Teréz Julianna Szabó
Capstone Project Summary
08/06/2021

Automating a current dashboard and completing it with a forecast model

Introduction

Computers are becoming more and more important in the world and I'm proud I was able to contribute to this change at one company during my studies here in Budapest in the past year. This capstone project is the final project bringing together most of what this program covered in the past year. It was a challenge to do this alone and remember all the aspects of the course, but I am very proud of the work that was created in the end. The project focused mainly on visualization and automatization, but also included some modeling and forecasting. I created an HTML version of an existing report as a flexdashboard with some interactive values. Along with that, I automated the data combination process that goes into creating the dashboard. Lastly, I added a forecasting model that would help estimate the number of incoming tickets for the next month.

The company are currently working on automating and moving to the cloud companywide. Part of this process is automating processes and reports that have been done manually so far. They have a weekly report that is sent out to Management about the activity in a department. The report shows the status of tickets that are coming in. This report is currently put together by one person by hand into an excel document from multiple other excel reports every Monday.

The goal

The goal was to create a dashboard that can be filtered, so that it shows the information from the report as one dashboard that everyone could use and filter to find the data for their region. The process to create the dashboard should save time, so I set myself the goal of it taking maximum 30 minutes or even less to create the report if possible. It should automatize as much of the combination process as possible since this is the hardest and most time-consuming part. Furthermore, the dashboard should include all the information and graphs currently available in the report so that there is not a huge problem with switching.

In terms of technical aspects, the project should be completed using R since that is what the company is currently using. It was suggested by the company to use flexdashboard to create the dashboard, but other options will also be considered. Documentation of the project will be done mostly within the code so that it is available for anyone who opens the RMarkdown. Along with that, this document is created for all the technical aspects.

The process

The process of creating the final product included about five main areas. First the data needed to be combined into complete files instead of multiple excels by week. It was important to create complete files with data for at least the whole year so that it is easier to work with especially for the time series forecasting. Furthermore, reading date in directly from the original document created by the system is very important to avoid any form of corruption to the data because of human involvement. This process is then automated so it adds the current weeks data whenever the script is run on Monday morning.

The second important area was creating the visualizations that will be in the dashboard. This sounds like a relatively simple task, but these graphs were complex and created with excel which is hard to track in terms of steps taken to create. To create the interactivity and the ability to pick the region that was desired, I decided to use the crosstalk package. This would allow me to add a selector to the dashboard that could be changed. However, when I looked into the package a bit more it turned out that I used the wrong package to graph the data (ggplot instead of plotly). This created the additional work from transforming my existing ggplot objects into plotly ones, but in the end the interactive element was worth it.

I then developed a time series forecasting model. This model predicts the number of tickets that are going to come in that week and for the next three weeks. This is interesting because so far, all the information they were seeing was from the past or the current week. I believe this is a valuable addition to the dashboard and one they will truly appreciate once they found the best way to utilize it. I'm using three years (2018, 2019, 2020) worth of data to build the model.

I've spent some time deciding what kind of model would make sense to build. The original data only includes what I would consider the dependent variable and none of the other tables seemed like helpful information for independent variables. I considered adding additional data sources, but that could be confusing and seem unrelated to the rest of the dashboard. After some research I found some time series models that predict new values based on its own previous values. The final model, after analyzing multiple and checking their statistical significance, was an AR(2) model which stands for Auto Regressive model so looking back. Because it is an AR(2) model it looks back 2 time periods.

Once I had all the elements I've built the final dashboard in R using the flexdashboard package. Flexdashboard uses RMarkdown as its base makes it ideal for these repetitive processes. This way it could be rerun once a week and sent out to anyone who needed it. The

static HTML version would always be available for anyone who wanted to look at it. Another point is that it can be important to save older versions of the HTML for reference.

Outcome

Overall, this first version of the dashboard is running and can be used for the intended purpose. It is supposed to be prototype as a proof-of-concept for management that are currently using the PDF report. They should be shown that this automated report that is much faster to produce for the BI and Reporting team and has much less error due to the automation could work just as well than the current PDF report. I believe the final dashboard can be easily used for this purpose and makes it simple to imagine what the final product could look like.

Since this is only the first version as used a s proof-of-concept some things need to be updated when the next version is created. An important element that needs to be researched further and if possible updated are the text elements in the plotly functions. Another would be being able to combine all the different data tables used for the graphs into one so it could be filtered at once with the crosstalk package method.

Takeaways

While it has its frustrating moments, I thoroughly enjoyed this project. It was a challenge that allowed me to put the skills I learned in the past year to the test. I learned new ways to apply them and also completely new things along the way. I learned about new packages such as flexdashboard, crosstalk and how to graph in plotly. This project showed me what it is like to work for a real client and how demanding it can be. Furthermore, I learned how to structure and work a full data science project from beginning to end. I am very grateful to CEU for having these kinds of opportunities.