



PREDICT CUSTOMER PERSONALITY TO BOOST MARKETING CAMPAIGN BY USING MACHINE LEARNING

www.linkedin.com/in/fadillahakbar



Fadillah Akbar

Data scientist with a strong mathematical foundation and problem-solving abilities. Experienced in building projects in data mining, data processing, business performance analysis, data visualization, and predictive modeling across multiple industries. Motivated to use data science to improve business impact through analytics, statistics, and machine learning.

www.linkedin.com/in/fadillahakbar



BUSINESS OVERVIEW

To manage a company's product sales strategy, it is crucial to comprehend how client profiles and purchasing behavior. With that, we can offer the appropriate remedy based on each issue the consumer encounters to offer a better experience in the future. This will have an impact on the rise in transaction satisfaction and may boost sales performance going forward. Customers can be divided up into several groups for this analysis using a clustering approach. You'll learn more about it through this project.



Table of Content

- **Conversion Rate Analysis Based on income, Spending and Age**
.....
- **Data Cleaning & Preprocessing**
.....
- **Data Modeling**
.....
- **Customer Personality Analysis for Marketing Retargeting**
.....

Conversion Rate Analysis

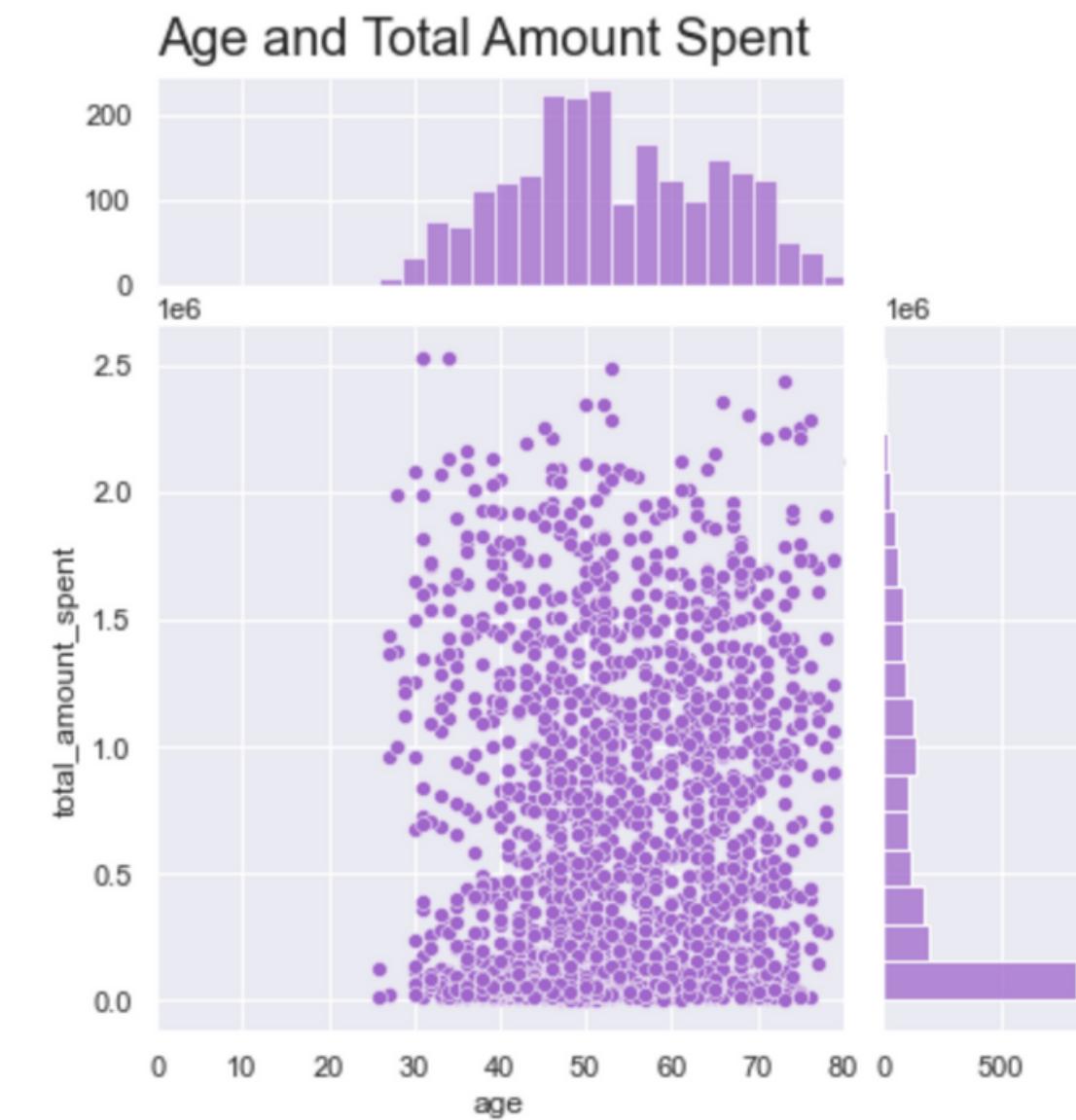
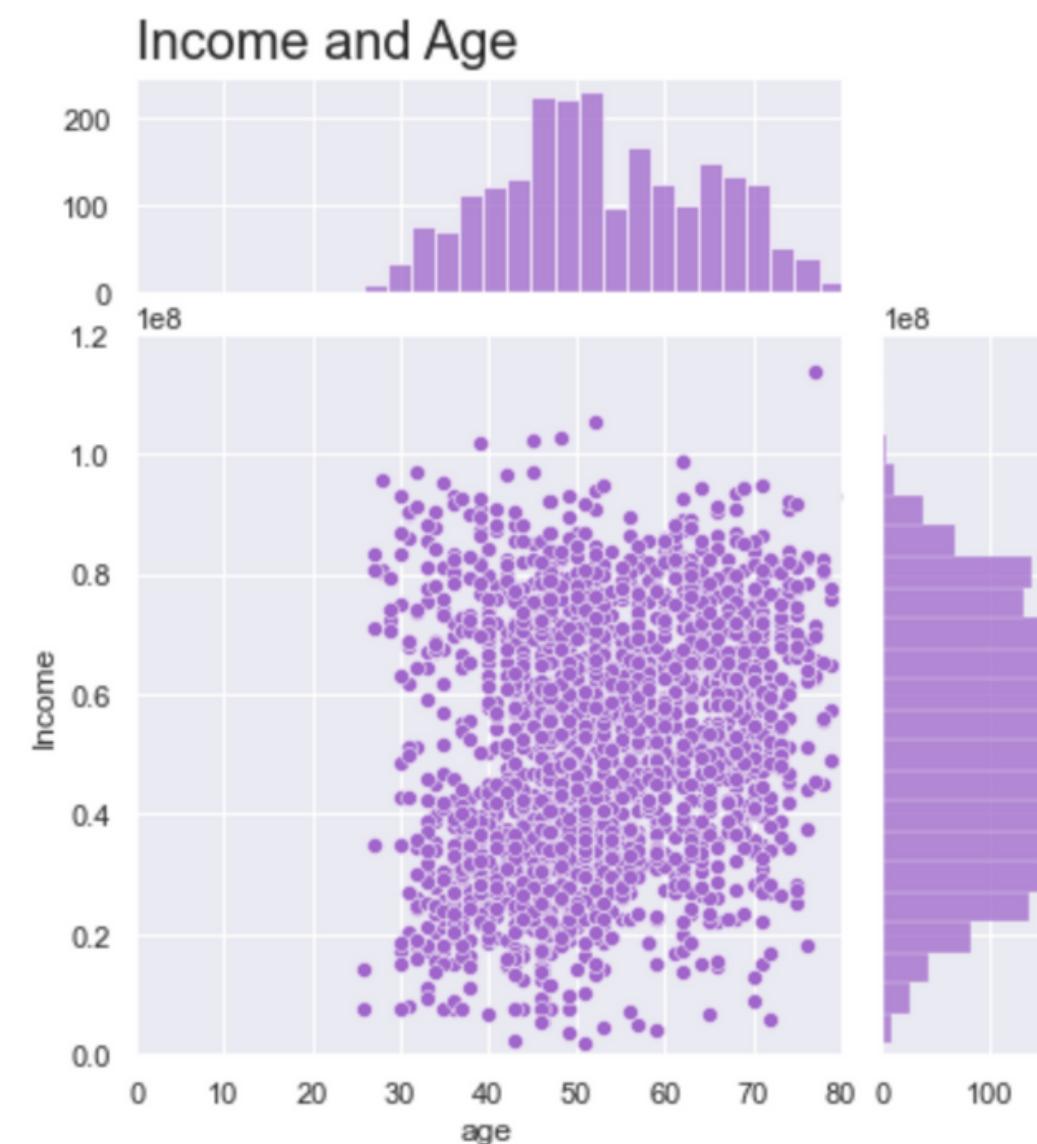
Based on income, Spending and Age

Conversion rate analysis is a search for data insight into the percentage of website visitors and what actions they take while visiting the site, and whether their actions result in a purchase transaction or not while visiting the website, this can be done by performing feature engineering on the data variables presented, so as to produce a new column, namely Conversion rate.



Relationship Between Income, Age and Total Spending

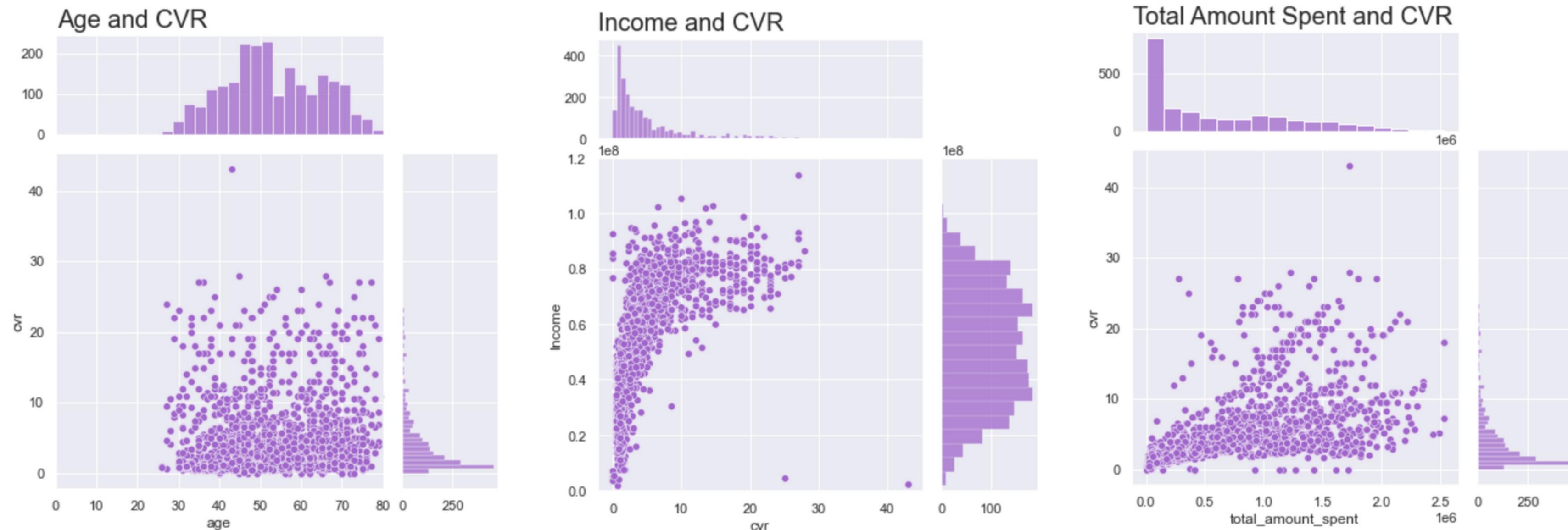
www.linkedin.com/in/fadillahakbar



At this stage, we analyze how the relationship between total spending, income or age. The results of the data exploration obtained, the greater the income, there is a tendency to have a greater total expenditure. Meanwhile, the older the age of our customers does not affect the amount of expenses and income.

Conversion Rate Analysis Based on Income, Spending and Age

www.linkedin.com/in/fadillahakbar



At this point, we investigate the connections between feature conversion rate, income, age, and overall spending. The results of the data research revealed that there is a tendency for overall spending on our platform to increase with income. Other characteristics, like age and overall spending, don't show this. We can choose clients with incomes of 60 milion and beyond for marketing campaigns by focusing on a sizable population with a CVR of 10% or above.

Data Cleaning & Preprocessing

Prepare raw data into clean and ready-to-process data.
Handle various data problems such as empty data,
incompatible data, and identify data that is not needed.



Data Cleaning & Preprocessing



Null Data

Null values were found in the Income column for 24 rows (1.07% of the data). since the number is quite low, we can safely delete these

```
dataset = dataset.dropna()
```



Duplicate Data

After performing duplicate detection no duplicate data was found

```
dataset.duplicated().sum()
```



Feature Encoding

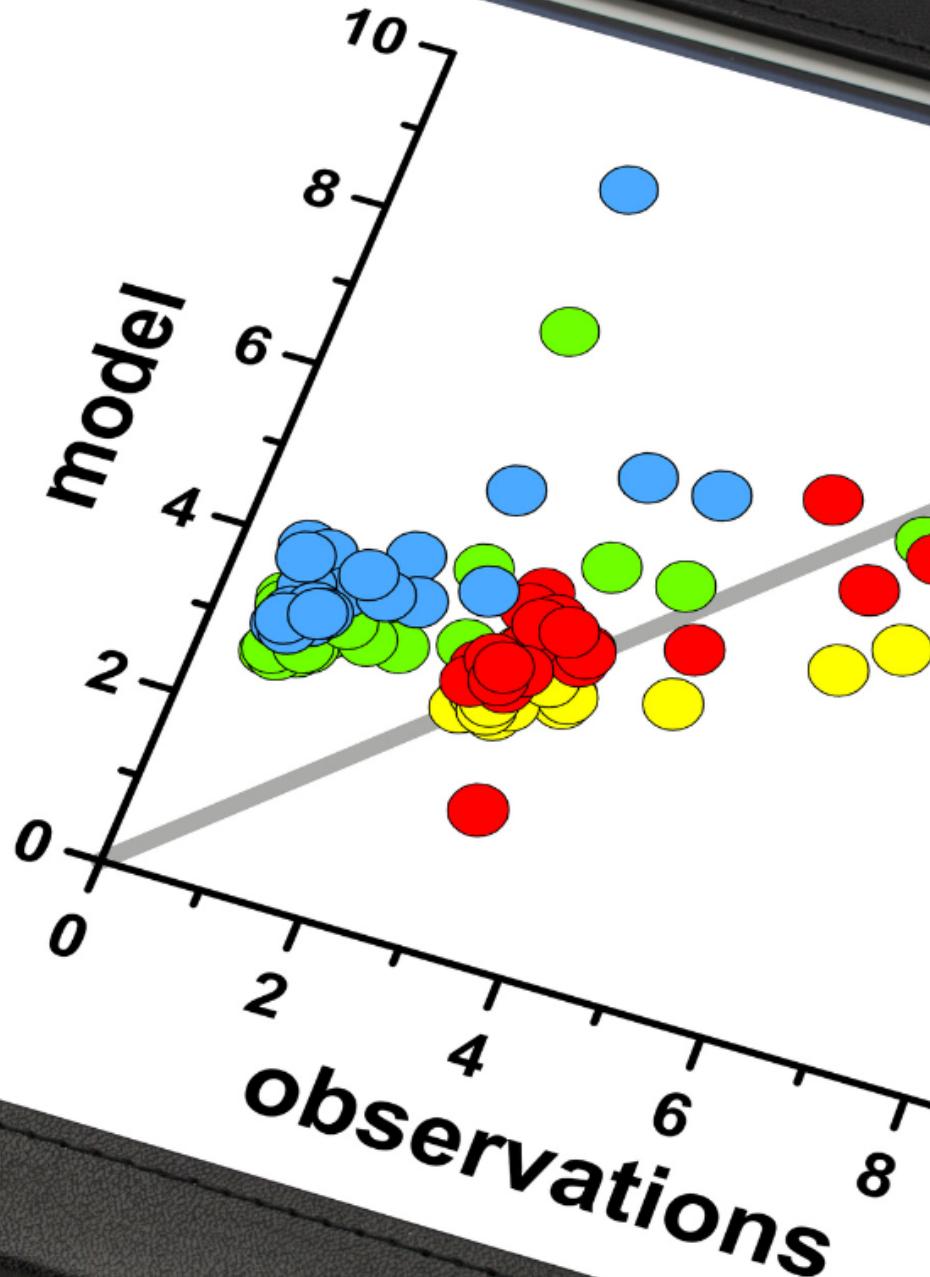
Education features using label encoding while for other features using one hot encoding.

```
# Label encoder
mapping_education = {'SMA' : 0, 'D3' : 1, 'S1' : 2, 'S2' : 3, 'S3' : 4}
dataset['Education_mapped'] = dataset['Education'].map(mapping_education)
```

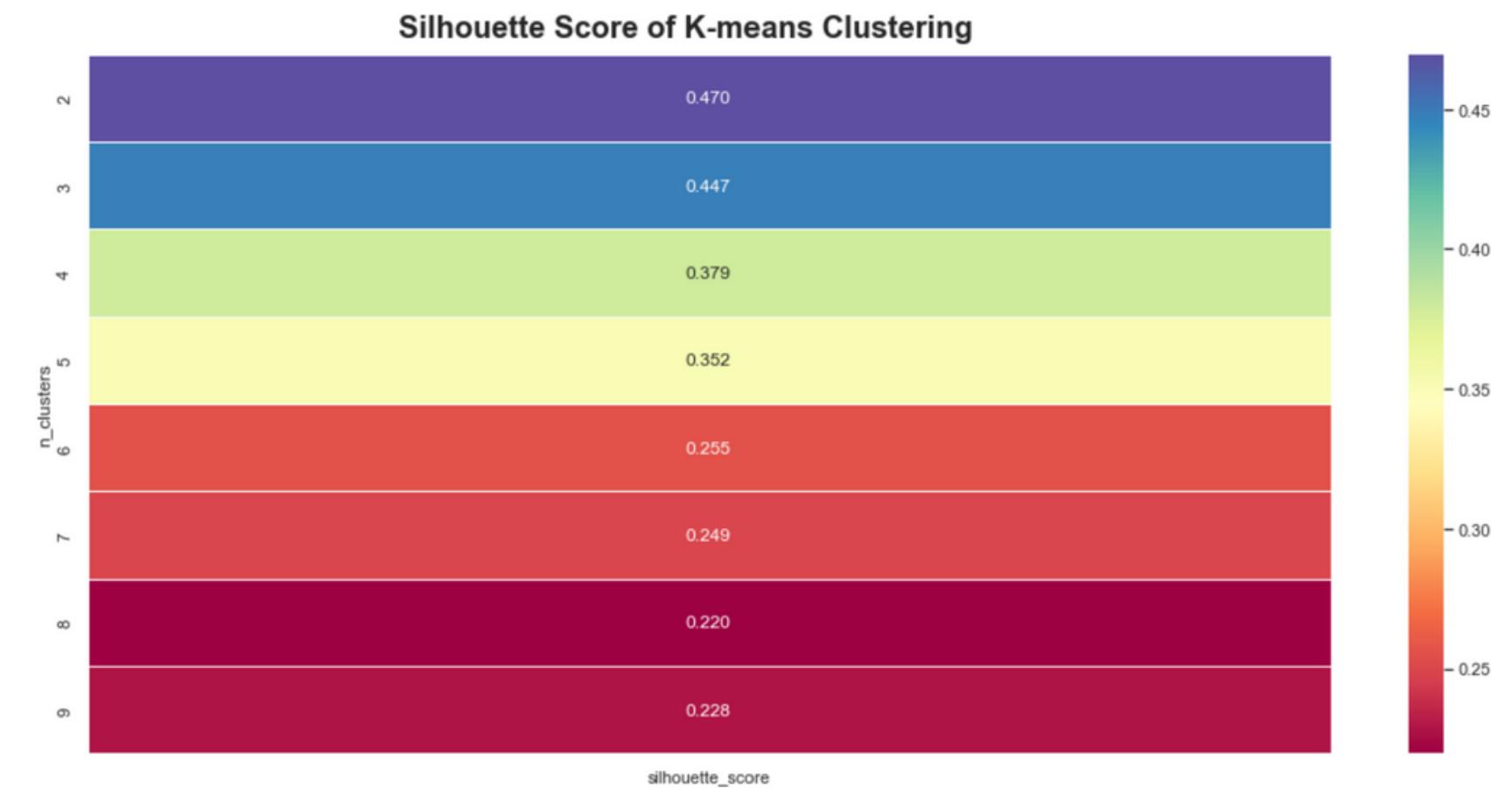
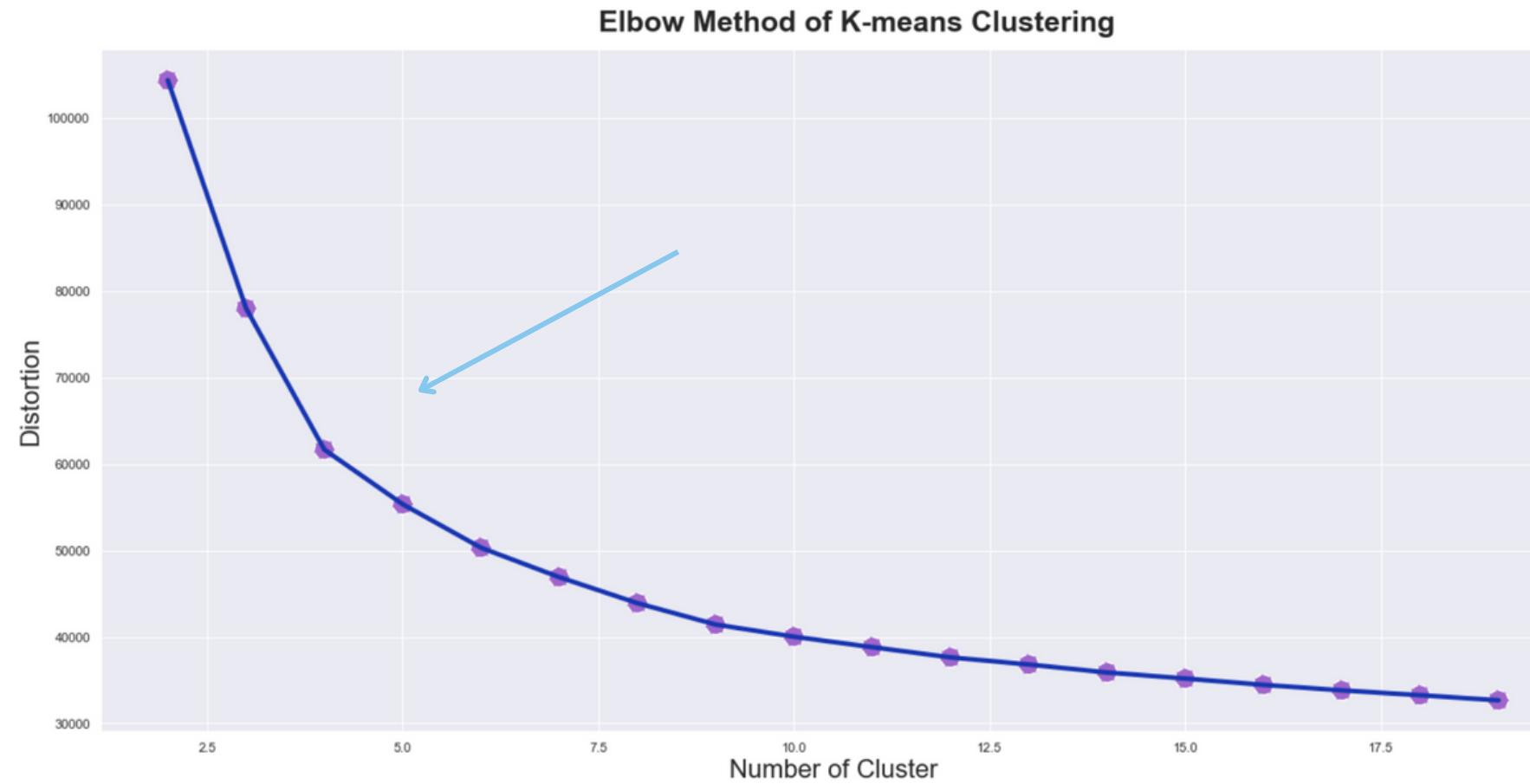
```
# One hot encoder
for cat in ['Marital_Status', 'age_range', 'is_parents']:
    onehots = pd.get_dummies(dataset[cat], prefix=cat)
    dataset_cats = dataset_cats.join(onehots)
```

Data Modeling

Apply the k-means clustering algorithm to the dataset that is ready to be processed, then choose the right number of clusters by looking at the elbow method, and evaluate using the silhouette score.



K-means Clustering



Insight/analysis

The elbow method and silhouette score analysis were conducted to see the number of clusters obtained from the k-means clustering algorithm. In the graph above, we can see that the appropriate number of clusters is 4. The number of clusters is sufficient to separate each cluster with the most optimal distance.

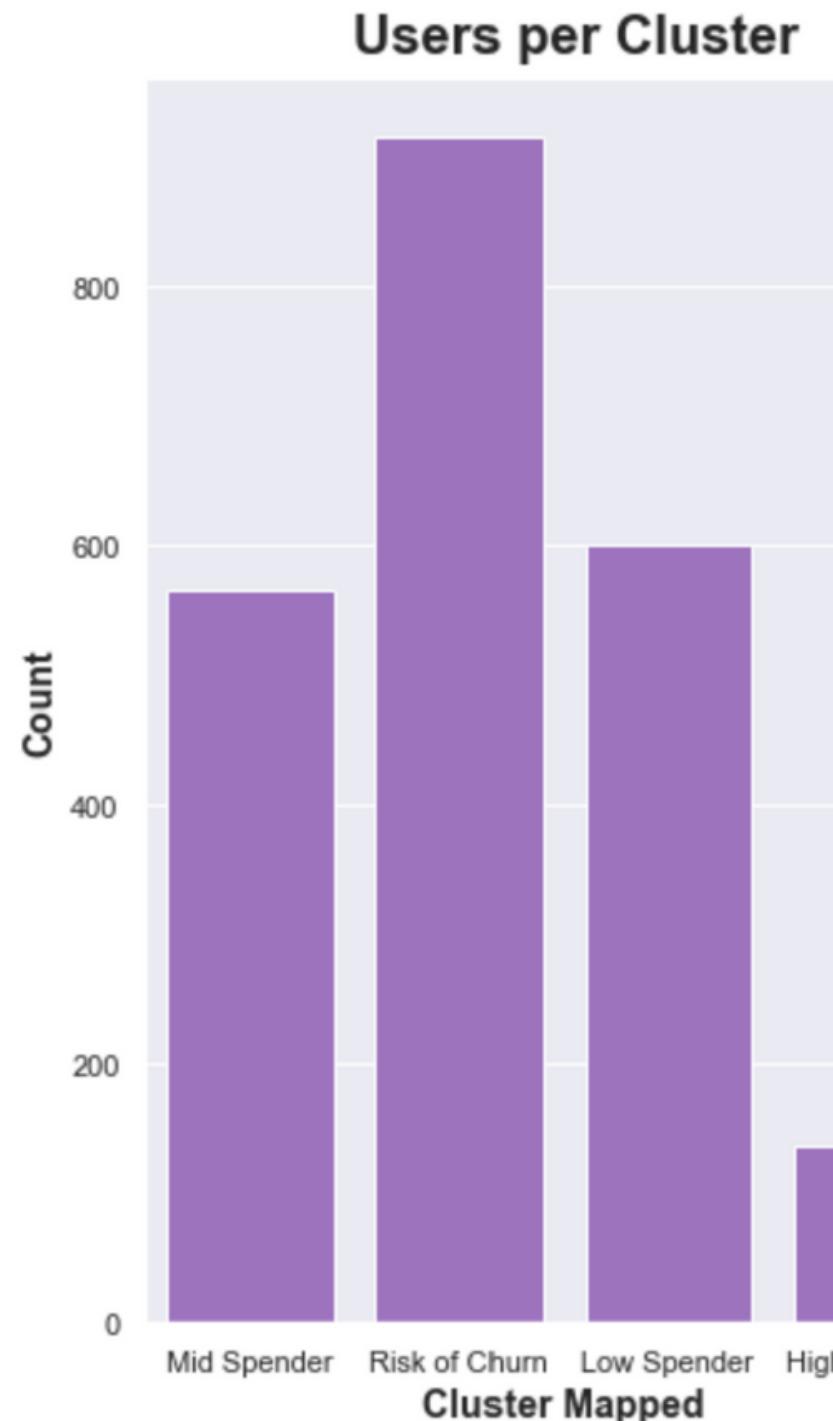
Customer Personality Analysis for Marketing Retargeting

The results of the clustering that has been done can be interpreted based on the characteristics of each group, how the tendency of the cluster to respond to existing marketing campaigns, and how the potential revenue results if we do retargeting marketing to the cluster.



Customer Personality Analysis for Marketing Retargeting

www.linkedin.com/in/fadillahakbar



Low Spender

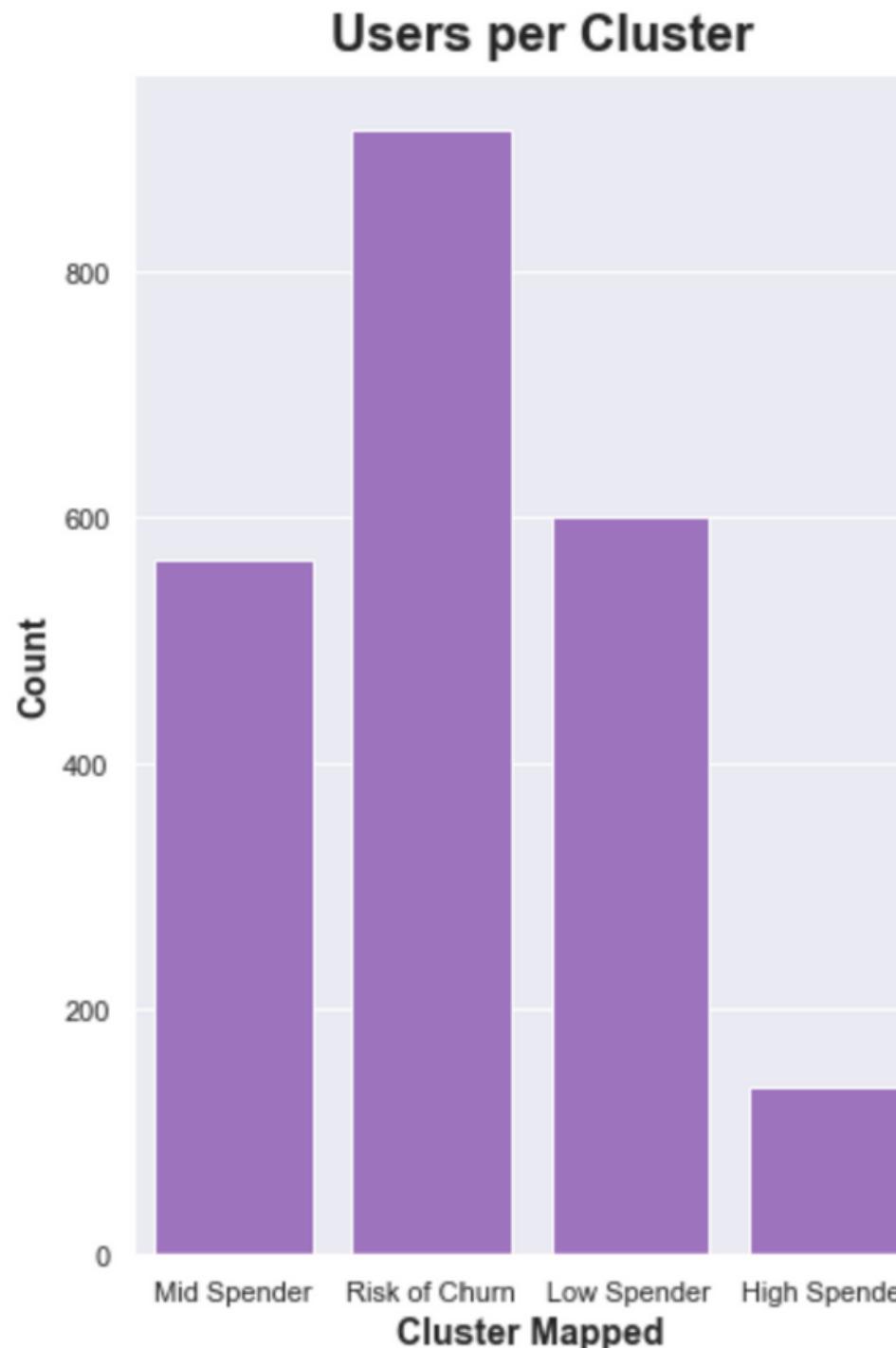
- This category is dominated by older adults (>55 years old) and middle aged adults (36-55 years old), who are mostly married and have one child.
- This group accesses the website frequently, on average 5 times per month; nevertheless, this group searches for promotions frequently, with each person purchasing promos 2 times per month.
- This category has the second lowest annual total income and spending, with IDR 57 Million for total revenue and IDR 506K for annual expenditure.

Risk of Churn

- This category has the most users, with up to 900 persons, and is dominated by middle aged adults (36-55 years), who are mostly married and have one child.
- This category has the lowest monthly income and spending, with total annual income of IDR 33.4 Million and annual expenditure of IDR 57K, respectively.
- This group visits the web the most, with a median of 7 visits per month, but they rarely make purchases and even employ promotions in their purchases.
- Not many respond to campaigns compared to other groups. They come organically

Customer Personality Analysis for Marketing Retargeting

www.linkedin.com/in/fadillahakbar



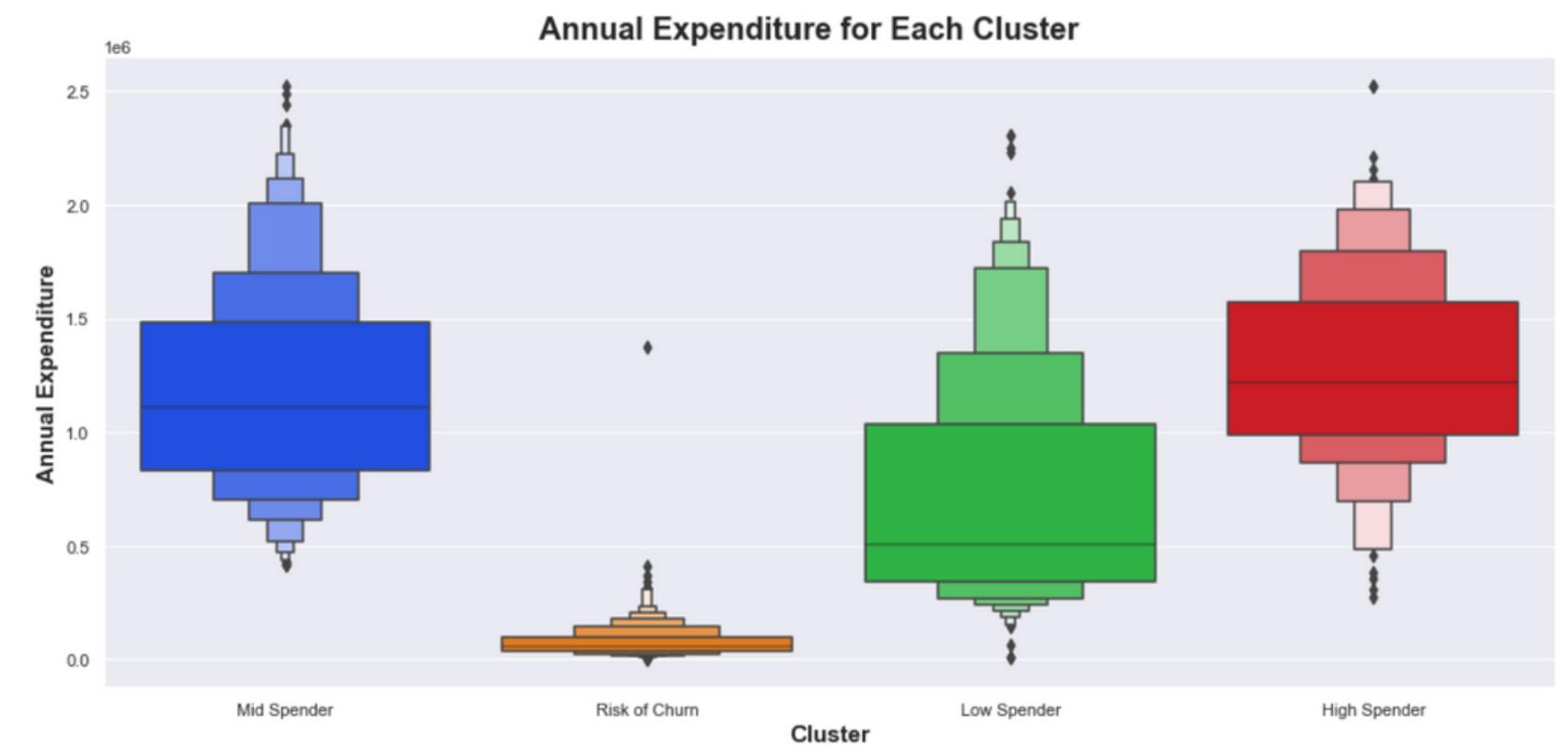
Mid Spender

- This category is dominated by older adults (>55 years) and middle aged adults (36-55 years), who are mostly married and have 0-1 children.
- this category has the second highest total income and spending, with IDR 68 million for total revenue and IDR 1.1 million for expenditure every year.
- Although they rarely use the web, this group responds to our ads and uses promos the most frequently in a month, with an average of three times each month.

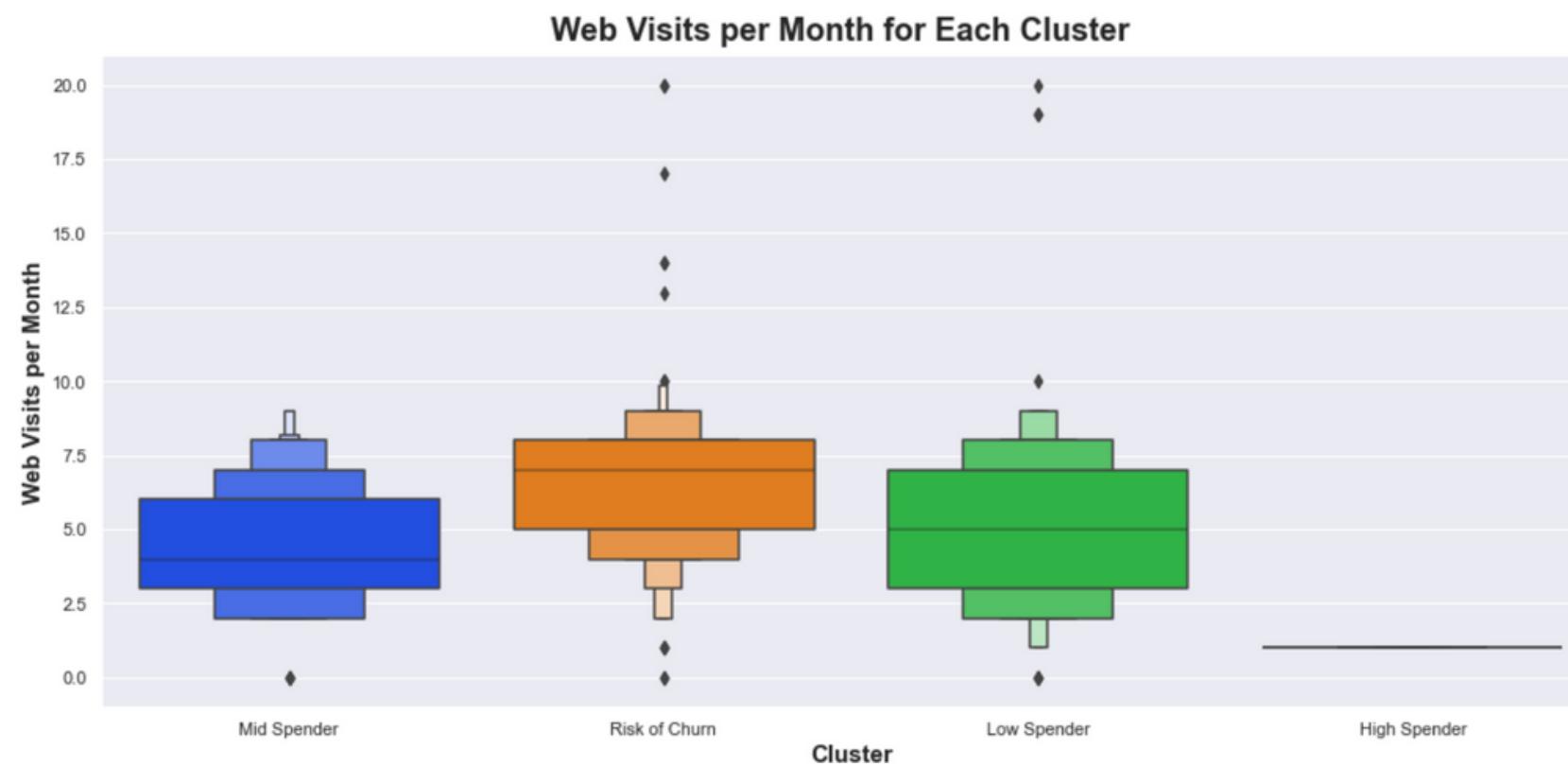
High Spender

- This category has the fewest users, with 137 persons, and is dominated by older adults (>55 years). middle aged adults (36-55 years), who are mostly married and do not have children
- This category has the greatest monthly income and expenditure, which is IDR 80 million for total revenue per year and IDR 1.2 million for expenditure per year, respectively.
- This cluster is almost entirely non-organic in terms of campaign response, although it has the least amount of promotional utilization compared to the others.
- This category has the highest conversion rate for our products and should not be overlooked.

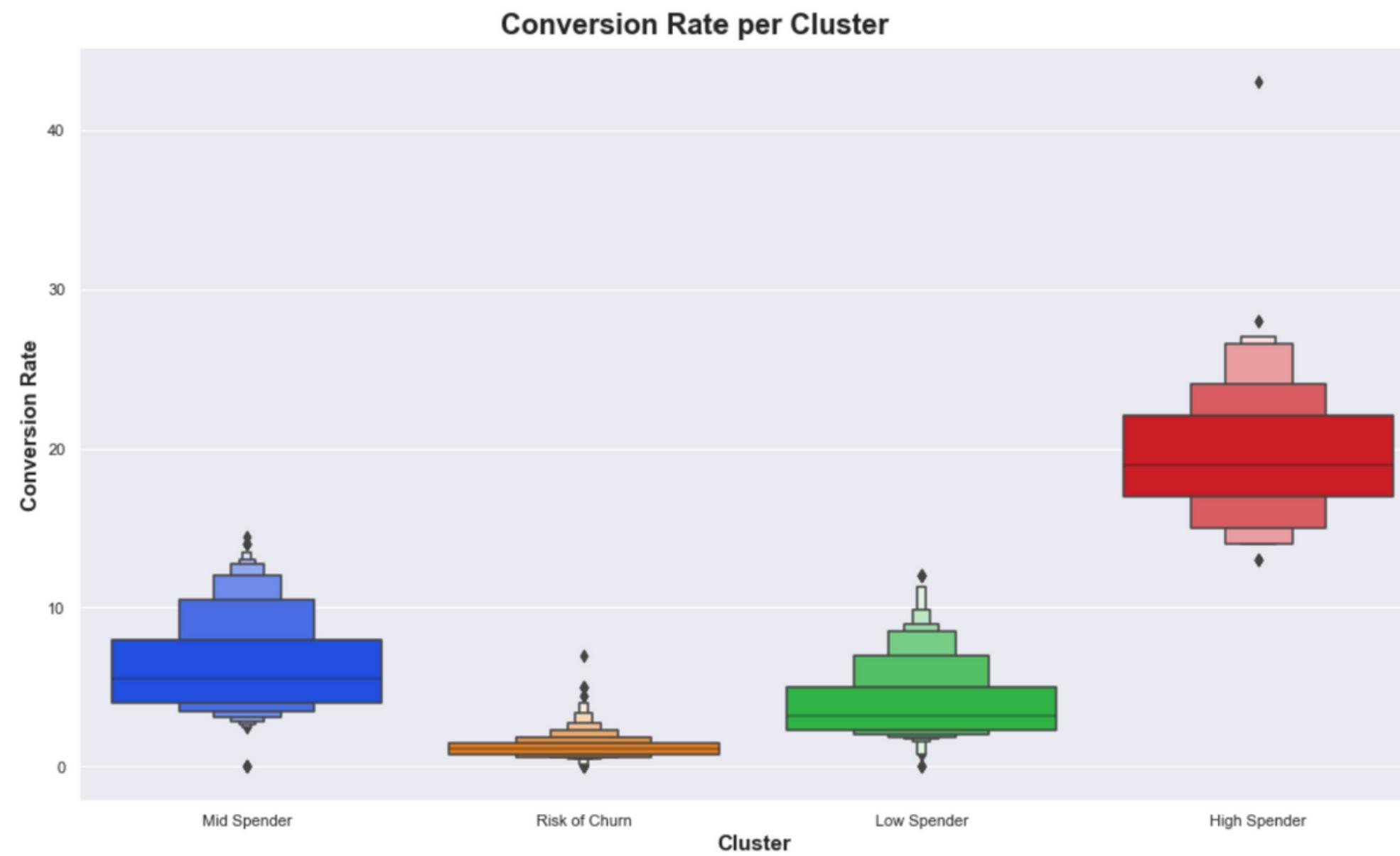
Customer Personality Analysis for Marketing Retargeting



Customer Personality Analysis for Marketing Retargeting



Recommendation & Potential Impact



■ Recommendation

- Continue to monitor transactions and retention of the High Spender category, and focus on enhancing service to prevent churn.
- Further research can be conducted for the Mid Spender group on how to boost their transactions by providing more targeted recommendations, as well as further research on how to improve promotions for this segment and keep them buying on our platform.
- Further analysis can be done on how to boost the conversion ratio of visits to transactions for the Low Spender and Risk to Churn groups, as they have a pretty high number of visits but do not make transactions. This can be caused by unsuitable products or prices.

■ Potential Impact

- If we continue to monitor the High Spender group, we will earn a possible GMV of IDR 176 Million, while the Mid Spender group would get IDR 66 Million.
- If we can optimize the promotion expenditure for the Mid Spender (assuming a 50% reduction), we can save IDR 50 million.



Thank you!

For more details, you can check out the jupyter notebook [here](#).

The dataset can be seen [here](#).