**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.

This document is dedicated to Jane and Jessica’s. It will explain the purpose and scope for the project.

**Business Objectives**

To showcase the skills that Jane and Jessica have acquired through the Data Science program. They will be using R, Python, Tableau, SQL, and other programs to wrangle, analyze, and visualize the “Breast Cancer Wisconsin” dataset made available by UCI Machine Learning on Kaggle.

At the end of the project, Jane and Jessica 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.

Jane and Jessica have chosen the “Breast Cancer Wisconsin” dataset because they are both interested in healthcare, and preventative care. They hope to glean insight from this document to make actionable suggestions on how to identify breast cancer types (benign and malignant.)

**Scope**

Jane and Jessica 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, Jane and Jessica 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: Jane and Jessica 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 Jane and Jessica 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**

Jane and Jessica are the two 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 and begin to analyze which dataset to use to perform each analysis for the questions that are proposed. Begin to review the data chosen for the proposed questions. Educate ourselves on the topic at hand. Set Up Github.*

Week 2: *Study the dataset and begin to ask questions. Break up the questions proposed and identify the methods used to analyze the data and see if there were any inconsistencies. Are there any correlations? What possible methods of presentation would best showcase the analysis? Whether to use R Studio or Python etc.*

*Conduct weekly meeting to discuss the proposed working going forward.*

Week 3: *Using the different components of the program from what we learned and demonstrate how to use; Data Wrangling and Visualization: Data Transformations. Intermediate Statistics- Single Sample T-Tests, Independent T-Test. Tableau to make visual presentation of the data.*

*Conduct weekly meeting to discuss the proposed working going forward.*

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

*Conduct weekly meeting to discuss the proposed working going forward.*

Week 5: *Begin to piece together the presentation. Reviewing each component to ensure it will be presented clear and concise. Work out any visual media to best present the project.*

*Conduct weekly meeting to discuss the proposed working going forward.*

Week 6: *Make final touches to the Power Point presentation. Jasmin, Lacey, and Ar’kajia will not attempt to come up with a brand-new analysis. There will not be enough time to verify their findings. We will practice presenting and determine suitable placement. Present the presentation at least once with their instructor.*

**Other requirements**

All programs used should be free of charge. Though Jane and Jessica may decide to use a paid service, such as a more advanced version of Tableau.

**Assumptions**

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

*All of the programs that will be used will be readily available to Jasmin, Lacey, and Ar’kajia.*

**Limitations**

*Limitations collectively will be scheduling. Respecting the schedules of Jasmin, Lacey, and Ar’kajia may be a limitation, but we are eager to take ownership and maintain the weekly scheduled deadlines. Availability and not being able to maintain the weekly schedule. We expect to have a timely delivery of the project.*

**Risks**

*There may be many risks associated with completing a group project. One of them being, failing to maintain a schedule and meeting deadlines. Jasmin, Lacey, and Ar’kajia are determined to be efficient in communication and conveying all discrepancies. Collectively we will be working diligently to complete the project in the proposed timeline.*