RFM Analysis for Box Booster.

Kanan Mammadli / 04.14.2022

RFM segmentation allows marketers to target particular groups of customers with communications that are considerably more relevant to their individual behaviors, resulting in higher response rates, increased loyalty, and increased customer lifetime value. RFM segmentation, like other ways to segmentation, is a useful tool for identifying groups of customers who should be addressed differently.

Preprocessing

Before to start Analysis, there is some crucial points that we have to consider. We have to start with data Preprocessing in order to get accurate result. Our dataset contains 18 variables and almost 300,000 observations and the first thing I did, is finding missing values and dropping vari ables where 30% of observations are missing. As a result I dropped "state", ""birthday_year" , "gender", "age" , and "ad_click_id_value" which are contain almost 100% missing values. Actually , the "first_name" , "surname" and some other variables are not good enough to use our analysis, we still keep name and used it in order to decide customer, just because of showing result and bring conclusion.

As date column's type was wrong, I changed it object to date type.

Lastly I checked, "conversion_name", " conversion_currency" and "country" columns and seemed there are no problem for go futher.

What info would enrich dataset?

For this kind of analysis price and quantity of products would enrich dataset and could help to find total amount of money that company earn from selling.

RFM Analysis

First of all, I created "recency" column that represent, how recent is the last purchases of the customers are. Then, in order to find in which countries have more costumer I made a graph that represent customers by countries. Moreover, after some analysis company earn its most revenue from Italy.

I am going to simulate an analysis I am doing in real time by setting the NOW date at one day after the last purchase. This date will be used as a reference to calculate the Recency score. To make things easier, I am going to add a column with the number of days between the purchase and now. To find the Recency values, I will just have to find the minimum of this column for each customer.

I have the values for the Recency, Frequency, and Monetary factors at this stage. For each parameter, each client will receive a note ranging from 1 to 5.

This may be accomplished by establishing ranges depending on predicted behavior. To rate Recency, for example, we may use the following scale:

1: 0 to 30 days

2 60 days

3: 61 and 90 days

4: 91 to 180 days

5. 181-365 days

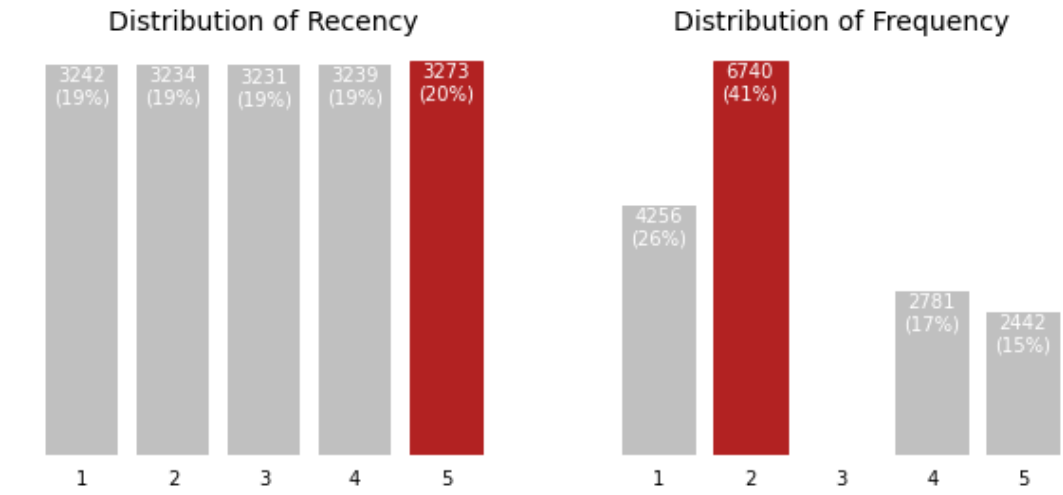
Quintiles might also be used. Each quintile is made up of 20% of the population. Using quintiles is more versatile since the ranges adjust to the data and may be used across sectors or if projected consumer behavior changes. The quintiles approach will be used. I begin by obtaining the quintiles for each parameter.

Then I provide ways for assigning ratings ranging from one to five. A lower Recency value is preferable to a greater Frequency and Monetary value. I'll have to write two different procedures.

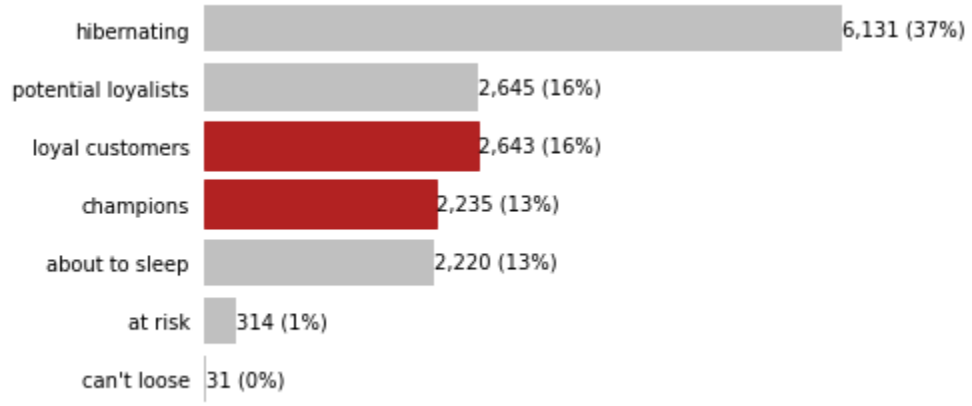
Then I provide ways for assigning ratings ranging from one to five. A lower Recency value is preferable to a greater Frequency and Monetary value.

Moreover, I created matrix for based on the R and F ratings, I'll work with 11 segments.

Segment	Description
Champions	Bought recently, buy often and spend the most
Loyal Customers	Buy on a regular basis. Responsive to promotions.
Potential Loyalist	Recent customers with average frequency.
Recent Customers	Bought most recently, but not often.
Promising	Recent shoppers, but haven't spent much.
Customers Needing Attention	Above average recency, frequency and monetary values. May not have bought very recently though.
About To Sleep	Below average recency and frequency. Will lose them if not reactivated.
At Risk	Purchased often but a long time ago. Need to bring them back!
Can't Lose Them	Used to purchase frequently but haven't returned for a long time.
Hibernating	Last purchase was long back and low number of orders. May be lost.



When we look at graphs that we visualized, it seems higher Recency score which would be better when it is lower and , we expect higher frequency score where it was just in 2. We can analyze from graph, 67% of our customers are not buying very often. For Recency it is evenly distributed and we can say, recent buys are not good as perfect and not bad as awful.



From last graph we can see that companies' 29% of customers are either loyal or champions while most of them (37%) are hibernating. So, most of customers are pretty loyal to company.

So which analysis could be alternative to RFM ?

Kernel density estimation is a technique that may be used as an alternative to RFM. It uses the same set of characteristics as the previous system, but gives each client a weighting based on their

expected future activity or utility value. To put it another way, it's more predictive than descriptive, highlighting clients who are about to become inactive or extremely valued.