MRC Data Science Program Final Project

Southern Careers Institute / Woz U

Executive Summary

At the end of the Data Science program, students are required to complete a final project of their choice. They are given six weeks to work on the project. Often times, they will be paired up with another fellow-student.

This document is dedicated to Matt, Richard, and Craig's Final Project.

It will explain the purpose and scope for the project.

Business Objectives

To showcase the skills that Matt, Richard, and Craig have acquired through the Data Science program. They will be using R, Python, Tableau, and other programs to wrangle, analyze, and visualize the "Effects of Covid-19 on the Navajo Nation Indian Tribe with a dataset made available by Johns Hopkins University. Additionally data collected by New Mexico Highlands University.

At the end of the project, Matt, Richard, and Craig 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

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

Matt, Richard, and Craig have chosen the "Effects of Covid-19 on the Navajo Indian Tribe because COVID-19 has taken over the world in 2020, and the Navajo Nation people have been one of the hardest hit communities in the entire USA.

Scope

Matt, Richard and Craig will be using the data collection, analysis methods taught in the program to complete the project. They will be intentional on using tools of their interest or tools that may aid them in finding a job after they graduate from the program. 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, Matt, Richard and Craig should consider sub-setting the dataset in a proper manner, meaning the subset should be a random selection of the data. The data types for each column should also be converted to a usable format for the needed analysis.

Data Analysis: Matt, Richard, and Craig 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 Matt, Richard and Craig 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: Working with school leaders, Jane and Jessica will schedule a time to present their findings via Zoom. They 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

Matt, Richard, and Craig are the developers. 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 COVID-19 cases in the USA and Navajo Nations Set up Github.

Week 2: Study the dataset and ask questions. TBD...Visualize the data to see if there is any interesting findings.

Week 3: Modeling/Optimization - 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. Matt, Richard and Craig should not attempt to come up with a brand-new analysis. There will not be enough time to verify their findings. They 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. Though Matt, Richard and Craig may decide to use a paid service, such as a more advanced version of Tableau.

Assumptions

The software programs and platforms Matt, Richard, and Craig use should be available, up-to-date, and not broken.

Limitations

If something should come up for Matt, Richard, and Craig (like moving…) during this six-week period, the project may be delayed. If the instructor or mentor have scheduled or unscheduled time-off, the project may be delayed as well. Matt, Richard, and Craig may experience a roadblock in their work, which may push back the completion date.

Risks

The risks that may arise are such like natural disasters, power outages, family emergencies or broken software/hardware. Matt, Richard, and Craig are eager to complete the program so there should be no motivation issues. 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!