**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. This will be a one person project, not a group effort project.

This document is dedicated to me, Cing San Lun. It will explain the purpose and scope for the project.

**Business Objectives**

I would like to showcase my skills that I have acquired through the Data Science program. I will be using Excel, R, Python, and Jupyter Notebook to heavily focusing on analytic and visualization. I will be using my personal bank statement for the year of 2022 “Credit Card Year End 2022” dataset made available by Wells Fargo as a CSV file.

At the end of the project, I should be able to explain what issues I have with budgeting in my personal finance. And also, if I have impulse spending on random things. I will then present my 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.

I have chosen this budgeting dataset because I am interested in this area of working environment where it is heavily involve with using analytic and visualization. I hope to glean insight from this document on how to budget my expenses and extrapolate my spendings.

**Scope**

I will be using different types of 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 might 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, I 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: I will familiarize myself with the dataset. I should have a good understanding of what each column means, and how the values are measured. I 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 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. I should be able to communicate in a clear and easy-to-understand manner. The presentation should be kept around 20 minutes. I should be dressed professionally for this occasion.

**Personnel requirements**

I will be working by myself as a developer and ask questions when needed to finish this project successfully. I will touch base with others such as instructors or mentors via Zoom or Slack to problem-solve or to check in on work progresses. Once a week, I will review the past week workload and plan out the next week. I will report my progress to my instructor.

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

I will also consult with my coding mentor.

**Delivery schedule**

Week 1: Import dataset from Wells Fargo into preferred software to begin data wrangling. Any unnecessary columns should be removed. Educate myself on finance and how to budget. 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 I can make from it? Visualize the data to see if there is any interesting findings.

Week 3: Modeling/Optimization

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 or 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 should practice presenting at least a couple times with my instructor or others if I can.

**Other requirements**

All programs used should be free of charge. Though I might decide to use a paid service, such as a more advanced version of Tableau.

**Assumptions**

The software programs and platforms that 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 my 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 if I do everything on time successfully.