

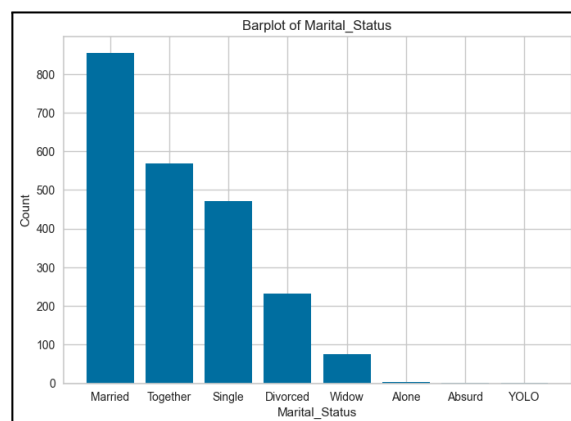
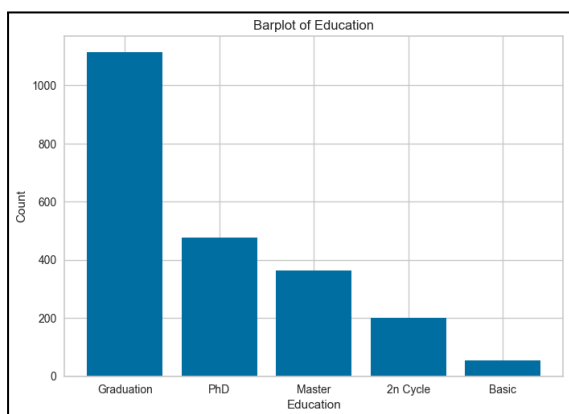
# Customer Segmentation Analysis Report

## Introduction

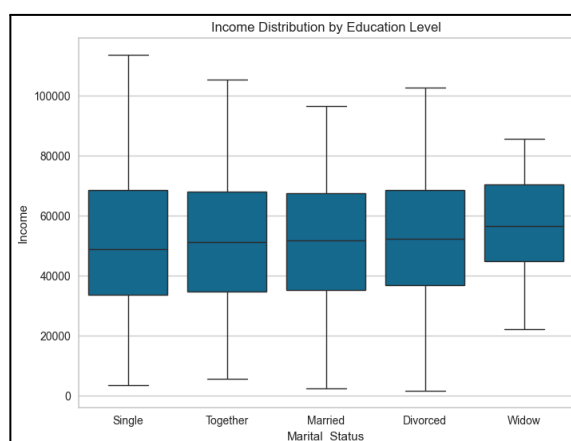
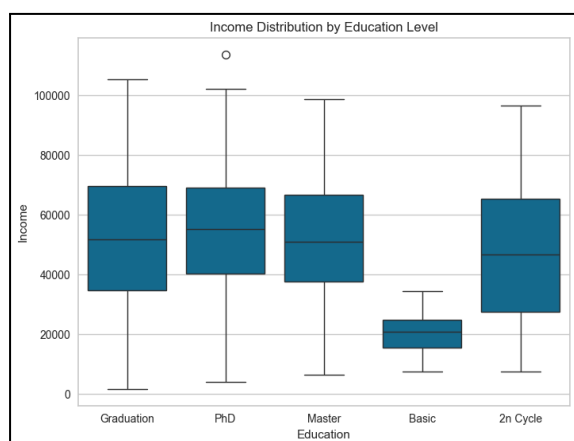
Customer segmentation is a critical process for businesses that enables more effective marketing and personalized services. By dividing customers into distinct groups based on shared characteristics, businesses can target each group with tailored offers and strategies, improving customer satisfaction and boosting profitability.

## Data analysis

At the start of the project, we examined types of marital status and education. There are 5 types of education and 8 types of marital status. The 'Absurd' and 'YOLO' types of marital status were excluded, and the 'Alone' value was merged with 'Single'.

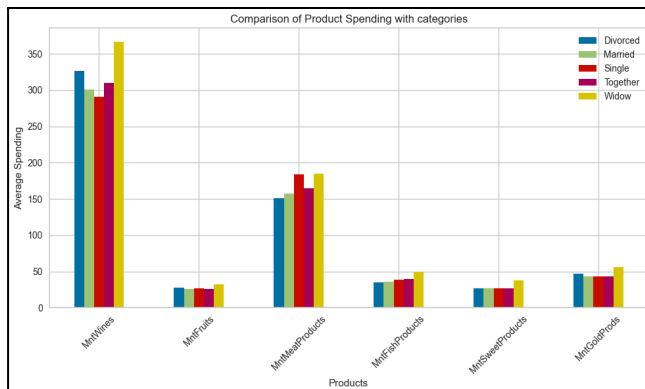
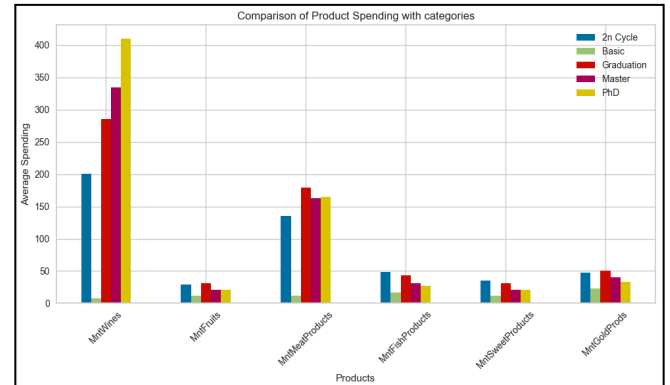
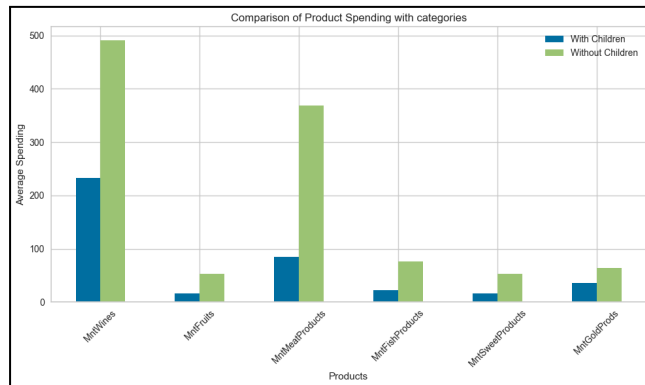


After that, we examined the income of different groups using box plots, and the analysis shows that PhDs and widows have the highest income.



Then, we examined the correlation between income and the number of products bought. The analysis shows a positive correlation between income and other sold products, with the strongest correlation found in sold wines (correlation coefficient of **0.73**). Additionally, there is a high correlation between sold wines and sold meat (correlation coefficient of **0.59**). It seems that customers who buy wine are more likely to purchase meat products as well.

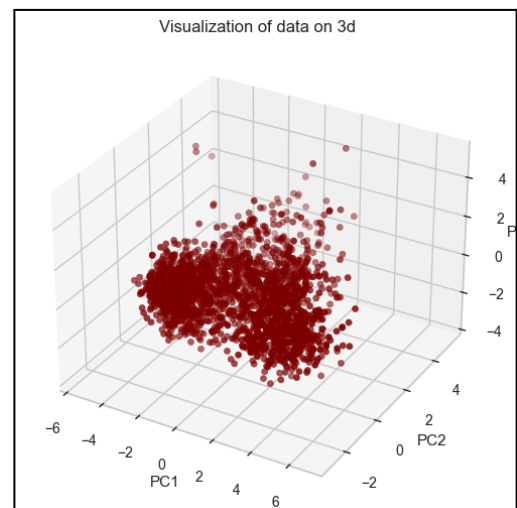
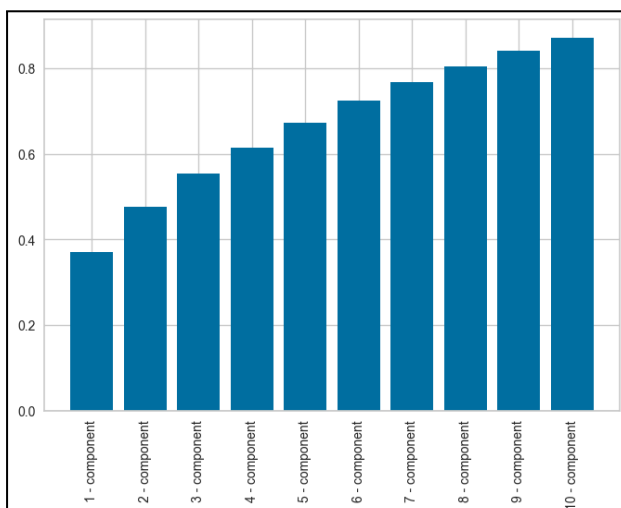
Then, we examined the mean spending amount on products by different groups, and the analysis shows that customers without children spend more. Surprisingly, customers with children purchase fewer sweet products than those without children. PhDs and Masters are the main consumers of wine, while customers with a 'Graduation' education level are the primary consumers of meat products. Widows and divorced customers buy more wine than others.



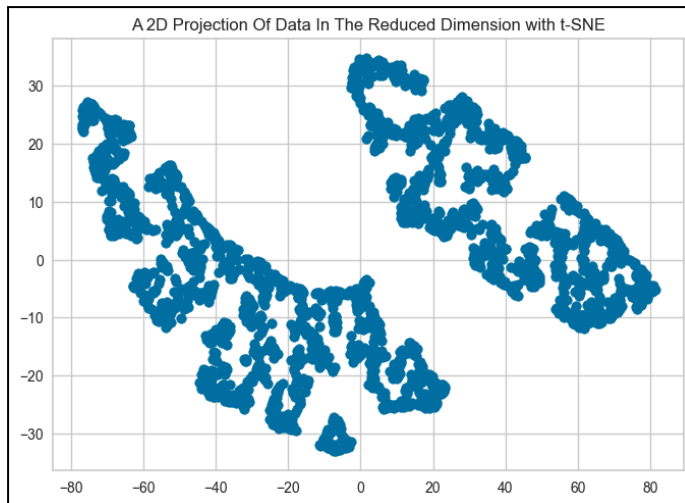
## Dimensionality Reduction

To apply dimensionality reduction algorithms to the data, we first encode the categorical columns. Before performing PCA on the dataset, we standardize the data matrix, ensuring it has a mean of 0 and a standard deviation of 1.

Below, you can see the explained variance ratio of the 10 principal components from the PCA results. It shows that with 3 principal components, it is possible to retain more than 50% of the variance, but the result does not provide meaningful insights about clusters.

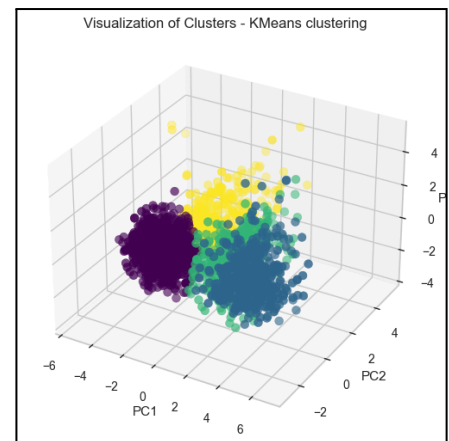
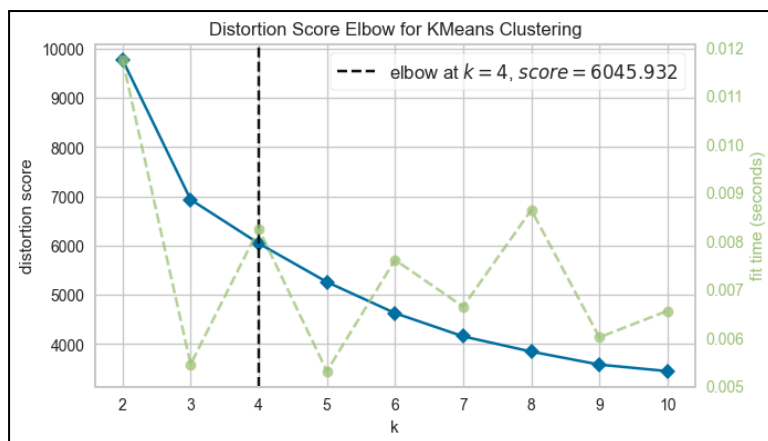


After that, we apply the t-SNE dimensionality reduction algorithm with 2 components to the dataset, and it clearly reveals the clusters.

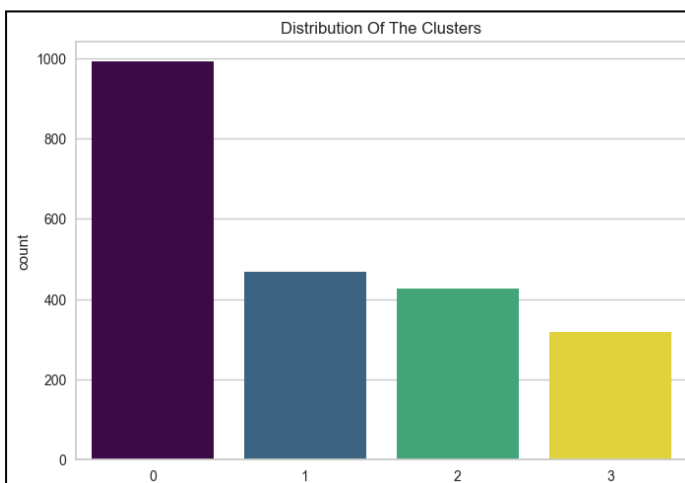


## Clustering

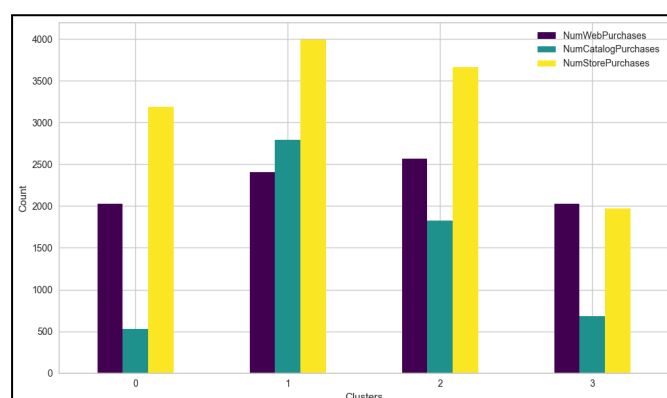
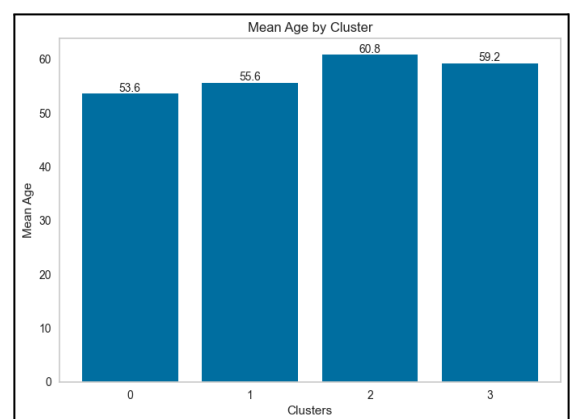
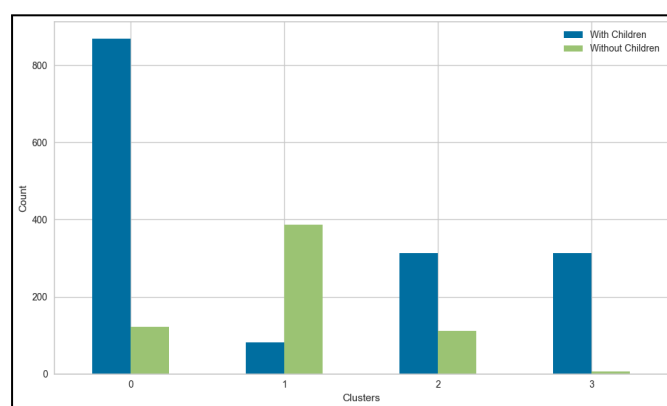
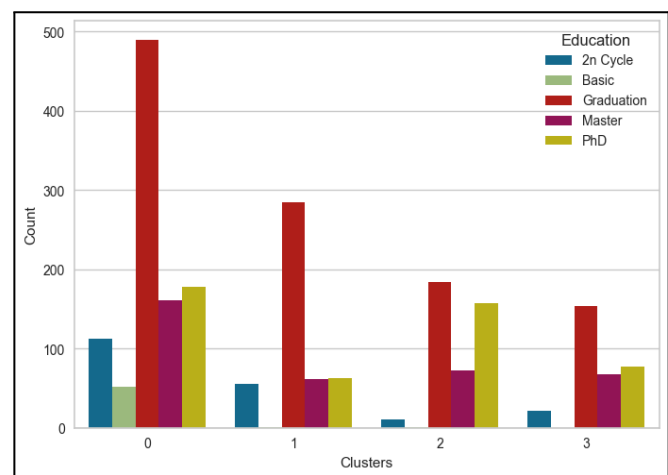
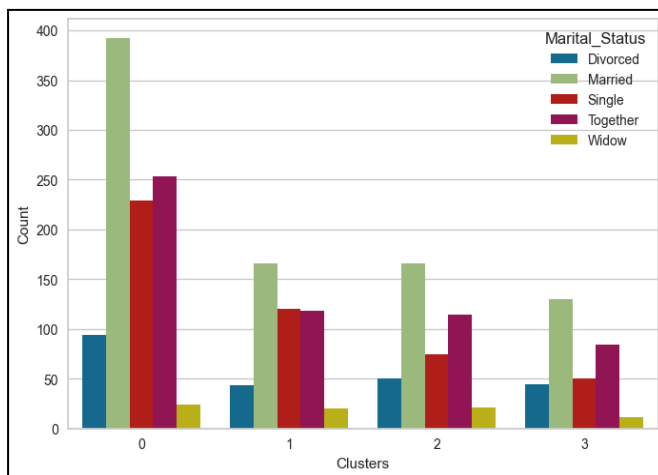
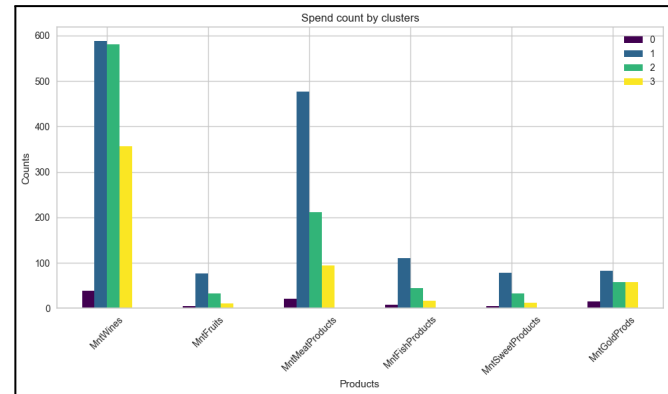
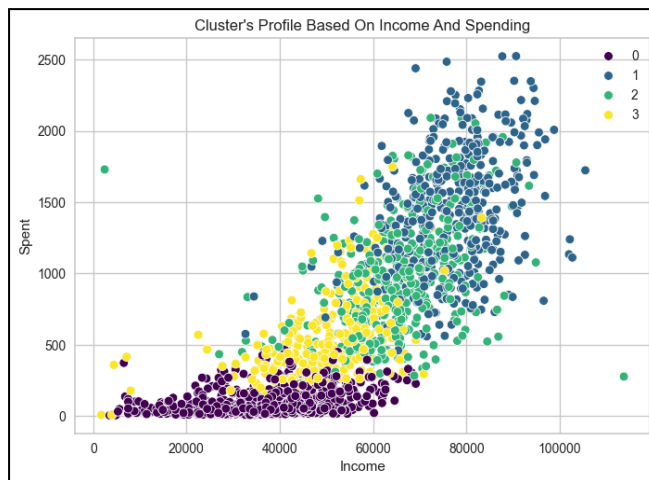
For clustering, we used the k-means algorithm with k-means++ initialization on the PCA results. To determine the optimal number of clusters, we applied the elbow method, which indicates that there are 4 effective clusters.



Below, you can find the distribution of the clusters in the dataset. Cluster 0 is the most frequent one.



After that, we examined various features of the clusters, such as income versus spending rate, the count of each purchased product, the distribution of education levels by cluster, the distribution of marital status levels by cluster, the distribution of parents versus non-parents by cluster, the mean age of each cluster, and the distribution of purchase types by cluster.



## Conclusion

---

### Cluster 0

- Low income, low spending
  - Majority: Mostly graduated
  - Mostly married and with children
  - Younger cluster (mean age ~53.6)
- 

### Cluster 1 [Most Valuable]

- High income, high spending
  - Enthusiastic about wine and meat
  - Mostly married and without children
  - Mean age ~55.6
  - Highest catalog purchases, but prefers shopping in stores
- 

### Cluster 2

- Mean income, mean spending
  - Predominantly PhDs and graduates
  - Enthusiastic about wine
  - Mostly parents
  - Older cluster (mean age ~60.8)
  - Prefers shopping in stores
- 

### Cluster 3

- Mean income, low spending
- Primarily parents
- Mean age ~59.2
- Primarily shops on the company website