**Data Science Program Final Project**

Executive Summary

At the end of the Data Science program, students must complete a final project of their choice. There are six weeks to work on the project. Typically, they will work with other student(s) to complete the assignment.

Philip, Charissa, and Edith' created this document. It will clarify the objective of the project.

Business Objectives

To demonstrate the abilities Philip, Charissa, and Edith' have obtained throughout the Data Science program. The team will use Rstudio to wrangle, analyze, and visualize the "New York Health Rankings" dataset made available by https://www.countyhealthrankings.org/app/new-york/2022.

At the end of the project, Philip, Charissa, and Edith should be capable of illustrating their outcome in layman's terms and presenting their discoveries to the pupils, faculty, and possible employers, along with other curious audiences through Zoom.

Background

The final project is a way to demonstrate how students can use data science in real-life scenarios. Essentially, it is creating a beneficial use of all the academics we specifically learned. Philip, Charissa, and Edith have decided to analyze the "New York Health Rankings." because of the

lack of attention to other factors throughout covid. They anticipate learning from the missed premature deaths and the most significant influences on longevity.

Scope

Philip, Charissa, and Edith will be using the software taught in the program to complete the project. They may use additional software/tools, but that is not required.

Functional requirements

Data Wrangling: The downloaded dataset should be clean for analysis. In order to clean our data columns and unusable columns will be removed. As the dataset is relatively large, Philip, Charissa, and Edith should consider properly sub-setting the dataset, 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: Philip, Charissa, and Edith will familiarize themselves with the dataset. They will brainstorm questions to ask and what they might gather from the dataset. Then, they will identify the proper operations to construct ideals, projections, and solutions.

Data Visualization: Once they understand the data and have gathered conclusions and insight from the dataset, they will work on illustrating the outcomes. They may decide to use Tableau or other graphing programs and compile the visuals and texts in a PowerPoint slideshow.

Presentation: Working with institute administrators, the team will schedule a time to deliver their conclusions through Zoom. They should be able to convey and present their findings in an easy-to-understand presentation. The project should be presented for approx. 20 minutes in professional attire.

Personnel requirements

The three developers are Philip, Charissa, and Edith; they intend to work closely for this assignment to succeed. They will touch base daily through text or Slack review on work progress. They will examine the past week's workload once a week and plan out the following week. They will assume the role of scrum master and report their progress to their educator (Product Owner.)

They should be prepared to ask for guidance from their instructor or coding mentor once a week.

Week 1: Import the dataset into the selected program and begin data wrangling. Any unneeded columns should be cleared. Educate ourselves on death rates in new york and why. Set up Github.

Week 2: Review the dataset and question the best ways to analyze the data set. What are some likely correlations? Is the data commonly dispersed? What are some predictive measures we can assemble the data? Produce visualizations of data to detect if there are any intriguing conclusions.

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: Assemble conclusions into a PowerPoint slideshow. Practice presenting the data with friends and family. Reach out to the instructor to review the presentation to ensure the assignment is cohesive and shows the correct direction that the team is striving for.

Week 6: Complete the last edits to the PowerPoint presentation. The team should not try to start any new analysis. There will not be adequate time to affirm their conclusions. The team should be using this time to practice and perfect their discoveries.

Other requirements

All of the software to present should be free of cost.

Assumptions

The software programs and platforms the team will use should have the latest versions to prevent corrupting the data.

Limitations

The assignment may be postponed if something should come up for Philip, Charissa, and Edith throughout these six weeks. The project may be delayed to another date if the educator or mentor has scheduled or unscheduled time off. They may experience a roadblock in their work, moving the fulfillment date.

Risks

The risks that may occur are Tornados, Earthquakes, Thunderstorms, power outages, family difficulties, or broken software/hardware. Charissa, Philp, and Edith are enthusiastically ready to complete the program, so there should be no motivation issues. The instructor and mentor are fantastic and have great response times. The risk of this project being insufficient is marginal.