

Rody Bertolini

CSPB 4502

Final Project - Individual Evaluation

12/5/2025

This project was interesting to develop and frustrating in its results. While I think many components of the project went well, I had minimal issues with the coding aspect, the topic was one that interested me, and I was able to learn new skills while implementing the project, I don't feel like my results are very clear or meaningful. While the preprocessing, cleaning, etc. went well and was fairly standard, my initial EDA revealed a mess, the data was so disparate and confused that I wasn't able to draw any meaningful conclusions. I realized this was because the rule changes have dramatically impacted things like pit stops, and the change in power unit from standard combustion to hybrid also had a dramatic impact. I decided to limit the years I was analyzing to focus on the modern era and help at least demonstrate some trends in the data.

Throughout this project I applied the concepts we learned in class, but was also forced to think about them more robustly to solve the aforementioned issue. This issue helped me realize the importance of early EDA in determining the usefulness of our results, and that datamining might not always be a linear process. I also learned more about modeling via constructing my linear regression model, which had me using resources like the Rielly's book on Data Science.

If I could do this project again I would pick a different subject matter. I realized too late that the aspects of F1 are too interconnected to try and determine which aspects of a race are most important. For instance, the constructors ultimately helps determine qualifying, a faster car means faster lap times which means a higher qualifying position. There was no way to isolate the various factors I was studying and so the results aren't as useful or interesting as I would like.