E-commerce Customer Behavior Data Analysis

- ## Dataset overview

This data set contains one-year trade information for all online orders placed on an e-commerce online trade website. The owner of this website focuses on gift sales and has many clients acting like wholesalers at the same time.

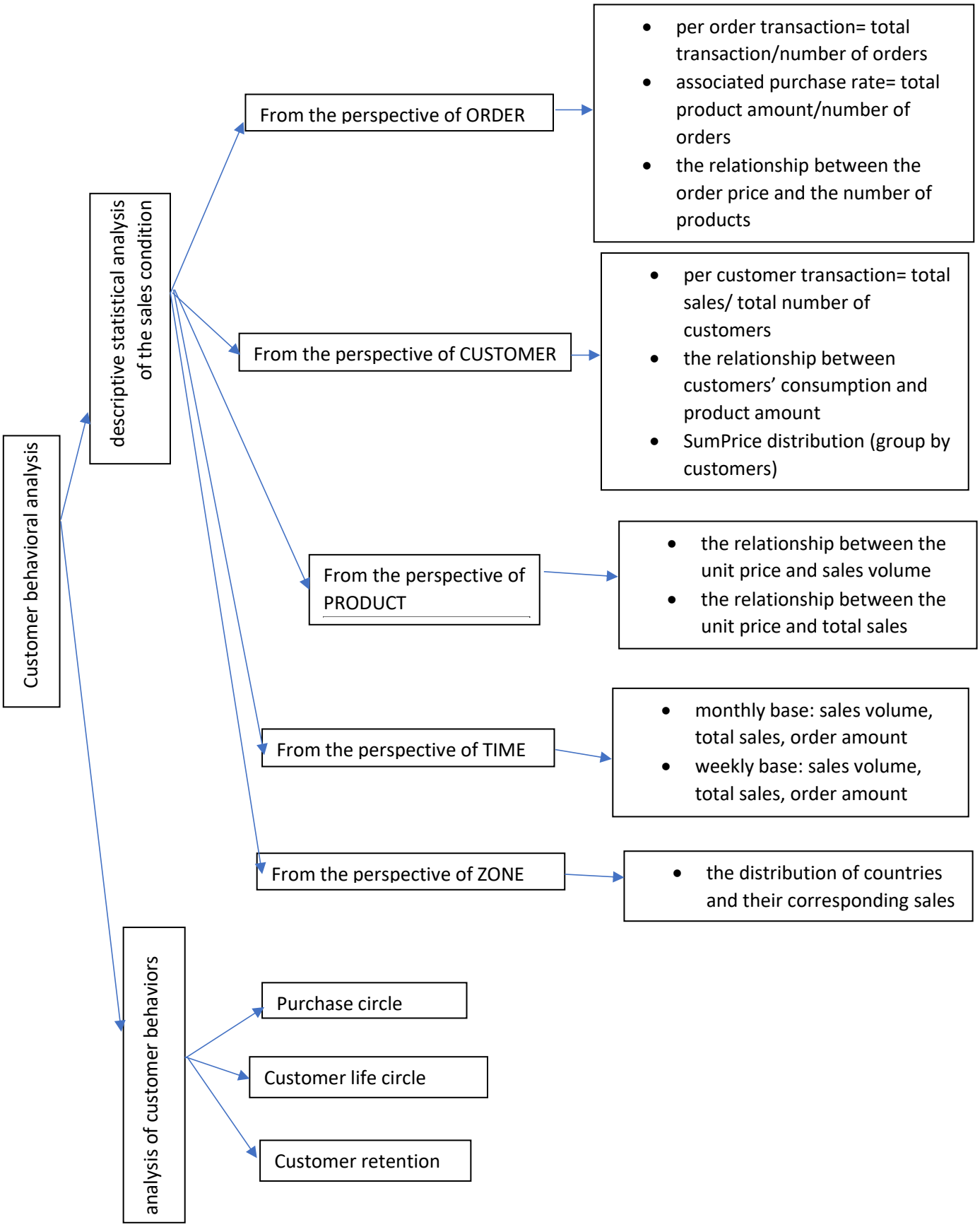
The spreadsheet contains 8 fields and 50000+records. The fields include InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country.

InvoiceNo: The serial number of invoices. If begins with C, it means this order was canceled.

StockCode: The product code. Each product has a unique code.

CustomerID: Every customer has a unique 5 digits code.

- **Ideas for this analysis**



Conclusion:

Descriptive statistical analysis

From the perspective of ORDER:

- During the 1 yr statistical period, we have 19960 effective orders, per order transaction was £533.17, associated purchase rate was around 279. It indicates that the dominating business of this e-commerce website was wholesale.
- The overall difference among all orders were apparent, and customers with super-strong purchasing power existed.

From the perspective of CUSTOMER：

- Over 25% of customers only placed their order once and didn't stay retained.
- Most of the customers didn't have strong purchasing power.

From the perspective of PRODUCT:

- Price of one specific product could be various in different months
- The value proposition of this e-commerce website was focusing on the wholesale of small commodities with low unit prices.
- Products with a price lower than £5 got the customer's preference.
- Products with a low unit price also contributed to high sales. This kind of commodities became the leading source of total sales. Although some goods had a high unit price, the high unit price didn't result in high sales.
- We suggest the purchasing department select more low-price products to enlarge product types in low price area.

From the perspective of TIME:

- The sales conditions remained stable relatively from January to August. And then kept growing between September and November. Consider the dominating business of this e-commerce platform is selling gifts, the sales would be affected obviously by holidays, such as Halloween, Thanks Giving Day, Black Friday, and Boxing day.
- On 9th, Dec, there was an apparent decrease in order quantity, however, we got the highest sales in this sampling interval, which indicates there should be some orders with large goods quantity on that day.
- It would be better if the e-commerce platform to pay more attention to high-value customers by taking actions like setting fixed customer service staff to these clients.

From the perspective of ZONE:

- most of the customers were from the UK, and the primary source of overseas income was mostly from the neighboring countries of the United Kingdom. This phenomenon might relate to logistic costs and language factors, or it might because the influence of this e-commerce platform was attenuated gradually by distance.
- We may try to increase the overseas popularity by launching more advertisements. At the same time, add more language choices for the website. And also, provide more transparent and convenient solutions to decrease the overseas logistics costs and optimize logistic procedures.

Customer behavior analysis

Customer life circle:

- Since we only have data for the one-year period, we don't know their buying actions before/after this period. Real-life cycles for some of them would be longer, and this brought limitations to our analysis.
- The date of initial consumptions occurred frequently at the beginning of the statistical period, and the date of last consumptions occurred mostly at the end of the statistical period. That means the actual life cycle of a large number of users must be longer.
- The minimum and Q1 values were both 0, thus more than 25% of clients only consumed once. The average life cycle was 130 days, the median was 93 days, which means some loyal customers made the mean value larger.
- This website needs to focus more on the improvement of customer's initial purchase experience. We may build an online mark system and do some phone investigations to get feedback. Also, activities aim at attracting second purchases should be considered, for instance: offering coupons with time limitations.
- The average life cycle for customers with more than twice purchases was 203 days, far higher than the total mean life cycle length of 103 days.

Customer retention condition:

- The 'customer life cycle' is the time difference between the first and last purchase. However, the 'customer retention period' focus on customers' periodic behaviors. Due to the dataset incompleteness we mentioned before, the outcome had limitations here as well.
- Customers came back in

  0~3 days:3.2%
  4~7 days:6.6%
  the second month:37.4%
  the third month:40.5%
  90~180 days:67%

- The high loyalty customers might not consumption frequently but their user-stickiness was pretty good.

Purchase period:

- The distribution for purchase period was a right-skewed distribution with the peak value of around 15~70 days, we suggest this company send promotion information to customers monthly.

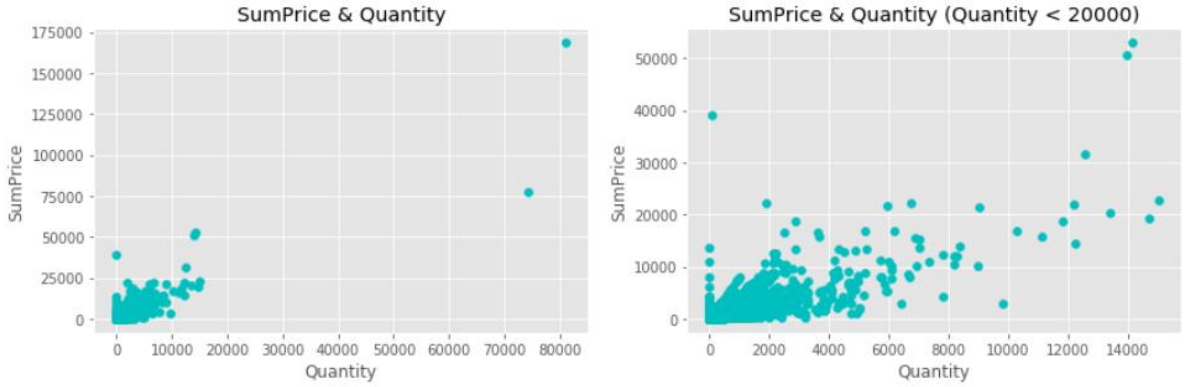
Appendix

part of the intermediate process:

From the perspective of ORDER:

	Quantity	SumPrice
count	19960.000000	19960.000000
mean	279.178758	533.170098
std	955.011964	1780.412701
min	1.000000	0.380000
25%	69.000000	151.695000
50%	150.000000	303.300000
75%	296.000000	493.462500
max	80995.000000	168469.600000

During the 1 yr statistical period, we have 19960 effective orders, per order transaction was £533.17, associated purchase rate was around 279. It indicates that the dominating business of this e-commerce website was wholesale.
The overall difference among all orders were apparent, and customers with super-strong purchasing power existed.



In general,the variable SumPrice has a positive correlation with Quantity. However,there are some outliers(high sumprice and low quantity). This is a point for further exploration.

From the perspective of CUSTOMER：

	InvoiceNo	Quantity	SumPrice
count	4338.000000	4338.000000	4338.000000
mean	4.272015	1187.641770	2048.679865
std	7.697998	5043.619358	8985.227179
min	1.000000	1.000000	3.750000
25%	1.000000	159.000000	306.482500
50%	2.000000	378.000000	668.570000
75%	5.000000	989.750000	1660.597500
max	209.000000	196915.000000	280206.020000

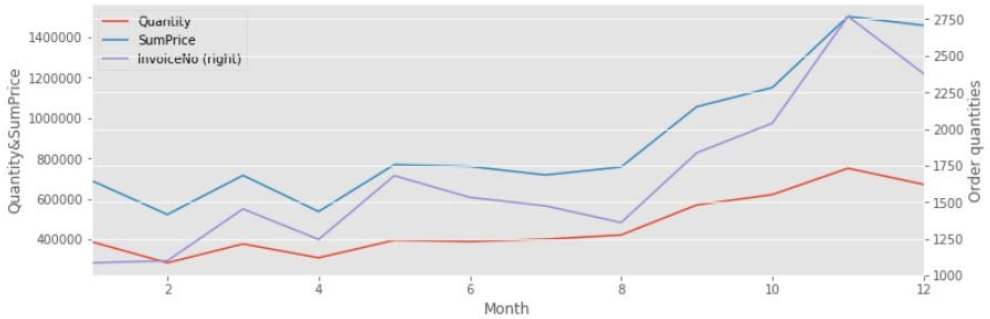
The mean value for per customer order amount was 4; the median was 2; also, over 25% of customers only placed their order once and didn't retain.
The per customer order amount was 1187, which exceeded the Q3 value. The customer who purchased most bought 196915 products in total.
The per customer transaction was £2049, which exceeded the Q3 value as well.
In short, noticeable purchasing power differences occurred among all customers. Customers with high expenditure made the mean value high.

From the perspective of PRODUCT:

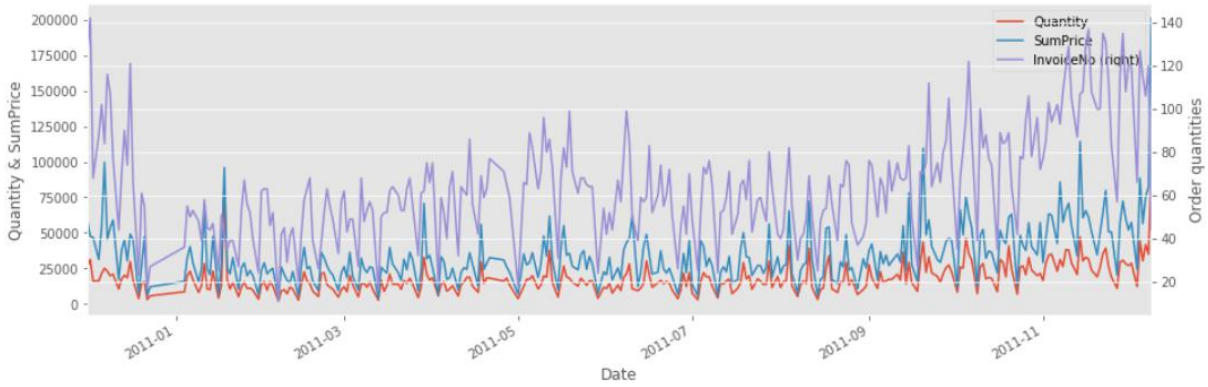
The peak value of products unit price was £1-2. A few products had their unit price higher than £10. Thus, the value proposition of this e-commerce website was focusing on the wholesale of small commodities with low unit prices.

Products with a low unit price also contributed to high SumPrice. This kind of commodities became the leading source of total sales. Although some goods had a high unit price, the high unit price didn't result in high sales. Here we suggest the purchasing department select more low price products to enlarge product types in low price area.

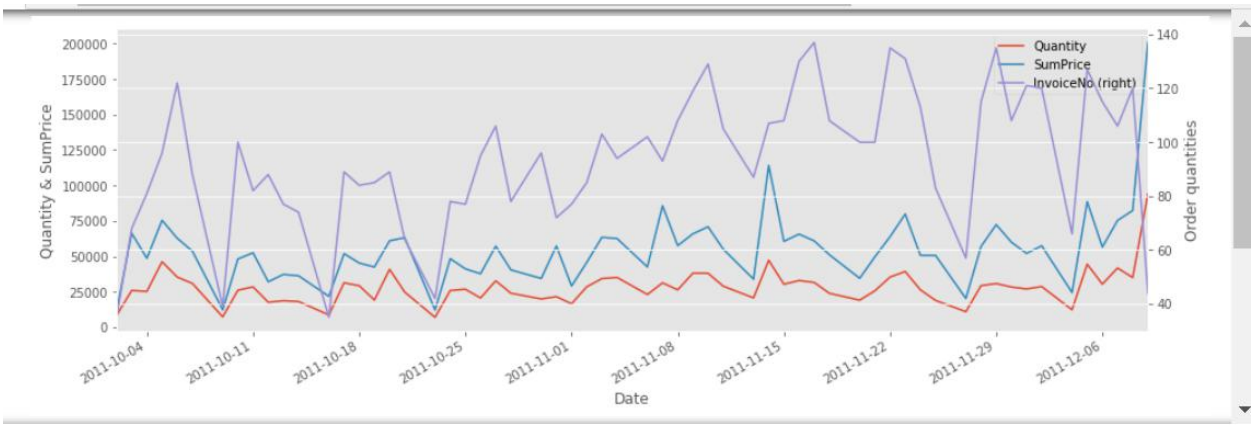
From the perspective of TIME:



We drew the plot with double axes on a monthly basis to show the wholesales, sales volume, and order amount per month. One point to mention here is that we only had data for 9 days in December and that's why we saw a 'decrease' in December which against the general trend. All three plots had similar trends. The sales conditions remained stable relatively from January to August. And then kept growing between September and November. Consider the dominating business of this e-commerce platform was selling gifts, the sales would be affected obviously by holidays, such as Halloween, Thanks Giving Day, Black Friday and Boxing day.



We drew the plot with double axes daily to show the sales, sales volume, and order amount per day. As we analyzed before, Quantity and SumPrice show similar trends. Here we saw a surge in sales in the last month. We drew a graph of this period separately.



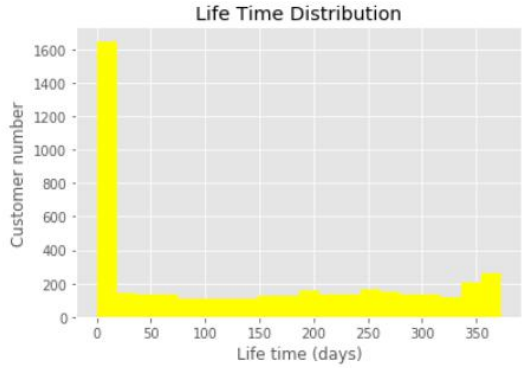
On 9th, Dec, there was an apparent decrease in order quantity, however, we got the highest sales in this sampling interval, that indicates there should be some orders with large goods quantity on that day. After browsing the sales data, a customer from the UK bought 80000+ papercraft on 9th, Dec, caused the sales surge. It would be better if the e-commerce platform to pay more attention to this high-value customer by taking actions like setting fixed customer service staff to this client.

From the perspective of ZONE:

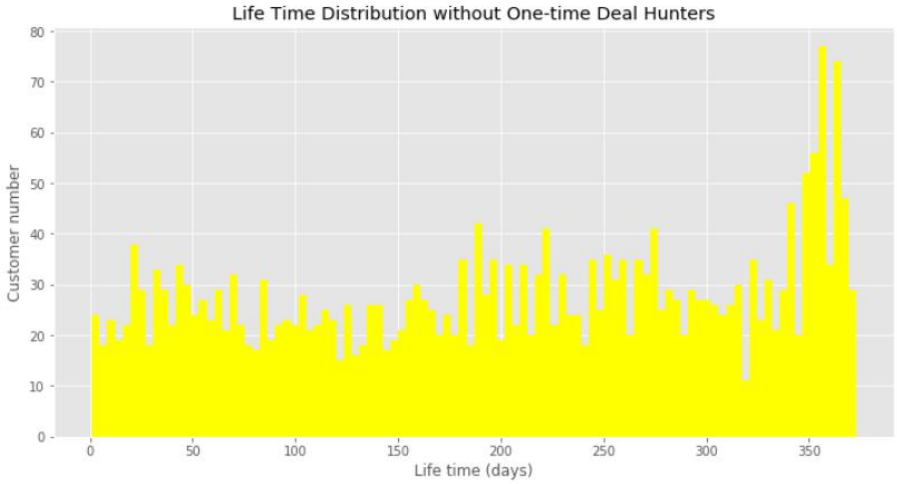
Country	SumPrice	CustomerID	AvgAmount
United Kingdom	7284989.004	3920	1858.415562
Netherlands	285446.340	9	31716.260000
EIRE	265262.460	3	88420.820000
Germany	228678.400	94	2432.748936
France	208934.310	87	2401.543793
Australia	139843.950	9	15538.216667
Spain	66470.260	30	2215.675333
Switzerland	57222.850	21	2724.897619
Belgium	47971.210	25	1918.848400
Sweden	38367.830	8	4795.978750
Japan	37416.370	8	4677.046250
Norway	36165.440	10	3616.544000
Portugal	33375.840	19	1756.623158
Finland	22546.080	12	1878.840000
Singapore	21279.290	1	21279.290000
Channel Islands	20440.540	9	2271.171111
Denmark	19774.400	9	2197.155556
Italy	17483.240	14	1248.802857
Austria	16775.840	11	1525.076364
Cyprus	16518.540	8	2064.817500
Poland	7334.650	6	1222.441667
Israel	7215.840	3	2405.280000
Greece	4760.520	4	1190.130000
Iceland	4310.000	1	4310.000000
Canada	3666.380	4	916.595000
USA	3580.390	4	895.097500
Malta	2725.590	2	1362.795000
Unspecified	2660.770	4	665.192500
United Arab Emirates	1902.280	2	951.140000
Lebanon	1693.880	1	1693.880000
Lithuania	1661.060	1	1661.060000
European Community	1300.250	1	1300.250000
Brazil	1143.600	1	1143.600000
RSA	1002.310	1	1002.310000
Czech Republic	826.740	1	826.740000
Bahrain	548.400	2	274.200000
Saudi Arabia	145.920	1	145.920000

It seems that most of the customers were still from the the UK and the primary source of overseas income was mostly from the neighboring countries of the United Kingdom. This phenomenon might relate to logistic costs and language factors, or it might because of the influence of this e-commerce platform was attenuated gradually by distance. We may try to increase the overseas popularity by launching more advertisements. At the same time, add more language choices for the website. And also, provide more transparent and convenient solutions for decreasing the overseas logistics costs and optimizing logistic procedures.

Customer life circle:

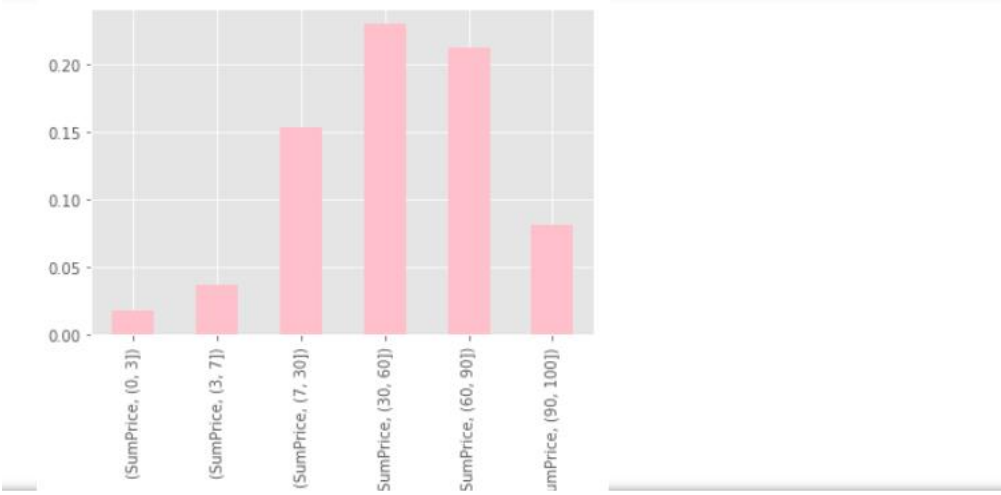


Many customers only consumed once and then left, this website needs to focus more on improving customer's initial purchase experience.
We may build an online mark system and do some phone investigations to get feedback. Also, activities aim at attracting second purchase should be considered, for instance: offering coupons with time limitation.



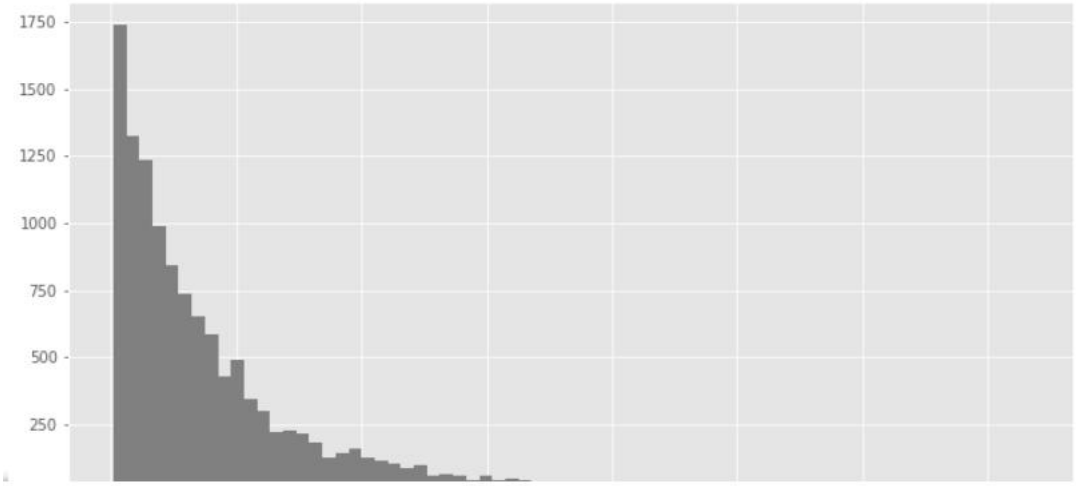
We filtered the customer with life cycle 0 and plot the remaining data again, about 25% of clients had a life cycle length between 170~330 days, this group had a high-quality life cycle. And customers with life cycle longer than 330-day can be viewed as having a good user sickness, we had a lot of cases located in this period, which was a good sign.

Customer retention condition:

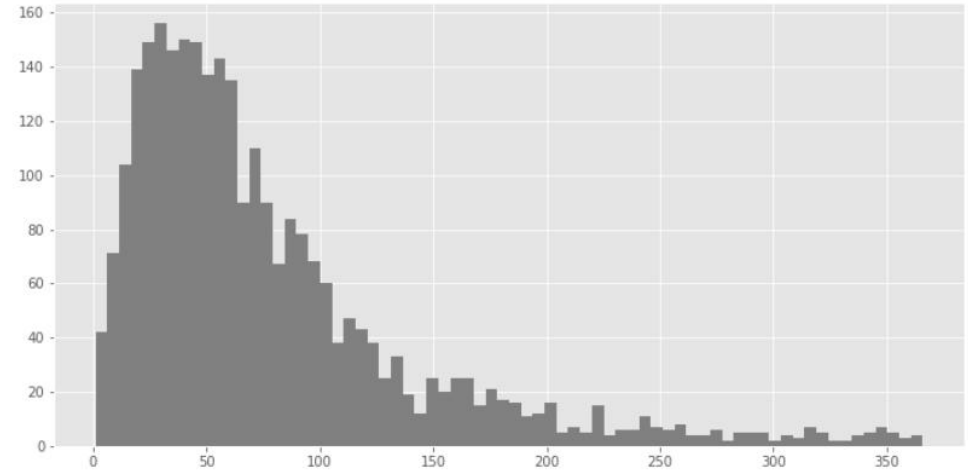


Customers came back in:
0~3 days:3.2%
4~7 days:6.6%
the second month:37.4%
the third month:40.5%
90~180 days:67%
The high loyalty customers might not consumption frequently but their user stickiness was pretty good.

Purchase period:



The purchase period for all orders showed a long-tail distribution.
most intervals were not long



Group the data by 'CustomerID' again. The plot was a right-skewed distribution with the peak value around 15~70 days, indicating most customers concentrated there. We suggest this company send promotion information to customers monthly.