PROJECT PRESENTATION

AGENDA

Business Problem Overview

Data Overview

EDA – Univariate / Multivariate Analysis

K Means Clustering

Cluster Profiling and Analysis

Business Recommendations

BUSINESS PROBLEM OVERVIEW

BUSINESS PROBLEM OVERVIEW

In today's market and its competitive environment , businesses must improve their marketing strategies , this company face a major challenge : How to better understand and engage with its diverse customer population ? , the company wants to uncover the personality traits and purchasing behaviors , by doing this they aim to segment the customer base into meaningful groups , through advanced segmentation (clustering techniques) the company hopes to build a foundation for the sake of engagement and long term customer loyalty

DATA OVERVIEW

• Data types include int64, float64, object, and datetime.

• Income column had 24 missing values (handled by row removal.)

• No duplicate rows found.

• Average Household Income: 52247.25

• we dropped every irrelevant columns which include:
'Dt_Customer','Year_Birth','ID','AcceptedCmp1', 'Z_CostContact',
'Z_Revenue' , and more

DATA OVERVIEW

index	Missing Before	Missing After
AcceptedCmp1	0	0
AcceptedCmp2	0	0
AcceptedCmp3	0	0
AcceptedCmp4	0	0
AcceptedCmp5	0	0
Age	0	0
Complain	0	0
Dt_Customer	0	0
Education	0	0
ID	0	0
Income	24	0
Kidhome	0	0
Marital_Status	0	0
MntFishProducts	0	0
MntFruits	0	0
MntGoldProds	0	0
MntMeatProducts	0	0
MntSweetProducts	0	0
MntWines	0	0
NumCatalogPurchases	0	0
NumDealsPurchases	0	0
NumStorePurchases	0	0
NumWebPurchases	0	0
NumWebVisitsMonth	0	0
Recency	0	0

1. This figure shows missing data before and after handling

EDA – UNIVARIATE / MULTIVARIATE ANALYSIS

EDA – UNIVARIATE / MULTIVARIATE ANALYSIS

Univariate analysis :

- Kidhome/Teenhome: These features show that a significant portion of customers have no children or teenagers at home, while others have one or two. This can influence spending patterns

- Most spending features are right-skewed, indicating that a large portion of customers spend little to no money on certain product categories, while a smaller group spends significantly. This suggests distinct customer segments based on purchasing habits
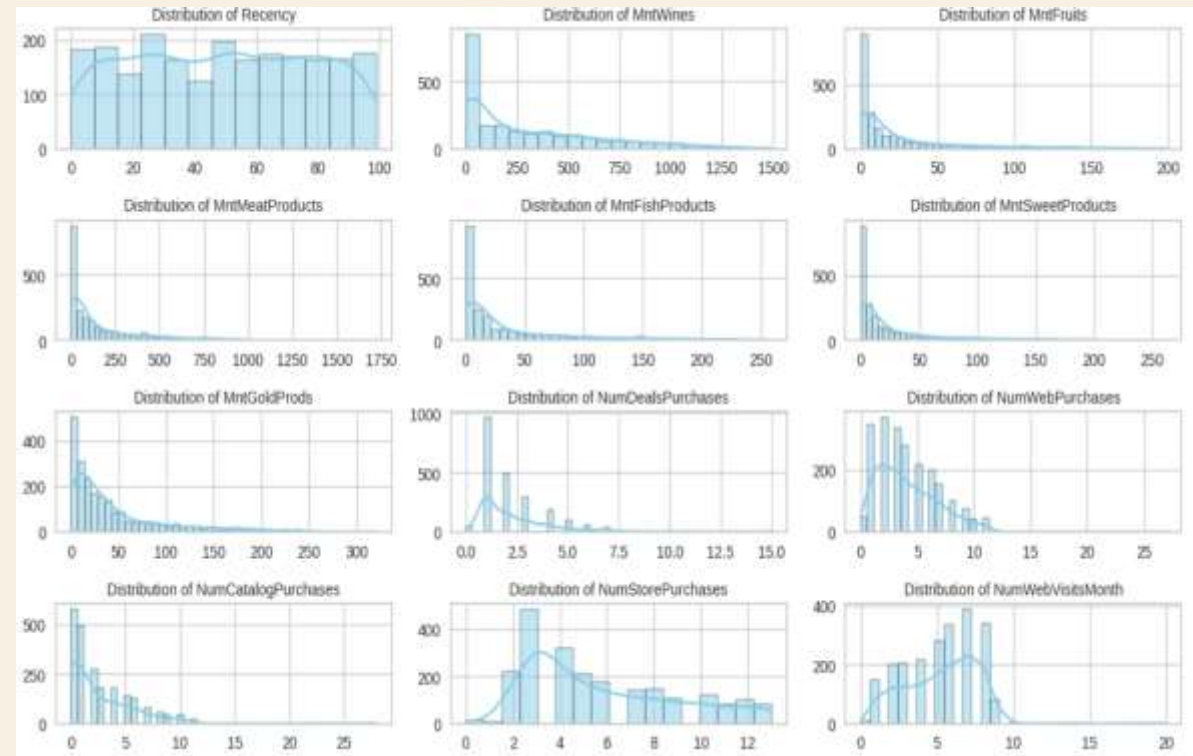
Multivariate/ Bivariate analysis :

- Correlation Heatmap: The heatmap reveals strong positive correlations among the spending features (e.g., MntWines with MntMeatProducts, MntFruits with MntSweetProducts). This suggests that customers who spend on one category tend to spend on others. Income also shows a positive correlation with most spending categories, indicating that higher-income customers tend to spend more.
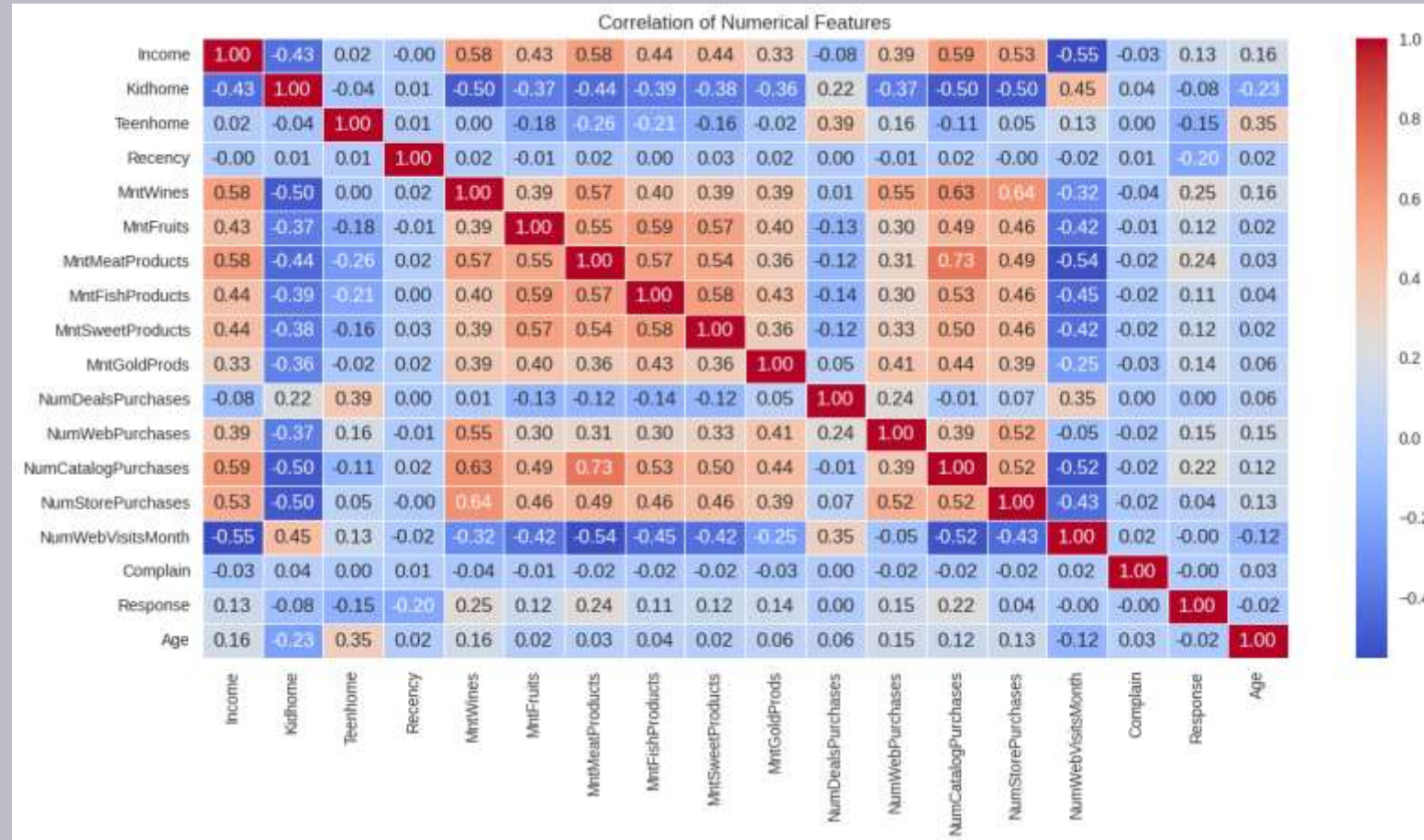
EDA – UNIVARIATE / MULTIVARIATE ANALYSIS
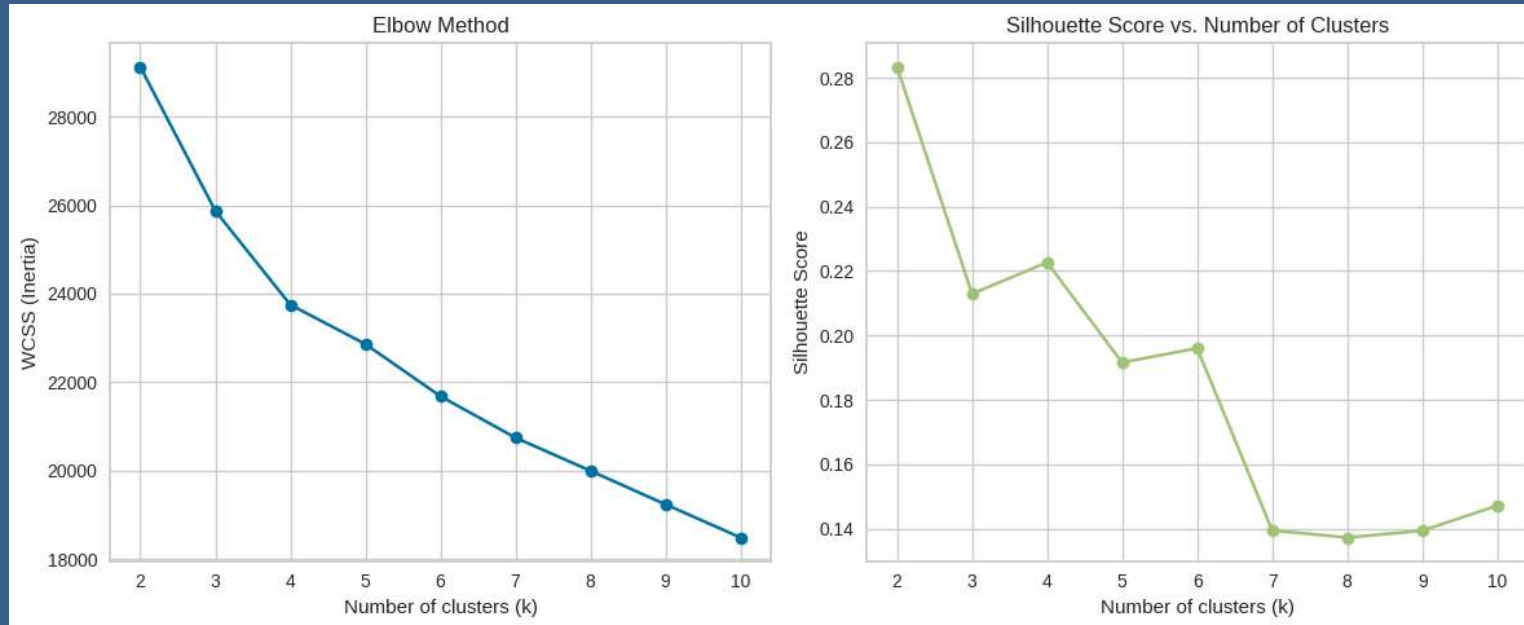

3. Univariate figure (boxplot)

2.The Univariate figure (histogram)

EDA – UNIVARIATE / MULTIVARIATE ANALYSIS

4.The Multivariate figure (heatmap)

K MEANS CLUSTERING

Number of Clusters: 2	WCSS: 29127.70	Silhouette Score: 0.2834
Number of Clusters: 3	WCSS: 25873.79	Silhouette Score: 0.2129
Number of Clusters: 4	WCSS: 23749.54	Silhouette Score: 0.2227
Number of Clusters: 5	WCSS: 22862.10	Silhouette Score: 0.1916
Number of Clusters: 6	WCSS: 21681.77	Silhouette Score: 0.1960
Number of Clusters: 7	WCSS: 20746.06	Silhouette Score: 0.1395
Number of Clusters: 8	WCSS: 19996.86	Silhouette Score: 0.1373
Number of Clusters: 9	WCSS: 19239.07	Silhouette Score: 0.1394
Number of Clusters: 10	WCSS: 18489.21	Silhouette Score: 0.1472

Based on Silhouette Score, the best number of clusters is k = 2 , anyway There's a moderate "elbow" between k = 3 and k = 5 , we will go with 4 because it have the best Silhouette Score in that range which have total time of 9.4 ms & 9.79 ms wall time

5.Figures show elbow method & Silhouette score

CLUSTER PROFILING AND ANALYSIS

- Cluster 1 has some teen/kids at home more than the others

- Cluster 3 is the ideal customer segment

- Cluster 2 is not very active on website (purchases)

cluster/ category	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Income	Low	Moderate	Low to mid	High
spending	Minimal across all	Higher than 0 high spending on **wines and gold**	Medium	High
Engage & response	Very low	Engaged more in store purchases	Low	High store/Web activity
Recency	Inactive / uninterested	Medium	High	Highest

12

CLUSTER PROFILING AND ANALYSIS

Barplots of All Numerical Variables by Cluster
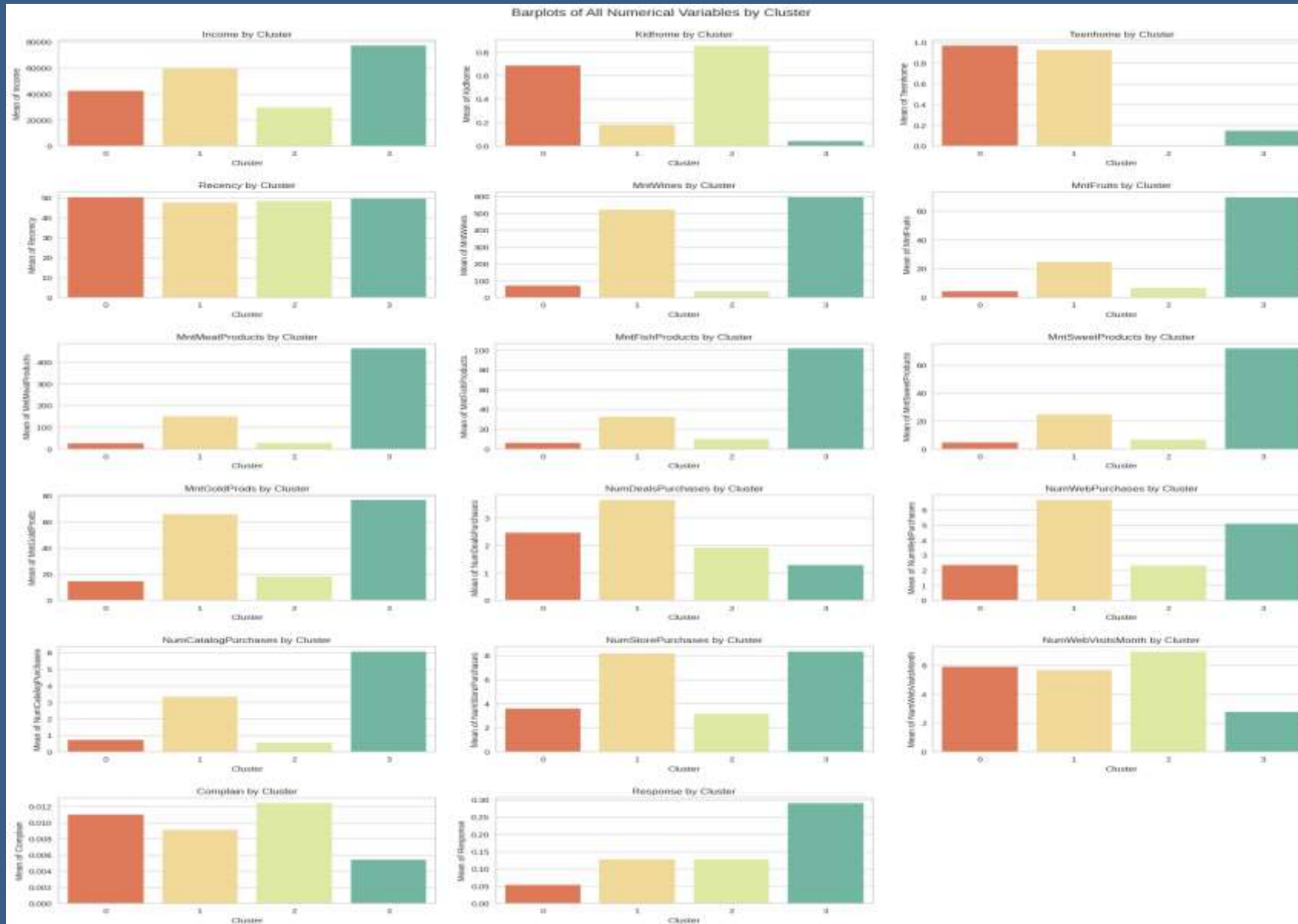
6. figure shows all barplots of all numerical variables by cluster

CLUSTER PROFILING ANALYSIS SUMMARY

Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Income	Low	Medium	Low	Very High
Spending	Very Low	Medium-High	Medium	Very High
Purchases	Low	Medium	Medium	High
Web Visits	Low	Medium	Medium	High
Campaign Response	Very Low	Medium	Low	High
Recency	High (inactive)	Medium	High	Low (active)

BUSINESS RECOMMENDATIONS

Cluster 0 "**cold leads**"

Profile : Low income , low spending , high recency with little interaction

Recommendations :

✓ Don't heavily invest here

✓ Use low cost mass campaigns

Cluster 1 "**Shy engagers**"

Profile : Moderate income & spending , more store engagement

Recommendations :

✓ Consider cross-channel nurturing

✓ Offer product bundles , seasonal discounts

BUSINESS RECOMMENDATIONS

Cluster 2 "**Cation spenders**"

Profile : Lower spending and engagement than Cluster 1 but better than 0

Recommendations :

✓ Send reactivation offers if they are inactive

Cluster 3 "**High value customers** "

Profile : very high income & spending , low recency

Recommendations :

✓ Treat them as VIPs (premium services , early access, etc.)

✓ Create predictive models to suggest products

✓ Give them priority in service , inventory , and targeting

BUSINESS RECOMMENDATIONS - GENERAL

✓ Focus investments on cluster 3 and nurture cluster 1

✓ Reactive cluster 2 cautiously

✓ Minimize efforts on cluster 0

THANK YOU

Nawaf Alzahrani

Nawaf4zh5@gmail.com