Project Proposal: F1 Analysis Tool

Samuel Mata · 300574763 · DATA-472 · Project Proposal

Tracking sports can be notoriously difficult as it often requires in-depth domain knowledge of rules and regulations. This is particularly notable in Formula 1 *(F1)*, where rules and regulations change often, and the scoring system can be obtuse for those who are not familiar with the sport. This project aims to provide a tool to allow for the easy and quick analysis of various aspects of F1.

1. Background

F1 is a complex sport with various interacting systems for measuring performance. As such, analysing F1 can be done through various aspects, including seasonal scoring *(both for drivers and teams),* lap times *(both for qualifying and race pace),* overall rankings *(across multiple seasons),* pit-stop timing, circuit times, and more. Due to the highly technical nature of the sport, data is plentiful and public. Unfortunately, this data is often difficult to understand by eye or without special parsing. This leaves space for development of tooling to improve the clarity and usability of such analysis.

2. Goals

The goal for this project is to develop a solution for the described problem above. This will involve the development of an interactive web application *(built with R and Shiny)* to allow for users to perform queries and analysis on current and historical F1 data. The application should be intuitive to control and aesthetically pleasing to allow for users that may not be necessary familiar with F1 to perform analysis regardless of domain knowledge. To further this focus on usability, features should be implemented to guide understanding of the tool and its results.

3. Data

This project will be using the Formula 1 World Championship (1950 – 2023) dataset made publicly available on Kaggle. This dataset extensively details data on all races, drivers, constructors, qualifying sessions, race sessions, circuits, lap times, pit stops, and championships from 1950 *(F1's inaugural season)* to 2023 *(The most recent complete season)*. The data is compiled from the commonly used community-run Ergast API.

4. Outcome

The outcome, being an interactive web app loaded with the described dataset, should contain two pages for different aspects of analysis:

1. **A season page**, where users can select a given season to analyse, with options for viewing both driver and team standings evolve over time, with each race detailed including results and other events *(e.g. crashes or cancellations)*.
2. **A driver page**, where users can view and compare drivers among various measures. These include age, nationality, points, positional results *(including podiums and wins)*, and team. This will also contain some aspects of team analysis.

Interfaces should be made to allow for computing detailed analysis, including ranking total and average points per season for teams and driver, win proportions, and rate of failures. Tooltips should be provided for both inputs and results to better aide user understanding in the systems provided.