Business Analytic Report

Fereshteh Rabbani

19-06-2020

1 KPI Metrics

In the following, we will do some exploratory Data Analysis on our data and try to gain some insights that will help in some decision making of the business.

1.1 Data integrity

- **Missing Values:** The missing values are identified by Pandas. There are only 0.028 percent missing values in "User-ID" column.
- **Dropping irrelevant data or Null values:** There may be data included that is not needed to improve our results. In this data set we have dropped Null values.
- **Converting Features From Object:** The date columns identified as objects by Pandas were converted to date type.
- **Removing Duplicates:** We may have to deal with duplicates, which will skew our analysis. The duplicate records removed by Pandas drop function.
- **Merging two tables** The two tables were merged on the 'conv-id' primary key.

1.2 Monthly Revenue

Monthly revenue data can allow us to analyze short-term company performance and seasonal trends. In fig.1, Monthly Revenue significantly jumped from April to May and went back to previous levels afterward.

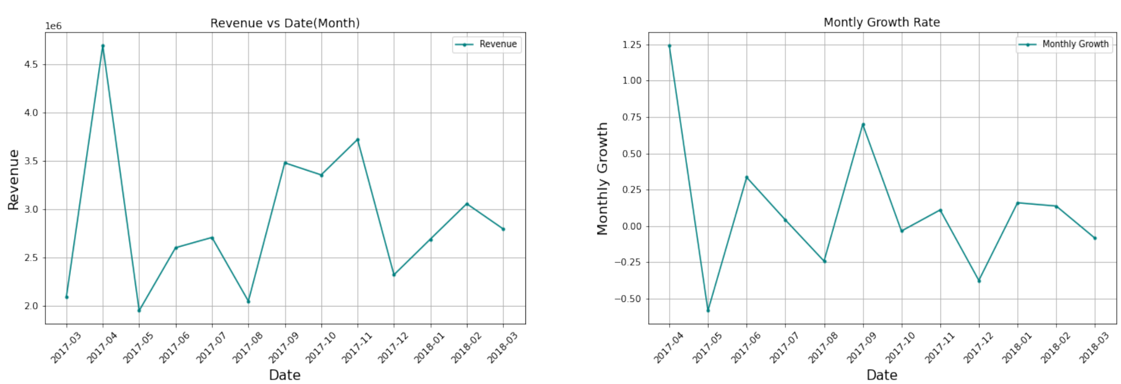


Figure 1: Monthly Revenue

1.3 Monthly Active Customers

From a business perspective, it is important to know that the product is actually being used and is useful to the customers. The presence of Active users might indicate that people are interacting with the service or product. Tracking the number of active users over time helps with assessing the effectiveness of the marketing campaigns and customer experience. We can get monthly active customers by counting unique User-IDs. As we can see in May 2017, the numbers dropped drastically from 8308 to 4397.

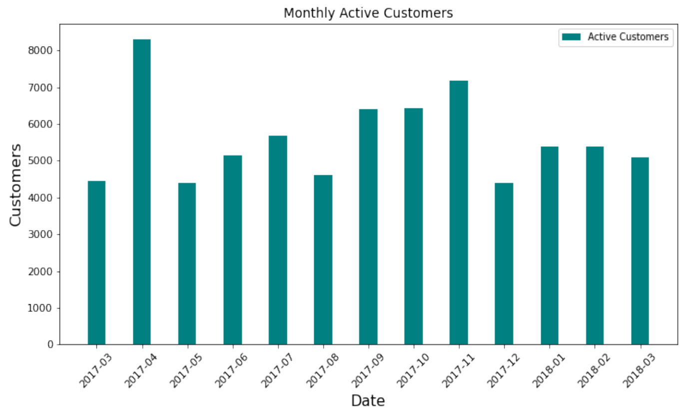


Figure 2: Monthly active customers

1.4 Customers Ratio

In our dataset, we can assume a new customer is whoever did his/her first purchase in the time window (monthly) we defined. We have been using ".min()" function to find our first purchase date for each customer and define new customers based on that. The line plot below (3) shows us the ratio for each group monthly. A similar trend is observed for existing and new customers and indicates that our customer base is declining in August and December.

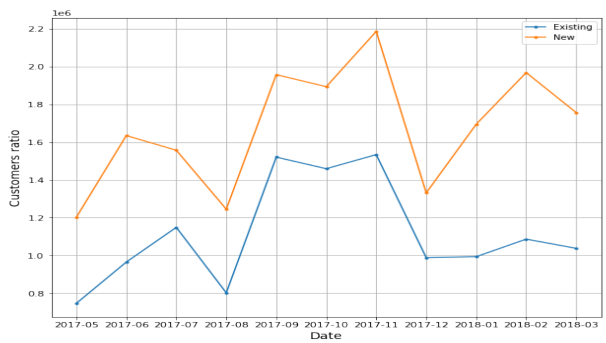


Figure 3: Customer ratio over time

1.5 Performance and impact of different channels

The fig.4 shows our data broken down by channel. We can see that channel "A" and "G" generated the highest amount of revenue and a significantly higher conversion rate than the other channels that contribute significant revenue. This information tells us that users coming from these sites are much more likely to convert into purchasers. By aggregating revenue per channel, we can get the most effective channels.

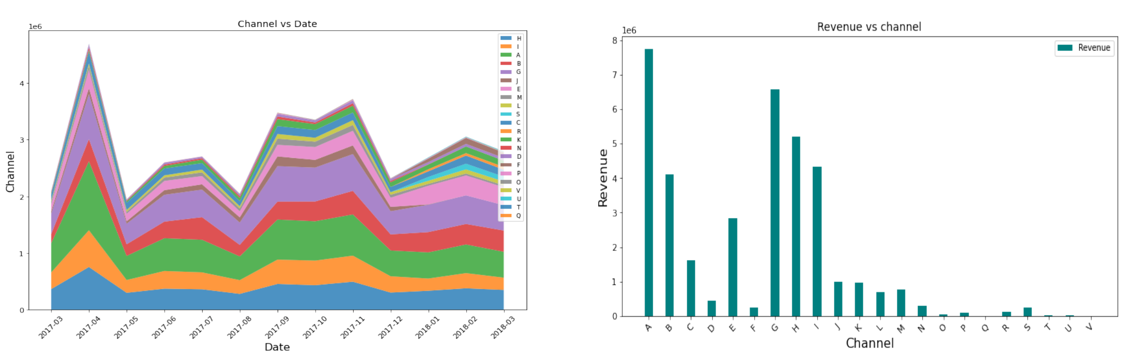


Figure 4: the performance of different channels

2 Cohort Analysis

Cohort analysis is a useful tool to measure user engagement and spot behavioral differences or similarities between them over time. The fig.5 is showing customer retention N months after acquisition.

- the Y-axis shows months in which the customers were acquired (the cohorts).
- the X-axis shows the number of months after the first interaction (1,2,3, up to 13).
- each square represents the percentage of returning users in a specific month.

Notice that the first period of each cohort is 100 percent because our cohorts are based on each user's first purchase, meaning everyone in the cohort purchased in month 1.

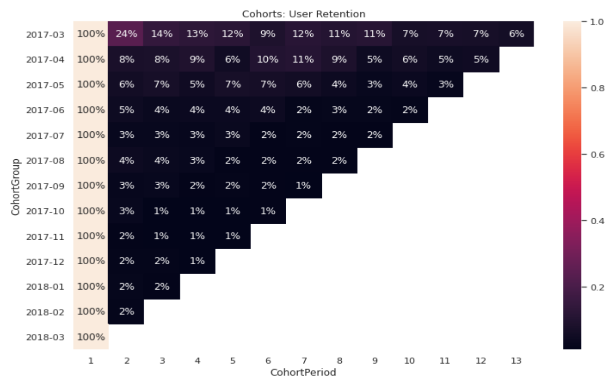


Figure 5: The impact plot

We can see from the above chart that fewer users tend to purchase as time goes on. However, we can also see that the 2017-03 cohort is the strongest, which enables us to ask targeted questions about this cohort compared to others.

3 Customer Segmentation

Customer segmentation is the process of grouping customers by common attributes or characteristics, such as demographic or psychographic. The objective is to have a deeper understanding of each segment so it can market and message effectively.

3.1 An analysis of the main customer segments using RFM

RFM stands for *recency, frequency and monetary value (revenue)*. It is a customer segmentation technique that uses past purchase behavior to divide customers into groups.

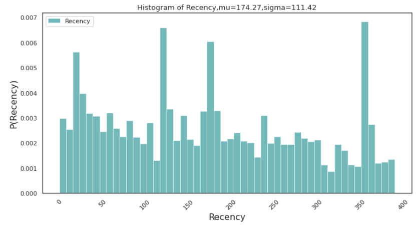
3.1.1 Calculate RFM score and segment customers

We have calculated RFM parameters and applied unsupervised machine learning (K-means clustering) to assign a score and identify different groups (clusters) for each. The overall score (cluster numbers) is classified as:

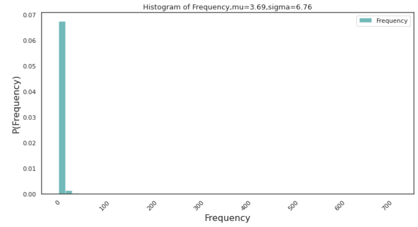
- 0 to 2: Low Value (not very frequent buyer/visitor and generates very low revenue.)
- 3 to 4: Mid Value (fairly frequent and generates moderate revenue.)
- 5+: High Value (High revenue, frequency and low inactivity.)

The distribution of R, F and M (fig.6) can help us to get a better idea of our customer's portfolio.
In fig.7, We can see how the segments are clearly differentiated from each other in terms of RFM. We can start taking actions with this segmentation. The main strategies are quite clear:

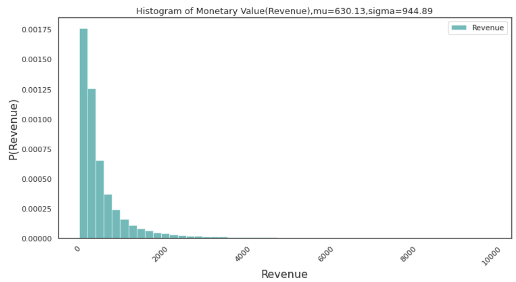
- High Value: Improve Retention
- Mid Value: Improve Retention + Increase Frequency
- Low Value: Increase Frequency



(a)

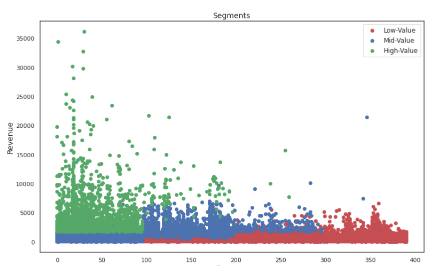


(b)



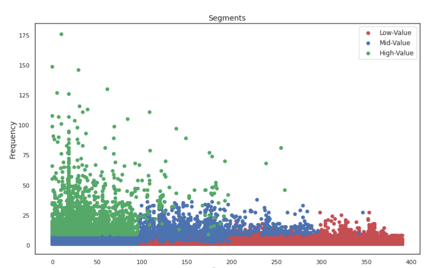
(c)

Figure 6: Distribution of RFM across customers



(a)

(b)



(c)

Figure 7: Segmenting customers using K-means clustering