

Capstone project public summary

Szabolcs Kerekes – 2019.06.16.

The following document details the short overview of the capstone project, the results delivered, and the professional experiences gained from the work. Due to the non-disclosure requirement only certain details of the project will be shared in this document.

Project description

The topic of the project was analysing the relationship between asset flows into certain type of investment funds and the performances of stocks and market sectors, and key short- and long-term interest rates. The project was done for a client in the financial services industry, where the results would be taken into consideration related to financial investments.

The project ran from early April to mid-June in 2019 and involved many rounds of iterations on the data to achieve the desired outcomes. The work started with data cleaning and feature selection, fortunately as the input data was sourced from well curated data sources this step mostly involved transformations and the creation of new derived variables, after the exploration of the data was done. The project work involved trying different datasets to fit models on and select the best performing data and model combination.

One challenging aspect of the project was to find the evaluation criteria for the model selection. The challenge here was finding a metric that incorporates the most business sense. Due to this the model selection was primarily based on back-tested investment strategies compared with simple benchmark strategies, which due to their relevance proved to be more useful in most of the cases than simply using different statistical error metrics.

The project was done in Python with chart and excel type outputs. The analysis could have been implemented in any other programming languages (with sufficient statistical capabilities), the choice of Python was based on the client's specific ask for using this coding language.

Results of the project

The primary projects results were statistical models that can be used for making predictions for several different target variables based on the respective model. The results also detail the characteristics of the models, as well as the steps of the process for getting from the raw data to the final predictions. These results should help the client assessing scenarios related to investment decisions, which can be value generating for the client.

Furthermore, the results can be the basis for additional even deeper analysis of the research areas under scope. The client can use the results to evaluate whether any specific related areas or approaches would be reasonable to examine or build on. Recommendations were also provided along with results about the areas where the current analysis could be expanded, or new analysis could be conducted.

Lastly, due the implementation, the analysis can be easily reproduced via the notebooks created in Python. This allows the client to later-on repeat the analysis or fine-tune any parts or aspect of the modelling. This can be also the basis of creating a fully automated report later-on, which in that case could serve as an additional tool for the client.

Knowledge and skills gained

Overall, the capstone project provided an outstanding opportunity to implement the knowledge and skills gained during the MS in Business Analytics program. The project helped gaining knowledge and developing skills in the following areas.

Analytical skills

The project required using techniques for the full data analytics activity cycle from data transformation to modelling, while making the process easily reproducible. Gaining hands-on experience in how to approach and handle a data-oriented problem was very useful for developing an even more systemic and analytical thinking.

In this aspect I benefitted from the several Data Analysis and the Data Science courses, which helped in understanding the statistical concepts required for the project, as well as the implementation in a coding environment.

Domain knowledge

Dealing with asset flows was not entirely a new area for me, however this depth and the approach was still something that I haven't completely dealt with before. The project was helpful on gaining a better understanding of what data is related to investment funds, and what are the key aspects that can be interesting for any analysis in the financial services industry.

Project management knowledge

The project required heavy use of project management in order to properly time and structure the work. At certain stages decisions were required for the optimal usage of the available resources and to deal with certain possible dead-ends in the analysis. For this the project management course of the Business Analytics programme proved to be the most useful source for insights.

Technical skills

Although the Business Analytics programme has a stronger focus on using R, the course I took on data analysis and management with Python proved to be very useful, given that the capstone project partner asked for the work to be done specifically in Python.

From my side this required doing additional extracurricular Python courses (which was done on DataCamp) and putting the extra effort into creating working and understandable codes and scripts. I also utilized many of the 'Stack Overflow' articles, which helped a great deal with the coding of the analysis.

Personal takeaways

This project provided a deep insight into how complex data analysis projects can get. Doing the work also highlighted that a good combination of the aforementioned knowledge and

skills is required for successful projects. I think it is important to strike the right balance for each, in terms of not having a bottleneck area that limits the effectiveness in the others.

I think it is important to start with and maintain a consistent and structured workflow, while remaining flexible for new ideas and approaches as eventually the initial vision of the project can get fairly different at closing (even if the aim and the goals remain the same, the implementation and execution can easily shift as the work progresses).

I believe that the capstone project was an optimal closing of the Business Analytics programme, as it required extensive use of the materials taught during the academic year, while offering the chance to gain real life experience in data analysis via the corporate project.

