

Predict Customer Personality to Boost Marketing Campaign Using Unsupervised Machine Learning

Chantika Azzamy S. Yahya





Hi!

I'm Chantika Yahya

Having recently graduated with a degree in Management and completed the Data Science Program at Rakamin Academy, I have developed a strong interest in data analysis and visualization. My goal is to leverage my skills in data analysis and statistics to help businesses create value and make informed decisions.



chantikayahya1@gmail.com



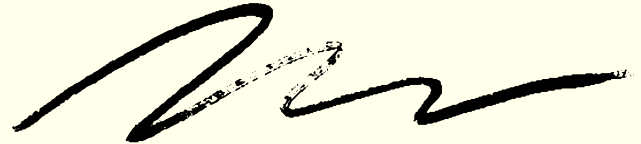
[Chantika Yahya](#)

I am skilled in data visualization and statistical analysis. I also possess expertise in a range of tools including Excel, Google Sheets, SQL, Python, Tableau, and Google Data Studio. With excellent communication and teamwork skills, I am able to clearly and effectively communicate complex information to stakeholders. My strong analytical, problem-solving, and critical thinking abilities coupled with my attention to detail enable me to efficiently handle data and create meaningful insights. I am excited to bring my skills and passion for data analysis to contribute to the growth and success of a data-driven organization.



Overview

Understanding customer behavior and personalities is key to a company's rapid development, as it allows the company to offer enhanced services and benefits to customers who may become loyal in the future. By analyzing past marketing campaign data, companies can improve performance and target the right customers for transactions on their platforms. Our main objective is to create a cluster prediction model from these data insights, which can enable companies to make better-informed decisions.



Disclaimer

If you're interested to see the full notebook, you can access it here:

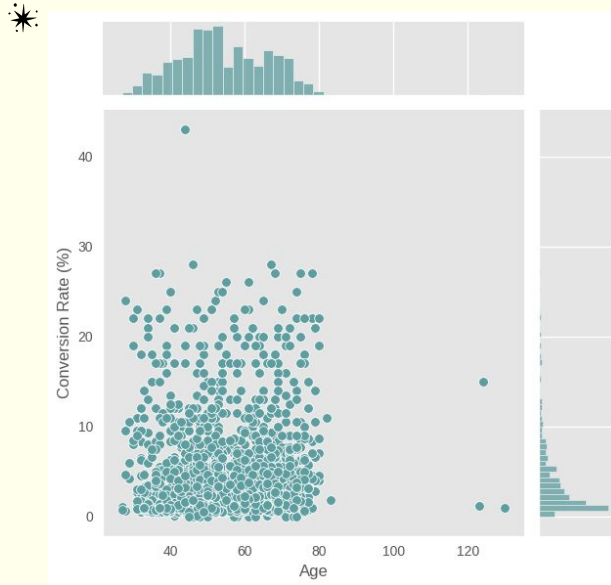
[Google Colab](#)



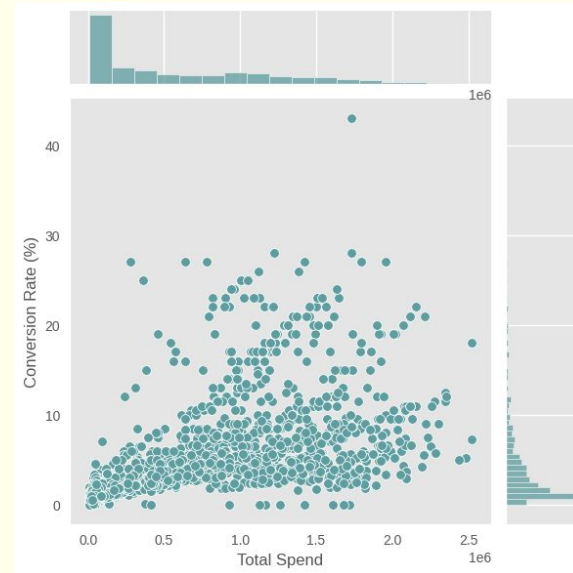
Conversion Rate Analysis Based on Income, Total Spend, and Age



Age vs Conversion Rate



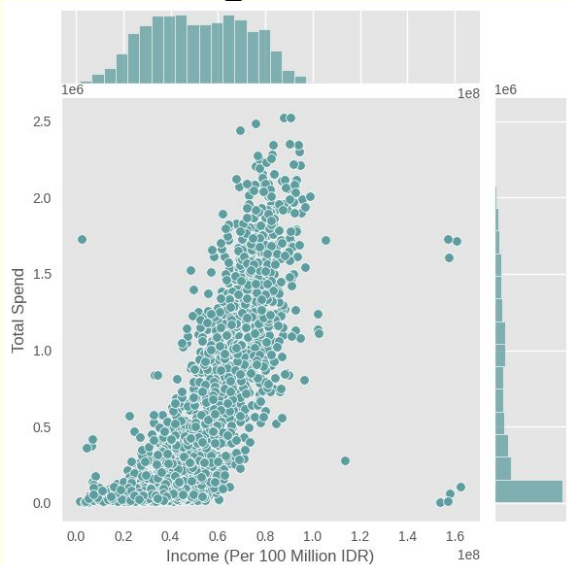
Total Spend vs Conversion Rate



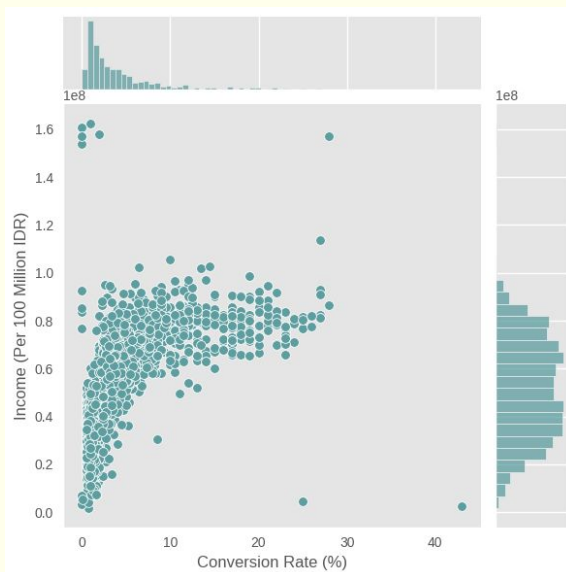
Conversion Rate Analysis Based on Income, Total Spend, and Age



Income vs Total Spend



Income vs Conversion Rate



In this phase, we analyze the correlation between conversion rates and income, total spend, and age. Our exploratory data analysis indicates that customers with higher incomes tend to spend more, while age does not show a similar correlation.

Based on the conversion rate analysis, we suggest a campaign that targets customers with an income of over 60 million and a conversion rate above 10%.



Data Cleaning and Preprocessing



- There are 24 missing values in the 'Income' column. **The missing value is deleted.**
- There are **no duplicate data.**

There are many outliers in the dataset. However, we don't want to delete too many data. Therefore, we just want to **remove the irrelevant rows in 'age' column and too far outlier in 'Income' column**. Then, the rest of the outliers if they are considered skewed, we will **perform log transformation** to those columns.

Perform **label encoding** for 'Education' and 'age_group' columns. Also, performing **one hot encoding** to the 'living_with' column.

Handling Missing Values and Duplicates

Handling Outliers

Feature Encoding

```
df.dropna(subset=['Income'], inplace=True)
```

```
df.duplicated().sum()
```

```
0
```

```
# Remove too far outlier in Income column and Age that's already not relevant anymore
df_fs = df_fs[df_fs['age'] < 90]
df_fs.drop(df_fs[df_fs['Income'] == df_fs['Income'].max()].index, inplace=True)
print("The total number of data-points after removing the outliers are:", len(df_fs))

The total number of data-points after removing the outliers are: 2212

for i in range(0, len(log)):
    skew = df_log[log[i]].skew(axis = 0, skipna = True)
    print(f'{log[i]}: {skew}')
```

```
mapping_education = {'SMA' : 0,
                     'D3' : 1,
                     'S1' : 2,
                     'S2' : 3,
                     'S3' : 4}

df_fs['Education'] = df_fs['Education'].map(mapping_education)

mapping_age = {'Young Adult' : 0,
              'Adult' : 1,
              'Senior' : 2}

df_fs['age_group'] = df_fs['age_group'].map(mapping_age)

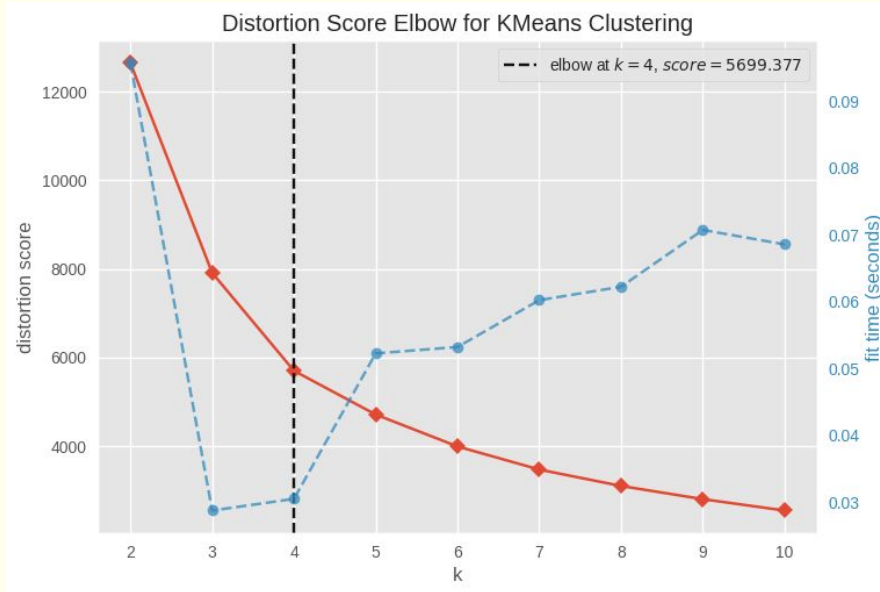
prefix_living = pd.get_dummies(df_fs['living_with'], prefix = 'is')

df_fs = df_fs.join(prefix_living)

df_fs = df_fs.drop(['living_with'], axis = 1)
```

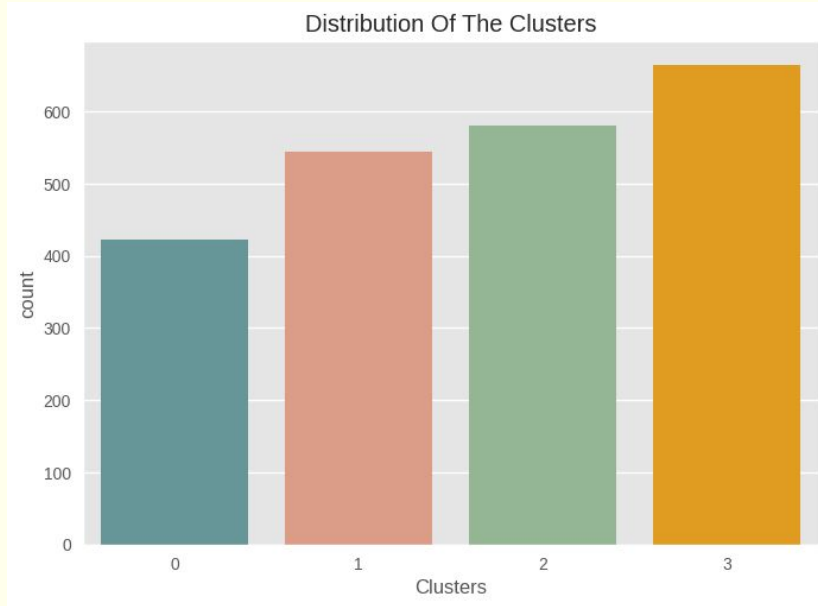


Data Modelling



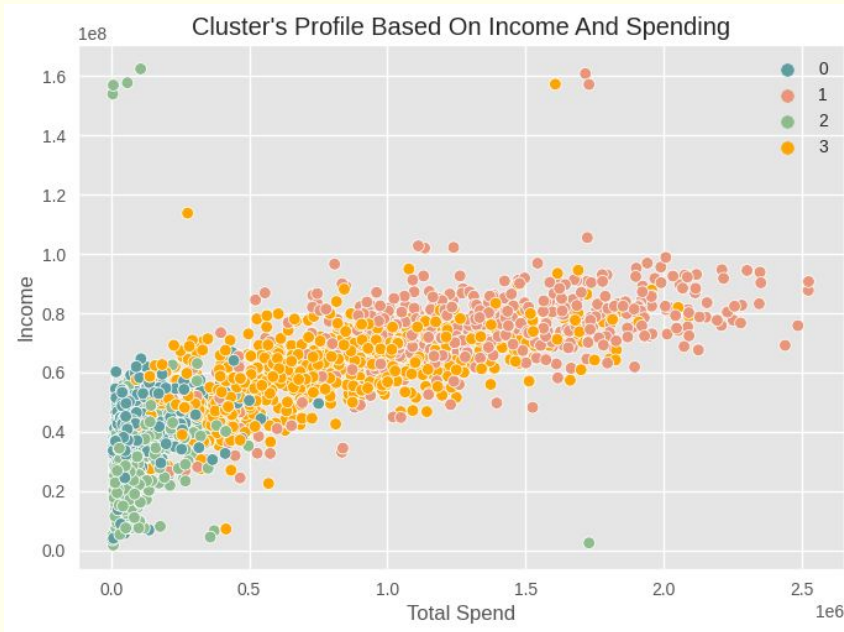
By using Elbow Method, we now know that the best number of cluster for this analysis is 4.

Model Evaluation



The clusters is fairly distributed.

Model Evaluation

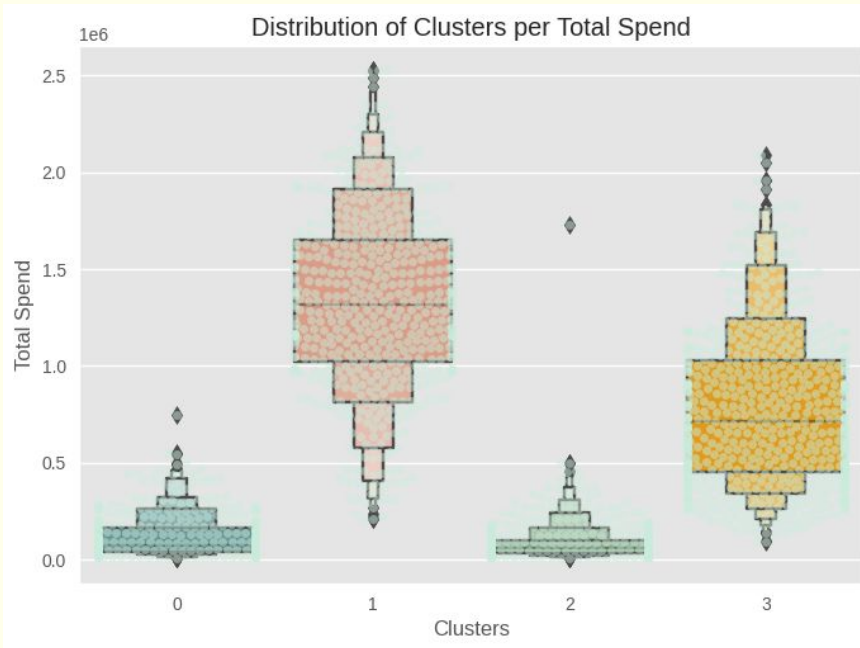


Income vs Spending plot shows the clusters pattern

- Group 0: Average Income - Low Spend
- Group 1: High Income - High Spend
- Group 2: Low Income - Low Spend
- Group 3: High Income - Average Spend



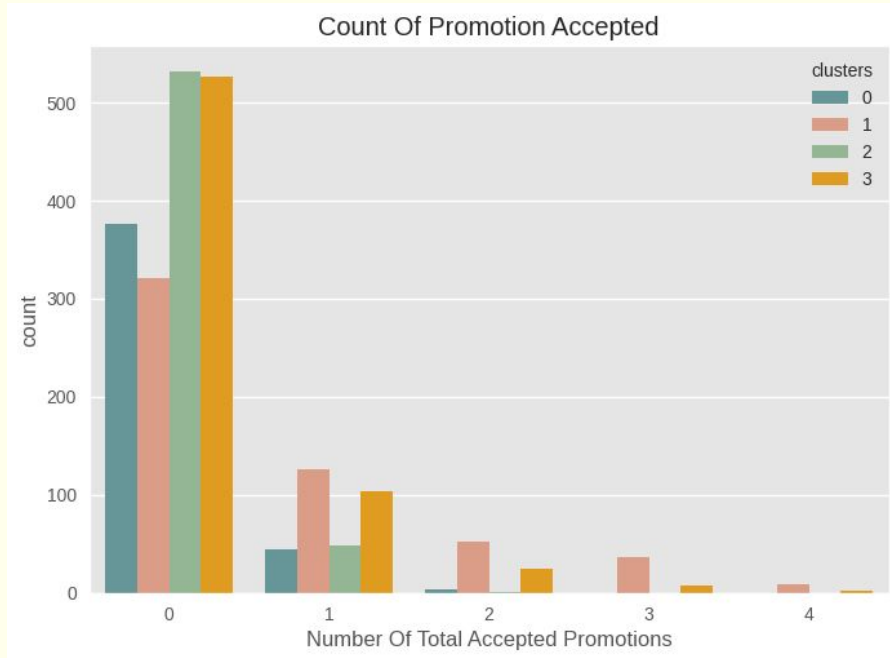
Model Evaluation



From the plot, it can be seen that cluster 1 is our biggest set of customers closely followed by cluster 3. We can explore what each cluster is spending on for the targeted marketing strategies.



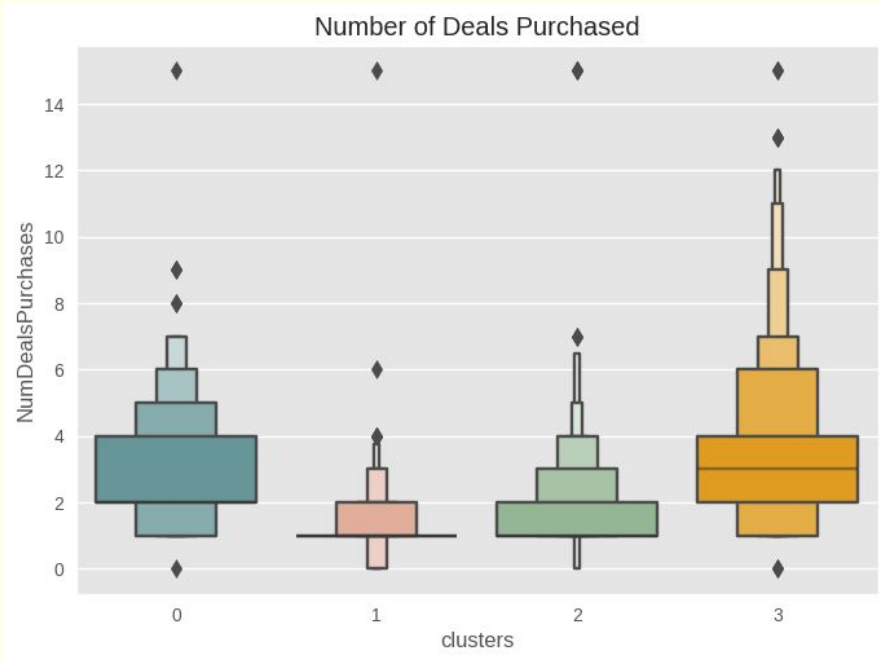
Model Evaluation



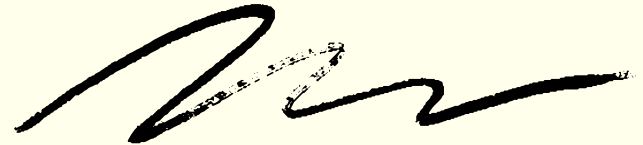
There has not been an overwhelming response to the campaigns so far. There are only a few people accepted the campaign(s). Moreover, no one part take in all 5 of them. Therefore, better-targeted and well-planned campaigns are required to boost sales.

A stylized handwritten signature in black ink, followed by a small black star icon.

Model Evaluation



Unlike campaigns, the deals offered did well. It has best outcome with cluster 3 and 0. However, our star customers cluster 1 are not much into the deals.



Profiling



Group 0:

- Definitely a parent with 1 kid and/or 1 teen.
- Either an adult or senior. But mostly adults.
- Majority has 2 children and some has 3.
- Household capacity mainly 4 and some has 5 family members.



Group 2:

- Majority is a parent with 1 or 2 children.
- Majority has a kid in the house.
- Span in all ages but mainly an adult.
- Member of the household are mainly 2, with the maximum of 4.



Group 1:

- Not a parent.
- Either an adult or senior.
- Majority live alone and some has 2 members in the household.



Group 3:

- Definitely a parent, with a majority of a teen in the house.
- Either an adult or senior.
- Mainly have just 1 child.
- Mainly have 2 or 3 family members, and some have 4.



THANK



YOU!



Am

