Data Science Program Final Project

Executive Summary

At the end of the Data Science program, students are required to complete a final project of their choice. The teams are given six weeks to work on the project.

This document is dedicated to the Final 3 Team. It will explain the purpose and scope for the project.

Business Objectives

To showcase the skills that the Final 3 Team have acquired through the Data Science program. They will be using R, Python, Tableau, SQL, and other programs to wrangle, analyze, and visualize" Forecasting Daily Coffee Price" dataset made available by UCI Machine Learning on Kaggle.

At the end of the project, Final 3 Team should be able to explain their work in layman's term, and present their findings to the students, faculty, staff, and potential employers, along with other interested parties via Zoom.

Background

To activate and put practical use to what the students have learned, doing a final project is a good way to demonstrate that.

Team Final 3 have chosen the "Forecasting Daily Coffee Price" dataset because we all love coffee! They hope to gain insight on variables that affect the daily price of coffee.

Scope

Team Final 3 will be using the software taught in the program to complete the project. They will be intentional on using tools of their interest or tools that may aid finding a job. They may choose to use additional software/tools, but that is not required.

Functional requirements

Data Wrangling: The downloaded dataset should be successfully cleaned up for analyzing. Columns and unusable columns should be removed. As the dataset is fairly large, the Team should consider sub-setting the dataset in a proper manner, meaning the subset should be a random selection of the data. The datatypes for each column should also be converted to a usable format for the needed analysis.

Data Analysis: The Team will familiarize themselves with the dataset. They should have a good understanding of what each column means, and how the values are measured. They will brainstorm on questions to ask, and what they might gather from the dataset. Then, they will identify the proper functions to create models, predictions, etc.

Data Visualization: Once the team have a comprehensive understanding of and insight gathered from the dataset, they will work on visualizing the findings. They may decide to use Tableau or other graphing programs and compile the visuals and texts in a Power Point slideshow.

Presentation: The Team should be able to communicate in a clear and easy-to-understand manner. The presentation should be kept around 20 minutes. They should be dressed professionally for this occasion.

Personnel requirements

Gustavo, Chelsea and Henrietta are the three developers in Team Final 3. They will need to work closely for this project to succeed. They will touch base once a day via Zoom or Slack to problem-solve or to check in on work progresses. Once a week, they will review the past week workload and plan out the next week. They will take turns being the scrum master, and report their progress to their instructor (Product Owner.)

Once a week, they will meet with their instructor. They should be prepared to ask questions and seek guidance for the next steps.

They may also consult with their coding mentor.

Delivery schedule

Week 1: Import dataset into preferred software to begin data wrangling. Any unnecessary columns should be removed. Educate ourselves on breast cancer. Set up Github.

Week 2: Study the dataset and ask questions. What are some possible correlations? Is the data normally distributed? What are some predictive models we can make from it? Visualize the data to see if there is any interesting findings.

Week 3: Modeling/Optimization (Combined Stepwise - Forward and Backward Selection) and Machine Learning (Random Forest.)

Week 4: Review and validate findings from the previous week, and draw insights/conclusions.

Week 5: Compile findings into a Power Point slideshow. Go over it with their instructor and friend/family member to ensure that the presentation is clear and logical. Work on the style and layout of the presentation so it is delightful on the eyes.

Week 6: Make final touches to the Power Point presentation. The team should practice presenting at least a couple times with the three of them, and at least once with their instructor.

Other requirements

All programs used should be free of charge.

Assumptions

The software programs that are used should be available, up-to-date, and not broken.

Limitations

If something should come up for any of the team members during this six-week period, the project may be delayed.

Risks

The risks that may arise are such like natural disasters, power outages, family emergencies or broken software/hardware. The instructor and mentor are phenomenal so there is no concern of no help from them. The risk of this project being incomplete is minimal. They will be successful in completing this project!