Dec 2023

iFood Market Analysis

Descriptive and Predictive Analysis of Campaign Data

SQL, R, Tableau, Microsoft Excel

Introduction (from Kaggle):

"The new campaign, sixth, aims at selling a new gadget to the Customer Database. To build the model, a pilot campaign involving 2.240 customers was carried out. The customers were selected at random and contacted by phone regarding the acquisition of the gadget. During the following months, customers who bought the offer were properly labeled. The total cost of the sample campaign was 6.720MU and the revenue generated by the customers who accepted the offer was 3.674MU. Globally the campaign had a profit of -3.046MU. The success rate of the campaign was 15%. The objective is of the team is to develop a model that predicts customer behavior and to apply it to the rest of the customer base. Hopefully the model will allow the company to cherry pick the customers that are most likely to purchase the offer while leaving out the non-respondents, making the next campaign highly profitable. Moreover, other than maximizing the profit of the campaign, the CMO is interested in understanding to study the characteristic features of those customers who are willing to buy the gadget."

Objectives:

1. Customer profile analysis - identify prevalent characteristics & behaviors among our customers
2. Create correlation matrix for correlation analysis of variables
3. Customer segmentation
4. Predictive model and recommendations for next campaign to target select customers

Data Information and Collection:

Marketing dataset from Kaggle:
https://www.kaggle.com/datasets/jackdaoud/marketing-data

About the data:
Dataset includes information about:
Customers'

- Age
- Income
- Marital status
- Education
- Days since last purchase
- Complaints

- Days of being enrolled with the company

Number of teens and number of kids in the customer's household recorded.

Spend amount on the following 5 categories recorded:

1. Wines
2. Fruits
3. Meat Products
4. Fish Products
5. Sweet products

Items purchased fall under broader categories of being 'regular' or 'gold' (premium) products.

Purchases from the following channels of sale recorded:

1. Web
2. Catalog
3. Store

Data about if the customer accepted the offer from the company on the Xth campaign included.

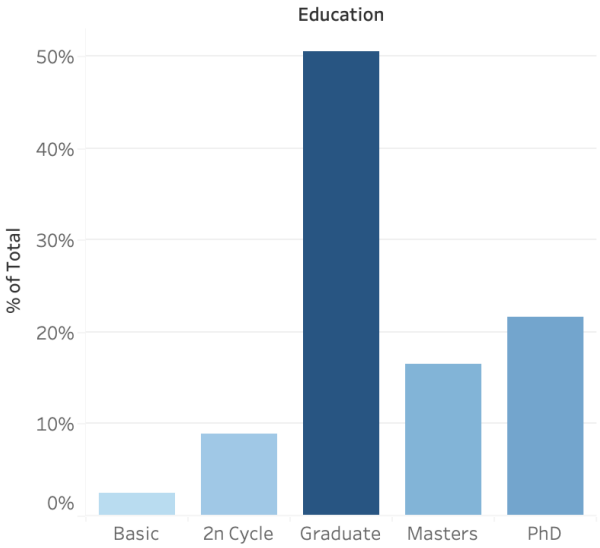
Data Cleaning and Preprocessing:

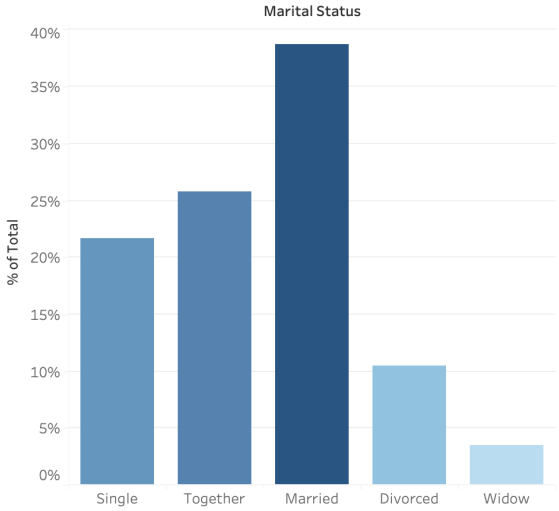
For Tableau

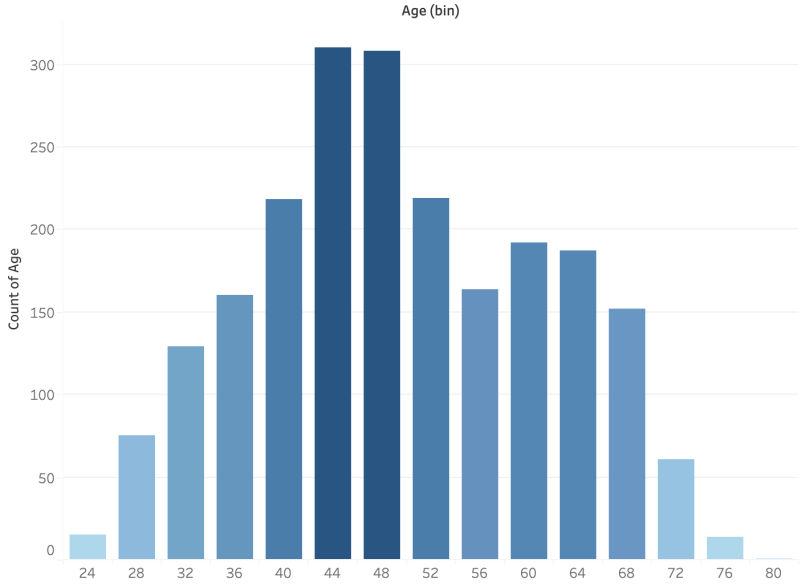
1. Merge education columns into one
2. Merge marital columns into one
3. Add Customer ID column

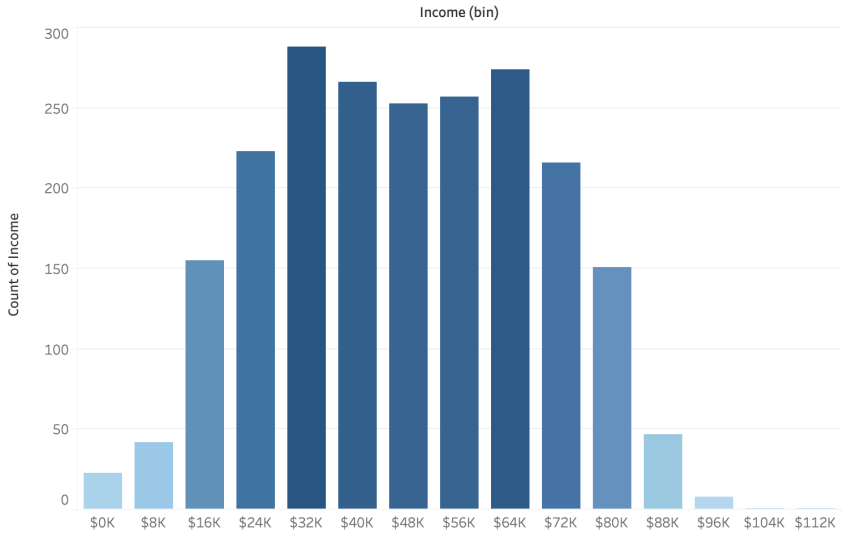
Exploratory Data Analysis:

Customer Profile:









Customer Profile Summary: Most customers tend to be married and/or graduated. Customers' income level ranges from 1.7K to 113K. Customers' age ranges from 24 to 80.

Summary Statistics using SQL:

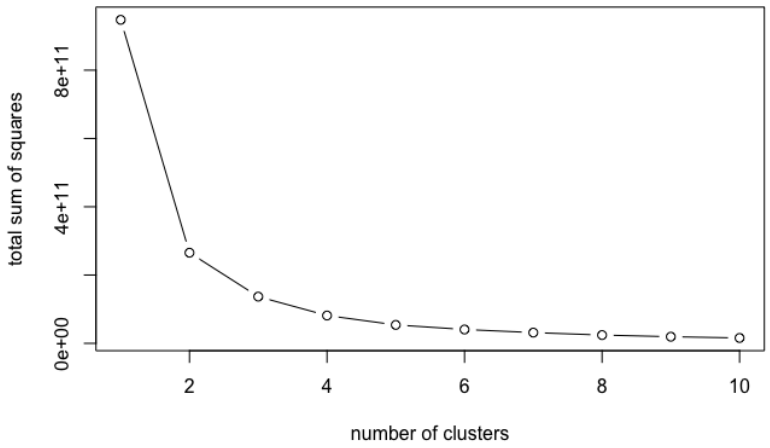
	MEAN Income	MAX Income	MIN Income
▶	51622.0948	113734	1730

	MEAN Age	MAX Age	MIN Age
▶	51.0957	80	24

Customer Segmentation using R:

K-means clustering, elbow method

3 clusters are optimal



Cluster 1: Low Income (AVG $28,000)

1 Kid

Most web visits per month

Cluster 2: Average Income (AVG $51,500)

1 Teen

Cluster 3: High Income (AVG $75,500)

No kids on average

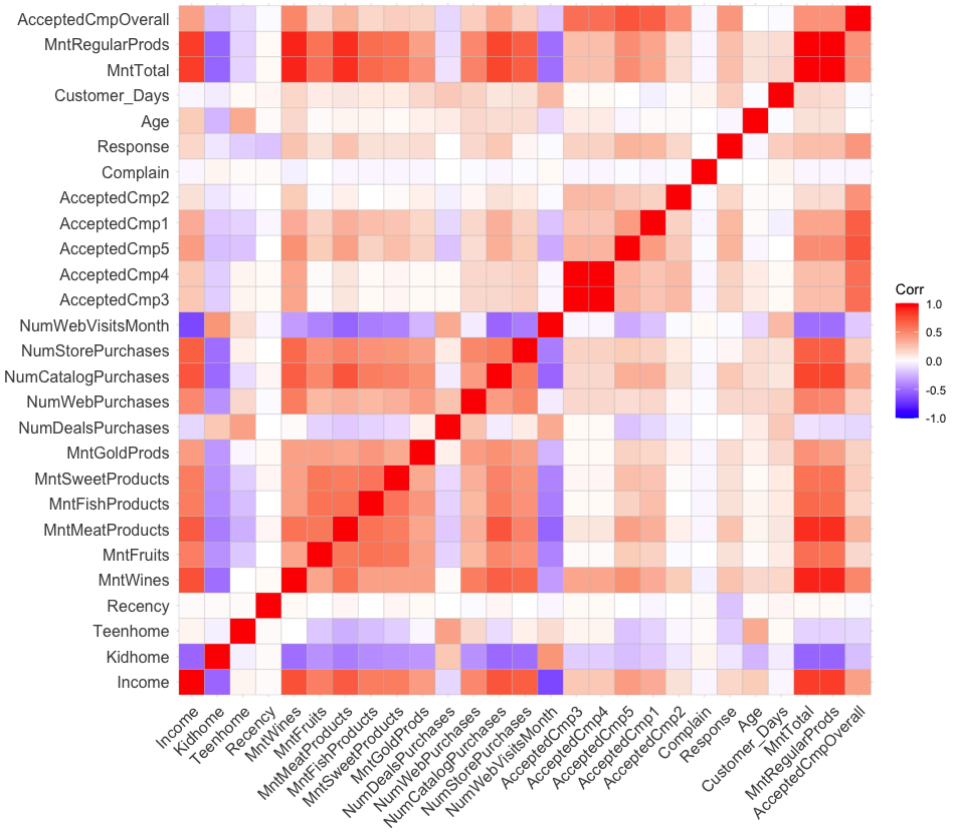
Accepts the most campaigns

Most number of purchases overall

```
> k3 <- kmeans(clustering_ifood,3,iter.max = 100,nstart = 50,algorithm = 'Lloyd')
> print(k3)
K-means clustering with 3 clusters of sizes 724, 766, 715

Cluster means:
    Income      Age    Kidhome  Teenhome  Recency     Wines    Fruits
1 28077.80 46.79282 0.81215470 0.3080110 48.41436  29.86188  5.872928
2 51572.19 53.73107 0.41906005 0.8146214 49.57180 271.36423 17.592689
3 75516.22 52.62937 0.09230769 0.3776224 49.00839 623.22797 56.630769
  MeatProducts FishProducts SweetProducts GoldProds NumDealsPurchases
1     25.30525     9.063536      6.022099  17.50276          2.146409
2     91.58355    24.113577     16.938642  43.96214          3.093995
3    386.06853    81.426573     59.416783  71.04755          1.661538
  NumWebPurchases NumCatalogPurchases NumStorePurchases NumWebVisitsMonth
1        2.142265           0.5290055          3.067680          6.911602
2        4.612272           2.1422977          5.934726          5.744125
3        5.535664           5.3272727          8.495105          3.306294
  AcceptedCmp1 AcceptedCmp2 AcceptedCmp3 AcceptedCmp4 AcceptedCmp5
1  0.001381215   0.00000000   0.08425414  0.004143646  0.000000000
2  0.016971279   0.01436031   0.06657963  0.083550914  0.003916449
3  0.179020979   0.02657343   0.07132867  0.135664336  0.220979021
     Complain Customer_Days   MntTotal MntRegularProds AcceptedCmpOverall
1 0.015193370      2511.544   76.12569        58.62293         0.08977901
2 0.005221932      2522.389  421.59269       377.63055         0.18537859
3 0.006993007      2503.547 1206.77063      1135.72308         0.63356643
```

Correlation Matrix using R (Pearson Method):



Correlation Coefficients Table:

Var1	Var2	Freq
AcceptedCmpOverall	MntWines	0.5099130
Income	MntTotal	0.8230660
	MntRegularProds	0.8168792
	MntWines	0.7304952
	NumCatalogPurchases	0.7100565
	MntMeatProducts	0.7024996
	NumStorePurchases	0.6872057
	MntSweetProducts	0.5556010
	MntFishProducts	0.5517580
	MntFruits	0.5379203
	NumWebPurchases	0.5031842
	Kidhome	-0.5316989
	NumWebVisitsMonth	-0.6483063
Kidhome	NumStorePurchases	-0.5065432
	NumCatalogPurchases	-0.5198133
	Income	-0.5316989
	MntRegularProds	-0.5398280
	MntTotal	-0.5511520
MntFishProducts	MntTotal	0.6350381
	MntRegularProds	0.6206264
	MntMeatProducts	0.5956730
	MntFruits	0.5925564
	MntSweetProducts	0.5829742
	NumCatalogPurchases	0.5631735
	Income	0.5517580

MntSweetProducts	MntTotal	0.6045143
	MntRegularProds	0.5953936
	MntFishProducts	0.5829742
	MntFruits	0.5709861
	MntMeatProducts	0.5565110
	Income	0.5556010
	NumCatalogPurchases	0.5243692
MntTotal	MntRegularProds	0.9965693
	MntWines	0.9023096
	MntMeatProducts	0.8613915
	Income	0.8230660
	NumCatalogPurchases	0.7911869
	NumStorePurchases	0.6778928
	MntFishProducts	0.6350381
	MntFruits	0.6066577
	MntSweetProducts	0.6045143
	NumWebPurchases	0.5210860
	NumWebVisitsMonth	-0.5016390
	Kidhome	-0.5511520
MntWines	MntTotal	0.9023096
	MntRegularProds	0.9018484
	Income	0.7304952
	NumCatalogPurchases	0.6732338
	NumStorePurchases	0.6393728
	MntMeatProducts	0.5931189
	NumWebPurchases	0.5523421
	AcceptedCmpOverall	0.5099130

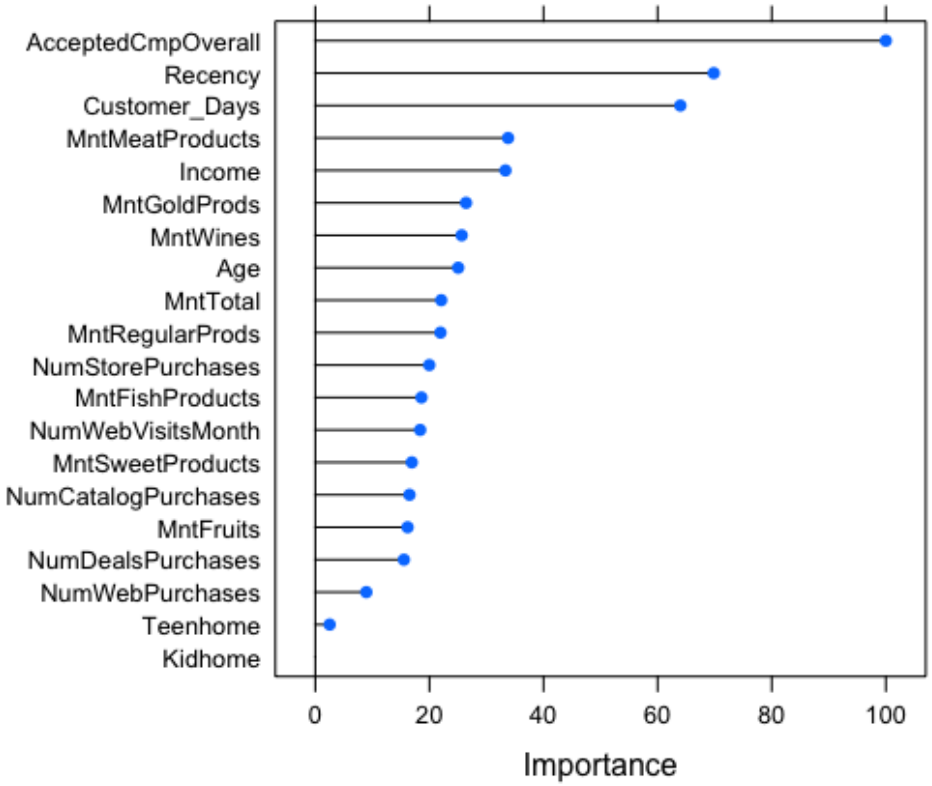
MntFruits	MntTotal	0.6066577
	MntRegularProds	0.5941803
	MntFishProducts	0.5925564
	MntSweetProducts	0.5709861
	MntMeatProducts	0.5681001
	Income	0.5379203
	NumCatalogPurchases	0.5136863
MntMeatProducts	MntTotal	0.8613915
	MntRegularProds	0.8606634
	NumCatalogPurchases	0.7143823
	Income	0.7024996
	MntFishProducts	0.5956730
	MntWines	0.5931189
	MntFruits	0.5681001
	MntSweetProducts	0.5565110
	NumStorePurchases	0.5172449
	NumWebVisitsMonth	-0.5433874
MntRegularProds	MntTotal	0.9965693
	MntWines	0.9018484
	MntMeatProducts	0.8606634
	Income	0.8168792
	NumCatalogPurchases	0.7787424
	NumStorePurchases	0.6686322
	MntFishProducts	0.6206264
	MntSweetProducts	0.5953936
	MntFruits	0.5941803
	NumWebPurchases	0.5039472
	Kidhome	-0.5398280

NumCatalogPurchases	MntTotal	0.7911869
	MntRegularProds	0.7787424
	MntMeatProducts	0.7143823
	Income	0.7100565
	MntWines	0.6732338
	MntFishProducts	0.5631735
	NumStorePurchases	0.5612613
	MntSweetProducts	0.5243692
	MntFruits	0.5136863
	Kidhome	-0.5198133
	NumWebVisitsMonth	-0.5306229
NumStorePurchases	Income	0.6872057
	MntTotal	0.6778928
	MntRegularProds	0.6686322
	MntWines	0.6393728
	NumCatalogPurchases	0.5612613
	MntMeatProducts	0.5172449
	NumWebPurchases	0.5121086
	Kidhome	-0.5065432
NumWebPurchases	MntWines	0.5523421
	MntTotal	0.5210860
	NumStorePurchases	0.5121086
	MntRegularProds	0.5039472
	Income	0.5031842
NumWebVisitsMonth	MntTotal	-0.5016390
	NumCatalogPurchases	-0.5306229
	MntMeatProducts	-0.5433874
	Income	-0.6483063

Predictive Model using R:

5-fold Random Forest classification method, repeated 5 times

Importance of factors ranked to see if customer will respond to campaign



rf variable importance

	Overall
AcceptedCmpOverall	100.000
Recency	69.838
Customer_Days	63.989
MntMeatProducts	33.793
Income	33.329
MntGoldProds	26.424
MntWines	25.660
Age	25.051
MntTotal	22.077
MntRegularProds	21.944
NumStorePurchases	19.985
MntFishProducts	18.609
NumWebVisitsMonth	18.382
MntSweetProducts	16.918
NumCatalogPurchases	16.515
MntFruits	16.186
NumDealsPurchases	15.533
NumWebPurchases	8.946
Teenhome	2.521
Kidhome	0.000

Results:

Key Findings: Accepted Campaigns Overall ranks as the top variable of importance to determine whether a customer will respond to a campaign, followed by Recency and Days of enrollment with the company. Amount spent on wines and meat products highly correlates with income and amount spent total.

Interpretation: Next campaign should target customers who have accepted at least one campaign, while also taking into consideration how recent their last purchase was (30 days or less) and how long they have been enrolled with the company (min. 2650 days)

Meta data:

Feature	Description
AcceptedCmp1	1 if costumer accepted the offer in the 1st campaign, 0 otherwise
AcceptedCmp2	1 if costumer accepted the offer in the 2nd campaign, 0 otherwise
AcceptedCmp3	1 if costumer accepted the offer in the 3rd campaign, 0 otherwise
AcceptedCmp4	1 if costumer accepted the offer in the 4th campaign, 0 otherwise
AcceptedCmp5	1 if costumer accepted the offer in the 5th campaign, 0 otherwise
Response (target)	1 if costumer accepted the offer in the last campaign, 0 otherwise
Complain	1 if costumer complained in the last 2 years
DtCustomer	date of customer's enrollment with the company
Education	customer's level of education
Marital	customer's marital status
Kidhome	number of small children in customer's household
Teenhome	number of teenagers in customer's household
Income	customer's yearly household income
MntFishProducts	amount spent on fish products in the last 2 years
MntMeatProducts	amount spent on meat products in the last 2 years
MntFruits	amount spent on fruits in the last 2 years
MntSweetProducts	amount spent on sweet products in the last 2 years
MntWines	amount spent on wines in the last 2 years
MntGoldProds	amount spent on *gold* products in the last 2 years
NumDealsPurchases	number of purchases made with discount
NumCatalogPurchases	number of purchases made using catalogue
NumStorePurchases	number of purchases made directly in stores
NumWebPurchases	number of purchases made through company's web site
NumWebVisitsMonth	number of visits to company's web site in the last month
Recency	number of days since the last purchase

Table 1: Meta-data table