

Perqara Data Scientist

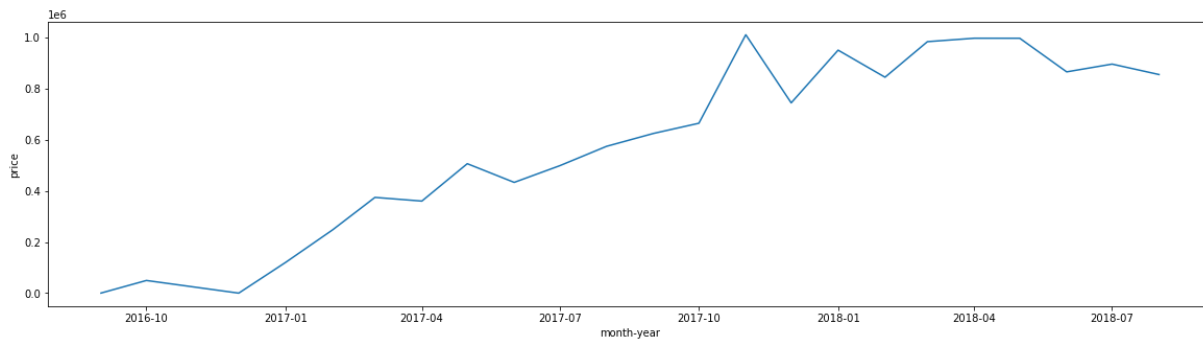
Background

This dataset contains data on E-commerce sales from 2016-2017, details of products sold, customer geography and payment information.

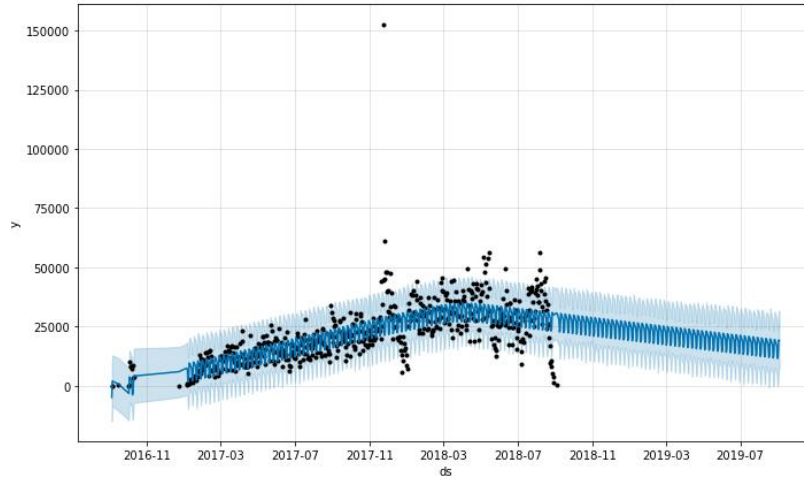
In Year on Year (YoY) growth 2017-2018 there was a decrease to YoY growth 2016-2017. The decline came to \$4.875.777. Therefore the company wants to increase sales by making marketing strategies more effective and efficient. As a data scientist team, I will make customer segmentation based on recency, frequency and monetary to help the marketing team

I am working as data scientist for e-commerce brazil company and want to give business recommendation and segment our customers to help the effectivity and efficiency marketing. Therefore we will save our time and money by marketing a spesific group of customers.

Exploratory Data Analysis



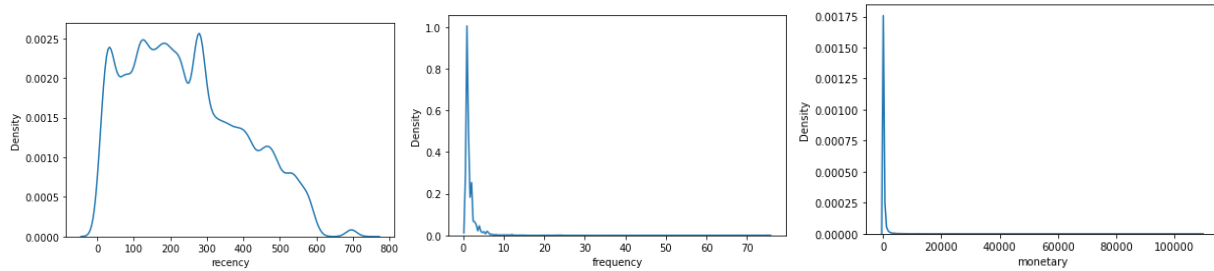
Sales data start from 2016 to 2018. From October 2016 to April 2018 sales usually increasing month to month but in the beginning May 2018 to August 2018 sales slightly deacresing month to month.



Forecast use fbprophet with daily sales from 2016 to 2018 showed us 2019 sales will decrease. Therefore to prevent decrease need action and preparation. One of the ways is increasing sales with effective and efficient marketing.

Recency, frequency, monetary value is marketing analysis tool used to identify customer shopping behavior by measuring and analyzing spending habits. Recency showed us how recently a customer has made purchase. Frequency, how often a customer make purchase. Monetary, how much money a customer spends on purchases.

	recency	frequency	monetary
count	95420	95420	95420
mean	242.6003773	1.232487948	212.8289112
std	153.1603197	0.823079148	641.7600887
min	0	1	0
25%	118	1	63.9075
50%	223	1	113.15
75%	352	1	202.5425
max	728	75	109312.64



Total of the data to customer segmentation is 95.420 customer. Based on table and kdeplot above, frequency and monetary have extreme skewness and outlier. 75% of customer buy the product just one time. To mitigate this problem we can categorize RFM to three dan two category. This category based on quantile distribution of RFM.

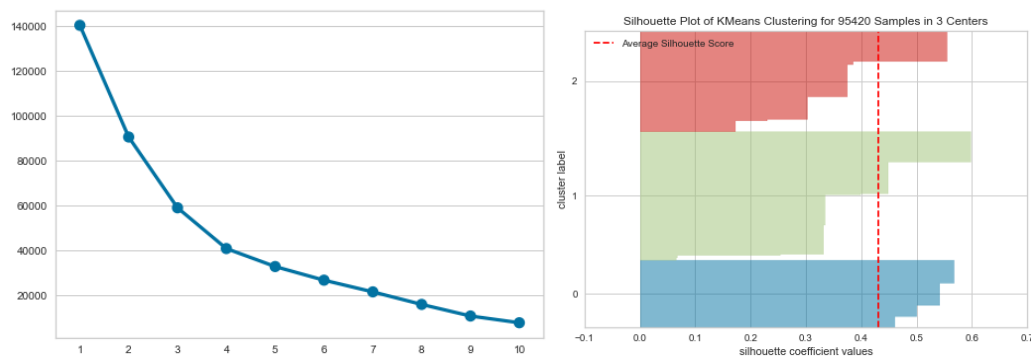
recency_score	recency			recency_score	recency		
	min	max	count		min	max	count
1	292	728	32351	1	292	728	32351
2	153	291	31405	2	153	291	31405
3	0	152	31664	3	0	152	31664

frequency_score	frequency		
	min	max	count
1	1	1	81088
2	2	75	14332

Monetary and recency separated by three category. Frequency separated by two category because value one in frequency very dominant.

Result

Clustering for customer segmentation using Kmeans with elbow method to optimize the number of the cluster and silhouette score for evaluation.



Result of Kmeans modelling, optimal number of clusters is three. There cluster have near equal of the amount of customer.

cluster	recency			frequency			monetary			
	mean	min	max	mean	min	max	mean	min	max	count
0	150.4706	0	291	1.535643	1	24	459.4772	161.38	44048	21673
1	151.3364	4	291	1.069089	1	22	82.20006	9.59	161.37	41396
2	421.1016	292	728	1.238478	1	75	214.7424	0	109312.6	32351

Kmeans modelling produce three category of customer :

1. Cluster : 0

Category : Best Customer

RFM Characteristics: Recent and frequent customers who heavy spending money to bought product in e-commerce

Action on RFM characteristic : Offering new products to customers and providing a referral code to implement word of mouth marketing

2. Cluster : 1

Category : Average Customer

RFM Characteristics: Recent customers who buy products infrequently and less spending money to bought product in e-commerce

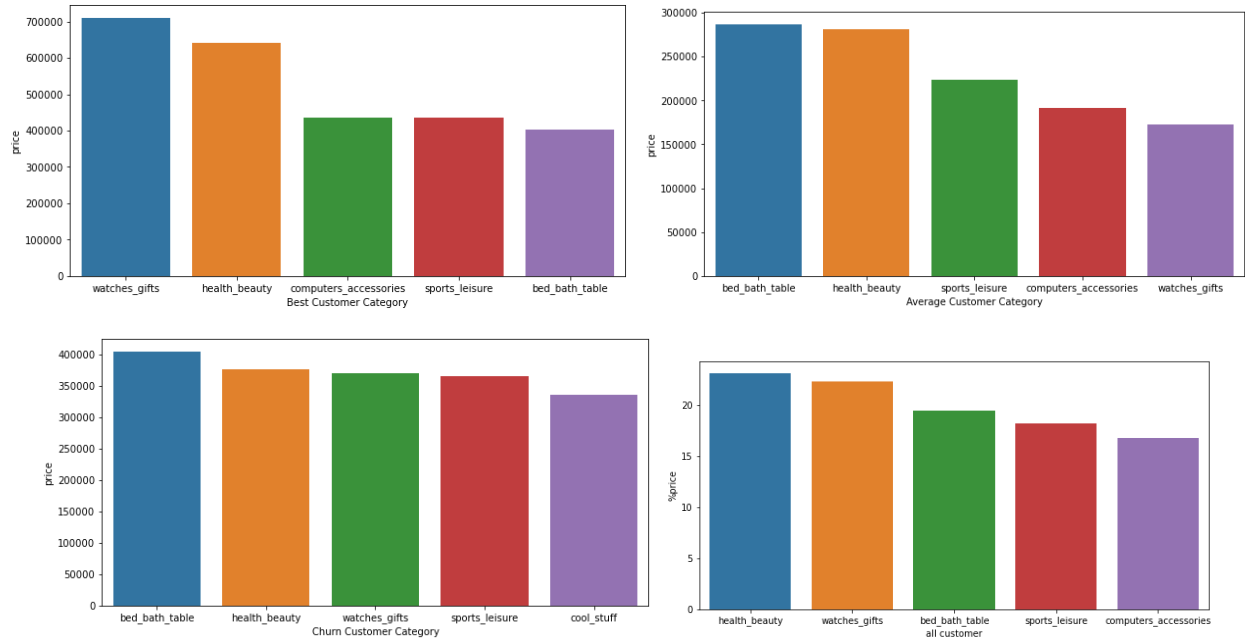
Action on RFM characteristic : Develop Customer Relationship Management to improve shopping experience

3. Cluster : 2

Category : Risk to Churn Customer

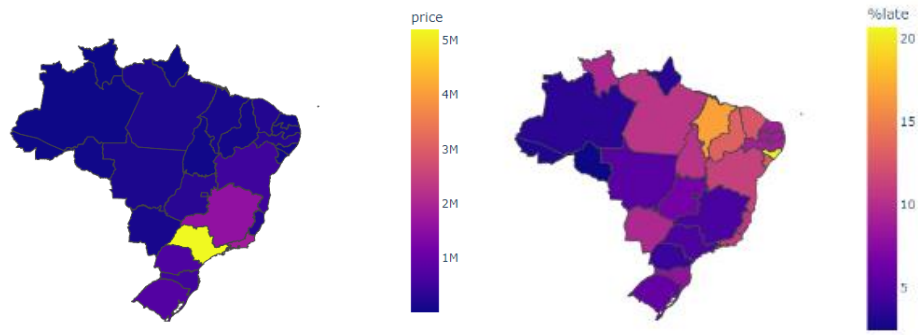
RFM Characteristics: Customers who frequently and average spending money bought products in the past but haven't transacted in e-commerce for a long time

Action based on RFM characteristic : Research why they don't transact again and create marketing strategies to encourage shop again



Business recommendation based on category product sales:

1. Encourage customers to buy products with the top 5 categories
2. Offering new furniture décor products to best customer
3. Give recommendation product of watches gift to churn and average customer with attractive offers



Business recommendation based on geoanalysis:

1. Mitigate delayed delivery in north brazil to increase concern in north brazil
2. Develop marketing strategies for north brazil region to increasing attention and sales

cluster	credit_card	boleto	debit_card
0	\$ 5,352,089.79	\$ 1,011,304.19	\$ 113,305.65
1	\$ 2,141,581.81	\$ 518,626.86	\$ 49,283.13
2	\$ 3,935,466.37	\$ 899,044.14	\$ 27,238.58

Business recommendation based on payment :

1. marketing team can develop marketing strategies for credit payments

Summary

Customer segmentation used Kmeans modeling with RFM feature. RFM chosen because it's easily interpret and make business recommendation. Business recommendation given with category product, geography and payment method consideration. This business recommendation have objective to increasing effective and efficient marketing strategies.