



Customer Personality Analysis

BUSINESS STRATEGY FOR
DATA SCIENTIST FINAL
PROJECT

LAMISS GARGOURI

Agenda

Introduction

Understanding the dataset

Modeling

Evaluating Models

Summary

Conclusion



A large, irregular pink brushstroke shape on the left side of the slide, serving as a background for the section header.

I/ Introduction

Customer personality analysis is a thorough examination of a business' ideal clients. It makes it simpler for businesses to adapt their goods to the unique wants, habits, and concerns of various consumer types. It also helps businesses better understand their clients.



THE SCOPE OF A DATA PROBLEM IS WHERE WE LOOK FOR STRUCTURE INITIALLY. A PROJECT SCOPE HAS FOUR COMPONENTS. THE PROJECT'S BACKDROP, THE NEEDS IT SEEKS TO ADDRESS, THE IDEALIZED PICTURE OF SUCCESS, AND ULTIMATELY THE INTENDED RESULTS ARE THE FOUR COMPONENTS.

A large, irregular pink brushstroke shape that serves as a background for the text.

II/ Understanding the dataset

df.info

<class 'pandas.core.frame.DataFrame'> RangeIndex: 2240 entries, 0 5 Kidhome 2240 non-null into 2239

Data columns (total 29 columns):

# Column		Non-Null Count	Dtype
---	-----	-----	----
0	ID	2240 non-null	int64
1	Year_Birth	2240 non-null	datetime64(ns)
2	Education	2240 non-null	object
3	Marital_Status	2240 non-null	object
4	Income	2216 non-null	float64
5	Kidhome	2240 non-null	int64
6	Teenhome	2240 non-null	int64
7	Dt_Customer	2240 non-null	datetime64(ns)
8	Recency	2240 non-null	int64
9	MntWines	2240 non-null	int64
10	MntFruites	2240 non-null	int64
11	MntMeatProducts	2240 non-null	int64
12	MntFishProducts	2240 non-null	int64
13	MntSweetProducts	2240 non-null	int64
14	MntGoldProds	2240 non-null	int64
15	NumDealsPurchases	2240 non-null	int64
16	NumWebPurchases	2240 non-null	int64
17	NumCatalogPurchases	2240 non-null	int64
18	NumStorePurchases	2240 non-null	int64
19	NumWebVisitMonth	2240 non-null	int64
20	AccptedCmp3	2240 non-null	int64
21	AccptedCmp4	2240 non-null	int64
22	AcceptedCmp5	2240 non-null	int64
23	AccptedCmp1	2240 non-null	int64
24	AcceptedCmp2	2240 non-null	int64
25	Complain	2240 non-null	int64
26	Z_CostContact	2240 non-null	int64
27	Z_Revenue	2240 non-null	int64
28	Response	2240 non-null	int64

Dtypes: datetime64(ns)(2), float64(1), int64(24), object(2)

Memory usage : 507.6+KB



Cleaning the data

- Basic data cleaning:

We don't sure what two variables (Z CostContact and Z Revenue) correspond to, so what should we do?

- *notation: They Only Hold One Value.
- *Single observation or value columns are probably not useful for modeling. These predictors or columns are known as zero-variance predictors because they would be eliminated if the variance (the average value from the mean) were assessed.

Always verify that the data rows' IDs match, as there occasionally are duplicate IDs.

- Marking and remove missing data
- Outlier identification and removal:

Despite the fact that all of these columns are numerical, some of them indicate categorical values, therefore we must choose columns in accordance with how we interpret the data.

A large, horizontal, pink brushstroke that serves as a background for the text. It has a soft, painterly texture with visible bristles and a slight gradient of pink color.

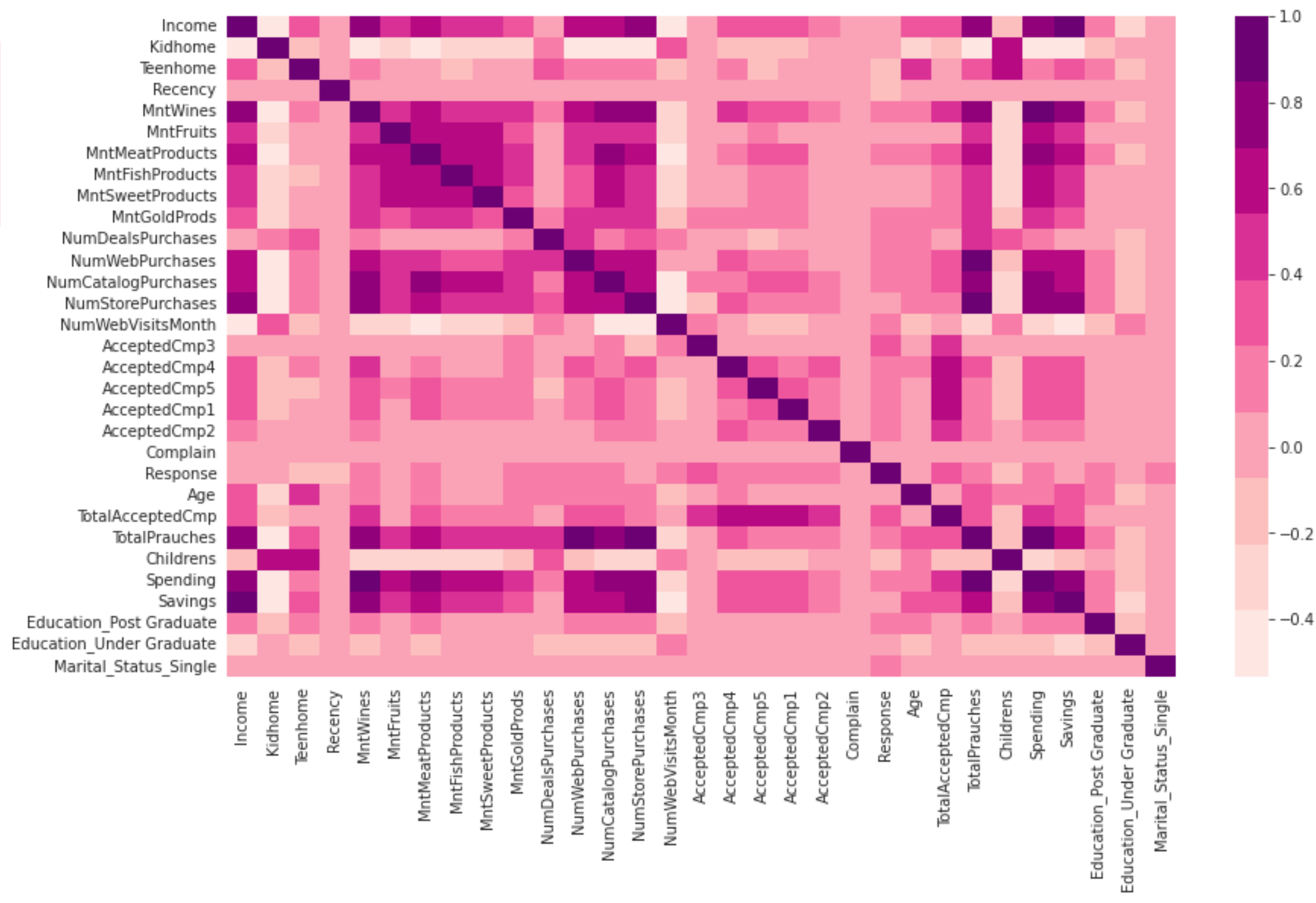
III/ Modeling

A large, horizontal, pink brushstroke graphic with a textured, hand-painted appearance. It is wider on the left and tapers slightly towards the right, with irregular, feathered edges.

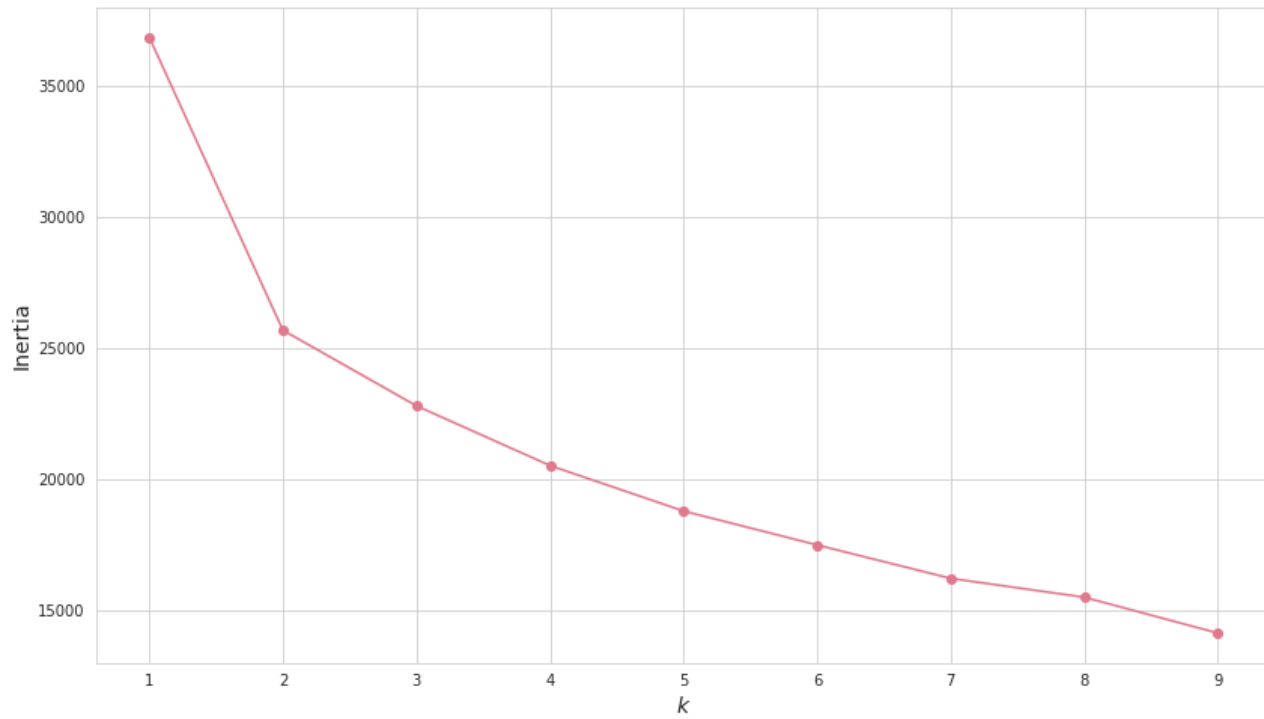
Modeling

- Dimensionality Reduction
- Searching for the Best Numbers Of Clustering:

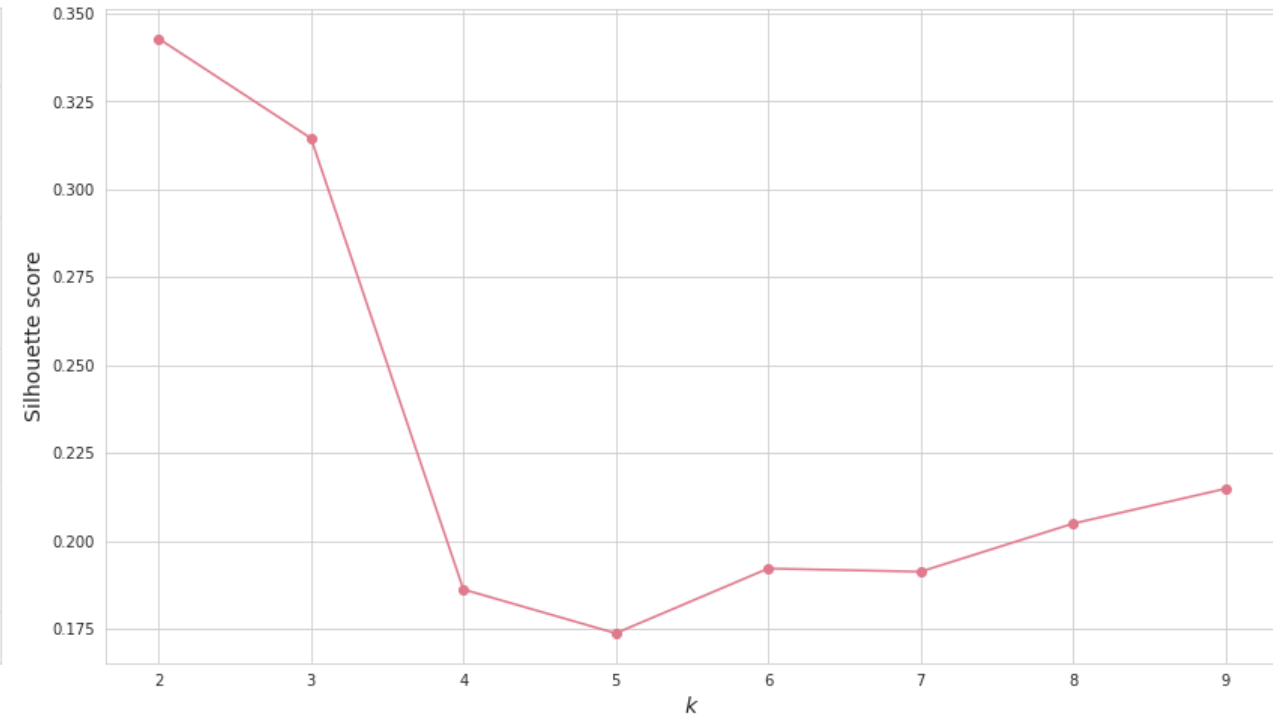
The elbow approach is a heuristic for figuring out how many clusters there are in a data collection. The process is graphing the explained variance as a function of the number of clusters, then choosing the number of clusters to employ at the elbow of the curve.



→ You can see that our features are strongly related to one another, thus we're going to use PCA to minimize the dimension.



➔ A statistic known as the silhouette score is used to evaluate how effective a clustering method is. Its value is between -1 and 1.



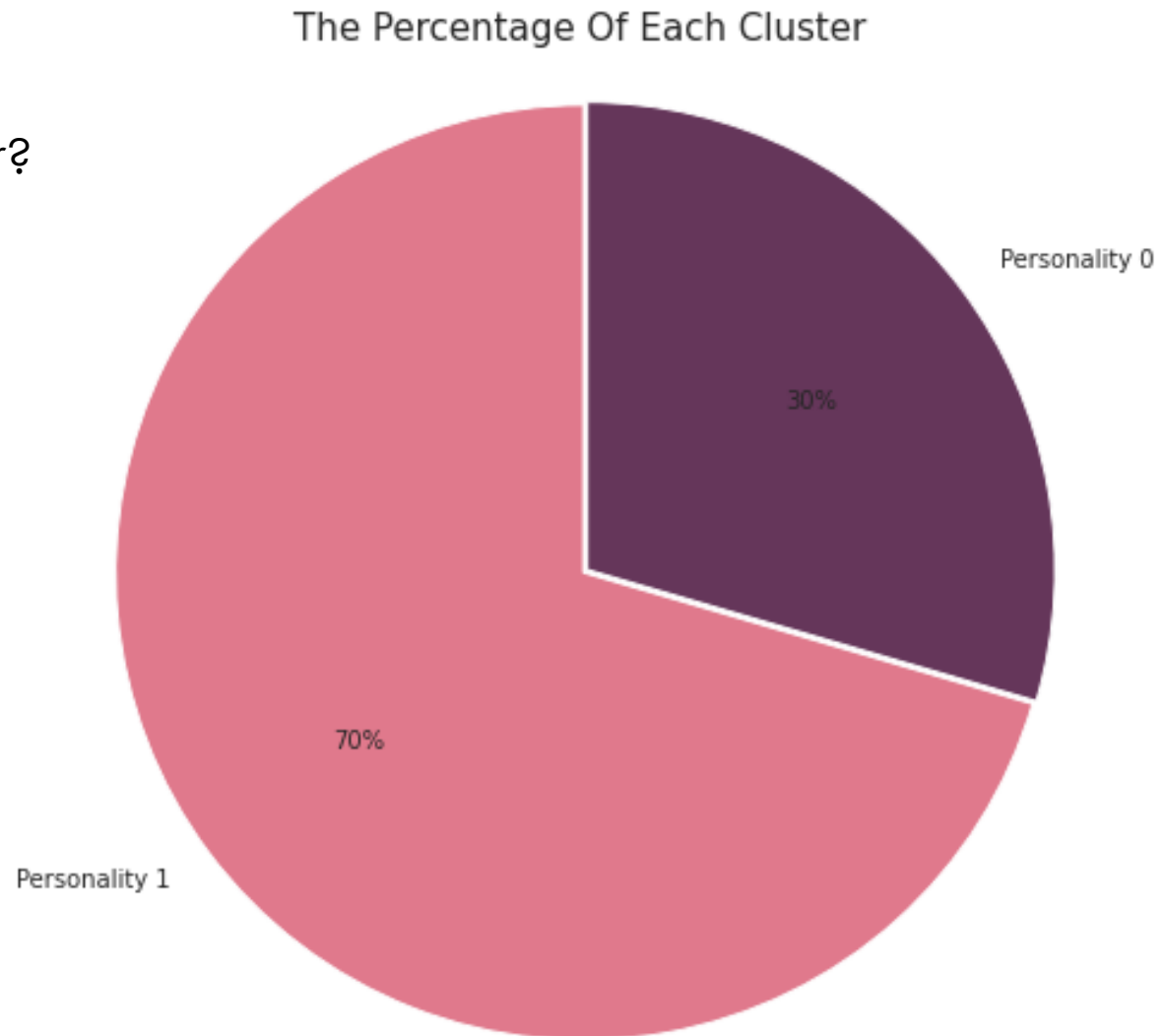
➔ As a result, although the elbow was unsure whether it was 2 or 3, the silhouette arrived and said it was 2.

A large, irregular pink brushstroke shape that serves as a background for the title text. It has a soft, painterly texture with varying shades of pink.

IV/ Evaluating Models

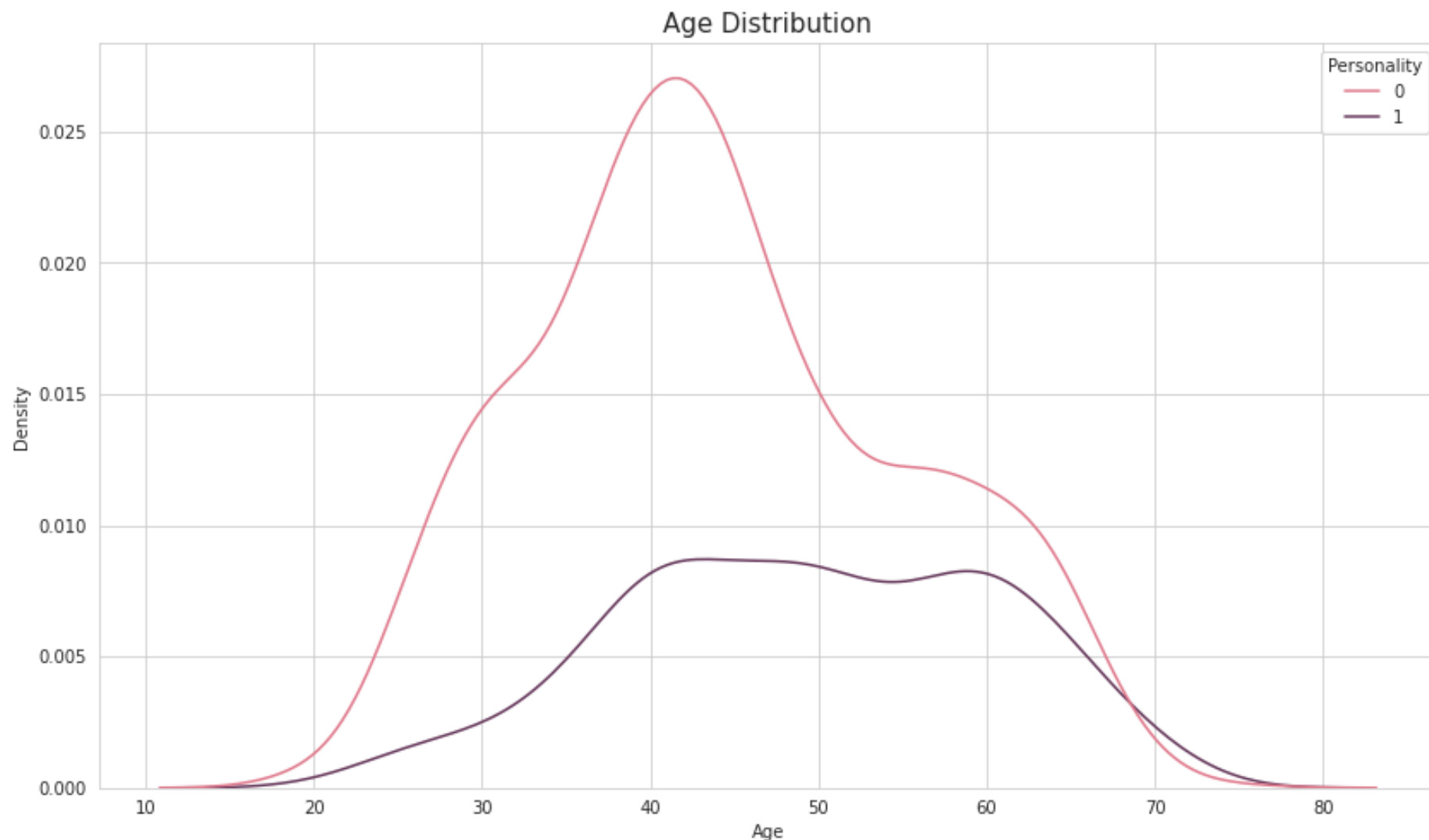
➔ Now, we must examine the patterns in the clusters that have developed and ascertain their nature.

What is the Percentage of each Cluster?



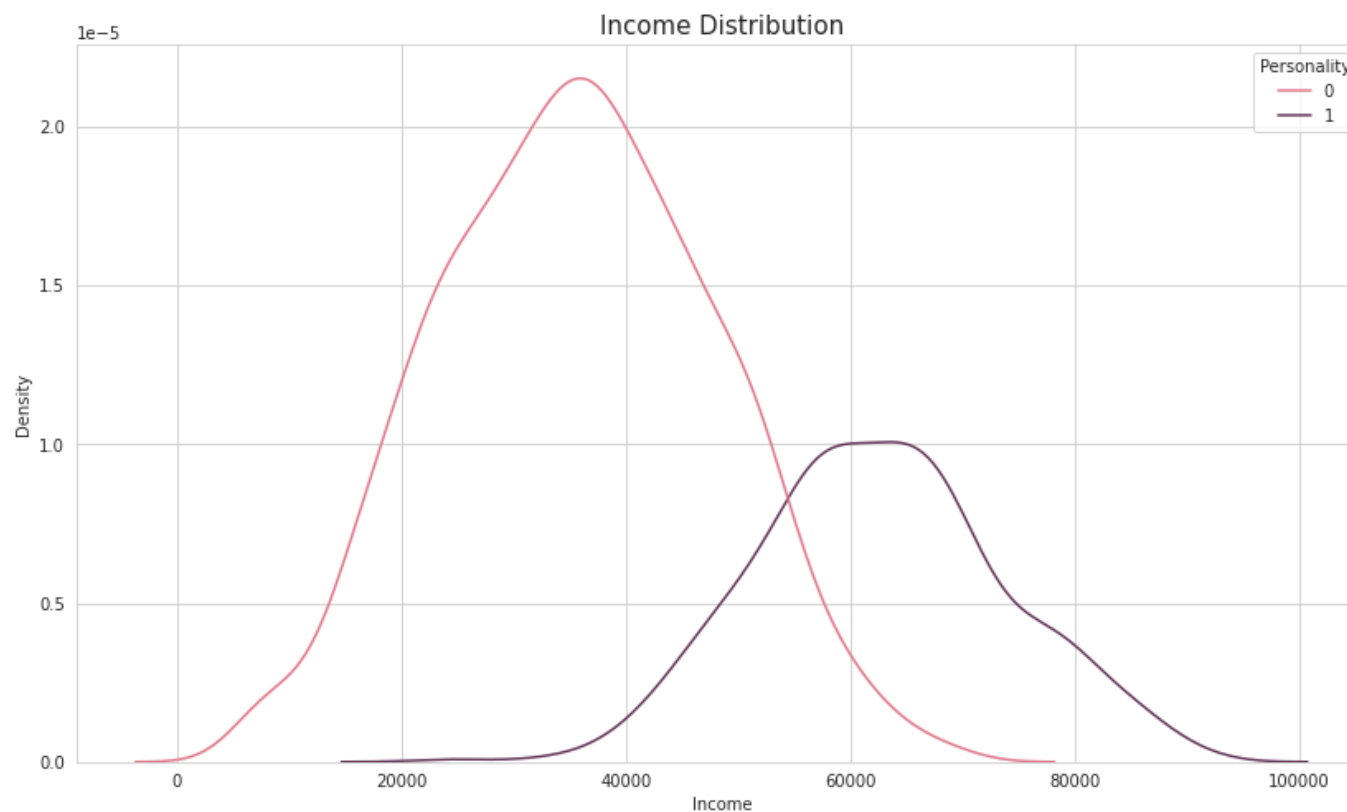
What is the age distribution for each cluster?

- Cluster 0 has a greater age diversity and more youth.
- Those in cluster 1 are primarily older folks.
- proportion of patrons who are over 40 at cluster 0: 0.59
- 0.76 percent of cluster 1's customers are over 40.



What is the income distribution of each cluster?

- Cluster 0 earns greater money with an average of 35,615 dollars.
- Cluster 1 makes higher money on average, at 62,760 dollars.

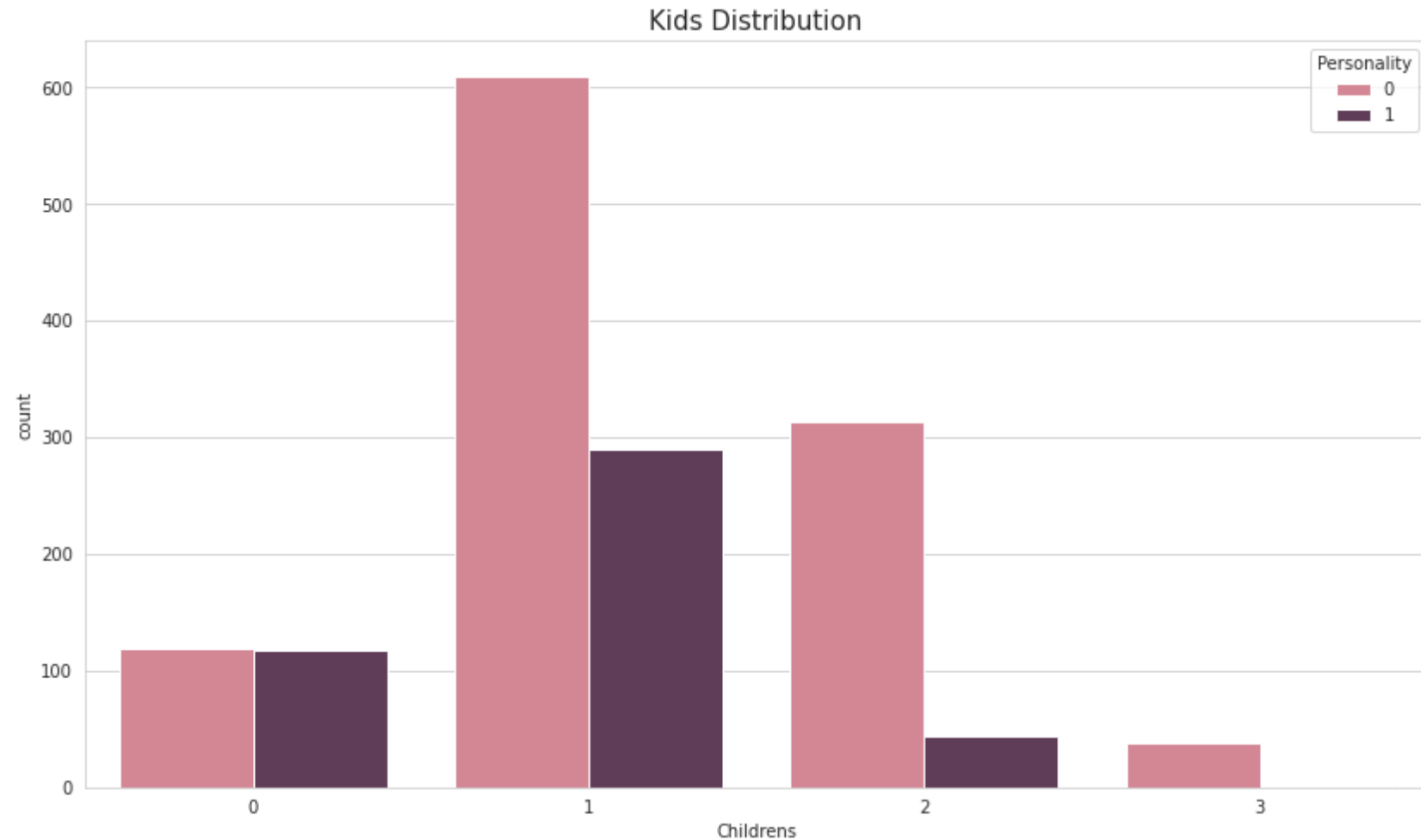


	count	mean	std	min	25%	50%	75%	max
Personality								
0	1081.0	35552.766883	12156.203429	5305.0	26759.0	35684.0	44322.0	69139.0
1	453.0	62674.678808	11073.939261	24401.0	55357.0	62220.0	69674.0	90933.0

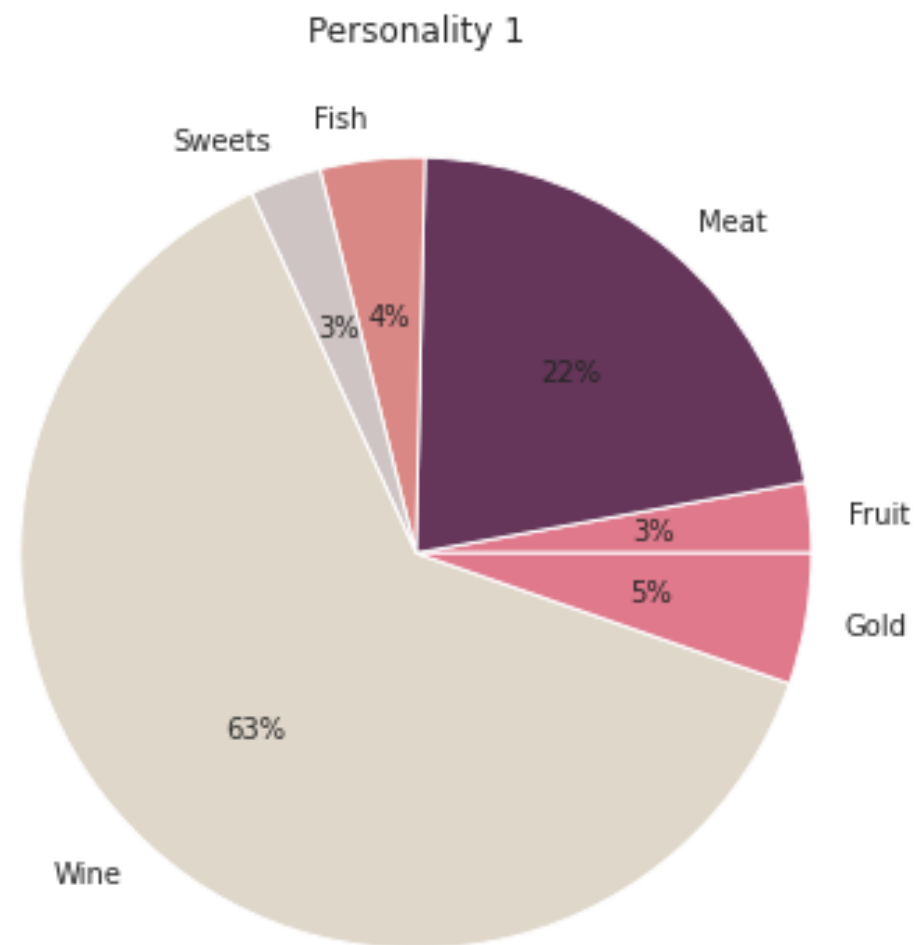
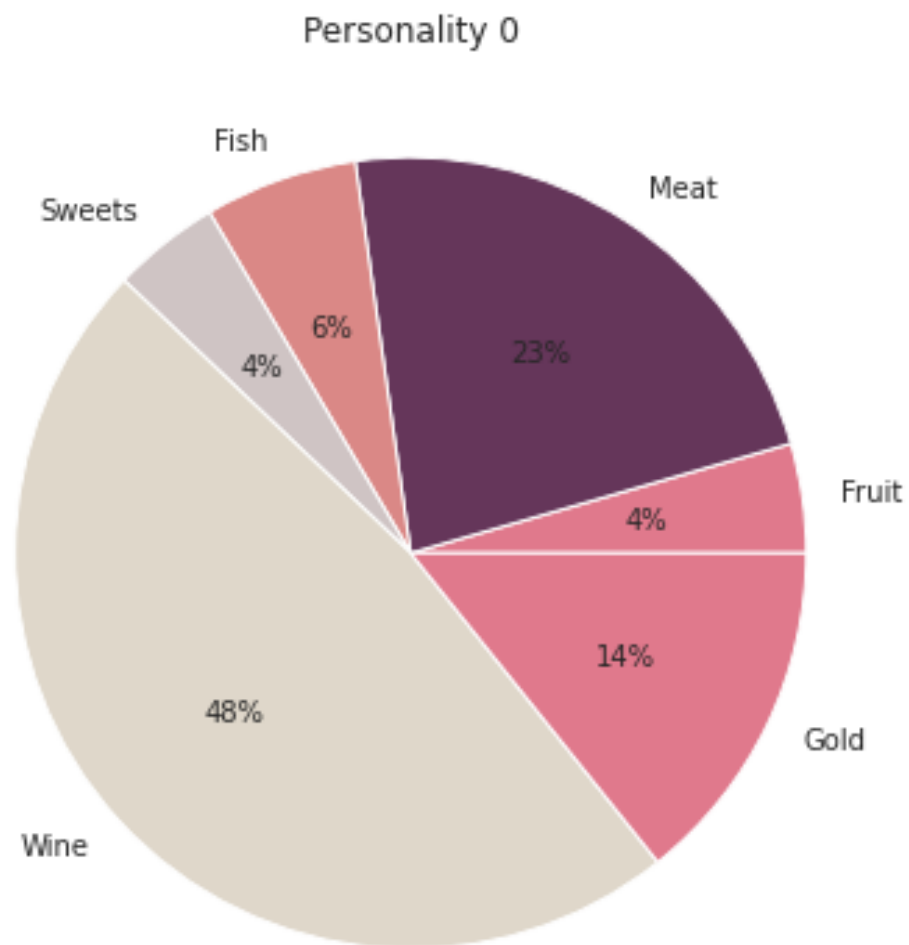
What is the kids number of each cluster?

- At cluster 0, the likelihood of having children is 0.63 percent.
- At cluster 1, the likelihood of having children is 0.22 percent.

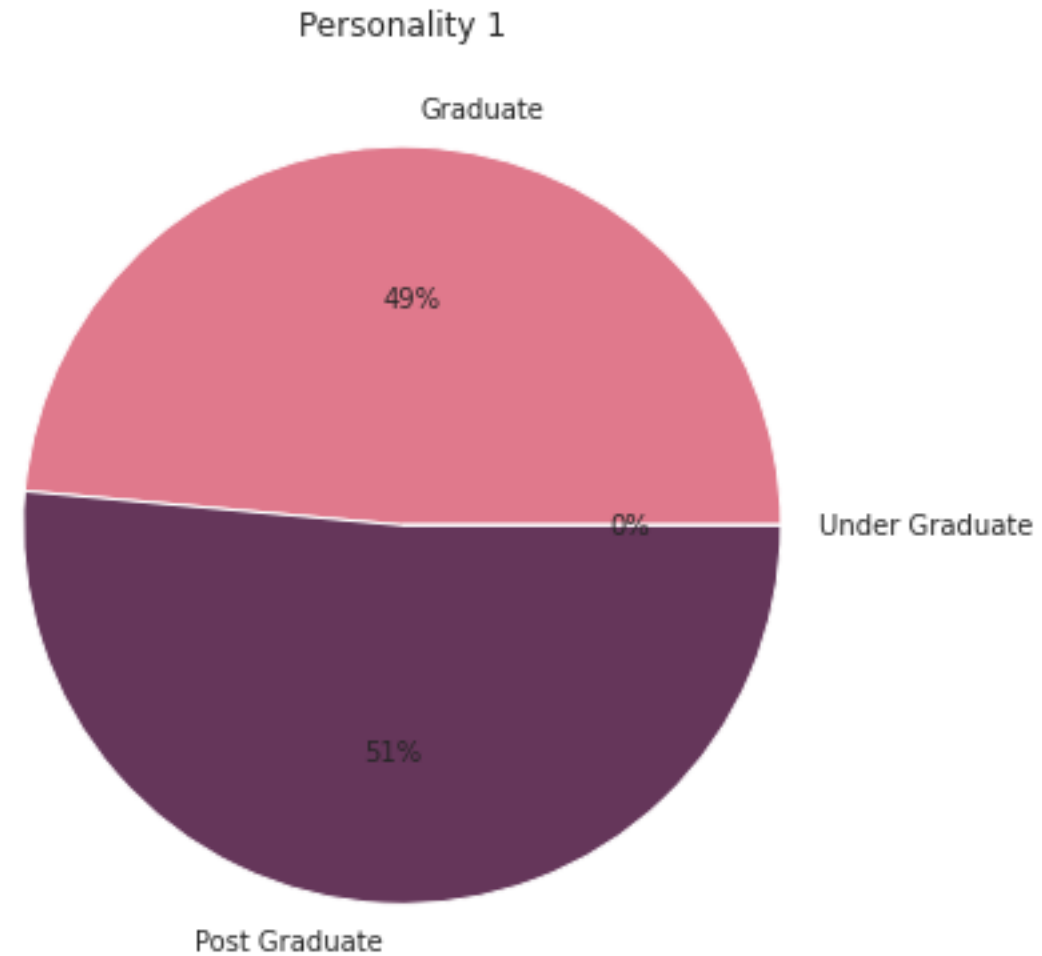
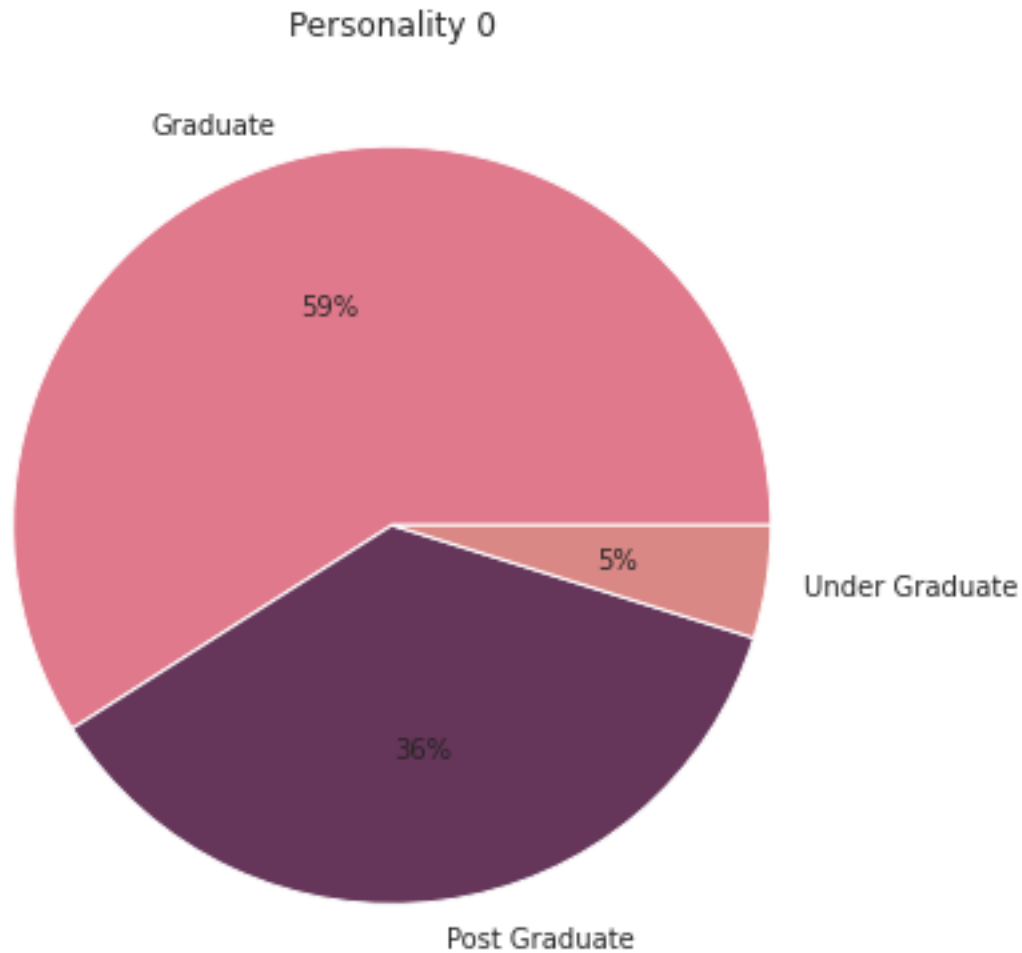
➔ Cluster 0 often has more children.



What is the main product of each cluster?



What is the education of each cluster?



➔ Those in cluster0 are more educated, with 51 percent holding master's or doctoral degrees, compared to only 36 percent in cluster1.

V/ Summary

WE HAVE TWO CLUSTERS

FIRST CLUSTER

- represent 70%
- more income with mean of 62,760\$
- probability of having kids is 23%
- 52% are master and PHD
- relative older with 76% older than 40
- more spending

SECOND CLUSTER

- represent 30%
- less income with mean of 35,615\$
- probability of having kids is 63%
- 36% are master and PHD
- wide range with 59% older than 40
- less spending

VI/ Conclusion



Conclusion

Consumer personality analysis assists a corporation in tailoring its offering to its target market from various customer categories. For instance, a firm may assess which customer group is most likely to purchase the product and then promote the product only to that specific segment rather than investing money to market a new product to every consumer in the company's database.



Thank you

Lamiss Gargouri

Lamiss.gargouri@bahcesehir.edu.tr

