**Customer Segmentation: Customer personality analysis**

Customer segmentation is the process of dividing a company's customer base into smaller groups with common characteristics, based on demographic, behavioral, or other variables. It is a powerful tool for understanding the needs, preferences, and behaviors of different groups of customers, and for designing targeted marketing strategies.

One aspect of customer segmentation is personality analysis, which involves identifying the personality traits of different groups of customers and using this information to tailor marketing efforts. Personality analysis can be performed using various techniques, such as psychological testing, surveys, or machine learning algorithms.

Below are the steps involved in performing customer personality analysis through customer segmentation:

1. Data Exploration: The first step in any unsupervised learning project is to explore the data. This involves looking at the variables in the dataset and understanding their meanings and relationships.

About the dataset: It consists of 2240 datapoints and 29 columns; this dataset can be categorized into the following four subsets:

Customer’s Information:

* ID: Customer's unique identifier
* Year\_Birth: Customer's birth year
* Education: Customer's education level
* Marital\_Status: Customer's marital status
* Income: Customer's yearly household income
* Kidhome: Number of children in customer's household
* Teenhome: Number of teenagers in customer's household
* Dt\_Customer: Date of customer's enrollment with the company
* Recency: Number of days since customer's last purchase
* Complain: 1 if the customer complained in the last 2 years, 0 otherwise

Products : Amount spent on each product in last 2 years

* + MntWines
  + MntFruits
  + MntMeatProducts
  + MntFishProducts
  + MntSweetProducts
  + MntGoldProds

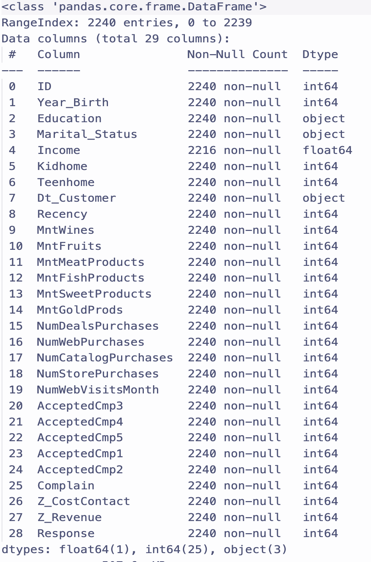
Promotion:

* + NumDealsPurchases: Number of purchases made with a discount
  + AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
  + AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
  + AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
  + AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
  + AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
  + Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

Place:

* NumWebPurchases: Number of purchases made through the company’s website
* NumCatalogPurchases: Number of purchases made using a catalogue
* NumStorePurchases: Number of purchases made directly in stores
* NumWebVisitsMonth: Number of visits to company’s website in the last month

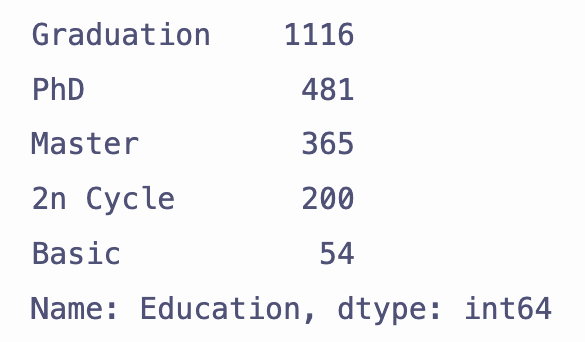
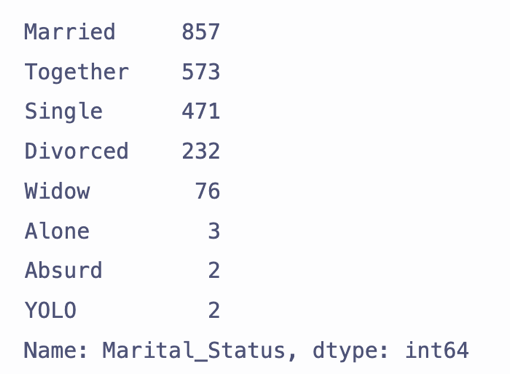
1. Data Cleaning & Feature Engineering: The next step is to clean and prepare the data for analysis. This may involve removing missing or irrelevant data, and creating new features or variables that may be more relevant to the analysis.
2. Missing Values in Income: drop missing values: (2216 rows after removing nulls)



1. Dt\_Customer is object, needs to convert to datatime.

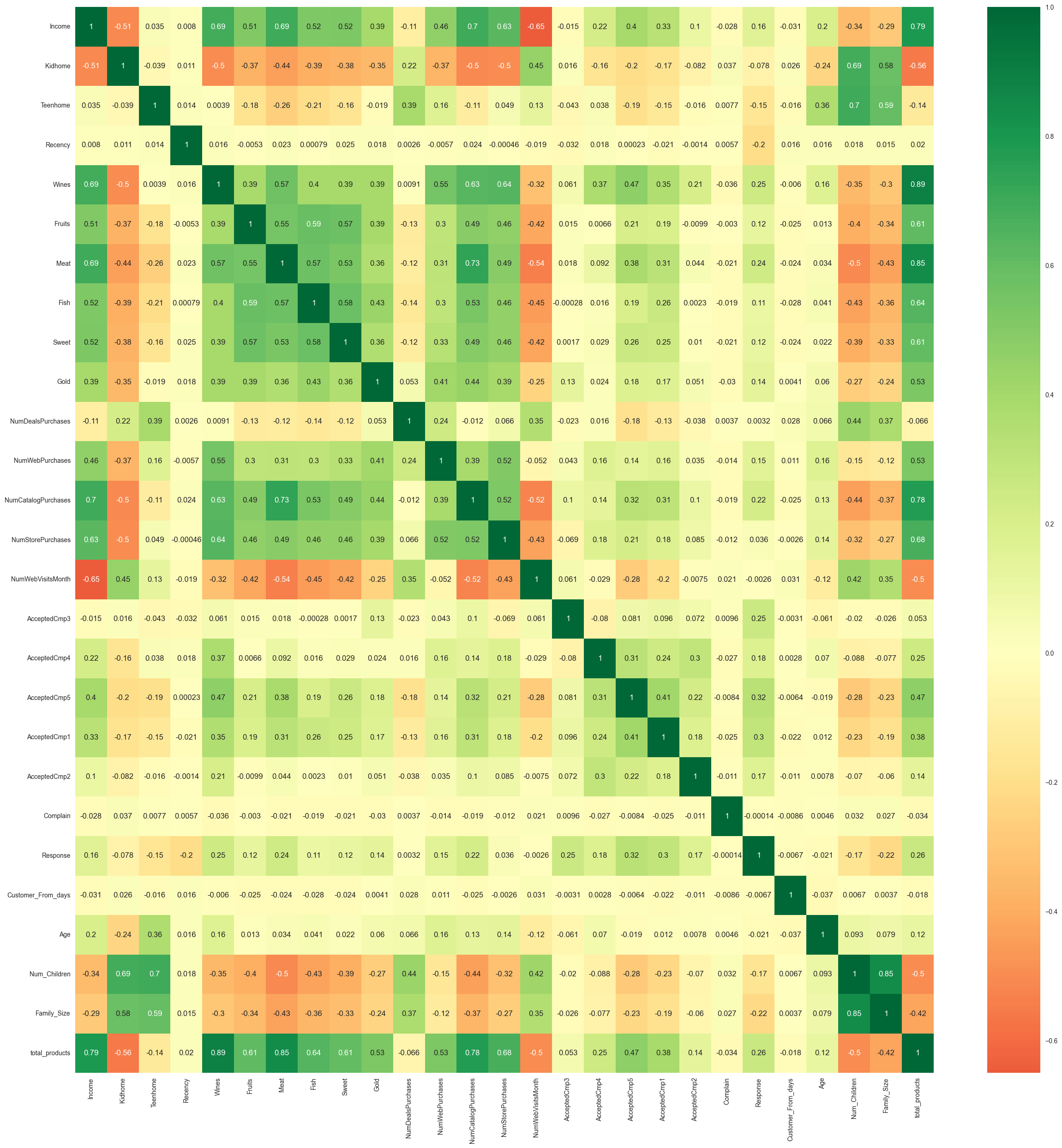


1. Exploring what are the values in the categorical features (Education, Marital\_Status)



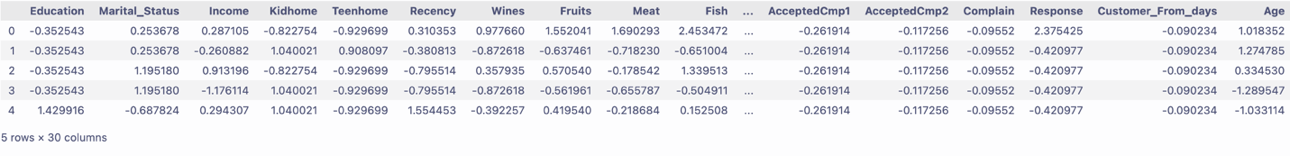
Feature Engineering:

* + Add new features and modify some features to better clarify the data
    - Customer\_From\_days; Age, Marital\_Status, Living\_with, Num\_Children, Family\_Size
  + Remove outliers in age and income.
  + Heatmap for the correlation matrix after adding new features and outliers are removed.

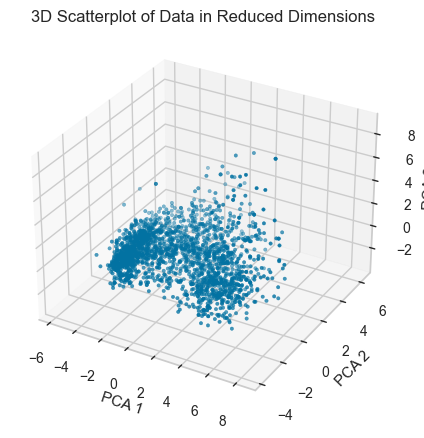


1. Data Preprocessing: Once the data is clean and ready for analysis, it needs to be transformed and standardized to be used in a machine-learning model. This may involve scaling the data, normalizing it, or applying other transformations.

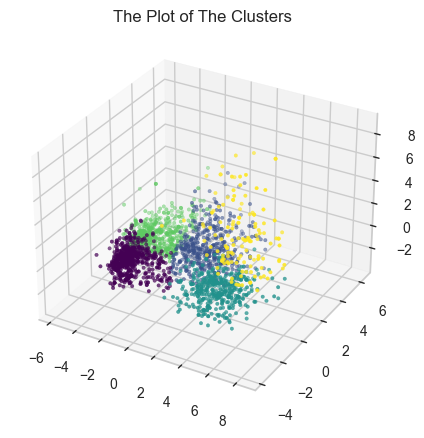
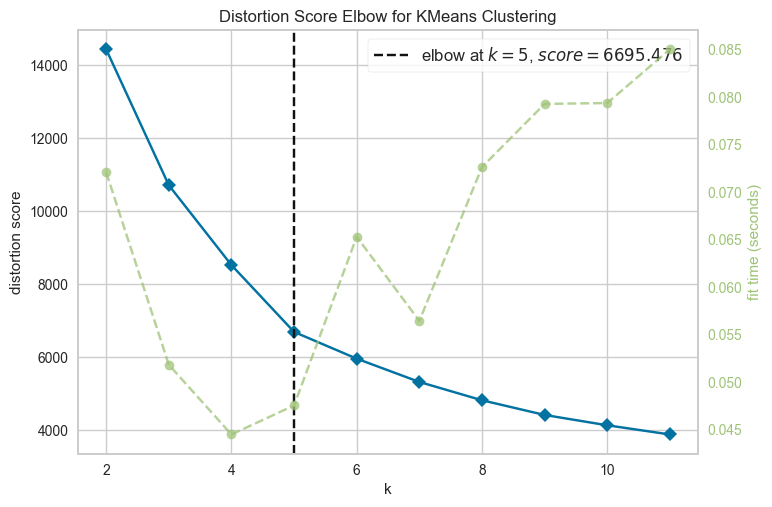
Data after StandardScaler



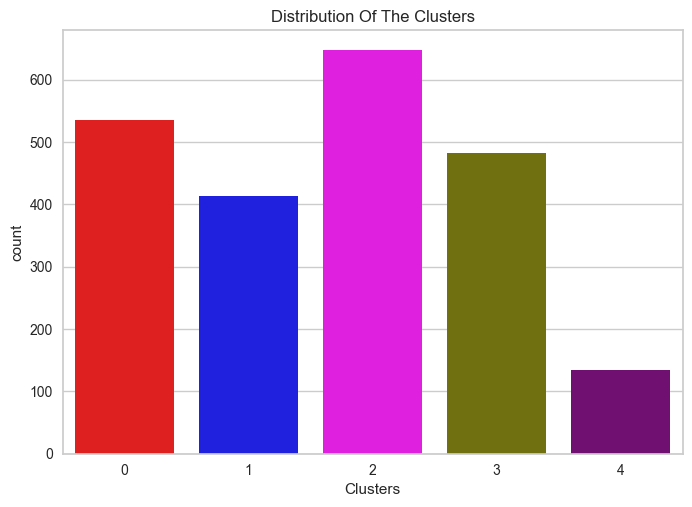
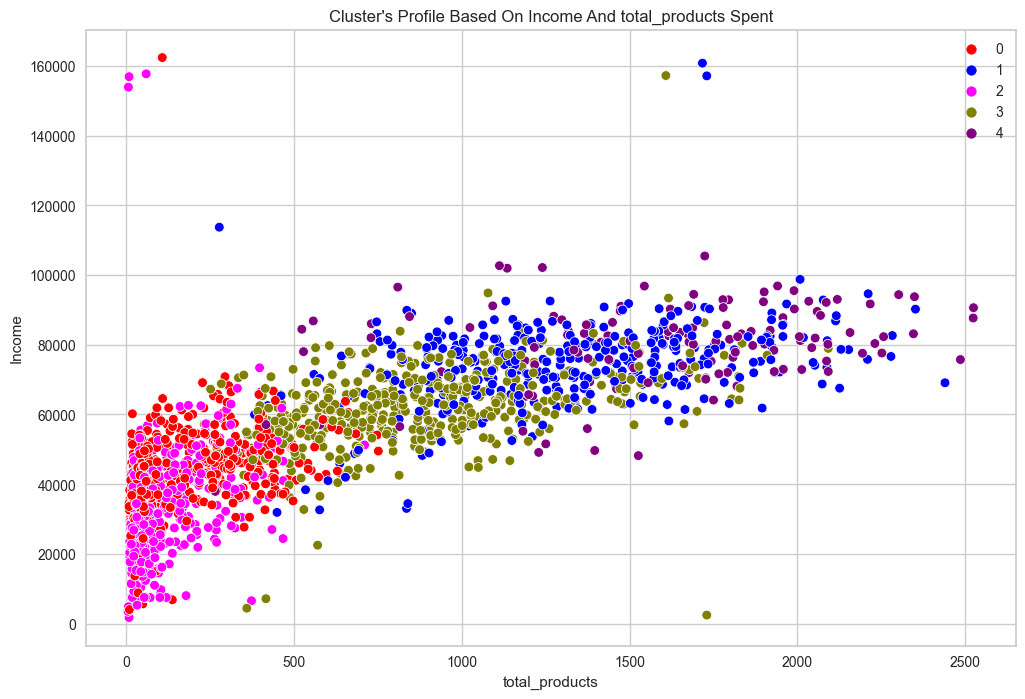
1. Dimensionality Reduction: In some cases, the dataset may contain a large number of variables, which can make it difficult to analyze. To simplify the dataset, it may be helpful to use dimensionality reduction techniques such as principal component analysis (PCA) to reduce the number of variables.

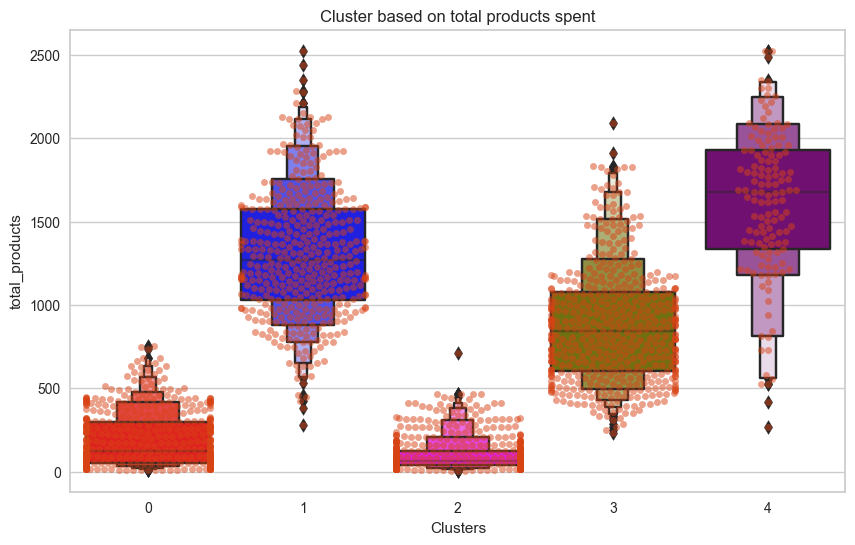
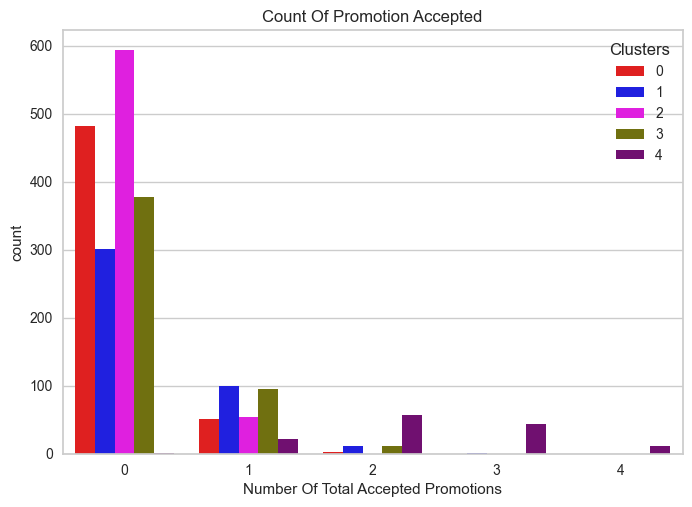


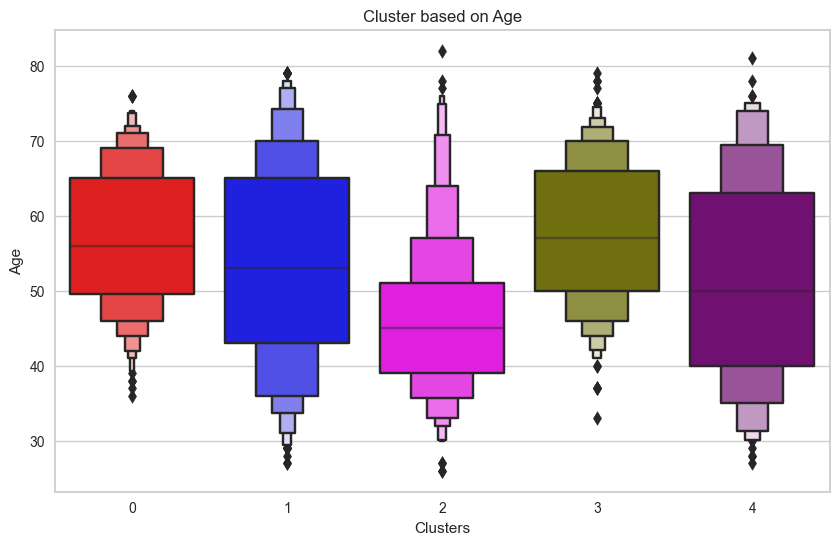
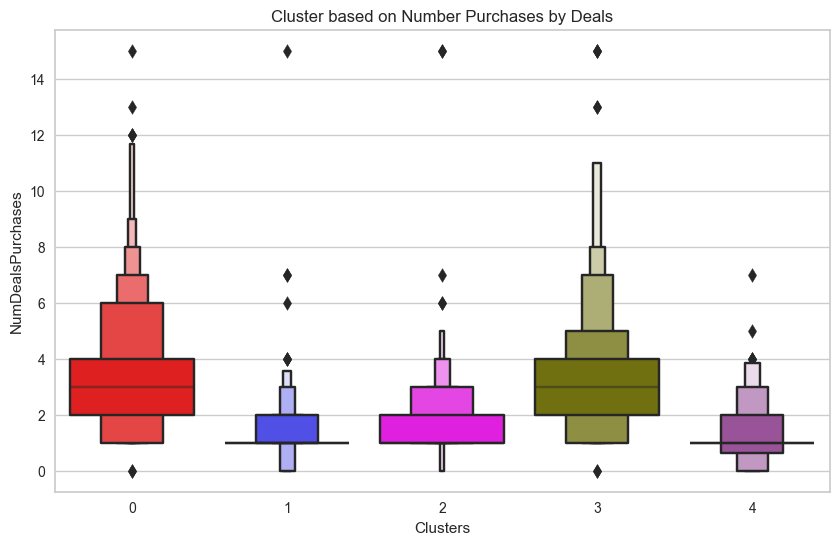
1. Clustering : After the data has been preprocessed, it can be used to perform clustering, which involves grouping the data into clusters based on their similarities. There are various clustering algorithms that can be used, such as k-means, hierarchical clustering, or density-based clustering.

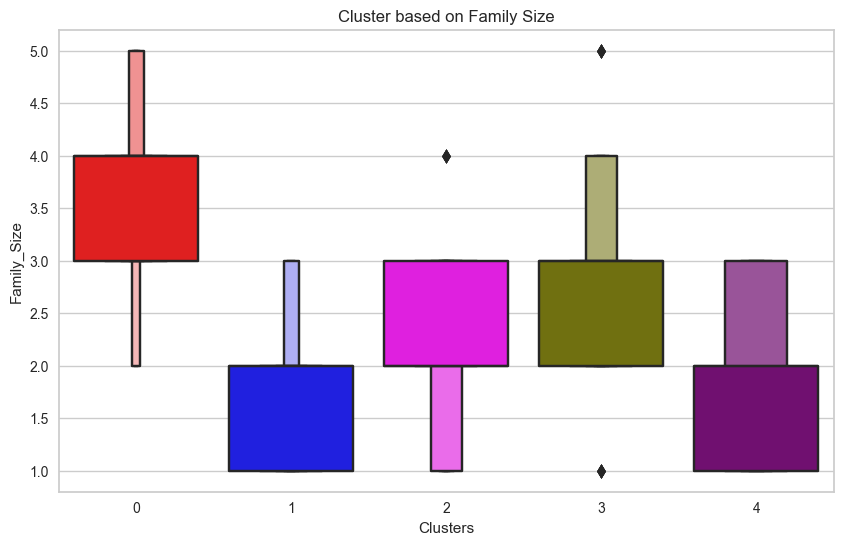


1. Evaluating Models: Evaluating the performance of the clustering model is a crucial step in customer personality analysis through customer segmentation. This involves using various metrics to measure the accuracy and meaningfulness of the model. Below are the steps I used to evaluate the model.

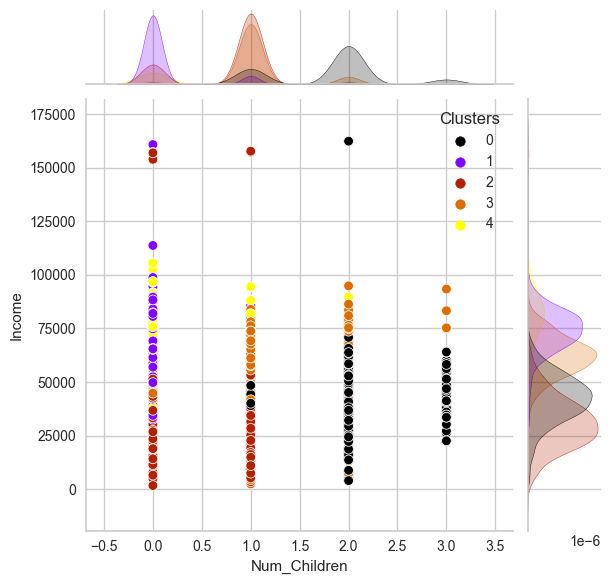
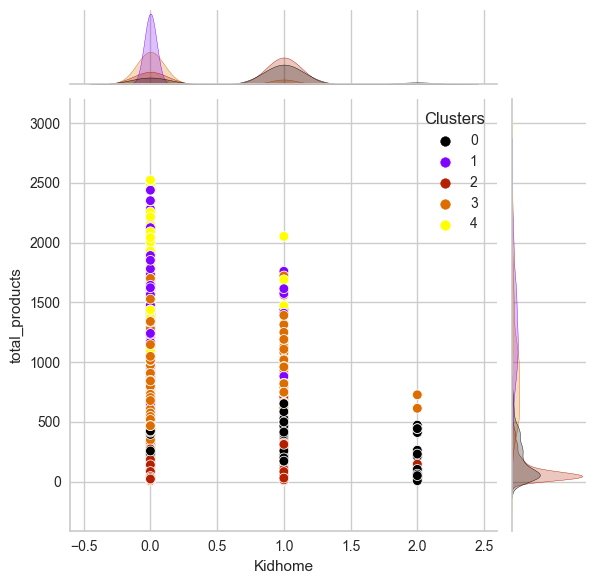
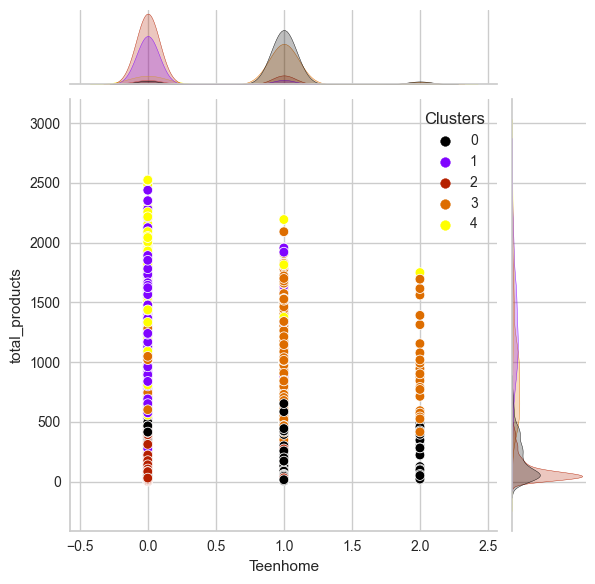
 

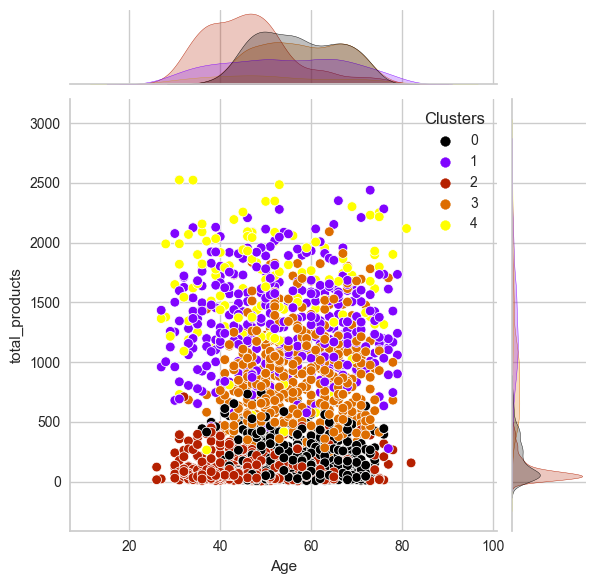
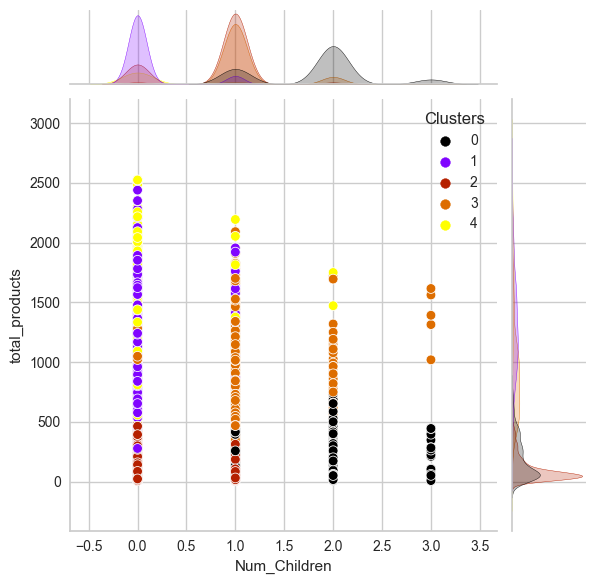
 

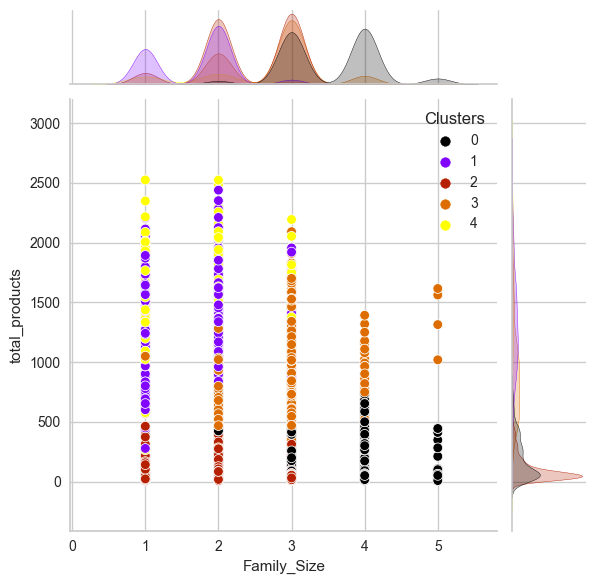
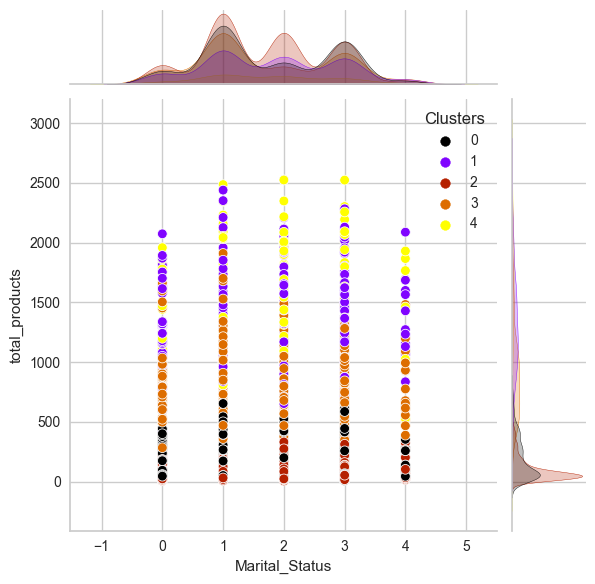
 



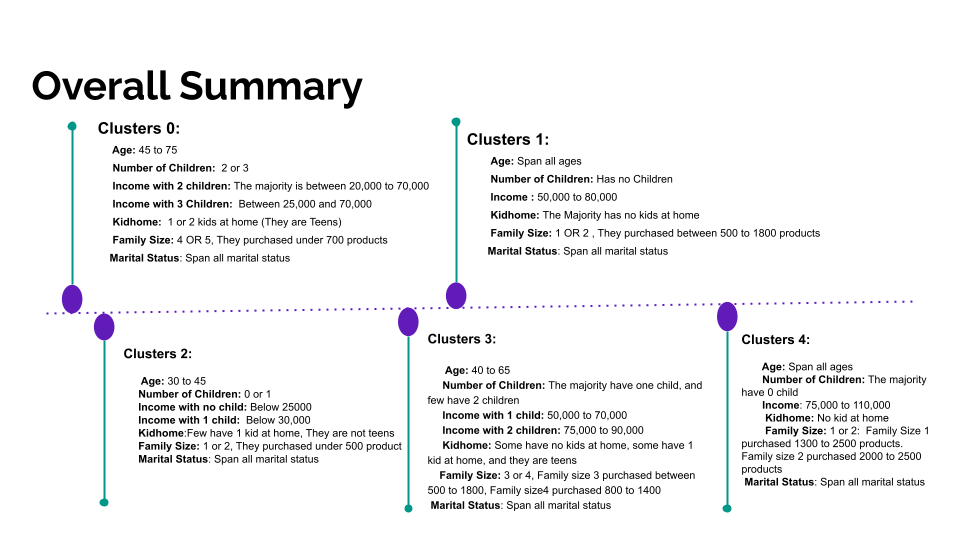
1. Profiling: The final step is to profile the clusters and identify the characteristics of each group. This may involve analyzing each cluster's demographic, behavioural, or other variables to understand the needs and preferences of the customers in that group.

Overall Summary:



In conclusion, customer personality analysis through customer segmentation is a powerful tool for understanding and targeting different groups of customers. By following the steps outlined above, companies can gather and analyze data, reduce dimensionality, cluster customers, and evaluate and profile the resulting clusters to better understand the personalities of their customers.