State Street Global Advisors UConn Senior Design Project Summary

Advisors: John Iacovacci, Dr. Phillip Bradford, Joel Turcotte, Patrick Bernard

Student Team: Agasti Mhatre, Kate Piotrowski, Taamir Khan, Souleyman Bane

Project Goal

- The students aim to build a full-scale model that can intelligently read large amounts of financial information and derive correlations. The students have successfully completed this and implemented so a cloud environment.

- For the Fall semester, the students focused on the correlation between security price and Morgan Stanley Capital International (MSCI) scores.

- For the Spring semester, the students focused on the correlation between security price, Content Quality (CQ), Global Industry Classification Standard (GICS) Sector, and TruCost scores.

- When building out this project, the students split it up into three components: building the model in Python using the Pandas library, building the model in the R programming language, which is a standard language used by quantitative analysts, and building out the cloud infrastructure.

Project Components

R-Model (developed by Kate Piotrowski and Taamir Khan)

- The students began constructing their model in R, as is utilized by the quantitative teams at State Street. The students leveraged correlation methods within R to determine a correlation between MSCI scores and price data of various securities.
- The students were unfamiliar with R before the start of this project and have grown their expertise in the language.
- The students and analysts preferred working with the R language over Python because it gives them a lot of information on how the numbers are being utilized to derive conclusions.
- The students used statistical analysis and data mining techniques to identify patterns and relationships between ESG data and security prices, created graphs to help visualize and communicate these findings in a more intuitive way, allowing the team to make more informed investments decisions.

Python / Pandas Model (developed by Agasti Mhatre)

- The students used the Pandas library as a starting point for building out their correlation model. Pandas is a widely used library with a reputation for being easy to use and having accurate outputs.
- The first step was to create mock data and find correlations before getting their hands on any financial data.
- Once the students could efficiently use the Pandas correlation modules, they were given MSCI and security price data.
- The next step was to merge all files so that the model could read just one dataset and output all the desired information.
- Once the data was merged, the students could output a table of security prices and their correlation with MSCI scores.

Cloud Infrastructure (headed by Agasti Mhatre and Souleyman Bane)

- The students created a Google Cloud account to host virtual machines which could be accessed by anyone on the team.
- The students uploaded their code base to the virtual machine by cloning their git repository.
- The students then created a SQL server in the cloud which would contain all of the financial data their model would need in order to output correlations.

Project Conclusion

- The students learned how to build a full-scale model capable of intelligently reading large amounts of financial information and deriving correlations.

- They gained experience in using the Pandas library in Python and the R programming language, which is used by quantitative analysts.

- The students gained an understanding of various ESG data sets and their implications in the market, providing insights into price movement.

- They gained experience in building out cloud infrastructure to implement their model.

- The project allowed the students to apply their knowledge of statistics and data analysis to real-world financial data, providing insights into how ESG data can impact security prices.

- The students work has been handed off to the Fundamental Growth & Core Equities Research team and will be utilized as reusable component within the firm.