Rakamin Academy

Predict Customer Personality to boost marketing campaign by using Machine Learning

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

Created by:
Saila Fikriyya
sailafs@gmail.com
linkedin.com/in/sailafs1203/

I'm a health psychology graduate turned data scientist, bridging the gap between human behavior and data-driven solutions. With a Master's in Health Psychology from the University of Stirling and ongoing training in data science, I combine expertise in research methods, data analysis, and understanding human behavior. My experience spans from studying post-traumatic growth in cancer survivors to applying data science techniques in digital healthcare and career development. I'm passionate about leveraging interdisciplinary skills to create evidence-based strategies that improve health outcomes and overall well-being, always prioritizing a human-centric approach in collaborative environments.

Overview

“A company can grow rapidly when it understands its customer personality, allowing it to provide better services and benefits to customers who have the potential to become loyal customers. By analysing historical marketing campaign data, it helps to increase performance and target the right customers to make transactions on the company's platform, the focus from these data insights is to create a cluster prediction model to facilitate decision-making for the company."

Conversion Rate Analysis Based on Income, Spending and Age

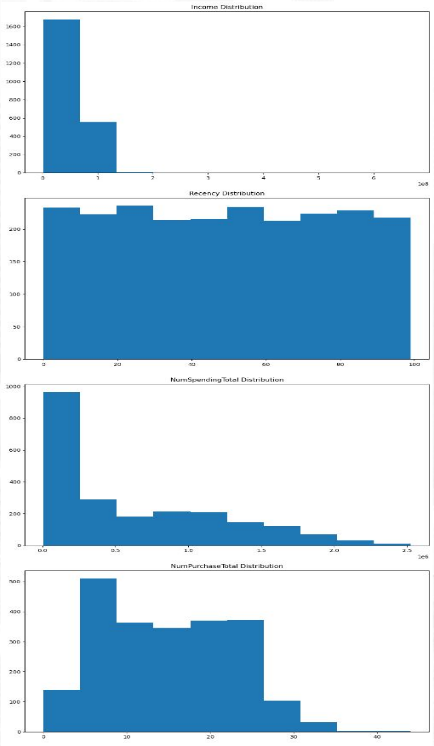
Feature engineering has been performed by calculating the conversion rate defined as (#response / #visit). Not only the conversion rate, but also other representative features such as,

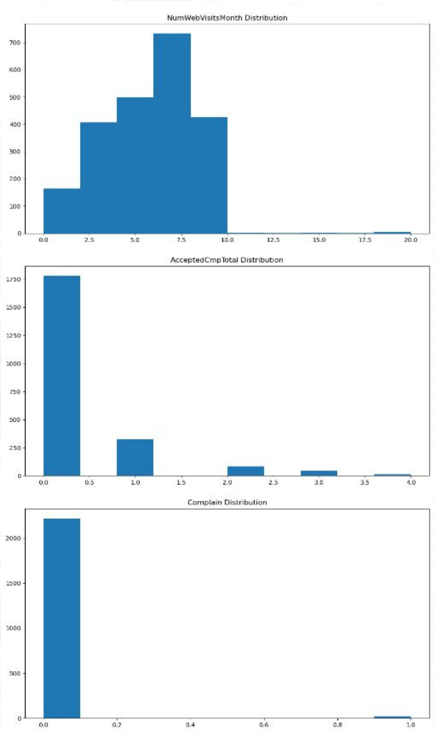
- Creating a new feature 'Age' from 'Year_Birth'
- Creating a new feature for age grouping
- Creating a new feature to sum up accepted campaigns
- Creating a new feature to sum up the number of children
- Creating a new feature to sum up total transactions
- Creating a new feature ‘Spending’ from some of products bought

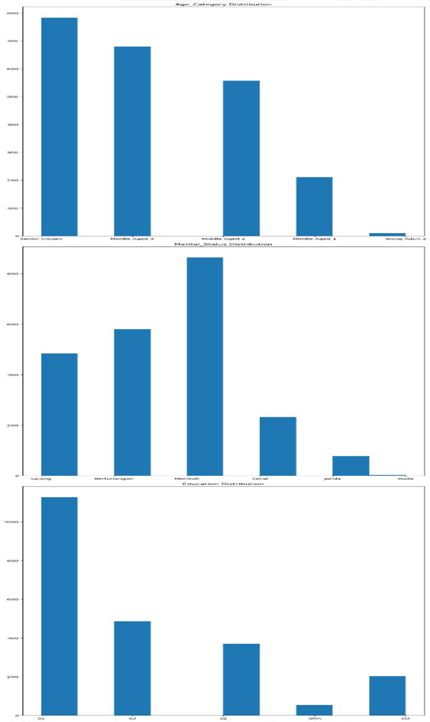
For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Univariate Analysis → most of the features are positively skewed except recency and total number of purchase which looked relatively uniform.



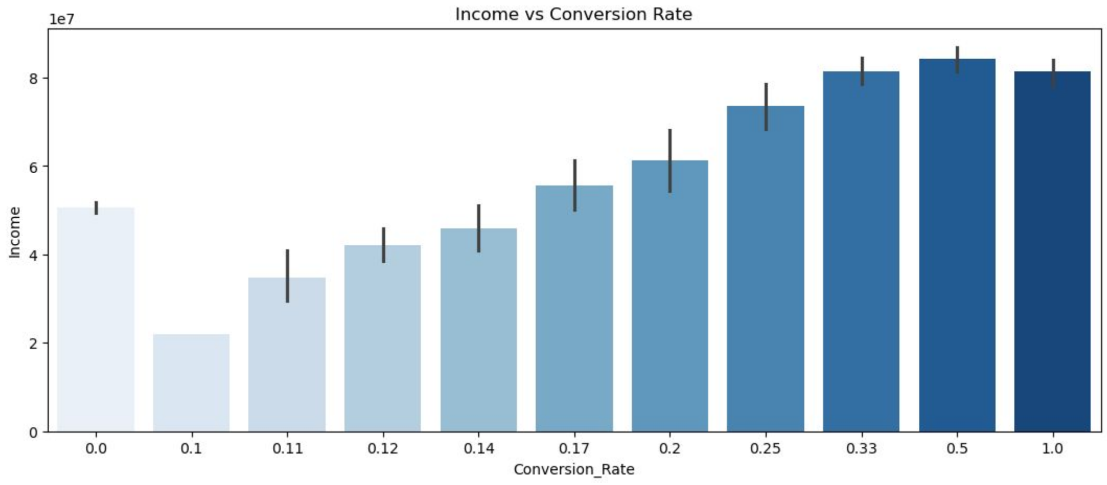




For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis (Income and Conversion Rate)



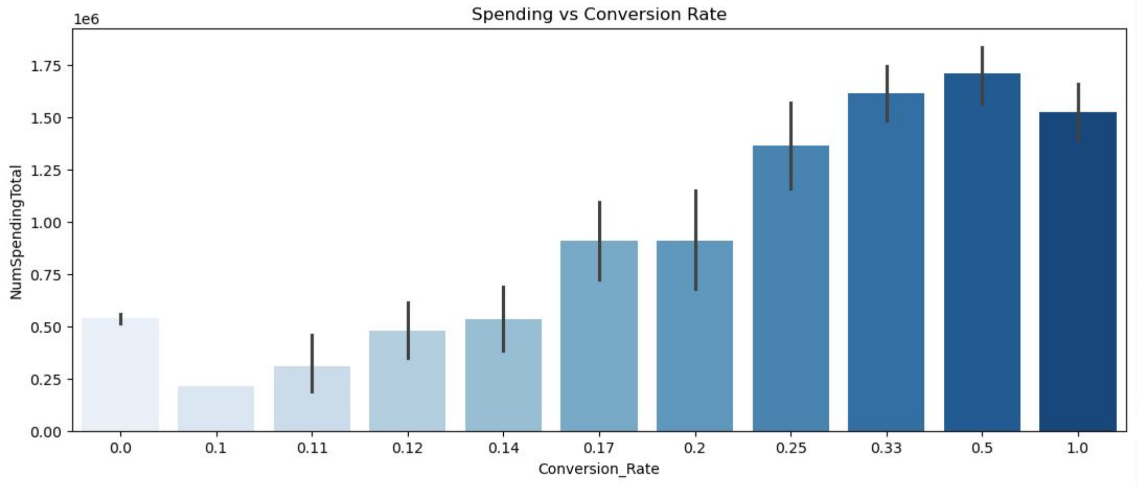
There's a slight positive trend, with income generally increasing as conversion rate increases. This implies that the higher the income, higher conversion rate is observed.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

Rakamin Academy

EDA: Bivariate Analysis (Spending and Conversion Rate)

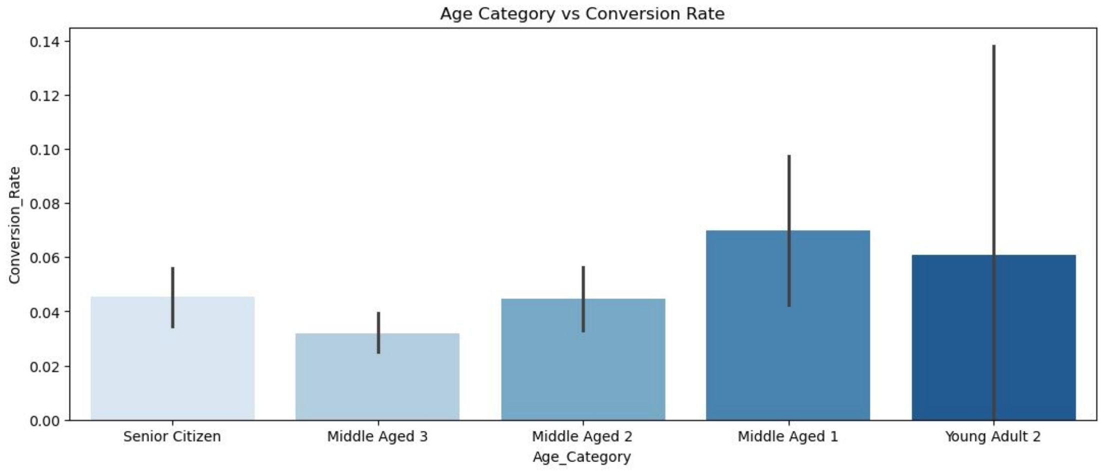


As the conversion rate increases from 0 to 1.0, there's a general trend of higher total spending, with the highest spending levels observed at conversion rates between 0.33 and 1.0. The relationship suggests that customers who are more likely to convert (make a purchase) tend to spend more overall, though there's some variability in spending amounts within each conversion rate category.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis (Age and Conversion Rate)

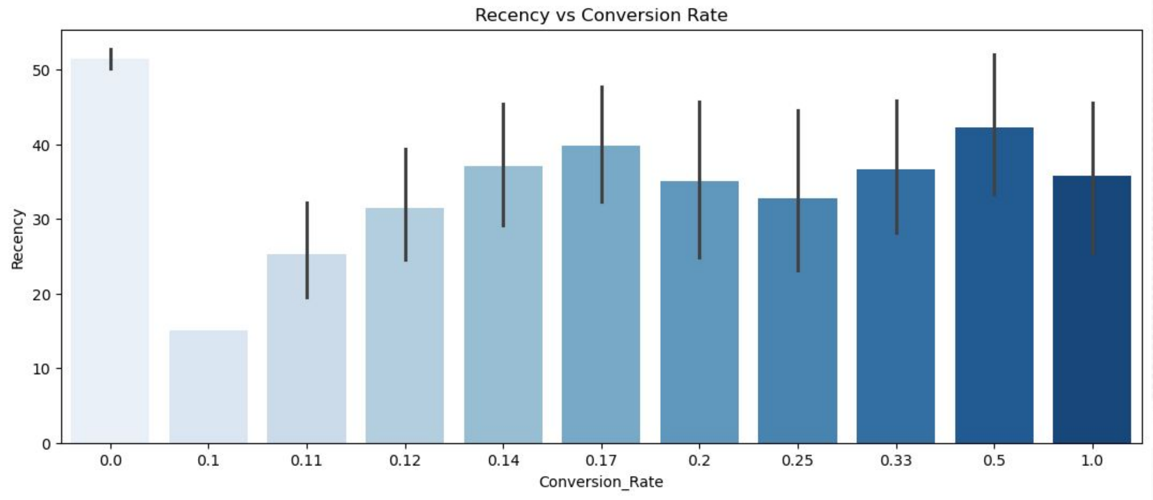


The middle-aged groups, particularly Middle Aged 1 (31-39), have the highest conversion rates, while Senior Citizen (above 60) and Middle Aged 3 (50-59) have the lowest. The younger customers are, the higher the conversion rate.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis

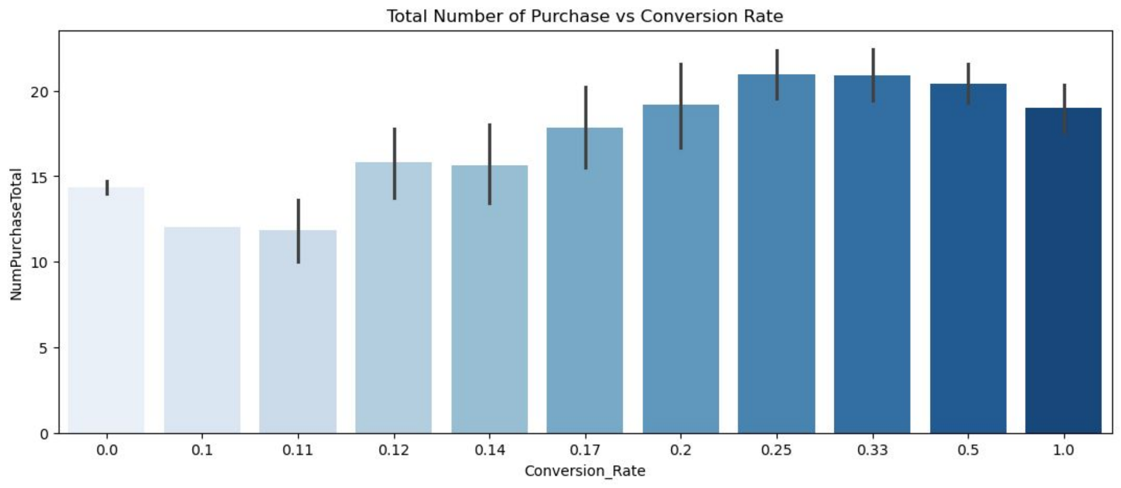


In terms of recency, trend fluctuated across conversion rate. The sweet spot for higher conversion rates seems to be customers with moderate Recency values. This could imply that nurturing relationships over time, rather than focusing solely on the most recent or oldest customers, might be beneficial for improving conversion rates.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis

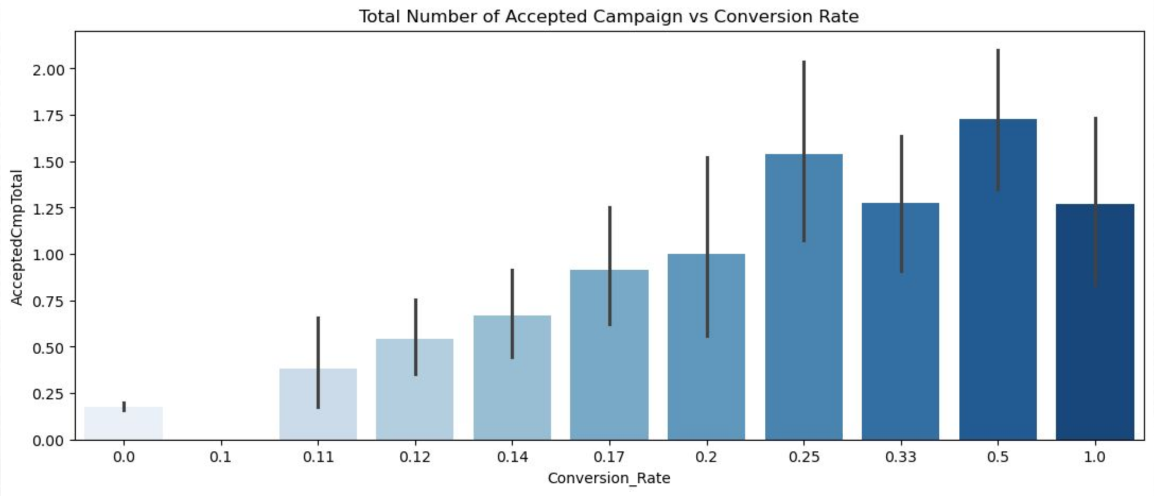


In terms of total purchase, a plateau or slightly decrease for very high conversion rates (0.33 to 1.0) is observed despite steadily increasing. This could suggest that beyond a certain point, higher conversion rates don't necessarily lead to more total purchases. Maximum purchase is between 20-21 and minimum purchase is about 12.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis



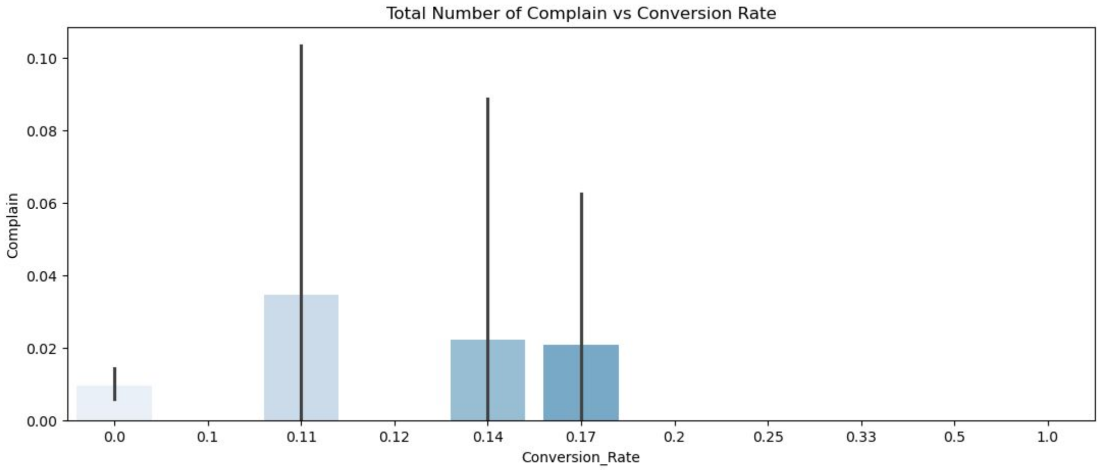
Similar to total purchase, total campaign increased steadily and fluctuate from 0.25 to 1. However, it can be concluded that higher accepted campaigns associated with higher conversion rate.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

Rakamin Academy

EDA: Bivariate Analysis



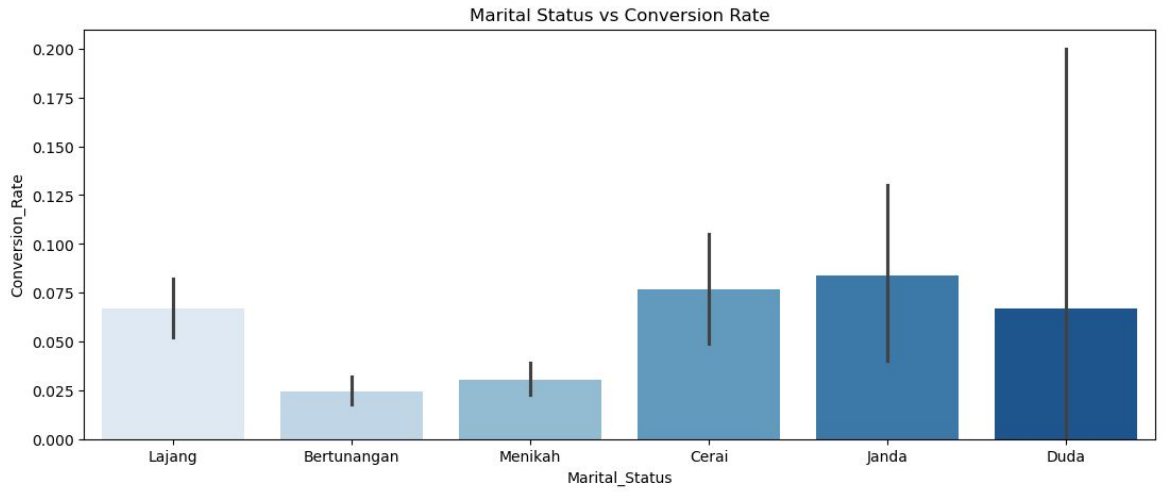
There is a slight negative correlation between complain and conversion rate, suggesting that campaigns with higher conversion rates tend to have fewer complaints.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

Rakamin Academy

EDA: Bivariate Analysis

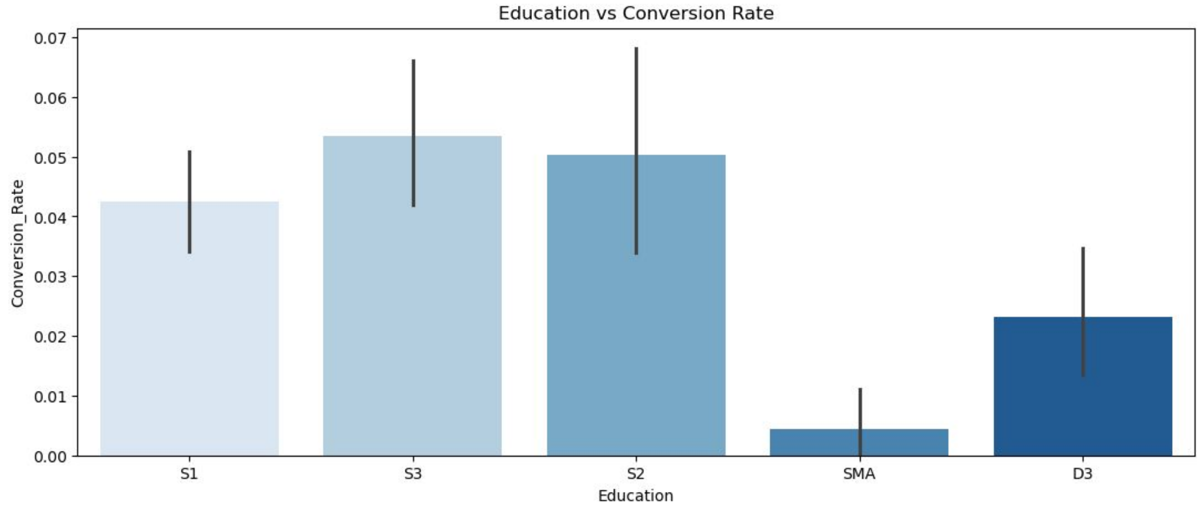


In terms of relationship between marital status and conversion rate. The highest conversion rates are observed among those who are married and widowed, while the lowest rates are seen among the single and divorced individuals. However, there is significant overlap among the groups, suggesting that marital status alone may not be the sole determinant of conversion rate.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Bivariate Analysis

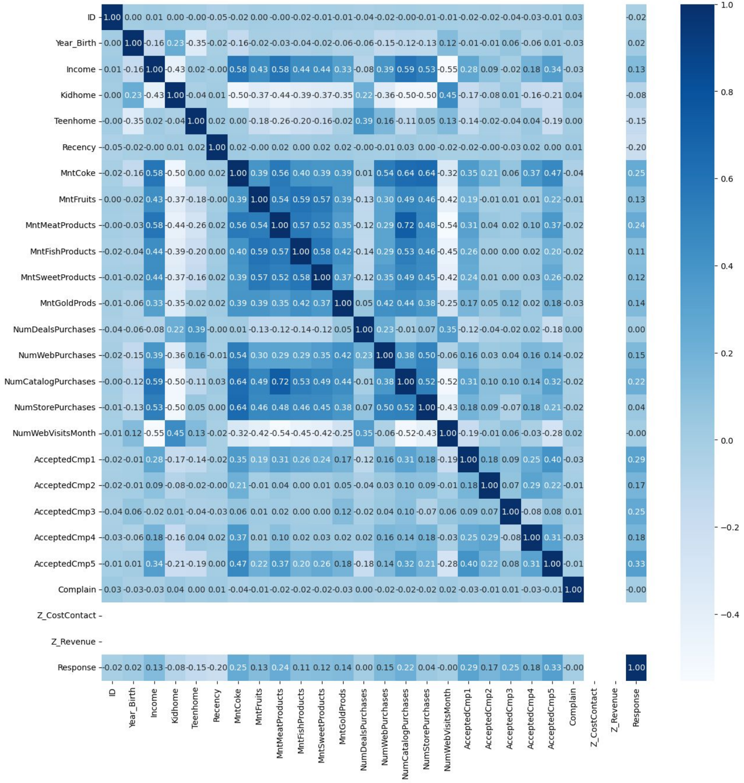


In terms of relationship between education level and conversion rate. The highest conversion rates are observed among individuals with S1 and S3 education levels, while the lowest rates are seen among those with SMA and D3 education levels. Alas, those who are having tertiary education are more likely to convert.

For more information, see jupyter notebook [here]

Conversion Rate Analysis Based on Income, Spending and Age

EDA: Multivariate Analysis



	ID	Year_Birth	Income	Kidhome	Teenhome	Recency	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	AcceptedCmp1	AcceptedCmp2	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	Complain	Z_CostContact	Z_Revenue	Response
ID	1.00	0.00	0.01	0.00	-0.00	-0.05	-0.02	0.00	-0.00	-0.02	-0.01	-0.01	-0.04	-0.02	-0.00	-0.01	-0.01	-0.02	-0.02	-0.04	-0.03	-0.01	0.03			-0.02
Year_Birth	0.00	1.00	-0.16	0.23	-0.35	-0.02	-0.16	-0.02	-0.03	-0.04	-0.02	-0.06	-0.06	-0.15	-0.12	-0.13	0.12	-0.01	-0.01	0.06	-0.06	0.01	-0.03			0.02
Income	0.01	-0.16	1.00	-0.43	0.02	-0.00	0.58	0.43	0.58	0.44	0.44	0.33	-0.08	0.39	0.59	0.53	-0.55	0.28	0.09	-0.02	0.18	0.34	-0.03			0.13
Kidhome	0.00	0.23	-0.43	1.00	-0.04	0.01	-0.50	-0.37	-0.44	-0.39	-0.37	-0.35	0.22	-0.36	-0.50	-0.50	0.45	-0.17	-0.08	0.01	-0.16	-0.21	0.04			-0.08
Teenhome	-0.00	-0.35	0.02	-0.04	1.00	0.02	0.00	-0.18	-0.26	-0.20	-0.16	-0.02	0.39	0.16	-0.11	0.05	0.13	-0.14	-0.02	-0.04	0.04	-0.19	0.00			-0.15
Recency	-0.05	-0.02	-0.00	0.01	0.02	1.00	0.02	-0.00	0.02	0.00	0.02	0.02	-0.00	-0.01	0.03	0.00	-0.02	-0.02	-0.00	-0.03	0.02	0.00	0.01			-0.20
MntCoke	-0.02	-0.16	0.58	-0.50	0.00	0.02	1.00	0.39	0.56	0.40	0.39	0.39	0.01	0.54	0.64	0.64	-0.32	0.35	0.21	0.06	0.37	0.47	-0.04			0.25
MntFruits	0.00	-0.02	0.43	-0.37	-0.18	-0.00	0.39	1.00	0.54	0.59	0.57	0.39	-0.13	0.30	0.49	0.46	-0.42	0.19	-0.01	0.01	0.01	0.22	-0.01			0.13
MntMeatProducts	-0.00	-0.03	0.58	-0.44	-0.26	0.02	0.56	0.54	1.00	0.57	0.52	0.35	-0.12	0.29	0.72	0.48	-0.54	0.31	0.04	0.02	0.10	0.37	-0.02			0.24
MntFishProducts	-0.02	-0.04	0.44	-0.39	-0.20	0.00	0.40	0.59	0.57	1.00	0.58	0.42	-0.14	0.29	0.53	0.46	-0.45	0.26	0.00	0.00	0.02	0.20	-0.02			0.11
MntSweetProducts	-0.01	-0.02	0.44	-0.37	-0.16	0.02	0.39	0.57	0.52	0.58	1.00	0.37	-0.12	0.35	0.49	0.45	-0.42	0.24	0.01	0.00	0.03	0.26	-0.02			0.12
MntGoldProds	-0.01	-0.06	0.33	-0.35	-0.02	0.02	0.39	0.39	0.35	0.42	0.37	1.00	0.05	0.42	0.44	0.38	-0.25	0.17	0.05	0.12	0.02	0.18	-0.03			0.14
NumDealsPurchases	-0.04	-0.06	-0.08	0.22	0.39	-0.00	0.01	-0.13	-0.12	-0.14	-0.12	0.05	1.00	0.23	-0.01	0.07	0.35	-0.12	-0.04	-0.02	0.02	-0.18	0.00			0.00
NumWebPurchases	-0.02	-0.15	0.39	-0.36	0.16	-0.01	0.54	0.30	0.29	0.29	0.35	0.42	0.23	1.00	0.38	0.50	-0.06	0.16	0.03	0.04	0.16	0.14	-0.02			0.15
NumCatalogPurchases	-0.00	-0.12	0.59	-0.50	-0.11	0.03	0.64	0.49	0.72	0.53	0.49	0.44	-0.01	0.38	1.00	0.52	-0.52	0.31	0.10	0.10	0.14	0.32	-0.02			0.22
NumStorePurchases	-0.01	-0.13	0.53	-0.50	0.05	0.00	0.64	0.46	0.48	0.46	0.45	0.38	0.07	0.50	0.52	1.00	-0.43	0.18	0.09	-0.07	0.18	0.21	-0.02			0.04
NumWebVisitsMonth	-0.01	0.12	-0.55	0.45	0.13	-0.02	-0.32	-0.42	-0.54	-0.45	-0.42	-0.25	0.35	-0.06	-0.52	-0.43	1.00	-0.19	-0.01	0.06	-0.03	-0.28	0.02			-0.00
AcceptedCmp1	-0.02	-0.01	0.28	-0.17	-0.14	-0.02	0.35	0.19	0.31	0.26	0.24	0.17	-0.12	0.16	0.31	0.18	-0.19	1.00	0.18	0.09	0.25	0.40	-0.03			0.29
AcceptedCmp2	-0.02	-0.01	0.09	-0.08	-0.02	-0.00	0.21	-0.01	0.04	0.00	0.01	0.05	-0.04	0.03	0.10	0.09	-0.01	0.18	1.00	0.07	0.29	0.22	-0.01			0.17
AcceptedCmp3	-0.04	0.06	-0.02	0.01	-0.04	-0.03	0.06	0.01	0.02	0.00	0.00	0.12	-0.02	0.04	0.10	-0.07	0.06	0.09	0.07	1.00	-0.08	0.08	0.01			0.25
AcceptedCmp4	-0.03	-0.06	0.18	-0.16	0.04	0.02	0.37	0.01	0.10	0.02	0.03	0.02	0.02	0.16	0.14	0.18	-0.03	0.25	0.29	-0.08	1.00	0.31	-0.03			0.18
AcceptedCmp5	-0.01	0.01	0.34	-0.21	-0.19	0.00	0.47	0.22	0.37	0.20	0.26	0.18	-0.18	0.14	0.32	0.21	-0.28	0.40	0.22	0.08	0.31	1.00	-0.01			0.33
Complain	0.03	-0.03	-0.03	0.04	0.00	0.01	-0.04	-0.01	-0.02	-0.02	-0.02	-0.03	0.00	-0.02	-0.02	-0.02	0.02	-0.03	-0.01	0.01	-0.03	-0.01	1.00			-0.00
Z_CostContact																										
Z_Revenue																										
Response	-0.02	0.02	0.13	-0.08	-0.15	-0.20	0.25	0.13	0.24	0.11	0.12	0.14	0.00	0.15	0.22	0.04	-0.00	0.29	0.17	0.25	0.18	0.33	-0.00			1.00

Strong positive correlations are observed among different product spending categories (e.g., MntCoke, MntFruits, MntMeatProducts), suggesting that customers who spend more in one category tend to spend more in others as well. There are also notable correlations between income and various spending categories, and between different types of purchases (web, catalog, store), indicating consistent purchasing behavior across channels.

For more information, see jupyter notebook here

Conversion Rate Analysis Based on Income, Spending and Age

Insights

- In terms of income, higher income levels are associated with higher conversion rates, indicating that wealthier customers are more likely to make purchases.
- In terms of spending, higher conversion rates are associated with increased total spending, indicating that customers who convert more frequently are also likely to be higher-value customers in terms of overall expenditure.
- In terms of age, younger to middle-aged adults tend to have higher conversion rates compared to older age groups, suggesting that marketing strategies may need to be tailored differently for various age demographics to maximise conversions.
- The heatmap reveals strong interconnections between different types of customer spending, with positive correlations across product categories and purchase channels, suggesting that high-value customers tend to engage more across all aspects of the business.

For more information, see jupyter notebook here

Data Cleaning & Preprocessing

- In terms of cleaning data, checking on ***missing values*** and ***duplicate*** rows are done
  - Before executing EDA, some missing values had been addressed as such 'Year_Birth' and 'Income'.
  - After thorough checking, 'Conversion_Rate' seemed to have 11 missing values as the calculation was supposed to be 0. Hence, filling in missing values with 0 is done.
  - No duplicates detected.

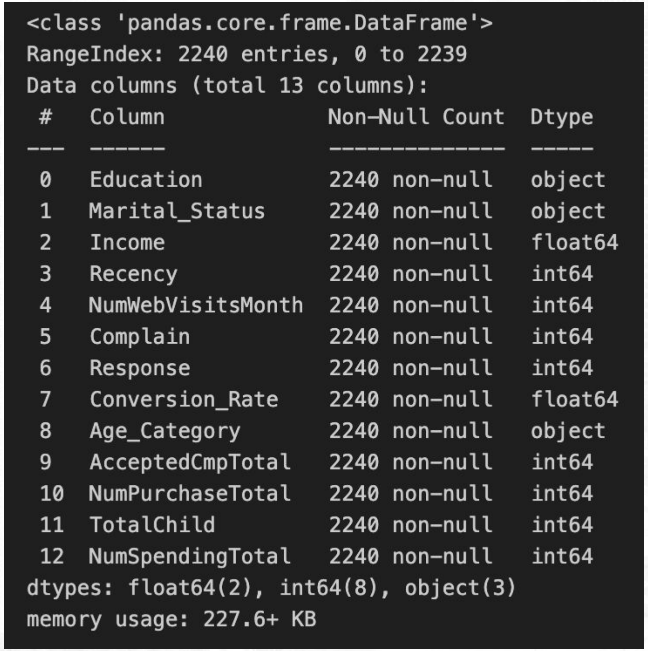
```
rows_with_na = df[df['Conversion_Rate'].isna()]
print(rows_with_na)
df['Conversion_Rate'] = df['Conversion_Rate'].fillna(0)

duplicate_row = df[df.duplicated(keep=False)]
duplicate_row
```

For more information, see jupyter notebook here

Data Cleaning & Preprocessing

As some new features are made, those features that wont be used for machine learning will be drop and only features below are retained:

- NumSpendingTotal
- TotalChild
- NumPurchaseTotal
- AcceptedCmpTotal
- NumWebVisitsMonth
- Age_Category
- Conversion Rate
- Complain
- Recency
- Education
- Marital_Status
- Income
- Response (target variable)



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 13 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Education          2240 non-null   object
 1   Marital_Status     2240 non-null   object
 2   Income             2240 non-null   float64
 3   Recency            2240 non-null   int64
 4   NumWebVisitsMonth  2240 non-null   int64
 5   Complain           2240 non-null   int64
 6   Response           2240 non-null   int64
 7   Conversion_Rate    2240 non-null   float64
 8   Age_Category       2240 non-null   object
 9   AcceptedCmpTotal   2240 non-null   int64
 10  NumPurchaseTotal   2240 non-null   int64
 11  TotalChild         2240 non-null   int64
 12  NumSpendingTotal   2240 non-null   int64
dtypes: float64(2), int64(8), object(3)
memory usage: 227.6+ KB
```

For more information, see jupyter notebook here

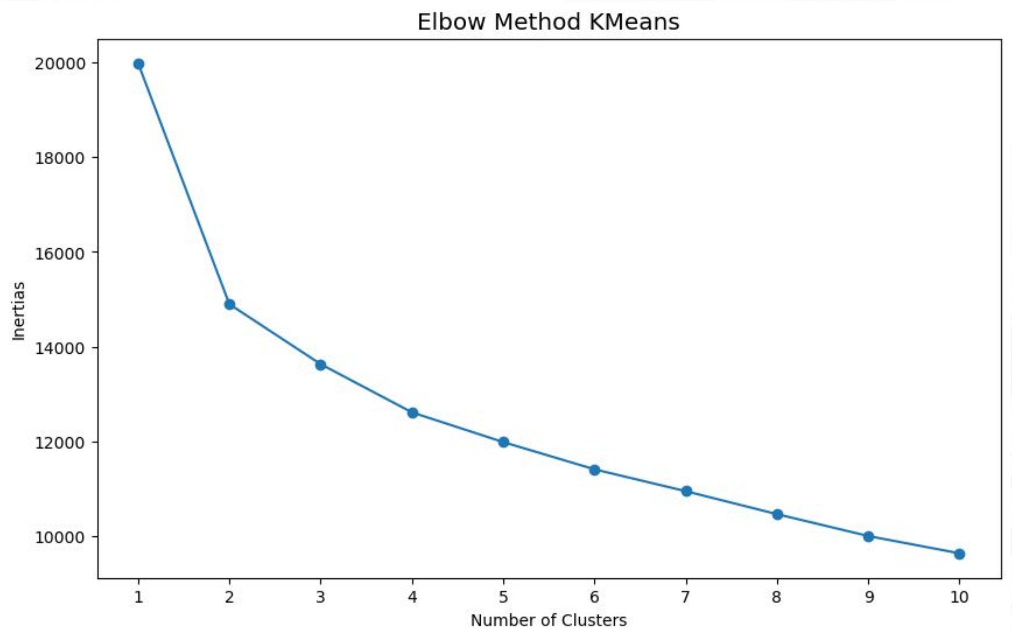
Data Cleaning & Preprocessing

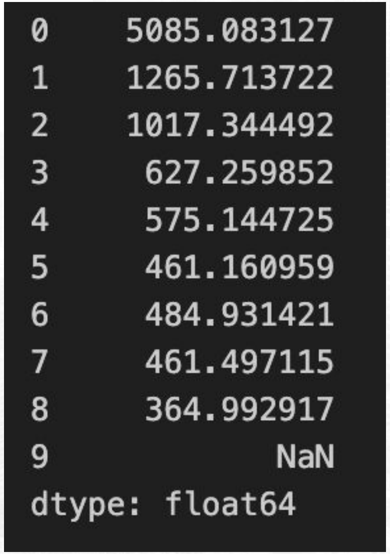
- Next for data preprocessing, ***feature encoding*** and ***feature standardisation*** are done.
  - Feature encoding for categorical features include:
    - Label encoding for ordinal data (Education, Age_Category)
    - One hot encoding for non-ordinal data (Marital_Status)
  - Feature standardisation using StandardScaler, features has been checked for standardised
    - Check if features were not encoded before
    - Check if features have different scales with other features

For more information, see jupyter notebook here

Data Modeling

- The ***Elbow Method*** using ***K-Means Clustering*** is performed to get better visualisation





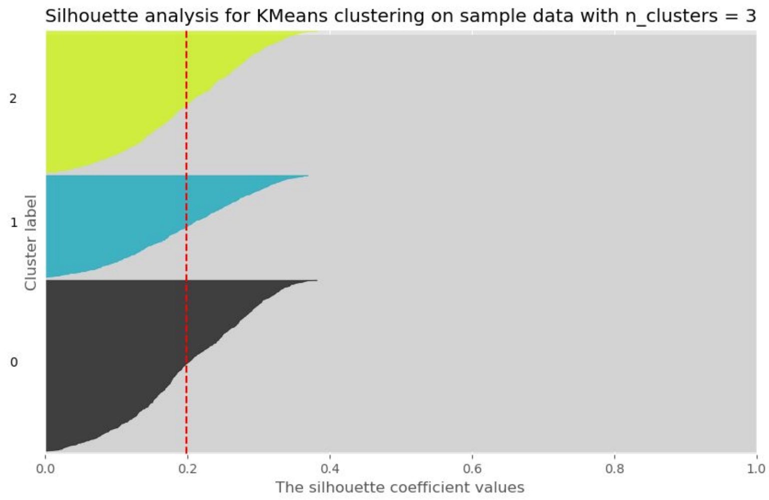
```
0    5085.083127
1    1265.713722
2    1017.344492
3     627.259852
4     575.144725
5     461.160959
6     484.931421
7     461.497115
8     364.992917
9            NaN
dtype: float64
```

- From the elbow method, the best cluster is 3. Although some confusion with 4 clusters, the difference between 3 to 4 is smaller rather than 4 to 5. However, keeping in mind 4 clusters to be check for modelling is cautious.

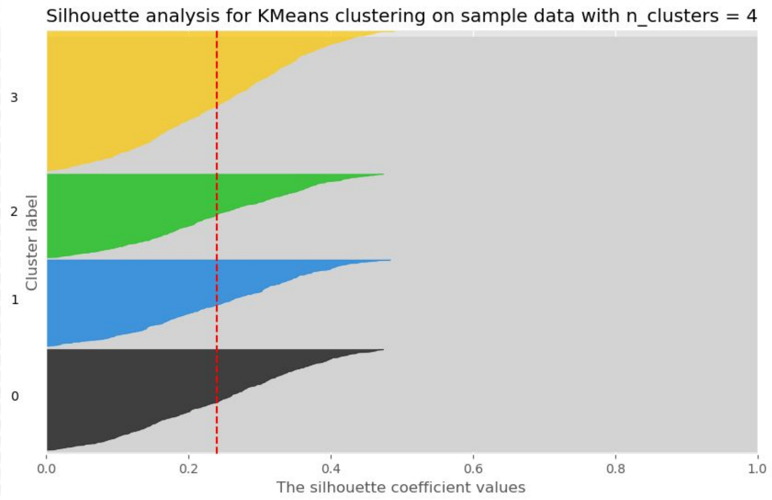
For more information, see jupyter notebook here

Data Modeling

- Evaluation using ***Silhouette Score*** can be seen below between 3 and 4 clusters.



Silhouette analysis for KMeans clustering on sample data with n_clusters = 3



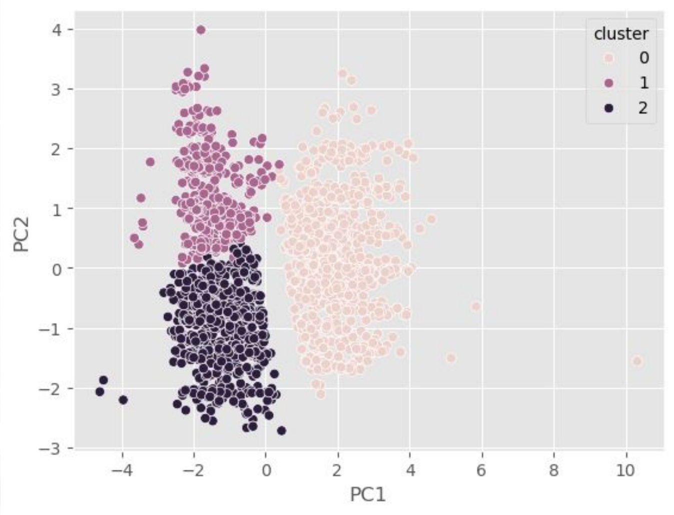
Silhouette analysis for KMeans clustering on sample data with n_clusters = 4

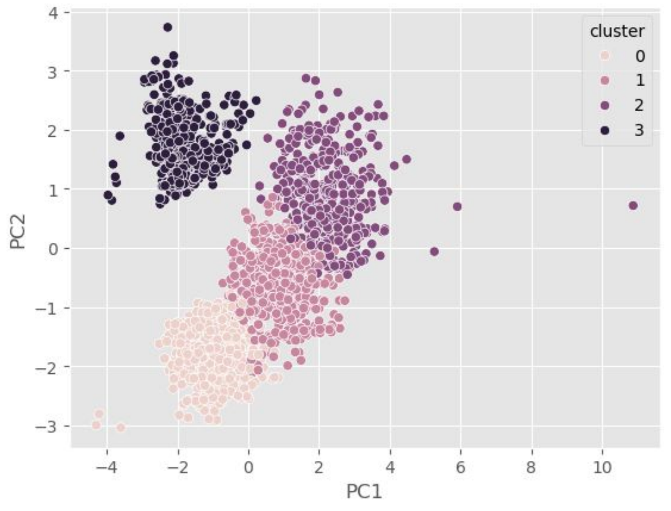
- From the graph, it can be seen that 1 and 0 in n_cluster=3 is being separated into three clusters in n_cluster=4 with more uniform shape.
- Silhouette coefficient is also higher for n_cluster=4 than n_cluster=3.

For more information, see jupyter notebook here

Data Modeling

- Evaluation using ***PCA*** can be seen below between 3 and 4 clusters.

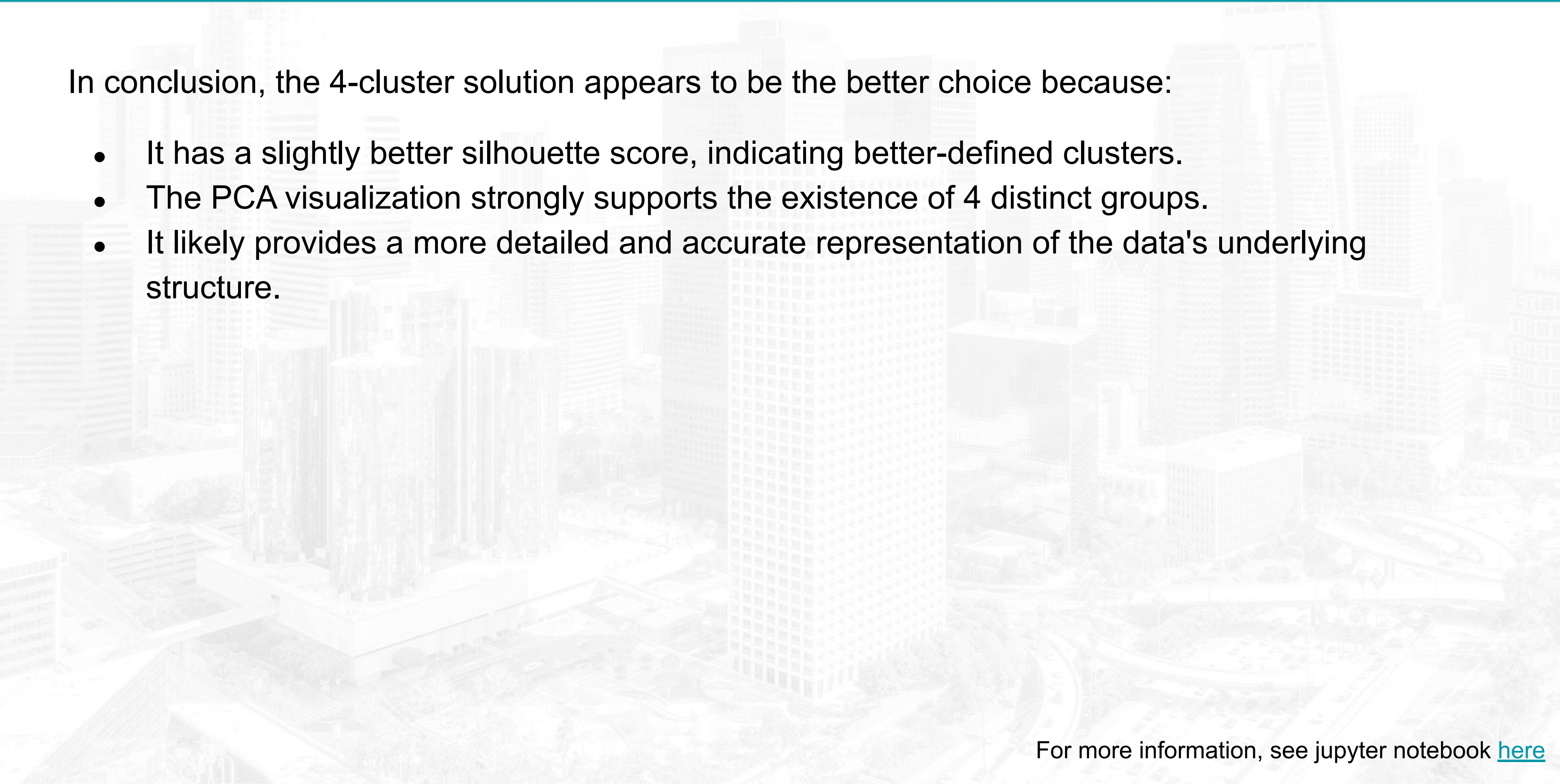




- From the graph, it can be seen that a slight difference in clustering with n_cluster=3 seems to be more spread than n_cluster=4
- In n_cluster=4, although cluster 2 and 3 seems to be overlap, there’s a clear distinct as such cluster 2 is high values on PC2 but low values on PC1. While the opposite is for cluster 3.

For more information, see jupyter notebook here

Data Modeling



In conclusion, the 4-cluster solution appears to be the better choice because:

- It has a slightly better silhouette score, indicating better-defined clusters.
- The PCA visualization strongly supports the existence of 4 distinct groups.
- It likely provides a more detailed and accurate representation of the data's underlying structure.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Customer Segmentation Cluster Analysis

- *Cluster 0: Low-Value, Older Customers*
  - Education: Highest average (2.90)
  - Income: Lowest (-0.45)
  - Web Visits: Slightly above average (0.33)
  - Age: Oldest (5.41, likely 55+ years)
  - Purchases: Lowest (-0.72)
  - Spending: Lowest (-0.78)
  - Marital Status: Highest percentage of widowed (3.8%) and divorced (13.2%)

Interpretation: This cluster represents older customers with higher education but lower income and spending. They visit the website occasionally but have the lowest purchase and spending rates.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Customer Segmentation Cluster Analysis

- *Cluster 1: Average-Value, Middle-Aged Customers*
  - Education: Above average (2.65)
  - Income: Slightly above average (0.35)
  - Web Visits: Slightly above average (0.12)
  - Age: Second oldest (5.28, likely 45-54 years)
  - Purchases: Above average (1.00)
  - Spending: Slightly below average (0.45)
  - Marital Status: Highest percentage of married (40.4%)

Interpretation: This cluster represents middle-aged, married customers with above-average education and income. They have average web engagement and purchasing behavior.

For more information, see jupyter notebook here

Customer Segmentation Cluster Analysis

- *Cluster 2: High-Value, Young Adult Customers*
  - Education: Above average (2.50)
  - Income: Highest (1.10)
  - Web Visits: Lowest (-1.23)
  - Age: Second youngest (4.82, likely 35-44 years)
  - Purchases: Above average (0.66)
  - Spending: Highest (1.31)
  - Campaign Acceptance: Highest (0.82)
  - Marital Status: Highest percentage of engaged (26.3%)

Interpretation: This cluster represents young adult customers with high income and spending. They have the highest campaign acceptance rate but visit the website less frequently.

For more information, see jupyter notebook here

Customer Segmentation Cluster Analysis

- *Cluster 3: Low-Middle Value, Young Customers*
  - Education: Lowest (1.74)
  - Income: Second lowest (-0.87)
  - Web Visits: Highest (0.63)
  - Age: Youngest (4.00, likely 25-34 years)
  - Purchases: Slightly below average (-0.87)
  - Spending: Second lowest (-0.82)
  - Children: Highest average (0.96)
  - Marital Status: Highest percentage of single (26.8%)

Interpretation: This cluster represents young, single customers with lower education and income. They have the highest web engagement but lower purchasing and spending rates.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Key Insights:

- Age and life stage significantly influence customer behavior across clusters.
- Income levels correlate with spending patterns and campaign responsiveness.
- Web engagement doesn't always translate to higher purchase rates or spending.
- Education levels vary across clusters but don't directly correlate with spending.
- Marital status and the presence of children appear to influence customer behavior.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Insights	Action	Example
Age and Life Stage Influence	Develop age-specific marketing campaigns and product offerings.	- Create a loyalty program for older customers in Cluster 0, focusing on value and reliability. - Design trendy, innovative products for the younger Cluster 3, emphasizing digital engagement.
Income Levels Correlation	Tailor pricing strategies and product ranges to each cluster's income level.	- Offer premium, high-end products to Cluster 2 (high-income group). - Develop budget-friendly options and value deals for Clusters 0 and 3.
Web Engagement vs. Purchase Rates	Optimize the website to convert high engagement into sales, especially for Cluster 3.	- Implement personalized product recommendations and targeted promotions on the website. - Use retargeting ads for Cluster 3 to convert their high web visits into purchases.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Insights	Action	Example
Education Level Variations	Adjust communication styles and product information to suit each cluster's education level.	- Provide detailed, in-depth product information for the highly educated Cluster 0. - Create more visual, easy-to-understand content for Cluster 3 with lower average education.
Marital Status and Children Influence	Develop products and services that cater to different family structures.	- Create family-oriented promotions for Cluster 3, which has the highest average number of children. - Develop "couples" packages or services for the predominantly married Cluster 1.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Insights	Action	Example
Cross-Cluster Strategies	Implement strategies to move customers up to higher-value clusters.	- Create a mentorship program where high-value customers from Cluster 2 share experiences with those in Cluster 3, potentially increasing engagement and spend. - Develop a tiered loyalty program that encourages customers to increase their purchasing behavior to reach higher tiers with better benefits.
Customer Lifetime Value Focus	Implement strategies to increase the lifetime value of customers in each cluster.	- For Cluster 0, focus on retention through personalized service and age-appropriate products. - For Cluster 2, emphasize exclusive experiences and early access to new products to maintain their high-value status.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Rakamin Academy

Comparison of EDA and Clustering Insights

- *Similarities*
    - Income and Spending Correlation
        - EDA: Higher income levels are associated with higher conversion rates and increased total spending.
        - Clustering: Income levels correlate with spending patterns and campaign responsiveness (Key Insight 2).
        - Similarity: Both analyses confirm that higher income is associated with higher value customers in terms of conversions and overall spending.
    - Age-Based Differences
        - EDA: Younger to middle-aged adults tend to have higher conversion rates compared to older age groups.
        - Clustering: Age and life stage significantly influence customer behavior across clusters (Key Insight 1).
        - Similarity: Both analyses highlight the importance of age in customer behavior, suggesting the need for age-specific marketing strategies.
    - Interconnected Customer Behavior
        - EDA: Strong interconnections between different types of customer spending, with positive correlations across product categories and purchase channels.
        - Clustering: Web engagement doesn't always translate to higher purchase rates or spending (Key Insight 3).
        - Similarity: Both analyses suggest complex relationships between different aspects of customer behavior, although they highlight different specific relationships.

For more information, see jupyter notebook here

Customer Personality Analysis for Marketing Retargeting

Comparison of EDA and Clustering Insights

- *Potential Contradictions or Nuances*
  - Web Engagement and Purchases
    - EDA: Suggests a positive correlation between engagement and spending across channels.
    - Clustering: Indicates that high web engagement doesn't always lead to higher purchase rates (Key Insight 3).
    - Nuance: This suggests that while there's generally a positive relationship between engagement and spending, this relationship may vary across different customer segments.
  - Age and Value
    - EDA: Suggests younger to middle-aged adults have higher conversion rates.
    - Clustering: Identifies the oldest cluster (Cluster 0) as having the highest education but lowest income and spending.
    - Nuance: This highlights the complexity of age-based segmentation, where factors like education and income interact with age to influence customer behavior.
  - Education and Customer Value
    - EDA: Doesn't explicitly mention education.
    - Clustering: Notes that education levels vary across clusters but don't directly correlate with spending (Key Insight 4).
    - Nuance: This suggests that while education is an important factor in segmentation, its relationship with customer value is not straightforward and interacts with other factors like age and income.

For more information, see jupyter notebook here

Additional Insights from Clustering

1. The clustering analysis provides additional insights into the role of marital status and presence of children in customer behavior.
2. The clustering approach allowed for the identification of distinct customer segments (e.g., high-value young adults, low-middle value young customers), providing a more nuanced view than the general trends identified in the EDA.

For more information, see jupyter notebook here