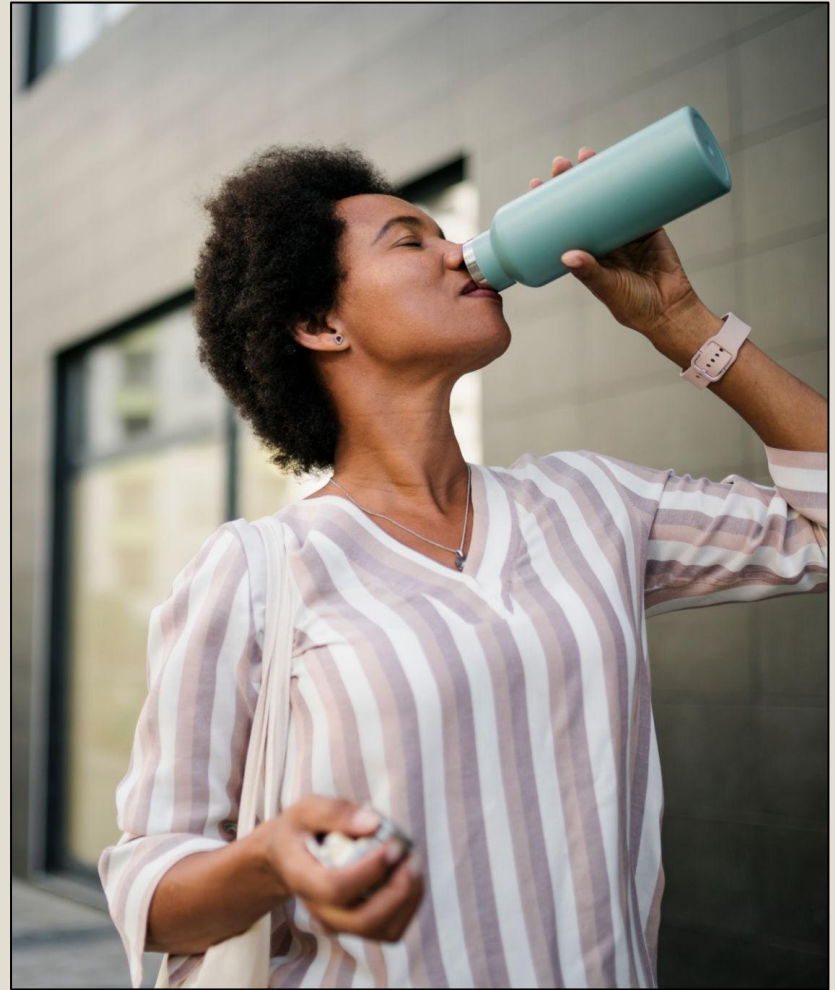# Marketing Analysis

Customer Segmentation

Paula BD.

Intro

Data downloaded from [Kaggle](#) – Marketing Analytics by Jack Daoud (2205 Rows and 38 Columns)

Data contains information on:
1) Customer profiles
2) Product preferences
3) Campaign successes/failures
4) Channel performance

For this analysis I will focus on 1) and 2) in order to create customer segmentation and show their preferences.

Created new columns for the analysis

MtnTotal (Total spent)	*Marital_status*	*Education_status*
- MntWines	- marital_Divorced	- education_2n_Cycle
- MntFruits	- marital_Married	- education_Basic
- MntMeatProducts	- marital_Single	- education_Graduation
- MntFishProducts	- marital_Together	- education_Master
- MntSweetProducts	- marital_Widow	- education_PhD
- MntGoldProds		

Children_Binary; Kidhome and Teenhome > 0	Marital_Status_Binary; Together = 1, Single = 0

Income correlations
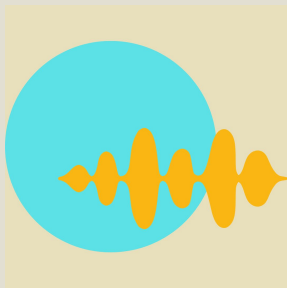
|	
Income	1.000000
MntTotal	0.819310
Age	0.216372
education_PhD	0.083590
marital_Widow	0.047589
Teenhome	0.044418
education_Maste	0.025343
education_Graduation	0.022483
marital_Divorced	0.015109
marital_Together	-0.004510
marital_Married	-0.007835
marital_Single	-0.018309
education_2n Cycle	-0.062488
education_Basic	-0.241868
Kidhome	-0.540774

Name: Income, dtype: float64

K-means Clustering for consumer segmentation

1) Standardization; Min Max Scaler
2) Perform Dimensionality Reduction with PCA
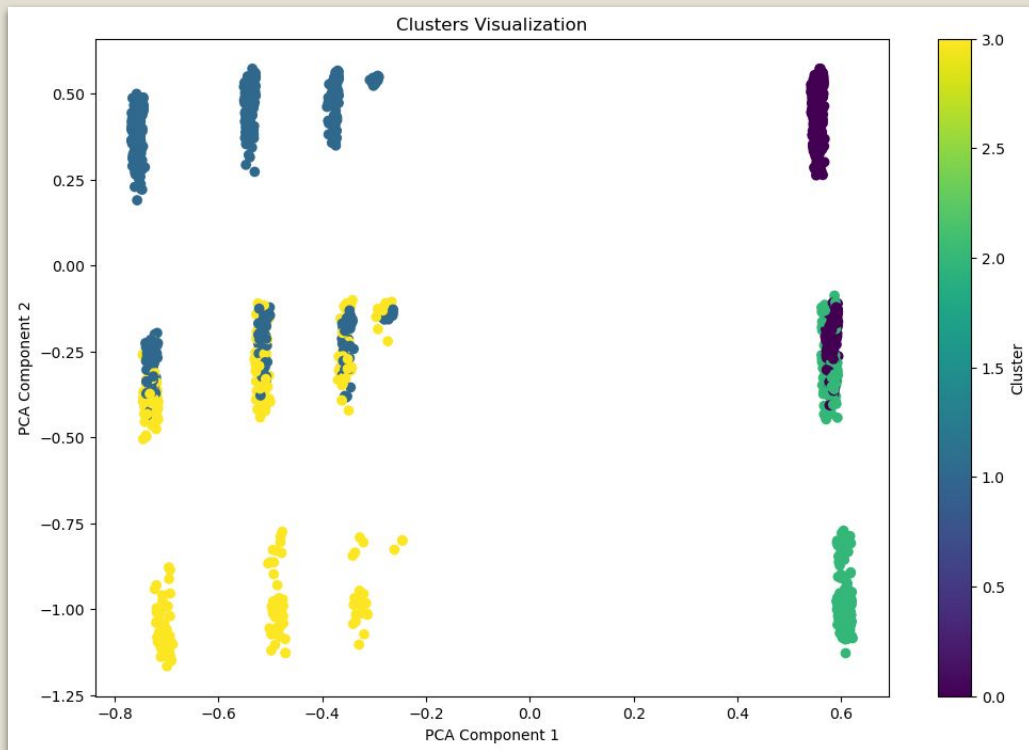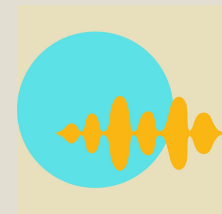3) Fit to K-means
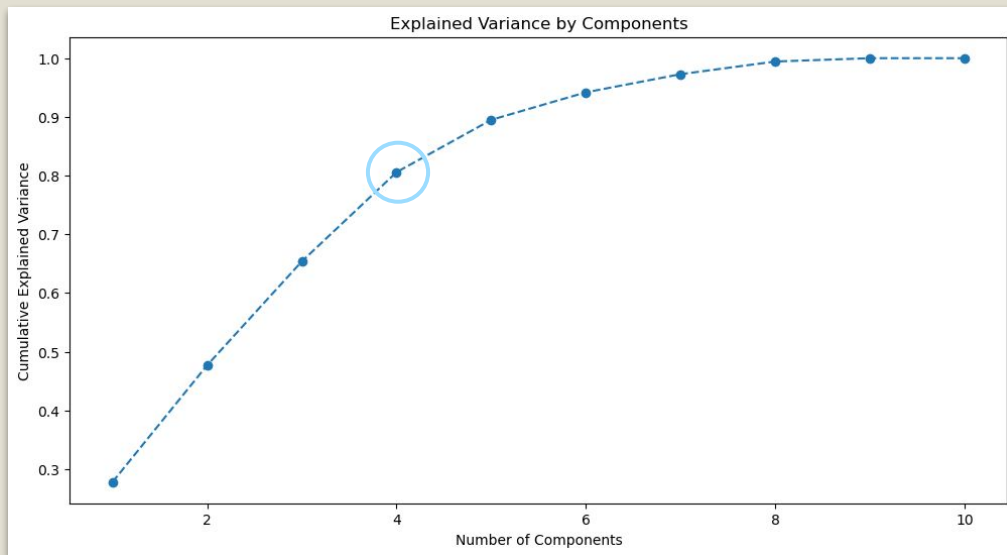
In order to preserve 80% of our data set,
4 components are chosen.

Running Inertia and Silhouette Score on our PCA results to determine the number of clusters for the model
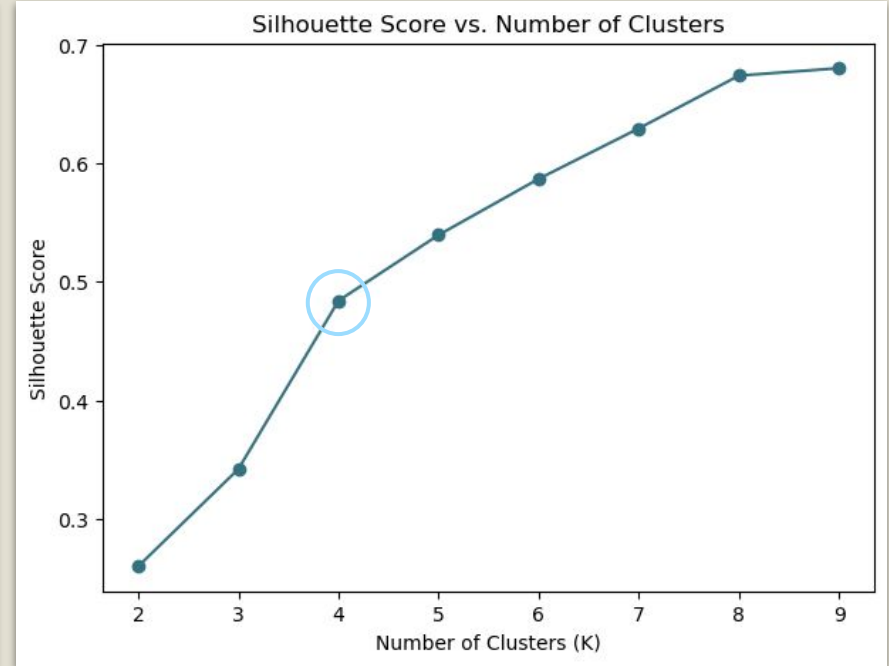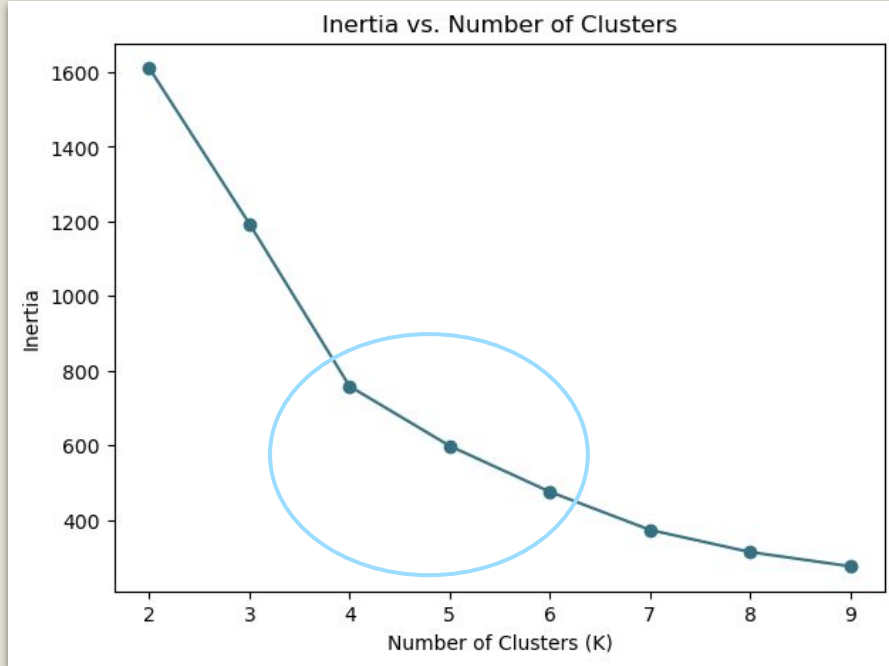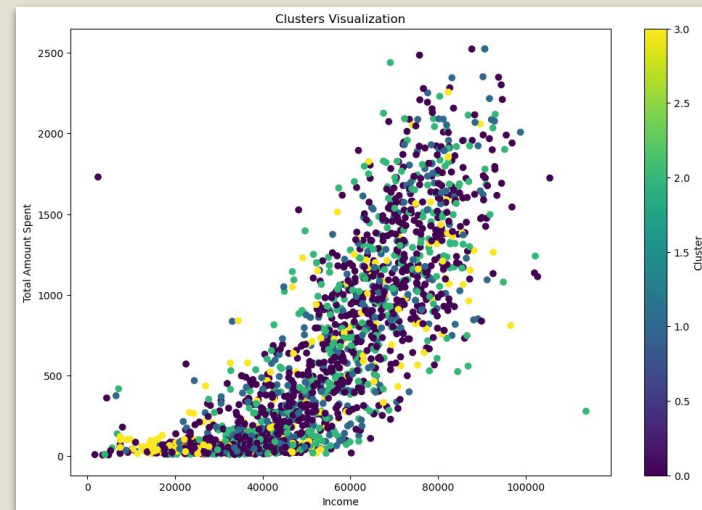

4 segmentation groups are recommended

PCA clustering visualization!

Clusters with only scaled data- not PCA

Exploring our customer segmentation

Cluster 0
- 728 Customers
- 25.23 Percentage spent
- At least 1 child in the household
- 65% in a relationship and 35% single

Cluster 1
- 726 Customers
- 23.04 Percentage spent
- At least 1 child in the household
- 66% in a relationship and 34% single

Cluster 2
- 287 Customers
- 26. 33 Percentage spent
- No children in the household
- 57% in a relationship and 43% single

Cluster 3
- 280 Customers
- 25.40 Percentage spent
- No children in the household
- 60% in a relationship and 40% single

Mean values per cluster - Sorted by Income

Cluster	Income	Age	MntTotal	Children_Binary	Marital_Status_Binary
2	66421.73	51.25	1126.96	0.0	0.57
3	64869.02	51.69	1113.91	0.0	0.60
0	46524.70	50.21	425.71	1.0	0.65
1	45955.37	51.74	389.71	1.0	0.66

- Cluster 2 is a premium cluster with higher spending, followed by cluster 3.
- Cluster 3 with 280 customers brings a similar revenue percentage as cluster 0 with 728 customers.

Clusters 2 and 3 are pretty similar and cluster 0 and 1 as well.
What separates them as clusters?

Mean values per cluster - Sorted by Income

	Income	Basic	Graduation	Master	PhD
Cluster					
2	66421.76	0.00	1.0	0.00	0.00
3	64869.029	0.05	0.0	0.33	0.41
0	46524.77	0.00	1.0	0.00	0.00
1	45955.37	0.04	0.0	0.33	0.44

As we analyse the variable of education we can see a clear distinction between the groups.

Now that we know premium cluster is #2 and cluster #0 includes at least one child
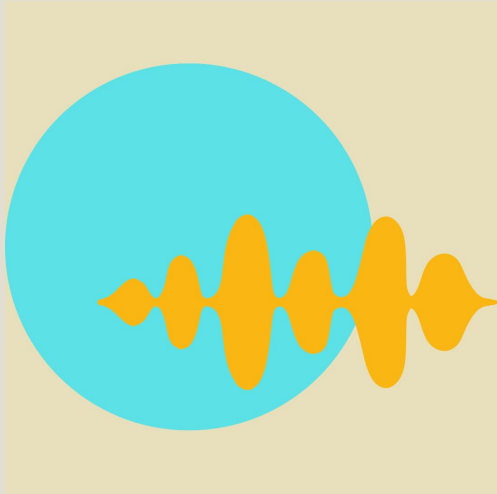in the household let's see spending preferences.
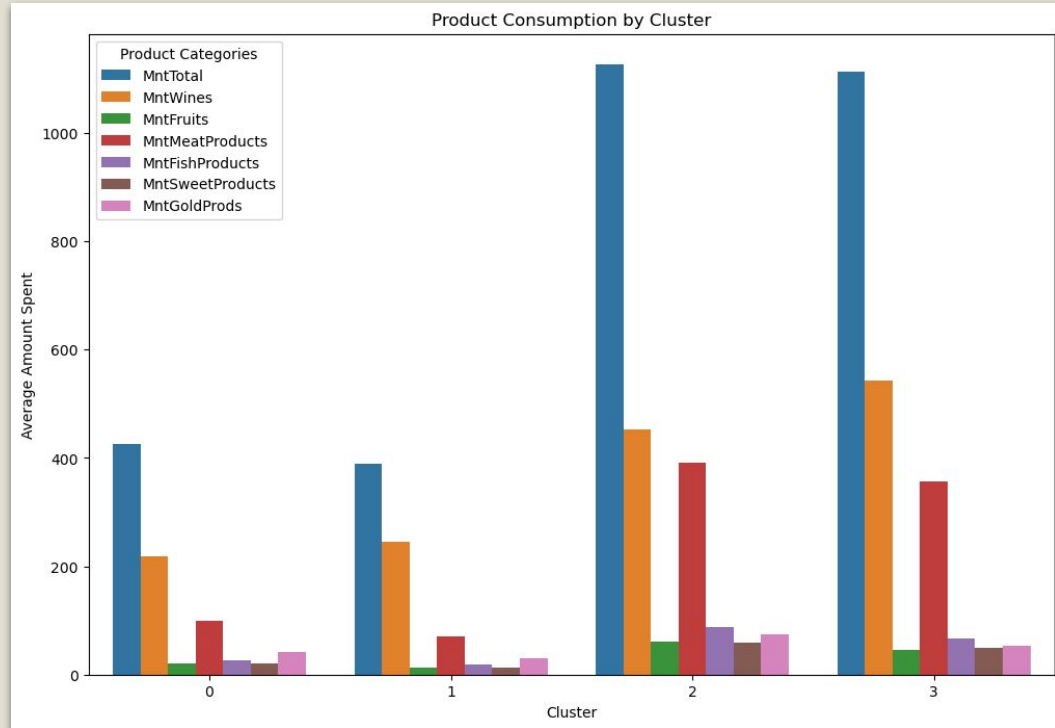
Conclusion

- Groups 2 and 3 are the highest spenders. The company should focus on the products that are preferred by these groups and try to cross-promote other products.

- We can also see that two categories of products are the top sellers among the four groups. The recommendation is ensure a steady supply of both categories to prevent stockouts, and negotiate better deals with suppliers due to high demand.

  *This analysis did not include promotions on products.

Thank you!