



Samsung Innovation Campus

| Artificial Intelligence Course

Customer Personality Analysis

Presented by:

Omar Khan

Supervised by:

D. Doaa Mahmoud Abdel-aty

Eng. Shimaa Othman

Data used

Agenda

1. Objective.
2. Business Questions
3. Data Description
4. EDA.
5. Data Preprocessing.
6. Clustering.
7. Evaluation Models.
8. Conclusion



Objective.



The data that we will analyze in this presentation is data and information for the clients of a particular company that wants to analyze the personality of the clients.



This analysis helps businesses better understand their customers and makes it easier for them to adjust products according to the specific needs, behaviors, and interests of different types of customers.

Business Questions .



1. What is the relationship between the client's joining date of the company and marital status?
2. Does marital status affect the number of purchases?
3. Is the size of the family and the number of children related to the number of purchases?
4. What does age have to do with the number of purchases?
5. Do clients with better education get more income?
6. What is the relationship between income, family size and number of children?
7. Do customers with more income make more purchases?
8. Do customers who visit the website make purchases through it or not?
9. What is the most popular way for customers to buy at a discount?
10. Are the promotions campaigns carried out by the company successful?
11. Are there complaints from customers, if yes, are they old or new customers?



Data Description



Rows

→ 2240

columns

→ 29 Before →

1-ID 2-Year_Birth 3-Education 4-Marital_Status
5-Income 6-Kidhome 7-Teenhome 8-Dt_Customer
9-Recency 10-MntWines 11-MntFruits
12-MntMeatProducts 13-MntFishProducts
14-MntSweetProducts 15-MntGoldProds
16-NumDealsPurchases 17-NumWebPurchases
18-NumCatalogPurchases 19-NumStorePurchases
20-NumWebVisitsMonth 21-AcceptedCmp3
22-AcceptedCmp4 23-AcceptedCmp5
24-AcceptedCmp1 25-AcceptedCmp2 26-Complain
27-Z_CostContact 28-Z_Revenue 29-Response

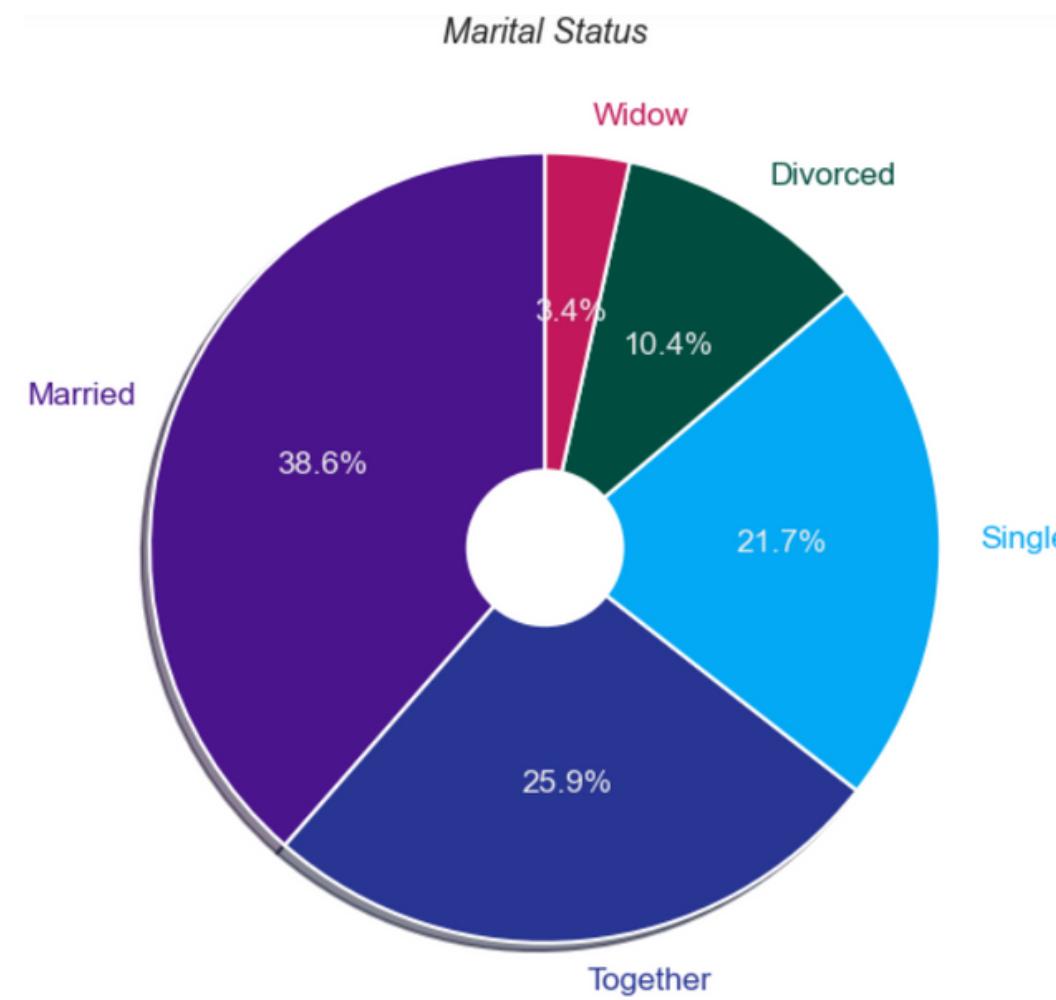
30-Customer_From_days 31-Age
32-Living_With 33-Num_Children
34-Family_Size 35-total_purchases
36-purchase_quantity 37-Total_Promos

columns: 37

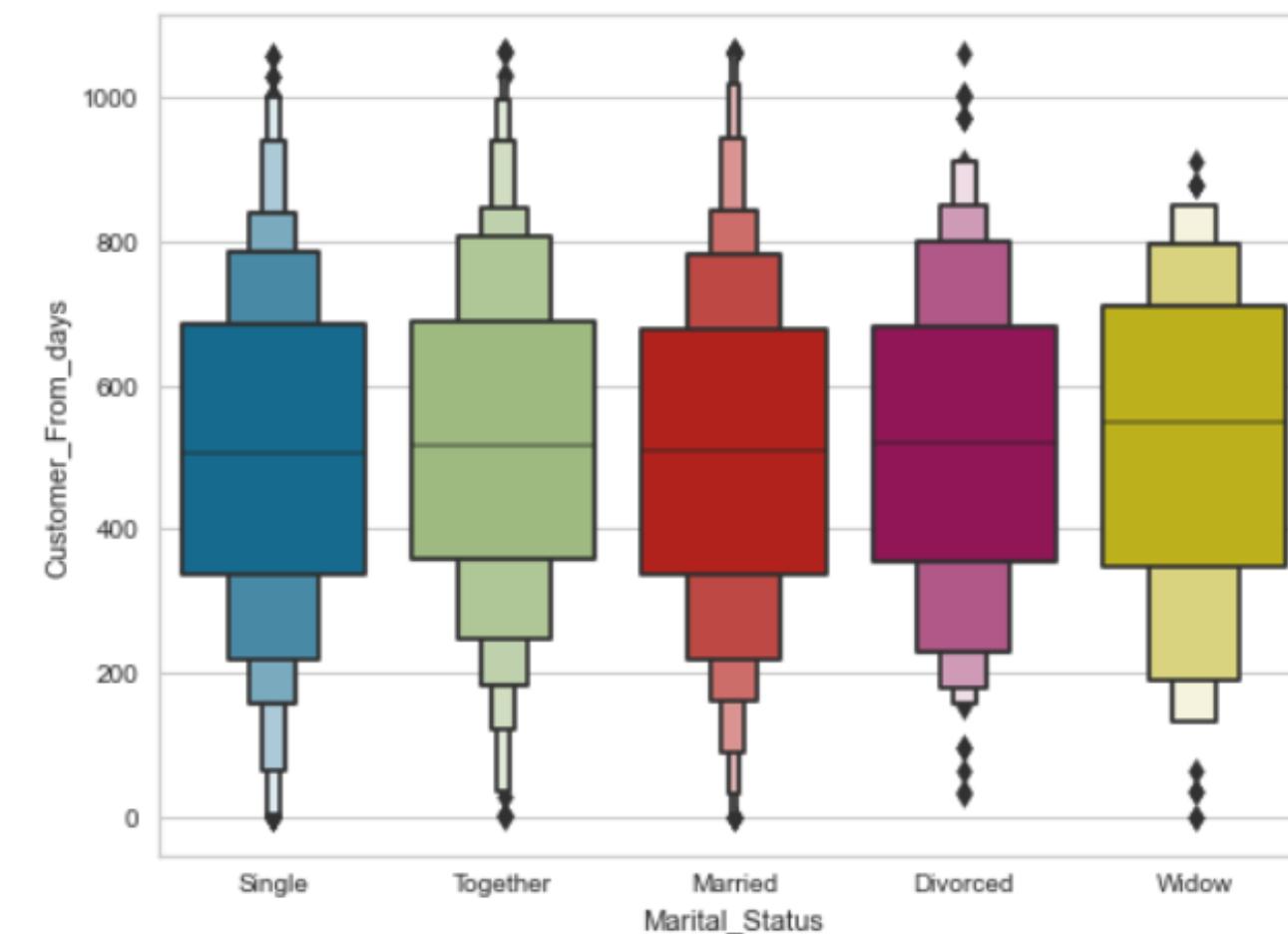
After



- Q1: What is the relationship between the client's joining date of the company and marital status?



The largest proportion of the company's customers are people who live with partners

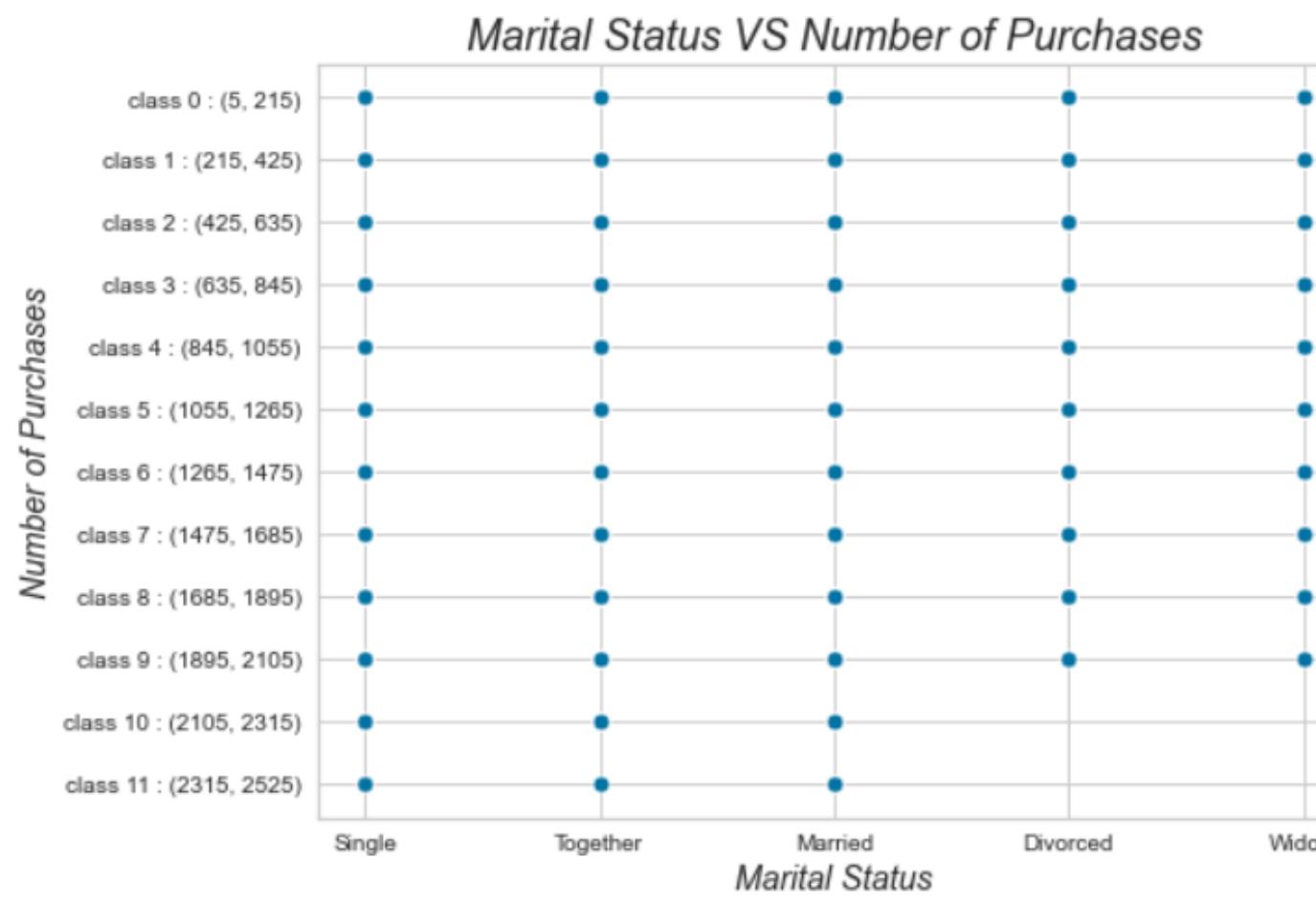
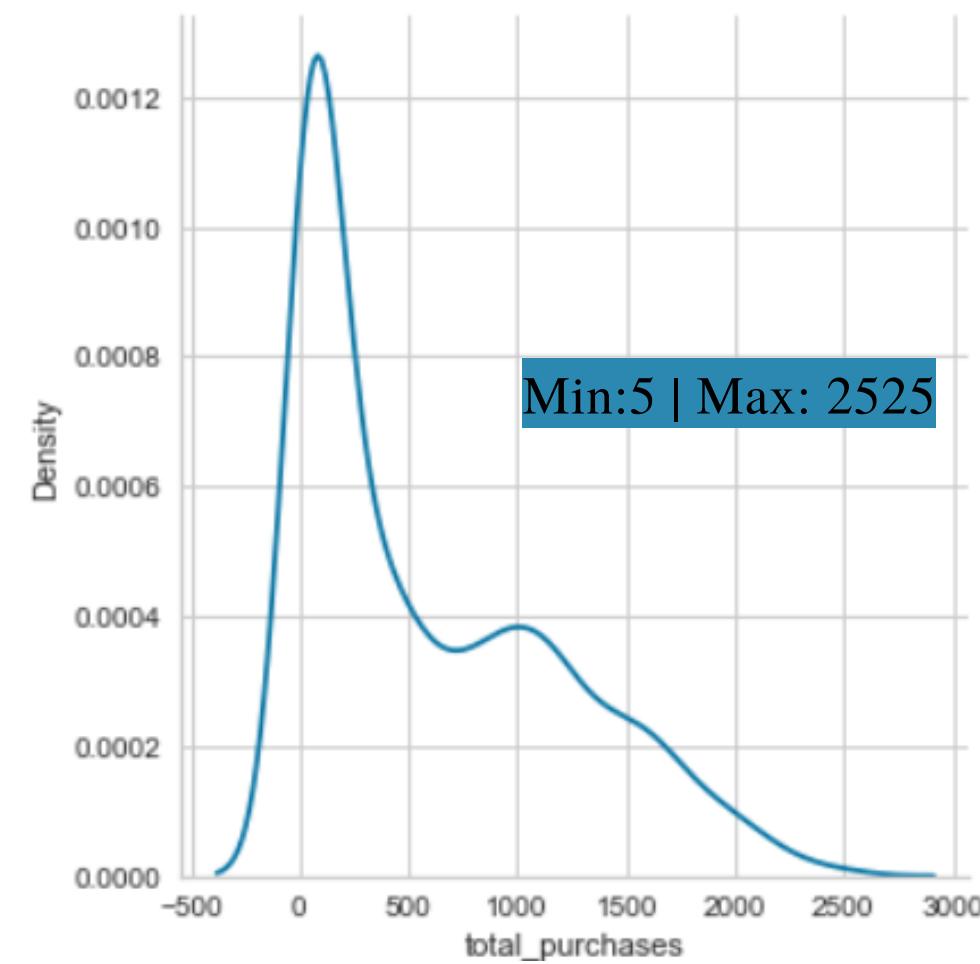


There is no relationship between marital status and the date of joining the company



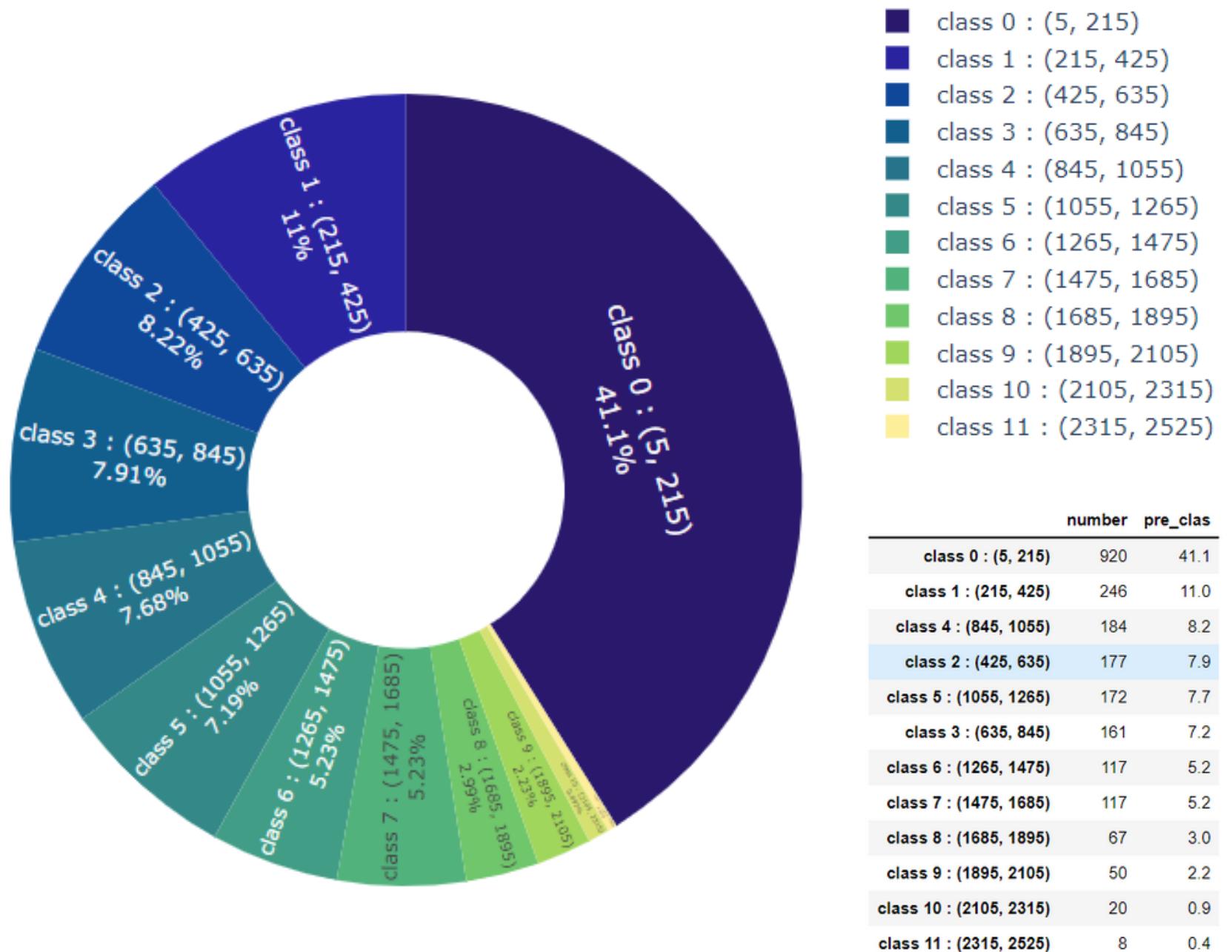
- Q2: Does marital status affect the number of purchases?

Divorced and widowed clients are not included in the categories 10 and 11





- Q2: Does marital status affect the number of purchases?

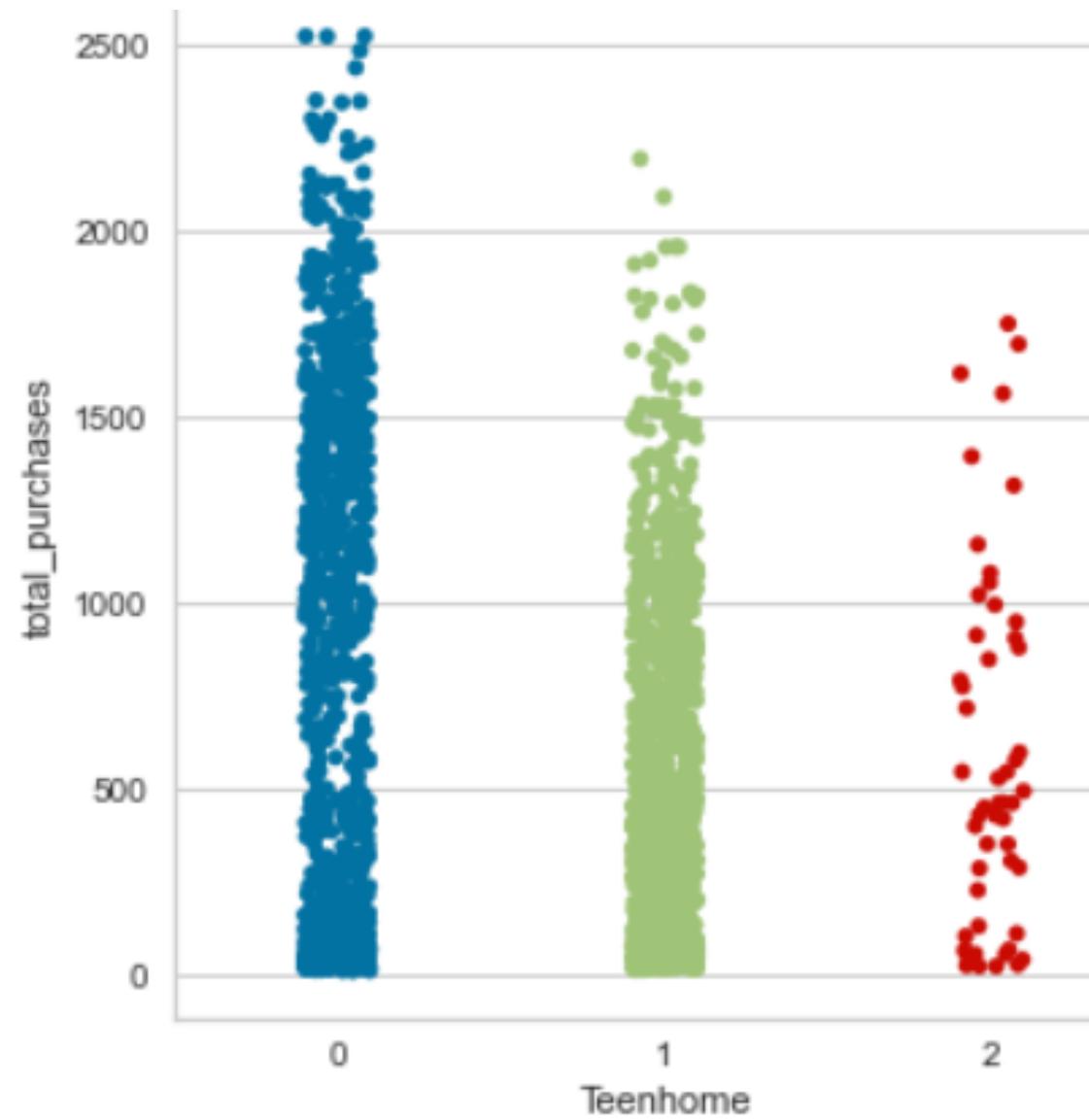


That the largest percentage of the number of purchases made by customers and up to 41.1% was between 5 to 215 purchases.

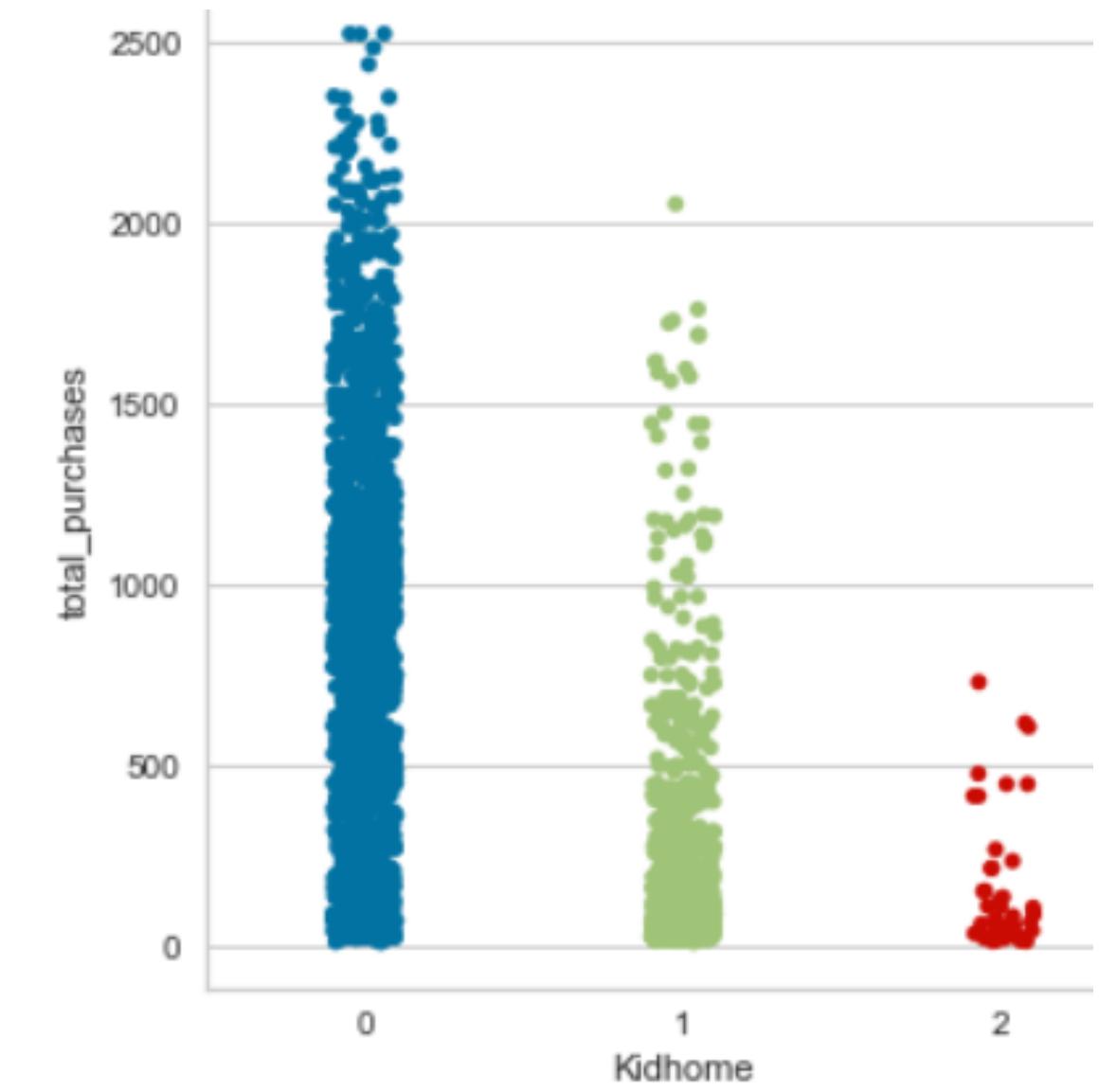
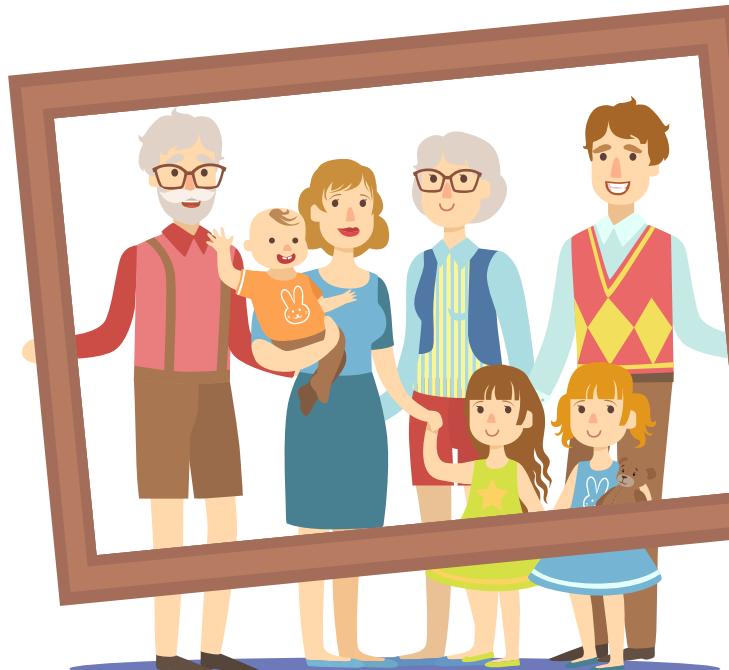




- Q3: Is the size of the family and the number of children related to the number of purchases?

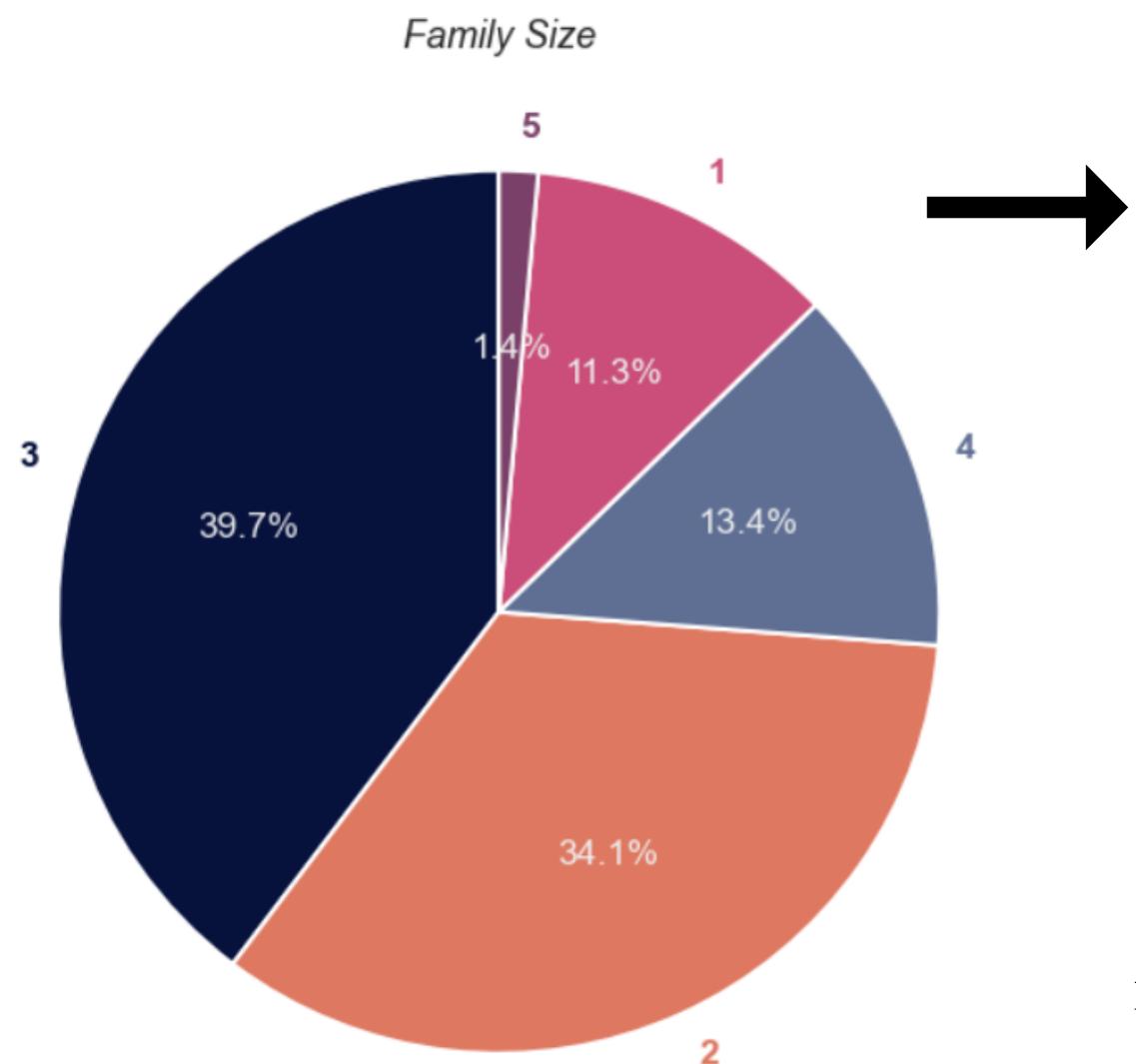


We see that customers with teenagers have more purchases than customers with children



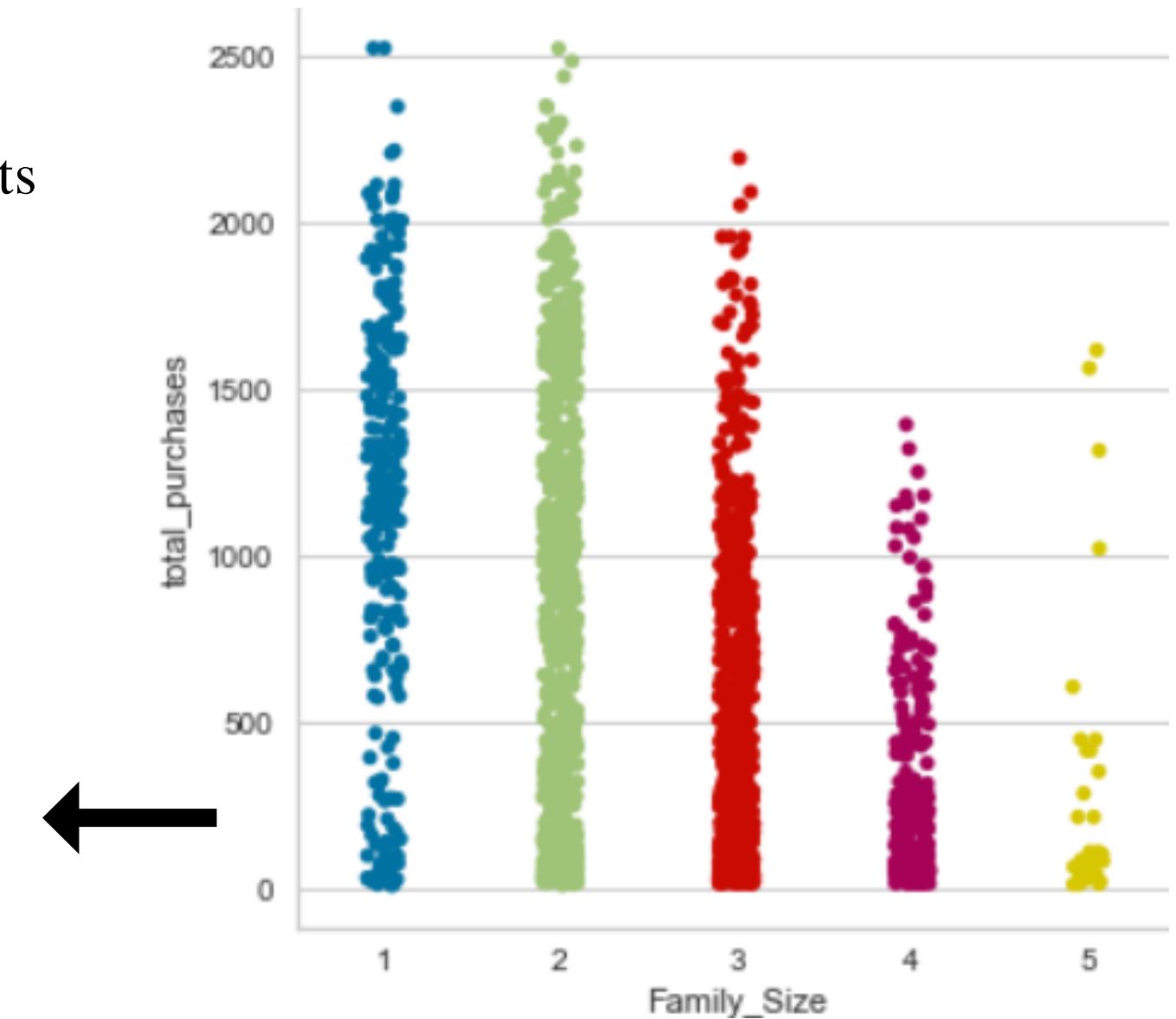


- Q3: Is the size of the family and the number of children related to the number of purchases?



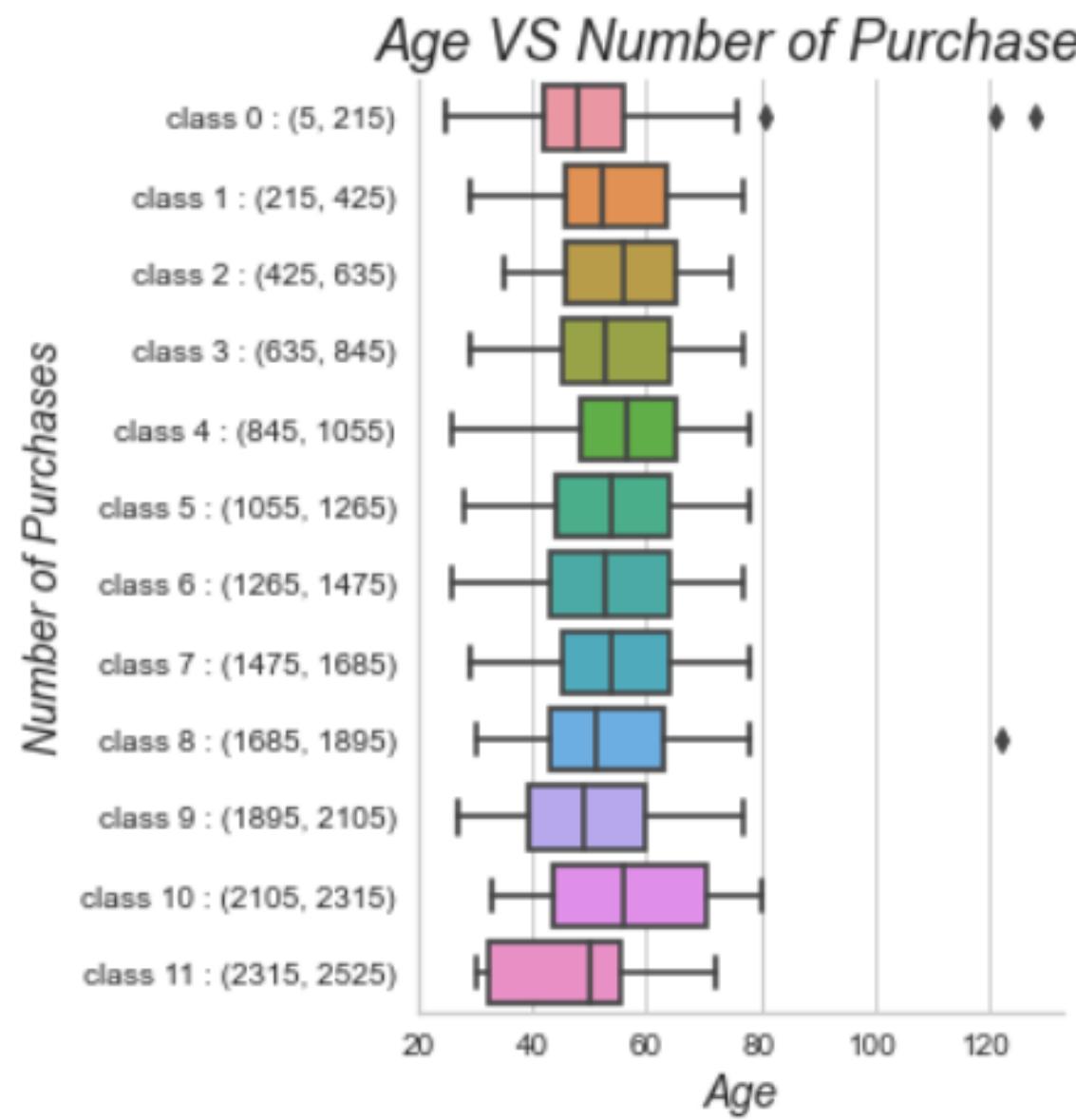
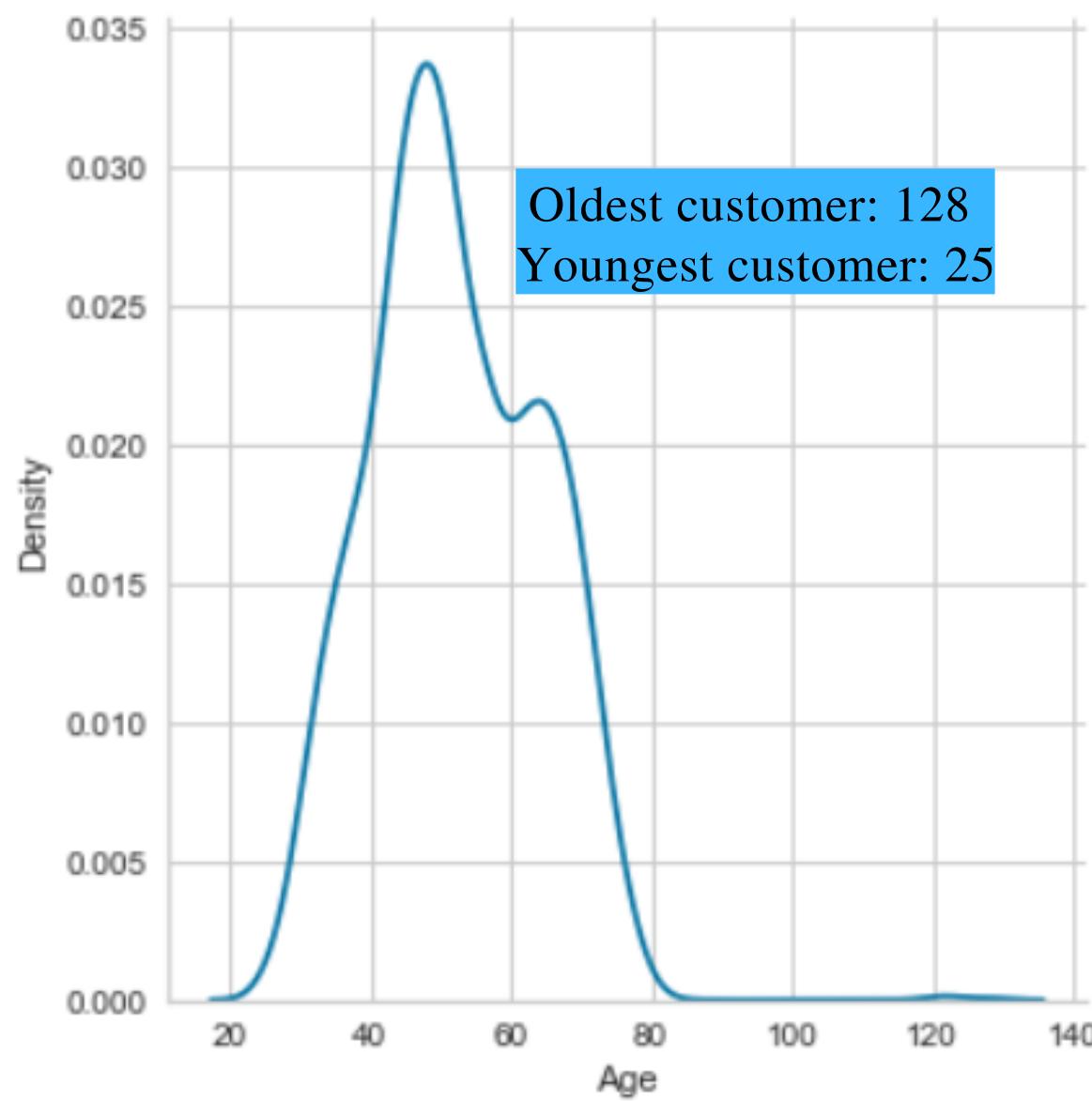
→ Most of the company's clients
are families of two or 3

The figure shows that the
higher the number of family
members, the fewer purchases





- Q4: What does age have to do with the number of purchases?

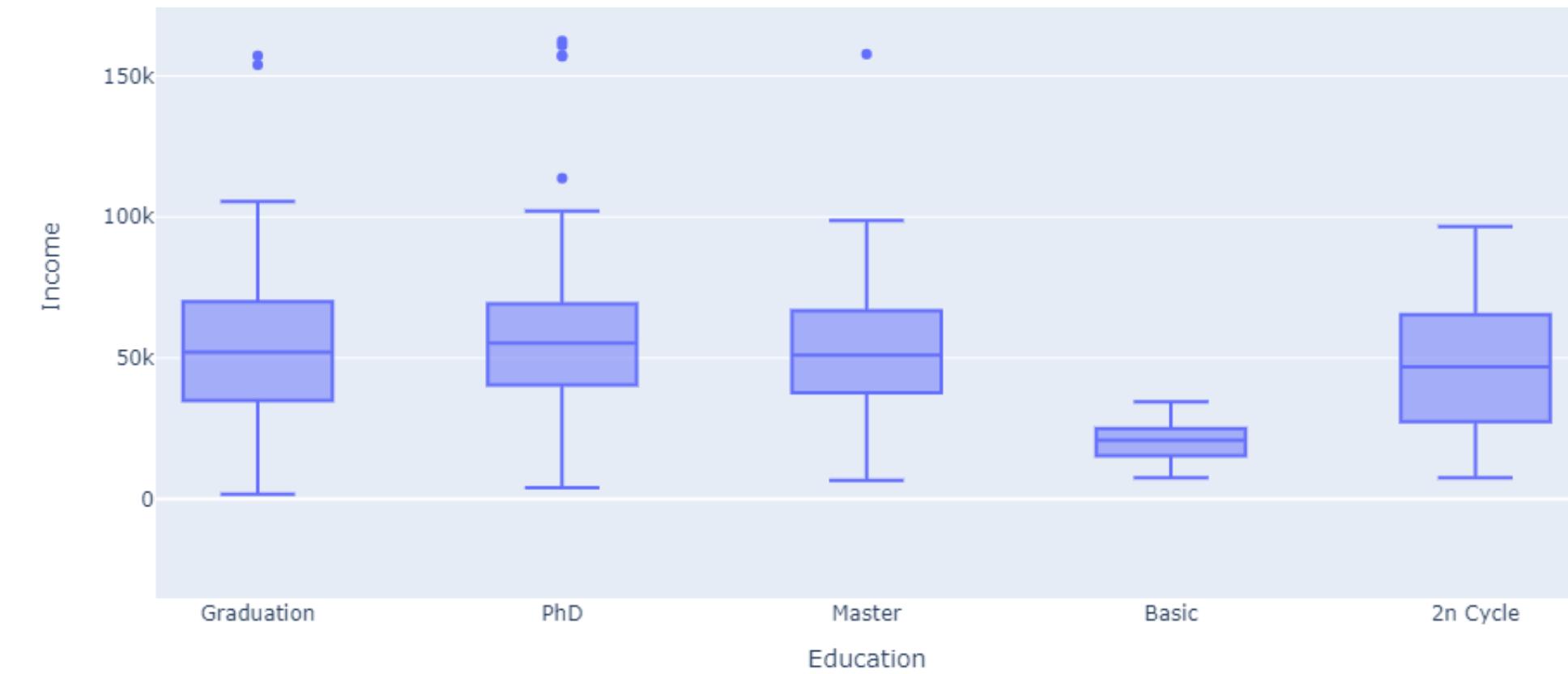
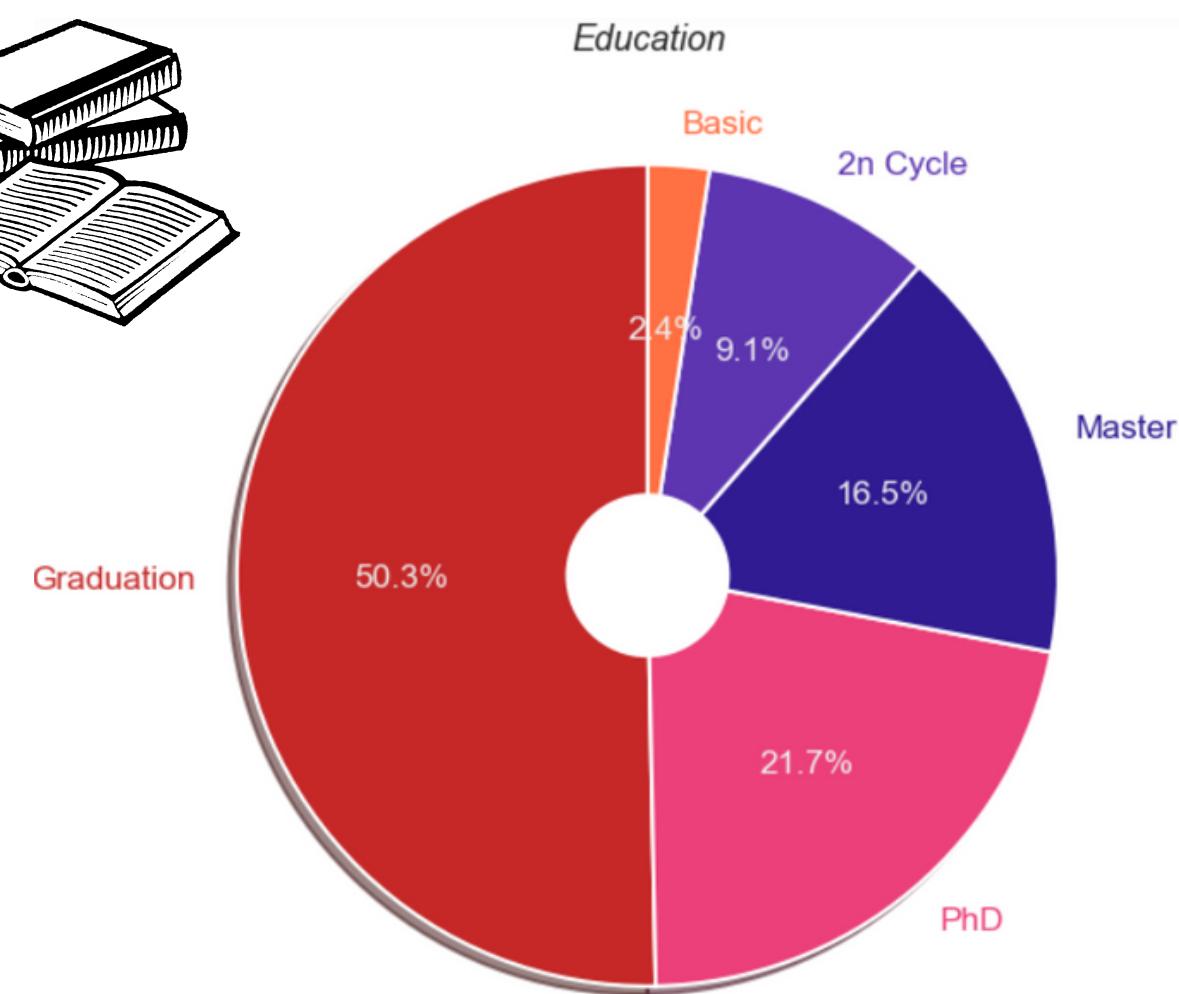
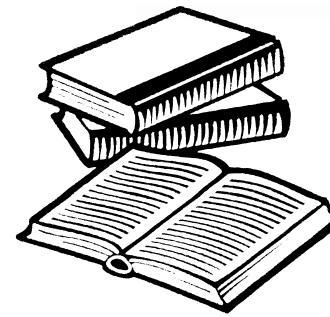


There is no relationship between age and number of purchases



- Q5: Do clients with better education get more income?

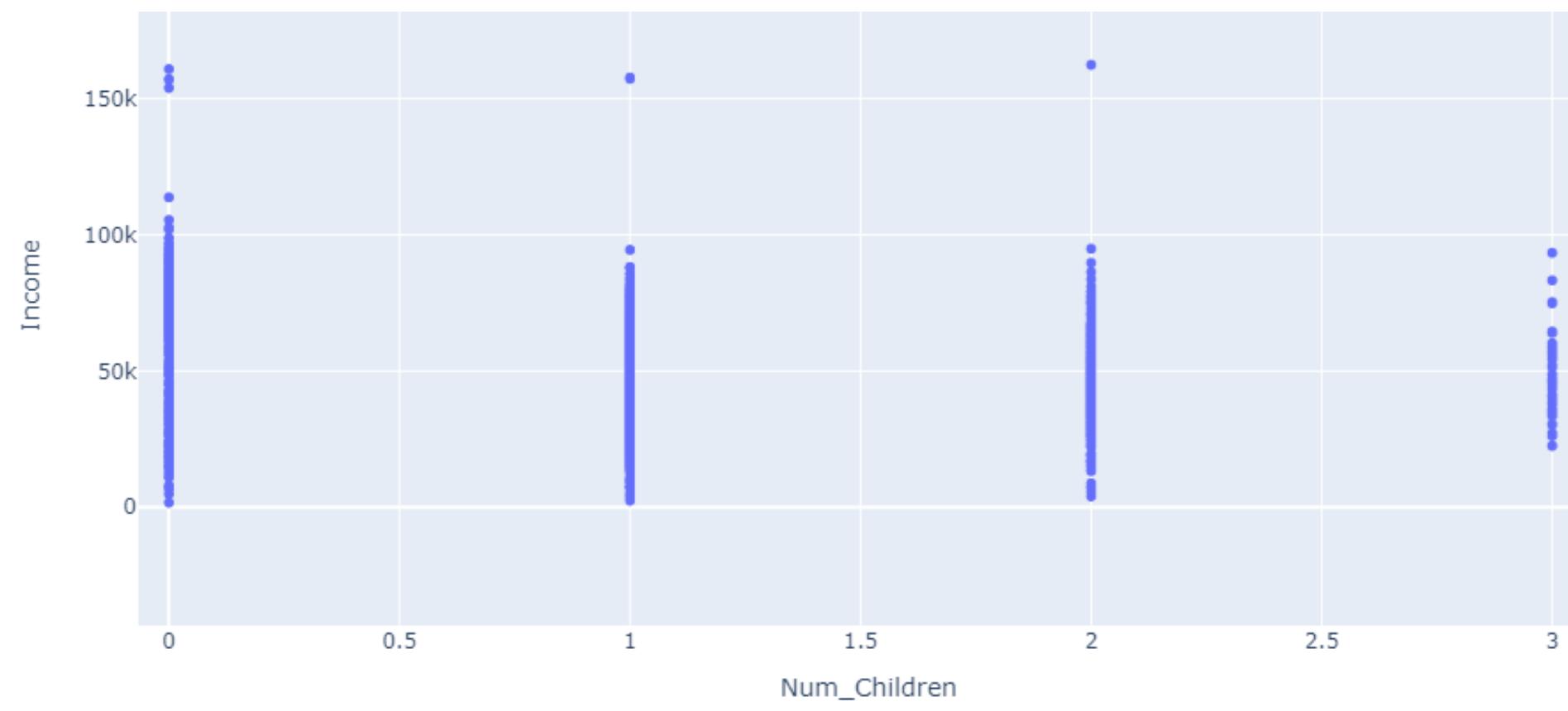
The percentage of clients with university degrees reaches 97.6%



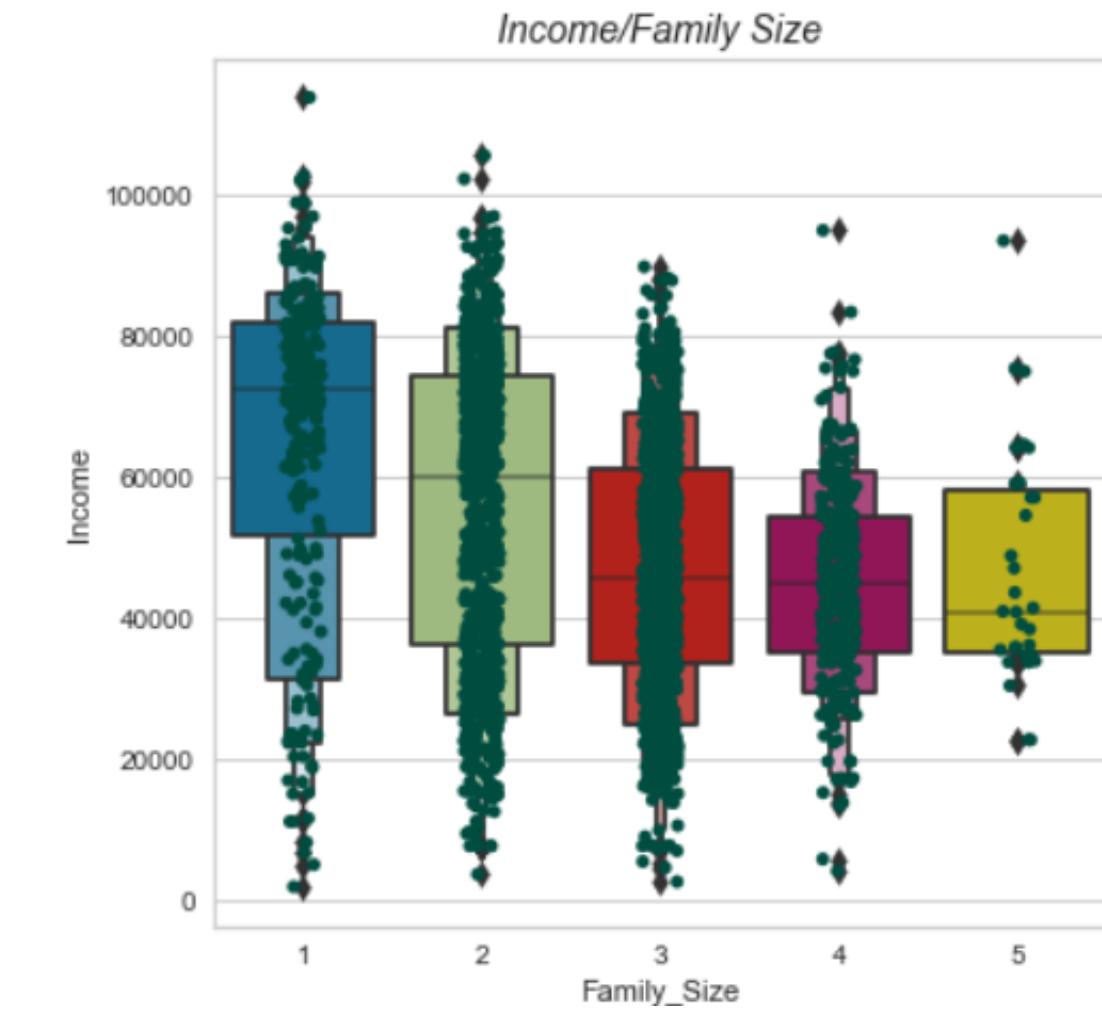
The average income of all clients is very similar except for Basic



- Q6: What is the relationship between income, family size and number of children?



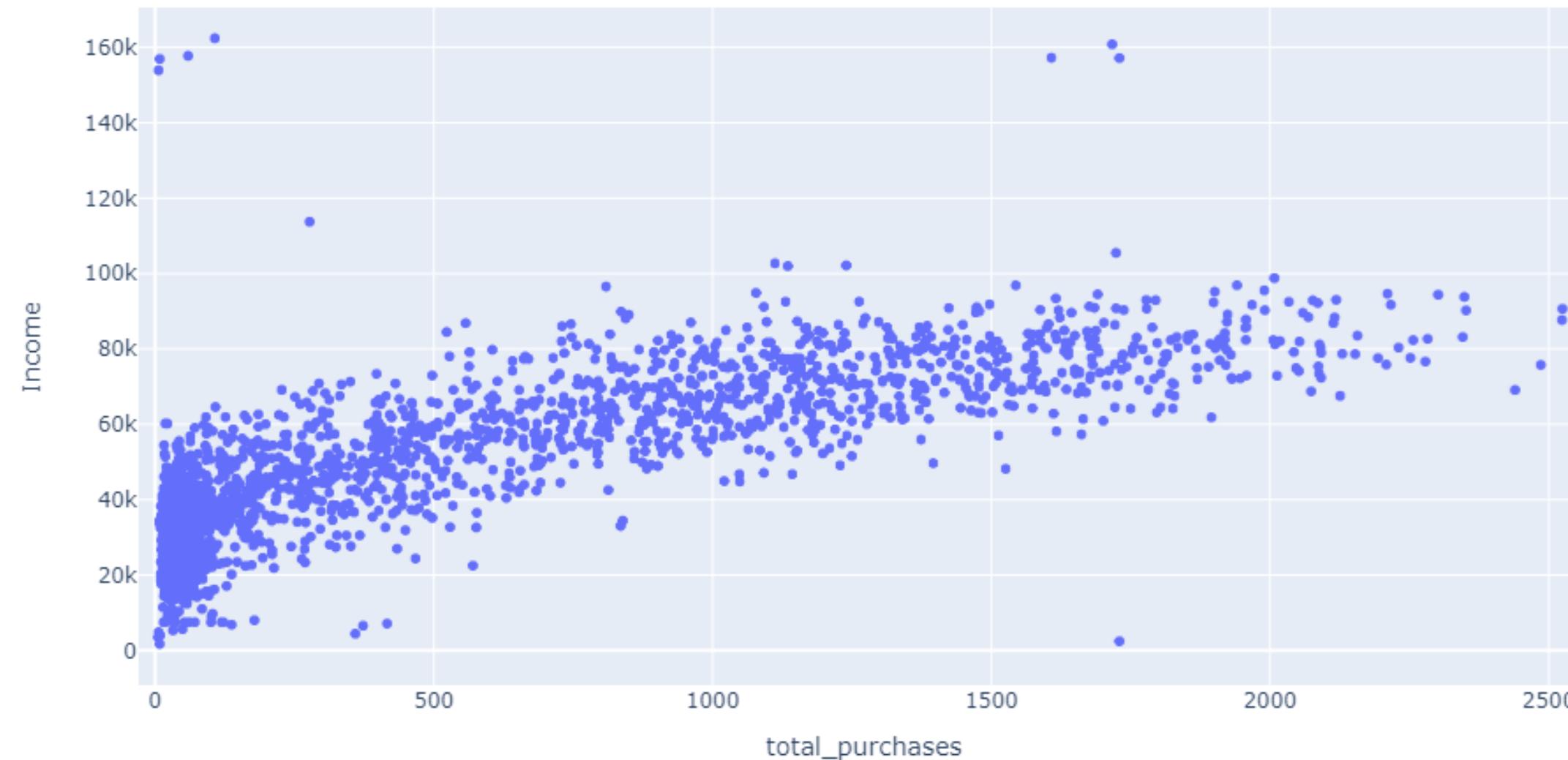
All clients have approximately the same income, despite the difference in the number of children, except for those who have 3



The average income of clients is approximately the same, although the size of the family is not equal.



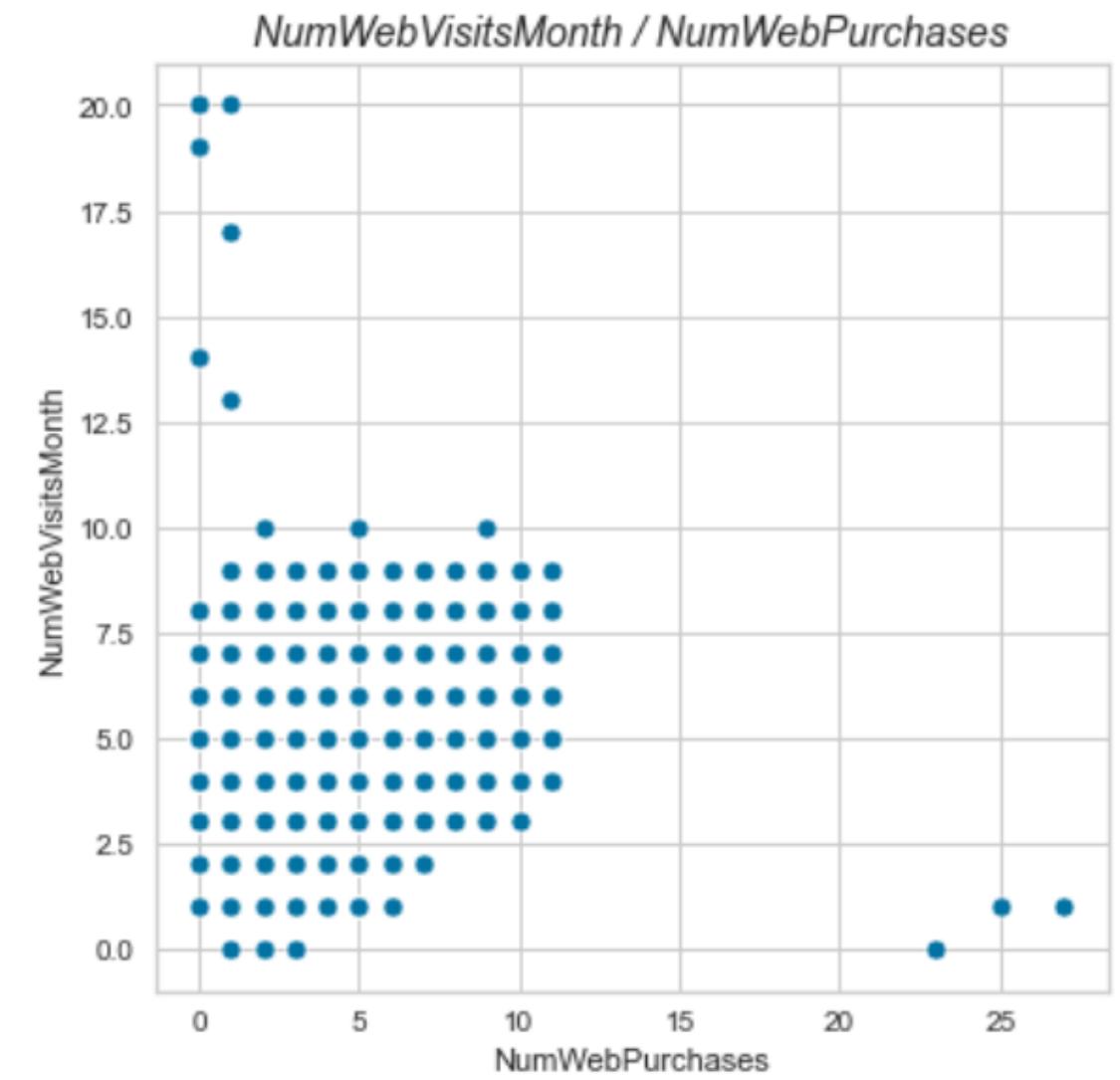
- Q7: Do customers with more income make more purchases?



There is a near-linear relationship between income and the number of purchases

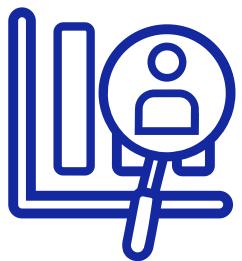


- Q8: Do customers who visit the website make purchases through it or not?

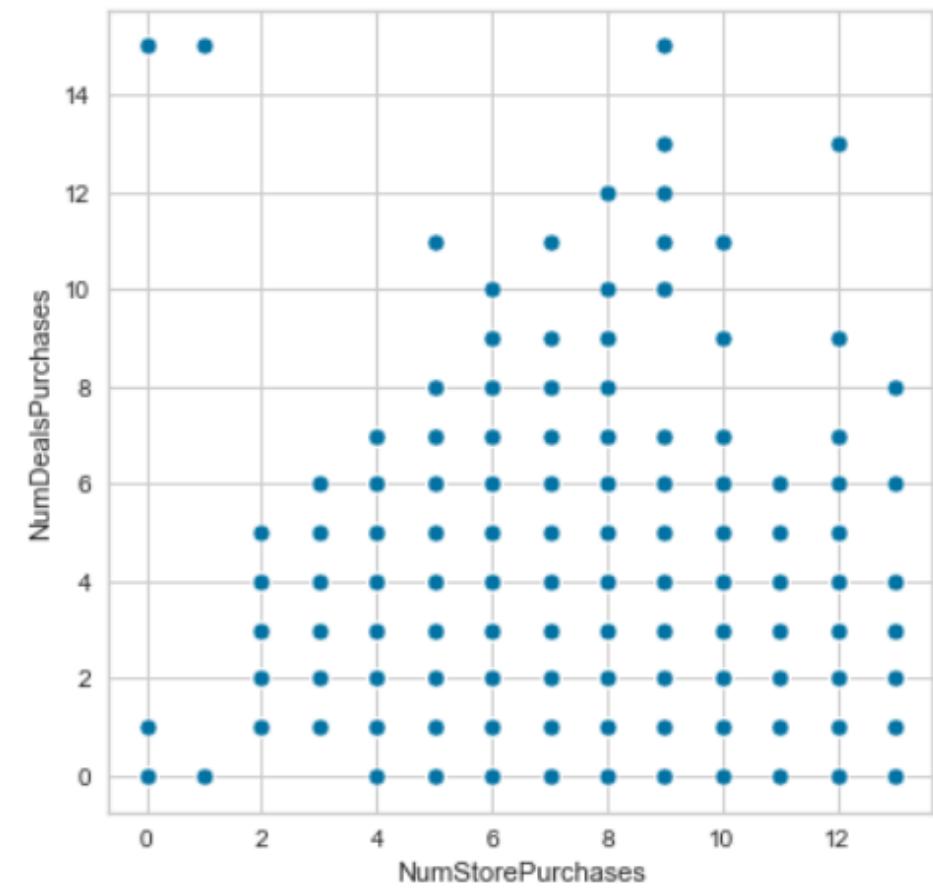
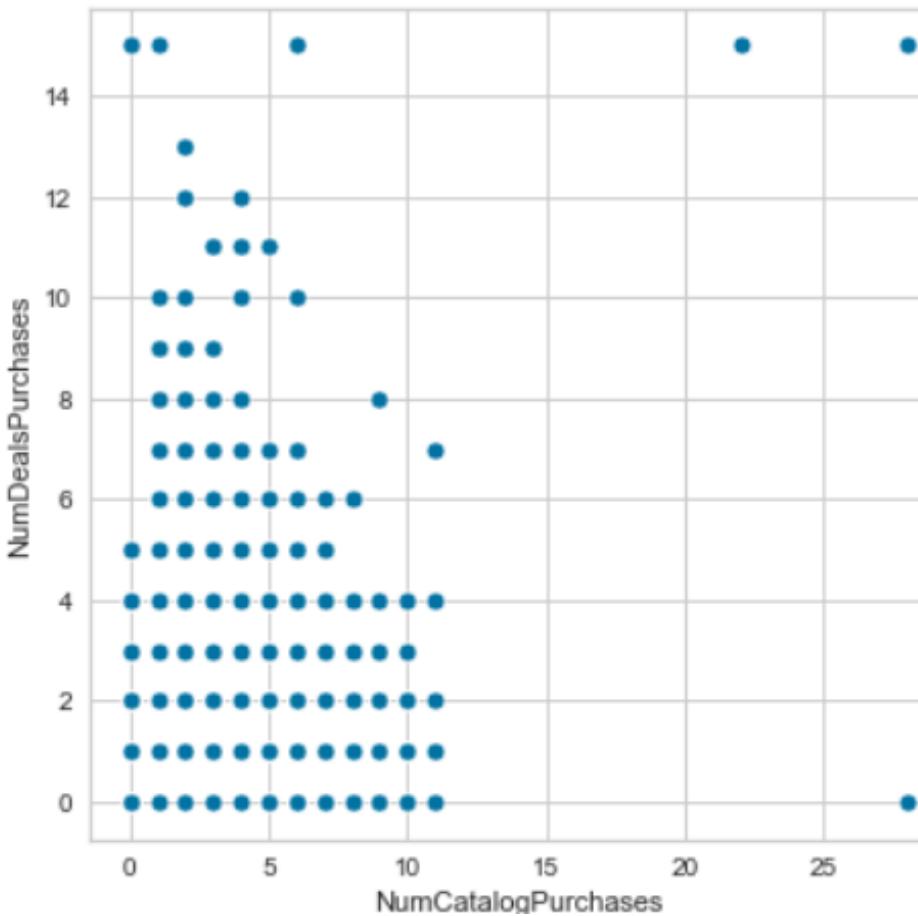
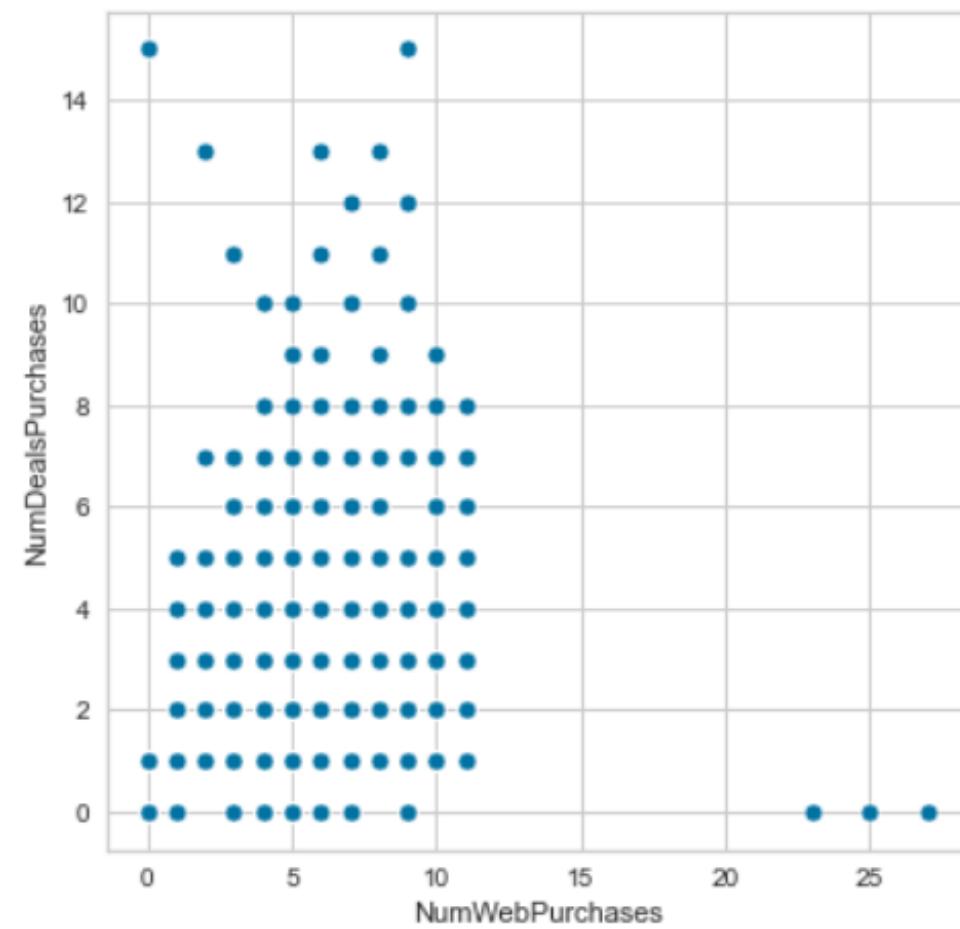


In general, there is a linear relationship between the number of visits to the site and the number of purchases from it

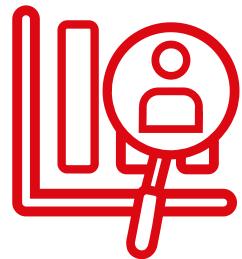
EDA



- Q9: What is the most popular way for customers to buy at a discount?

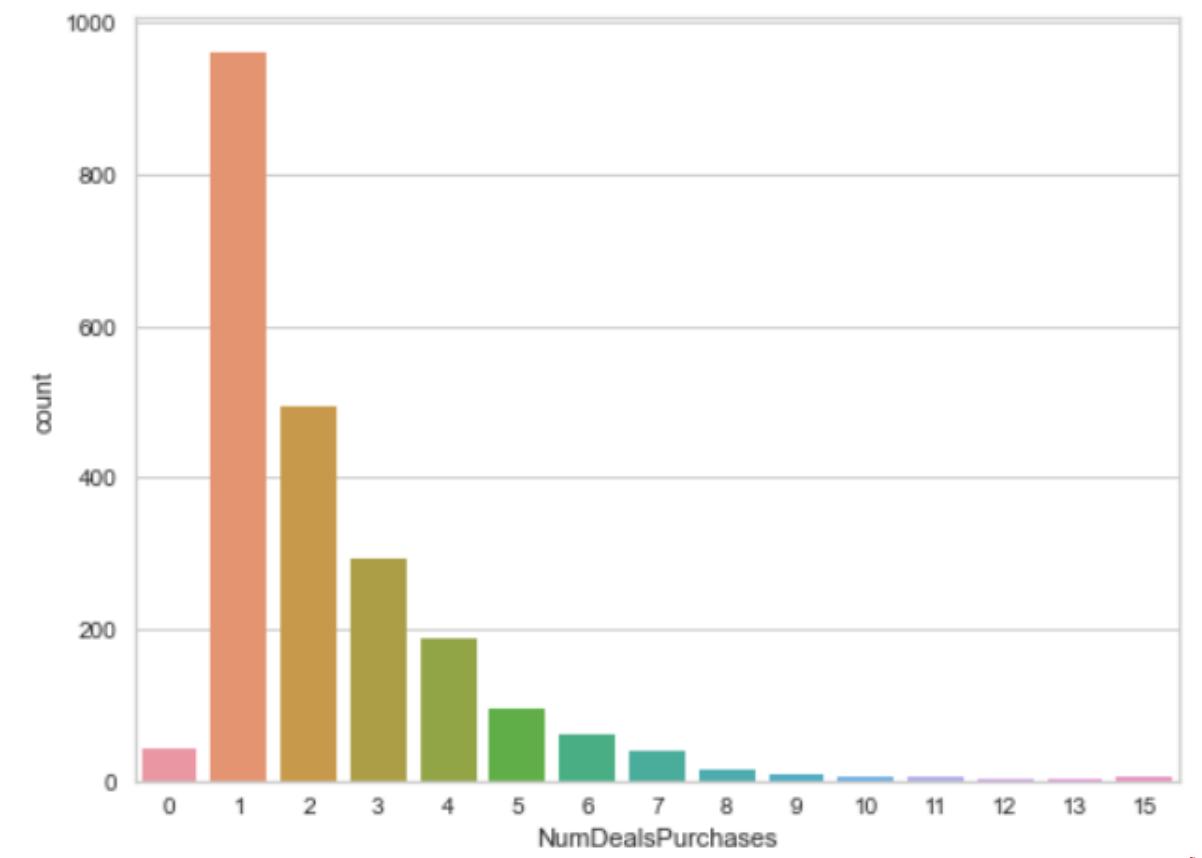


We conclude that, in general, there is no relationship between the method of purchase (from the store, the website, the catalog) and the times of purchase at a discount

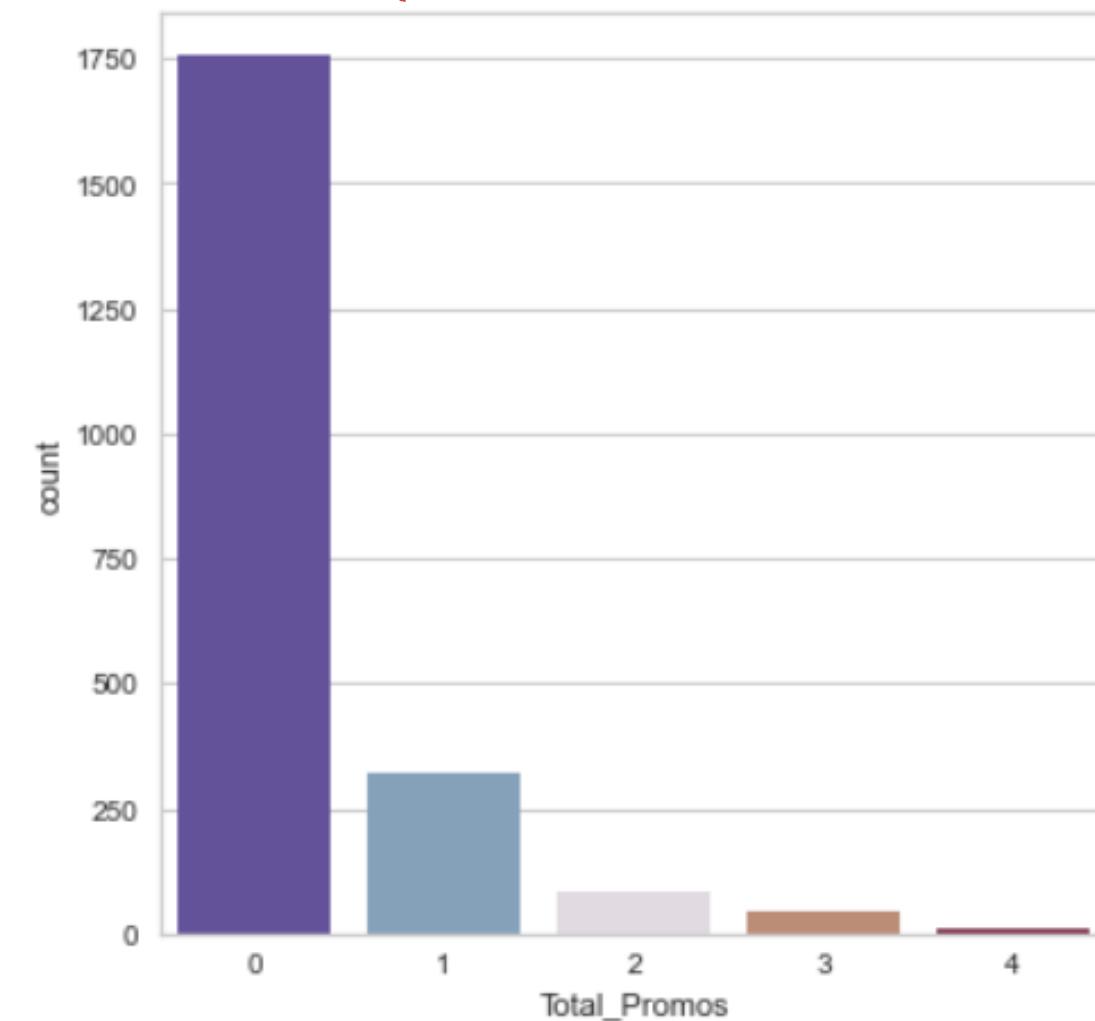


- Q10: Are the promotions campaigns carried out by the company successful?

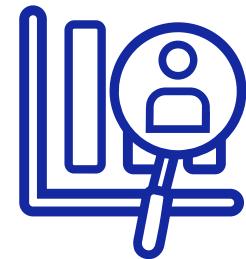
✓ COMPLIANT



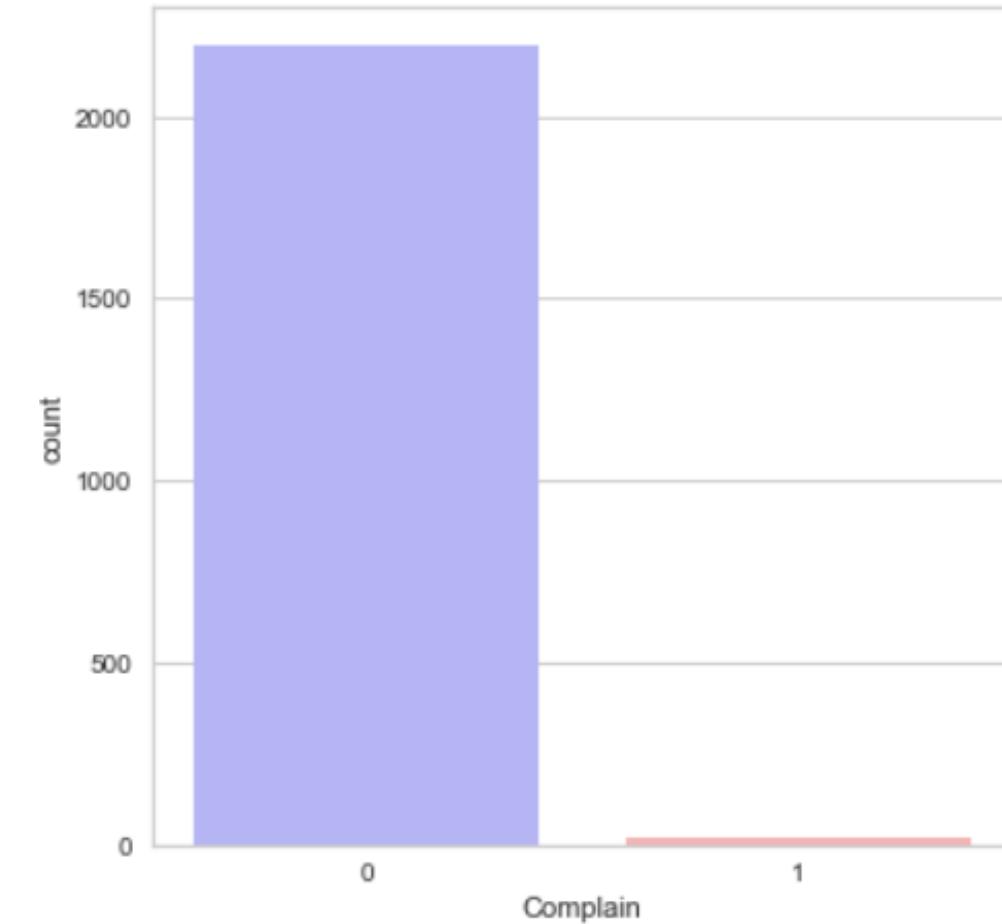
Almost all customers have made a purchase at least once



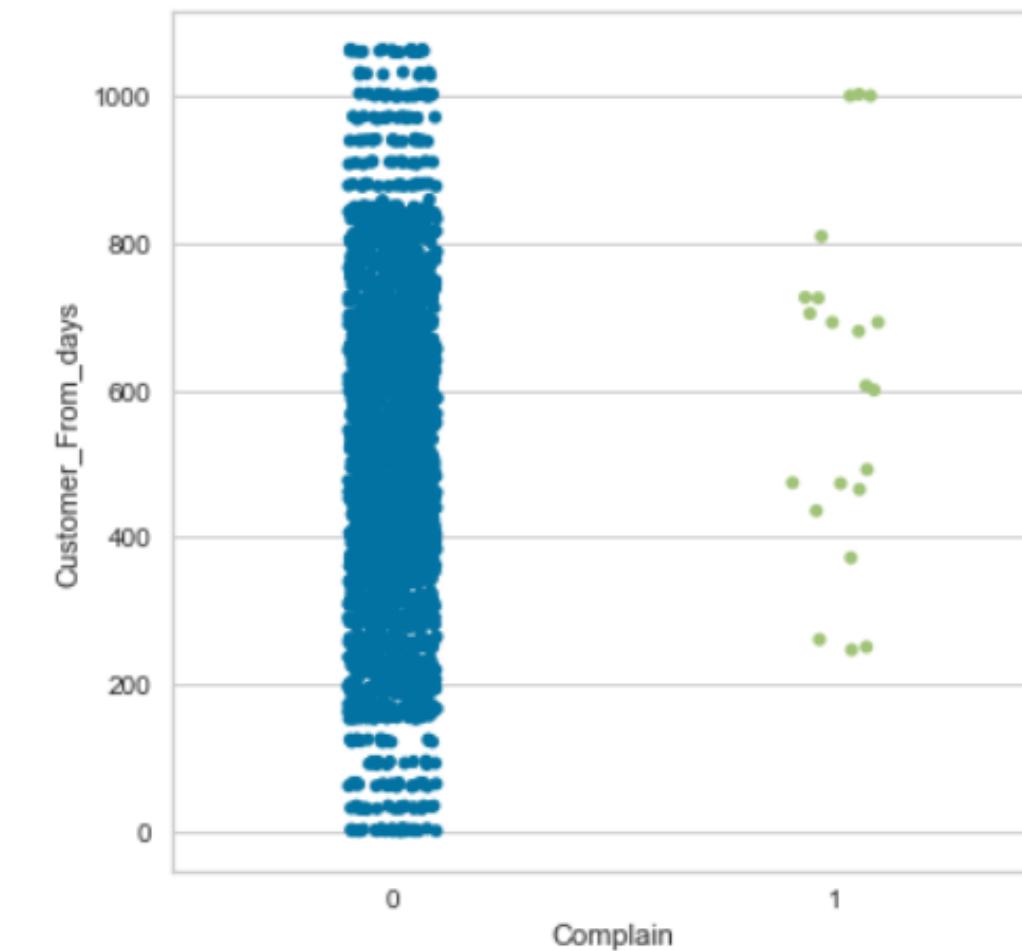
The number of customers who did not accept the offers is very large, up to 80%



- Q11: Are there complaints from customers, if yes, are they old or new customers?

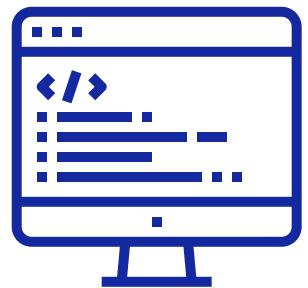


Very few customers
who complained



All customers who filed a complaint as if they
were with the company more than 200 days ago

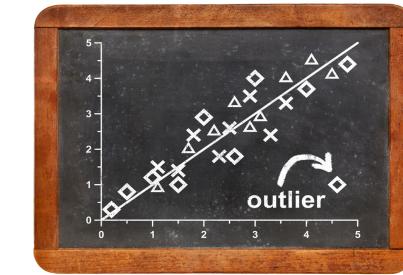
Data Preprocessing.



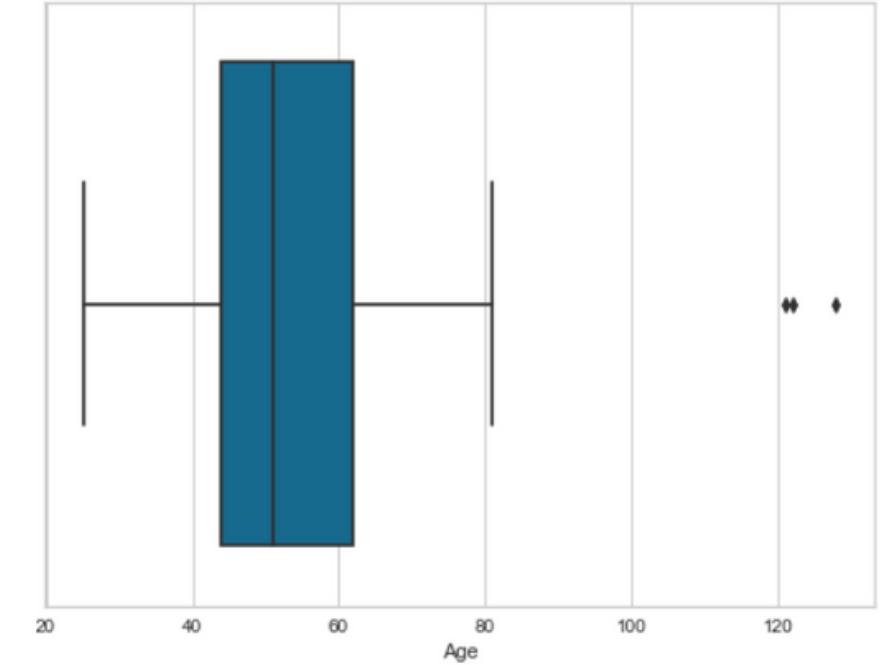
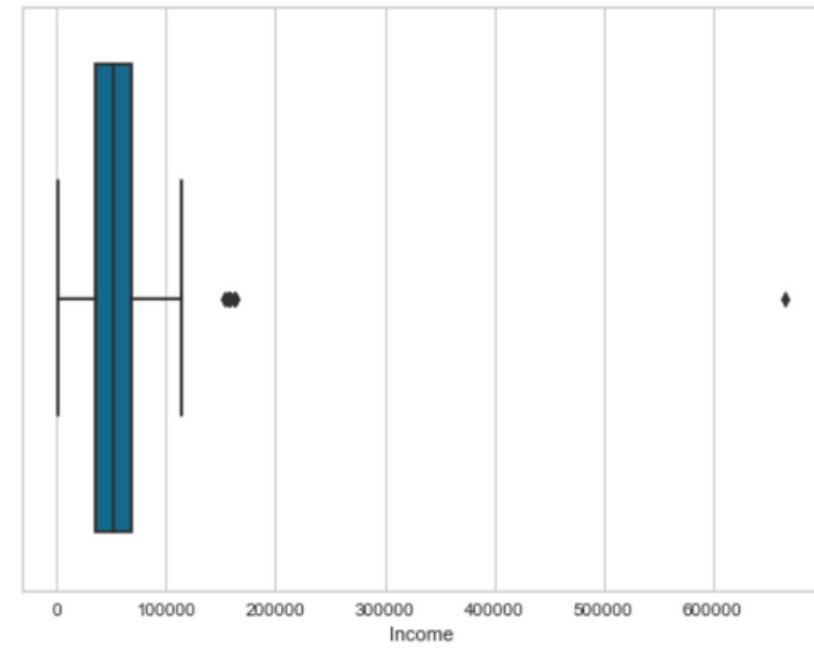
We dropped 13 features

1-ID 2-Year_Birth 3-Dt_Customer
4-AcceptedCmp3 5-AcceptedCmp4
6-AcceptedCmp5 7-AcceptedCmp1
8-AcceptedCmp2 9-Complain
10-Z_CostContact 11-Z_Revenue
12-Response 13-Living_With

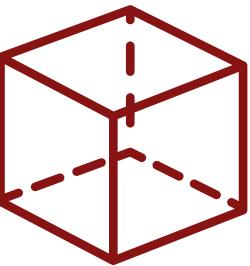
Check if there are any outliers



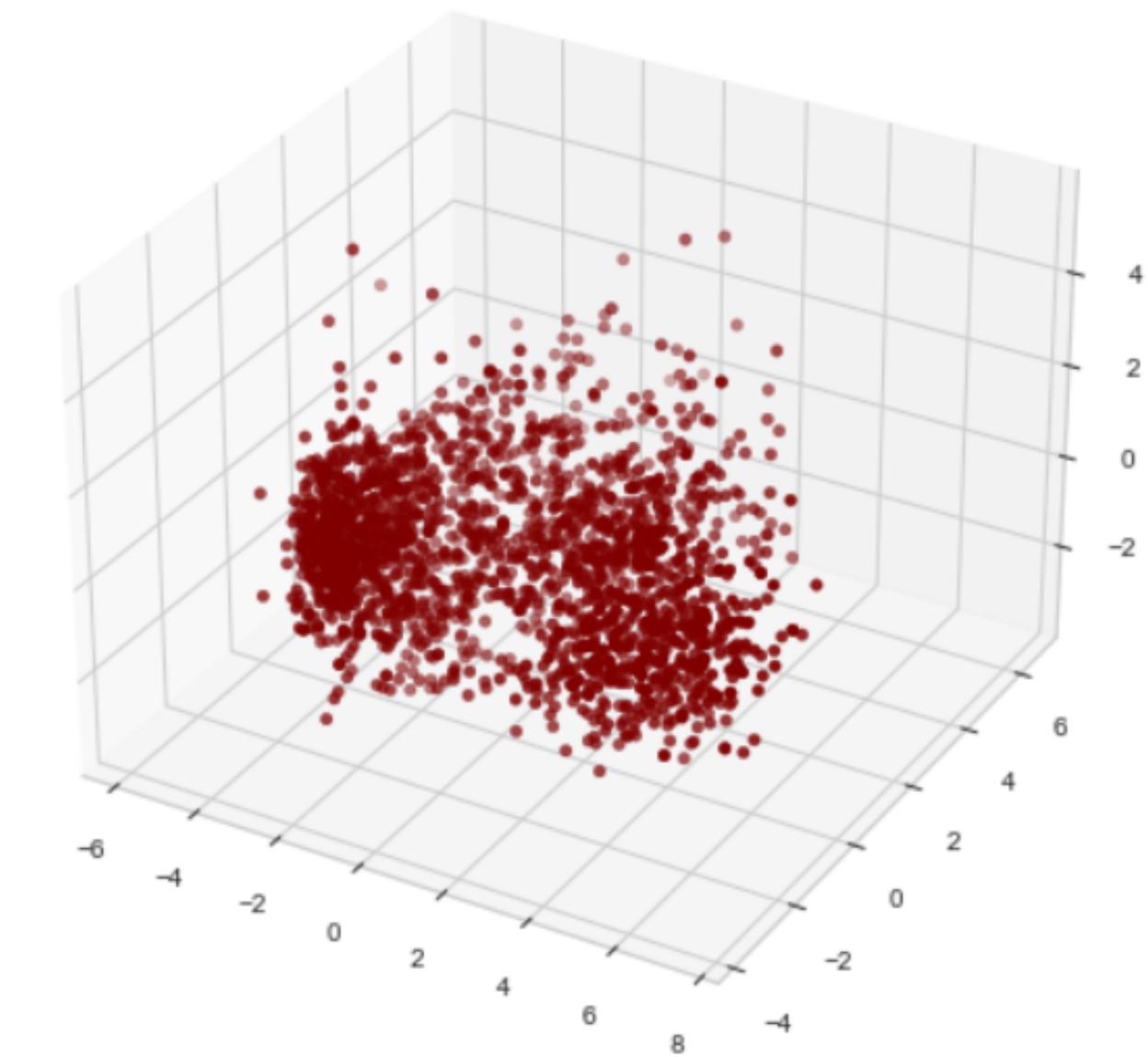
We eliminated any income over 200,000
and any age over 100 years old



Data Preprocessing.



A 3D Projection of Data After Dimensional Reduction

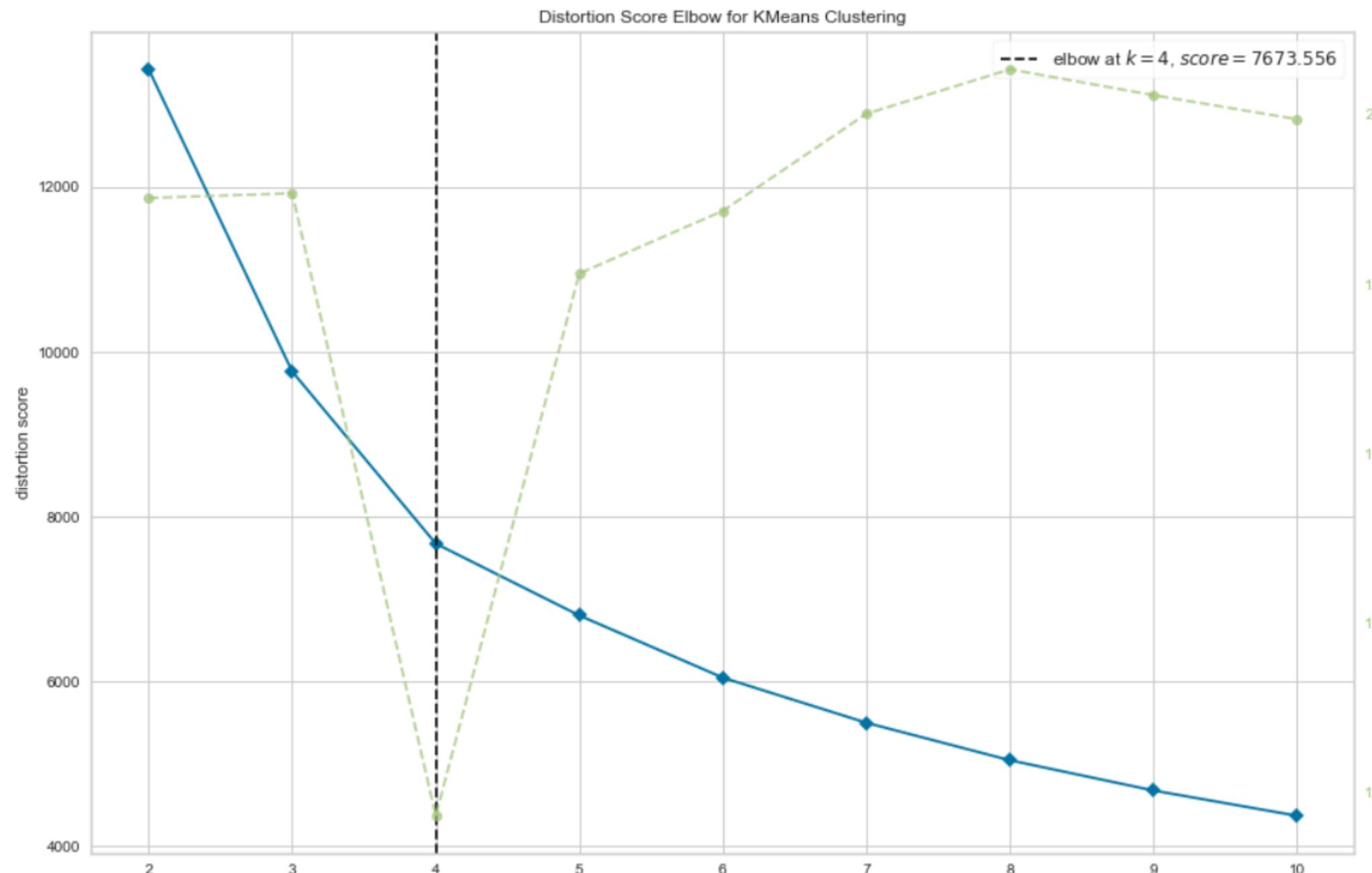


Label encoding the categorical features.
Scaling the features using the StandardScaler.
Using PCA to reduce the dimensions of the data.

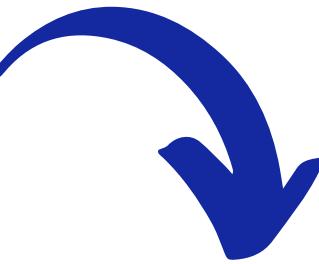
Clustering



Using the elbow method to find the optimal number of clusters



The optimal number of clusters is 4 .

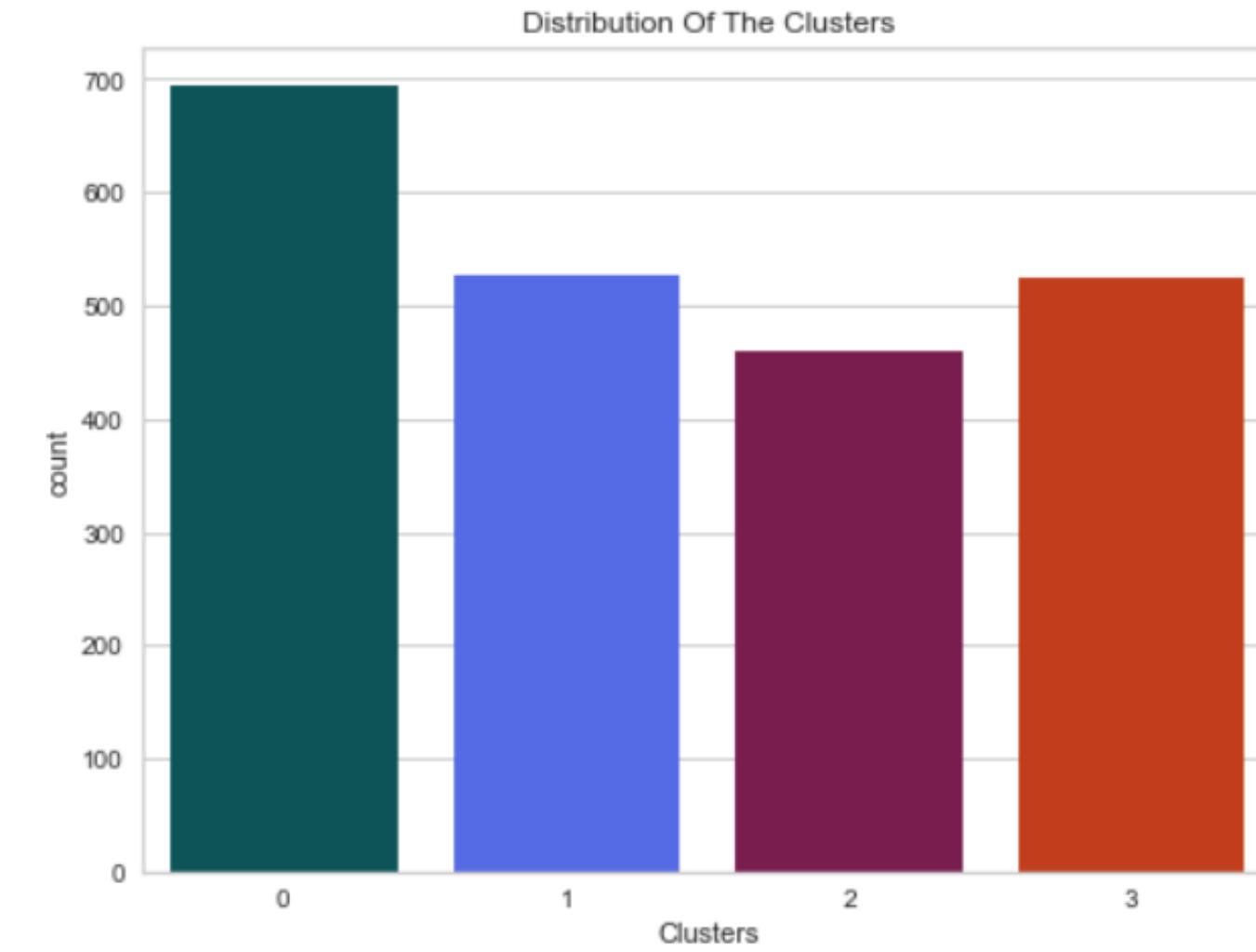
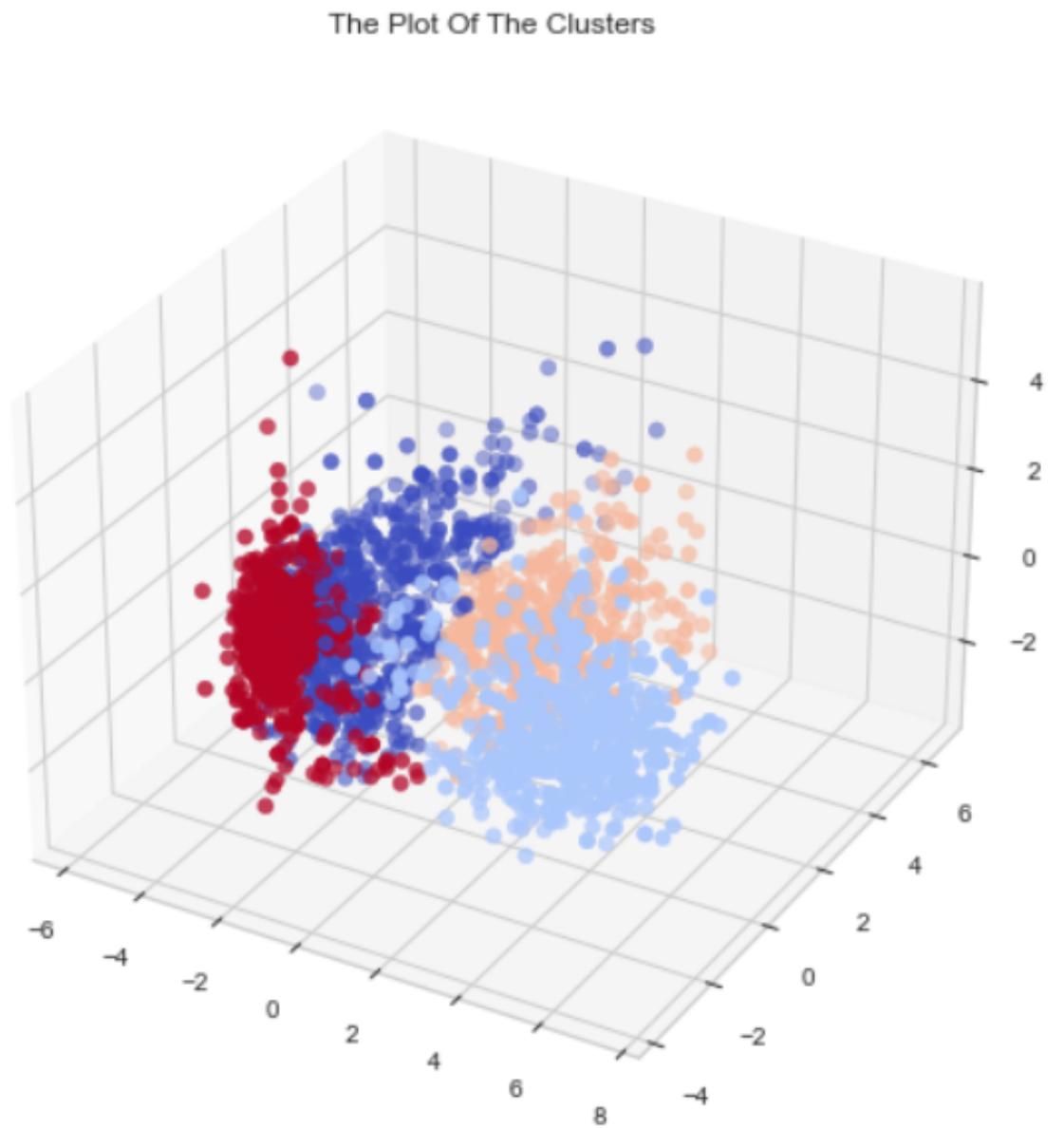


Using Agglomerative Clustering model

Evaluation Models.



A 3D Projection of Data After clustering using AgglomerativeClustering



Show numbers of clustering

Evaluation Models.



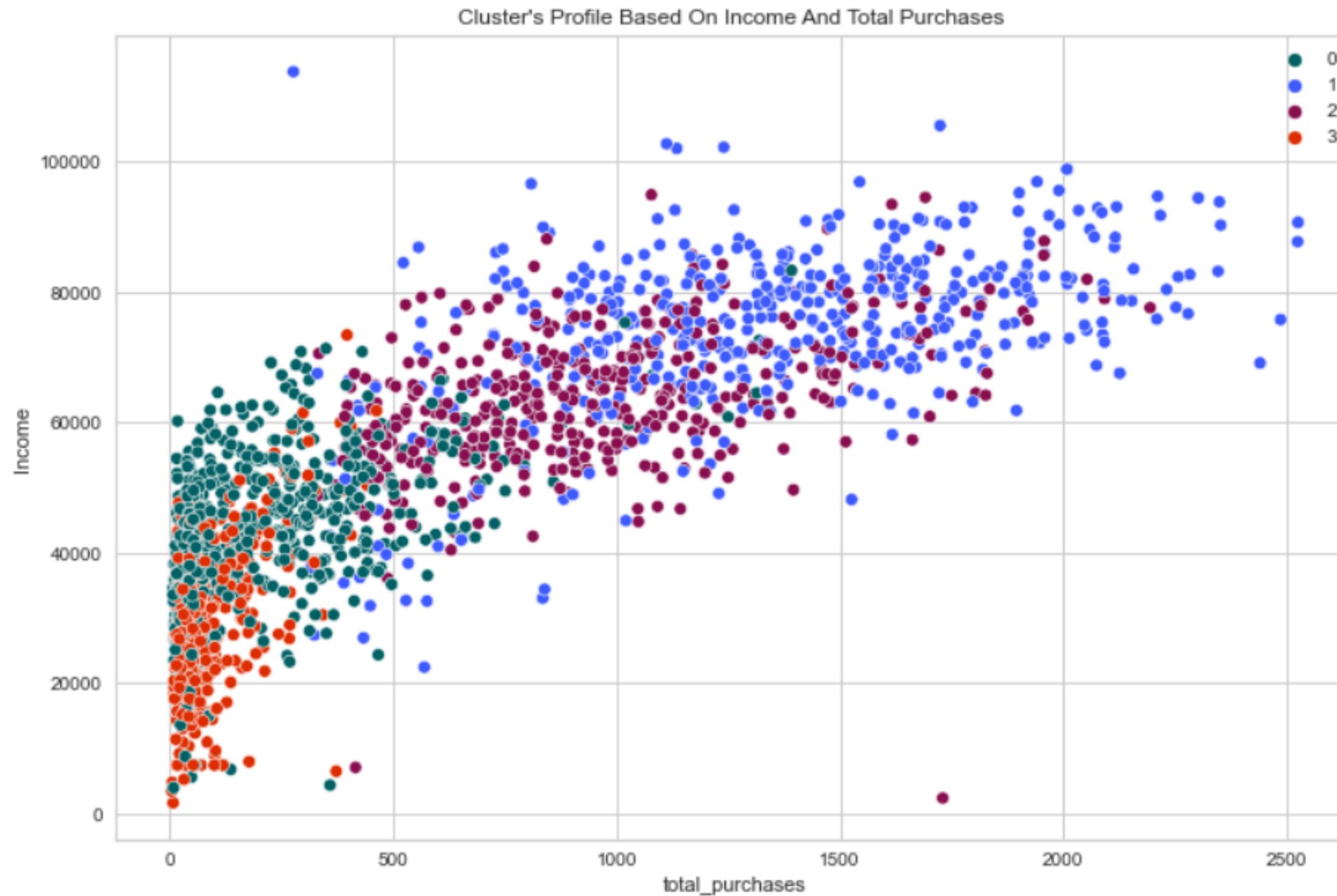
	Education	Marital_Status	Income	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	Kidhome	Teenhome	Num_Children	MntGoldProducts	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	Customer_From_days	Age	Family_Size	total_purchases	purchase_quantity	Total_Promos	Clusters
Education																									
Marital_Status	0.0032																								
Income	0.14	0.011																							
Recency	-0.011	0.018	0.0067																						
MntWines	0.2	0.0075	0.73	0.016																					
MntFruits	-0.085	0.0033	0.54	-0.0049	0.38																				
MntMeatProducts	0.032	0.03	0.7	0.026	0.59	0.57																			
MntFishProducts	-0.12	0.044	0.55	0.0012	0.4	0.59	0.6																		
MntSweetProducts	-0.11	0.015	0.56	0.026	0.39	0.57	0.56	0.58																	
Kidhome	-0.044	-0.02	-0.53	0.012	-0.5	-0.37	-0.45	-0.39	-0.38																
Teenhome	0.12	-0.0079	0.042	0.0140	0.0028	-0.18	-0.27	-0.21	-0.16	-0.04															
Num_Children	0.056	-0.02	-0.35	0.019	-0.36	-0.4	-0.51	-0.43	-0.39	0.69	0.7														
MntGoldProducts	-0.099	0.0085	0.42	0.019	0.39	0.39	0.38	0.43	0.36	-0.36	-0.02	-0.27													
NumDealsPurchases	0.019	-0.017	-0.130	0.00048	0.012	-0.14	-0.17	-0.15	-0.12	0.23	0.39	0.45	0.057												
NumWebPurchases	0.085	0.0025	0.5	-0.0051	0.55	0.3	0.33	0.3	0.33	-0.38	0.16	-0.15	0.41	0.25											
NumCatalogPurchases	0.063	0.02	0.71	0.03	0.67	0.51	0.71	0.56	0.52	-0.52	-0.11	-0.46	0.47	-0.059	0.42										
NumStorePurchases	0.070	0.0082	0.69	0.00046	0.64	0.46	0.52	0.46	0.45	-0.51	0.047	-0.33	0.39	0.073	0.51	0.56									
NumWebVisitsMonth	-0.035	0.034	-0.65	-0.018	-0.33	-0.42	-0.54	-0.45	-0.43	0.45	0.13	0.42	-0.25	0.36	-0.061	-0.53	-0.45								
Customer_From_days	-0.070	0.0046	0.031	0.034	0.15	0.06	0.069	0.078	0.077	-0.0580	0.0095	0.035	0.15	0.2	0.17	0.089	0.11	0.26							
Age	0.18	0.06	0.21	0.014	0.16	0.013	0.042	0.041	0.021	-0.24	0.36	0.093	0.059	0.072	0.16	0.14	0.14	-0.12	-0.02						
Family_Size	0.036	0.029	-0.3	0.016	-0.3	-0.34	-0.44	-0.36	-0.33	0.58	0.59	0.85	-0.24	0.38	-0.12	-0.39	-0.27	0.35	-0.0290	0.079					
total_purchases	0.091	0.021	0.82	0.022	0.9	0.61	0.86	0.64	0.61	-0.56	-0.14	-0.5	0.53	-0.079	0.53	0.8	0.68	-0.5	0.14	0.12	-0.43				
purchase_quantity	0.083	0.014	0.81	0.014	0.84	0.57	0.77	0.6	0.58	-0.56	-0.083	-0.47	0.51	-0.026	0.57	0.78	0.71	-0.48	0.14	0.13	-0.4	0.93			
Total_Promos	0.0370	0.0011	0.39	-0.013	0.51	0.16	0.32	0.17	0.2	-0.21	-0.13	-0.25	0.19	-0.13	0.2	0.37	0.2	-0.17	-0.0190	0.0015	-0.2	0.46	0.38		
Clusters	-0.12	-0.04	-0.240	0.0081	-0.05	-0.033	-0.1	-0.0530	-0.0374	-0.044	-0.41	-0.33	-0.053	-0.2	-0.08	-0.0660	-0.061	0.13	0.025	-0.28	-0.29	-0.0780	-0.0550	-0.036	
	Education	Marital_Status	Income	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	Kidhome	Teenhome	Num_Children	MntGoldProducts	NumDealsPurchases	NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	Customer_From_days	Age	Family_Size	total_purchases	purchase_quantity	Total_Promos	Clusters

The table shows how the features are related to each other.



Now we will see the relationship between some features and Clustering

Evaluation Models.



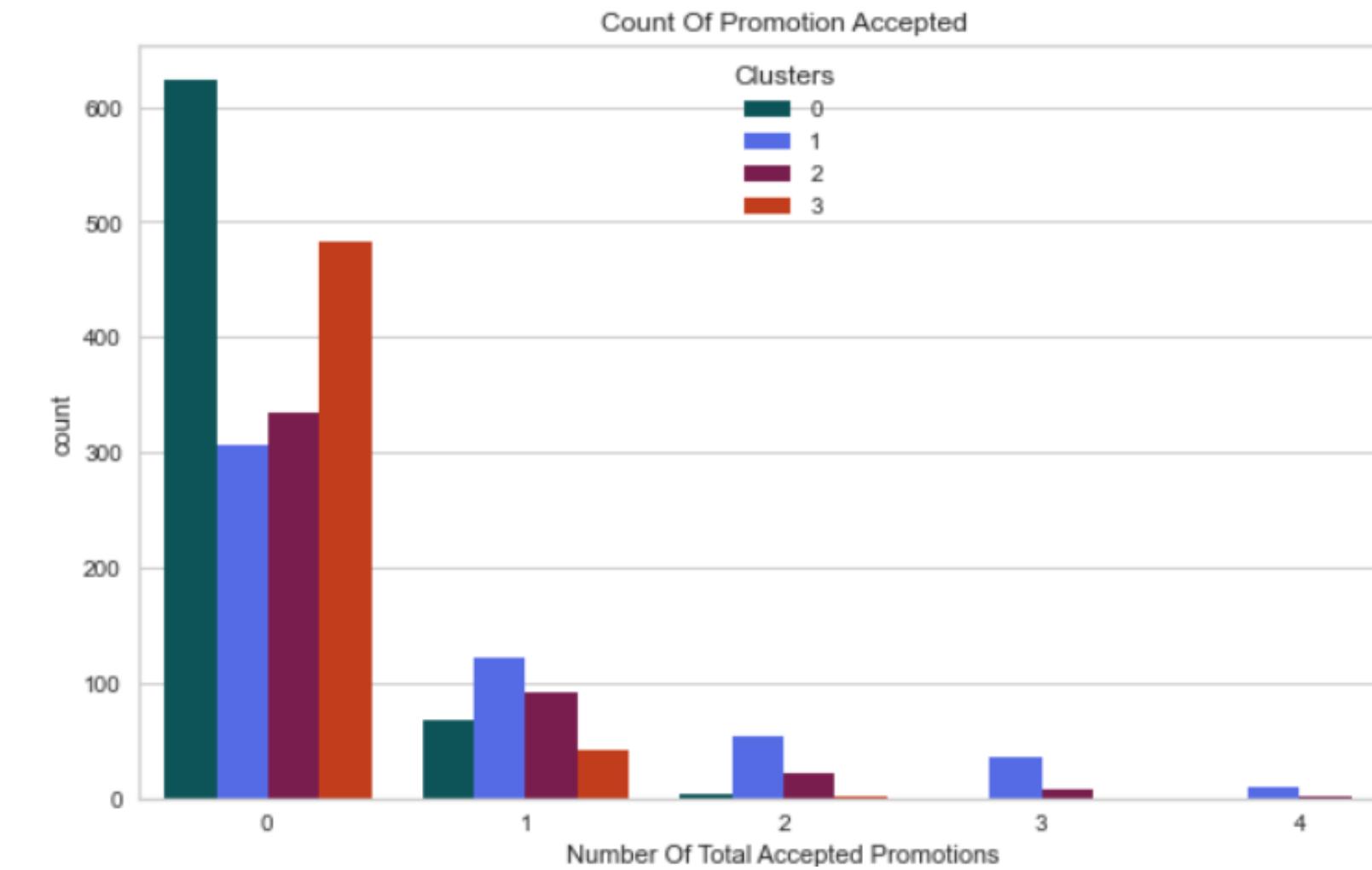
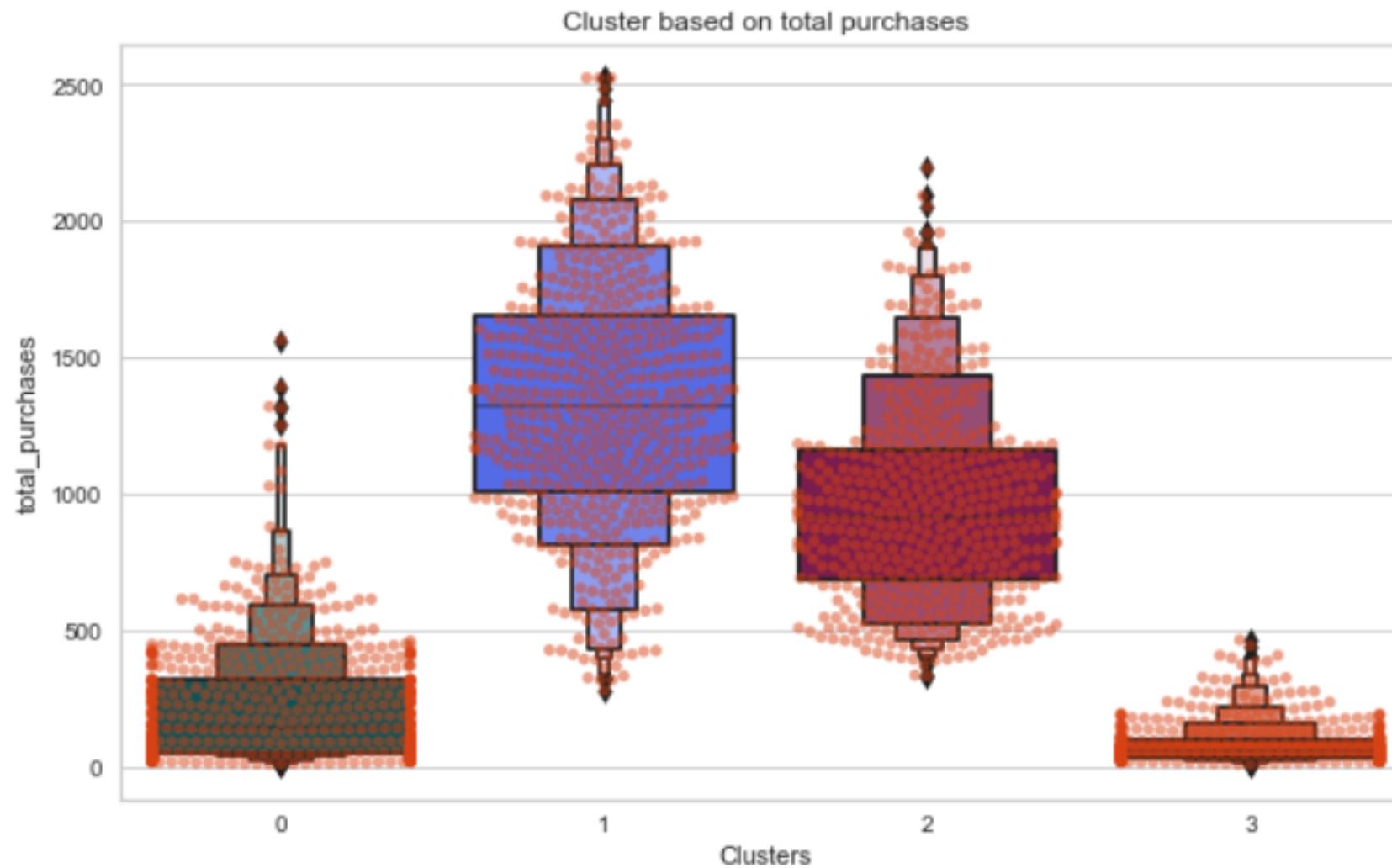
Agglomerative Clustering

The figure shows that the classification process is based on the linear relationship between income and the number of purchases.

Evaluation Models.



AgglomerativeClustering

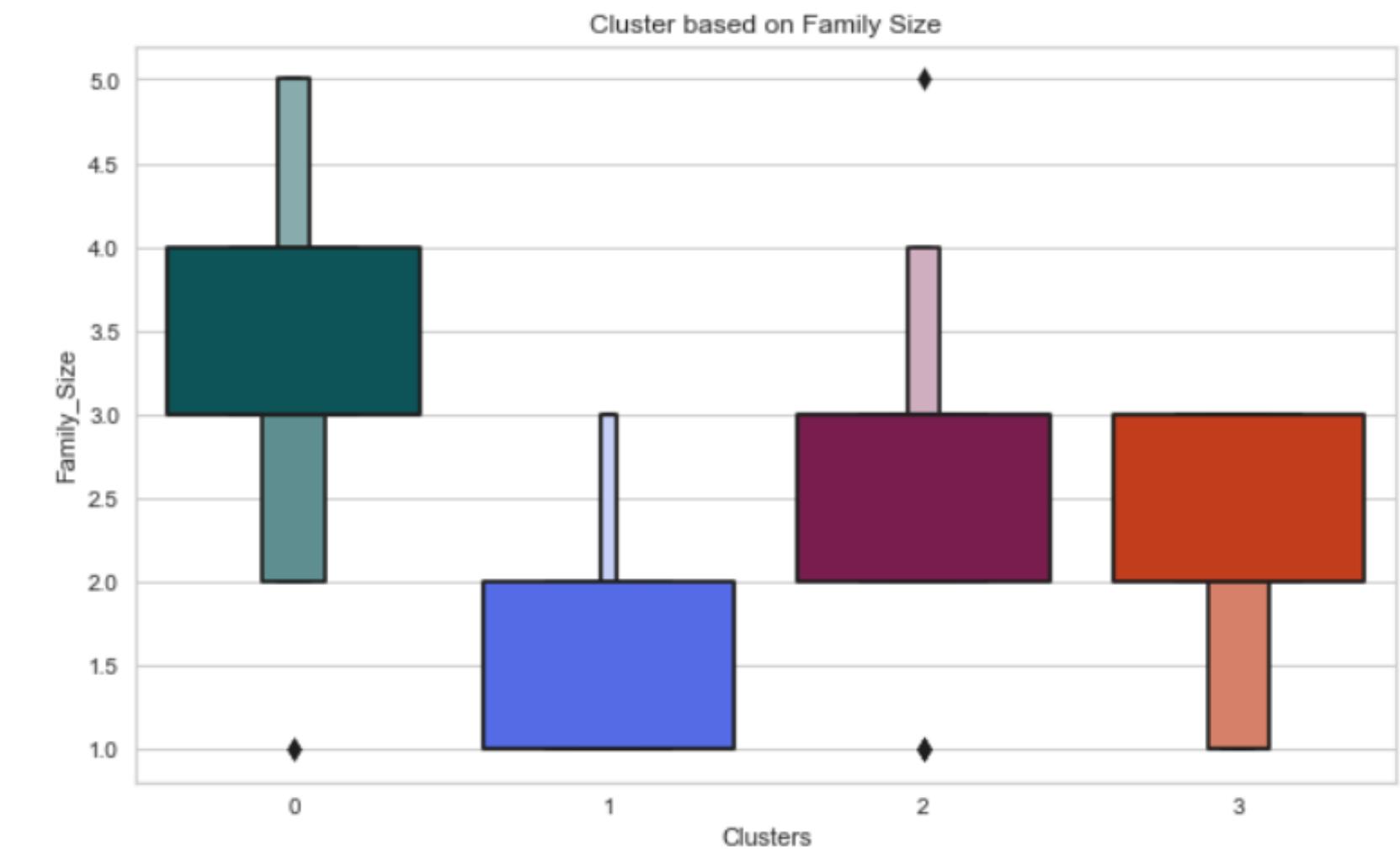
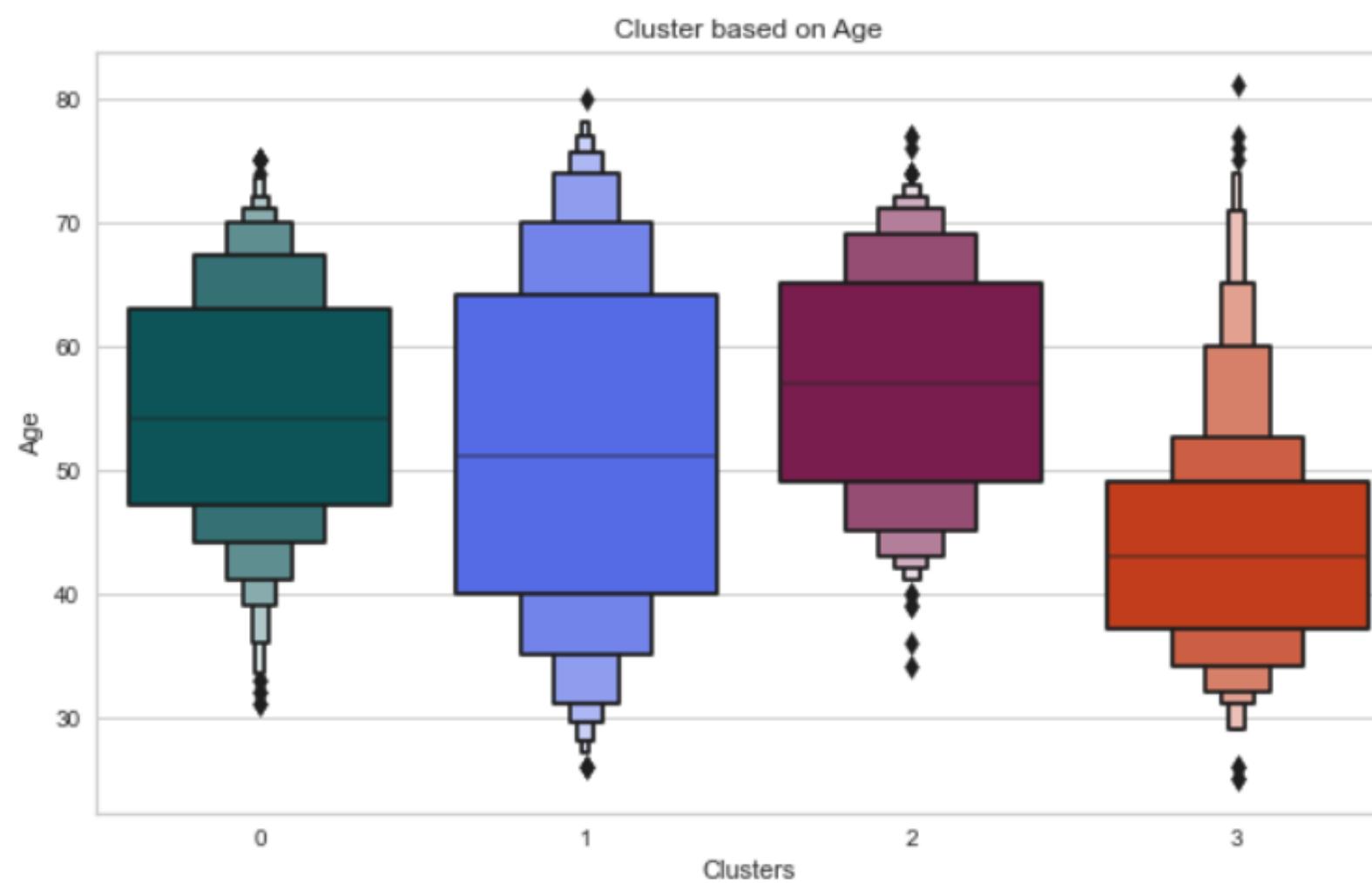


The numbers confirm that the customers were collected correctly.

Evaluation Models.

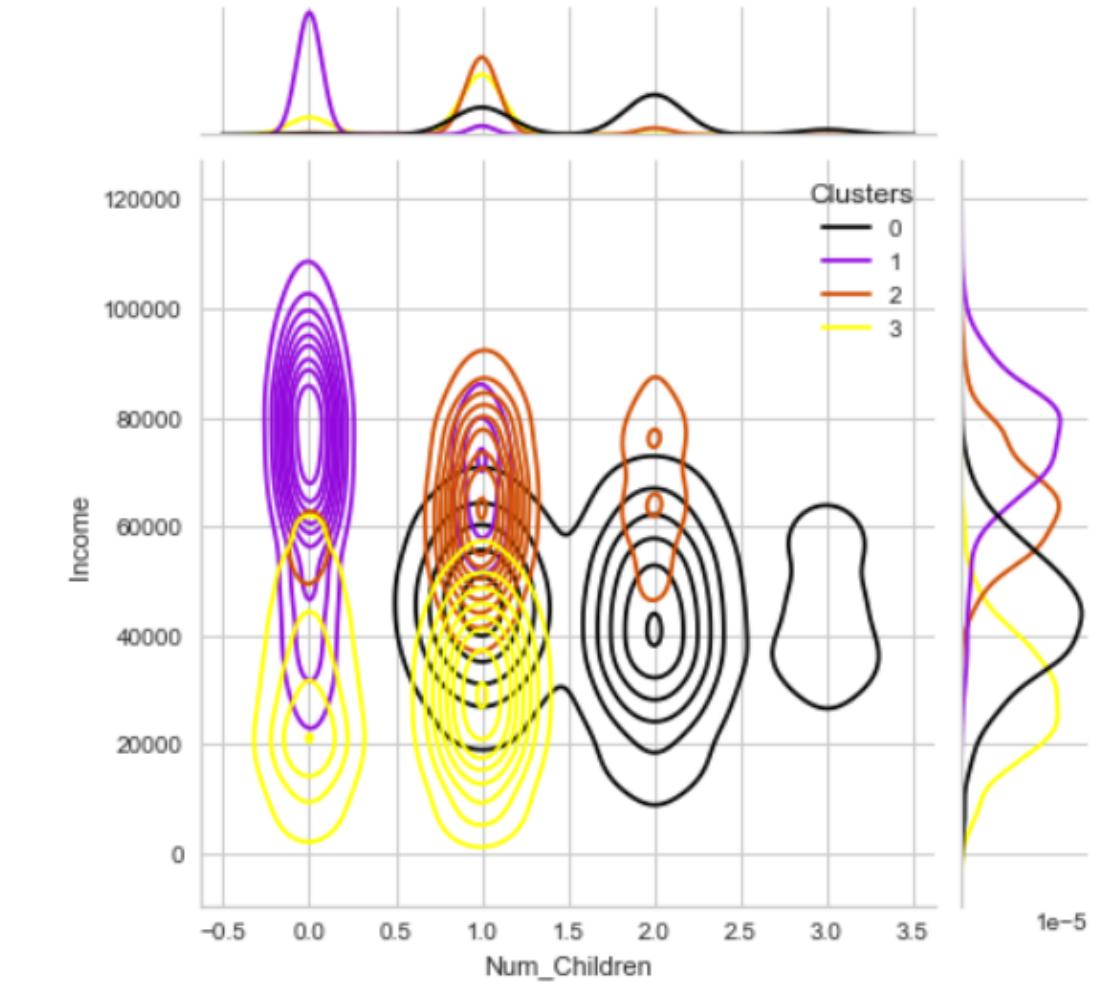
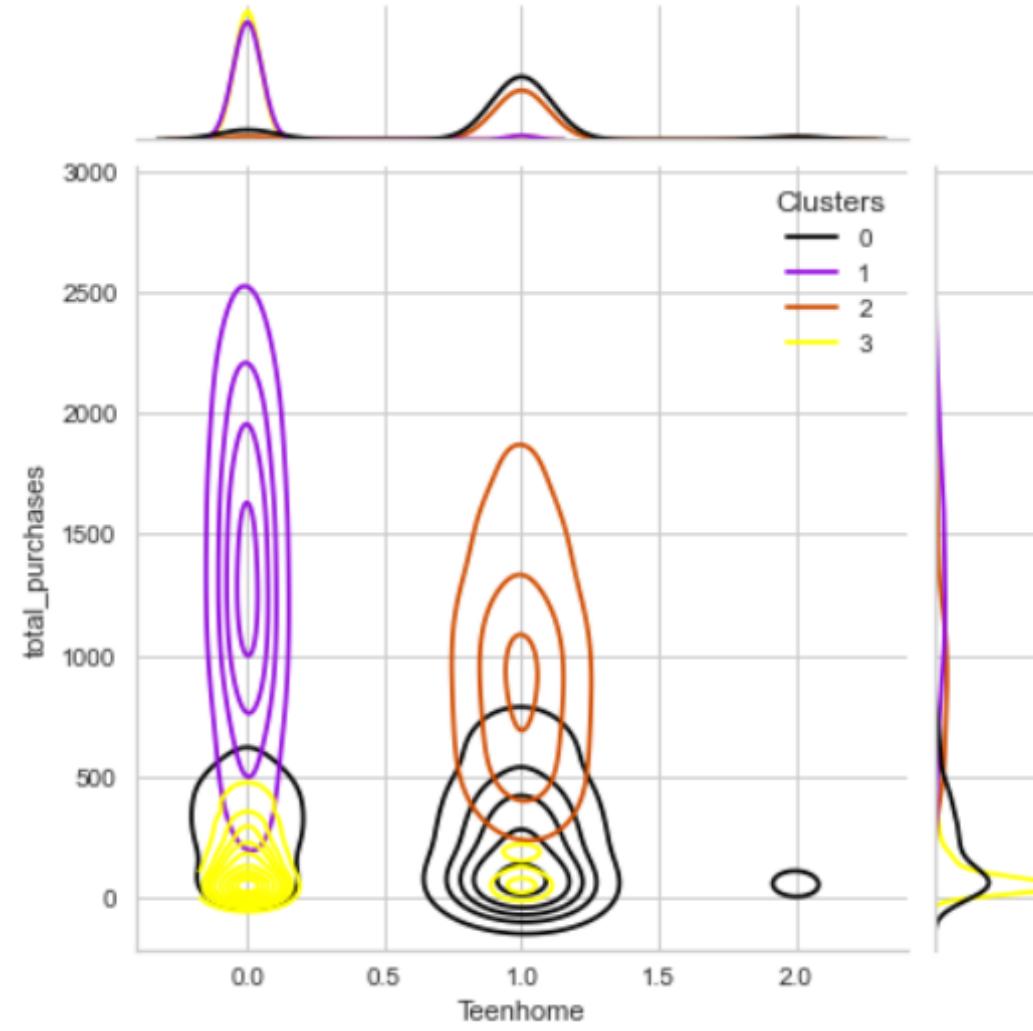
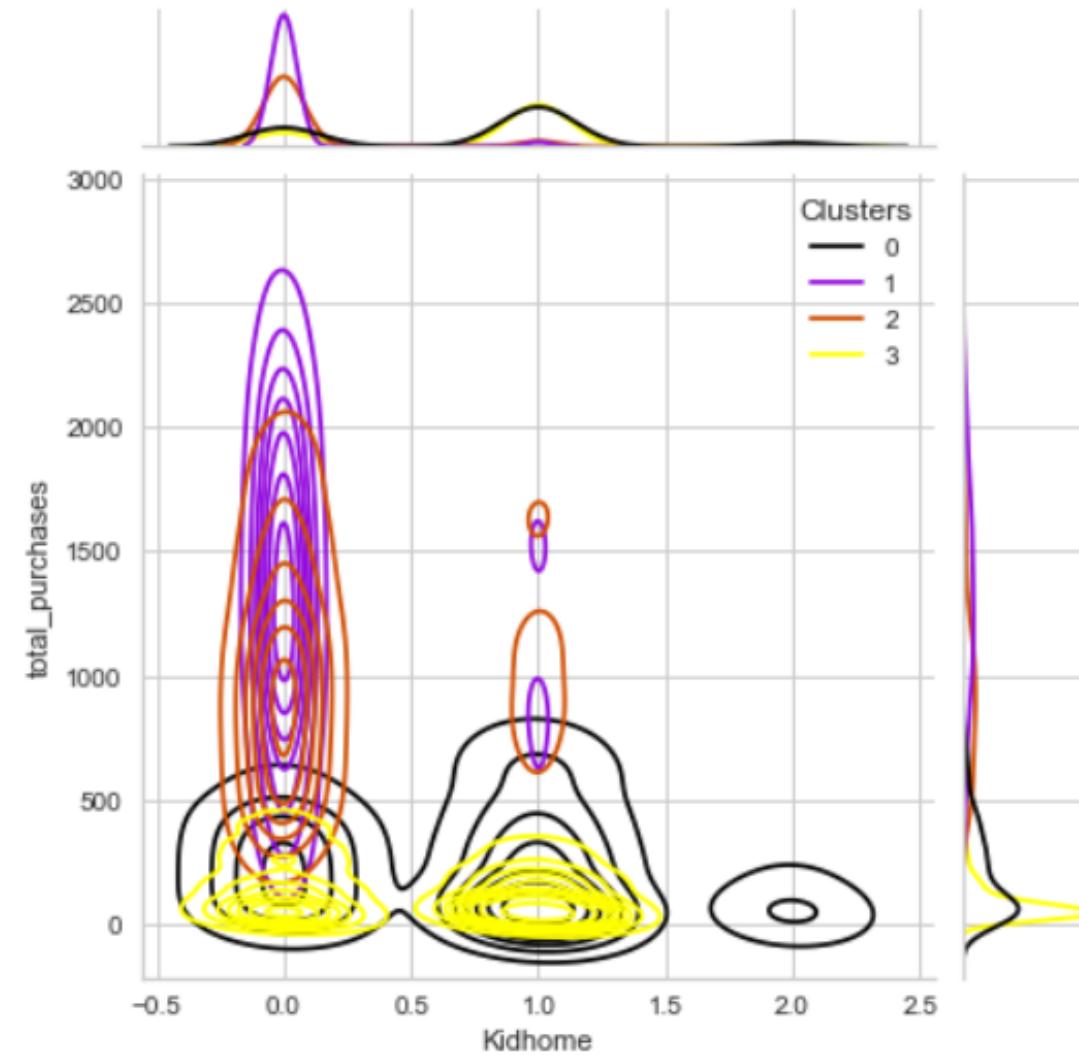


Agglomerative Clustering

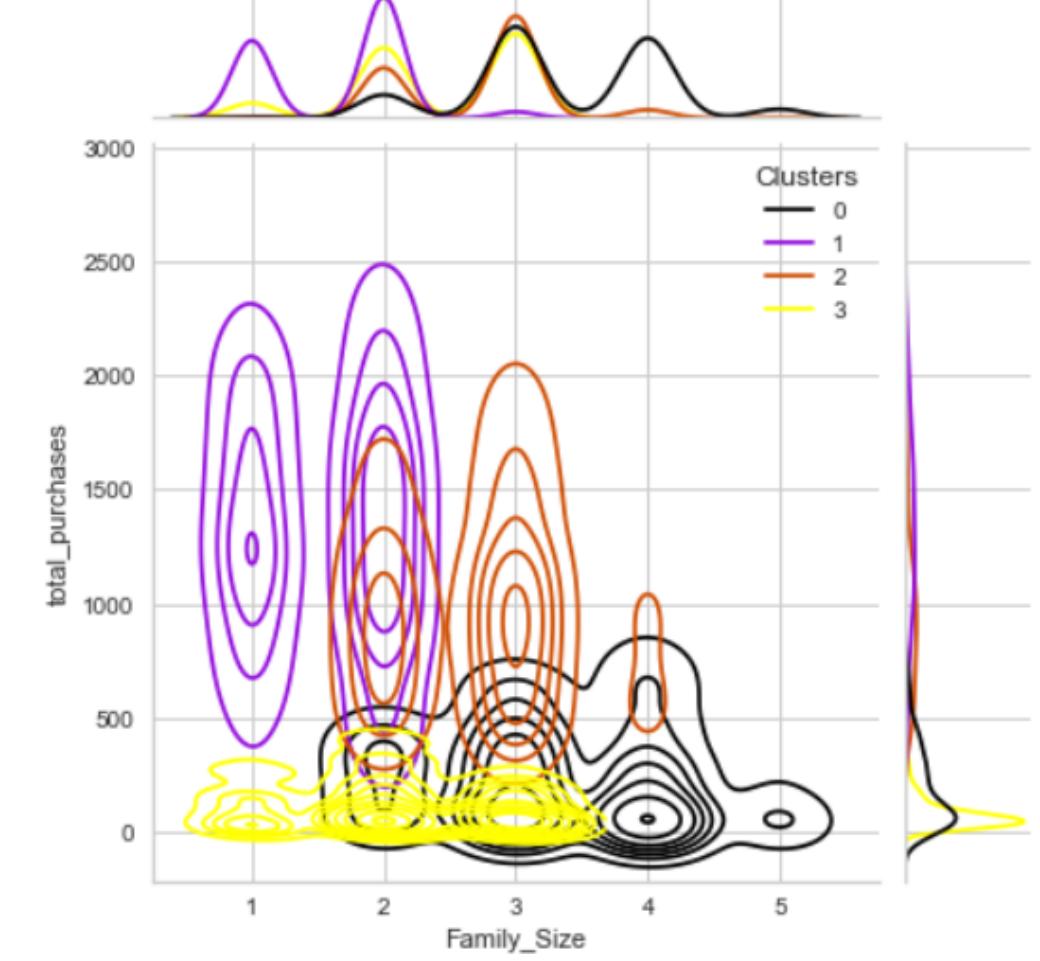
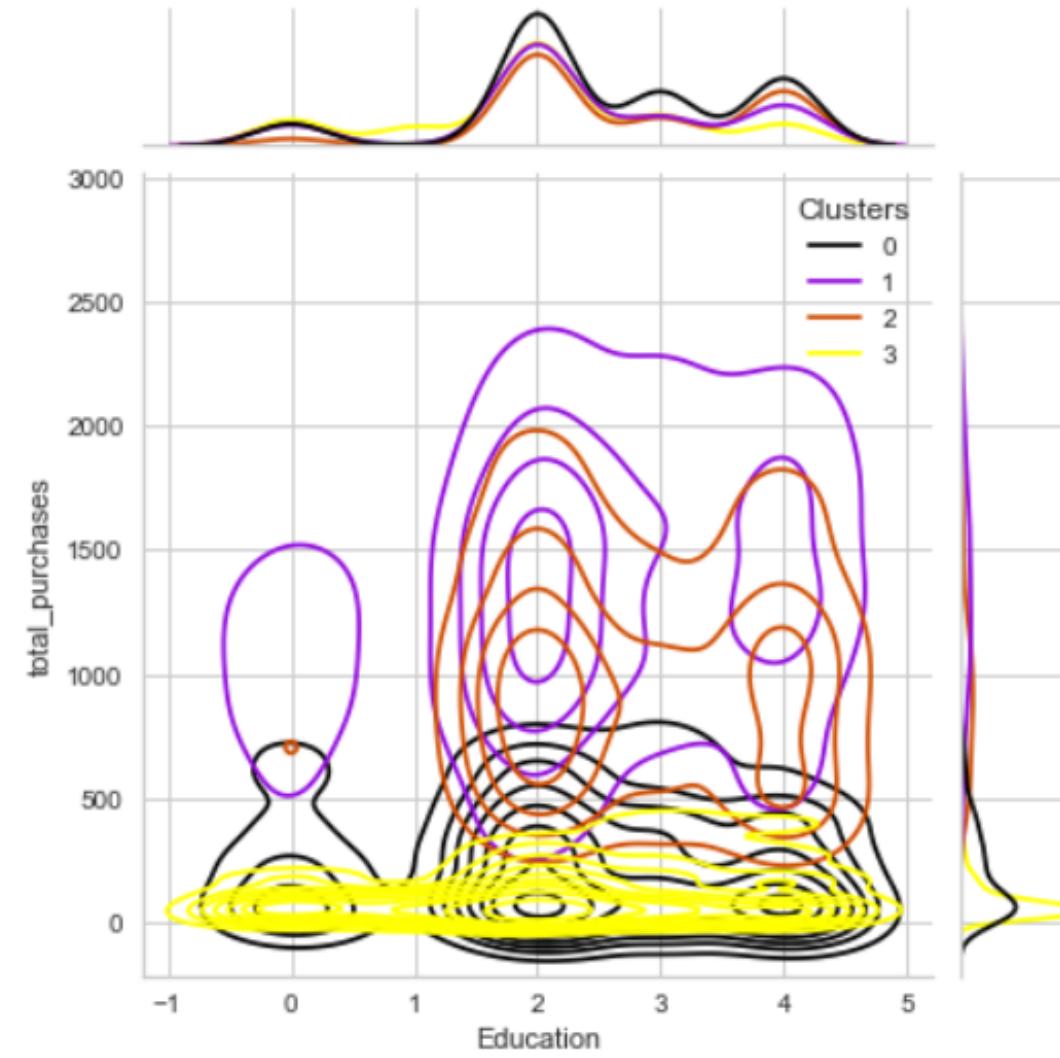
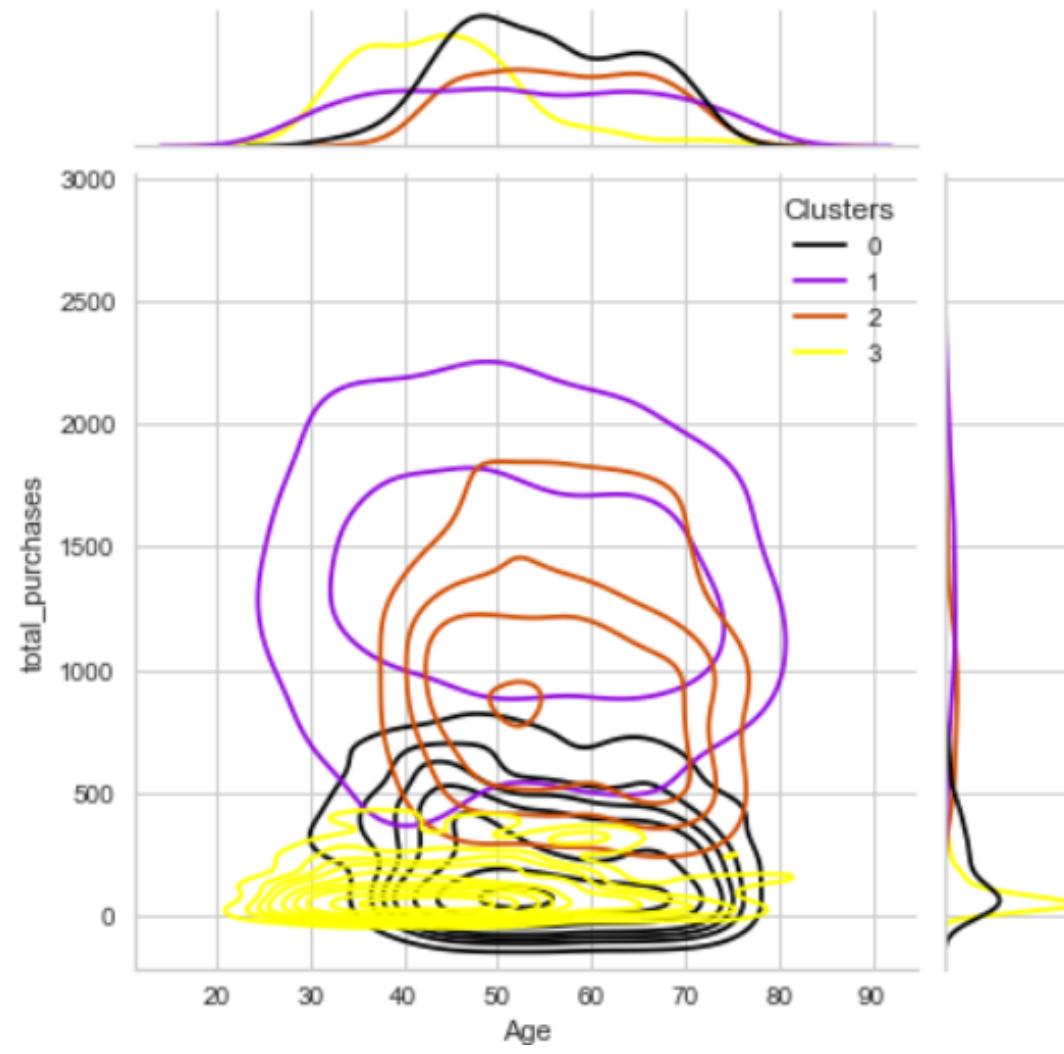


The numbers confirm that the customers were collected correctly.

Evaluation Models.



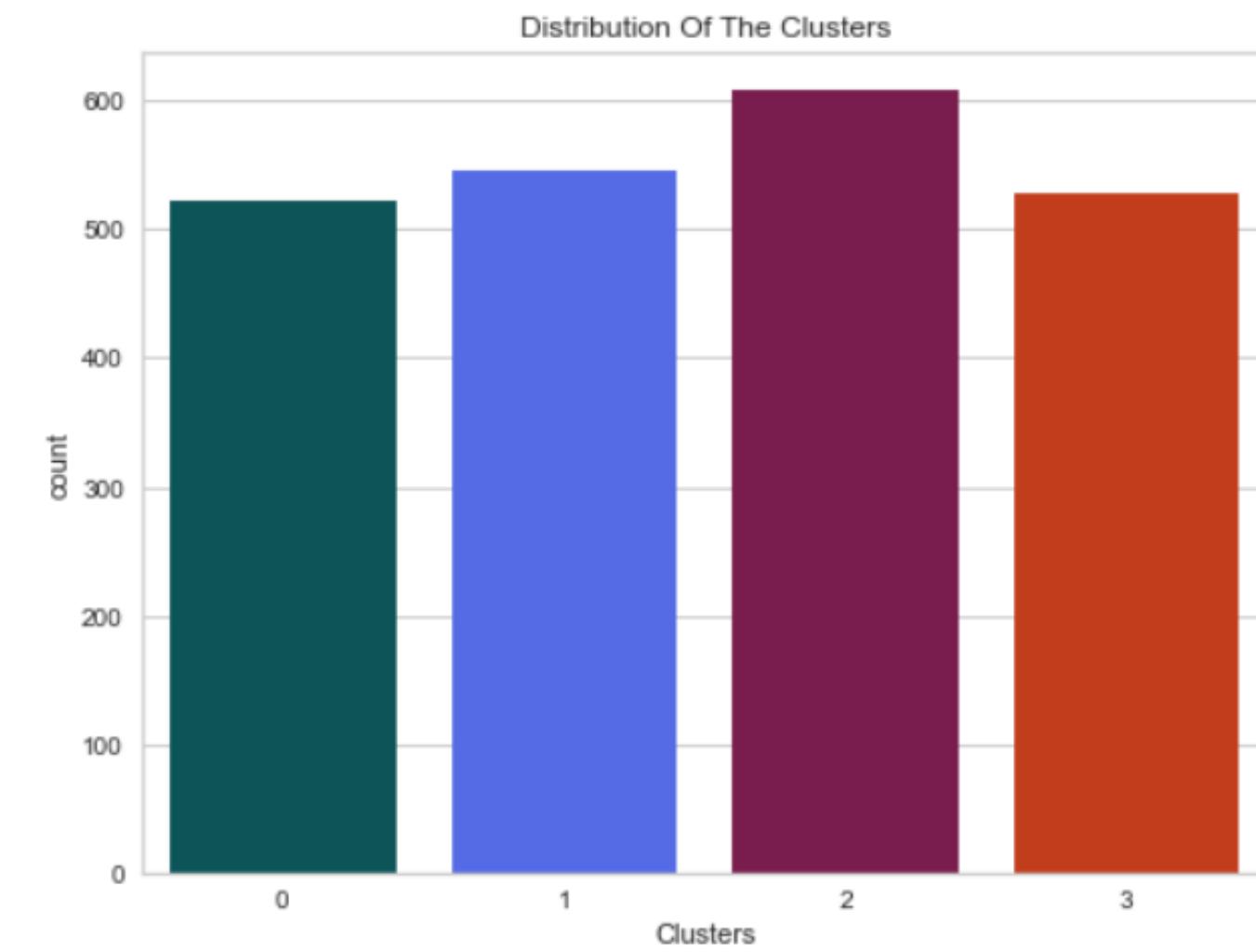
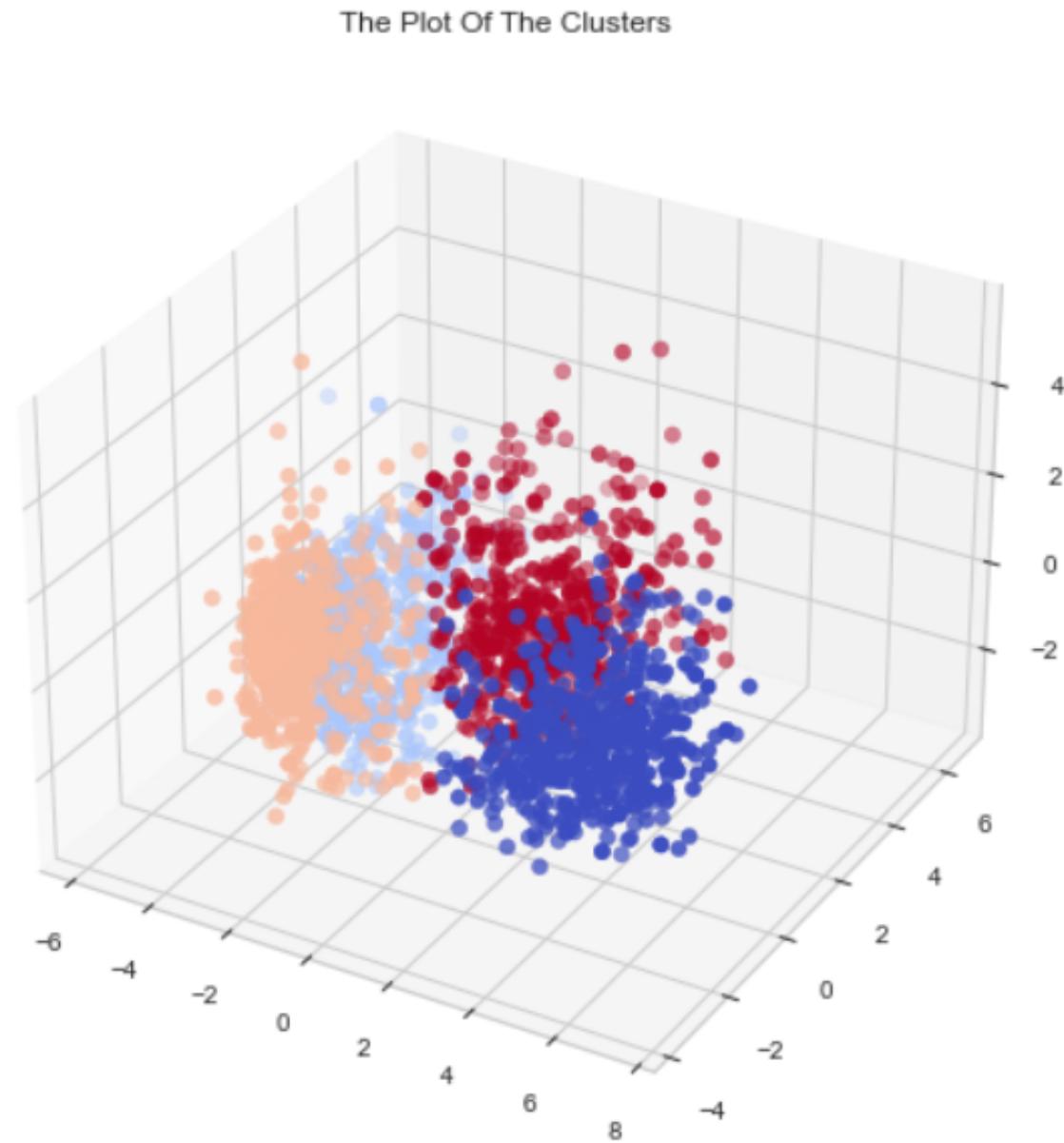
Evaluation Models.



Evaluation Models.

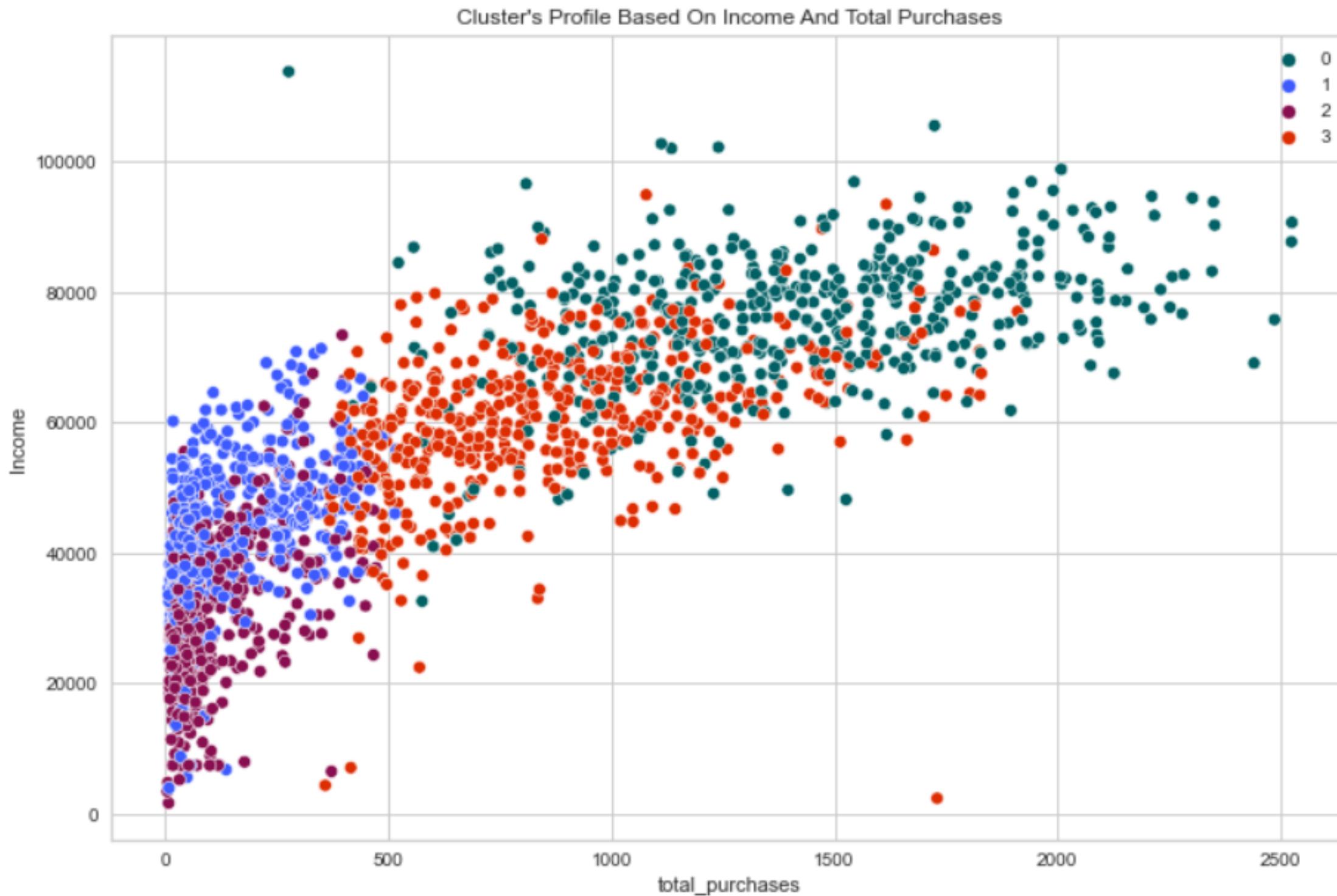


A 3D Projection of Data After clustering using KMeans



Show numbers of clustering

Evaluation Models.



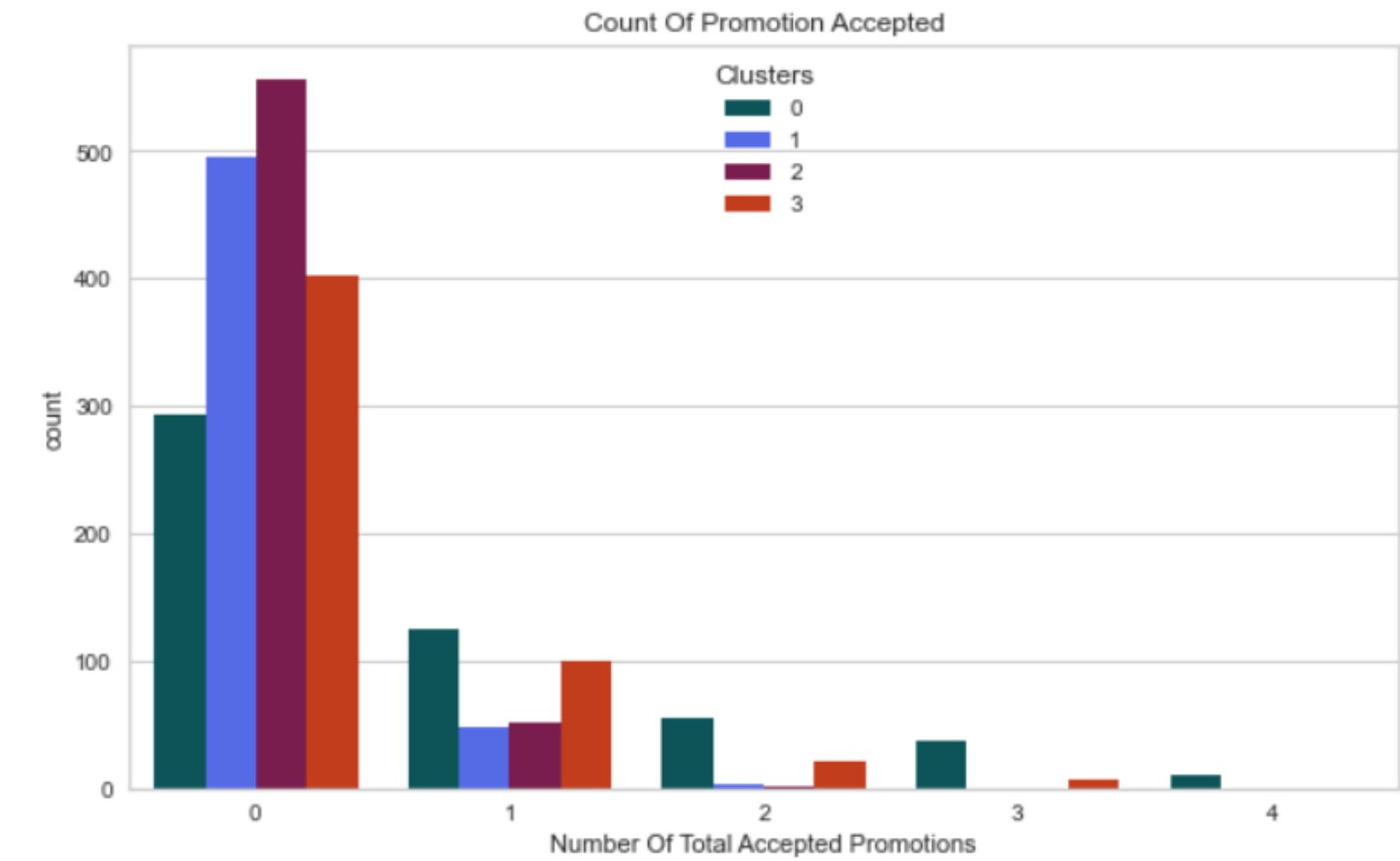
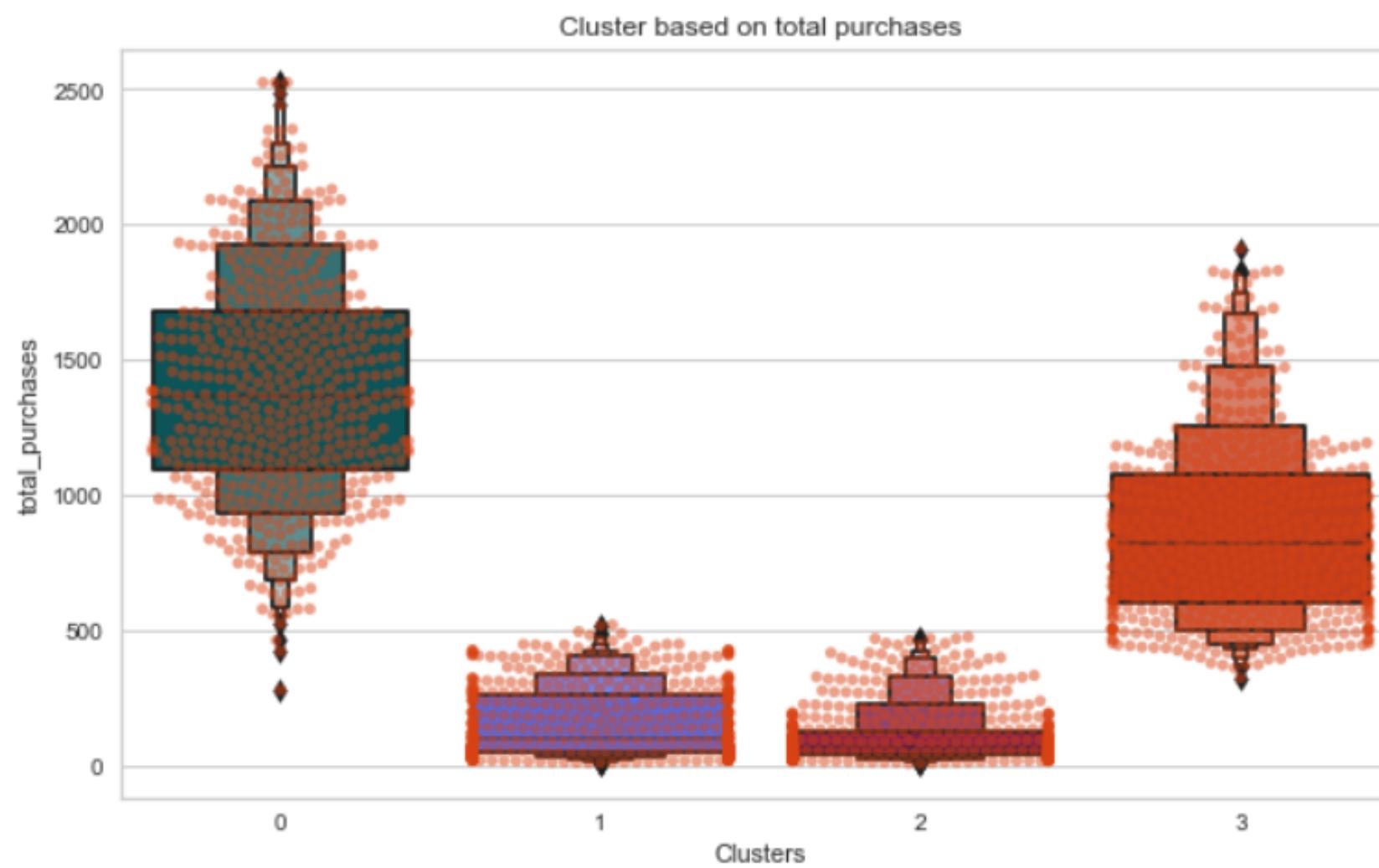
KMeans

The figure shows that the classification process is based on the linear relationship between income and the number of purchases.

Evaluation Models.



KMeans

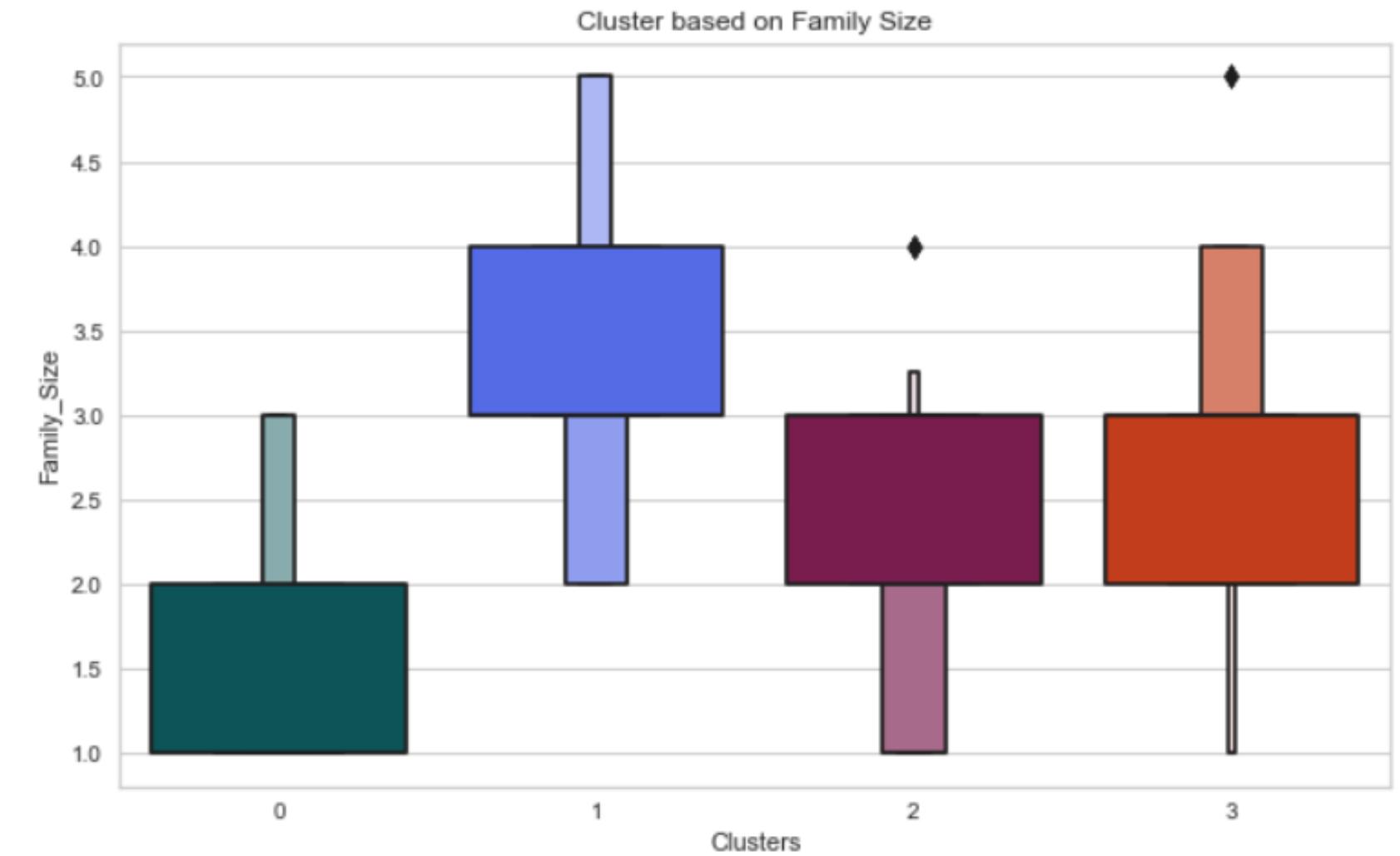
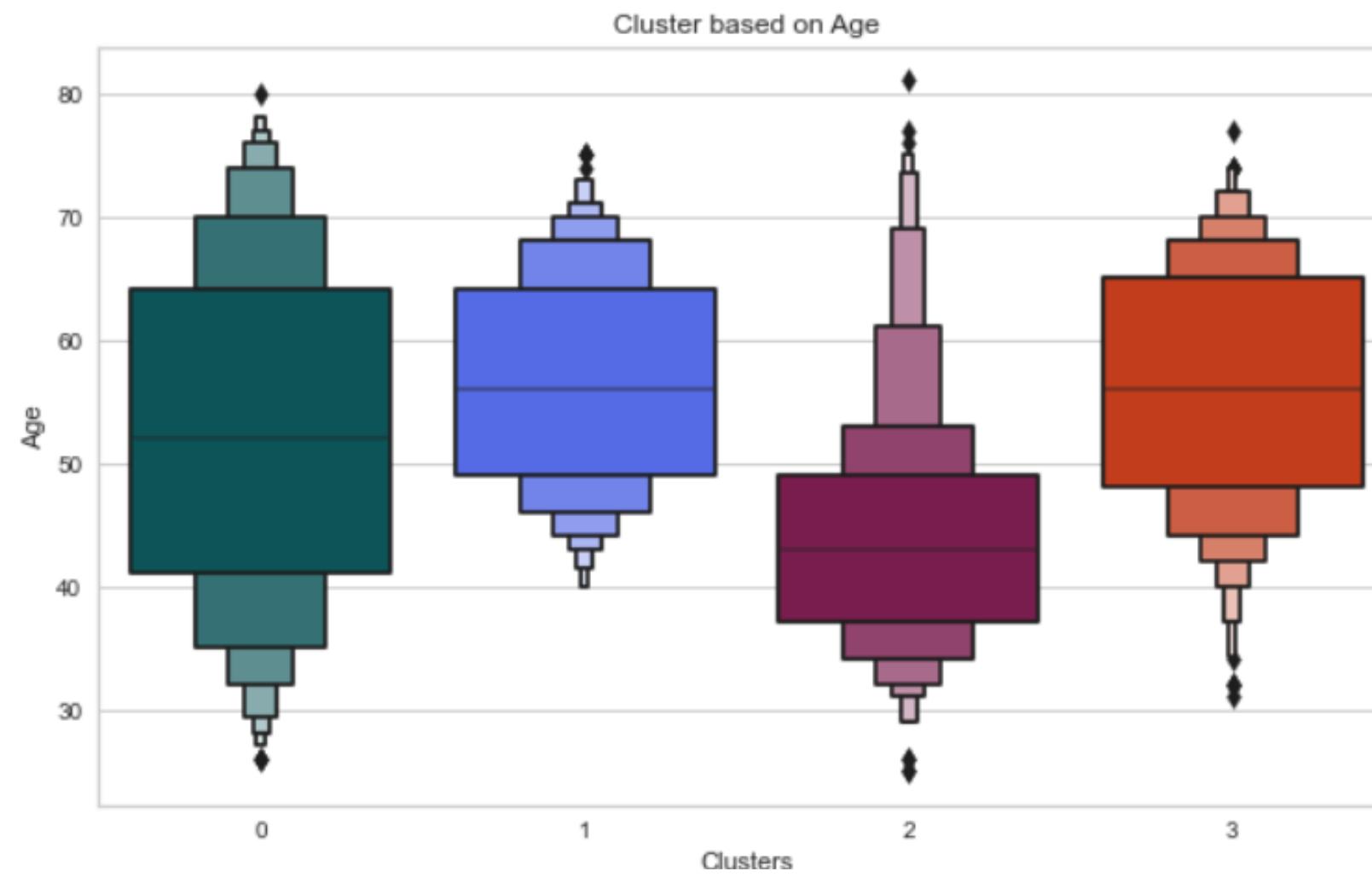


The numbers confirm that the customers were collected correctly.

Evaluation Models.

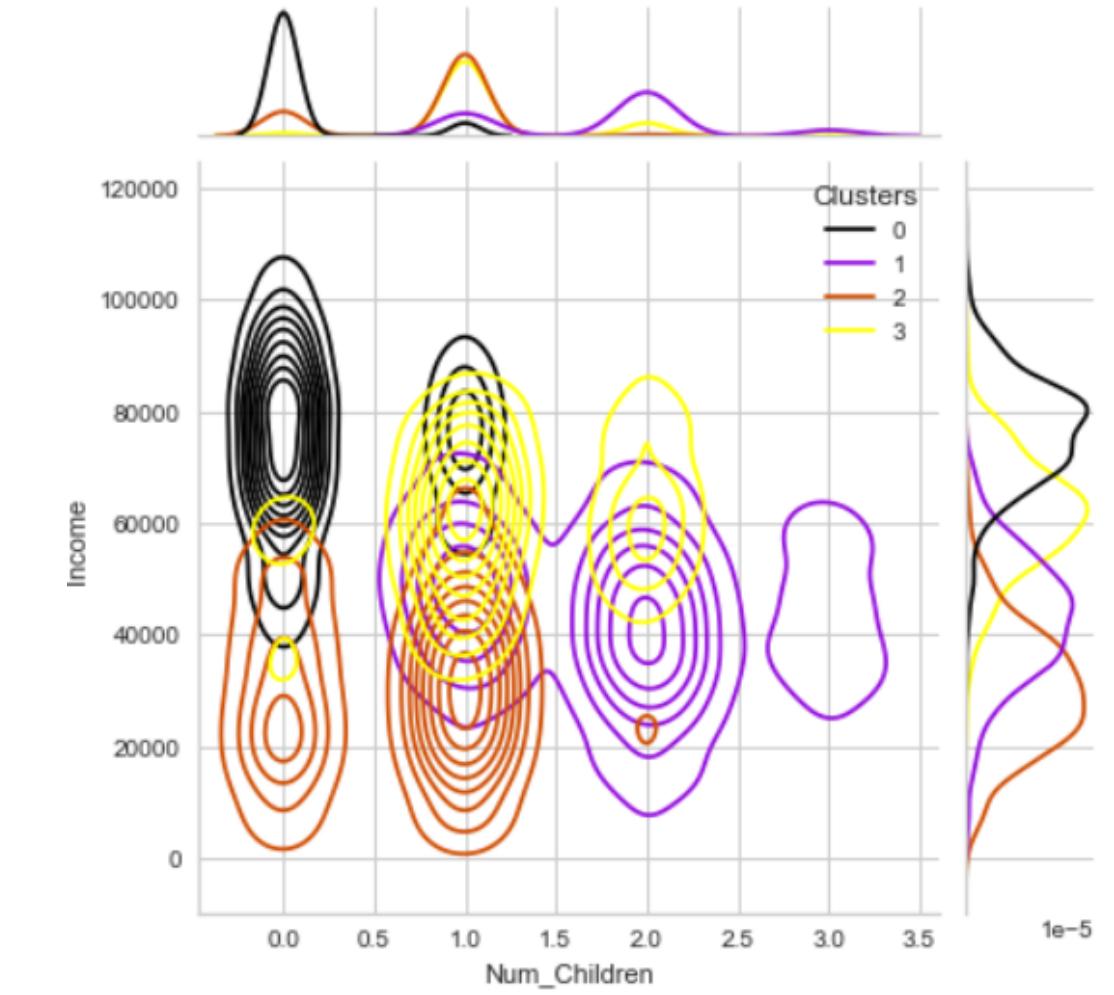
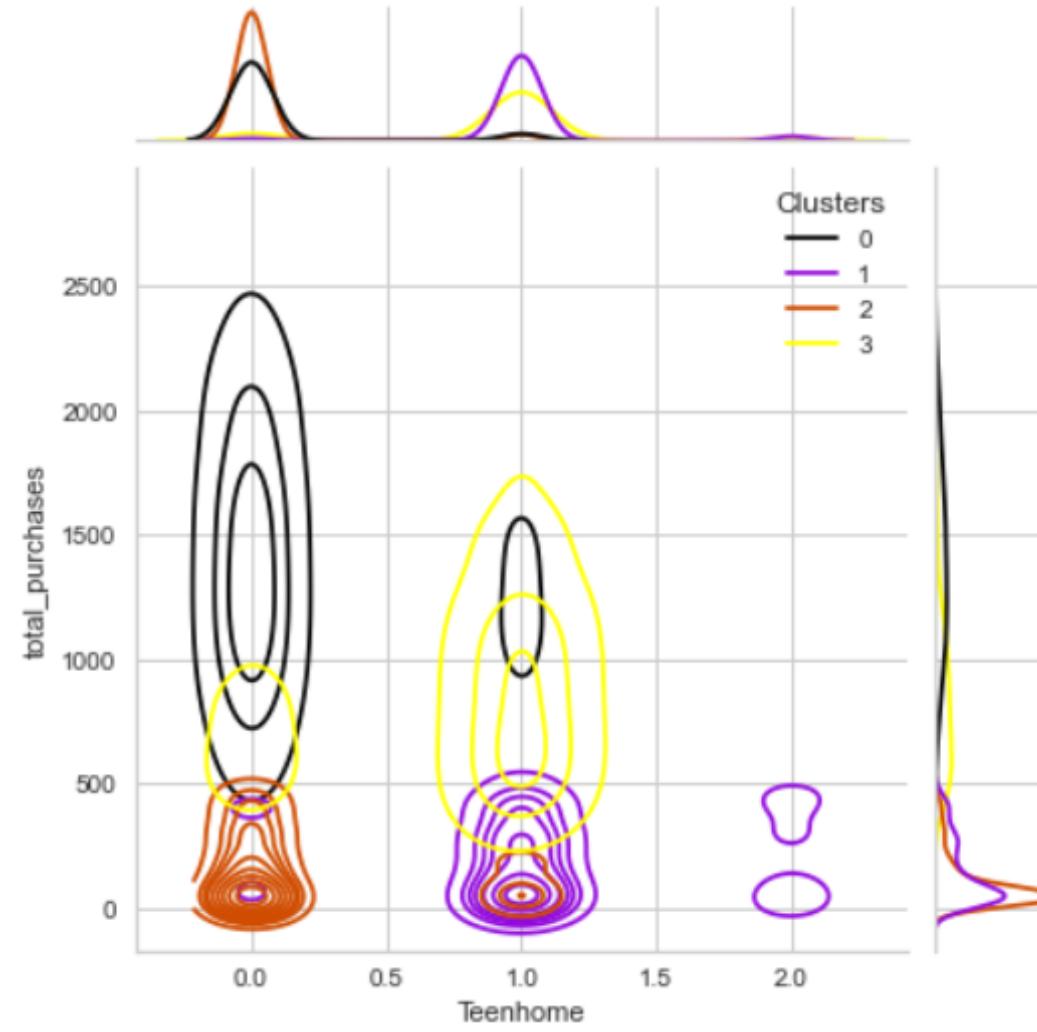
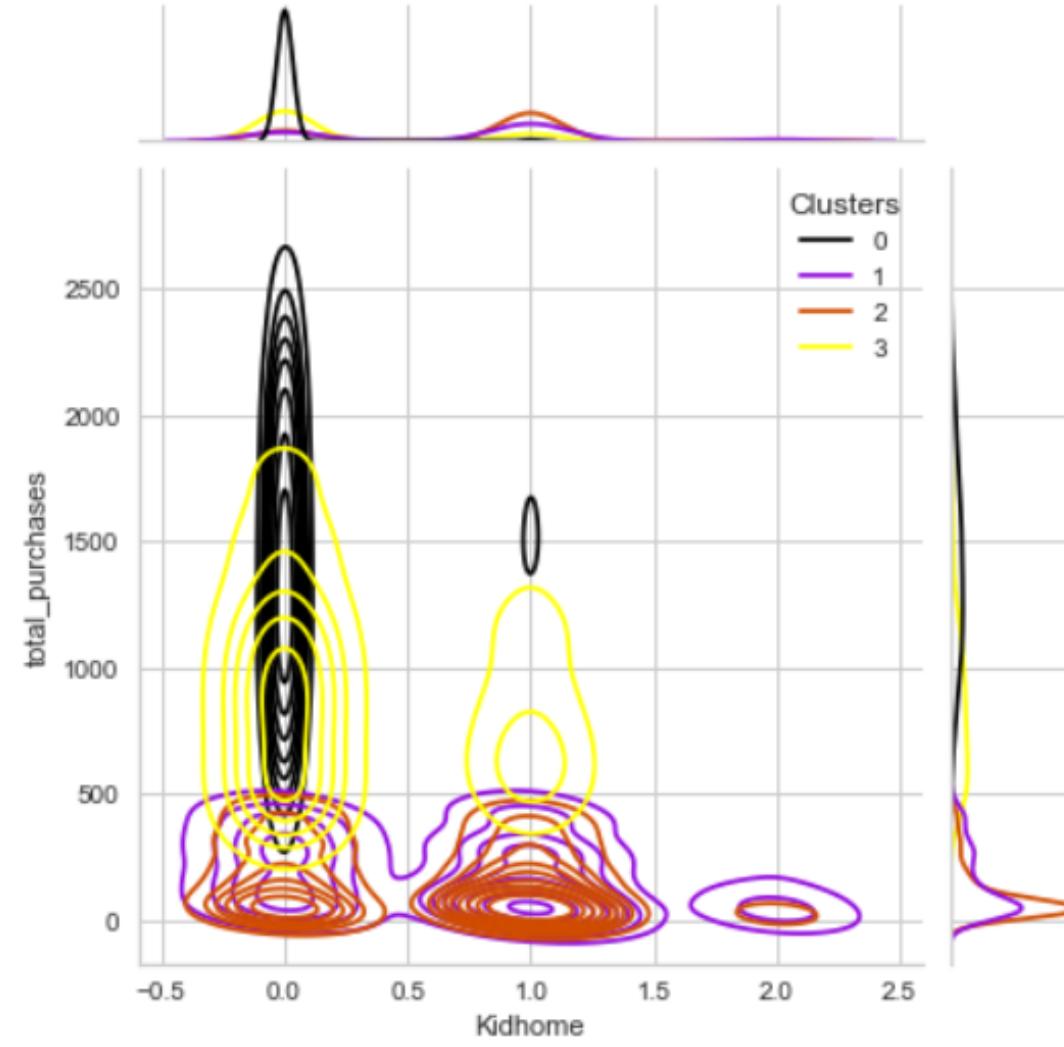


KMeans

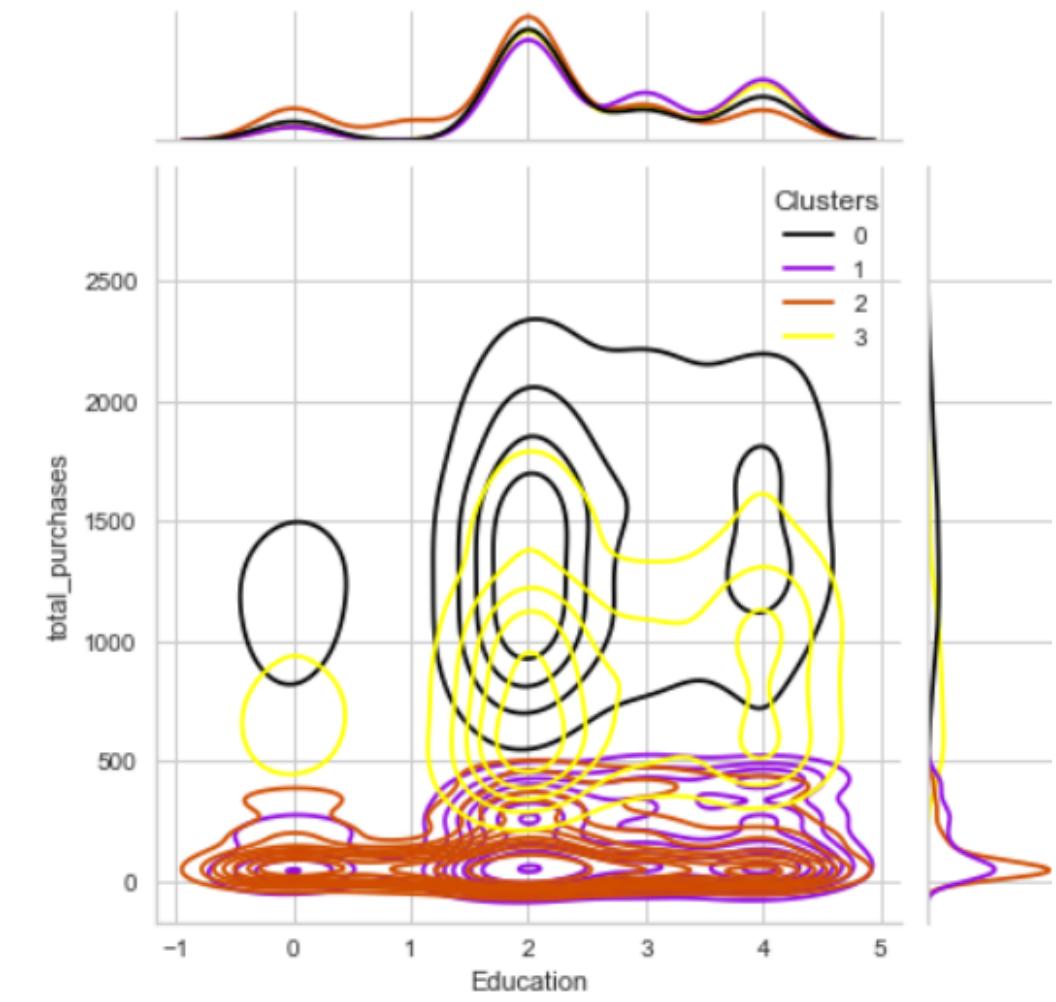
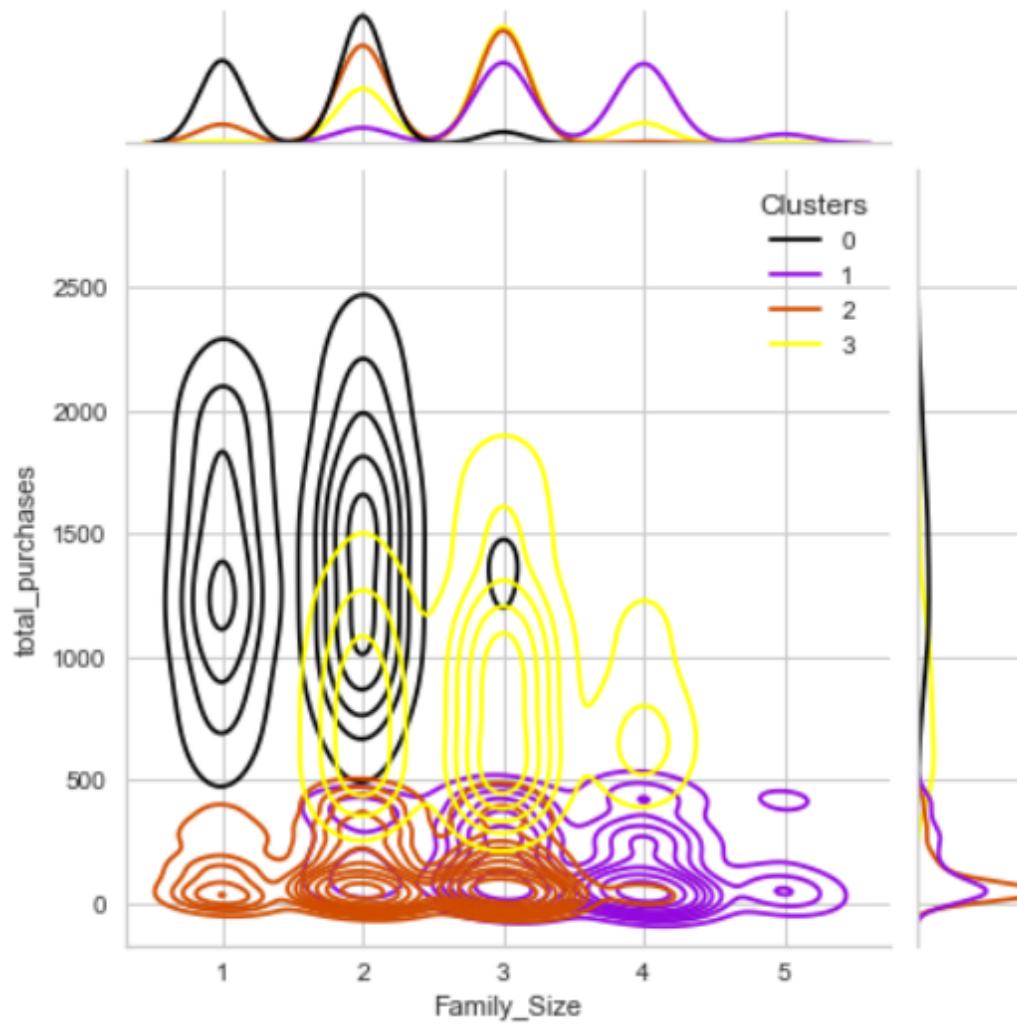
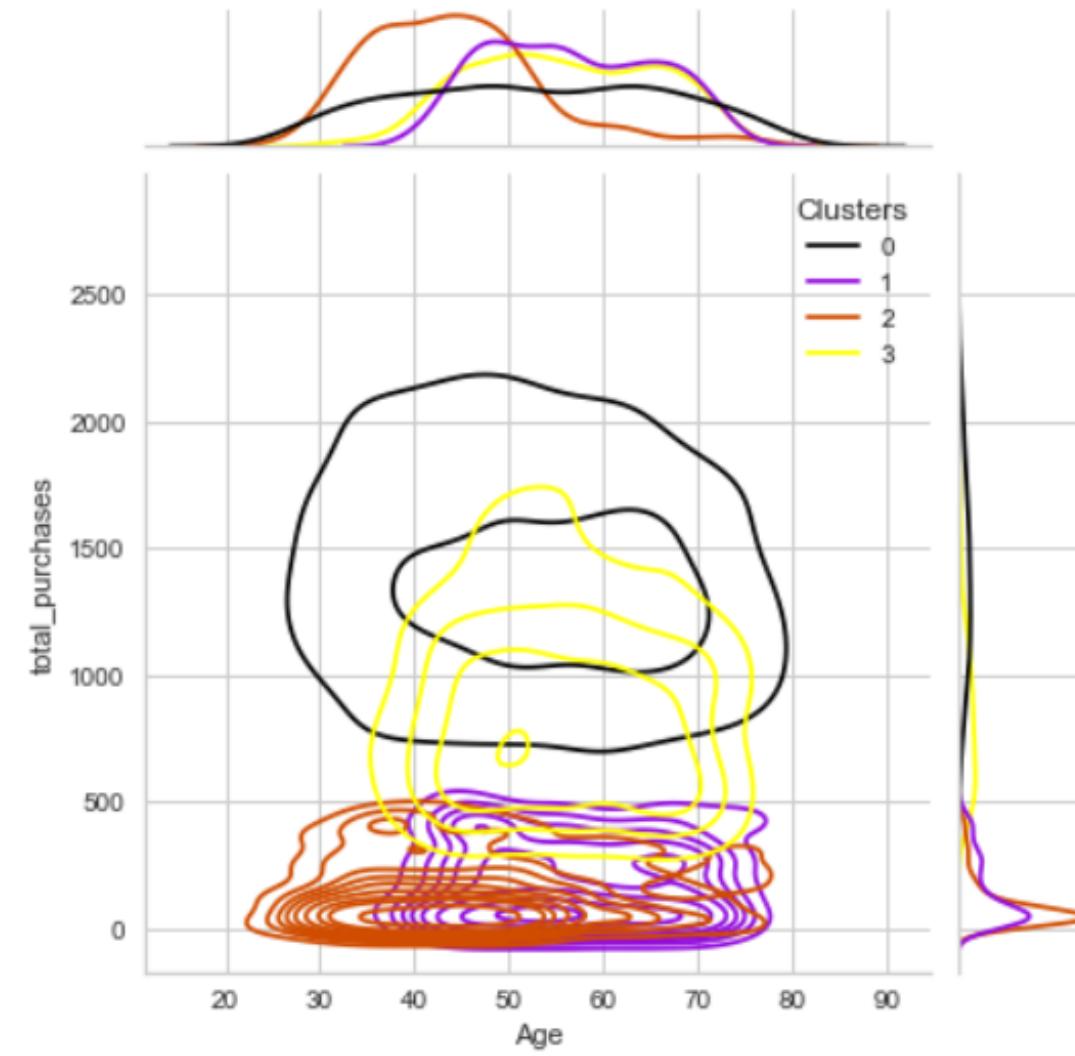


The numbers confirm that the customers were collected correctly.

Evaluation Models.



Evaluation Models.



Clusters 0

- The min age of the customer is 30 and the max age is 75.
- The min number of children is 0 and the max is 3.
- The min number of kids is 0 and the max is 2.
- The min number of teens is 0 and the max is 2.
- The number of purchases reaches 600.
- Income is between 20,000 and 70,000.

Clusters 2

- The min age of the customer is 35 and the max age is 77.
- The min number of children is 1 and the max is 2.
- The min number of kids is 0 and the max is 1.
- The min number of teens is 0 and the max is 1.
- The number of purchases starts from 500 to 2200.
- Income between 40,000 to 900,000.

Clusters 4

- The min age of the customer is 26 and the max age is 80.
- The min number of children is 0 and the max is 1
- The min number of kids is 0 and the max is 1.
- The min number of teens is 0 and the max is 1.
- The number of purchases starts from 500 to reaches the most purchases.
- The income is between 50,000 and 110,000.

Clusters 3

- The min age of the customer is 25 and the max age is 82.
- The min number of children is 0 and the max is 2.
- The min number of kids is 0 and the max is 1.
- The min number of teens is 0 and the max is 1.
- The number of purchases is less than 500.
- Income up to 60,000.

Agglomerative Clustering

Conclusion

Clusters 4

- The min age of the customer is 40 and the max age is 75.
- The min number of children is 1 and the max is 3.
- The min number of kids is 0 and the max is 2.
- The min number of teens is 0 and the max is 2.
- The number of purchases reaches 500.
- Income is between 15,000 and 70,000.

Clusters 3

- The min age of the customer is 31 and the max age is 77.
- The min number of children is 0 and the max is 2.
- The min number of kids is 0 and the max is 1.
- The min number of teens is 0 and the max is 1.
- The number of purchases starts from 350 to 1900.
- Income between 35,000 to 90,000.

Clusters 0

- The min age of the customer is 26 and the max age is 80.
- The min number of children is 0 and the max is 1
- The min number of kids is 0 and the max is 1.
- The min number of teens is 0 and the max is 1.
- The number of purchases starts from 500 to 2525.
- The income is between 40,000 and 120,000.

Clusters 2

- The min age of the customer is 25 and the max age is 81.
- The min number of children is 0 and the max is 2.
- The min number of kids is 0 and the max is 2.
- The min number of teens is 0 and the max is 1.
- The number of purchases is less than 500.
- Income up to 65,000.

Conclusion

KMeans



Thank you!





SAMSUNG

Together for Tomorrow! Enabling People

Education for Future Generation

©2021 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.