**DSC 520 Final Project**

Name: Steven Miller

Date: 2019-04-06

**Section 2**

* Provide an introduction that explains the problem statement you are addressing. Why would someone be interested in this?
  + By analyzing the results and other data from Formula One races and comparing this data to fan ratings of each race, I hope to better identify what factors contribute to a “good” race.
* Provide a concise explanation of how you plan to address this problem statement.
  + A large portion of the project will be focused on hypothesizing on factors that might be contributors to the quality of a race and seeing how those factors for each race correlate with the fan ratings.
* Discuss how your proposed approach will address (fully or partially) this problem.
  + By analyzing the correlations of these factors to the fan ratings, we can get a better understanding of which factors do and do not contribute to the quality of a race.
* List at least 6 research questions you aim to answer.
  + How have rule changes in recent years impacted the quality of races?
  + Are fans more engaged when more drivers are winning throughout a season or when a few drivers dominate?
  + Does weather play a factor in the quality of a race?
  + Are fans of a specific driver or team leading to biases in the data? Do we tend to find a higher score when specific drivers or teams win?
  + How does the circuit play a factor into the quality of a race? Are longer circuits more exciting?
  + Are races with more pit stops more or less exciting?
* Explain how your analysis may help the consumer of your research findings (recall you target audience from Section 1).
  + This analysis could help decision makers in Formula One adjust moving forward to provide a more engaging product for fans. For example, if a positive correlation is found between wet weather races and a good rating, adjustments to the calendar could be made to schedule races at times of the year when the weather is more likely to be rainy. If it is found that pit stops increase the excitement of a race, decision makers could partner with tire manufacturer Pirelli to develop softer tire compounds leading to higher degradation and more pit stops during an event.
* What types of plots and tables will help you to illustrate the ﬁndings to your research questions?
  + Histograms will be needed initially to get a better understanding of the initial data. In some cases, we will be looking at the impact a categorical value has on a quantitative value (for example, comparing the average rating of a race won by a specific rate team). In these situations, tables will be useful. Much of the analysis will focus on correlations. Correlation matrices will be helpful here, but so will scatterplots to allow the visualization of data.
* What do you not know how to do right now that you need to learn to answer your research questions?
  + The main area I need to learn more about is working with a relational database in R. The dataset I am using is spread across multiple tables and the data needed to fully present an analysis is not always present within the same table.
  + I will also need to get a better understanding of data transformations through the dplyr package. I expect dplyr to be sufficient for my needs, but if I find myself needing something different I will have to learn more about the packages available for R.