CLUSTER ANALYSIS

Understanding the Clusters given by our algorithm

Quick measure

CONTEXT

Previously, we carried out an exhaustive analysis in which we broke down our dataset to define different types of Buyer Personas, interpreting and analyzing all the information provided by the variables in the customer list. We determined what types of customers we had, according to their demographic characteristics, consumption patterns, and purchasing preferences.

As a second step, we used these same features to develop a Machine Learning model based on the Clustering technique (K-means) to see if, through the algorithm, we could group customers into automatic clusters that would reveal new patterns for our labeling task.

Therefore, after including the new clusters in our clean dataset, we proceed to analyze it through an interactive dashboard in order to better understand it.

DEMOGRAPHICS

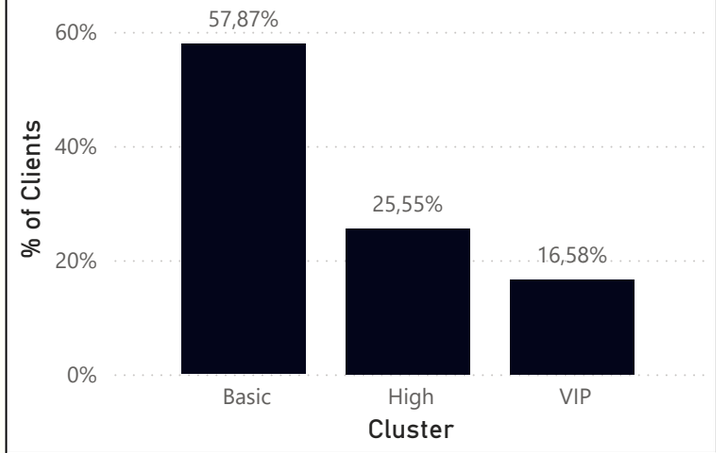
Total of Clients

1538

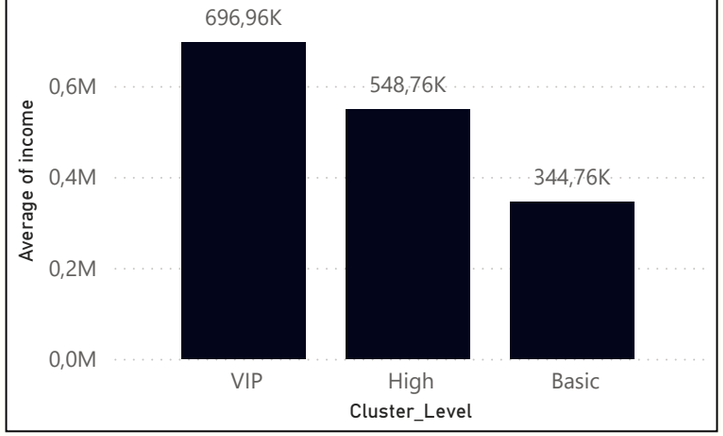
Cluster Filter

Basic | VIP | High

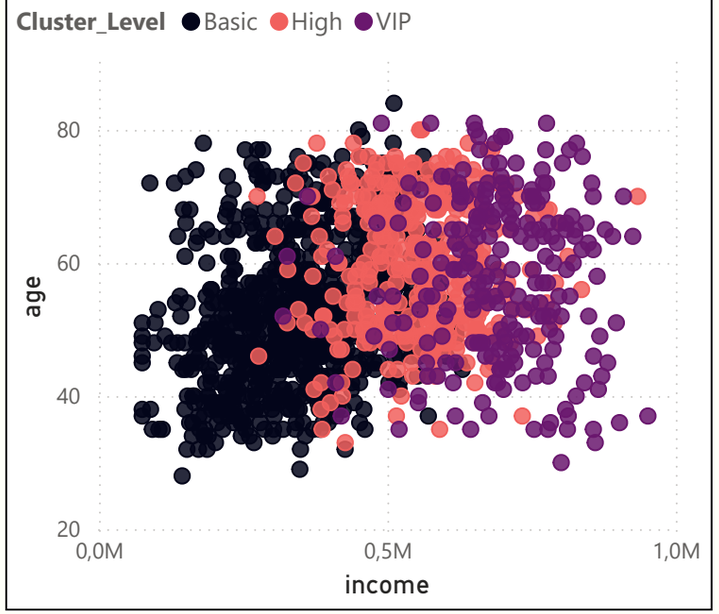
Clustered Clients



Average Income per Cluster Group

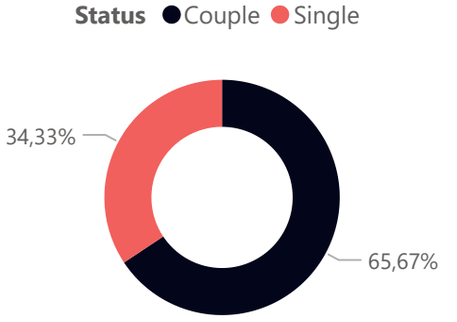


Relation between Income and Age

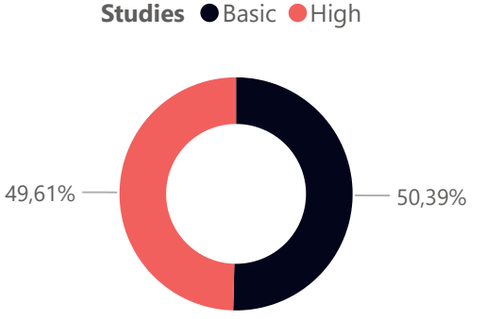


Cluster Type	Average of Age	Average of Children at Home
High	59,65	1,28
VIP	57,54	0,54
Basic	52,62	1,25

Marital Status



Level of Studies



Quick measure

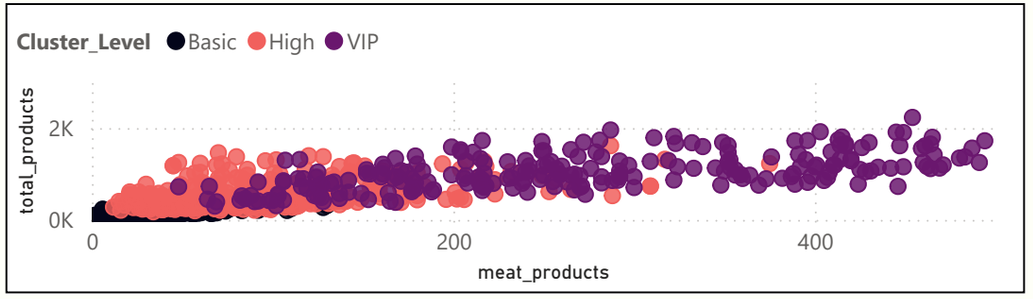
CUSTOMER PREFERENCES

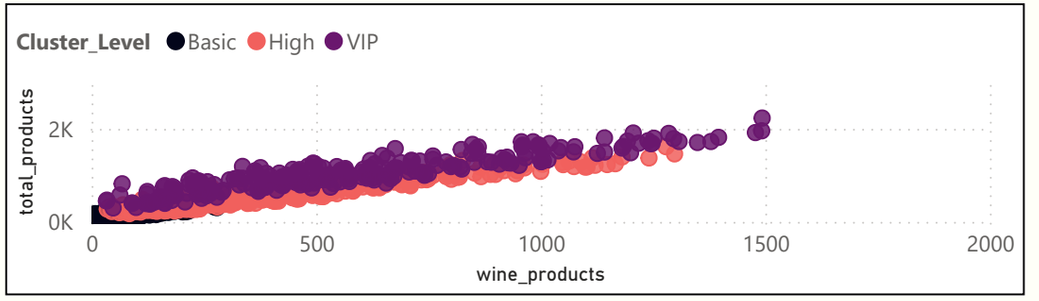
Total of Products

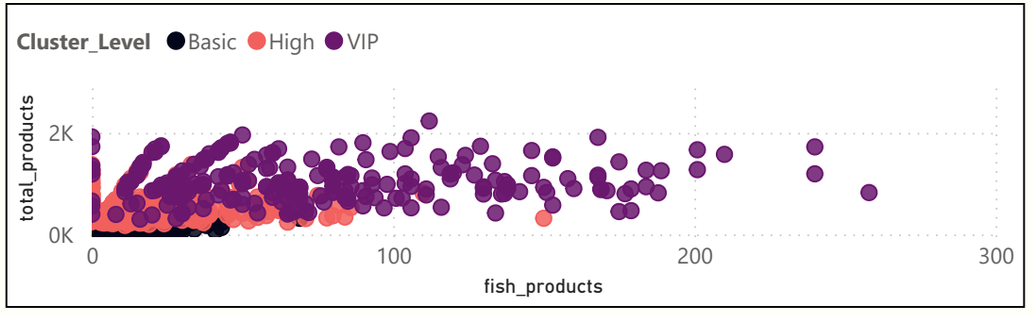
562K

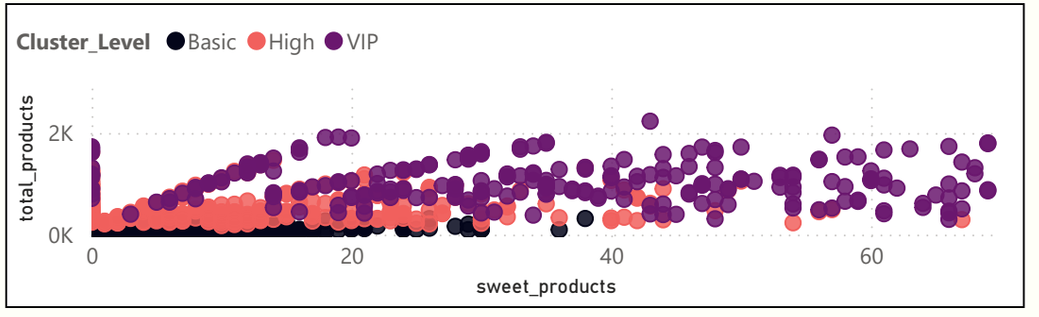
Cluster Filter

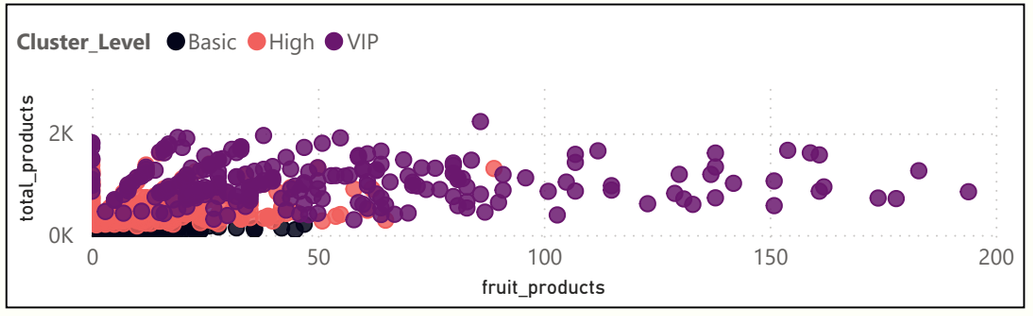
Basic | VIP | High

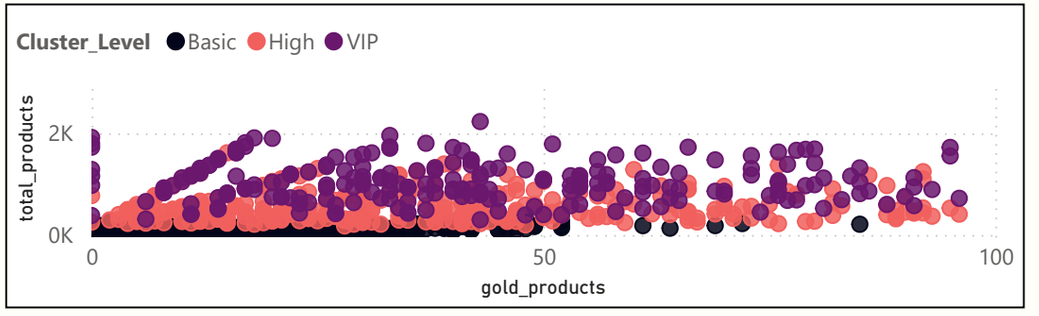












Quick measure

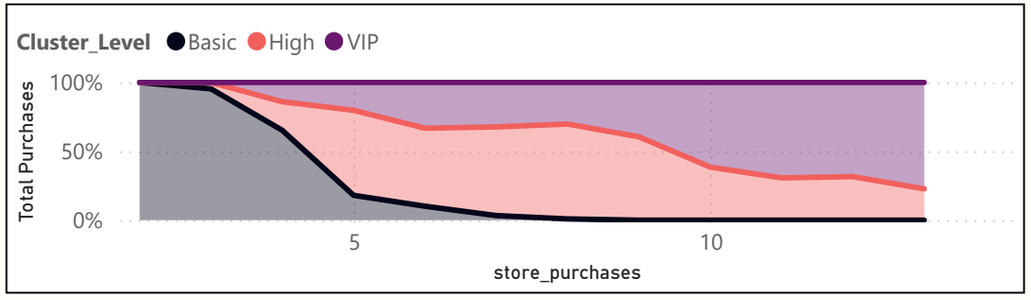
CUSTOMER BEHAVIOR

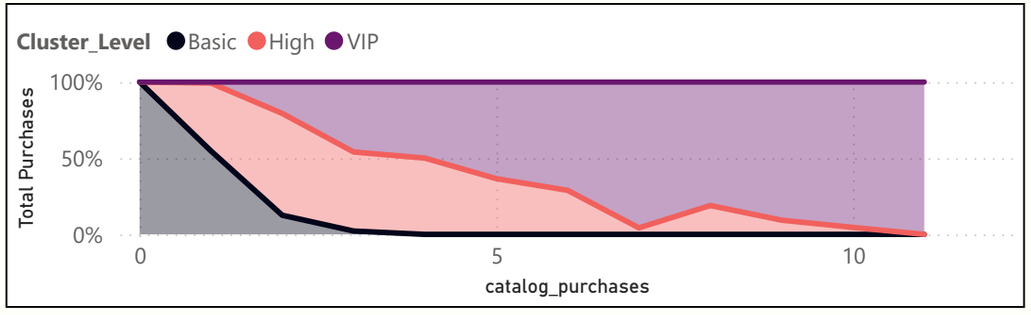
Total of Purchases

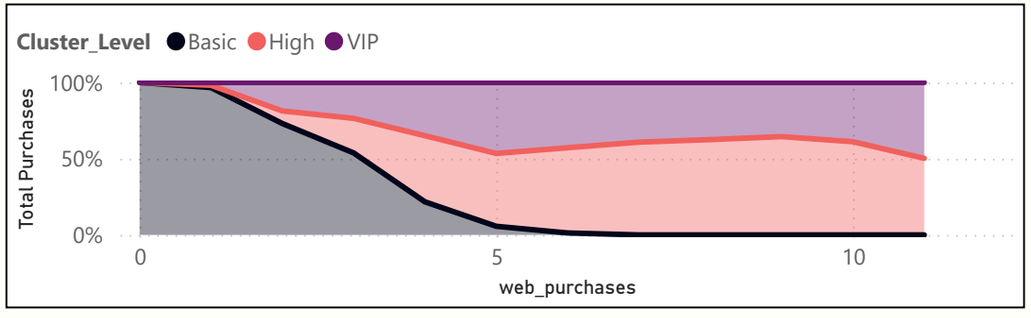
16K

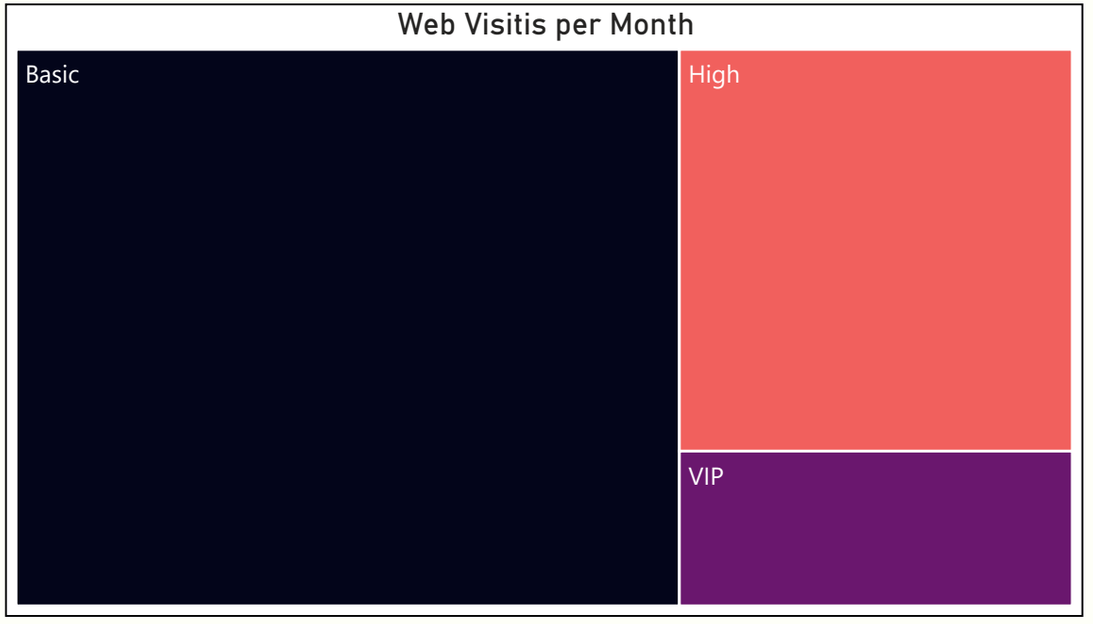
Cluster Filter

Basic | VIP | High









Cluster_Level	Total Purchases	Web Purchases	Store Purchases	Catalog Purchases	Deal Purchases
High	5932	2282	2716	934	1530
Basic	4896	1675	2810	411	1653
VIP	4862	1360	2219	1283	470
Total	**15690**	**5317**	**7745**	**2628**	**3653**

Quick measure

CONCLUSION

It is interesting to note how the patterns identified within the groups formed by the resulting clusters closely mirror the conclusions we had previously drawn from our analytical work. The three clusters (0, 1, and 2), which we renamed ‘Basic’, ‘High’, and ‘VIP’ due to their correlation with the ‘income’ characteristic of our customers, have segmented our entire customer base into three classes with clearly defined consumption habits.

The ‘Basic’ group is the largest segment in our records (57%); however, it is also the group that makes the fewest purchases and acquires the least number of products.

The ‘High’ and ‘VIP’ groups are more similar to each other and differ significantly from the Basic group in terms of purchasing habits and acquisitions. Although neither of these two groups matches the size of the largest cluster, together they account for the majority of products purchased and transactions made.

The main difference between ‘VIP’ and ‘High’ is that the ‘VIP’ group reaches higher milestones in both Customer Preferences and Customer Behavior.