Contoso inter-purchase time interval analysis

DATA ANALYTICS CAPSTONE PROJECT REPORT
UNE ZEMAITYTE

I. PROBLEM

Contoso Corporation is a multinational business with its headquarters in Paris. The company is an electronics manufacturing, sales, and support organization with more than 100,000 products, operating online as well as having physical stores all over the word. A general look at the Contoso customer base immediately tells us that around 98% of its individual customers come back, while the business client base relies on Contoso unconditionally, with 100% of them coming back.

Additionally, since Contoso is a non-contractual business that does not incorporate a subscription system, it is not possible to know if a customer is still a customer. For that reason, this analysis aims to figure out customer behavior in terms of coming back to Contoso and remaining a loyal client. The objective of this report is to analyze the interpurchase time interval dynamic among individual customers and answer the following questions: how long are the intervals between purchases and do they depend on the customer lifespan, as well test the hypothesis, that shorter periods of time between the first purchases will increase the frequency of a customer.

Inter-purchase time interval (ITI) describes a period of time between two consecutive orders by the same customer. Despite it being related to shopping frequency, with more frequent buyers naturally having shorter intervals during a given time period, the timing of the first few purchases might indicate whether the customer is likely to come back or not.

II. METHOD

DATA SOURCE

The database for the project was downloaded from the Microsoft website, loaded locally using Azure Data Studio. It has 25 tables and follows a similar outline as the AdventureWorks database (see here). Sales data spans over 12 million rows, so in order to make the datasets usable in BigQuery and later on, in Data Studio, only 8 percent of the sales data (Tables *FactSales* and *FactOnlineSales*) was used, eliminating the data for return orders as well. To

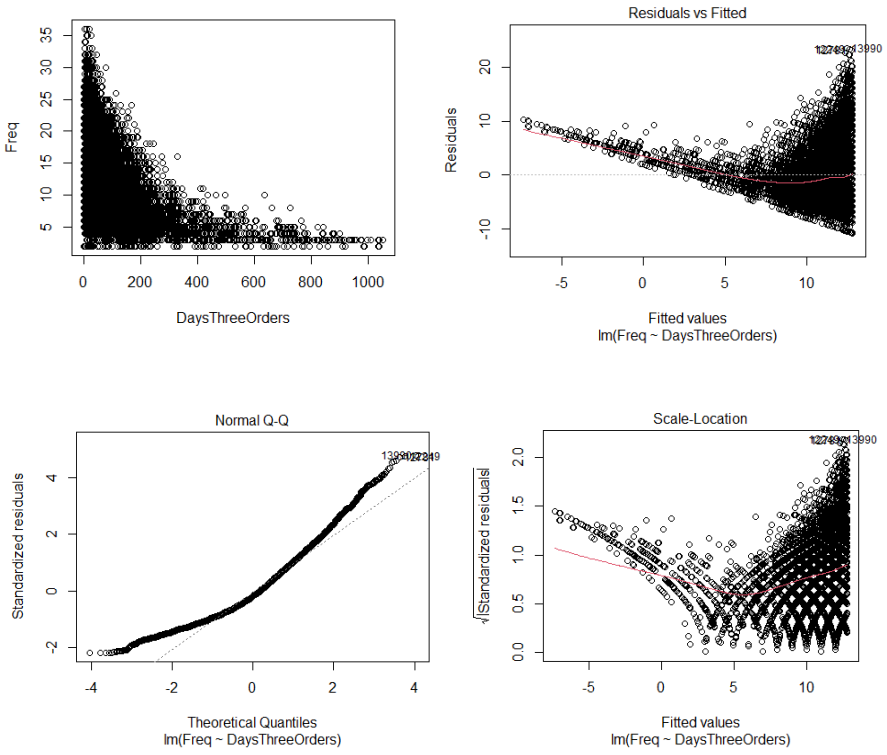
achieve this, a random sampling technique in Azure Data Studio was applied before exporting the .csv files.

10 tables were exported for the analysis: Customers, Geography, OnlineSales, Product, ProductCategory, ProductSubCatgeory, Promotion, SalesTerritory, Store, StoreSales.

ANALYSIS

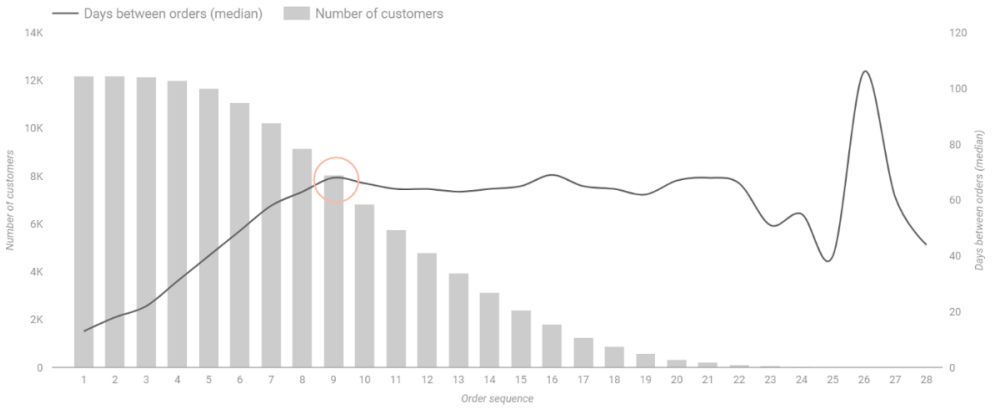
The necessary columns for the analysis were queried and aggregated with SQL in BigQuery, while aggregation for data visualization was automatically processed in Data Studio.

Simple regression analyses were conducted in R, using libraries *ggplot2, lm.beta,* and *dplyr*, between two variables: *frequency* (the number of orders a customer has made at Contoso) and *three-order time interval*, which is defined as the sum of days between making the 2nd and 3rd purchases at Contoso. In total, 18460 rows were analyzed. The variables met the assumptions of linearity, independence of errors, and normality of errors, but the assumption of equal variances is only partially met (see plots below).

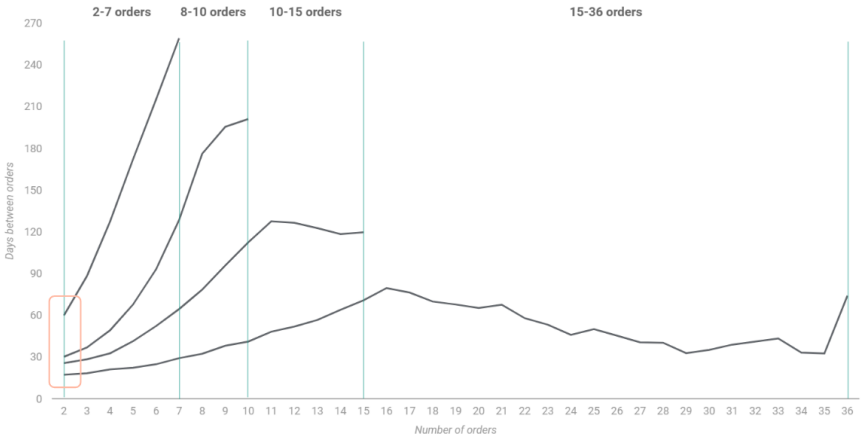


III. RESULTS

The first graph shows the timing of all customers throughout their lifespan, along with the number of orders they have made. We see that the median time between the initial purchases is around one month, later increasing to at least 2 months. Also, after 9 orders, the line tends to plateau, signifying that shoppers tend to come back more periodically after making a significant amount of purchases.



In terms of different less or more frequent buyers, there are differences in their ITIs, especially, during the first few orders. Initially, we see the “shortest” group spiking up rapidly, whereas for frequent buyers the line tends to plateau. Additionally, infrequent buyers seem to be waiting longer, around 2 months, to make a second or third purchase, whereas for other groups this interval is up to a month.



The time interval between first three purchases is a statistically significant ($p < 0.001$) and acceptable predictor ($\beta = -0.38$) of customer’s order frequency, but only explains 15% of the variance ($R^2 = 0.15$). The results indicate that for every increase in total days between the first three orders, the customer’s overall order frequency decreases around 0.4 times.

IV. CONCLUSIONS

The interval between the first ~10 orders tends to increase with every single order. Taking into account that the majority of users make between 10-15 orders, the plateau of the ITI at 2 months might indicate that after the 9th consecutive purchase, customers tend to shop periodically, with time intervals between purchases not differing too much.

Customers who have made the least number of orders, usually spent more time before making the initial purchases. The results revealed that the intervals for initial orders were smaller (around a month) for frequent customers, while for those who have only made up to 7 orders at Contoso overall, the initial purchases spanned around three months.

A similar finding was obtained by the regression analysis, indicating the possible tendency for shorter initial time periods to predict longer customer lifespan, and vice versa.

LIMITATIONS

Inter-purchase time interval analysis is not a heavily applied metric, likely because of its complex nature and a variety of possible variables affecting the time intervals. This analysis provides a very general outlook on the timing between purchases, but a more robust analysis technique (e.g. network analysis) could yield more accurate and actionable results, controlling for more outside variables.

FUTURE RECOMMENDATIONS

Based on the results in this analysis, the marketing team at Contoso should consider reaching the customers after they make a purchase, with a reminder to shop at Contoso again. This would potentially decrease the time between initial, and the subsequent purchases, while increasing the chances of gathering a larger base of loyal and frequent customers. The numbers show that the timing of such a reminder should not be overdue and be carried around a month after the purchase, if not sooner.