Customer Segmentation

I - Introduction

a) Problem Statement

The marketing team is struggling to determine how to create personalized campaigns according to customers' needs. As the data team, our goal is to gain insights from the customer data we have and eventually cluster them accordingly to help the marketing team develop more personalized campaigns.

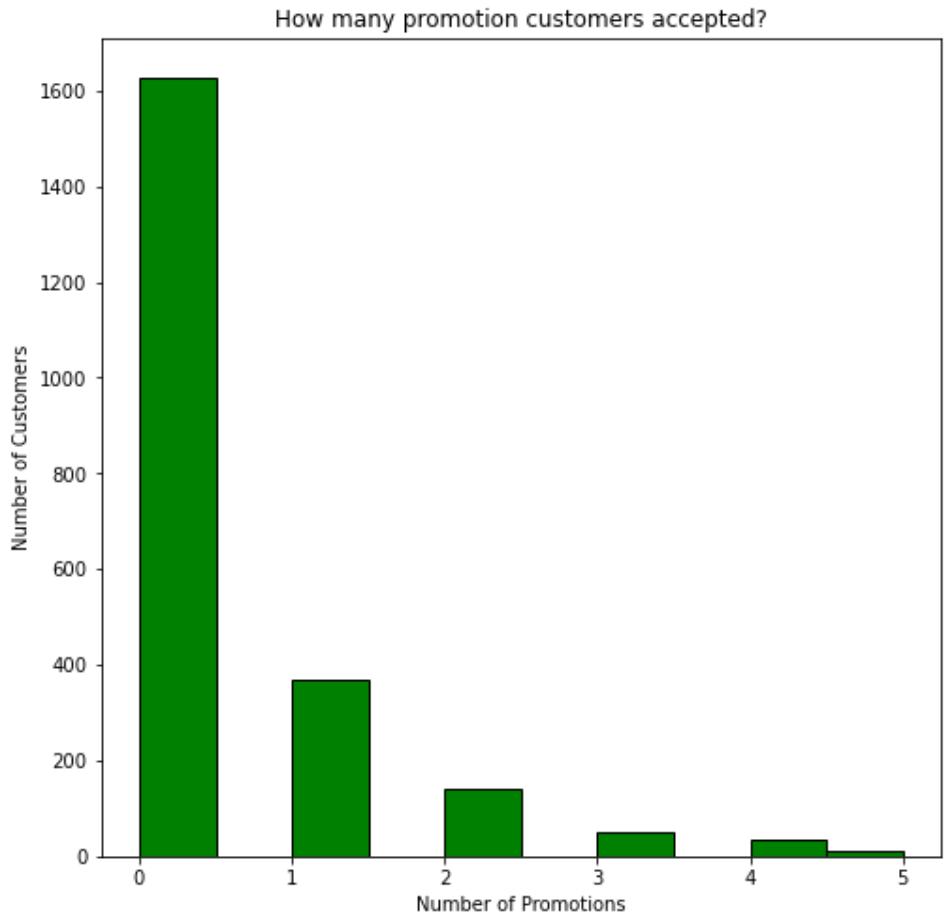
II - Data Wrangling

This project used customer data from Kaggle. The dataset contains 2240 customers and 29 features. Initially, some cleaning needed to be done to make the dataset more coherent. The income column had some missing values so they were filled with the mean of this column.

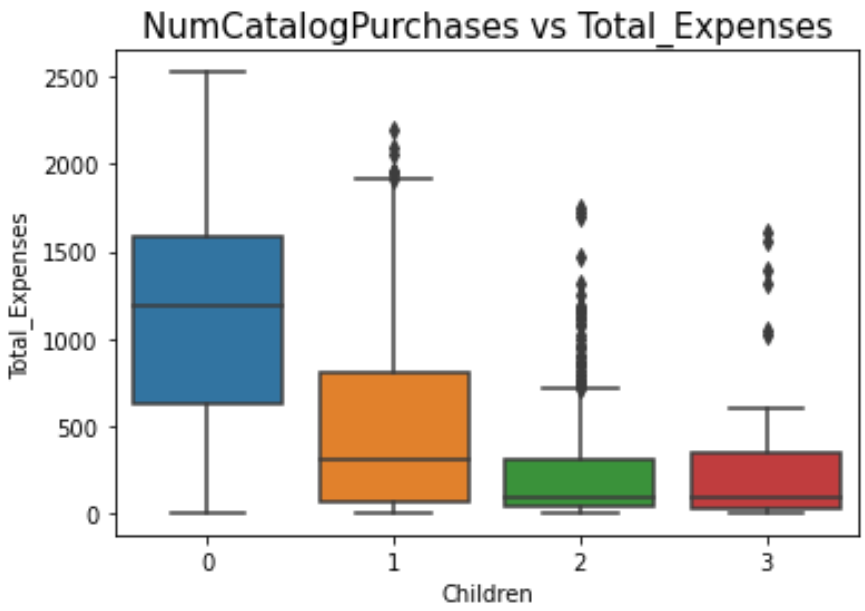
Also, some feature engineering was made. Expenses on different products were combined under the new column Total_Expenses. The age column was generated using customers' date of birth information. Values in Education and Relationship Status columns were edited to ease the analysis.

Finally, Age and Income column had some outliers which were removed to achieve more statistically significant results.

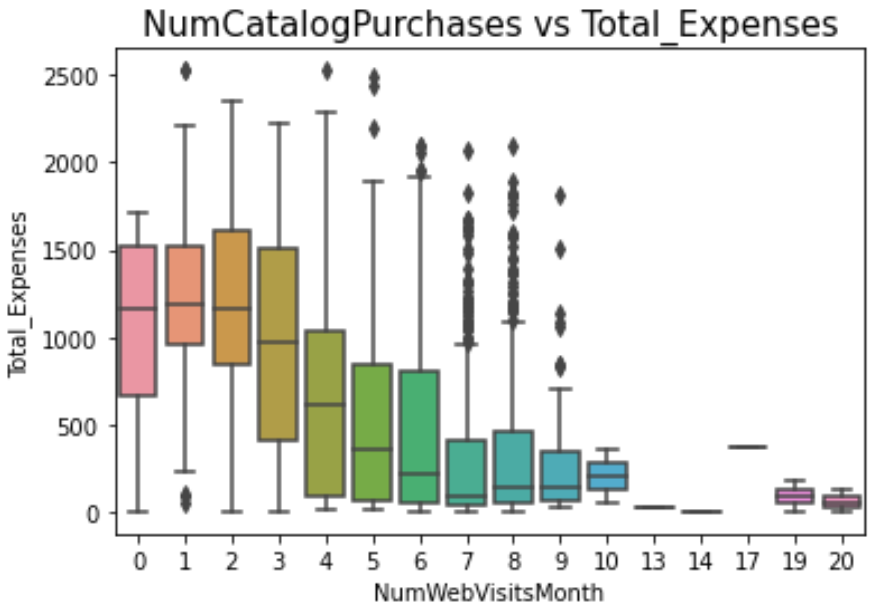
III - Exploratory Data Analysis



The graph indicates that %72.8 of customers didn't accept any promotion at all, 16.5% of customers accepted a promotion in the first offer. This shows that current promotions are not effective. Our goal is to increase this effectiveness by creating more personalized and data-driven campaigns.



Interestingly, there is a negative correlation between the number of children in a household and total expenses. The reason behind this might be having products more suitable for single people.



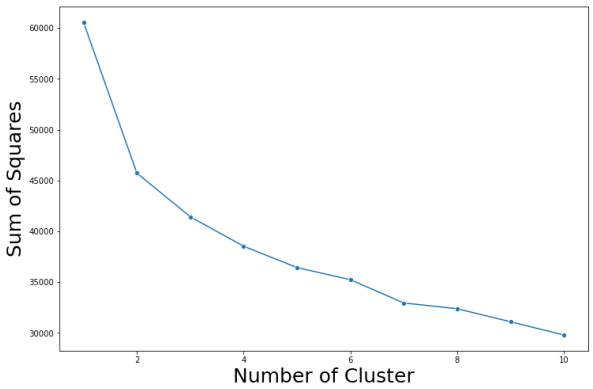
Another surprising correlation is between the number of monthly web visits and total expenses. Customers visiting the company's website more, spend less on the products. The reason may stem from ineffective online sales and marketing strategies.

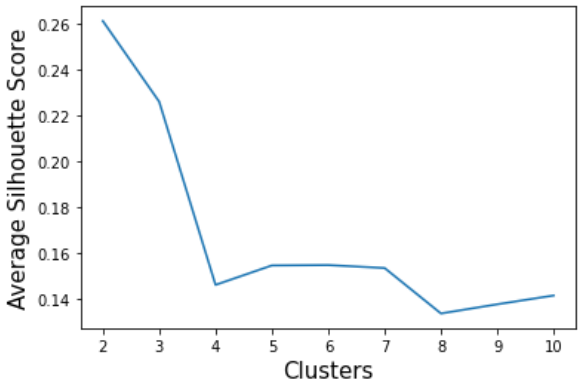
IV - Modeling

After getting insight and interesting information from data in EDA, the next step was applying unsupervised learning algorithms to form clusters.

After comparing silhouette scores for k-means and hierarchical clustering, I found that k-means performs better and decided to proceed with it.

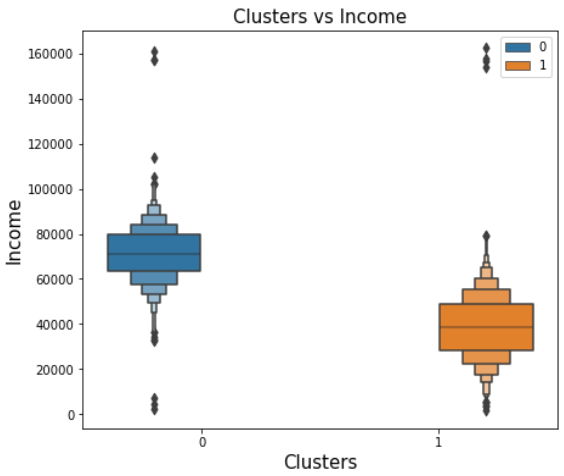
To find out what is the optimal k value for my algorithm, I used elbow and the silhouette methods.

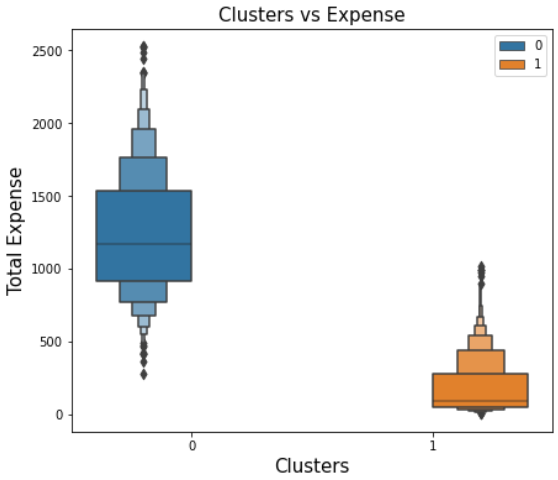


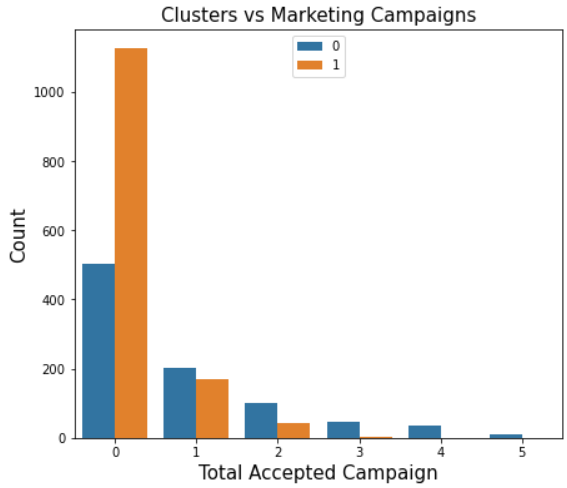


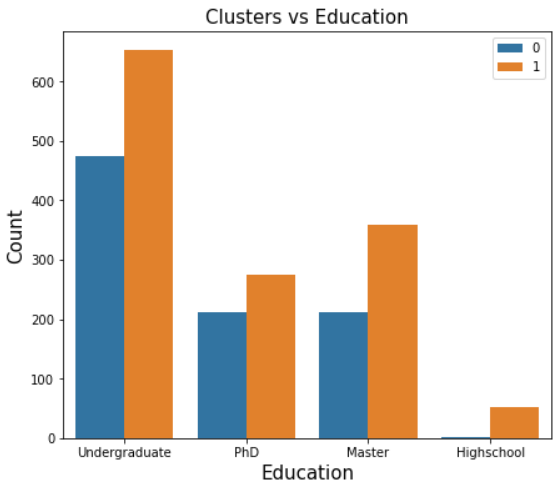
Both graphs indicated that 2 is the optimal number of clusters.

After clustering customers, I explored different features to decide what are the most important features we should focus on to produce more personalized marketing campaigns.

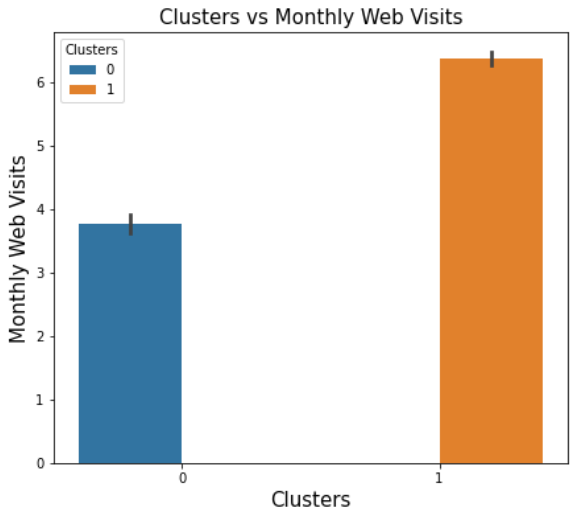


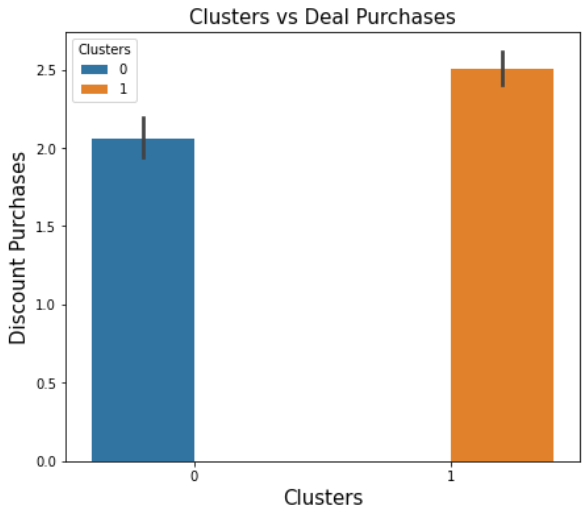



Clusters vs Marketing Campaigns
0
1
Count
Total Accepted Campaign

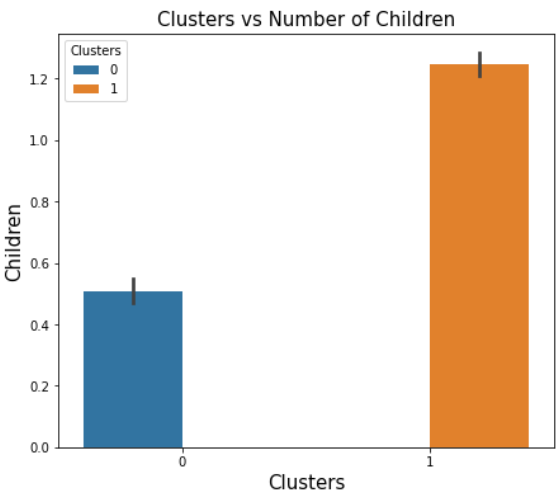

Clusters vs Education
0
1
Count
Undergraduate
PhD
Master
Highschool
Education

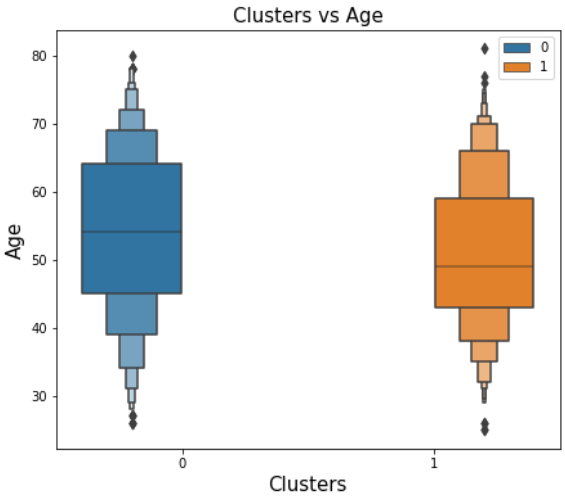
Clusters vs Web Purchases
Clusters
0
1
Web Purchases
Clusters


Clusters vs Monthly Web Visits
Clusters
0
1
Monthly Web Visits
Clusters


Clusters vs Deal Purchases
Clusters
0
1
Discount Purchases
Clusters

Clusters vs Store Purchases
Clusters
0
1
Store Purchases
Clusters


Clusters vs Number of Children
Clusters
0
1
Children
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
Clusters


Clusters vs Age
0
1
Age
30
40
50
60
70
80
0
1
Clusters

Cluster 0	Cluster 1
High socioeconomic status	Lower socioeconomic status
Very high total expenditure	Significantly lower total expenditure
Moderate marketing campaign acceptance rate	Low marketing campaign acceptance rate
More web purchases	Fewer web purchases
Less monthly web visits	More monthly web visits
Fewer discount purchases	More discount purchases
More store purchases	Fewer store purchases
Slightly more educated	Slightly less educated
Relatively older population	Relatively younger population

V - Conclusion

Cluster 0 - Marketing Strategy

- Promoting company's website
- Emailing the information of newly arrived items
- Creating customer loyalty rewards program
- Ask them why they are choosing our products

Cluster 1 - Marketing Strategy

- Increase the customer experience on the website and make the purchasing process less time consuming
- Promote the discounted items by sending emails and displaying them in the main page of the website
- Advertise how our products different from the competitors by using feedback of Cluster 0