Abi Glisson

8/12/22

Week 1 – Project Proposal

**Data Science Program Final Project**

**Executive Summary**

At the end of the Data Science program, students are required to complete a final project. In the current iteration, students are assigned a dataset and complete the entire project on individually.

This document is dedicated to Abi Glisson’s final project. It will explain the purpose and scope for the project.

**Business Objectives**

This project is intended to showcase the skills that Abi has acquired through the Data Science program. She will be using R and Python to wrangle, analyze, and visualize the “Daily Bike Sharing” dataset supplied to her by her instructor Dr. Mohamad Saleh.

At the end of the project, Abi will be able to explain her work in layman’s term, and present her findings to the students, faculty, staff, and potential employers, and any other interested parties via Zoom.

**Background**

A final project is a meaningful way to showcase the learning that students have acquired throughout their time in ENTITY’s Data Science Virtual Academy and to demonstrate that they are able to apply their skills to meaningful real-world datasets.

Abi has been assigned the “Daily Bike Sharing” dataset, an analysis of which may be helpful in bike sharing companies’ ability to make sound business decisions.

**Scope**

Abi will primarily be using R and Python to complete the data analysis. Tableau may be used in the data visualization stage. Other tools taught during the program may be used as necessary, but it is not the intent that Abi should have to learn any new programs. (Of course, she may need to learn additional features/code in the existing programs in order to fully execute on the evaluation question, which is not considered outside the scope.)

The dataset has been assigned by the instructor and is therefore not expected to change during this project. Any analyses that were taught during this program may be used to complete this data analysis project.

**Functional requirements**

Data Wrangling: The downloaded dataset should be successfully cleaned up for analyzing. Different subsets of the variables may be needed in order to perform different types of analyses to fully answer the evaluation question.

Data Analysis: Abi will familiarize herself with the dataset. She should have a good understanding of what each column means, and how the values are measured. She will brainstorm on what information might be gathered from the dataset. Then, she will identify the proper functions to create models, predictions, etc.

Data Visualization: Once Abi has a comprehensive understanding of the analyses and insight gathered from the dataset, she will visualize the findings. She may decide to use Tableau or other graphing programs and will compile the visuals and texts in a Power Point slideshow.

Presentation: Working with her instructor, Abi will schedule a time to present her findings via Zoom. She will communicate in a clear and easy-to-understand manner. The presentation should be kept around 20 minutes. She will dress professionally for this occasion.

**Personnel requirements**

Abi is the data scientist on this project. She will complete the bulk of the work completely independently. She will schedule regular work blocks in order to complete the project in the required time domain (6 weeks total). She is her own Scrum master.

Abi has been assigned to participate in a mastermind group of classmates who are also working on the same dataset. Although this group is available for help or brainstorming, Abi is entirely responsible for her own work product.

Once a week, she will meet with her instructor (Product Owner) Dr. Mohamad Saleh in a small group setting with her assigned classmate group. She will be prepared to ask questions and seek guidance for the next steps.

Abi may also consult with her coding mentor (Latrice Puckett) and/or the ENTITY Senior Mentor Devin Moya for technical assistance.

**Delivery schedule**

As this project is scheduled to occur over 6 weeks with intermediate deliverables, Abi is unable to predict the intermediate deliverables as the week’s requirements are not revealed until that week’s Lessons open in the Exeter LMS.

In Week 1, Abi has completed an initial familiarization process and determined the types of variables present in the dataset.

In Weeks 2-5, Abi will perform all parts of the data science pipeline, including data wrangling, analysis, modeling and optimization, and visualization.

By Week 6, Abi will prepare for and present her findings to her instructor and any other interested parties.

**Other requirements**

All data science programs used (R and Python) should be free of charge. Abi may choose to pay for advanced features of Tableau, though this is not required. Abi already has full access to the Microsoft suite of products to be used for creating the final presentation and report of findings.

**Assumptions**

It is assumed that the software programs to be used will be available, up-to-date, and not broken.

**Limitations**

Potential limitations are the instructor’s and/or group’s schedules for the weekly group meetings where they will receive updates and instructions on completing the project.

Additional limitations could be Abi’s availability to complete the required work. Although she is intending to work about 20 hours per week on this project, the potential for illness (hers or her family’s) or other unforeseen family circumstances remains. If temporary disruptions occur during the middle phases of the project, it is Abi’s intent to still work hard enough to make up for the lost time and complete the project by the 6 week deadline.

**Risks**

Potential unforeseen risks that may arise are natural disasters, power outages, family emergencies, and/or broken software/hardware. Abi is eager to complete the program in the same high-achieving manner that she has completed the coursework modules, so there should be no motivation issues. Any limitations that arise should be temporary in nature, so the risk of this project never being completed is nearly 0, while the risk of any delays in the project are low to minimal.