Customer Profiling Case Study

Case Study Overview

- This case study is about Customer Profiling & Targeted Marketing Strategy.
- A company wants to understand customers better, segment them, and predict which customers will accept marketing campaign offers.
- We were expected to use demographic, lifestyle, purchase, and engagement data to build customer behavior features, create segments via clustering, and build a prediction model to identify customers who are likely to respond to campaigns.

Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
ID	Year_Birtl	Educatior	Marital_S	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWine:	MntFruits	MntMeatl	MntFishP	MntSweet	MntGoldP	NumDeal:	NumWebl	NumCatal	NumStore	NumWeb\	Accepted(	Accepted(	Accepted(	Accepted(	Accepted(	Complain	Z_CostCor	Z_Revenu	Response
5524	1957	Graduatic	Single	58138	0	0	04-09-2012	58	635	88	546	172	88	88	3	8	10	4	7	0	0	0	0	0	0	3	11	1
2174	1954	Graduatic	Single	46344	1	1	08-03-2014	38	11	1	6	2	1	6	2	1	1	2	5	0	0	0	0	0	0	3	11	0
4141	1965	Graduatic	Together	71613	0	0	21-08-2013	26	426	49	127	111	21	42	1	8	2	10	4	0	0	0	0	0	0	3	11	0
6182	1984	Graduatic	Together	26646	1	0	10-02-2014	26	11	4	20	10	3	5	2	2	0	4	6	0	0	0	0	0	0	3	11	0
5324	1981	PhD	Married	58293	1	0	19-01-2014	94	173	43	118	46	27	15	5	5	3	6	5	0	0	0	0	0	0	3	11	0
7446	1967	Master	Together	62513	0	1	09-09-2013	16	520	42	98	0	42	14	2	6	4	10	6	0	0	0	0	0	0	3	11	0
965	1971	Graduatic	Divorced	55635	0	1	13-11-2012	34	235	65	164	50	49	27	4	7	3	7	6	0	0	0	0	0	0	3	11	0
6177	1985	PhD	Married	33454	1	0	08-05-2013	32	76	10	56	3	1	23	2	4	0	4	8	0	0	0	0	0	0	3	11	0
4855	1974	PhD	Together	30351	1	0	06-06-2013	19	14	0	24	3	3	2	1	3	0	2	9	0	0	0	0	0	0	3	11	1
5899	1950	PhD	Together	5648	1	1	13-03-2014	68	28	0	6	1	1	13	1	1	0	0	20	1	0	0	0	0	0	3	11	0
1994	1983	Graduatic	Married		1	0	15-11-2013	11	5	5	6	0	2	1	1	1	0	2	7	0	0	0	0	0	0	3	11	0
387	1976	Basic	Married	7500	0	0	13-11-2012	59	6	16	11	11	1	16	1	2	0	3	8	0	0	0	0	0	0	3	11	0
2125	1959	Graduatic	Divorced	63033	0	0	15-11-2013	82	194	61	480	225	112	30	1	3	4	8	2	0	0	0	0	0	0	3	11	0
8180	1952	Master	Divorced	59354	1	1	15-11-2013	53	233	2	53	3	5	14	3	6	1	5	6	0	0	0	0	0	0	3	11	0
2569	1987	Graduatic	Married	17323	0	0	10-10-2012	38	3	14	17	6	1	5	1	1	0	3	8	0	0	0	0	0	0	3	11	0
2114	1946	PhD	Single	82800	0	0	24-11-2012	23	1006	22	115	59	68	45	1	7	6	12	3	0	0	1	1	0	0	3	11	1
9736	1980	Graduatic	Married	41850	1	1	24-12-2012	51	53	5	19	2	13	4	3	3	0	3	8	0	0	0	0	0	0	3	11	0
4939	1946	Graduatic	Together	37760	0	0	31-08-2012	20	84	5	38	150	12	28	2	4	1	6	7	0	0	0	0	0	0	3	11	0
6565	1949	Master	Married	76995	0	1	28-03-2013	91	1012	80	498	0	16	176	2	11	4	9	5	0	0	0	1	0	0	3	11	0
2278	1985	2n Cycle	Single	33812	1	0	03-11-2012	86	4	17	19	30	24	39	2	2	1	3	6	0	0	0	0	0	0	3	11	0
9360	1982	Graduatic	Married	37040	0	0	08-08-2012	41	86	2	73	69	38	48	1	4	2	5	8	0	0	0	0	0	0	3	11	0
5376	1979	Graduatic	Married	2447	1	0	06-01-2013	42	1	1	1725	1	1	1	15	0	28	0	1	0	0	0	0	0	0	3	11	0

Dataset Overview

- Contains customer demographics, spending behaviour, and engagement data
- Includes Income, Age, Education, Marital Status, Kid/Teen home
- Tracks product-wise spending (Wines, Meat, Fish, etc.)
- Records channel usage: Web, Store, Catalog
- Includes campaign responses, complaints, and recency

Customer Profiling & Targeted Marketing

- Identifies different customer types based on behaviour and needs
- Helps tailor products, services, and marketing messages
- Avoids wasteful marketing by focusing on likely buyers
- Improves conversion, personalization, and ROI

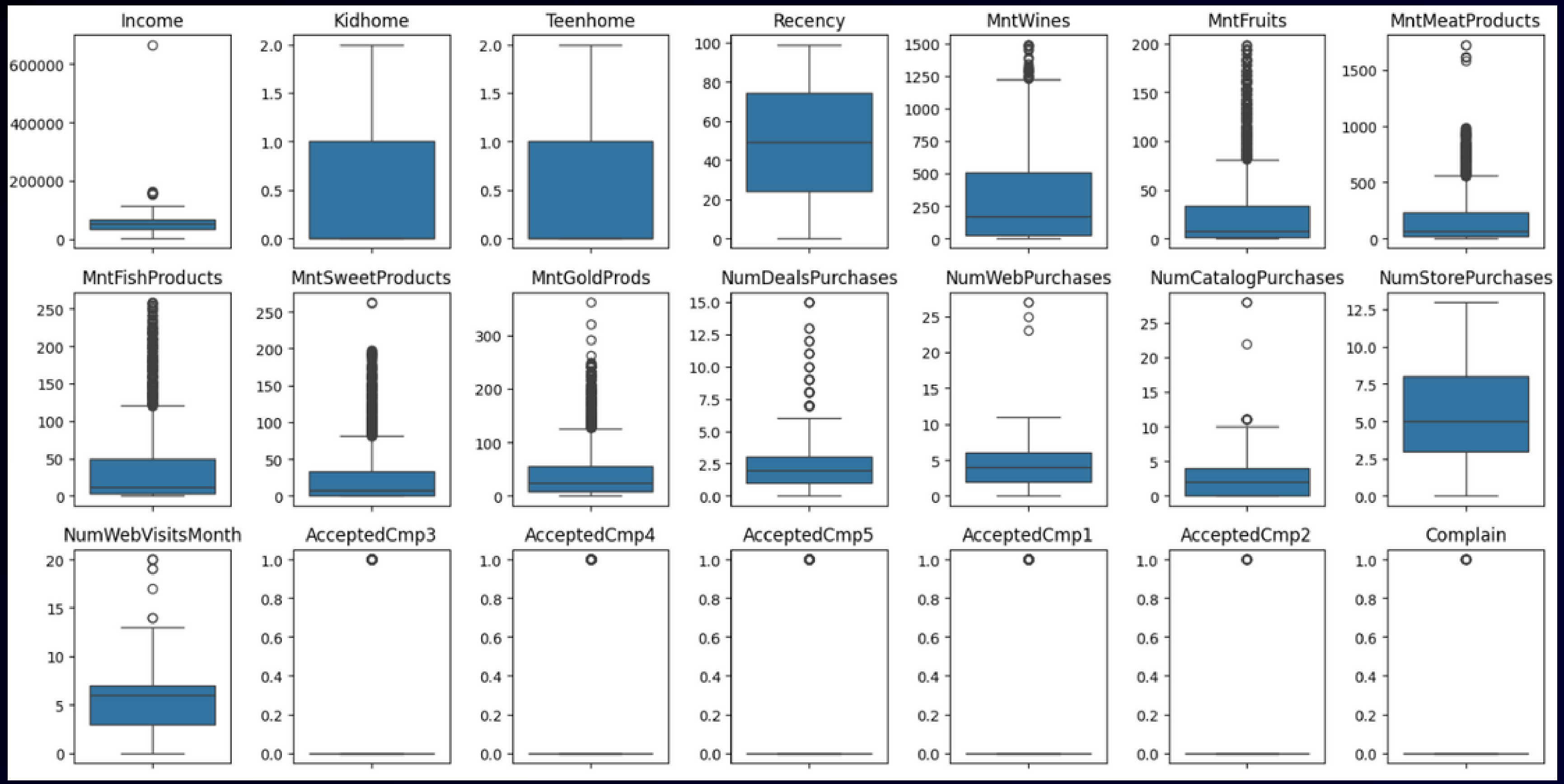
Imputing Missing Values

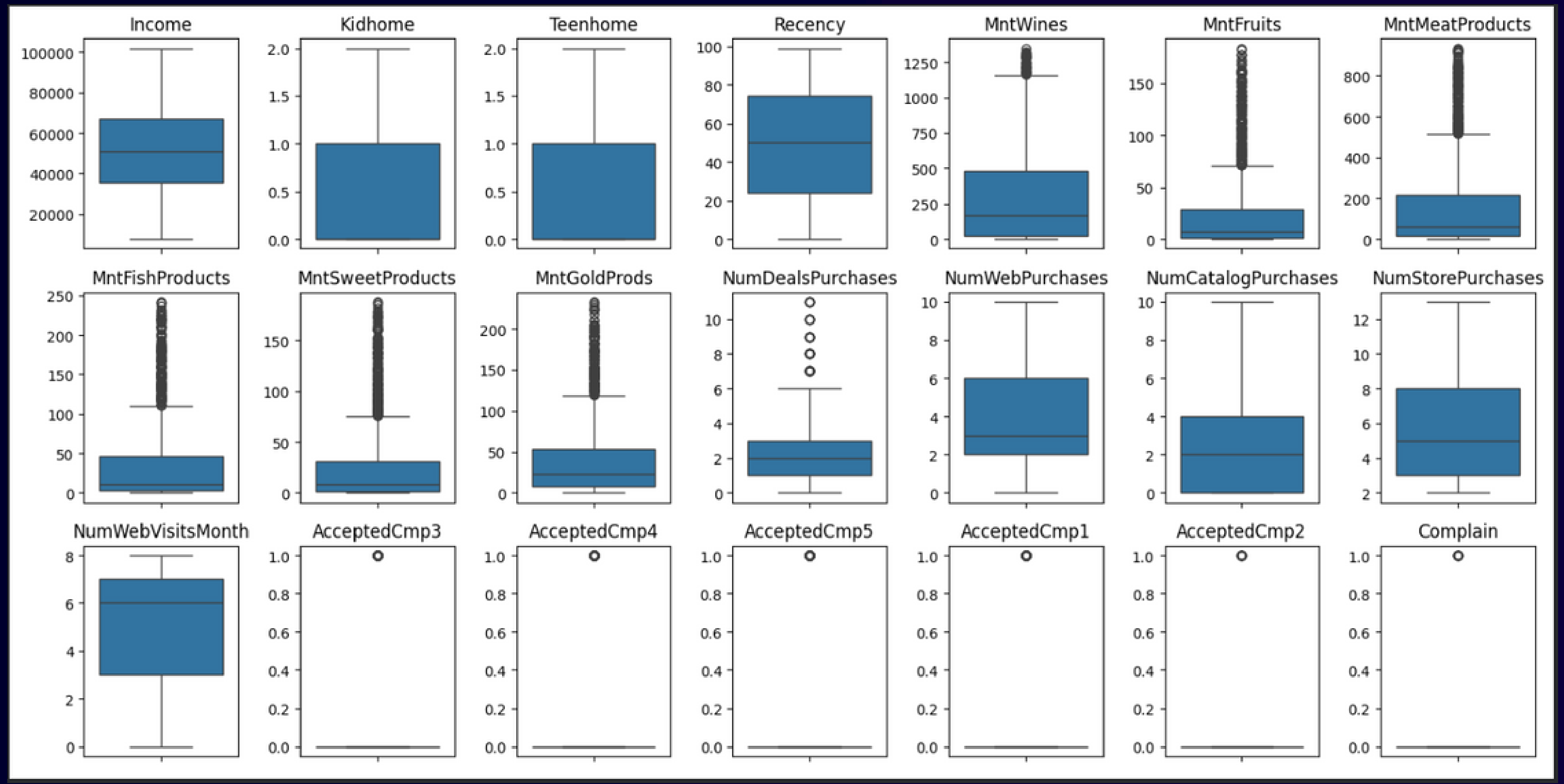
	0
ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	24
Kidhome	0
Teenhome	0
Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
NumDealsPurchases	0
NumWebPurchases	0
NumCatalogPurchases	0
NumStorePurchases	0
NumWebVisitsMonth	0
AcceptedCmp3	0

	0
ID	0
Year_Birth	0
Education	0
Marital_Status	0
Income	0
Kidhome	0
Teenhome	0
Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
NumDealsPurchases	0
NumWebPurchases	0
NumCatalogPurchases	0
NumStorePurchases	0
NumWebVisitsMonth	0
AcceptedCmp3	0

- **A total of 24 missing values were identified in the Income column.**
- **These missing entries were imputed to ensure data completeness and maintain analysis accuracy.**

Treating Outliers >99.5 percentile





- **Initial boxplot analysis revealed outliers in *Income, MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds, NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, and NumWebVisitsMonth.***
- **These outliers were treated by removing values above the 99.5th percentile, resulting in a cleaner and more reliable dataset.**

Converted Dates

Transformed the customer's joining date into how long they have been with the company, in both days and months, and also calculated how many years ago they joined.

```
Date conversion done. Sample:
```

	Dt_Customer	Customer_Tenure_Days	Customer_Tenure_Months	Enrollment_Year
0	2012-09-04	4830	158.7	13
1	2014-03-08	4280	140.6	11
2	2013-08-21	4479	147.1	12
3	2014-02-10	4306	141.5	11
4	2014-01-19	4328	142.2	11

Column	Value
Dt_Customer	9/4/2012
Customer_Tenure_Days	4830 days
Customer_Tenure_Months	158.7 months
Enrollment_Year	13

Customer Behavior Features

```
Using spend column: MntWines
Using spend column: MntFruits
Using spend column: MntMeatProducts
Using spend column: MntFishProducts
Using spend column: MntSweetProducts
Using spend column: MntGoldProds
```

	Total_Expenditure	Avg_Monthly_Spend	Dependency_Ratio	Engagement_Score
0	1617	10.189036	0.0	5.2
1	27	0.192034	2.0	3.2
2	776	5.275323	0.0	7.6
3	53	0.374558	1.0	4.8
4	422	2.967651	1.0	5.6

- **Total_Expenditure = sum of all category spending = Wines + Fruits + Meat + Fish + Sweet + Gold**
- **Average_Monthly_Spend = Total_Expenditure / Customer_Tenure_in_Months**
- **Dependency_Ratio = (Kids + Teens) / Adults**
- **Engagement_Score = (Web Visits * 0.4) + (Store Purchases * 0.6)**

Campaign Response Target Variable

```
Campaign columns found: ['AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1', 'AcceptedCmp2', 'Response']
Campaign_Response distribution:
 Campaign_Response
0    0.748248
1    0.251752
Name: proportion, dtype: float64
```

The distribution shows the proportion of customers who accepted vs not accepted:

0 → 74.87% (majority did NOT accept any campaign)

1 → 25.17% (only 1 in 4 accepted at least one campaign)

- Converted campaign acceptances (5 campaign indicators) into a single label
- Campaign_Response = 1, if any campaign was accepted. Else
- Campaign_Response = 0

Encode Categorical Variables and Logistic Regression and Random Forest

- **One-hot encode: Education, Marital Status**
- **Binary encode: Has_Kids, Has_Teenagers**
- **Normalize Income, Expenditure, Tenure**

```
... Logistic Regression AUC: 0.7876419748998311
                  precision    recall  f1-score   support

               0       0.80      0.91      0.85       299
               1       0.57      0.34      0.42       101

        accuracy                           0.77       400
       macro avg       0.68      0.62      0.64       400
    weighted avg       0.74      0.77      0.75       400

    Random Forest AUC: 0.8474287228053909
                  precision    recall  f1-score   support

               0       0.85      0.95      0.90       299
               1       0.78      0.50      0.61       101

        accuracy                           0.84       400
       macro avg       0.81      0.72      0.75       400
    weighted avg       0.83      0.84      0.82       400

    Top feature importances:
     Avg_Monthly_Spend          0.145387
    Total_Expenditure          0.132876
    Income                     0.122036
    Recency                    0.089352
    Customer_Tenure_Days       0.086033
    Engagement_Score           0.067692
    NumCatalogPurchases        0.066715
    NumStorePurchases          0.064039
    NumWebPurchases            0.057544
    NumDealsPurchases          0.045206
    Dependency_Ratio           0.028676
    Education_PhD              0.015230
    Education_Graduation       0.014134
    Marital_Status_Married     0.013478
    Marital_Status_Together    0.013101
    Marital_Status_Single      0.011760
    Education_Master           0.011064
    Marital_Status_Divorced    0.007544
    Marital_Status_Widow       0.004813
    Education_Basic            0.001931
    dtype: float64
```

- Random Forest performs best

AUC jumps from 0.78 → 0.84, outperforming Logistic Regression.

- Campaign acceptors are harder to predict

Class 1 recall low due to very few customers accepting campaigns.

- Spending behaviour drives predictions

Avg Monthly Spend, Total Spend, Income = top predictors.

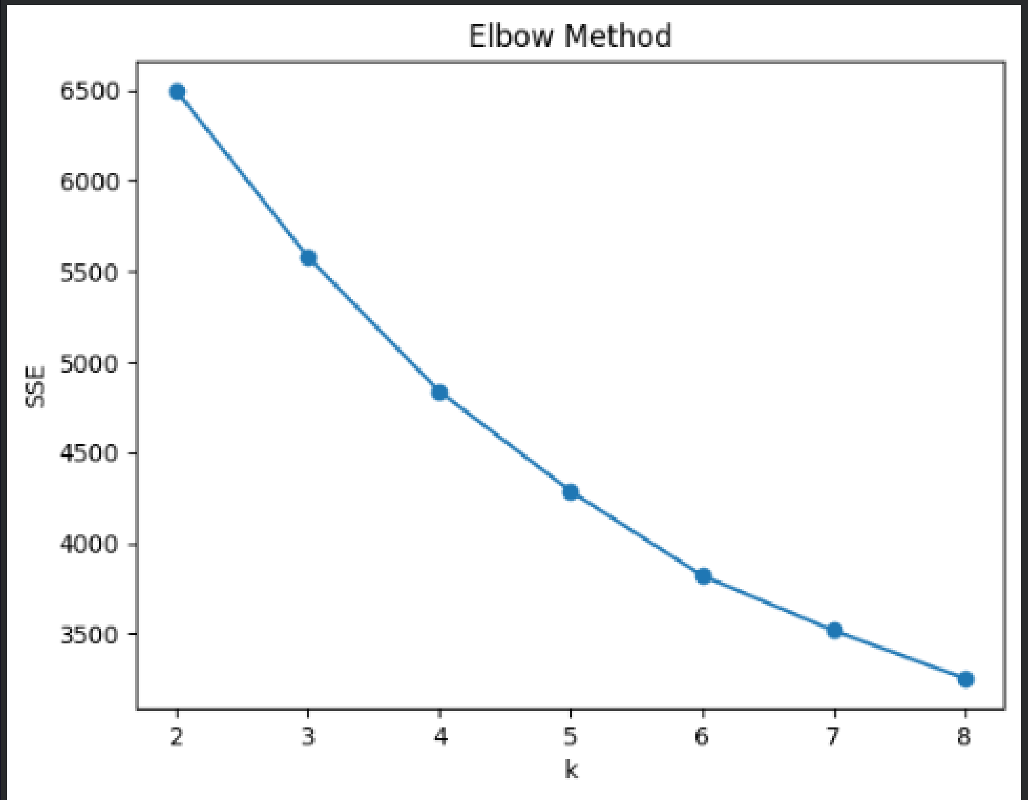
- Engagement & Recency matter

Recency, Tenure, Engagement Score, and purchase channels boost acceptance likelihood.

- Demographics add minimal value

Education & Marital Status have low importance vs behaviour data

Clustering based on behavior features




```
Silhouette score: 0.22566877016629966
```

cluster	count
2	592
1	589
0	459
3	358

dtype: int64

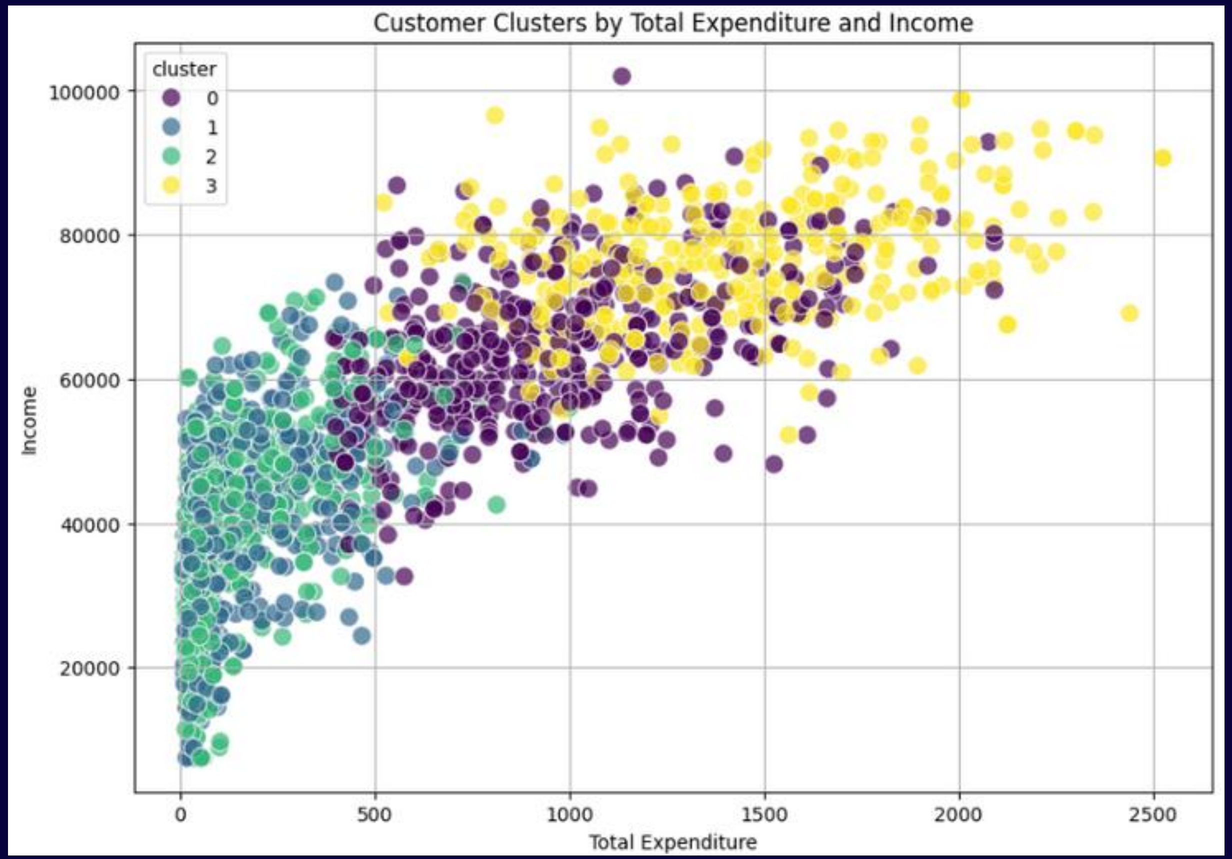
Elbow Method → Optimal K = 4
SSE drop flattens after 4, clearly showing the “elbow.”

Balanced Cluster Sizes
592, 589, 459, 358 — no cluster too large or too small.

Silhouette Score = 0.225 → Acceptable Clustering
Expected for customer behaviour; some natural overlap.

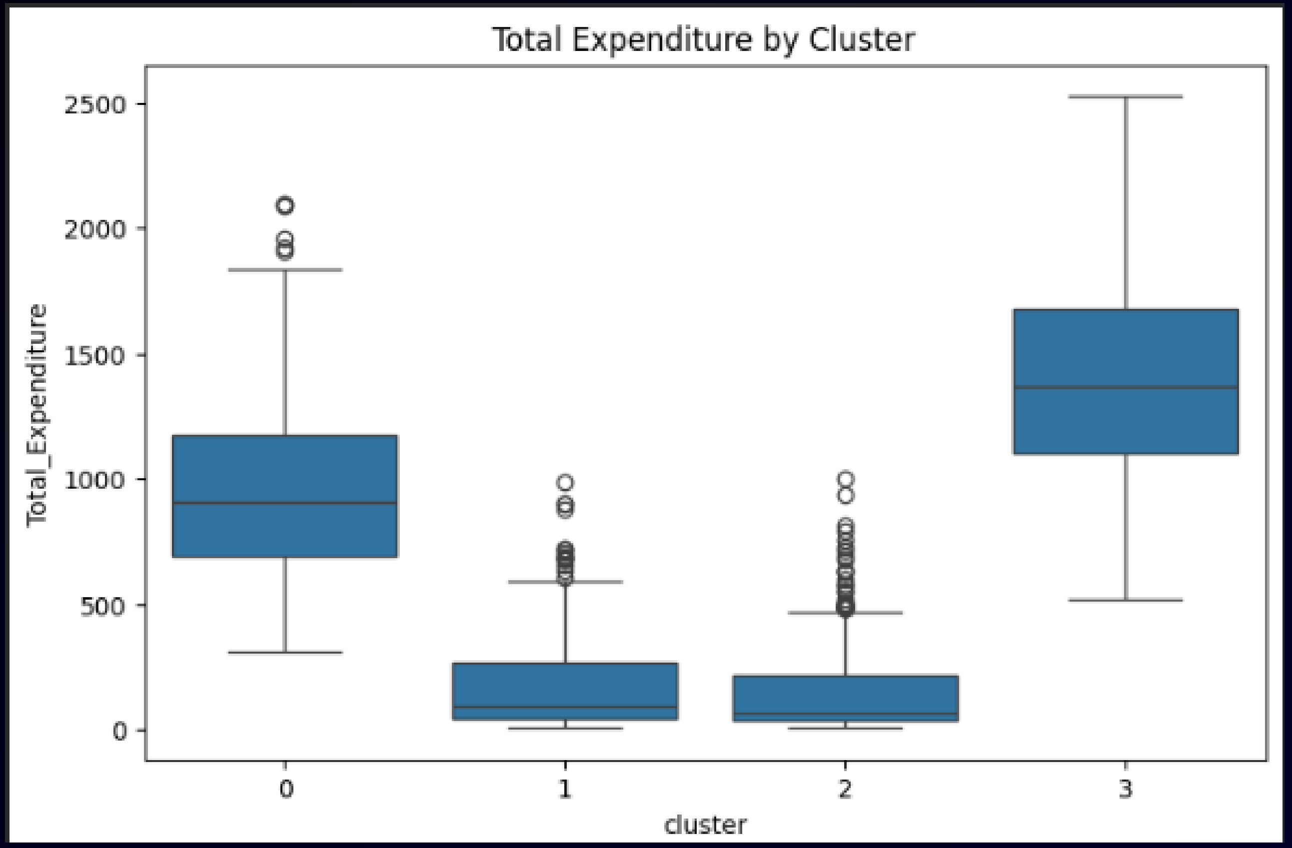
Enables Targeted Marketing Strategies
Each cluster can receive personalized offers → higher ROI.

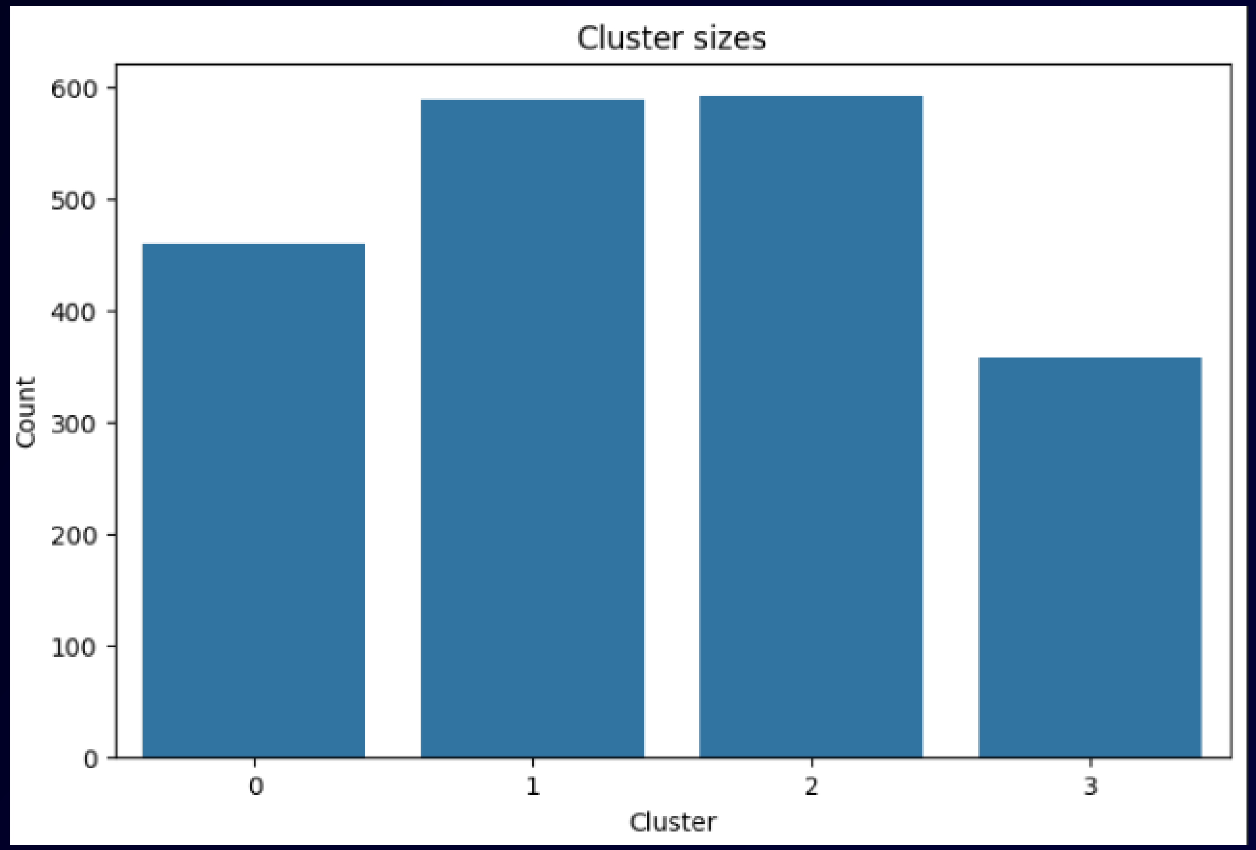
Clusters Capture Distinct Behaviour Patterns
Groups formed based on spending, engagement & campaign response.

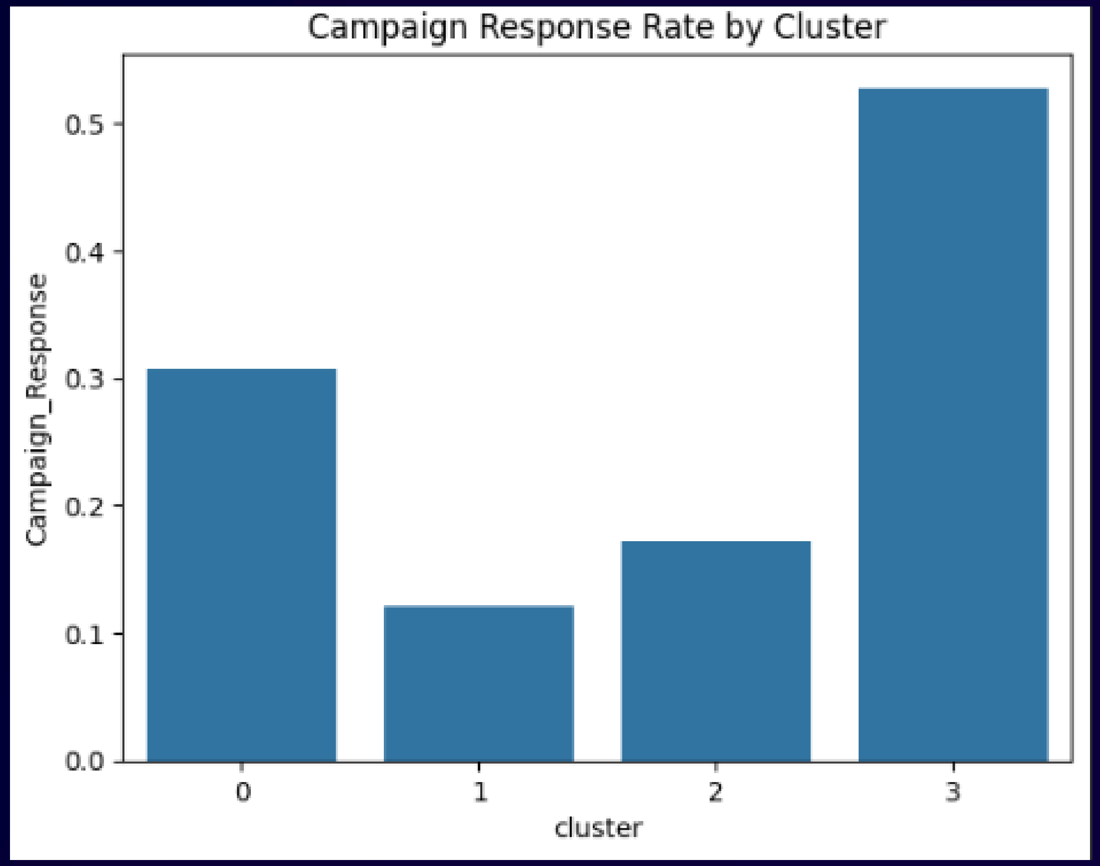



```
--- Cluster Definitions ---
Cluster 0 (459 members): High expenditure (~$965), good engagement (~8.06), moderate income (~$64.6K). Relatively high campaign response rate (~30.7%).
Cluster 1 (589 members): Low expenditure (~$174), low engagement (~4.76), moderate income (~$39K). Higher recency (not purchased recently). Lowest campaign response rate (~12%).
Cluster 2 (592 members): Low expenditure (~$146), low engagement (~4.61), moderate income (~$37.7K). Lower recency (purchased more recently). Moderate campaign response rate (~17.2%).
Cluster 3 (358 members): Highest expenditure (~$1406), highest income (~$76.3K). Moderate engagement (~5.22). Highest campaign response rate (~52.8%).
-------------------------
```

Campaign Response







- Cluster 3 shows the highest total expenditure and the strongest campaign response, making it the premium high-value segment.
- Cluster 0 has moderate spending and a fair response rate, indicating a reasonably engaged customer group.
- Cluster 1 has the lowest spending and the weakest response, representing low-value, low-engagement customers.
- Cluster 2 spends slightly more than Cluster 1 but still responds weakly, suggesting price-sensitive customers.

Conclusion

Cluster 3 – High Spenders & High Responders (Premium)

- Highest spend + strongest campaign response
- Loyal, engaged, premium-oriented customers
- Prefer exclusive offers & personalized recommendations
- Best target for upsell/cross-sell and VIP programs

Cluster 0 – Medium Spenders, Moderately Engaged

- Mid-level spenders with stable engagement
- Good potential for repeat purchases
- Responds to value-driven deals & reminders
- Suitable for loyalty points and combo offers

Cluster 2 – Low/Price-Sensitive Spenders

- Low-to-mid total spend; deal-driven buyers
- Purchase only during discounts/festive offers
- React better to catalog promotions
- Ideal for B1G1, budget bundles, and flash sale alerts

Cluster 1 – Low Engagement & Low Spending

- Lowest spending + weak campaign response
- Hardest to convert; minimal brand attachment
- Needs re-engagement or win-back offers
- Avoid heavy spend on campaigns; use low-cost channels

Thank You!