Customer Segmentation & Retention

Harsh Talajia

20/03/2021

Objectives

A) To find the subset of best customers who are eligible for receiving gift hampers.

B) To find a suitable set of customers to whom the company should provide coupons and offers so as to retain them and avoid their churn.

Methodology

Trying to understand each individual customer's behaviour is not feasible. Behavioural segmentation of customers would enable us to form groups of customers. Customers within a particular group would exhibit similar purchasing behaviour. Hence, each group can have its own separate marketing campaign which is most suitable for it.

Technique used

Based on the data provided, RFM segmentation seems suitable. Customer segments are formed on the basis of three factors:

Recency(R): How recent was customer's latest purchase.

Frequency(F): How frequently the customer purchased in the given time frame.

Monetary(M): How much the customer spent in the given time frame.

About Data

ï..Sno	Category	City	State	Country	Customer_ID	Order_Date
1	Baby Care	Los Angeles	California	United States	347	2018-04-01
2	Baby Care	Soledad DÃez GutiÃ©rrez	San Luis PotosÃ	Mexico	3500	2018-04-01
3	Baby Care	Detroit	Michigan	United States	6497	2018-04-01
4	Baby Care	San Diego	California	United States	642	2018-04-01
5	Baby Care	Amstelveen	North Holland	Netherlands	4809	2018-04-01
6	Baby Care	Sofia	Sofiya-Grad	Bulgaria	4411	2018-04-01

Order_Priority	Product_ID	Ship_Mode	Sub_Cat	Discount	Profit	Quantity	Unit_Price
Medium	TEC-PH-10003885	Standard Class	Grooming	20%	24	4	53
High	TEC-PH-10003405	Standard Class	Grooming	0%	7	4	45

Order_Priority	Product_ID	Ship_Mode	Sub_Cat	Discount	Profit	Quantity	Unit_Price
High	TEC-PH-10002660	First Class	Grooming	0%	136	8	68
Medium	TEC-PH-10002398	Standard Class	Grooming	20%	56	5	89
Medium	TEC-MA-10004897	Standard Class	School Supplies	50%	-74	4	19
Medium	TEC-EPS-10001129	Standard Class	School Supplies	0%	86	1	305

Transaction level data of 52,190 orders from 9,127 customers with 15 variables is provided.

Exploring Data

We add a new variable named 'Revenue' to our data, which is calculated as:
Revenue = Quantity * Unit_Price

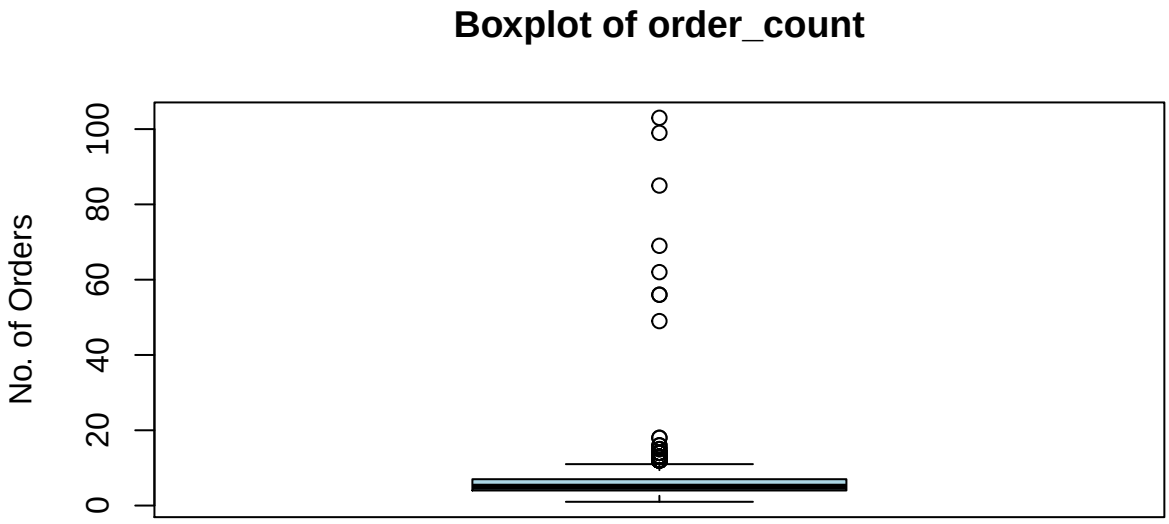
Checking Missing values

	x
ï..Sno	0
Category	0
City	0
State	0
Country	0
Customer_ID	0
Order_Date	0
Order_Priority	0
Product_ID	0
Ship_Mode	0
Sub_Cat	0
Discount	0
Profit	0
Quantity	0
Unit_Price	251
Revenue	251

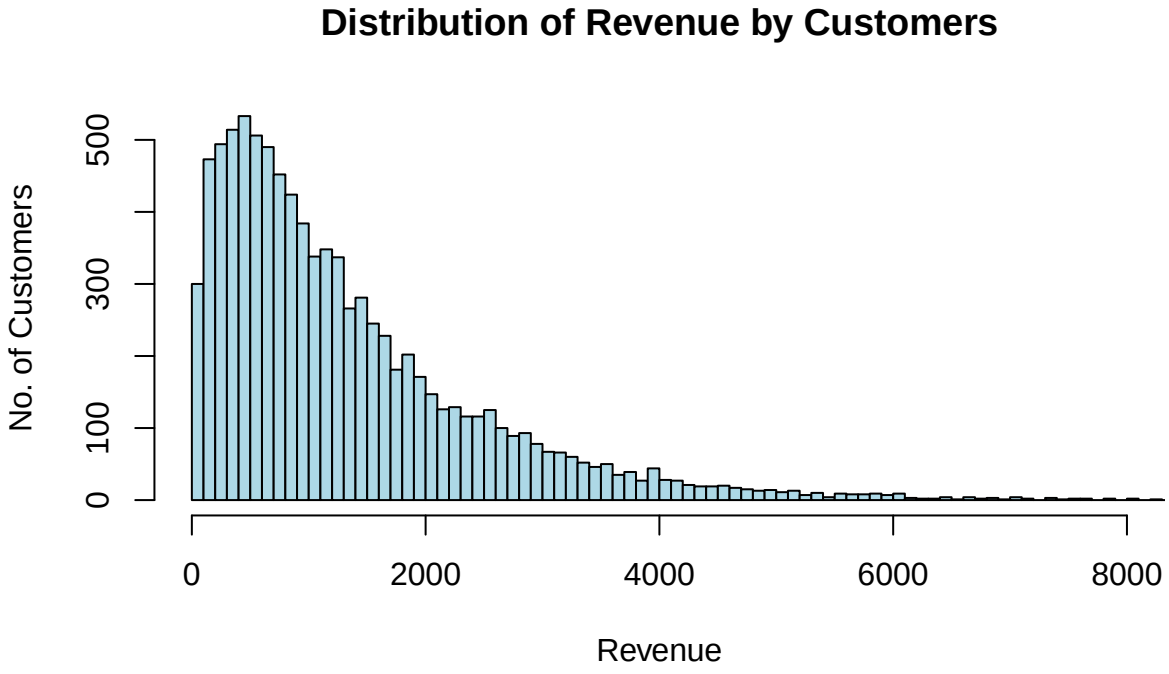
These 251 rows where Unit_Price is missing are **removed** from the data for subsequent analysis.

Checking Order frequencies

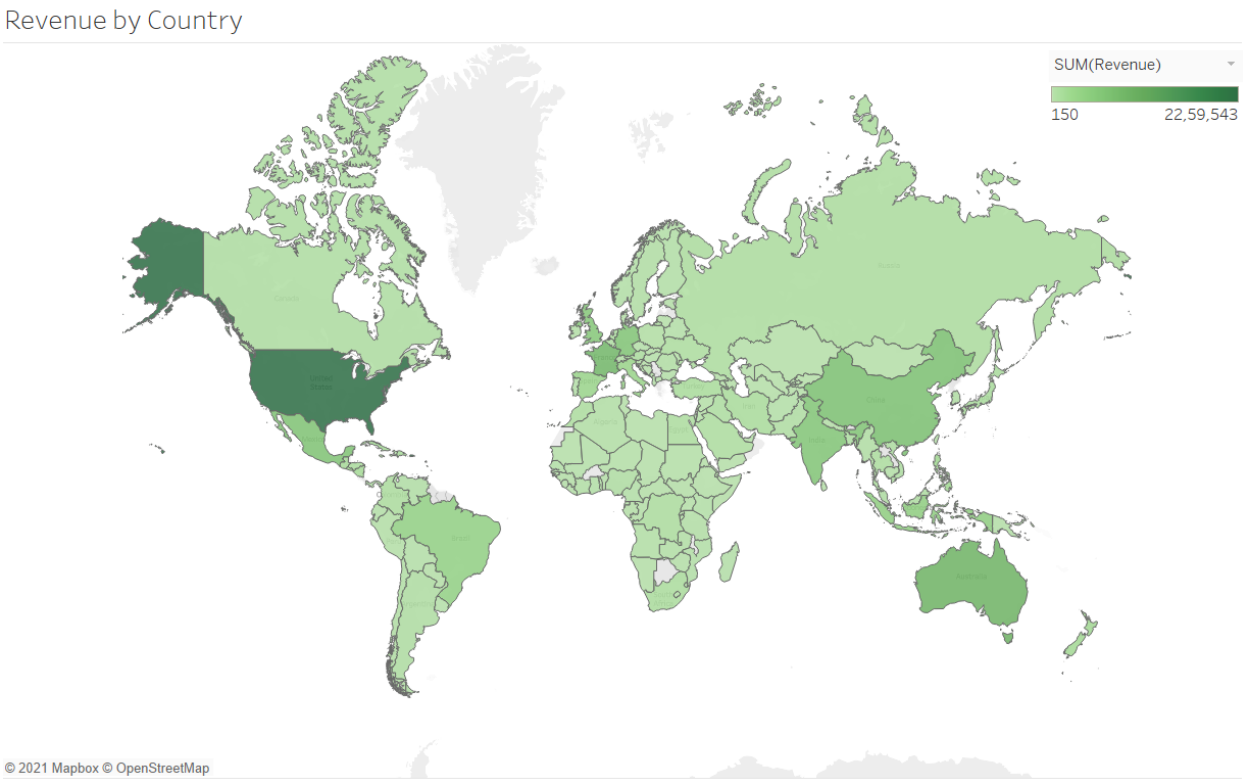
Customer_ID	order_count
0	9
1	6
2	3
3	4
4	4
5	7



Most of the customers have placed less than 20 orders. There are a handful of customers who have placed more than 40 orders.



Most of the customers have generated a revenue of less than 2000 units so far.



USA is the top revenue generating country, followed by China, Australia. and India.

Selecting the Required Variables

Our analysis focuses on customer behaviour. Though Profit is an important KPI, Revenue is more relevant for the analysis, since it directly measures the amount that a customer spends.

Only 3 variables are needed for RFM segmentation: Customer ID, Order Date, and Revenue of the Order. The other variables are dropped. There are total 9,126 unique customers in this data.

Final data used for segmentation:

Customer_ID	Order_Date	Revenue
347	2018-04-01	212
3500	2018-04-01	180
6497	2018-04-01	544
642	2018-04-01	445
4809	2018-04-01	76
4411	2018-04-01	305

RFM Segmentation

Metrics used:

Recency: How long has it been since the customer's last order? This is the number of days passed between the current date and customer's latest order date.

Frequency: How many times has the customer placed an order? i.e. Number of Orders per customer

Monetary: How much revenue has a customer generated till date? i.e. Sum of Revenue per customer

Assumptions

1. Customers belonging to the same segment exhibit homogeneous behaviour.
2. Recency: Lower the recency value, the better. We assume that the customer who has placed an order more recently will likely place another order in near future.
3. Frequency: The higher the better. Customers who purchased more frequently in past will show up again soon.
4. Monetary: The higher the better. The customers who have generated high revenue in the past will continue to do so in future.

Assigning Scores

A score from 1-4 for each Metric is given to a customer.

Recency values are arranged in ascending order, while the values of other two metrics are arranged in descending order. Top 25% most recent customers will be given a 'Recency' score of 4, then customers lying in 2^{nd} quartile will be given a score of 3, and so on.

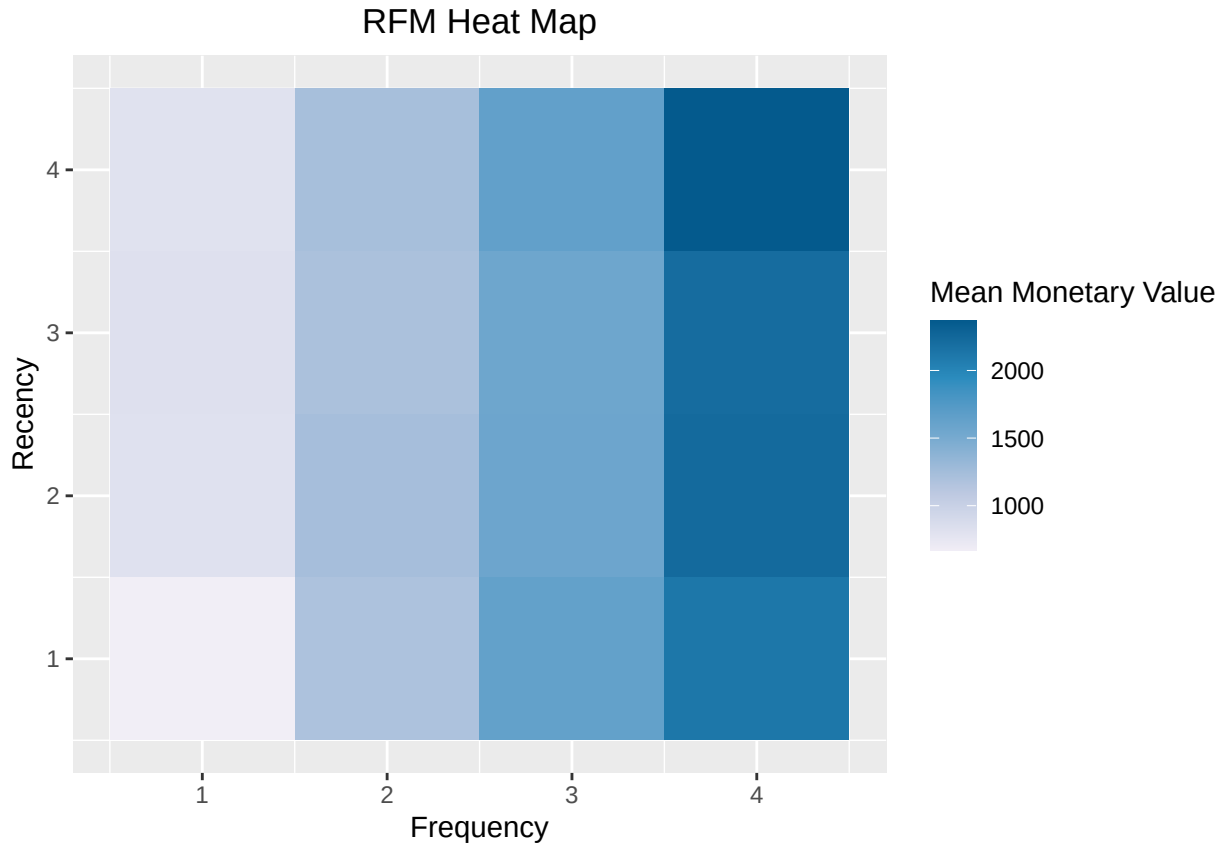
In similar way, top 25% most frequent buyers will be given a 'Frequency' score of 4, and top 25% highest revenue generators will have a 'Monetary' score of 4.

Customer_Quartile	Score
Top 25%	4
2nd Quartile	3
3rd Quartile	2
Bottom 25%	1

After assigning scores to individual metrics, we concatenate them to determine the overall segment membership of that customer.

We have 4 segments per metric. Hence, if we concatenate the scores of all 3 metrics, we have total $4^3 = 64$ segments.

customer_id	recency_days	order_count	revenue	R_score	F_score	M_score	RFM_score
0	553	9	2091	1	4	4	144
1	125	6	993	3	3	2	332
2	531	3	184	1	1	1	111
3	126	4	1687	3	1	3	313
4	67	4	393	4	1	1	411



The highest revenue generating cell is where Recency and Frequency scores are equal to 4. i.e. Segment 444 comprises of Best Customers.

Reducing the number of Segments

We merge some of the RFM segments together for the sake of feasibility and convenience, and name them as follows:

Main_Segment	RFM_segments	Customer_count
At Risk	112, 113, 114, 131, 132, 133, 142, 124, 123, 122, 121, 224, 223, 222, 221	1664
Loyal Customers	334, 342, 343, 344, 433, 434, 443	1481
Potential Loyalist	332, 333, 341, 412, 413, 414, 431, 432, 441, 442, 421, 422, 423, 424	1453
Customer Needing Attention	212, 213, 214, 231, 232, 233, 241, 314, 321, 322, 323, 324	1383
Can't Lose Them	134, 143, 144, 234, 242, 243, 244	797
Lost	111	796
Promising	311, 312, 313, 331	612
About to Sleep	211	352
Best Customers	444	350
Recent Customers	411	237
Hibernating	141	1

Best customers who deserve Gift Hampers

The segment **'Best Customers'** consists of 350 customers who have an **RFM score of 444**. (Score 4 in all 3 metrics)

These are the most loyal customers who buy frequently. They are also top revenue generators.

Therefore, these customers should be awarded with the gift hampers.

Customers that need to be Retained

The segment **'Cant lose them'** comprises of 7 RFM sub-segments: 134, 143, 144, 234, 242, 243, 244

This segment consists of 797 customers who have a **low recency score(1 or 2)**, and **high Frequency and Monetary Scores(3 or 4)**.

The high Frequency and Monetary scores of these customers indicate that they have bought often in the past, and they are top revenue generators too. But their low Recency score indicates that for some reason, they haven't purchased anything from the store since a long time.

A possible reason for this behaviour might be that these customers are exploring our competitor's products. To make sure that we do not lose these customers to our competition, we should try to retain them by providing coupons, offers and other incentives.

Subsets of Customers

The files 'best_customers.csv' and 'retain_customers.csv' contain the list of customer IDs of Best customers and customers that need to be retained respectively.