Finding relationships between sales and client features and building a predictive model for customer behaviour | Developing dashboard for sales metrics

Capstone 2019/2020 Spring

Gergely Futo (193297)

3 June 2020

Table of Contents

I.	Client background	.1
II.	Project introduction	.1
III.	Project Summary with findings and lessons learnt	.1
IV.	Recommendation	.3
V.	Conclusion	.3

I. Client background

My client's e-commerce company was founded in 2013 and the main profile of them is electronic products. The company has 40+ employees and its last year's net revenue falls in the range of 1-3 Billion HUF. This means 250000 packages delivered on annual basis. They sell only own branded products as they are a monobrand webshop company similar to Tchibo. At the moment they have three webpages for three countries including Hungary. On top of these the company has 3rd party channels, partnering with couple of companies in dropshipping model, including Hungarian eMag, where my client's company is one of the biggest marketplace partner. Finally, the company has B2B partners as well where resellers buy in huge portions from them.

II. Project introduction

In project kick-off we set two main goals, which we further specified later to maximize the value of the project:

- 1. Complete a descriptive analysis on sales historical data, find some patterns and build a predictive model for a certain product type, preferably Garden products.
- 2. Develop a dashboard for sales invoice and sales order metrics

III. Project Summary with findings and lessons learnt

Before COVID-19 outbreak I started to work on-site on a weekly basis therefore the data collection process went well as I could get structured data from ERP for sales invoices and sales orders and all additional manual

data was prepared for me within the first couple of weeks. The first challenge I faced was the limited data, as I only had one and half year of historical data which usually does not help in building models. On the other hand, the first positive surprise was an existing tracking mechanism for sales invoices, sales orders and other figures, however the process was a bit manual and less structured than expected, so I already saw at that point that my dashboard was going to be useful in a long run.

I worked on both parts of the project at the same time. I had regular calls and e-mail exchanges with my client to keep them updated. The first decision I made was to use Microsoft Power BI as an advanced analytics tool for the dashboard creation as it's a free software and number one tool in Gartner Magic Quadrant. The most challenging part of creating the dashboard was the data transformation which included transforming to tidy format the following datasets: daily sales order and invoice data, daily stock value and turnover data and 3rd party fee data. After having the necessary data structure, I created measures and additional calculated columns for each dataset so I could connect the tables and cover all the needs from my client. The next phase was designing the dashboard itself. I have created a landing page which includes the last refresh time of the dashboard, the logo of the company and some links to the other pages and highlights of results of four segments: invoice, order, stock value and turnover and 3rd party fee. All four segments have their own page with further details and charts, so I designed them one by one. The page for stock and turnover contains charts and highlights of daily average of stock value and turnover. The invoice page includes highlights of invoice amounts, number of unique partners and invoice amount deducted by 3rd party fee. The charts contain information about invoice amount by year, date, income source and product groups similarly to sales order page. Sales order page has highlights of number of customers and sales orders. 3rd party info page combines highlights of overall 3rd party results and charts with information about fee percentages by income source. Finally, based on the request of my client I created a page for yearly growth. It includes dynamic filtering option for date with date hierarchy. The charts on the last page comparing the metrics for invoice amount, orders, stock and one specific metric.

In addition to developing the Power BI dashboard, I was doing label engineering and started doing the exploratory analysis for the other part of the project, which showed increasing number of sales orders and invoice amount year by year and my client expected to grow further so this important finding resulted in yearly growth variable, which was built in my model as a regressor. Within a year all sales fluctuate within each month, so I included monthly dummies in the model, as well. Another finding was that Garden and Lighting had the second highest invoice amount with the highest value products on average. Not to mention awnings and garden chairs are the best selling products in the group. As aligned with my client I only focused on Garden related products, which was supported by the previous findings. At first, I was focusing on monthly/quarterly seasonality, however I recognized high volatility in daily orders and realized Monday and Tuesday are the strongest days of the week considering sales. In addition, there was big difference in daily sales therefore I included weekday numbers in my model. Significant finding was that garden products sales correlates with daily average temperature, therefore I considered adding this element to my model as well. I investigated another factor suggested by my client, precipitation however, surprisingly I found limited connection between raining and sales. The decline in sales in May last year was most probably due to temperature reduction. Furthermore, I have also explored BUX stock relationship as an economy related metric as in my opinion this stock could represent how the Hungarian economy performs and if the Hungarian economy performs well people are willing to spend more on retail webshops, however it showed no relationship so finally I dropped this possible predictor. Moreover, I have examined daily exchange rate between EUR and HUF as strong Hungarian Forint may also increase the willingness to buy certain products on webshops, however it showed no relationship either. Afterwards, finalizing the exploratory data analysis and investigating possible predictors, I started to develop models. Following training and testing several type of models my best model is a lasso model with the following regressors: last quarter's sales, yearly growth ratio, monthly dummies and seven weekday dummy variables are built in as well. My model performs well for the tested 3 months, it follows the volatility and it beat my benchmark model, which was calculating the last week's average invoice amount for the current week. My forecast model is the following with some clarification on the elements:

$$\begin{split} & Model_{Lasso} = Q_{n-1} + R * Q_{n-1} + R * Q_{n-1} * Weekday_1 + R * Q_{n-1} * Weekday_2 + R * Q_{n-1} * Weekday_3 + R * Q_{n-1} * Weekday_4 + R * Q_{n-1} * Weekday_5 + R * Q_{n-1} * Weekday_6 + R * Q_{n-1} * Weekday_7 + Month_1 + Month_2 + Month_3 + Month_4 + Month_5 + Month_6 + Month_7 + Month_8 + Month_9 + Month_{10} + Month_{11} + Month_{12} \end{split}$$

Yearly growth ratio: $R = \frac{Qn-1}{4*Qn-5} * \frac{Qn-4}{Qn-5}$

Qn-1: last quarter's sales; Qn-4: Current - 4 quarter's sales; Qn-5: Current - 5 quarter's sales

Weekday: dummy variable for each day of the week; Month: dummy variable for each month

IV. Recommendation

After completing my project, I came up with the following recommendations for my client:

- Lighting products tend to be more popular in winter season while garden products in summer and currently these products are part of one product group, therefore I suggested to split them into two separate groups: garden and lighting. This may help in reporting, visibility and reduce order management related errors.
- Currently, there are couple of system generated and manual errors in the ERP. My recommendation is, if possible, to correct the manual and system issues in the ERP even if it's extra workload for the team in the short term. It will be useful in reporting and in spend visibility in the long run. If it's not possible start using an error tracking file for different types of issues: wrong invoice numbers to be excluded from reports, recategorization of product groups, exchange rate table to be used for wrong currency conversions.
- The newly created PowerBI dashboard will help to track the previously mentioned metrics on a structured way. I advised to update the dashboard on a daily basis and put it on a screen in the office, so the team can come up with new insights and findings in the long term.
- During my limited time in the office I saw ad hoc reports sent via mail with different names. I would recommend to use a data repository for reports with naming convention, so there is less need to run ad-hoc reports. This may reduce the workload in the long run.
- This project was a good start, however, I see a project opportunity for next year to develop further my forecast model for garden products with more data and create new models for other product groups and to further develop the Power BI dashboard.

V. Conclusion

We set two goals with my client and even with challenges like COVID-19 outbreak and limited data I consider both goals are completed. Couple of months before my client was reporting their metrics in excel in an unstructured way and now they have a daily updated Power BI dashboard with well-structured data and great visibility on their most important metrics. On inventory management garden products sometimes caused problems to the company, however with my forecast model they may be able to manage their stock on an improved way in the near future. One personal note from my side: I really enjoyed working on this project as it covered data engineering, data analysis, business intelligence, data architecture and data visualization so most of the things we studied at the university and it was great opportunity to use my knowledge to real world problems with a collaborative client.