Lan Nguyen-Presentation

Part1: Introduction

A company in US produces Reference Standards (RS) for pharmaceutical manufacturers. The main customer is in India.
The global leaderships want to understand:

- How RS sales in India have trended since fiscal year 2016.
- Identify potential customer and customer need to pay more attention.

  * A fiscal year start sat July-1 and ends at June-30 next year.

Part2: Method

With a huge pool of customer, to identify and target customer, we use RFM method and build predictive model CLTV (Customer Life Time Value).
I will mainly use python to perform analysis with visualization.

Part3: Analysis

i. EDA:

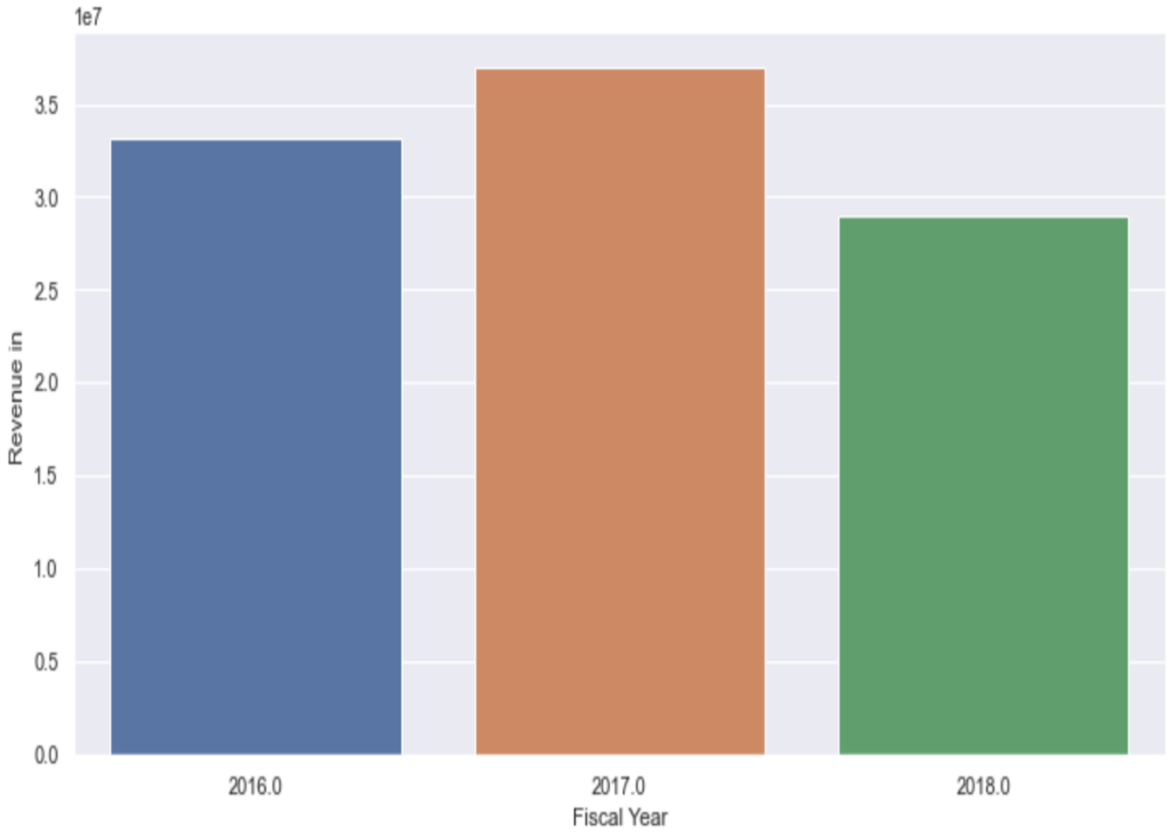
a. An overview of company business stage in the aspect of Sales, Unit Sold, Revenue and Customer acquired over time of period

```
The Sales Period: 2015-07-01  to  2018-03-30
The Total Units Sold: 256676
The Total Revenue: 99182865.47
Total Customer over Period: 319
```

b. Revenue movement over period of time (2016-2018)
There was an increasing from 2016 to 2017
The revenue decrease in 2018 due to some of reasons:

- The data for 2018 was not fully provided (the data only shows until the end of the 3rd quarter of 2018).

- Or we look at the aspect of number of customers, revenue trend to see the actual problem from existing dataset that cause the downtrend.

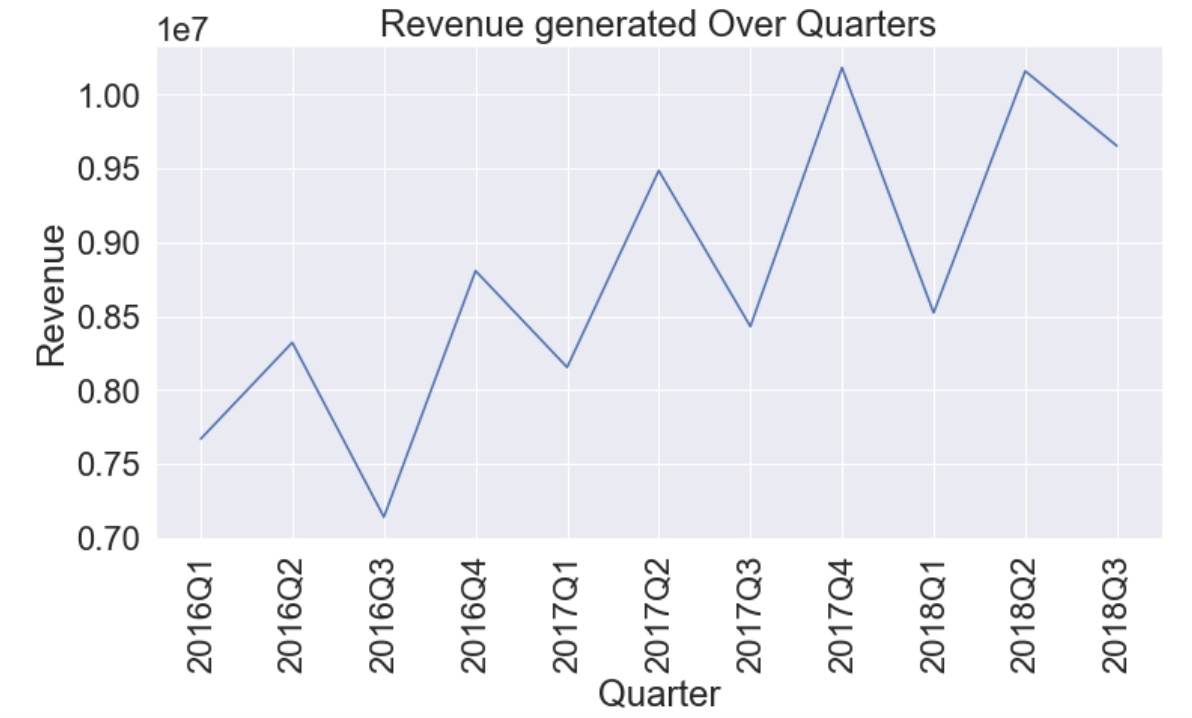


c. Revenue generated by quarter:

We look at the aspect of quarter so we can easily compare the revenue:

- Among all the 3rd Quarter of each year, the 2018Q3 is the highest one.

- Also the the revenue of the 4th quarter increased over year. Thus, there is possibility that the the downtrend in revenue of fiscal year 2018 may cause by the lack of data.

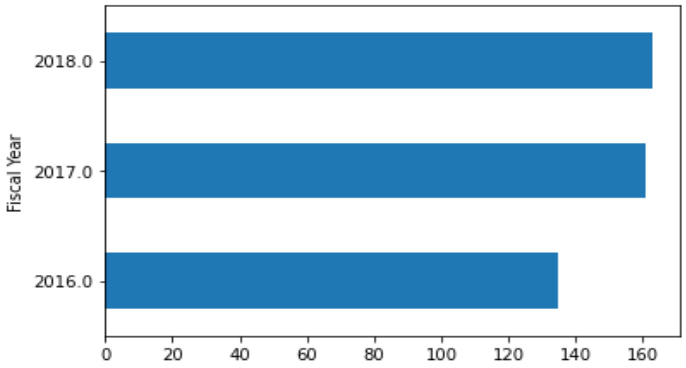


d. Revenue generated in each year:

There were not a huge different in revenue for each month of each years. However, the 3rd quarter of 2018 seems remain un changed.

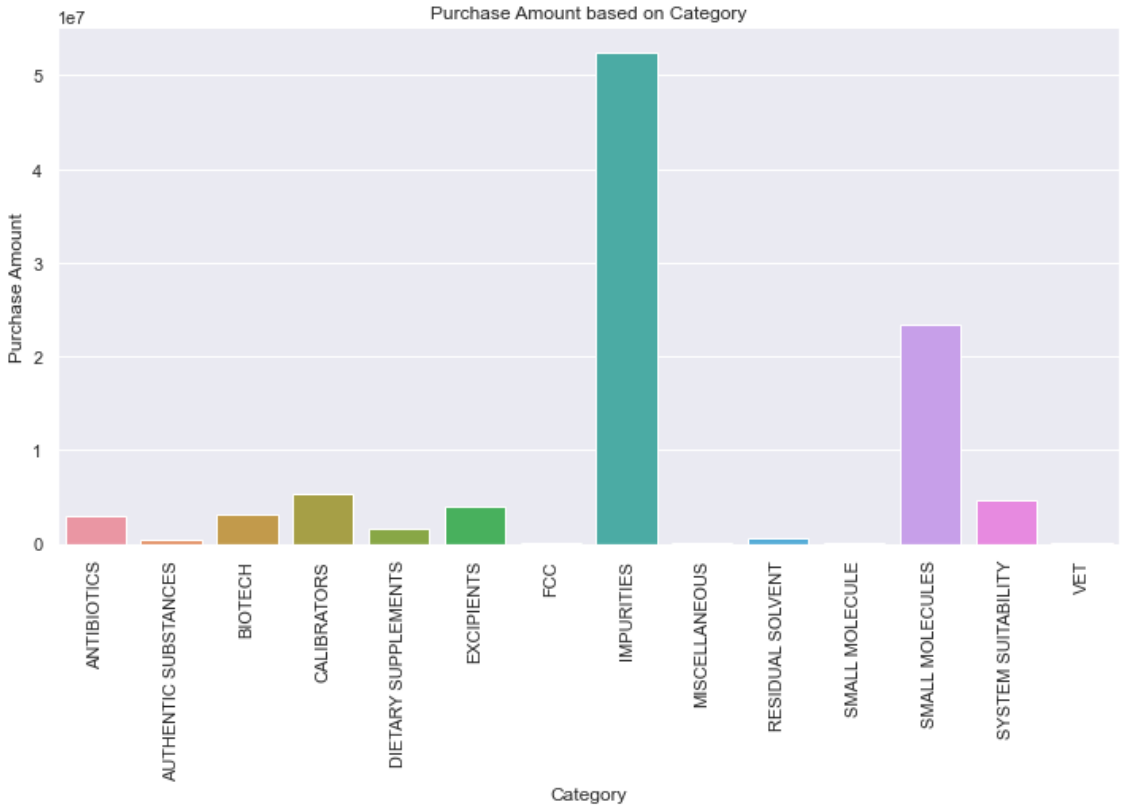
e. Customer acquired over time:

- Company had a big obtaining new customer from between 2016 and 2017.
- However, from 2017 until the end of 3rd quarter of 2018, there was only less than 5 new customers.



f. Revenue generated by RS categories:

“Impurities” and “Small Molecules” are the most selling categories

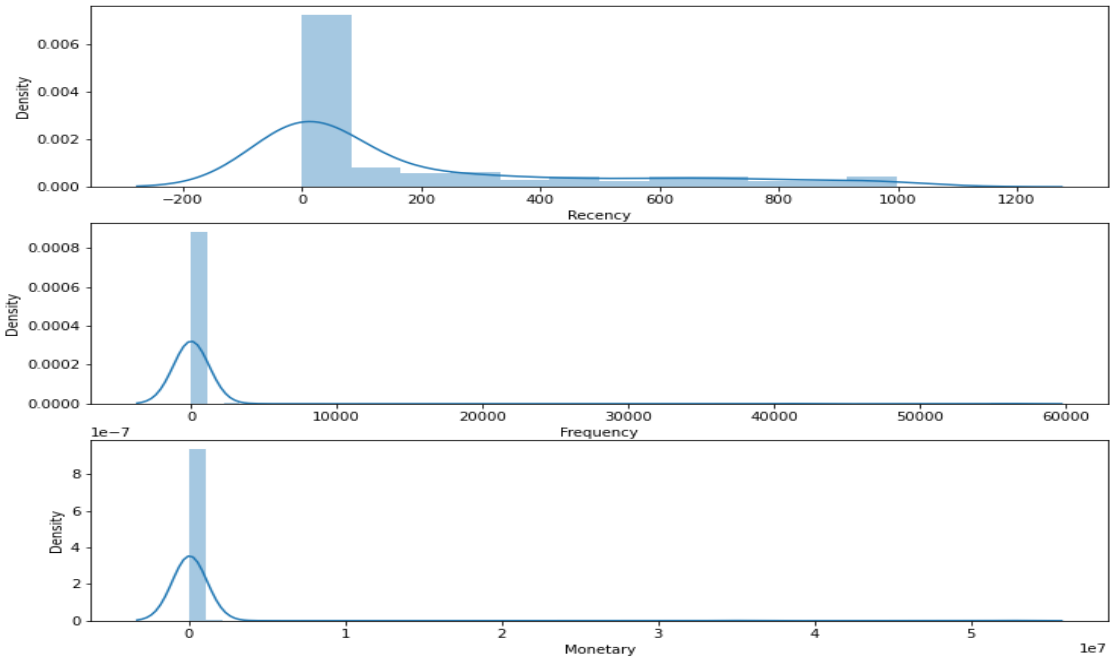


Category	Frequency	Unit Sold	Purchase_Amount	Average
ANTIBIOTICS	6038	13468	2995280	222
AUTHENTIC SUBSTANCES	1004	1880	486052	259
BIOTECH	1426	8130	3067397	377
CALIBRATORS	3296	24726	5321915	215
DIETARY SUPPLEMENTS	3352	6044	1685619	279
EXCIPIENTS	9744	19887	4037998	203
FCC	257	537	104756	195
IMPURITIES	42753	83596	52460242	628
MISCELLANEOUS	383	612	126773	207
RESIDUAL SOLVENT	1199	3246	643402	198
SMALL MOLECULE	3	4	821	205
SMALL MOLECULES	35587	85106	23382082	275
SYSTEM SUITABILITY	3827	8750	4712999	539
VET	424	690	157531	228

ii. RFM Analysis:

We will look at the Recency (Duration doing business), Frequency (How often doing business), Monetary(Spending power), to identify and label customer.

The lower recency, highest frequency and monetary amount are our best customer. However, that should be good if in short period of time. In this dataset we look at the aspect of the most recent customer with high frequency and high monetary value.



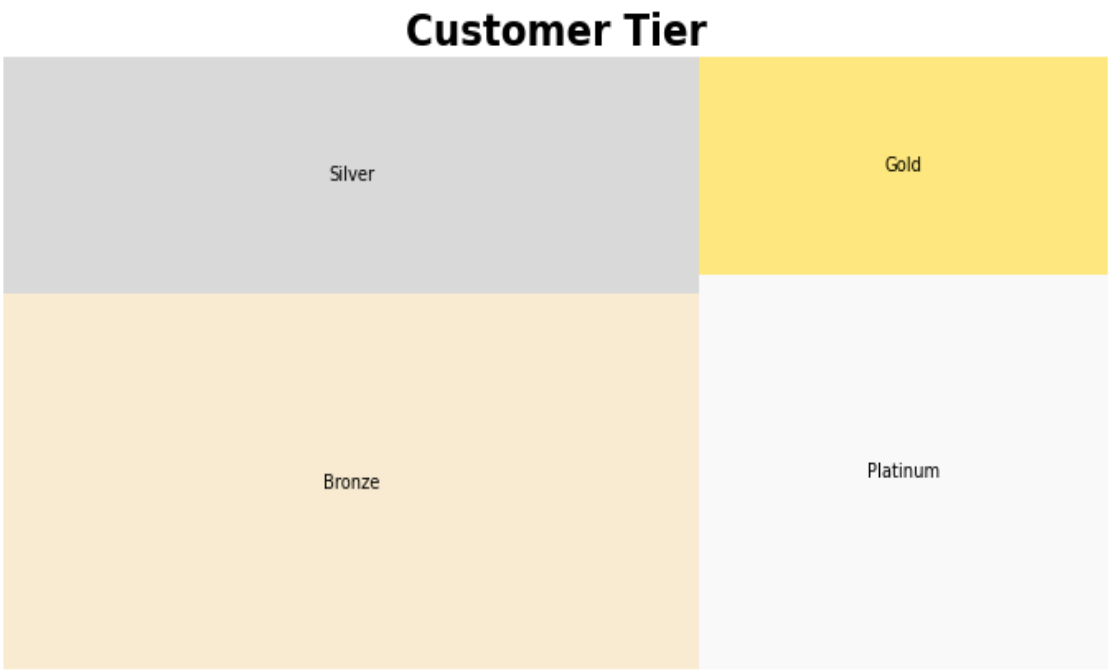
The level will range from 1 to 4 which represent the highest rate to lower west of RFM as the following example:

Customer	Recency	Frequency	Monetary	R	F	M	RFMSegment
A.O. Smith Corp	661	56030	52486605	1	1	1	111
ABIOMED Inc	942	491	367740	1	1	1	111
AES Corp	786	270	186809	1	1	1	111
AFLAC Inc	945	194	132754	1	1	1	111
AMETEK Inc.	764	55	45102	1	1	1	111

Labeling customer base of the total score of RFM (Platinum, Gold, Silver, Bronze)

Customer	Recency	Frequency	Monetary	R	F	M	RFMSegment	RFMScore	Customer_Tier
A.O. Smith Corp	661	56030	52486605	1	1	1	111	3	Platinum
ABIOMED Inc	942	491	367740	1	1	1	111	3	Platinum
AES Corp	786	270	186809	1	1	1	111	3	Platinum
AFLAC Inc	945	194	132754	1	1	1	111	3	Platinum
AMETEK Inc.	764	55	45102	1	1	1	111	3	Platinum
...	...	...	...	...	...	...	...	...	...
Newmont Corporation	0	1	310	4	4	4	444	12	Bronze
News Corp. Class A	0	2	2023	4	3	2	432	9	Silver
News Corp. Class B	0	1	1595	4	4	3	443	11	Bronze
NextEra Energy	0	1	729	4	4	3	443	11	Bronze
eBay Inc.	0	1	315	4	4	4	444	12	Bronze

There are more than 60% customer that need to pay more attention, the portion of platinum customer is likely 25%



iii. CLTV Analysis:

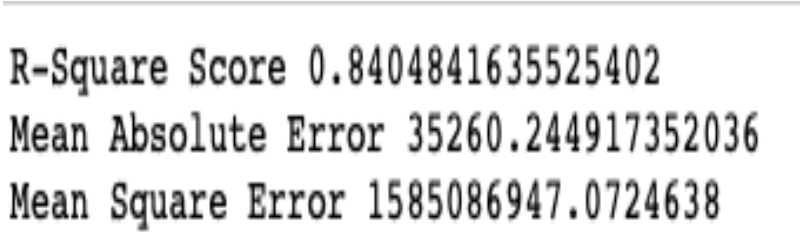
$$CLTV = \frac{AverageOrderValue * PurchaseFrequency}{ChurnRate} * ProfitMargin$$

*Even the dataset did not provide the profit margin, I assumed it as 40% in order to calculate the CLTV

By using that formula, we can calculate ethe Customer Lifetime Value as the following figure:

Customer	Recency	Frequency	Monetary	AvgOrderValue	ProfitMargin	CLTV
A.O. Smith Corp	284	28141	26240017	932	10496007	4343244285705
Alexandria Real Estate Equities	220	6792	5848153	861	2339261	893851339336
Alphabet Inc. (Class C)	330	518	726273	1402	290509	180757413810
Dish Network	8	3	50064	16688	20026	148303377509
AmerisourceBergen Corp	198	102	260012	2549	104005	117655025991
Advance Auto Parts	334	381	488645	1283	195458	111246789208
Advanced Micro Devices Inc	341	1105	766728	694	306691	94437909415
Alaska Air Group Inc	302	90	217753	2419	87101	93521470529
Broadridge Financial Solutions	269	35	109565	3130	43826	60883442747
Alliant Energy Corp	354	158	216504	1370	86602	52662358417

Using Linear Regression Model to predict CLTV of customer base on data of 2017. I pick the last 5 months to perform the process.
In order to perform the reliability of the model, I performed metrics to evaluate the process as the following figure:



Part4: Conclusion:

By using RFM analysis, it helps company to identify :

- The top tier customers to maintain relationship.
- Potential customer that we can pay more attention to promote to higher level (from bronze to gold, or gold to Platinum)
- Or we can stop maintaining with bronze customer to save the maintaining cost.

Also, we have the ability to perform the predictive model by using linear regression, it also helps us to predict in the long run to have specific plans to increase the revenue.