**Data Science Program Final Project:**

**Spotify Five-Year Analysis 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. Students will often pair up with another student to complete the project as a team..

This document will explain the scope and purpose of the final project of Bianca Serrano and Katie Ravenwood. The most current version of this document can be viewed on GitHub at:

https://github.com/Nuevachimi/Five-Year-Spotify-Analysis-Project/blob/main/Proposal

**Project Objectives**

To showcase the skills that Bianca and Katie have acquired through the Data Science program. They will be using R, Python, Tableau, SQL, and other programs to wrangle, analyze, and visualize the “Spotify Five Year Analysis Project” dataset they created for this project.

At the end of the project, Bianca and Katie should be able to explain their work in layman’s terms and present their findings via Zoom.

**Background**

The Data Science Final Project is a way to activate and put to practical use what the program’s students have learned.

Bianca and Katie have chosen the “Spotify Five Year Analysis Project” dataset because music is a vital part in their everyday lives. They are interested in understanding whether the general mood of the most-played albums and tracks has changed over the course of the last five years, particularly during the worldwide COVID-19 pandemic. The Five-Year Spotify Analysis will take a look at songs from top Billboard albums within the five year range from 2017 through 2021. These songs will be used in order to analyze correlations and trends between yearly charts and within different groups of tracks. Analysis will be performed to determine whether there are any overall changes in valence, the measurement of predicted song "mood", of the top albums' most popular tracks from year to year, as well as analyses to determine whether there

**Scope**

Bianca and Katie 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.

The scope of the data for this project will extend to analysis of tracks from the Top 200 Album charts for the past five years. Analysis will neither extend further back beyond that time scope or project forward more than a year beyond the current chart year. Analyses will include only features readily available and accessible through Spotify's public API and standard documentation.

**Functional requirements**

Dataset Creation: This dataset has been created by Bianca and Katie specifically for the purpose of analyzing the proposed questions. The dataset has been created from information in yearly Billboard Top Album charts and imported via the Spotify API.

Data Wrangling: The dataset be cleaned up and unnecessary columns will be removed. The data will then be subsetted to produce a more concise and manageable dataset. Each column will be converted to the appropriate datatype in order to successfully execute analysis. The main columns that will remain in the data set are Artist Name, Album Name, Genre, Duration, Tempo, Valence, Explicit, and Popularity.

Data Analysis: Bianca and Katie will familiarize themselves with the dataset. Data will be analyzed primarily in R with preliminary visualizations.

Data Visualization: Data visualizations will primarily be finalized in Tableau. Planned visualizations include year-to-year valence change as well as tempo and genre correlation to valence.

Presentation: Findings will be presented via Zoom/Webex.

**Personnel requirements**

Katie and Bianca are Data Science students. They will work closely and touch base daily to work for check in and project updates. These daily meetings will involve analyzing and problem solving. Daily communications will be through text message, slack, and zoom.

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. Set up Github.

Week 2: Data Wrangling and Dataset Exploration

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.

Week 6: Make final touches to the Power Point presentation.

**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.

**Limitations**

Shedule conflicts between Katie and Bianca and with the instructors may be possible, and shifts in the daily scheduled work time may be necessary. It is possible that one might need to shift work time to earlier or later in the day and that can push back the timeline. If something should come up for Katie and Bianca 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. Katie and Bianca may experience a roadblock in their work, which may push back the completion date.

**Risks**

The risks that may arise include force majeure events such as natural disasters, power outages, family emergencies or broken software/hardware. Bianca and Katie are 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!