E-commerce Behavioral Data Analysis

- ## data set overview

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

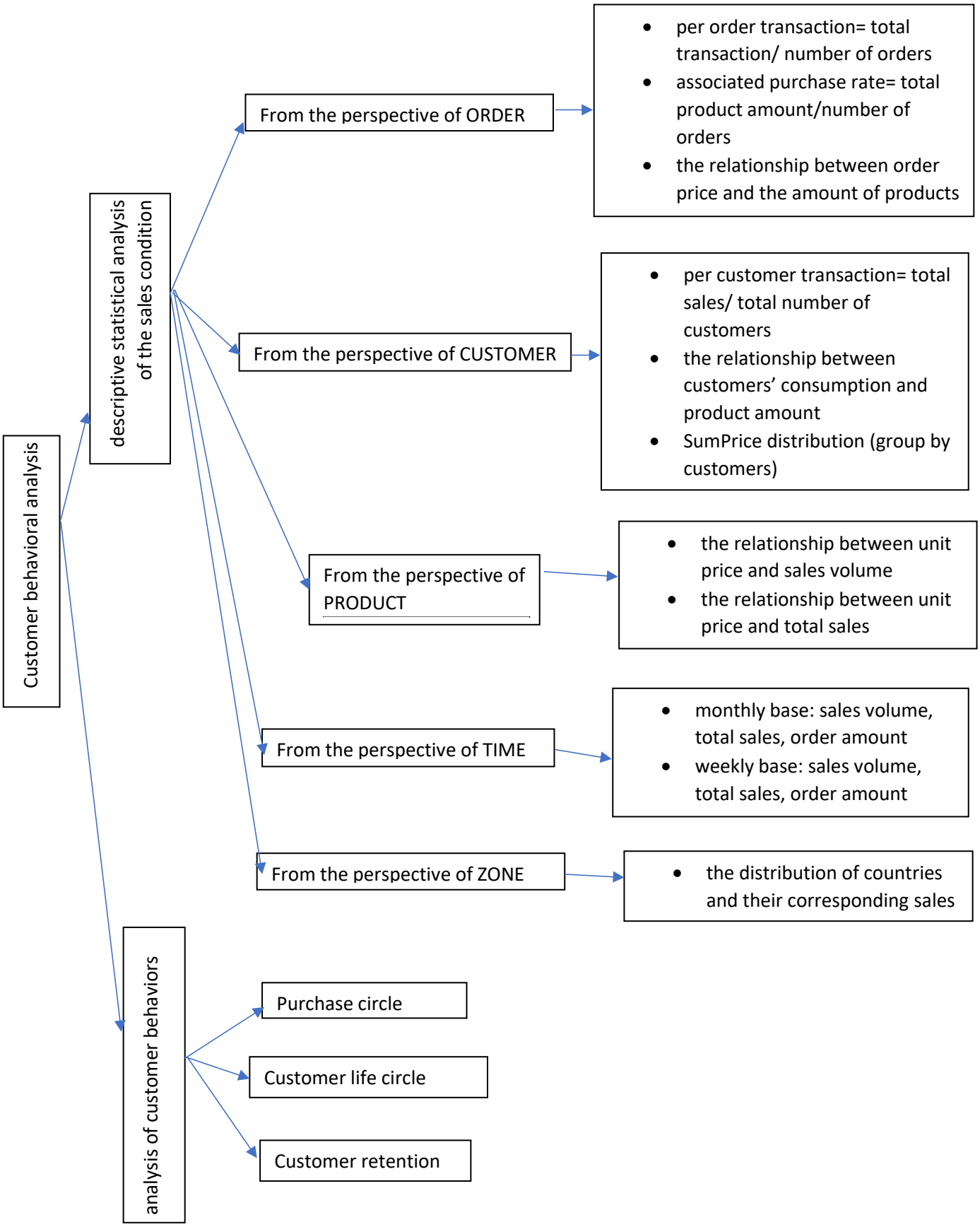
The form of our raw data is a xlsx file. The spreadsheet contains 8 fields and 50000+records. The fields include: InvoiceNo, StockCode, Description, Quantity, InvoiceDate , UnitPrice, CustomerID, Country.

InvoiceNo: The serial number of invoices. If begins with C, means this order has been cancelled.

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

CustomerID: Every customer has a unique 5 digits code.

- **Ideas for analysis**



Conclusion:

Descriptive statistical analysis

From the perspective of ORDER:

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

From the perspective of CUSTOMER:

- Over 25% customer only placed their order once and didn't retain.
- 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 price.
- Products with 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 the 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 was 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 get the highest sales in this sampling interval, that indicates there should be some orders with large goods quantity on that day.
- It's better for 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 can 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 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 brings 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 value are both 0, indicate more than 25% of clients only consumed once. The average life cycle is 130 days, the median is 93 days, that means some loyal customers make the mean value larger.
- This website needs to focus more on the improvement of customer's initial purchase experience. We can build online mark system and do some phone investigations to get feedback. Also, activities aim at attracting second purchase should be considered, for instance: offer coupons with time limitation.
- The average lifecycle for customer with more than twice purchase is 203 days, far higher than the total mean life cycle length 103 days.

Customer retention condition

- 'Customer life cycle' is the time difference between the first and last purchase. However, 'Customer retention period' focus on customers' periodic behaviors. Due to the dataset incompleteness we mentioned before, the outcome has 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-stickness were pretty good.

Purchase period

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