**Data Science Program Final Project**

**Executive Summary**

This document is dedicated to Les and Latrice’s final project for the Woz-U Data Science course. It will explain the purpose and scope for the project.

**Business Objectives**

To showcase the skills that Les and Latrice have acquired through the Data Science program. They have chosen to use R, Python, Tableau, SQL, and other programs to wrangle, analyze, and visualize the “Recidivism of Prisoners” dataset made available by the National Institute of Justice.

At the end of the project, Les and Latrice 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**

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

Les and Latrice have chosen the “Recidivism of Prisoners” dataset because they are both interested in the long terms causes for former felons returning to prison. This information can be used to help determine post-release support programs such as assessing mental health needs, addiction intervention, social environmental influences, and alternative sentencing to reduce recidivism. Reducing the number of people returning to prison also has an impact on prison overcrowding and costs.

**Scope**

Les and Latrice 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 large, Les and Latrice 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: Les and Latrice 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 Les and Latrice 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, Les and Latrice 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**

Les and Latrice 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 (Joe Raetano.)

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.

Les and Latrice will take turns as Scrum Master. Les has taken the lead in Week 1. Latrice will be taking Week 2 and so on.

**Delivery schedule**

Week 1: Import dataset into R to begin data wrangling. Any unnecessary columns and null data should be removed. Educate ourselves on recidivism. Set up Github, Trello, and Tableau.

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. Questions we plan to use for analysis are How does age play a factor? How does the kind of previous conviction affect recidivism? Does drug usage play a role? How does ethnicity play a factor?

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 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 Power Point presentation. Les and Latrice 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 times with the two of them, and at least once with their instructor.

**Other requirements**

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

**Assumptions**

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

**Limitations**

If something should come up for Les and Latrice 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. Les and Latrice may experience a roadblock in their work, which may push back the completion date. As Les and Latrice both work full time jobs outside of the project, they will coordinate how their work schedules coincide so they can avoid project delays.

**Risks**

The risks that may arise are such like natural disasters, power outages, family emergencies or broken software/hardware. Les and Latrice are eager to complete the program so there should be no motivation issues. They have a lot of support from the instructor so that shouldn’t be an issue. The risk of this project being incomplete is minimal. They will be successful in completing this project!