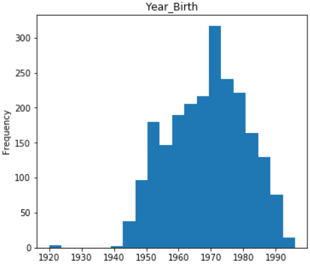
iFood - Campaign Analysis

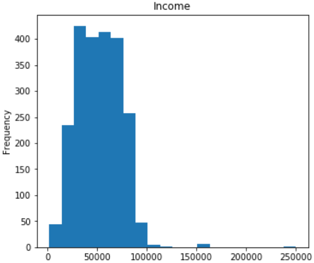
by Erik Davino Vincent

Summary

The user base

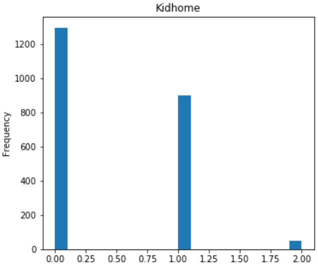
	Year Birth	Income
mean	1968.83	52061
std	11.83	21810
min	1920	1730
max	1996	250000

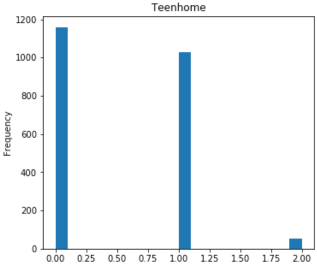




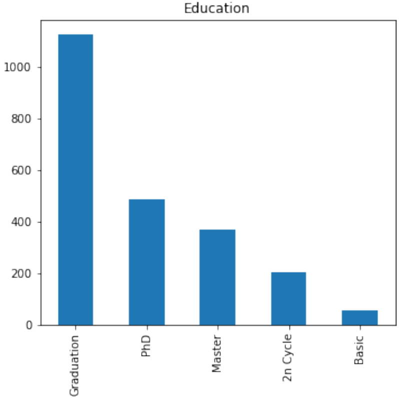
The user base

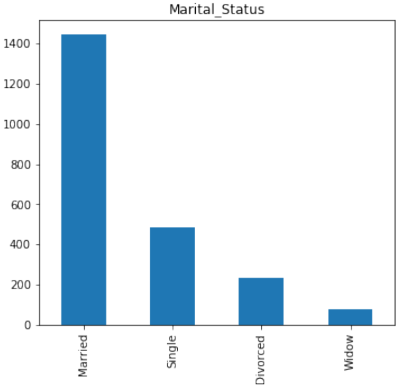
	Kid Home	Teen Home
mean	0.444196	0.506250
std	0.538398	0.544538
min	0.000000	0.000000
max	2.000000	2.000000





The user base

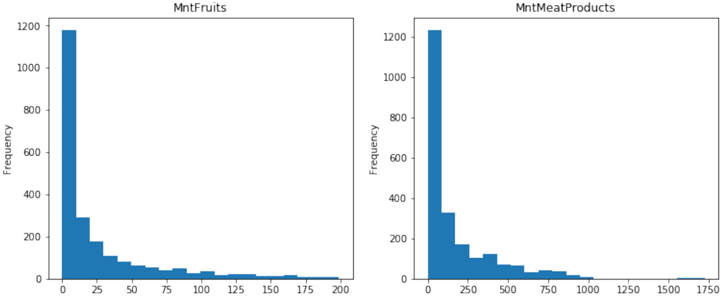




The user base

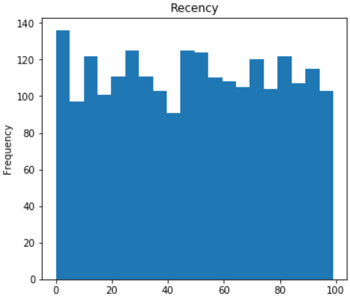
- High level of education - graduation or higher
- Most are married
- At their mid to late fifties, but also a significant amount of younger and older users.
- Almost half has at least one kid and/or teen at home
- Medium income amount - middle class

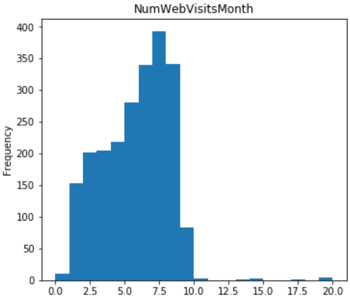
- Most users don't buy large amounts - There is still a considerable amount of users who buy a lot.



Userbase Segmentation - Behavior

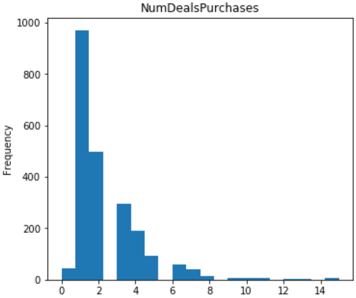
- Recency is uniformily distributed - There are as many users who come back often as there are users who don't
- Most users don't visit the site very often

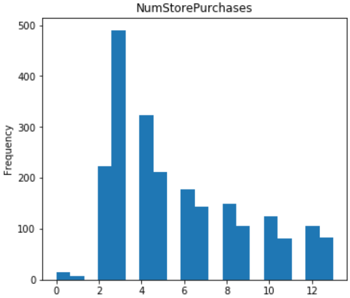




Userbase Segmentation - Behavior

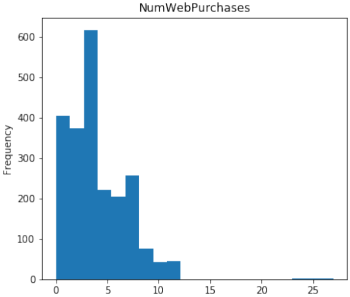
Preferred purchasing method

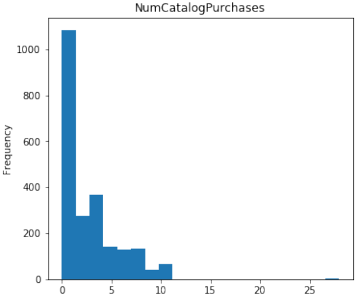




Userbase Segmentation - Behavior

Preferred purchasing method





Userbase Segmentation - Behavior

Correlation:

- Positive correlation between having kids at home and lower income
- Positive correlation between age and higher income
- Positive correlation between having teenagers home and number of web purchases
- Positive correlation between the amount of products bought
- Positive correlation between number of web visits and number of deals purchases
- Positive correlation between accepting different campaign offers, including the current one

Userbase Segmentation - Behavior

Based on the analysis of the data we can segment the users into the following categories:

Userbase Segmentation - Behavior

A.

- High income
- Purchases a lot o products
- More likely to purchase luxury products
- Older and have no kids, but may have teenagers
- Prefer store and catalog purchases

B.

- Low income - at the start of their careers or students
- Don't buy a lot o products
- A lot less likely to purchase luxury products
- Younger and have no kids
- Check the webservice more frequently

Userbase Segmentation - Behavior

c.

- Low income
- Don't buy a lot o products
- A lot less likely to purchase luxury products
- Younger - have kids and or teenagers at home
- Don't use webservice very often

Userbase Segmentation - Behavior

D.

- Medium income - majority of the userbase
- Buy moderate amounts
- More likely to use webservices and make deals purchases
- Neither young nor old - likely have kids/teenagers at home
- Use webservice more often

Userbase Segmentation - Behavior

E.

- Trusty user base
- Have taken offers brefore
- Likely to take a new offer
- Buy larger amounts
- Medium to high income

Userbase Segmentation - Clustering

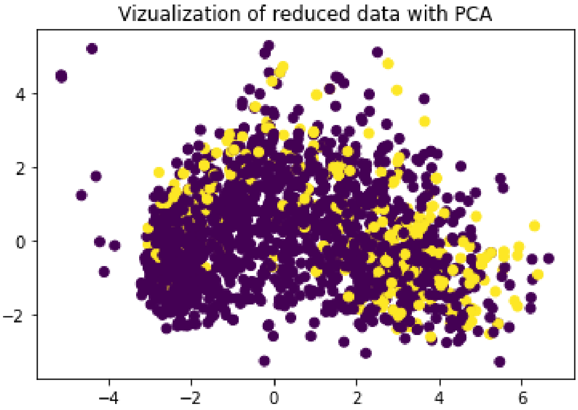


Figure 1: Legend: Yellow - Accepted Offer; Purple - Didn't Accept Offer

Userbase Segmentation - Clustering

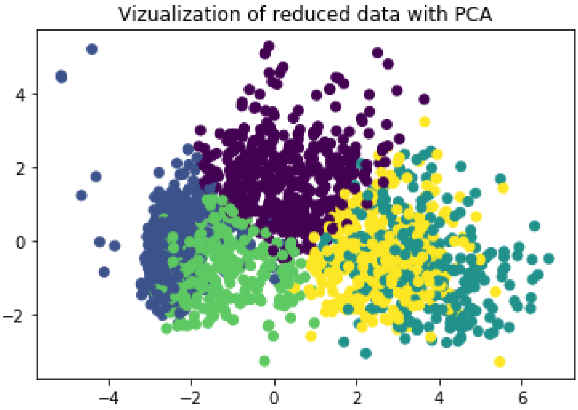


Figure 1: Different groups found by k-means algorithm

Userbase Segmentation - Clustering

Based on the K-means generated groups, we can try and define the main characteristics of these groups as the following:

Group 1

- Low income
- More likely to make deals purchases
- Buy smaller amounts
- Youngest userbase - have children home
- Lower recency
- Make more webvisists but don't buy a lot through the web.

Group 2

- Middle class
- Buy larger amounts
- Younger on average - have no kids or teenagers
- Lower recency
- More likely to give a positive response to the campaign offers, including the new one

Group 3

- Middle class
- Buy larger amounts
- Older on average - have no kids
- Prefer to make catalog purchases
- High recency time
- The most likely to give a positive response to the campaign offers, including the new one - 'loyal' userbase.

Group 4

- Low to medium income
- Buy smaller amounts
- Spend a lot on meats and wines despite lower average income
- Old group - have no kids but have teenagers
- Makes web purchases the most - it's likely that their children do make them

Group 5

- Lower income
- The oldest group
- Buy smaller amounts
- Have high recency time
- They have kids and/or teenagers at home
- The least likely to take campaign offers

Prediction Model

Machine Learning:

- Can make good predictions
- Can be very fast
- 'Learns' about the data by itself
- Feature importance can be used to better understand data and the algorithm's predictions

Prediction Model

Results on training set:

	Precision	Recall	f1-score	Support
0	0.93	0.99	0.96	1527
1	0.93	0.55	0.69	265
Accuracy			0.93	1792
Macro	0.93	0.77	0.82	1792
Weighted	0.93	0.93	0.92	1792

Prediction Model

Results on test set:

	Precision	Recall	f1-score	Support
0	0.90	0.98	0.94	379
1	0.76	0.42	0.54	69
Accuracy			0.89	448
Macro	0.83	0.70	0.74	448
Weighted	0.88	0.89	0.88	448

Prediction Results

Simulation Parameters:

- 448 users to make the offer or not
- Campaign cost for each user: 3 MU
- Campaign revenue for each user if the offer is accepted: 11 MU

Prediction Results

Simulation Results:

- Offer made to 50 users out of 448 MU
- Estimated cost: 150 MU
- Estimated revenue: 396 MU
- Estimated profit: 246 MU
- Gross profit margin ratio: 164%
- Campaign success rate: 72%

Final Considerations

Who accepted the campaign?

- Higher income on average
- Lower recency
- Smaller purchase amounts
- More web and catalog purchases
- Have accepted previous campaign offers

Final Considerations

Prediction model and results

- Easy to implement
- Made good predictions
- Can get better with more data
- Provided a good profit margin with 72% success rate on the test set

The End

- Thank you for your time -