UNIVERSITY OF OULU

521251S-3003: Special Course in Information Technology 3 - Data Mining Project

CUSTOMER PERSONALITY ANALYSIS

Group Members

Arash Nedaei Janbesaraei
Zohreh Yousefi Dahka
Kavinda Kulasinghe
Taufiq Ahmed
Md Mobusshar Islam

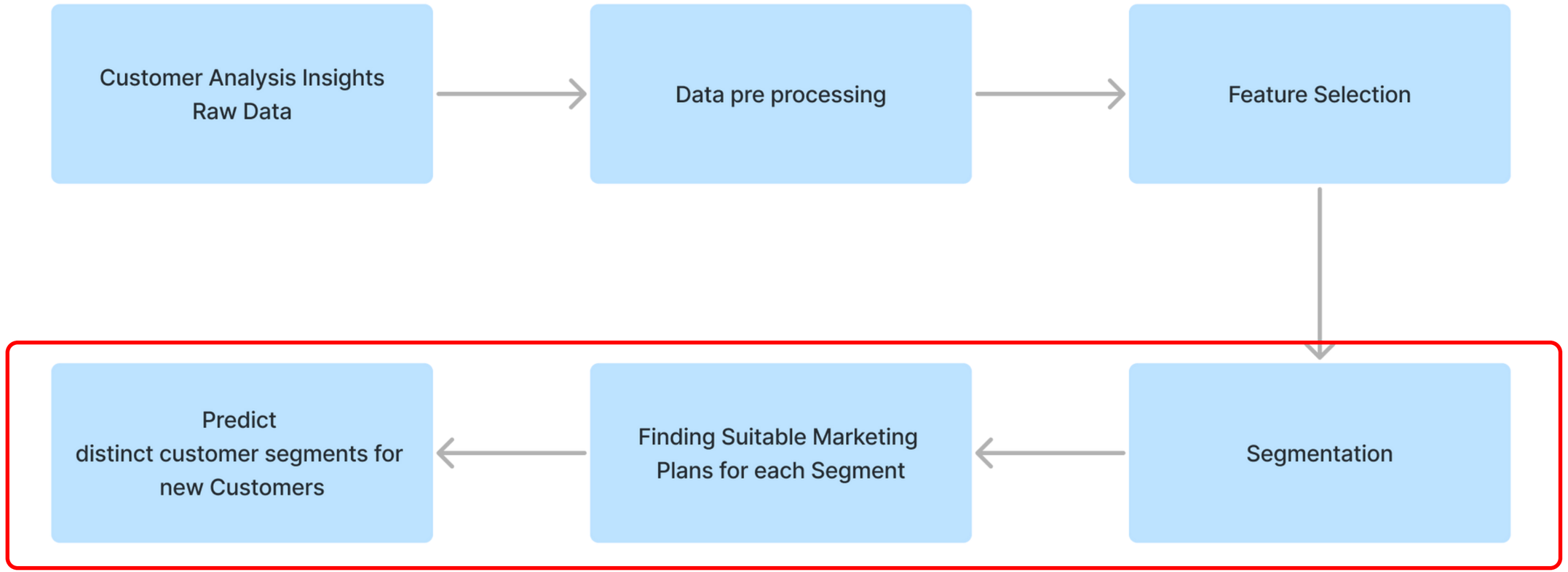
ABOUT THE DATASET

- Customer Personality Analysis is a detailed analysis of a company's ideal customers.
- Dataset has 2240 data points and 29 attributes
- It can be categorized into,
  - Customer's information
  - Products
  - Place
  - Promotioin

RESEARCH IDEA

- Identifying how marketing strategies can be optimized to effectively target and engage distinct customer segments.
  - What clustering techniques can be employed to accurately identify distinct customer segments based on relevant demographic, income, ... factors?
  - Predicting the customer segment to which a new customer belongs to devise tailored marketing strategies.

RESEARCH PLAN

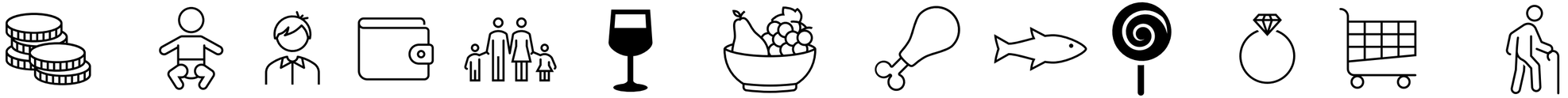


DATA ANALYSIS

CLUSTERING

CLUSTERING

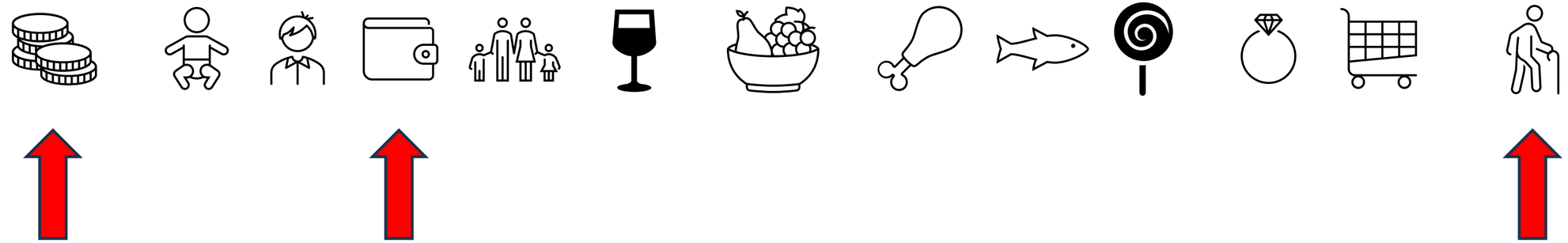
selected_attributes_01 =



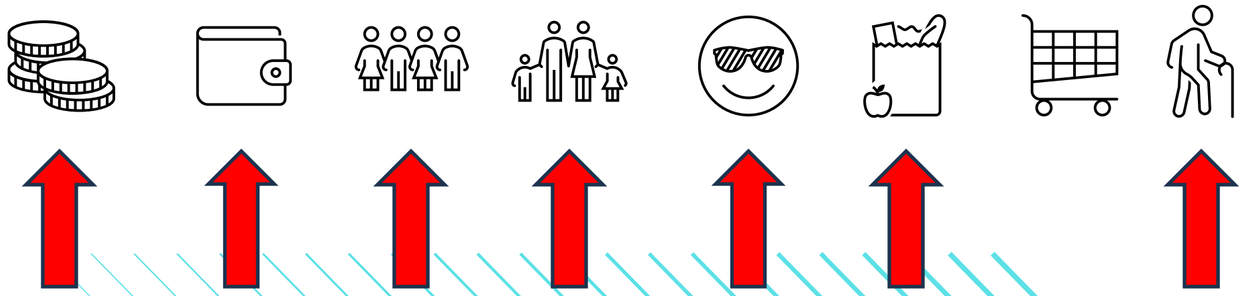
selected_attributes_02 =

CLUSTERING

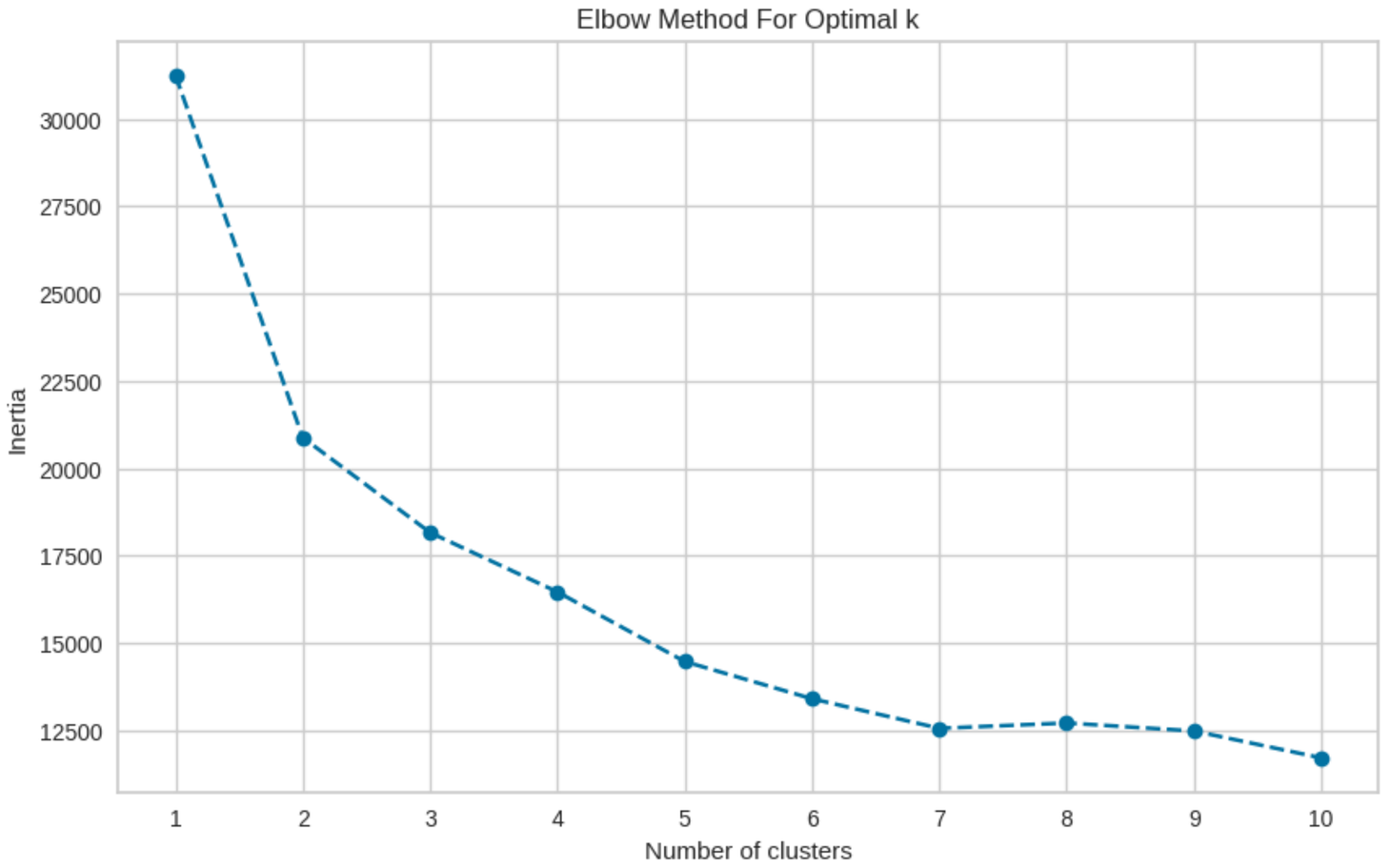
selected_attributes_01 =



selected_attributes_02 =



FINDINGS OF CLUSTERING



FINDINGS OF CLUSTERING

If we considered 2 cluster groups

Cluster	0	1
ID	5577.020462	5612.425384
Income	41813.030363	73330.138075
Kidhome	0.635644	0.041841
Teenhome	0.617162	0.273361
Recency	48.796700	49.909344
MntWines	158.572937	611.486750
MntFruits	8.530693	63.673640
MntMeatProducts	53.704290	406.259414
MntFishProducts	12.003300	91.111576
MntSweetProducts	8.897690	65.557880
MntGoldProds	26.803960	79.841004
NumDealsPurchases	2.664026	1.605300
NumWebPurchases	3.387459	5.560669
NumCatalogPurchases	1.287789	5.560669
NumStorePurchases	4.537294	8.451883
NumWebVisitsMonth	6.234323	3.387727
AcceptedCmp3	0.069307	0.080893
AcceptedCmp4	0.051485	0.124128
AcceptedCmp5	0.003300	0.217573
AcceptedCmp1	0.013201	0.171548

Cluster	0	1
AcceptedCmp2	0.005281	0.030683
Complain	0.009901	0.006974
Response	0.099670	0.252441
Total_Purchases	11.876568	21.178522
Education_2n Cycle	0.087789	0.094840
Education_Basic	0.034323	0.002789
Education_Graduation	0.485809	0.542538
Education_Master	0.174257	0.146444
Education_PhD	0.217822	0.213389
days_a_customer	501.555116	534.221757
Age	50.631023	52.125523
age_group	2.675248	2.615063
adults_in_home	1.669307	1.596932
Spent	268.512871	1317.930265
Living_With	0.669307	0.596932
Children	1.252805	0.315202
leisure_purchases	194.274587	756.885635
essential_purchases	74.238284	561.044630
Family_Size	2.922112	1.912134
Is_Parent	0.910231	0.304045
IncomePer	15306.902734	44909.978847
Total_Promos	0.142574	0.624826

FINDINGS OF CLUSTERING

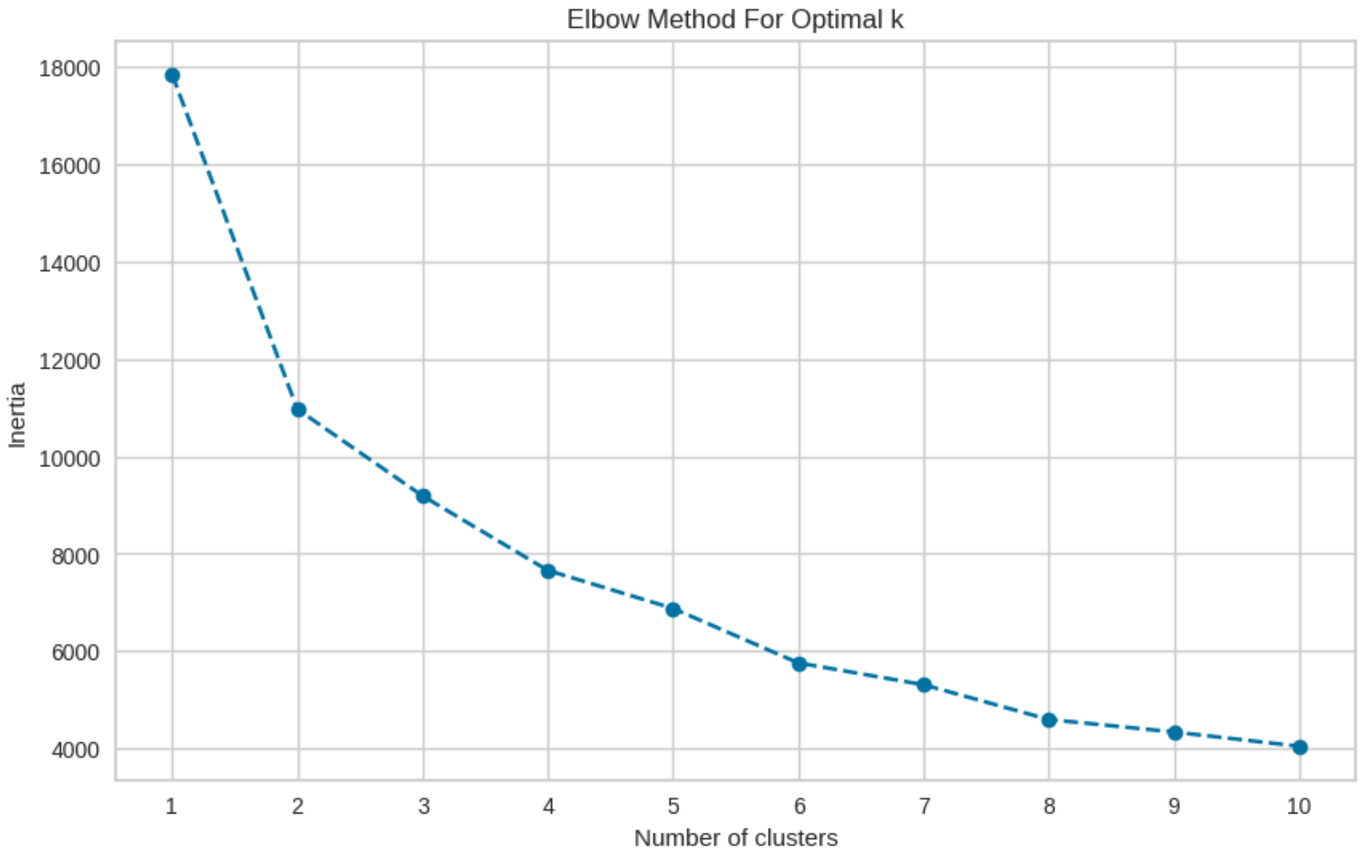
Cluster 0 Characteristics:

- Low income levels
- Larger family sizes
- Higher likelihood of parenthood
- Decreased responsiveness to promotions overall
- Reduced expenditure patterns
- Notable response surge in round 3 of campaigns, tapering off after round 4
- Prefer physical stores for shopping
- Tendency for seeking deals, indicating potential for targeted discount marketing
- Roughly purchases 1/7.5 ratio of products compared to Cluster 1
- Ratio for wine and gold purchases is approximately 1/3.5
- No substantial variance in education levels compared to

Cluster 1 Characteristics:

- High income bracket
- Smaller household sizes
- Decreased likelihood of parenthood
- Generally responsive to promotional initiatives
- Preference for in-store shopping experiences
- Actively engages with catalog and online shopping channels
- Balanced utilization of various shopping platforms, indicating potential for targeted marketing across multiple channels
- Roughly purchases 7.5 times more products compared to Cluster 0
- Ratio for wine and gold purchases is approximately 3.5 times more compared to Cluster 0
- No substantial variance in education levels compared to Cluster 0

FINDINGS OF CLUSTERING



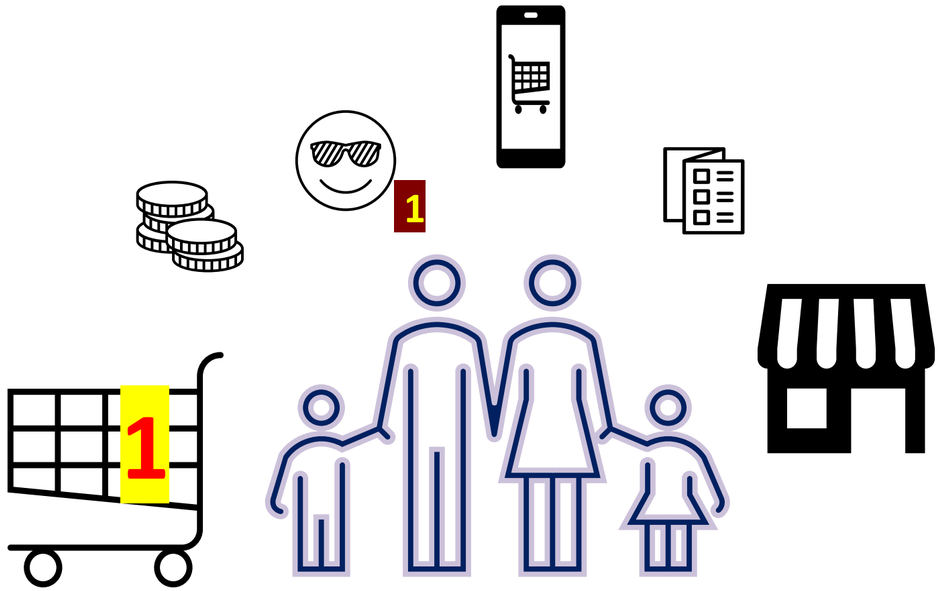
FINDINGS OF CLUSTERING

If we considered 4 cluster groups

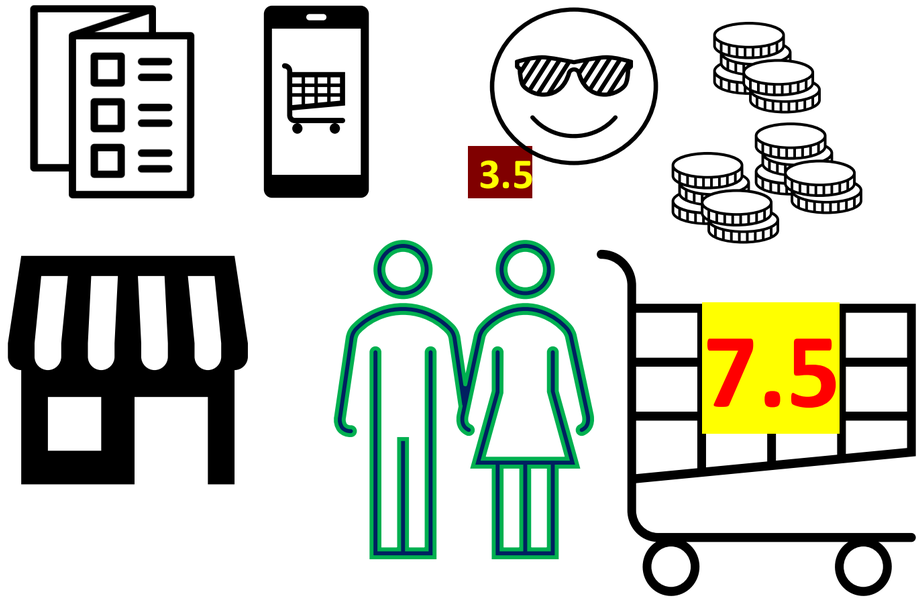
Cluster	0	1	2	3
ID	5596.819188	5601.920354	5752.134855	5398.936082
Income	35758.797048	43234.714602	76689.091286	62569.721649
Kidhome	0.555966	1.046460	0.004149	0.136082
Teenhome	0.263223	1.048673	0.024896	0.888660
Recency	48.168512	48.953540	49.497925	50.651546
MntWines	63.964330	112.453540	630.838174	560.367010
MntFruits	8.195572	5.632743	64.919087	37.274227
MntMeatProducts	33.782288	40.457965	485.265560	191.752577
MntFishProducts	12.297663	7.913717	95.927386	48.865979
MntSweetProducts	8.509225	6.356195	66.794606	38.142268
MntGoldProds	21.209102	21.378319	74.593361	72.152577
NumDealsPurchases	1.755228	3.511062	1.039419	3.447423
NumWebPurchases	2.638376	3.070796	5.016598	6.531959
NumCatalogPurchases	0.772448	0.951327	5.964730	4.134021
NumStorePurchases	3.698647	4.017699	8.313278	8.461856
NumWebVisitsMonth	6.269373	6.347345	2.643154	5.430928
AcceptedCmp3	0.065191	0.064159	0.082988	0.084536
AcceptedCmp4	0.022140	0.026549	0.134855	0.148454
AcceptedCmp5	0.003690	0.000000	0.284232	0.043299
AcceptedCmp1	0.006150	0.011062	0.226141	0.049485
AcceptedCmp2	0.001230	0.004425	0.033195	0.022680
Complain	0.007380	0.013274	0.004149	0.012371

Cluster	0	1	2	3
Response	0.091021	0.106195	0.317427	0.117526
Total_Purchases	8.864699	11.550885	20.334025	22.575258
Education_2n Cycle	0.113161	0.077434	0.085062	0.068041
Education_Basic	0.062731	0.004425	0.000000	0.002062
Education_Graduation	0.484625	0.477876	0.526971	0.538144
Education_Master	0.170972	0.188053	0.153527	0.146392
Education_PhD	0.168512	0.252212	0.234440	0.245361
days_a_customer	493.586716	484.475664	507.931535	572.785567
Age	46.712177	54.577434	51.634855	54.734021
age_group	2.404674	2.964602	2.489627	2.954639
adults_in_home	1.655597	1.670354	1.564315	1.688660
Spent	147.958180	194.192478	1418.338174	948.554639
Living_With	0.655597	0.670354	0.564315	0.688660
Children	0.819188	2.095133	0.029046	1.024742
leisure_purchases	93.682657	140.188053	772.226141	670.661856
essential_purchases	54.275523	54.004425	646.112033	277.892784
Family_Size	2.474785	3.765487	1.593361	2.713402
Is_Parent	0.815498	1.000000	0.029046	0.964948
IncomePer	15479.688704	11706.336007	53845.089730	23836.759656
Total_Promos	0.098401	0.106195	0.761411	0.348454

Customer Personas



A



B

RFM ANALYSIS

RFM Score = R quantile + F quantile + M quantile

Utilizing RFM to Segment Customers

- Recency: Measure the time since the last purchase
- Frequency: Evaluate the overall interaction frequency (combining various purchase channels)
- Monetary Value: Assess the total expenditure of each customer

Challenges in Frequency Calculation

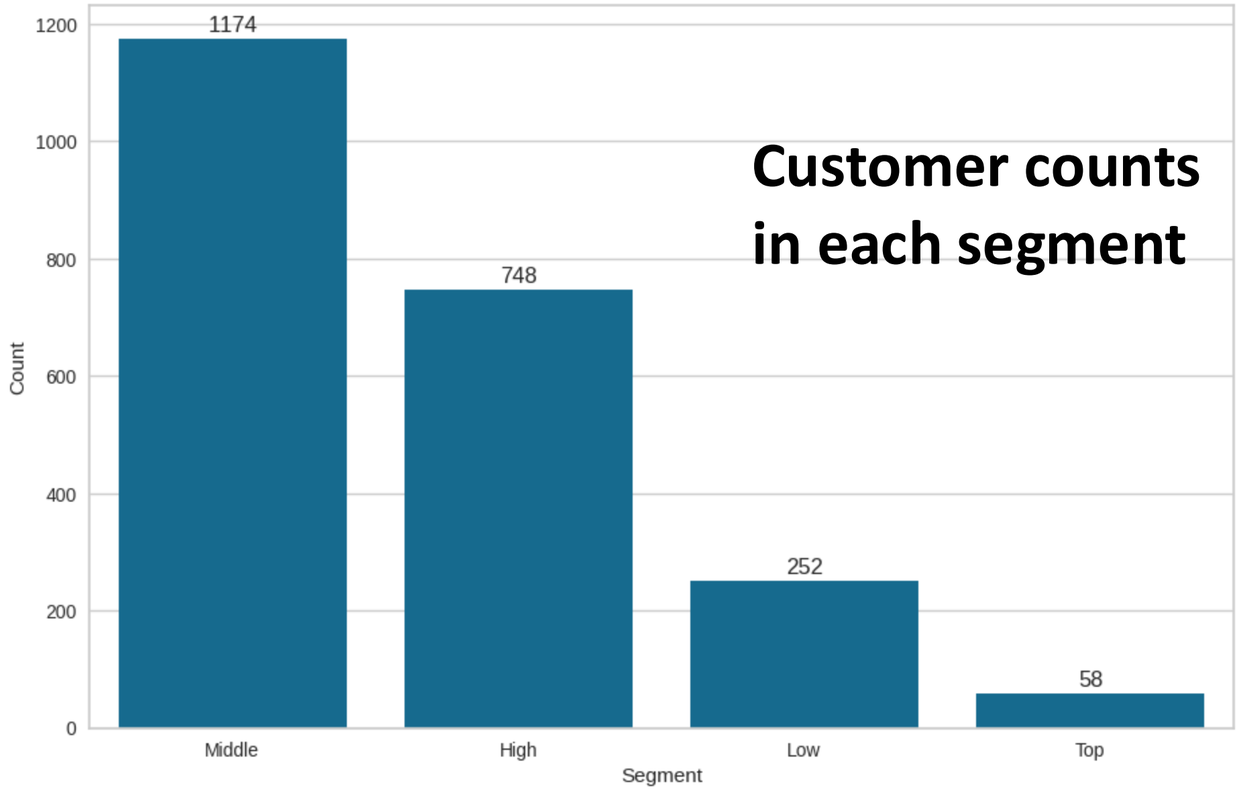
- Absence of direct time-wise purchase logging: All_purchases / days_a_customer (a proxy for frequency)
- Leveraging available data: NumWebPurchases, NumCatalogPurchases, NumStorePurchases

RFM ANALYSIS

Objectives

- Identify high-value customers for targeted marketing efforts
- Improve customer retention strategies by focusing on medium scoring customers
- Optimize resource allocation by understanding customer preferences and behaviors

FINDINGS OF RFM ANALYSIS



INSIGHTS

- **Top** : Strategies to retain and further engage these customers could include loyalty programs, personalized offers, and VIP treatment.

- **Middle** and **High**: segments are sensitive to segment change!

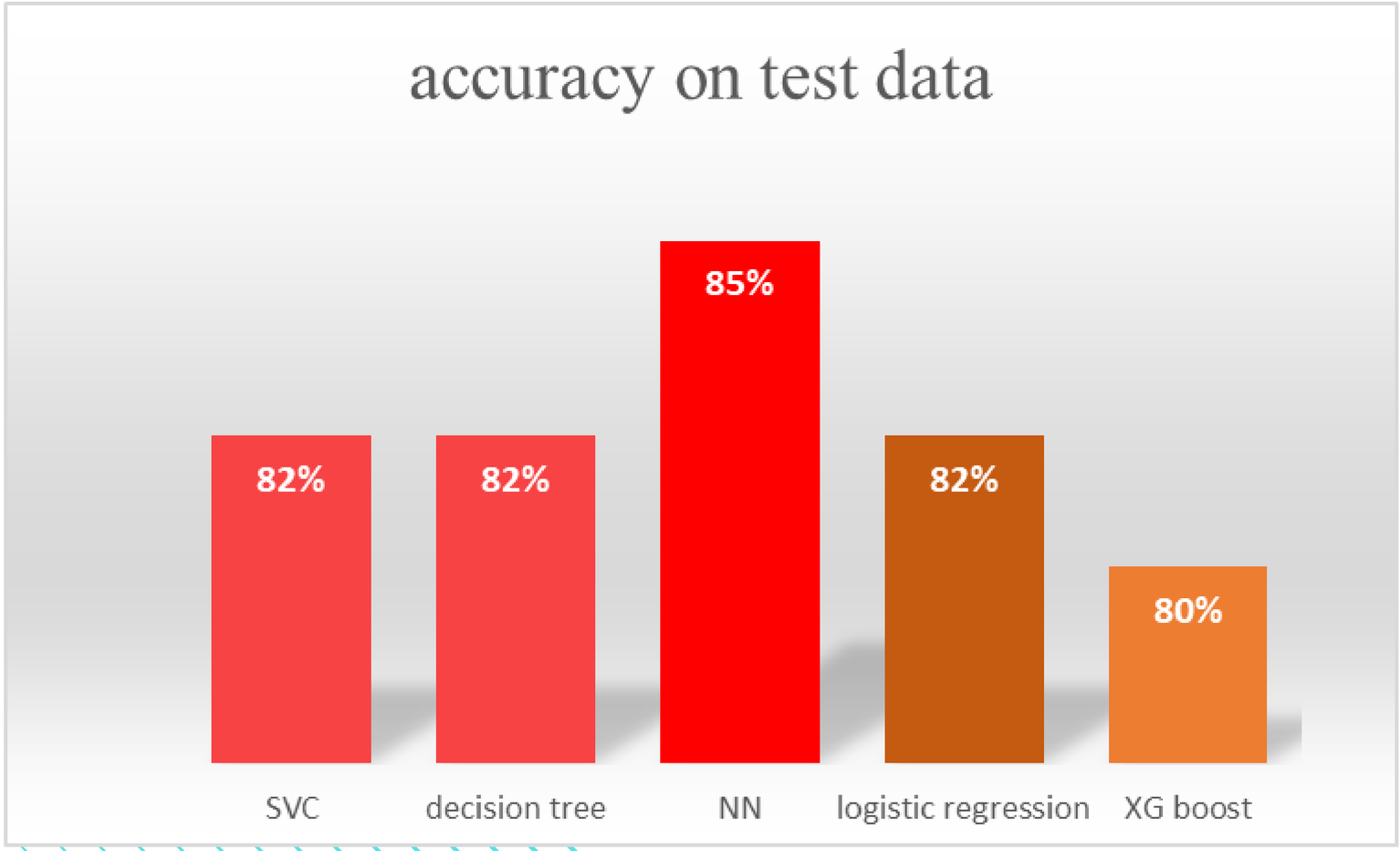
- **Low**: Strategies to re-engage these customers could involve targeted reactivation campaigns, special offers to incentivize purchases, and improving overall customer experience to encourage repeat business.

PREDICTION MODEL

- Prediction of total promotions

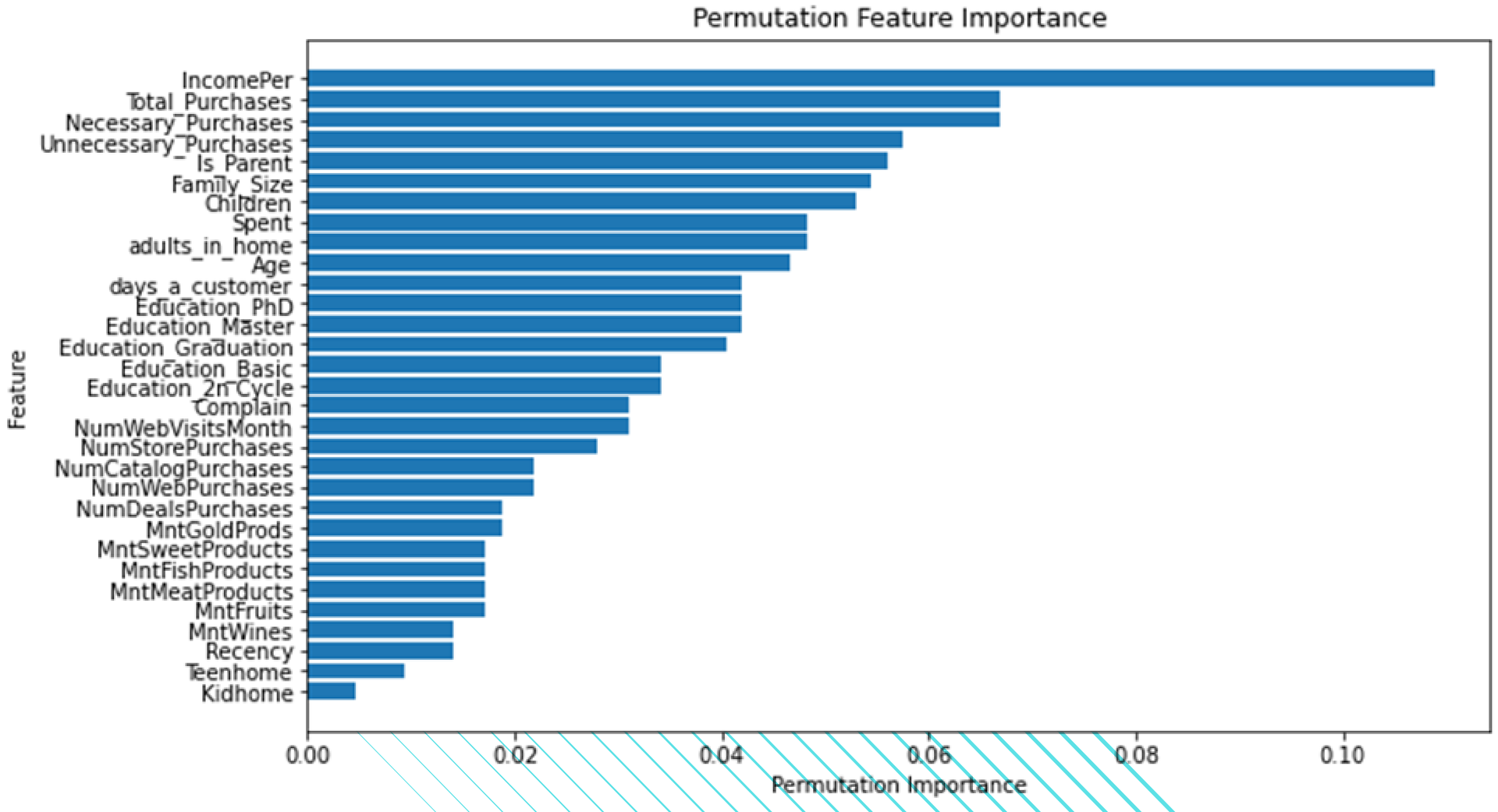
Attributes:
Kidhome, Teenhome, Recency, MntWines, MntFruits, 'MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds, NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth, Complain, Education_2n Cycle, Education_Basic, Education_Graduation, Education_Master, Education_PhD, days_a_customer, Age, adults_in_home, Spent, Children, Family_Size, Is_Parent, Unnecessary_Purchases, Necessary_Purchases, Total_Purchases, IncomePer

Target: accept promotion or not, 0 or 1
Balancing method: SMOTE



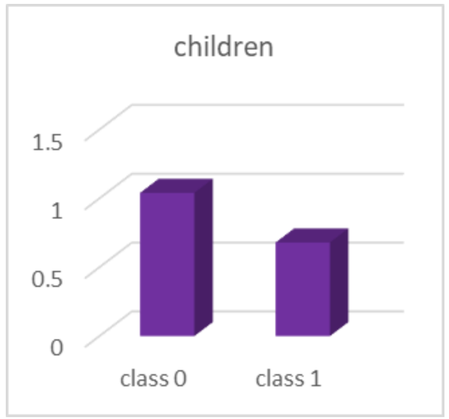
PREDICTION ANALYSIS

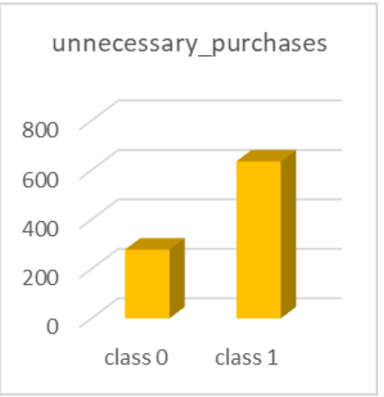
- Feature Importance

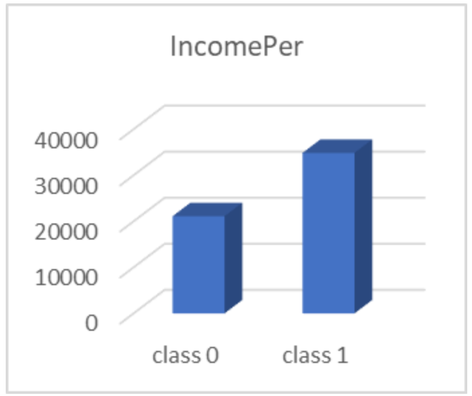


FINDINGS

- Important Features comparison





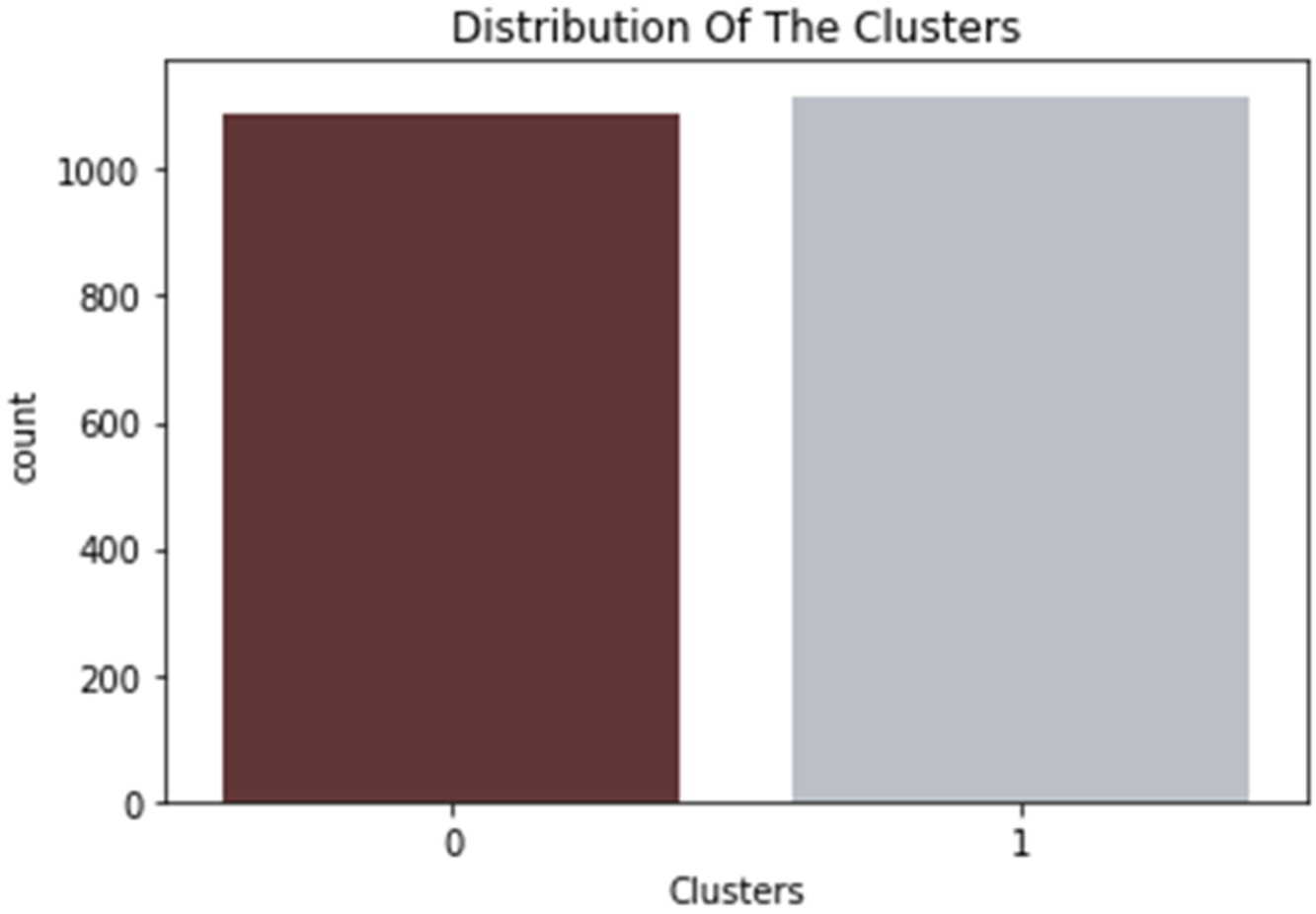


PREDICTION MODEL

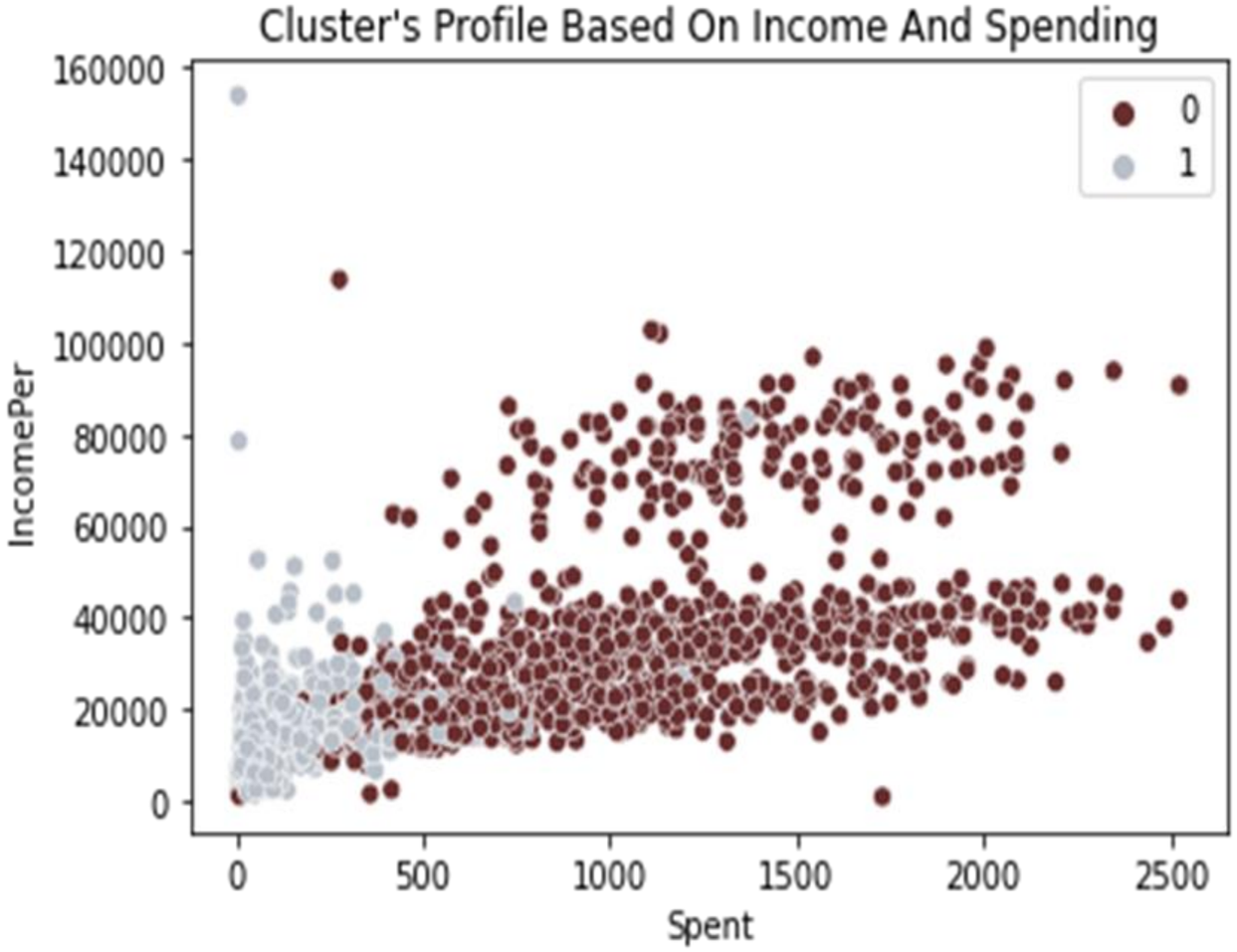
- Prediction of Customer Cluster

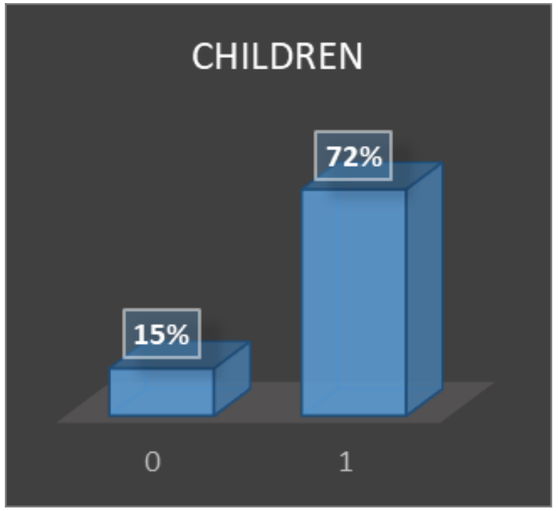
selected_attributes = NumDealsPurchases, NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth, Complain,Spent,Unnecessary_Purchases, Necessary_Purchases, Total_Purchases

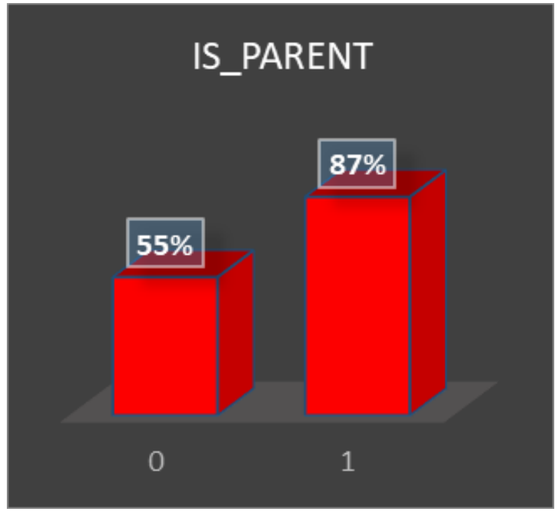
method: Agglomerative clustering
Number of clusters: 2, using Elbow method

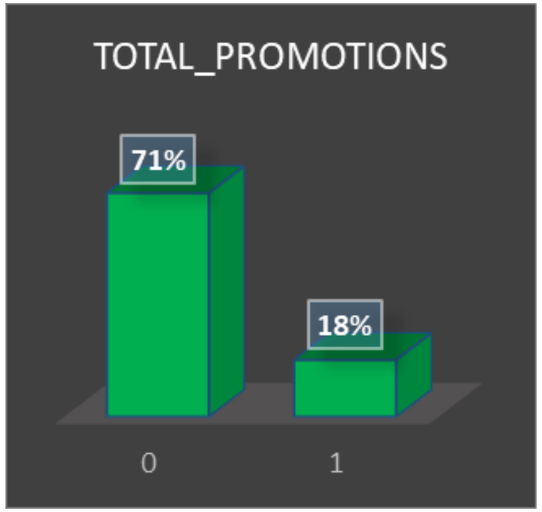


FINDINGS of Clustering









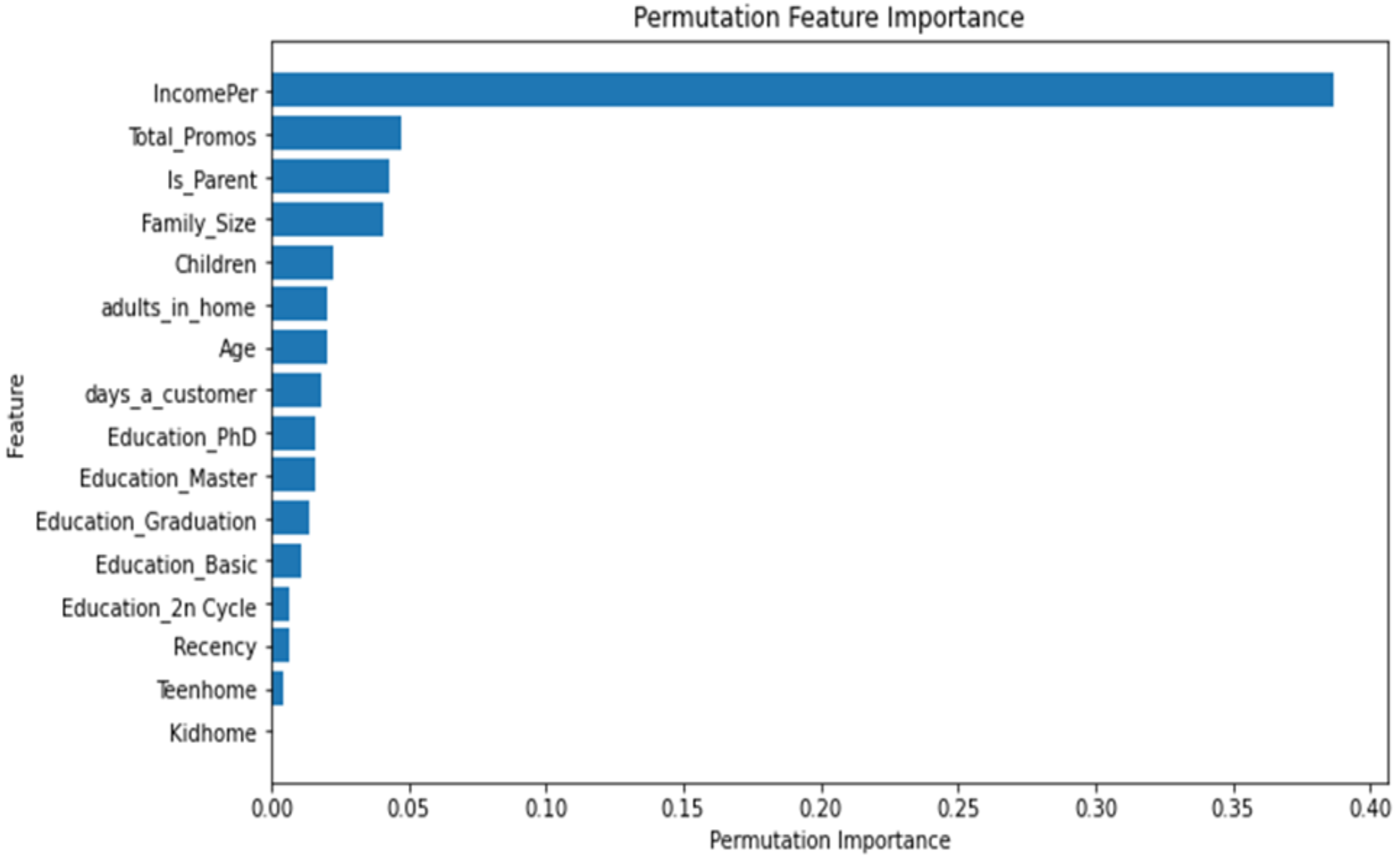
PREDICTION ANALYSIS

Attributes: Kidhome, Teenhome, Recency, Education_2n Cycle, Education_Basic, Education_Graduation, Education_Master, Education_PhD, days_a_customer, Age, adults_in_home, Children, Family_Size, Is_Parent, Total_Promos, IncomePer

Target: cluster number

Model: Neural Network

Accuracy: 89%



Insights

- Following the necessary and unnecessary purchases of a customer can help in the prediction of their behaviour in accepting campaigns or not
- Having the purchase behaviour of customers, which is available for stores, is sufficient for finding the cluster they belong to, which helps in having an idea about their income and if they will accept promotions or not
- Moreover, prediction of the cluster helps in predicting if the customer is parent, the probability if they have kids or not and also their spending on wine, sweet and gold products that has a significant difference between two clusters which can help in targeted marketing and making a recommender system
- Having weekly purchases of the customers in the two clusters, which was lacking in this dataset, can help in finding the cluster of the new customer and prediction of some demographics about them which helps in having a targeted marketing for new members in a short time.

INDIVIDUAL CONTRIBUTIONS

• **Arash Nedaei Janbesaraei**	Data Preprocessing, Analysis, Peer Review
• **Zohreh Yousefi Dahka**	Data Preprocessing, Research Plan, Analysis,
• **Kavinda Kulasinghe**	Data Preprocessing, Literature Review, Presentation Preparation
• **Taufiq Ahmed**	Data Preprocessing, Peer Review, Analysis,
• **Md Mobusshar Islam**	Data Preprocessing, Literature Review, Presentation Preparation

THANK YOU