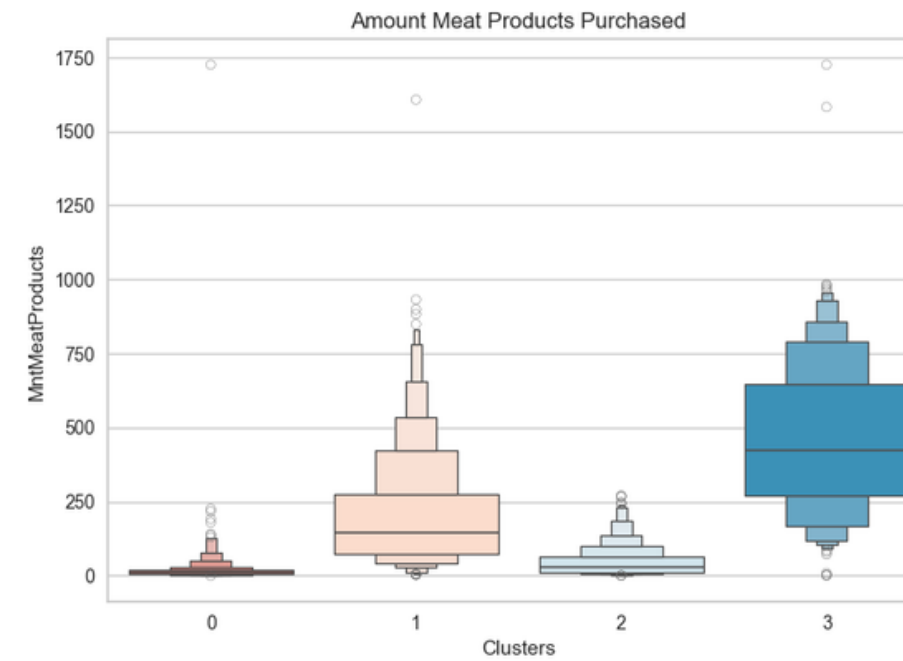
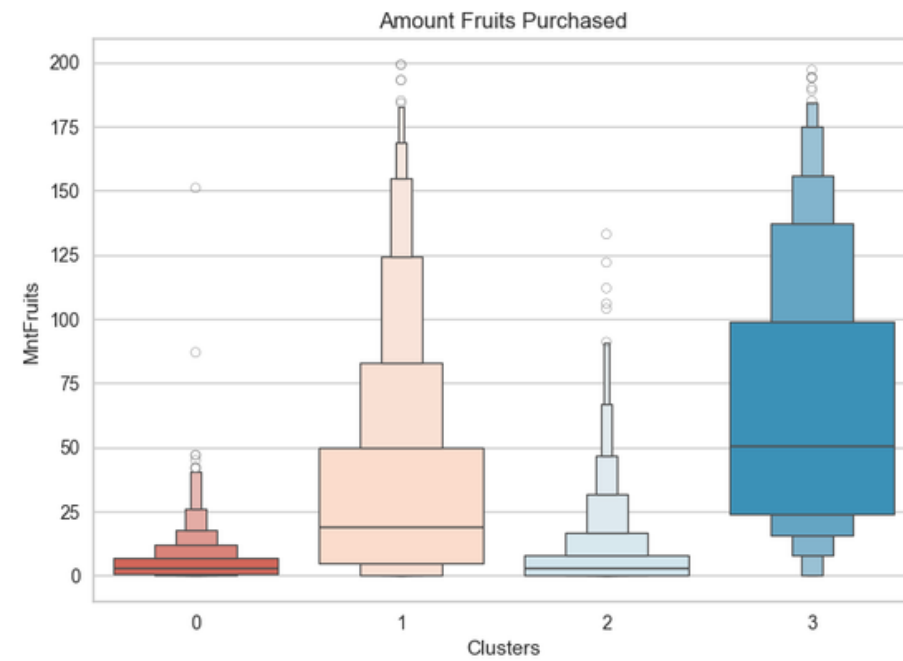
Clustering Model

Causal Inference Model

Conclusion & Future Steps

Segmenting the Market: Insights from Clustering Analysis

Behavioral Insights - Product Type



All segments love Gold Products, even for the lower income segments

Challenge & Hypothesis

Data Landscape

Exploratory data analysis

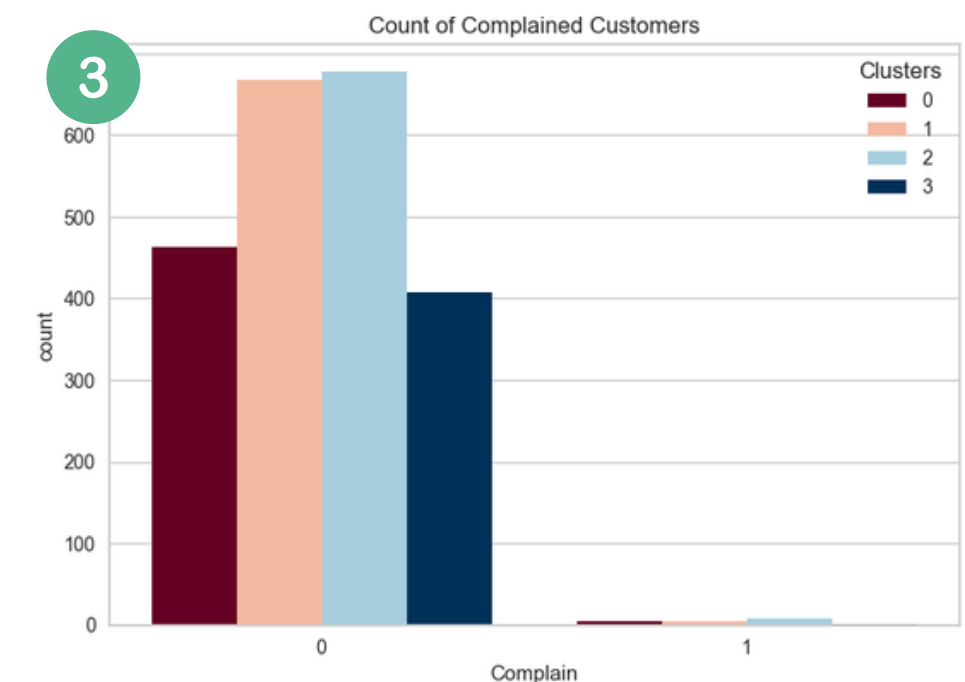
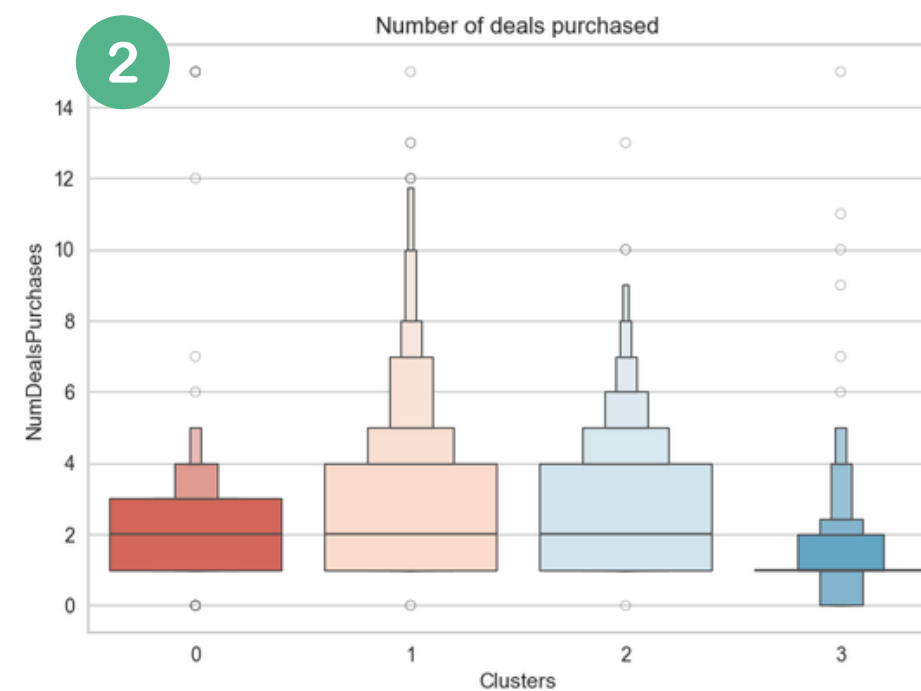
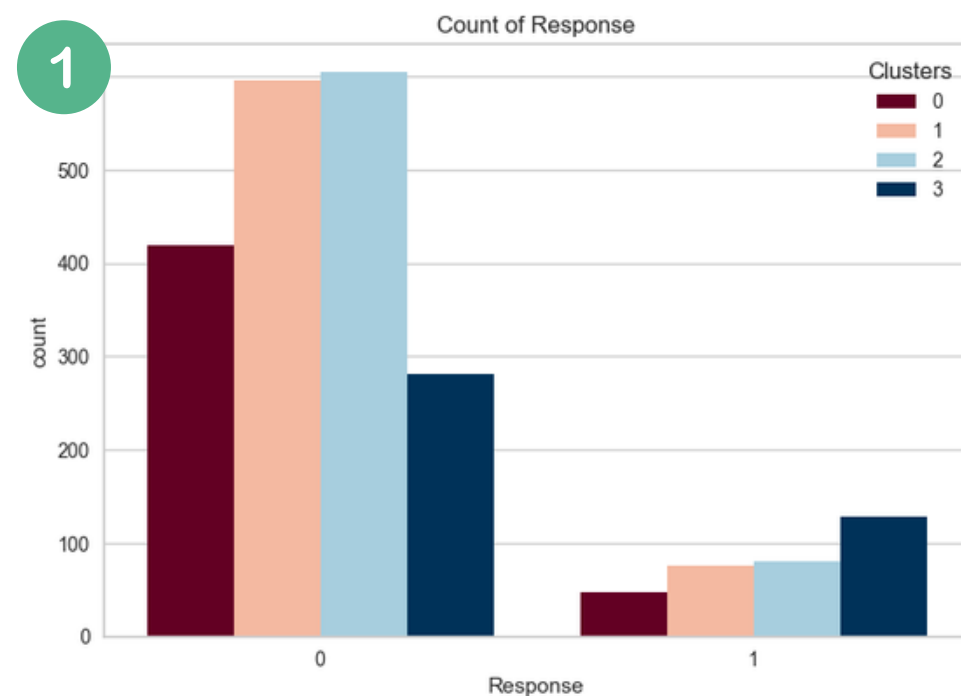
Clustering Model

Causal Inference Model

Conclusion & Future Steps

Segmenting the Market: Insights from Clustering Analysis

Behavioral Insights - Promotion



1. Only a small portion of people accept promotion offers, S3 is the most responsive group
2. Budget-Conscious Parents (S2) and Established Families (S1) are more price sensitive and like to purchase deals
3. Overall 1% complaint rate

Affluent Enthusiasts (S0)

Budget-Conscious Parents (S2)

Established Families (S1)

Prosperous Mature (S3)

Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Quantifying Impact: Causal Inference on Customer Purchasing Decisions

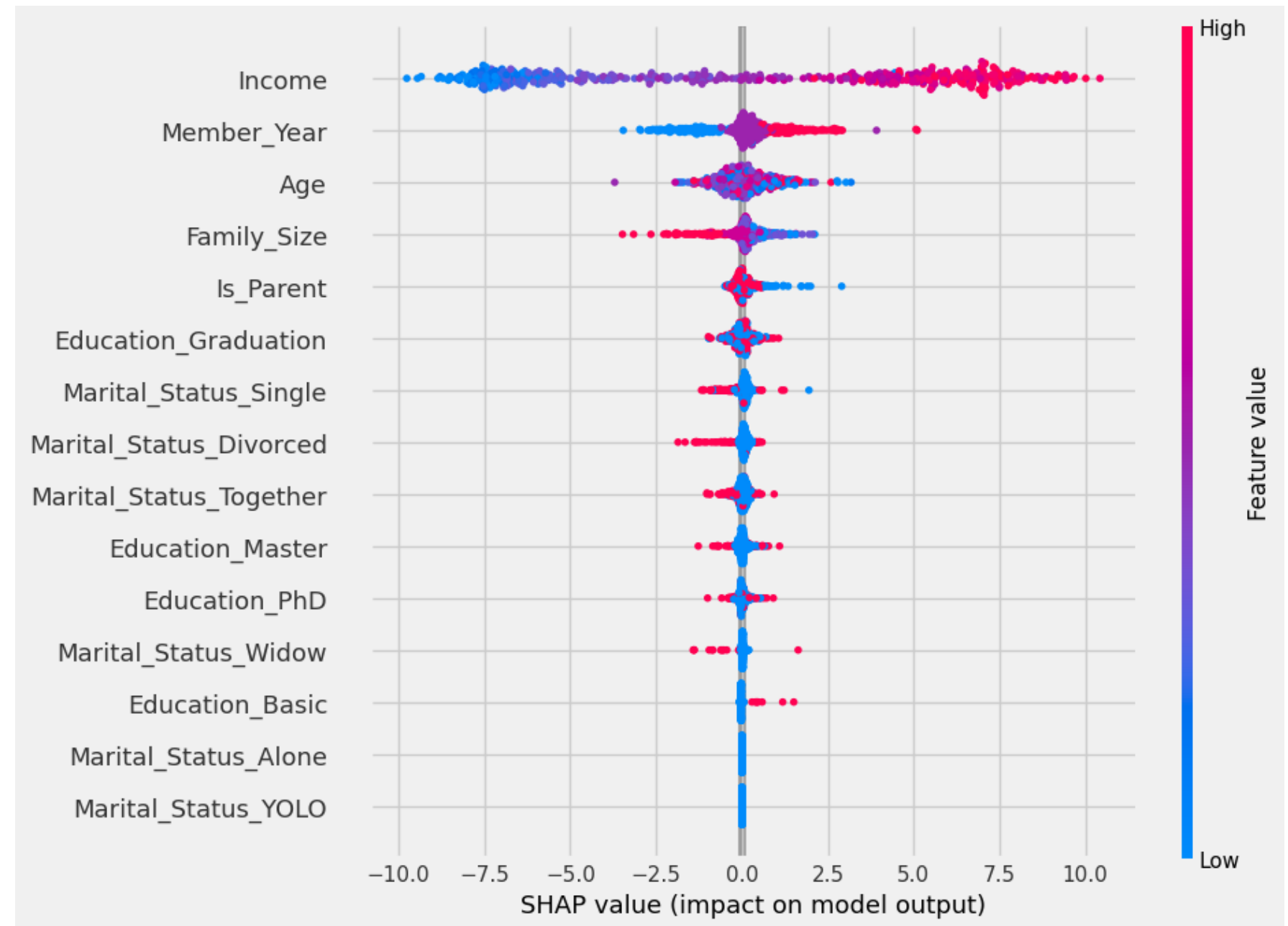
Treatment: Marital Status Married
Target: Total Purchased

LR Model RMSE: 12.34

XGB Model RMSE: 15.43

LRSRegressor ATE: 1.75 (± 0.962)

XGBRegressor ATE: 0.21(± 0.485)



Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Quantifying Impact: Causal Inference on Customer Purchasing Decisions

Treatment: Complain

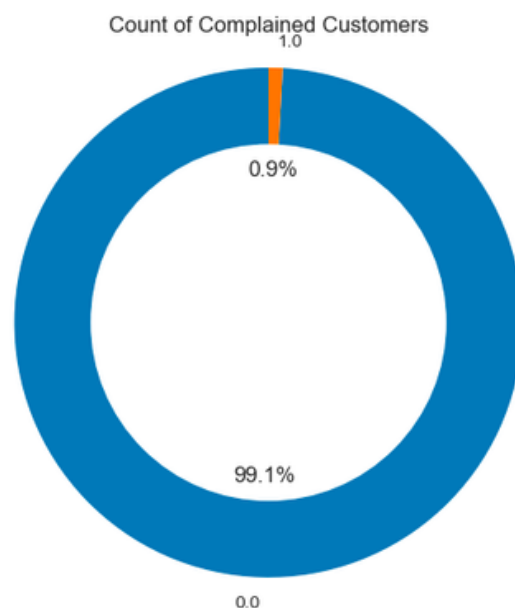
Target: Total Purchased

LR Model RMSE: 15.33

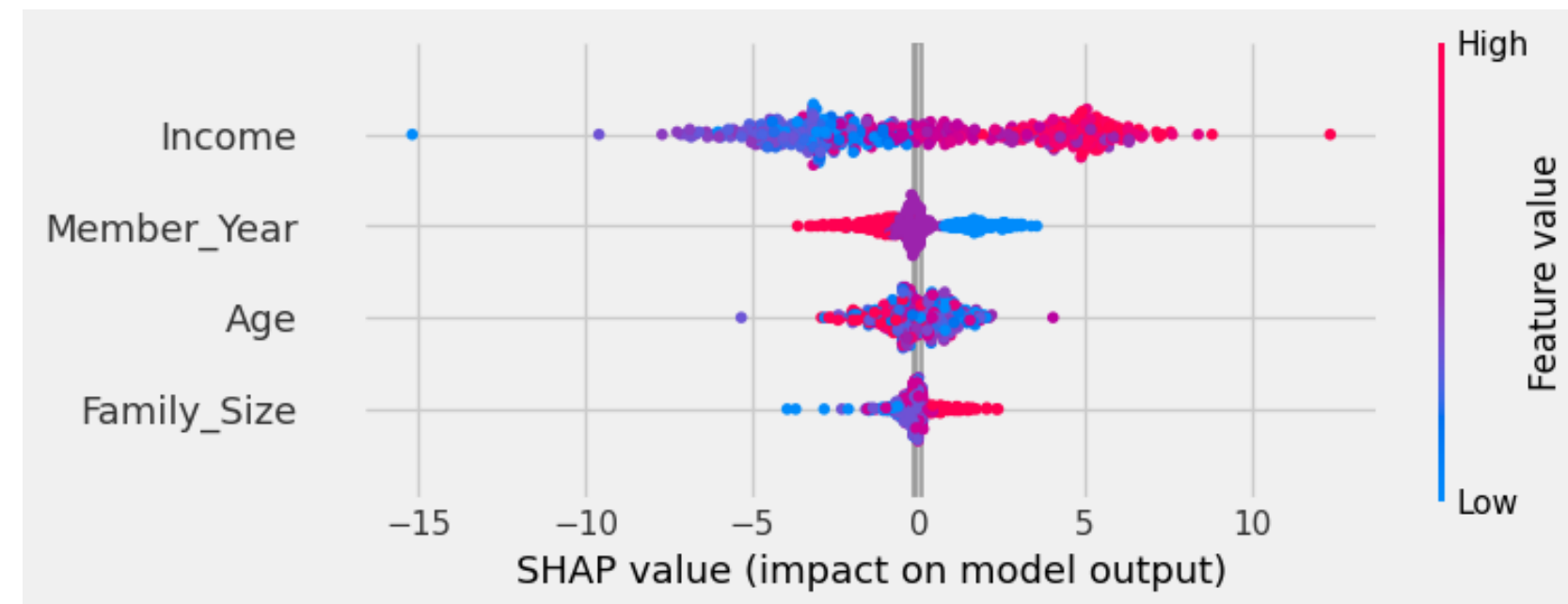
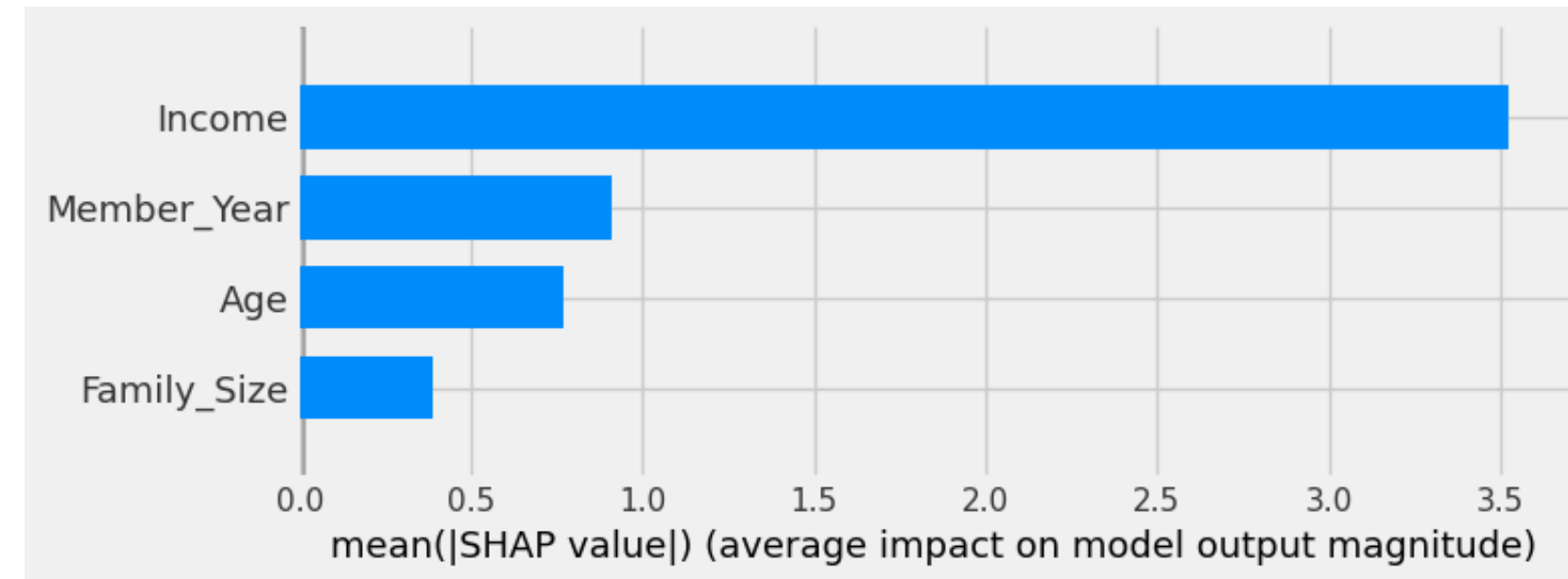
XGB Model RMSE: 13.83

LRSRegressor ATE: -0.56

XGBRegressor ATE: 1.06



Good Customer Satisfaction
1% complaint rate



Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Quantifying Impact: Causal Inference on Customer Purchasing Decisions

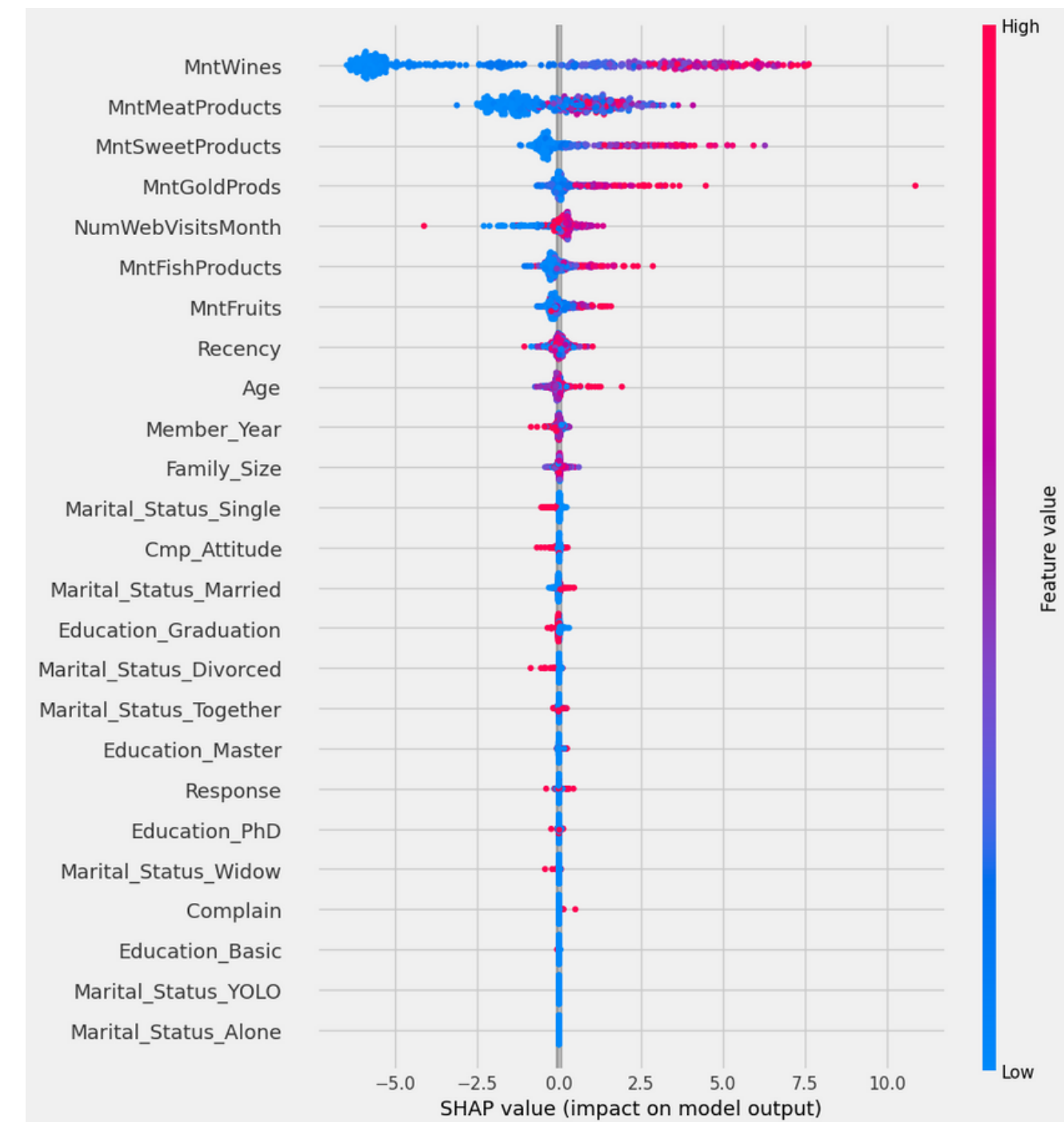
Treatment: Income

- High \geq Income Median
- Low $<$ Income Median
- Income Median = \$51,267

Target: All purchase (transaction counts)

LR Model RMSE: 11.14

LRSRegressor ATE: 4.07 (± 0.50)



Challenge &
Hypothesis

Data Landscape

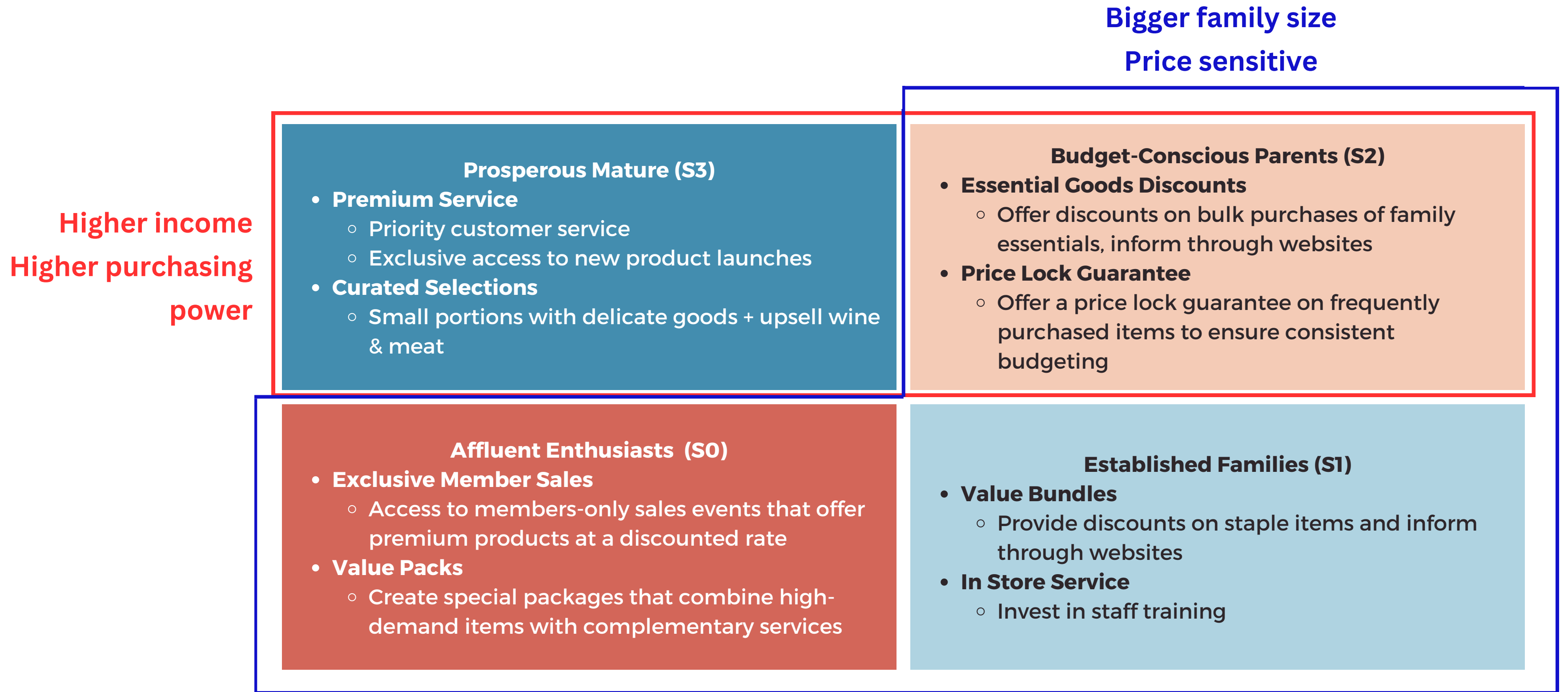
Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion & Future
Steps

Charting the Path Forward: Advice for Marketing Strategies



Challenge &
Hypothesis

Data Landscape

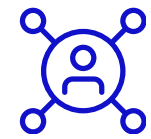
Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion &
Future Steps

Charting the Path Forward: Future Steps



Models Validity Enhancement

- Test with more unsupervised learning models without specifying K to gain more insights
- Explore semi-supervised learning with more data
- Explore self-supervised learning with images and texts
- Incorporate R learner for causal inference to estimate heterogeneous treatment effects



4Vs of Data

- Address data and business context limitations by updating the dataset to include recent information and engaging business SME
- Integrate diverse data types: product images, customer reviews, addresses, channel
- Increase data volume for a more robust analysis

Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion &
Future Steps

Charting the Path Forward: Advice for Store Management

Incorporate segments personas in these area:

**Targeted
Engagement**

Customized
Channel &
Approach

**Product
Selection &
Development**

Tailored to
customer needs

**Service
Excellence**

In-store & Online

**Technology
Optimization**

Leverage data
science & ERP
system

Challenge &
Hypothesis

Data Landscape

Exploratory data
analysis

Clustering Model

Causal Inference
Model

Conclusion &
Future Steps

Team 6

THANK YOU

WE ARE OPEN TO QUESTIONS