Market segmentation Project

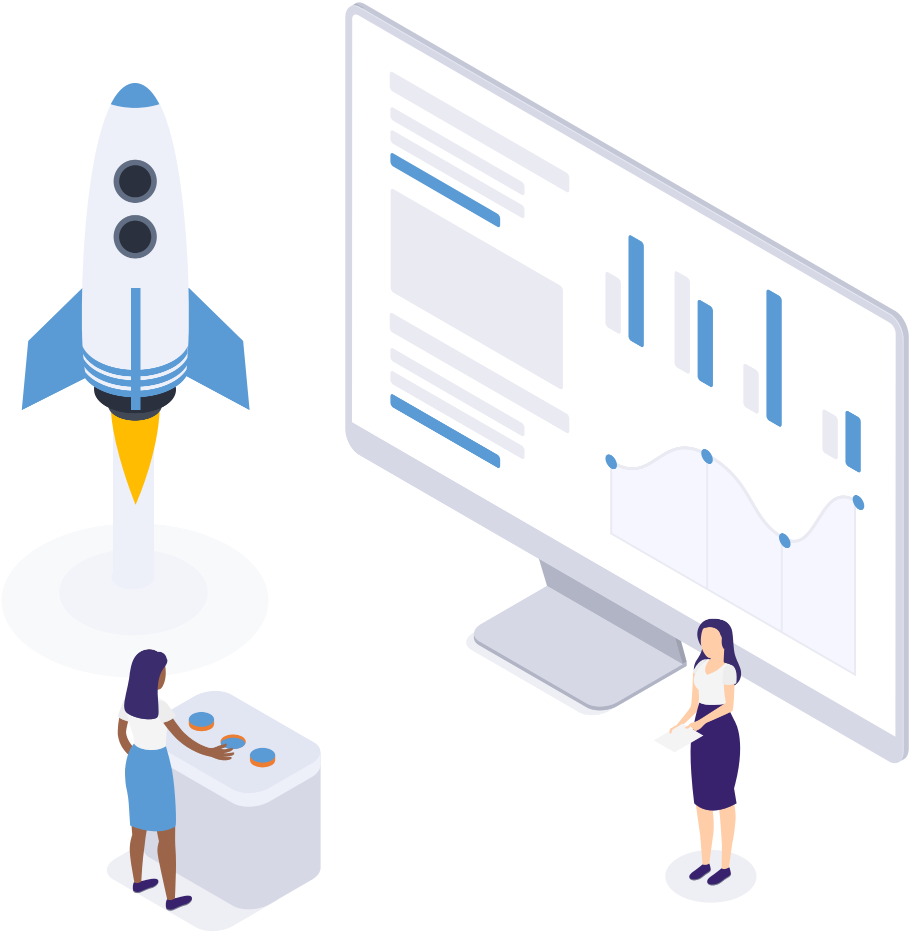
By Ibrahim Ahmad

Market segmentation

- The goal of this project is to segment a dataset of customers to enhance the effectiveness of future marketing campaigns.
- Each customer group shares specific demographic, psychographic and behavioral properties that make them suitable for a specific method of marketing
- We will use statistical techniques to segment the customers and verify the results on the data

1 Dataset source

What is the customers database that we wish to explore?



Dataset source

- Ifood is a Brazilian online food ordering and food delivery platform, operating in Brazil and Colombia. The company holds over 80% market share of the food delivery sector in Brazil.
- The dataset is a snippet of the total customer base that joined from August 2012 to July 2014 with a count of 2205 customers
- The dataset describes the various demographic details and economic activities of the customers.

Dataset's fields

The dataset contains the following fields:

- Demographic fields: Year of birth, Educational level, Marital status, Income, The number of children in the household and The number of teenagers in the household.
- Customer relationship fields: Date of customer enrollment with the company, the number of days since customer's last purchase and whether a customer filed a complaint within the last 2 years.
- Promotional campaigns engagement fields: number of purchases made with a discount, campaigns acceptance 1 through 6.

Dataset’s fields

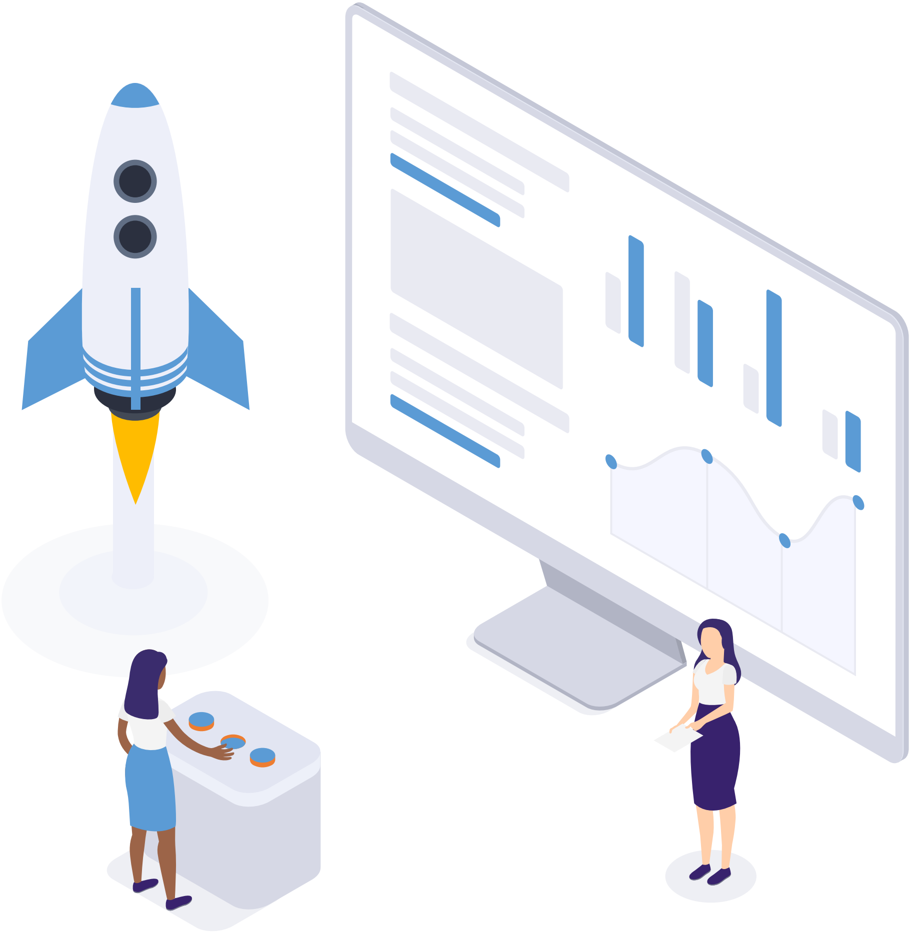
The dataset contains the following fields:

- Customer purchasing habits fields: The amount of spending on wine, fruits, meat products, fish products, sweet products and gold products.
- In addition to the customer method of purchasing: number of web purchases, catalog purchases, store purchases, purchases with deals and the number of web visits.

2

Dataset EDA and cleaning

What is the general distribution of the sample and what are the cleaning steps neccesary for analysis?



Dataset cleaning /future engineering

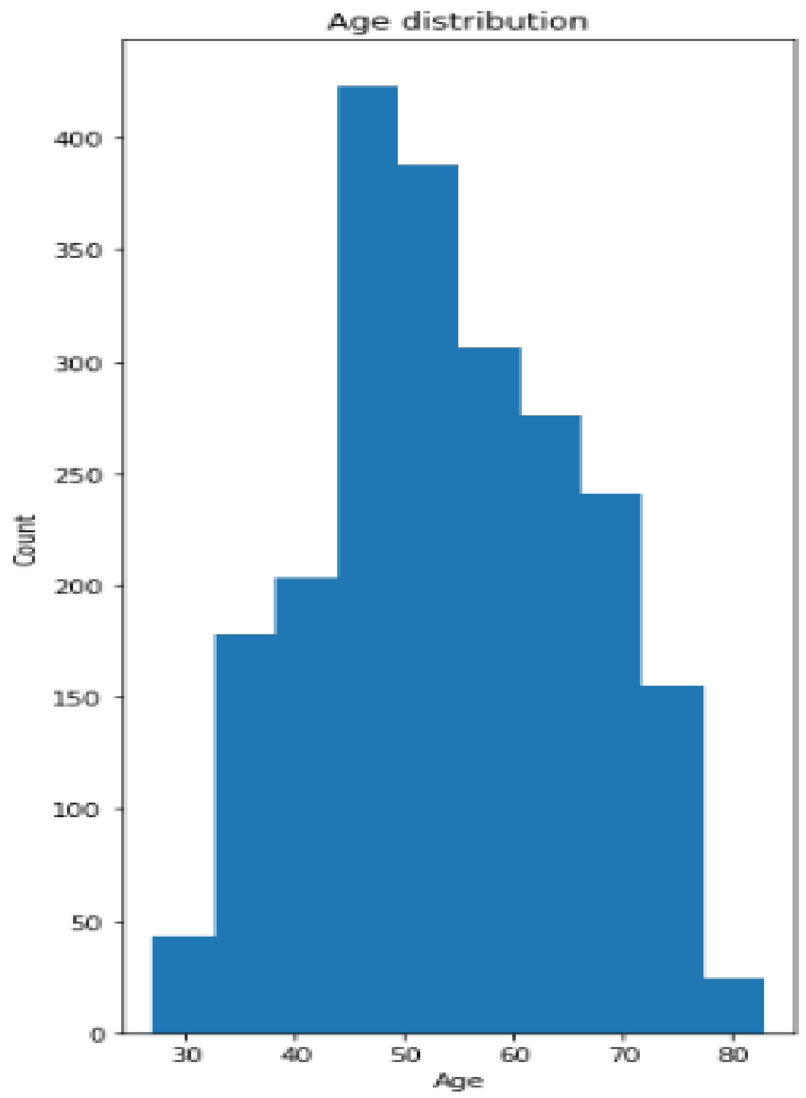
- Initially we find that we are missing 24 records in the Income fields, we impute these records with the respective average of each educational level.
- Exploring the Marital status field we find values such as "Alone", "Absurd" and "YOLO" these values indicate a single lifestyle, so they will be coded to "Single".
- We create the age field by subtracting the date of birth field from the last observed date of enrollment in the dataset, exploring this field shows some outliers with users age reaching more than 100 years. Those will be excluded from the analysis.

Dataset cleaning /future engineering

- In order to better signify the period of enrollment in the clustering analysis we convert the date of enrollment to days of enrollment.
- We replace the data dichotomous fields from 1/0 to Yes/No.
- Exploring the variables “Z_CostContact” and “'Z_Revenue” we see that they hold a constant value for all records and are useless for analysis, they will be dropped.

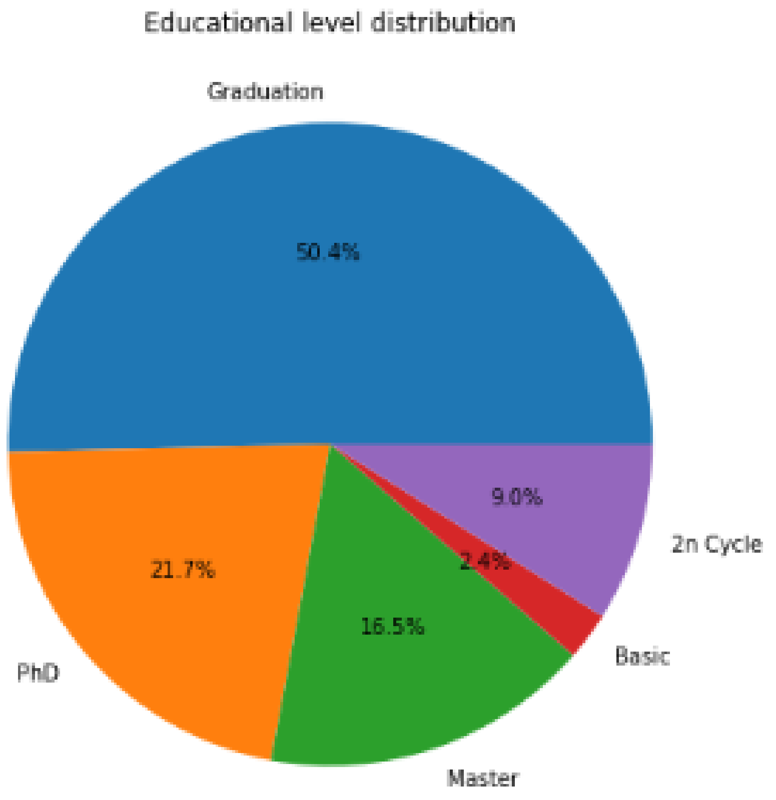
Dataset EDA

- The age distribution seems to be normal with the majority of the customer base within the 50-60 year old age.



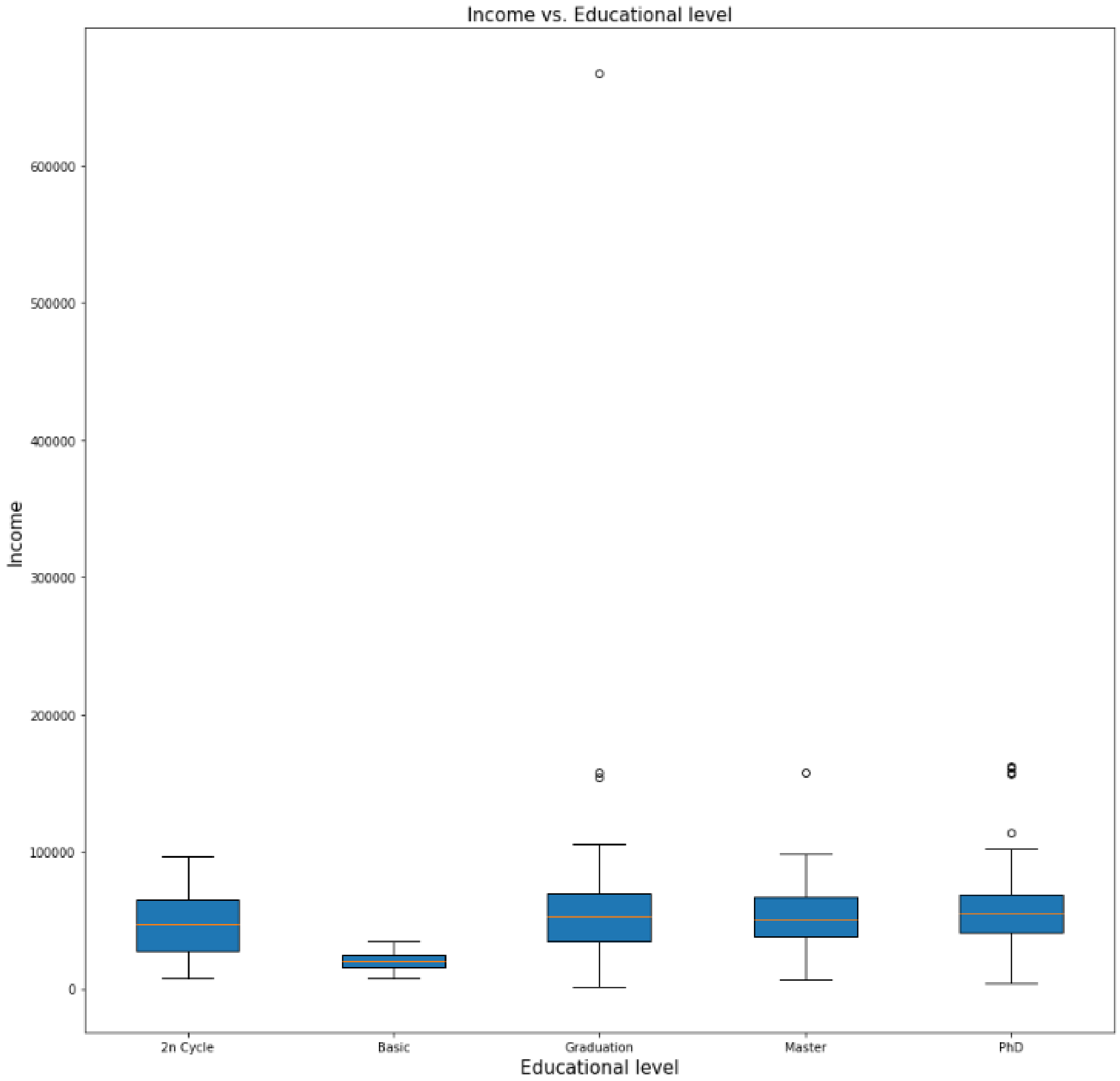
Dataset EDA

- Plotting the educational level and the income level distribution we conclude that the majority of the data set are highly educated.



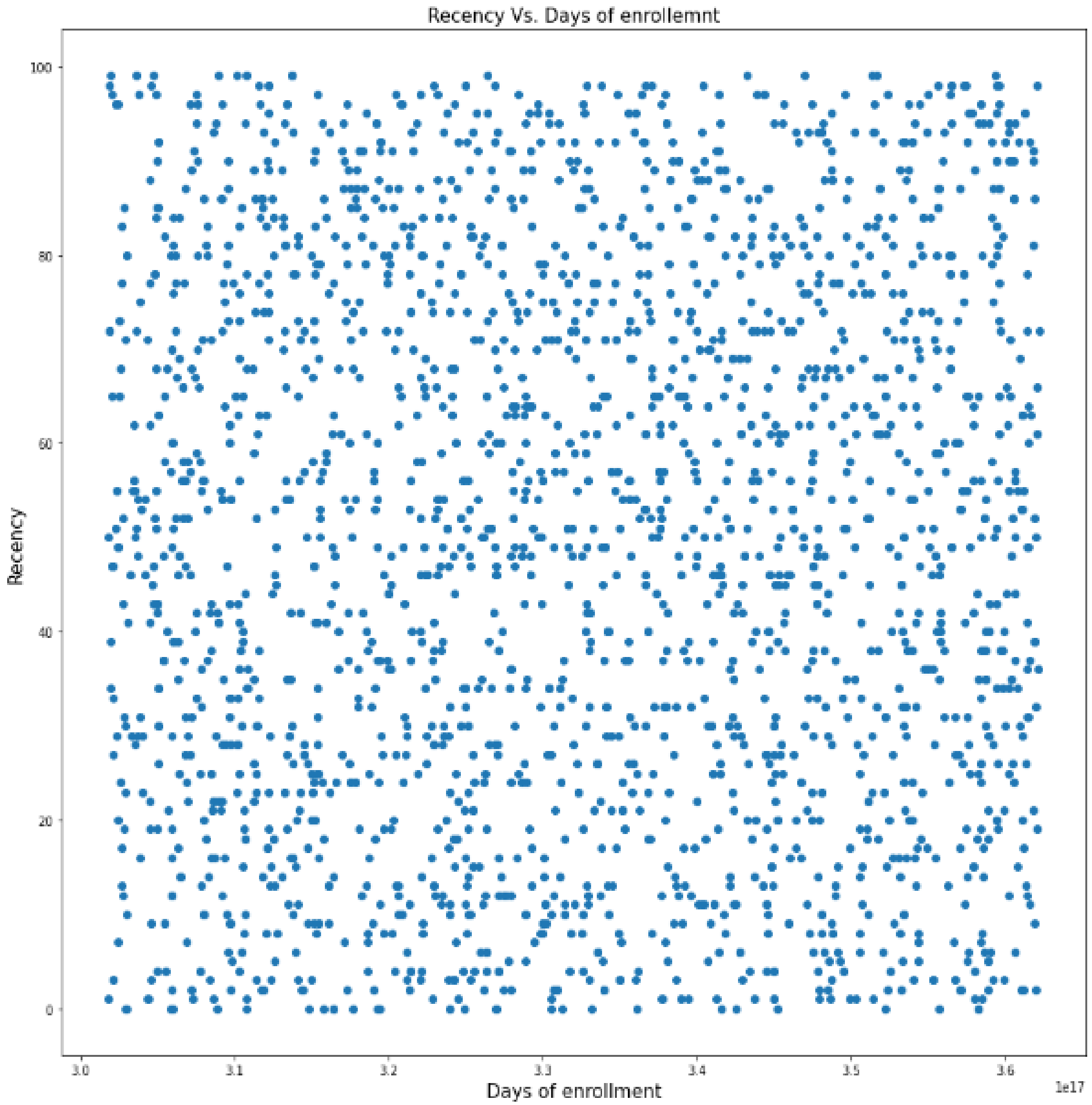
Dataset EDA

- The income level of the higher education users is somewhat similar in terms of average and spread while the lower education users tend to be poorer.
- We can also see a significant outlier in the graduation level category, it will be removed to improve clustering accuracy.



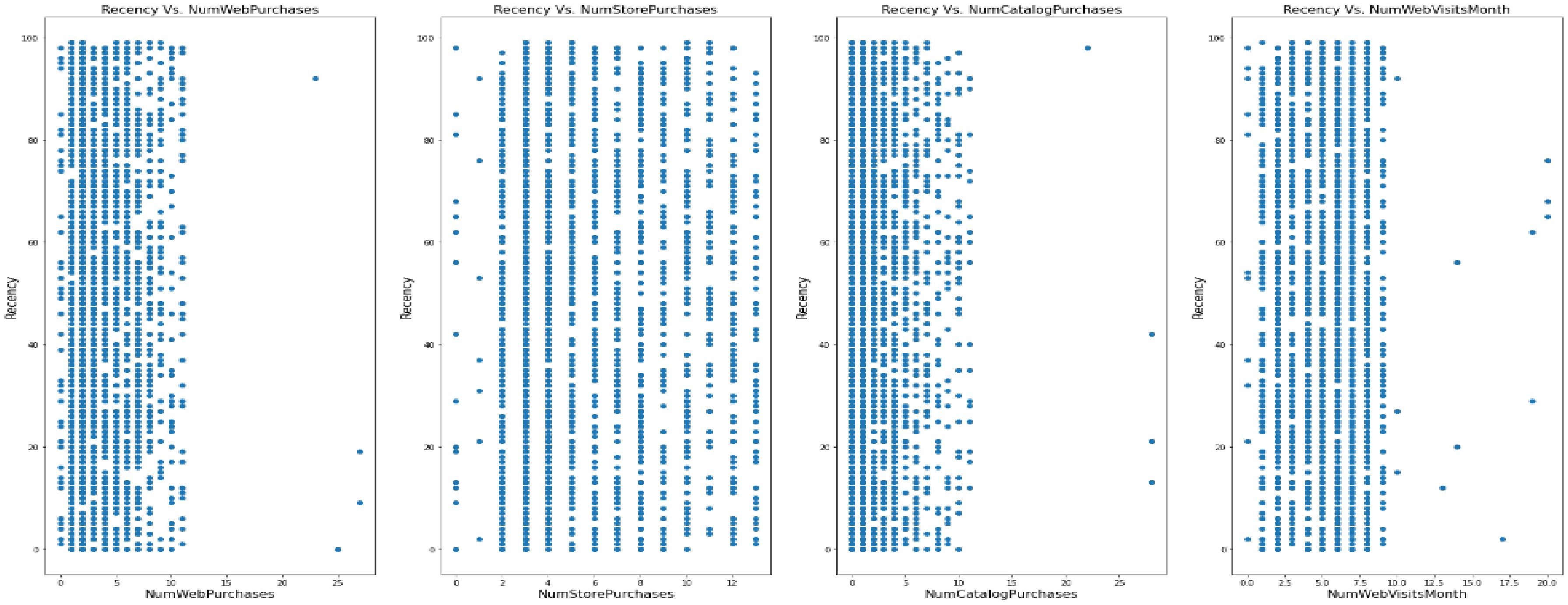
Dataset EDA

- Plotting the numerical variables against each other to find indications of correlations, we start with the number of days since customer's last purchase vs other spending habits:
- It shows very little correlation and isn't of great value to the next analysis steps therefore it will be dropped



Dataset EDA

- Plotting the numerical variables against each other to find indications of correlations, we start with the number of days since customer's last purchase vs other spending habits:
- It shows very little correlation and isn’t of great value to the next analysis steps therefore it will be dropped



Dataset EDA

- Another method of finding indications of correlations is the correlation matrix, it calculates the correlation coefficient between different variables.
- Calculating for the customer purchasing habits fields, we can find good to somewhat strong correlation between the amount of products purchased.
- The same can be carried out for the customer's method of purchasing field.
- This will be of great value for the feature reduction process.

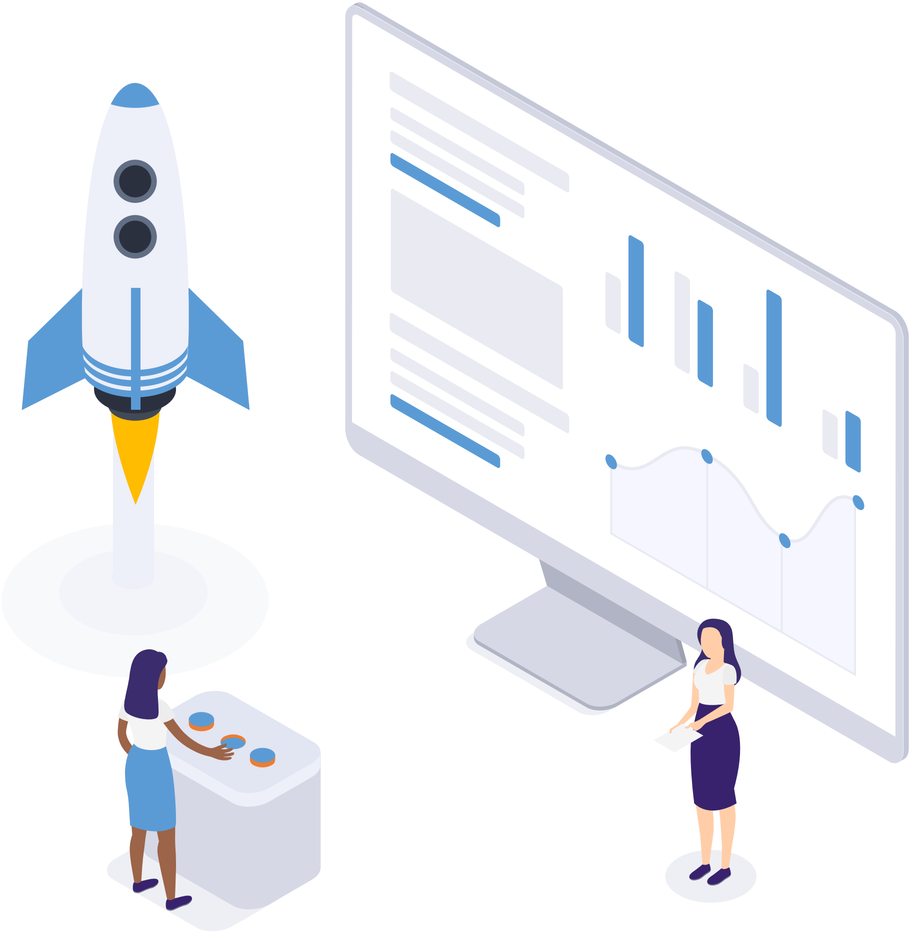
	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
MntWines	1.000000	0.388472	0.561886	0.398956	0.385845	0.386234
MntFruits	0.388472	1.000000	0.542030	0.594415	0.567030	0.389999
MntMeatProducts	0.561886	0.542030	1.000000	0.567813	0.523329	0.348727
MntFishProducts	0.398956	0.594415	0.567813	1.000000	0.579490	0.422018
MntSweetProducts	0.385845	0.567030	0.523329	0.579490	1.000000	0.369084
MntGoldProds	0.386234	0.389999	0.348727	0.422018	0.369084	1.000000

	NumWebPurchases	NumCatalogPurchases	NumStorePurchases
NumWebPurchases	1.000000	0.378049	0.502227
NumCatalogPurchases	0.378049	1.000000	0.518788
NumStorePurchases	0.502227	0.518788	1.000000

3

Clustering preperation

What is are the fields to be used for clustering?



Clustering preperation

- Clustering is the process of using statistical methods to analyze and find the connections that join groups of users within certain variables.
- Such variable can be categorical and numerical.
- The results of the clustering process is the classification of each user into a specific group (Cluster), the interpretation of the elements that create a cluster is up to the analyst.
- Considering the large number of variables it’s inconvenient to use all of them, so some reduction is necessary, variable reduction can be done when there is a correlation between them.
- This allows us to substitute multiple variables with a fewer number while maintain most of the information provided by the variables

Clustering preperation

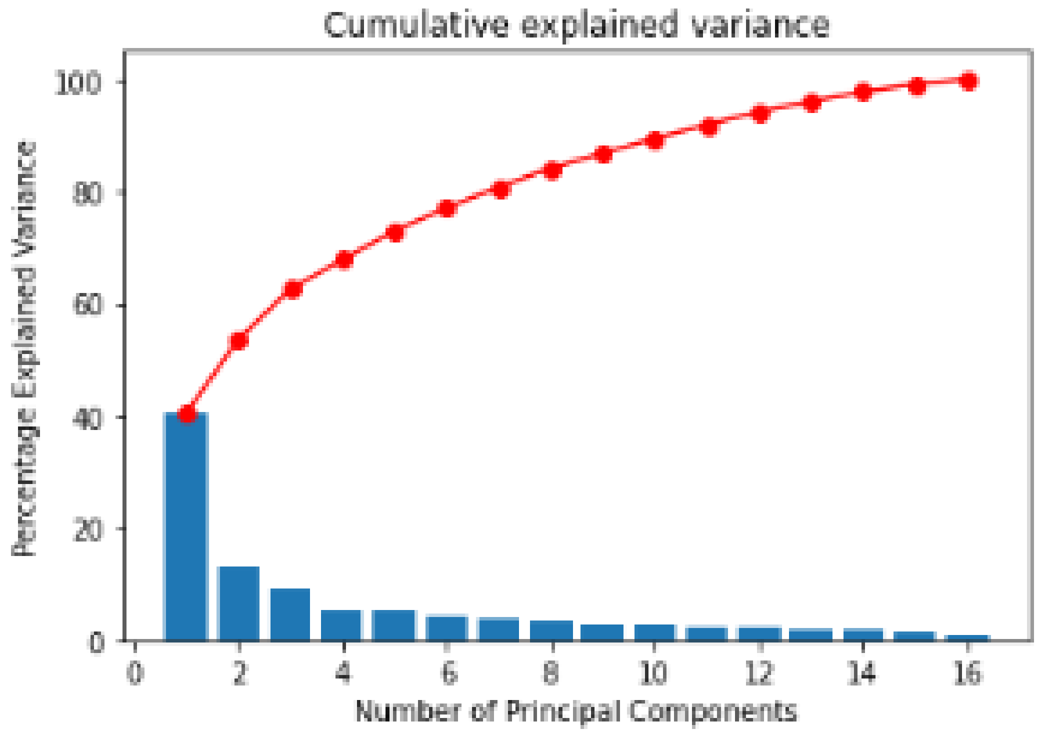
- The method used for variable reduction is called Principle component analysis commonly abbreviated as PCA, the method tries to fit a component through the variables that it’s trying to reduce.
- This component has a loading for the contribution of each variable, it signifies the weight and directionality of the variable’s effect on the component.
- Accordingly each component can be explained with the variables with the highest loading, as a rule of thumb we consider a loading greater that 0.3 to be significant.
- Note that PCA is only usable on numerical variables.

PCA

- The following variables will be used for PCA:
  - Income
  - Age
  - Number of kids at home
  - Number of teenagers at home
  - The amount of spending on wine
  - The amount of spending on fruits
  - The amount of spending on meat products
  - The amount of spending on fish
  - The amount of spending on sweet products
  - The amount of spending on gold products.
  - Number of web purchases
  - Number of catalog purchases
  - Number of store purchases,
  - Number of purchases with deals
  - Number of web visits.

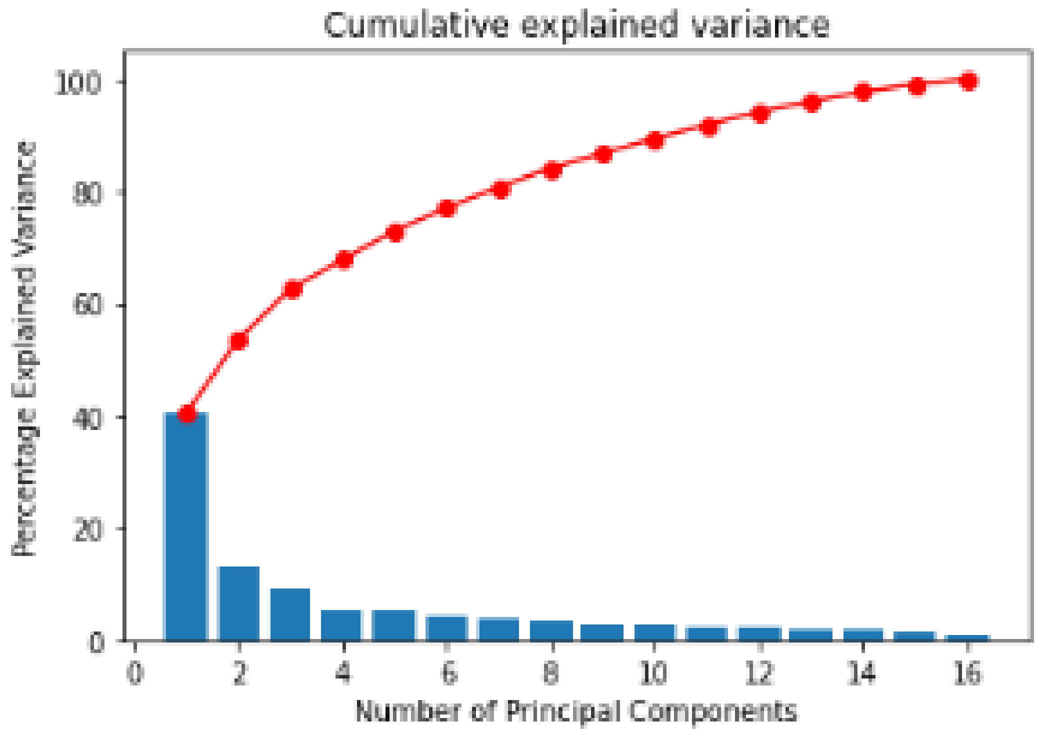
PCA

- Upon fitting the principle components upon the variables after standardization we can explore the cumulative explained variance of the principle components.
- The cumulative explained variance indicates the total coverage of variance in the data that a number of principle components has, we can choose any number we wish
- An 80% coverage is enough as any more components will not have any significant effect
- The following graph displays the cumulative variance of the components:



PCA

- At 7 principle components we see that we covered roughly 80% of the cumulative variance.
- As such we will refit the model with only 7 components.
- To start analyzing we will display the loading of each variable on each component only if it’s greater than 0.3



PCA

- The following table displays the loading for each component and will be used for analysis:
- PC1 has a matching loading between the Income, the amount of wines and the amount of meat products while the purchase method used the most is catalog purchasing, it expresses the **level of wealth of the individual**.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC1 has a strong loading in the number of teens in home and the number of deals purchases, with good loading in web purchases, this indicates a pattern of **thriftiness** with good knowledge in technology.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC2 has a strong loading in the number of teens in home and the number of deals purchases, with good loading in web purchases, this indicates a pattern of **thriftiness** with good knowledge in technology.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC3 has a strong negative loading with the days of enrollment and the number of web visits while having a strong loading with age, this indicates **older users that have recently joined the platform.**

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC4 has a strong loading with the number of kids at home and the number of deals purchases similar to PC2 except with kids, this is further verified by the strong negative correlation with the age and the days of enrollment **(essentially young new users).**

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC5 negative loading for wine spending while a positive loading for fruits and fish spending along with a positive loading for age, it describes **general spending habits** of older users.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC6 has a very strong loading for the amount of spending on gold and very small loading for other variables, as such this component offers little correlation between variables but describes **gold spending.**

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

- The following table displays the loading for each component and will be used for analysis:
- PC6 has a strong positive loading for the number of kids at home while having a negative loading of the number of teens at home along with the age, this describes the **familial age of the users.**

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Income	0.338414	NaN	NaN	NaN	NaN	NaN	NaN
Kidhome	NaN	NaN	NaN	0.405040	NaN	-0.309184	0.499309
Teenhome	NaN	0.467506	-0.389251	NaN	NaN	NaN	-0.371849
MntWines	0.301597	NaN	NaN	NaN	0.395020	NaN	NaN
MntFruits	NaN	NaN	NaN	NaN	-0.306519	NaN	NaN
MntMeatProducts	0.320413	NaN	NaN	NaN	NaN	NaN	NaN
MntFishProducts	NaN	NaN	NaN	NaN	-0.339198	NaN	NaN
MntSweetProducts	NaN	NaN	NaN	NaN	NaN	NaN	NaN
MntGoldProds	NaN	NaN	NaN	NaN	NaN	0.814859	NaN
NumDealsPurchases	NaN	0.523581	NaN	0.450747	NaN	NaN	NaN
NumWebPurchases	NaN	0.380717	NaN	NaN	NaN	NaN	NaN
NumCatalogPurchases	0.335380	NaN	NaN	NaN	NaN	NaN	NaN
NumStorePurchases	0.300950	NaN	NaN	NaN	NaN	NaN	NaN
NumWebVisitsMonth	NaN	NaN	0.336899	NaN	NaN	NaN	NaN
days_of_enrollment	NaN	NaN	0.486738	-0.568916	NaN	NaN	NaN
age	NaN	NaN	-0.513981	NaN	-0.490482	NaN	0.437753

PCA

After examining the principle components we can conclude the main direction and coverage of each component as following:

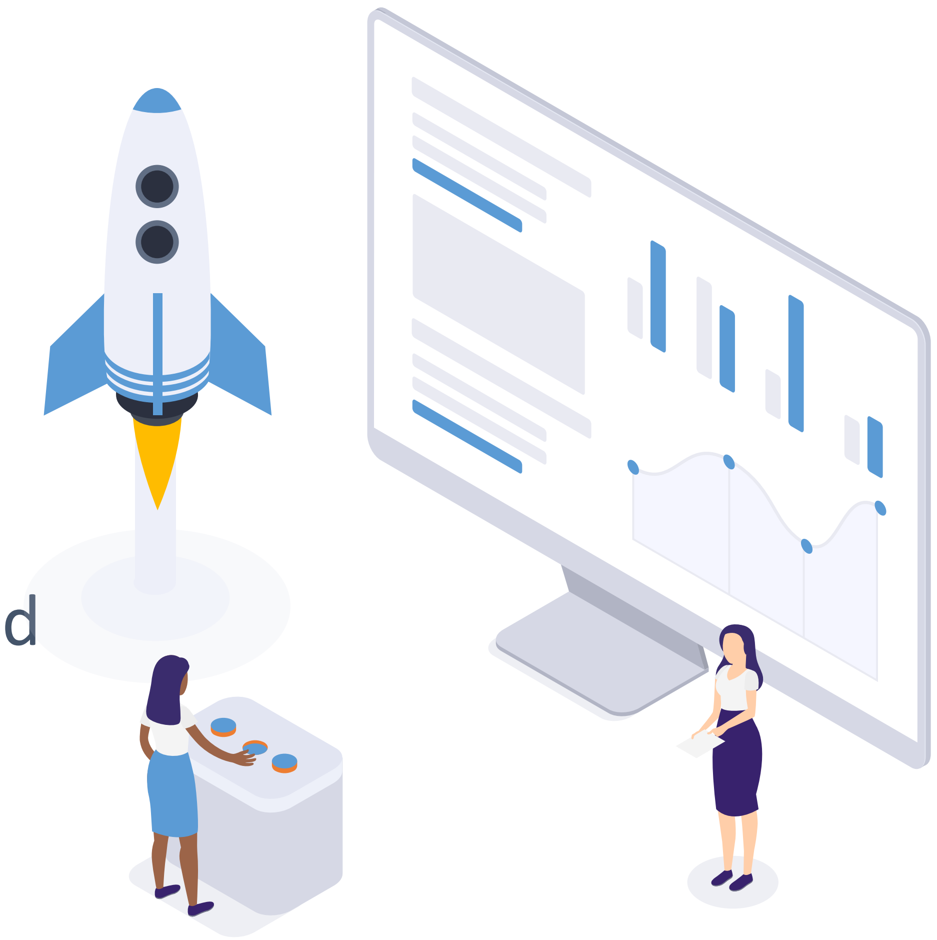
- PC1: Wealthiness
- PC2: Thriftiness
- PC3: Recently joined users habits
- PC4: Young users habits
- PC5: General purchasing habits
- PC6: Gold purchasing behavior
- PC7: Familial age of users

It’s worth noting that PCA is a dimensionality reduction technique and it’s main purpose is to establish elements that represent the correlation between the variables

4

Clustering

What is the clustering technique used and what are the insights extracted?

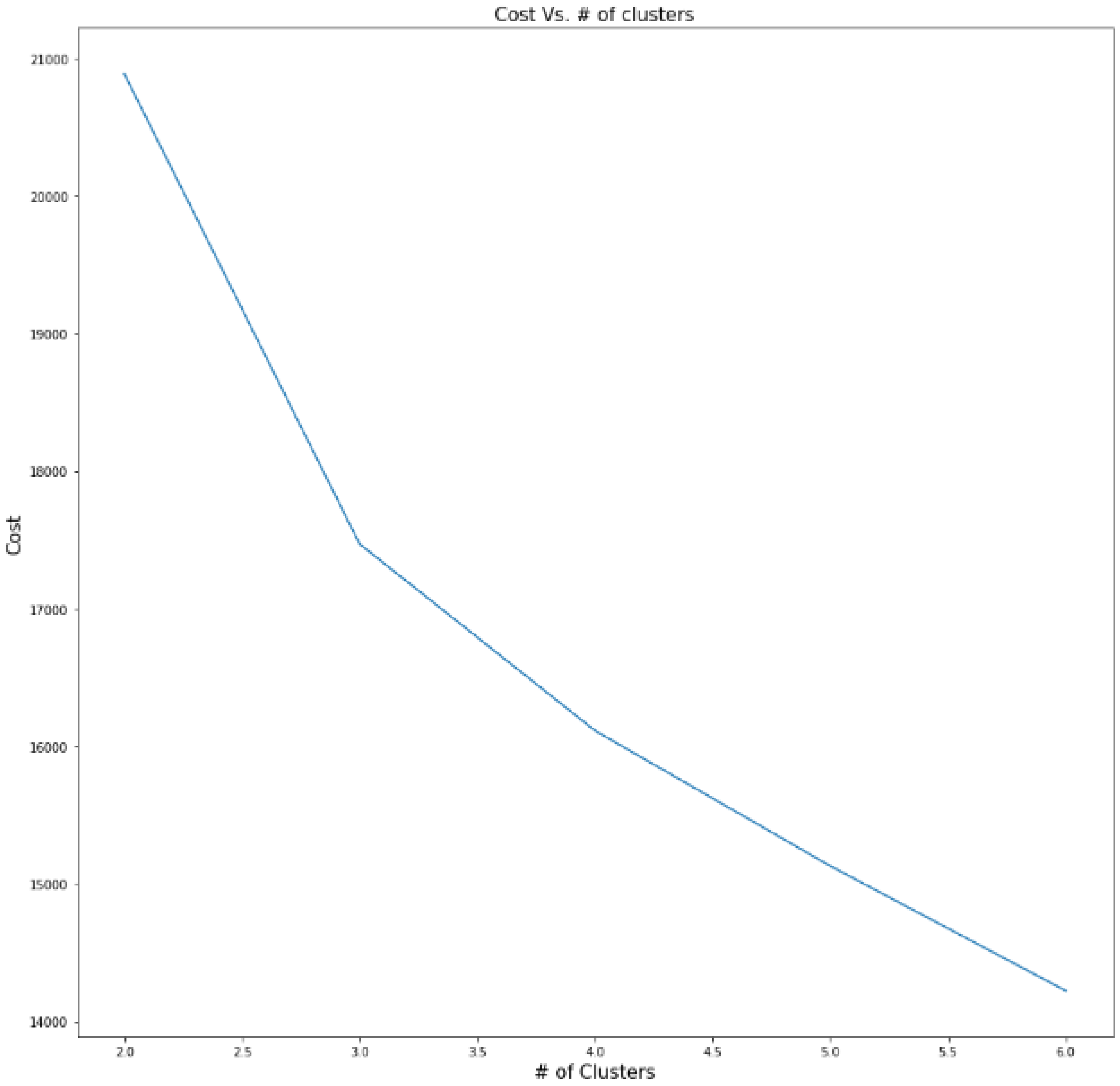


Clustering

- Considering we have both categorical data points and numerical data (The reduced Principle components) we can use the K-Prototype Algorithm that combines the k-means and k-modes algorithms.
- Note that we do not know the appropriate number of clusters that will minimize the cost and be reasonable for analysis.
- In order to identify the appropriate number of clusters we need to perform the clustering process with a different number of centroids and evaluate the appropriate cost
- A good spot is to compute up to 7 clusters as a start and plot the Cost Vs. Number of clusters plot.

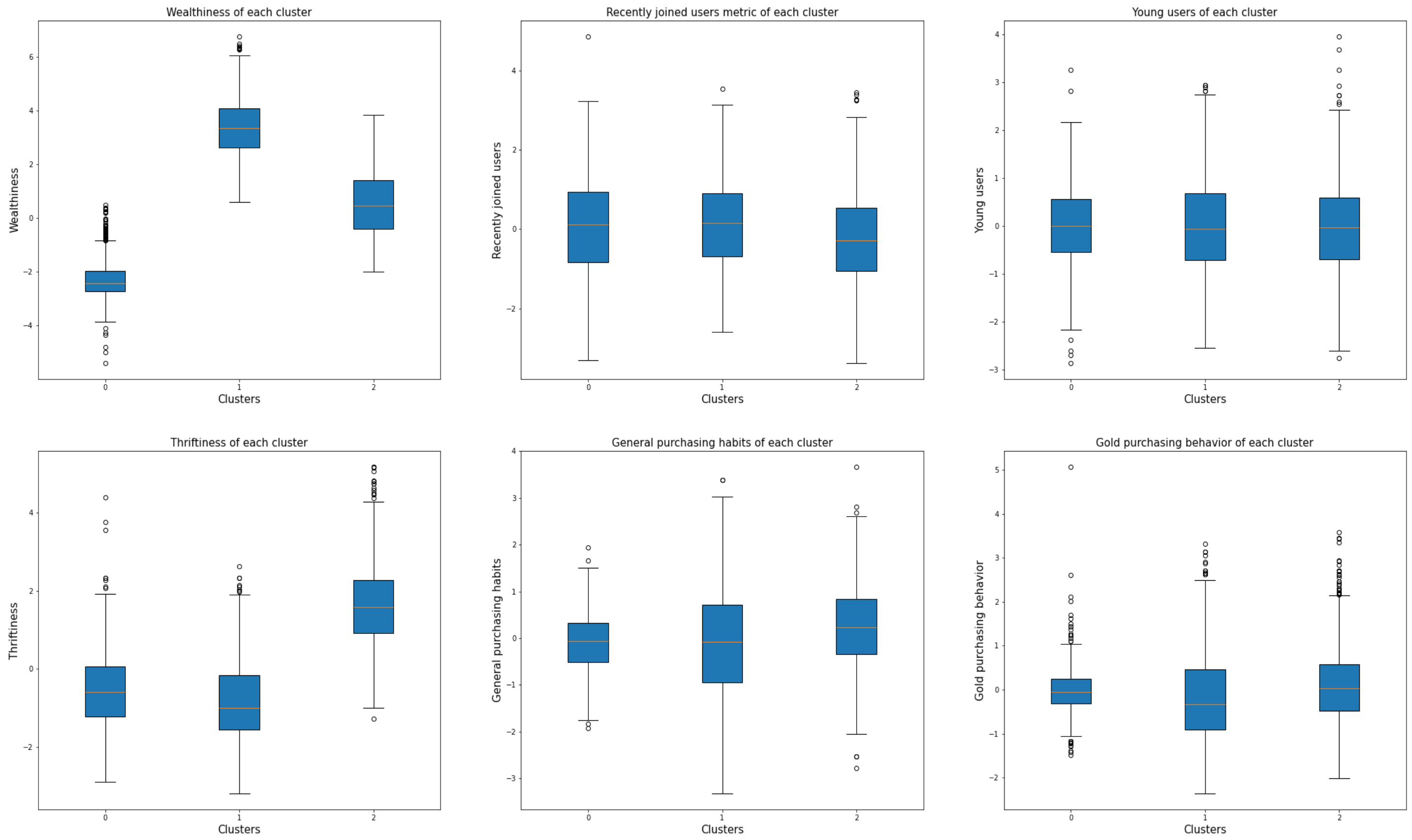
Clustering

- From the plot we can see that a good number of clusters is 3 as it has relatively low cost and can be general enough for our purposes.
- We then rerun the clustering function but with only 3 centroids.



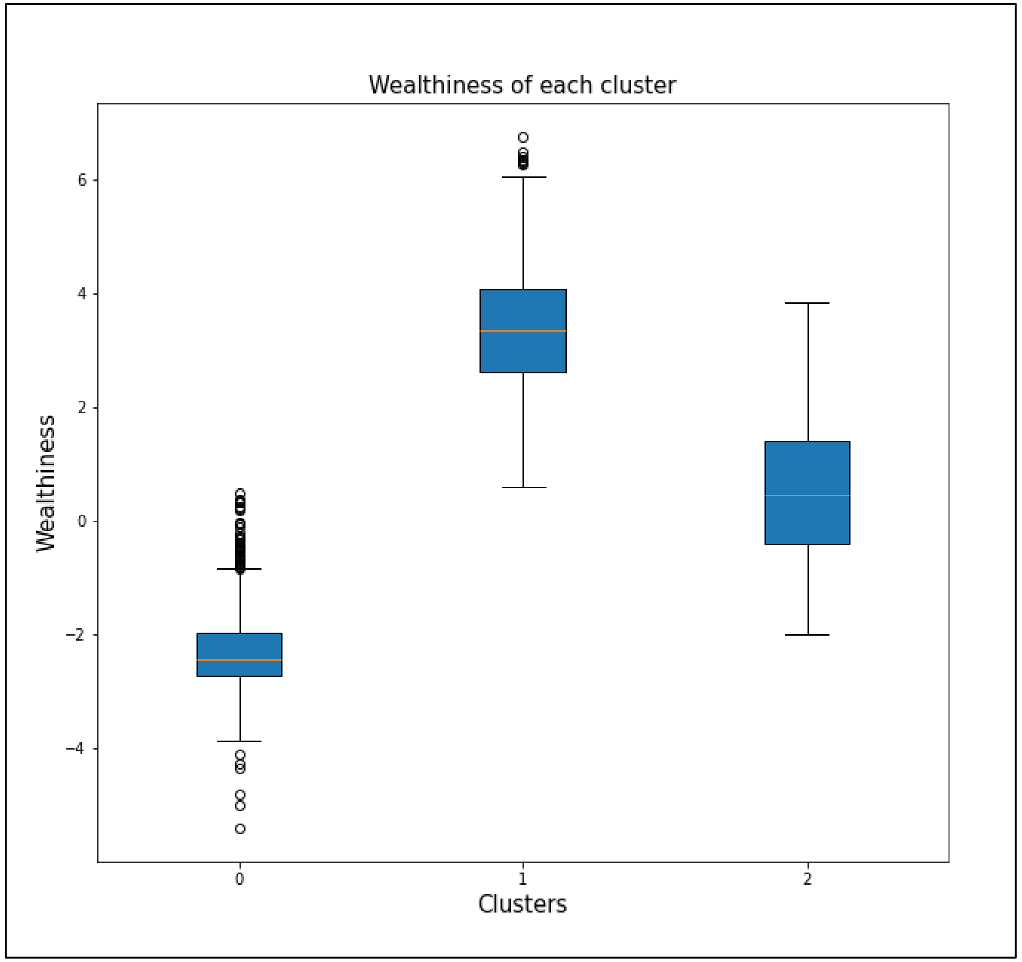
Clustering

- Once clustering Is complete we can describe each cluster by observing the characteristics of each cluster we can start with the Principle components.
- Plotting the Boxplots of the principle components we can see some significant differences for some components while having very little effect on others.



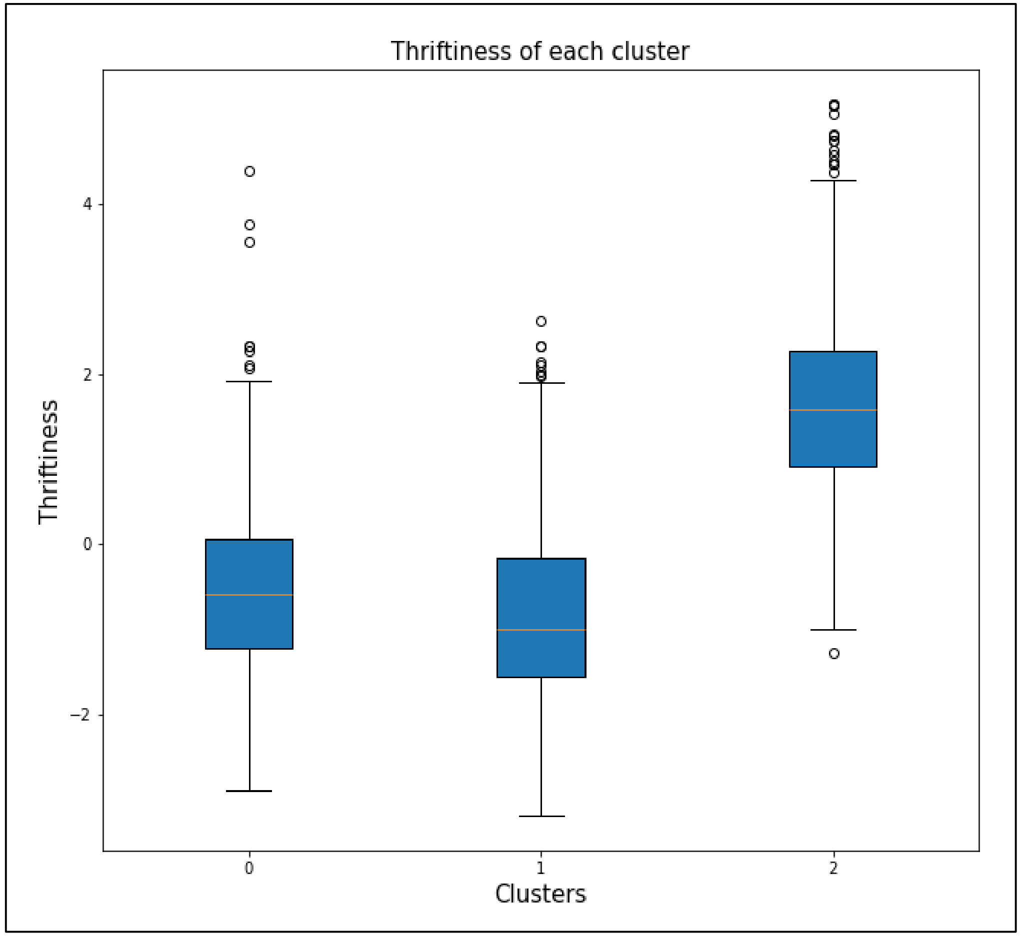
Clustering

- When examining wealthiness we note that cluster 0 is considerably poorer than the rest, while cluster 1 represents the users with the highest income.
- Cluster 2 can be categorized as users with middle income although some overlap has occurred with Cluster 1.
- It's worth noting that cluster 0 covers 45.8% of all users followed by 2 at 27.2% and 1 at 26.5



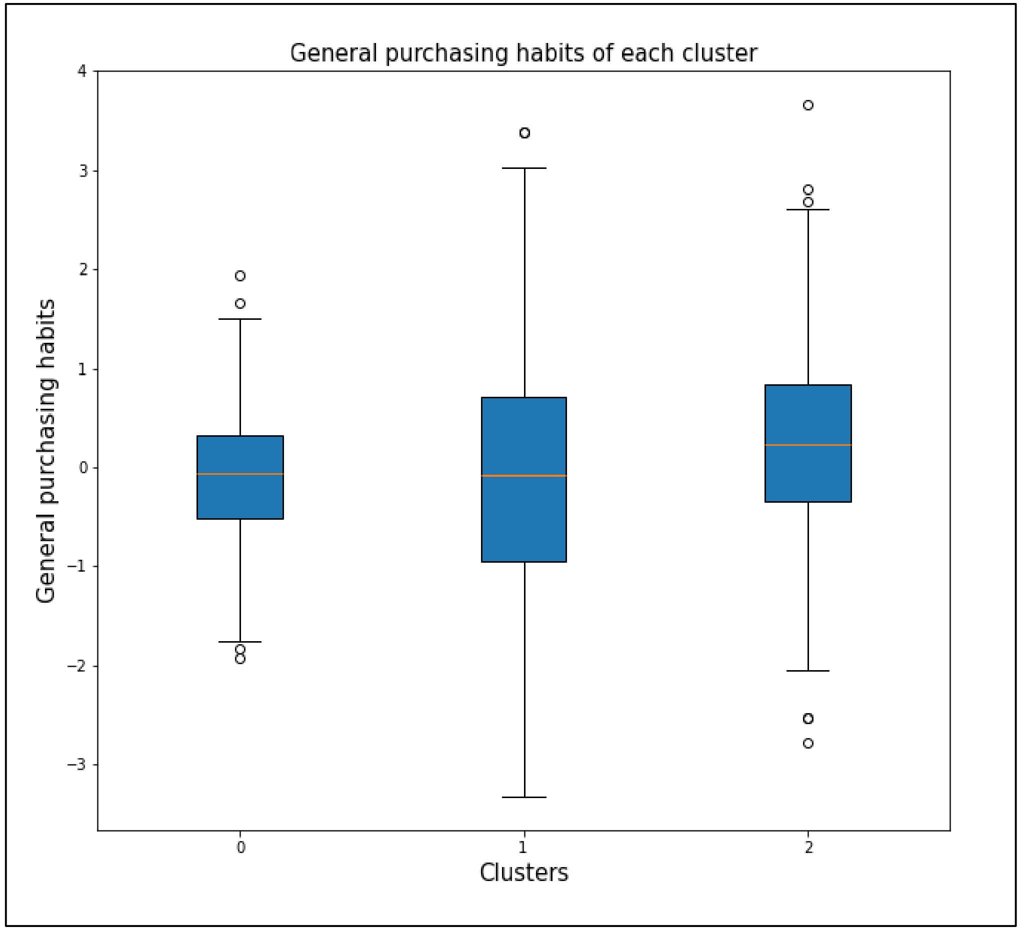
Clustering

- When examining thriftiness we note that cluster 2 is much more cautions regarding spending than the rest of the users.
- While Cluster 0 and 1 have somewhat similar median and spread of thriftiness.



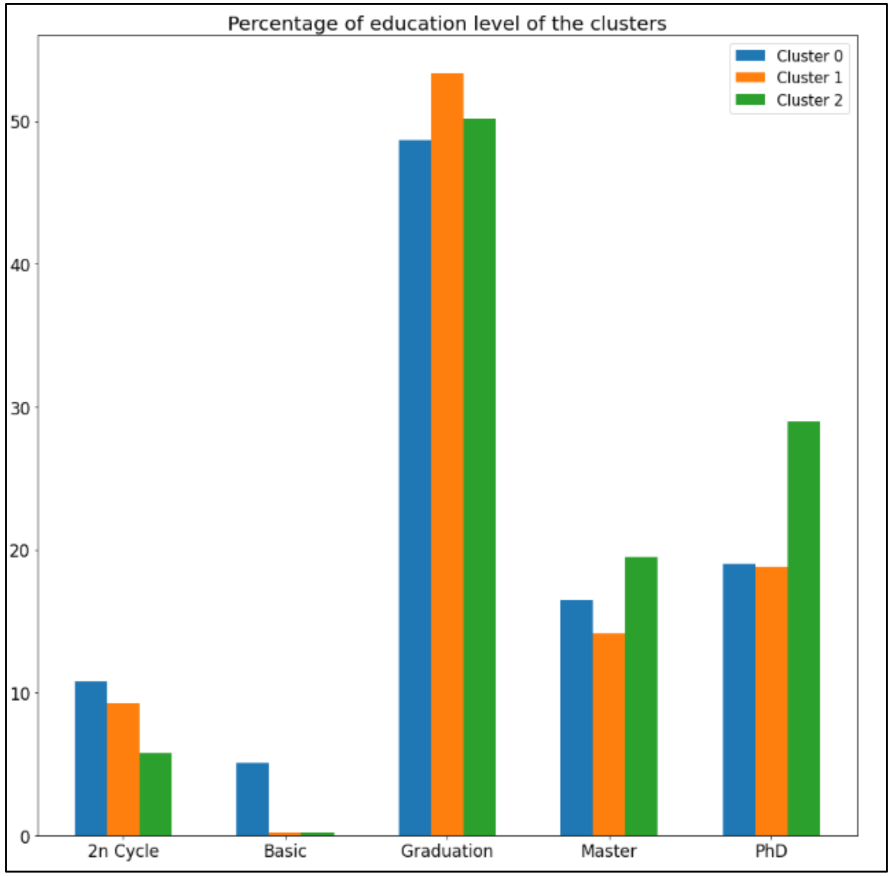
Clustering

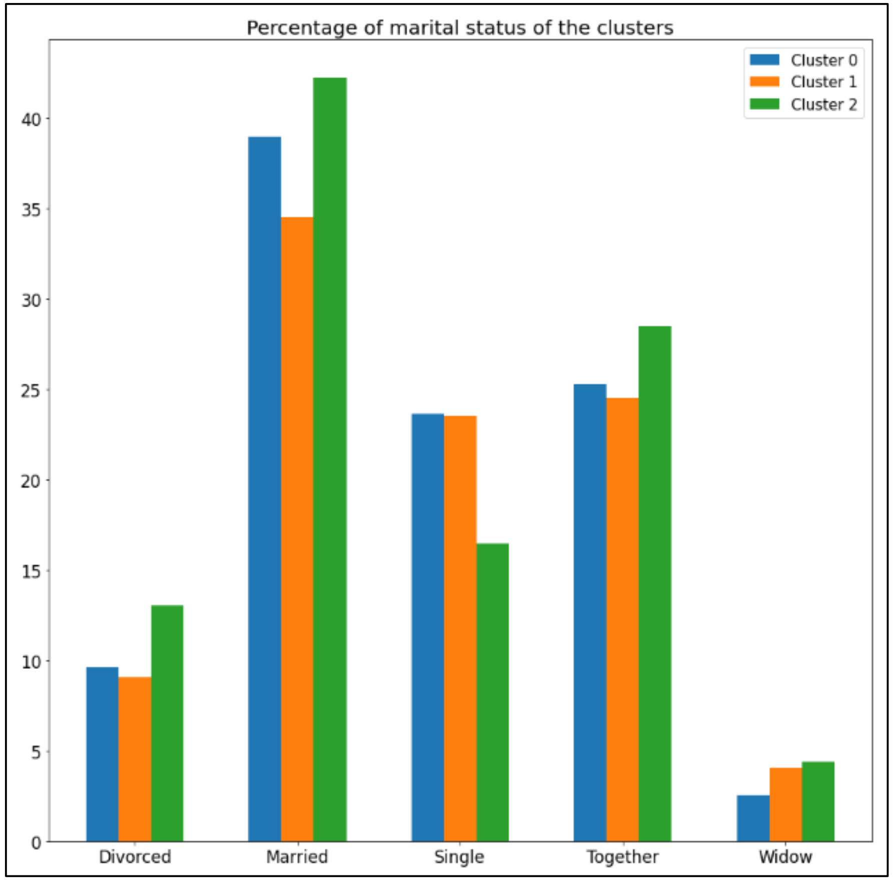
- When examining general purchasing habits, it’s worth noting that within cluster 1 and 2 there appear to be a large spread of purchasing activity compared to cluster 0 which covers the majority of the customers base.



Clustering

- Analyzing the demographic variables doesn't provide any new insights:





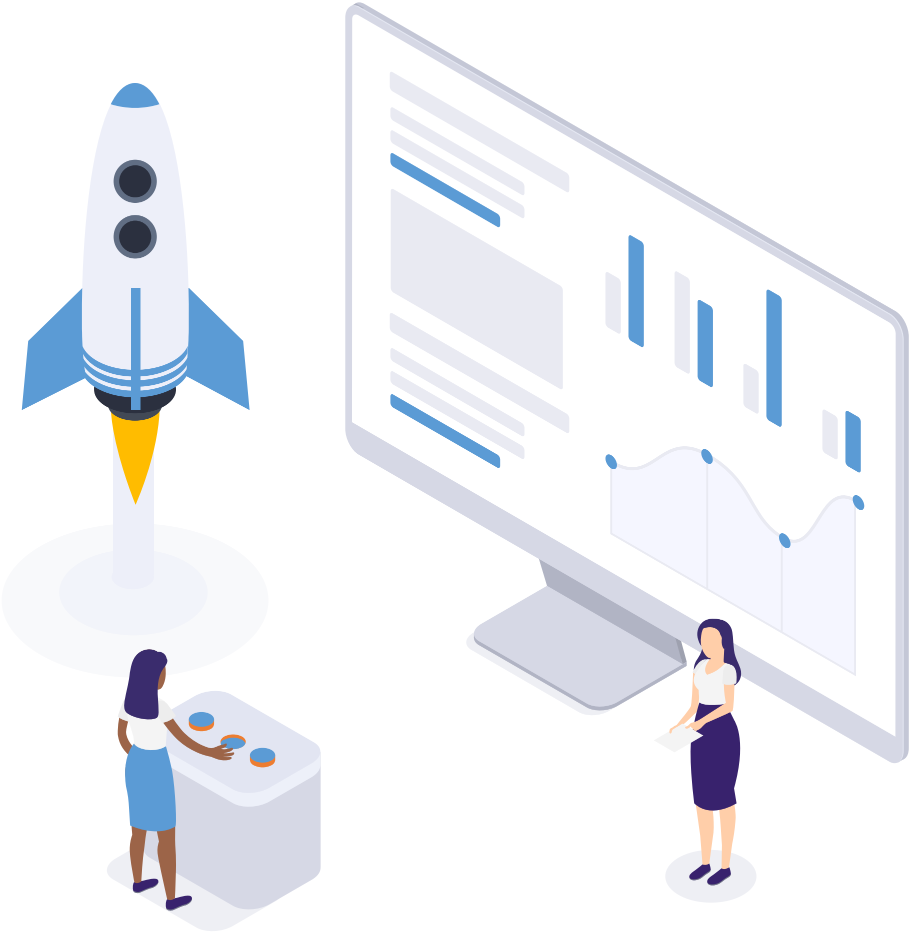
Clustering

After final examination we can categorize the clusters as follows:

- Cluster 0: This cluster covers 45.8 % of the customer base the large majority, in terms of wealthiness this cluster lies in the lower end of the scale relative to the other clusters, as for thriftiness this cluster has somewhat similar behavior to the “Average income/ low activity users” cluster. As such we will categorize this cluster as **Low income/ average activity users.**
- Cluster 1: This cluster covers 26.5 % of the customer base, in terms of wealthiness this cluster lies in the higher end of the scale relative to the other clusters, understandably In terms of thriftiness this cluster is less cautious regarding spending than cluster 0 (low income/activity users). As such we will categorize this cluster as **High income/activity users.**
- Cluster 2: This cluster covers 27.6% of the customer base, in terms of wealthiness this cluster lies in the middle of the scale relative to the other clusters with some small overlap with the High income cluster. In terms of thriftiness this cluster is considerably more cautious regarding spending and deals. As such we will categorize this cluster as **Average income/ low activity users.**

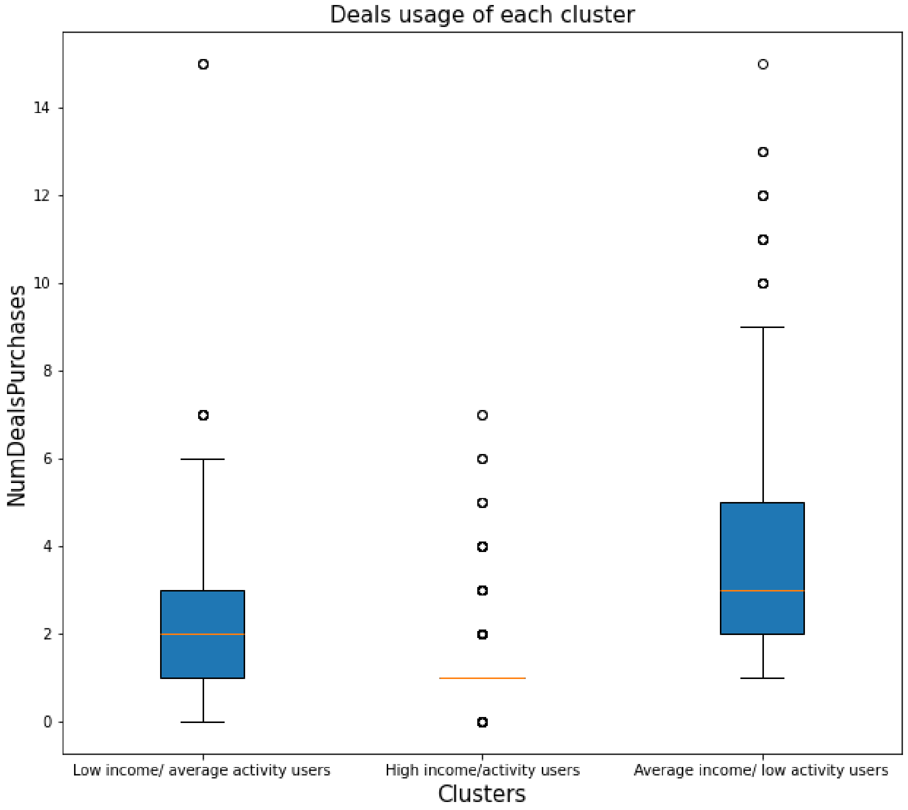
5

Recommendations for future campaigns



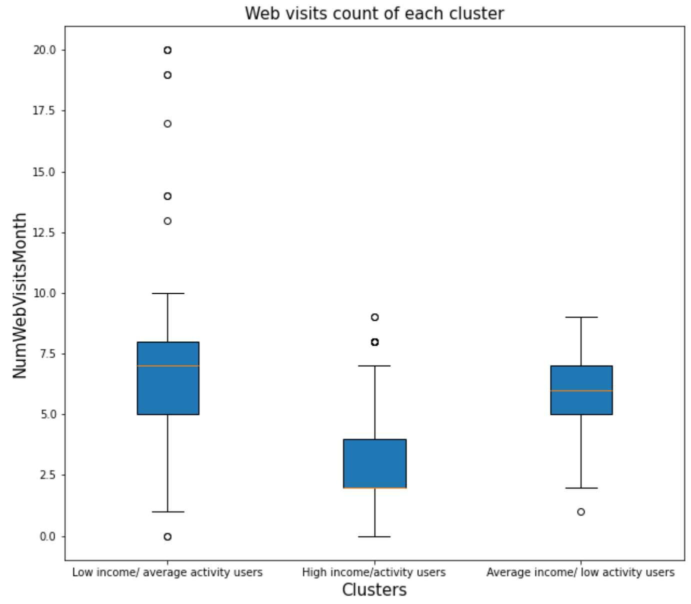
Marketing recommendations

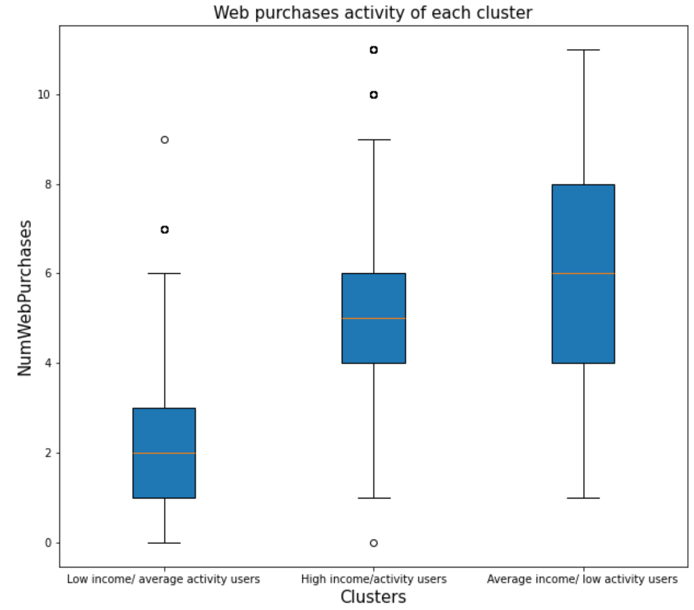
- Starting with the deals usage, it's apparent that while the Low income/ average activity users has a tendency to use deals the Average income/ Low activity users tend to use more deals on average.
- A campaign structured around deals and discounts will have better success than the other groups.
- As expected the High income/ activity users don't use deals that much.



Marketing recommendations

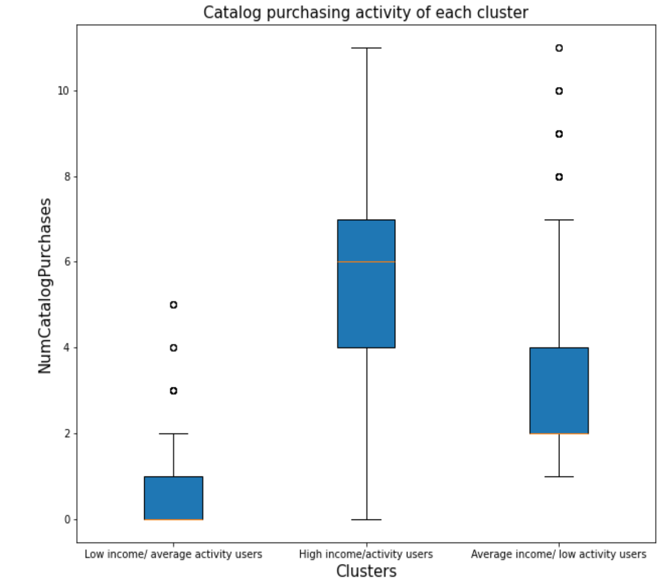
- Analyzing the web activities of the user groups we notice a somewhat high number of visits from the Low income/average activity users while forming the lowest volume of web purchases. Accordingly an online marketing campaign would benefit from this groups potential buyers.
- As for the Average income/ Low activity users they tend to have much larger purchasing activities relative to their web visits.
- And finally for the High income/ activity users they tend to use the web the least while forming the average of the web purchases activities.

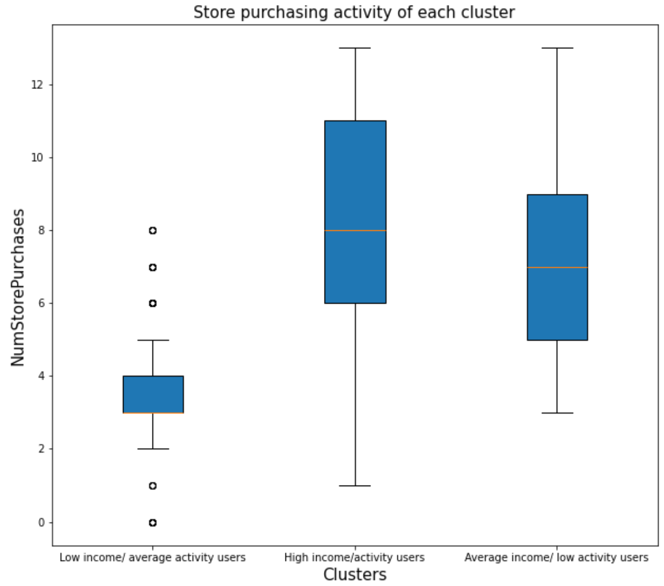




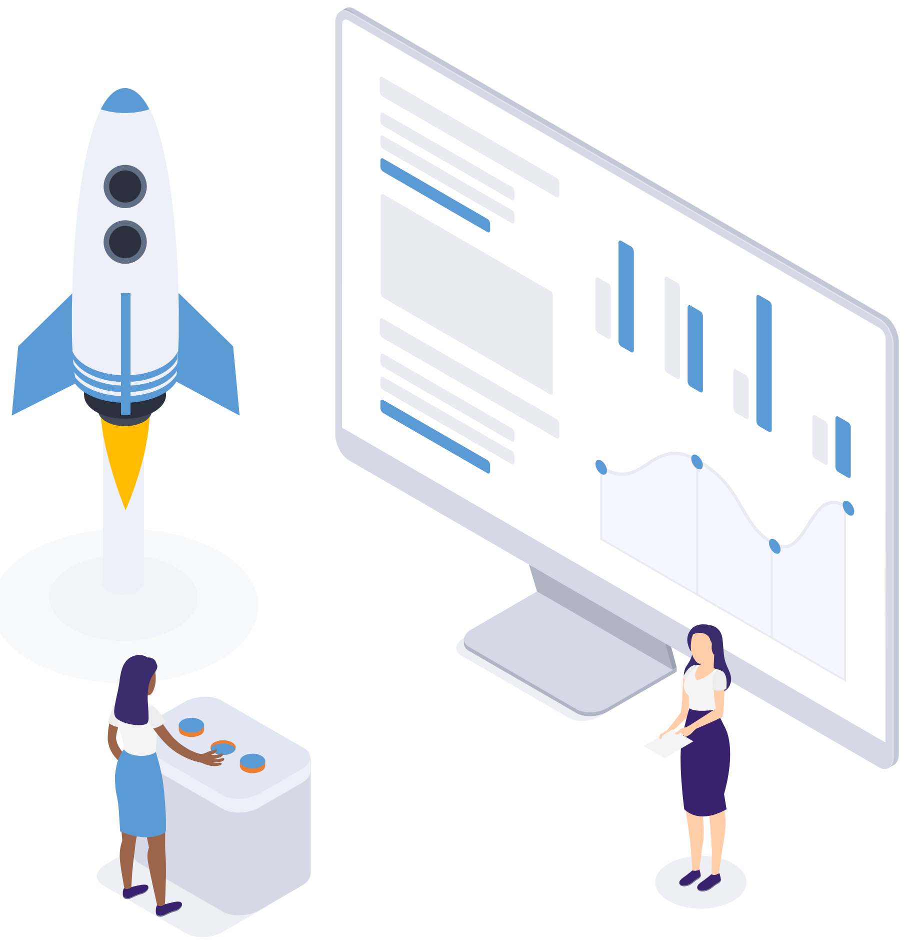
Marketing recommendations

- Examining the other sales outlets, we notice a considerable use of catalog purchasing in the High income/ activity users group followed by the Average income/ low activity users group, Along with the large spread of the purchasing activity it indicates a potential promotional marketing method.
- As for store purchasing, it appears to be the most used for the Average income/ low activity group.





6 Final notes



Final notes

- It’s worth noting that the used clustering method has some overlap in most of the users characteristics, as such we cannot state with absolute certainty that each cluster is completely unique.
- The insights we were able to find were merely a demonstration of the potential of clustering algorithms as there are more insights and notes that can be found.

Thank you