**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. They are given six weeks to work on the project. Often times, they will be paired up with another fellow-student. I will, however, be working by myself for this project.

This document is dedicated to myself to answer my own questions I and my family have been asking lately.

**Business Objectives**

To showcase the skills I have acquired through the Data Science program. I will be using R, Python, possibly Tableau, and other programs to wrangle, analyze, and visualize the multiple datasets I have chosen from Kaggle.

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

I will be examining the questions of whether or not it is more cost efficient to purchase an electric vehicle and how much you would either save or lose when purchasing an electric vehicle.

**Background**

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

I have chosen the multiple datasets I have chosen because I need multiple sources of information in order to find the answers to my questions.

**Scope**

I will be using the software taught in the program to complete the project. I will be intentional on using tools of their interest or tools that may aid finding a job. I 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 and I have multiple datasets, I might 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: I will familiarize myself with the dataset. I will have a good understanding of what each column means, and how the values are measured. I will figure out what questions to ask, and what I might gather from the dataset. Then, I will identify the proper functions to create models, predictions, etc.

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

Presentation: Working with school leaders, I will schedule a time to present their findings via Zoom to my mentor if they have time. I should be able to communicate in a clear and easy-to-understand manner. The presentation should be kept around 20 minutes or less. I will be dressed professionally for this occasion.

**Personnel requirements**

I am the solo developer. I will need to work carefully for this project to succeed.

Once a week, I will meet with my instructor. I will be prepared to ask questions and seek guidance for the next steps.

I can also consult with my coding mentor.

**Delivery schedule**

Week 1: Import dataset into preferred software to begin data wrangling. Any unnecessary columns will be removed. Educate myself on my topics. 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 are 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 my 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. I should not attempt to come up with a brand-new analysis. There will not be enough time to verify my findings. I will practice presenting at least a couple times by myself, my wife, and at least once with my instructor if they have time.

**Other requirements**

All programs used should be free of charge. Though I may decide to use a paid service, such as a more advanced version of Tableau, I probably will not.

**Assumptions**

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

**Limitations**

If something should come up for me 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. I 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. I am 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. I will be successful in completing this project!