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 document is dedicated to the Star Warriors. It will explain the purpose and scope of the project.

Star Warriors members include Tameka Gillett, Talia Hopkins, Alison Schnoes, Valerie Salazar, and Whitney Taylor.

Business Objectives

To showcase the skills these members have acquired through the Data Science program. They will be using R, Python, Tableau, SQL, and other programs to wrangle, analyze, and visualize the “PCOS Dataset" made available by Prasoon Kottarathil on Kaggle.

At the end of the project, the Star Warriors members should be able to explain their work in layman’s terms 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.

Star Warriors members have chosen the “PCOS Dataset” because they are interested in polycystic ovary syndrome which is a common hormonal disorder that affects ovaries in women during childbearing years. This results in an irregular menstrual cycle. They hope to glean insight from this document to make actionable suggestions regarding how different health metrics affect PCOS.

Scope

The Star Warriors will be using the software taught in the program to complete the project. They will be intentional in using tools of their interest or tools that may aid in 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, the Star Warriors 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: The Star Warriors 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 the Star Warriors 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 PowerPoint slideshow.

Presentation: Working with school leaders, the Star Warriors 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

The Star Warriors: Tameka Gillett, Talia Hopkins, Alison Schnoes, Valerie Salazar, and Whitney Taylor will be working closely for this project to succeed. They will touch base once a day using Google Meets and Chat to problem solve or to check in on the work progress. Once a week, they will review the past week's workload and plan out the next week. They will rotate being the scrum master and report their progress to their instructor (Product Owner.)

The scrum master schedule follows this rotation: Talia, Whitney, Valerie, Tameka, Alison, and Talia

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 Polycystic ovary syndrome (PCOS). 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: Follow the Analysis Flowchart for answering the evaluation questions

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

Week 5: Compile findings into a PowerPoint 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 PowerPoint presentation. The Star Warrior team 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 of times with the group, and at least once with their instructor.

Other requirements

All programs used should be free of charge.

Assumptions

The software programs and platforms the Star Warrior team use should be available, up-to-date, and not broken.

Limitations

If something should come up for the Star Warrior team during this six-week period, the project may be delayed. If the instructor or mentor has scheduled or unscheduled time-off, the project may be delayed as well. The Star Warrior team may experience a roadblock in their work, which may push back the completion date.

Risks

The risks that may arise are such as natural disasters, power outages, family emergencies, or broken software/hardware. The Star Warrior team is 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!