**Project Report: F1 Analysis Tool**

**Samuel Mata ∙ 300574763 ∙ DATA-472 ∙ Final Project Report**

**Table of Contents**

[1. Abstract 1](#_Toc168953160)

[2. Introduction and Motivation 2](#_Toc168953161)

[2.1 Goal 2](#_Toc168953162)

[2.2 Target Audience 2](#_Toc168953163)

[2.3 Use Case 2](#_Toc168953164)

[3. Data Overview 3](#_Toc168953165)

[3.1 Source 3](#_Toc168953166)

[3.2 Use 3](#_Toc168953167)

[3.3 Structure 3](#_Toc168953168)

[3.4 Use 3](#_Toc168953169)

[4. App Overview 4](#_Toc168953170)

[4.1 Project Structure 4](#_Toc168953171)

[4.2 App Structure 4](#_Toc168953172)

[4.3 Interface 4](#_Toc168953173)

[4.4 Output 4](#_Toc168953174)

[5. Insights 4](#_Toc168953175)

[5.1 Learnings 4](#_Toc168953176)

[5.2 Challenges 5](#_Toc168953177)

[5.3 Comparison to Initial Goals 5](#_Toc168953178)

[5.4 Future Changes 5](#_Toc168953179)

[6. Conclusion 5](#_Toc168953180)

[6.1 Project Summary 5](#_Toc168953181)

[6.2 Insights 5](#_Toc168953182)

# Brief

This project developed an interactive web-app allowing users to perform customized analysis on various aspects of current Formula 1 data. The final outcome, was largely successful in project aims, can be accessed online at [INSERT LINK HERE]. This report covers various aspects of the completed project, including initial goals, data sources, the application itself, and insights made from the development process.

# Introduction and Motivation

Briefly describe the scope of your application and who you would expect to use it. Explain why you decided on this project and why it was a worthwhile undertaking.

Formula 1 *(more commonly referred to simply as F1)* is an international motorsport league. Specifically, F1 is the highest class of open-wheel *(where wheels are exposed rather than internally contained within the chassis)* single seater racing, and widely seen as the most prestigious and popular form of motorsport.

Motorsports in general tends to have more complex procedures and rules compared to more traditional sports events. This is especially apparent in F1, mostly due to the highly technical nature of building and racing in such complex vehicles. Furthermore, the exact rules and specifications for F1 championships and entrants has changed drastically and continuously with various rule and regulation changes since its inaugural 1950 season.

It is this complexity and highly technical nature that lends towards a sport that is highly data-oriented yet often difficult to understand. In this context, this project *(simply titled “F1 Analysis tool”)* aims to provide a solution.

## Goal

The goal of this project is to utilize the large quantities of collected data surrounding F1 to create an interactive tool allowing users to manage and visualize various aspects of the sport. Specifically, this involves the creation of an interactive web application *(built with* [*R*](https://www.r-project.org/) *and* [*Shiny*](https://www.rstudio.com/products/shiny/)*)* to allow for users to perform queries and analysis on current and historical F1 data. The tool should be accessible to those who may not be familiar with F1 by providing a simple interface for various types of user-selected analysis.

## Target Audience

As the project aim is mostly focused on giving a tool for custom broad high-level analysis of F1, the target audience includes anyone who may need to analyse the sport. This includes those who may be unfamiliar with F1 and want to learn more about the sport, and those who need customized analysis of the sport. This userbase is broad and contains groups with varying degrees of domain knowledge on both statistics and F1, so interpretability and ease of use are highly important.

## Use Case

Talk about what app should do

# Data Overview

Describe the data used by your application. Include the source of the data and any steps you undertook to prepare the data.

Finding a suitable data source is crucial to allow for a wide range of accurate analysis to be performed. Such a dataset needs to be accurate, extensive, and publicly accessible for this project to be viable. Thankfully datasets are readily available for nearly every detail of F1 due to the inherently high levels of engineering and regulation.

## Source

This project uses the [Formula 1 World Championship (1950 – 2023)](https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020) 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](http://ergast.com/mrd/), which is updated live as seasons progress.

This dataset lists a multitude of records, including current and former circuits, constructors *(teams)*, drivers, season results, race results, lap times, pit stop timings, qualifying times, disqualifications, and crashes.

## Use

The aforementioned dataset provides 14 CSV *(comma-separated-value)* files, of which [THIS MANY] were used. [WHAT WAS CUT AND KEPT]

While various checks were made throughout development to confirm the authenticity and accuracy of the data, no errors were found. As such, no preparation steps were required for data cleaning or preprocessing.

## Structure

FIGURE OUT WHAT IS NEEDED FIRST

|  |  |
| --- | --- |
| File Name | Features |
| circuits.csv |  |
| constructor\_results.csv |  |
| constructor\_standings.csv |  |
| constructors.csv |  |
| driver\_standings.csv |  |
| drivers.csv |  |
| lap\_times.csv |  |
| pit\_stops.csv |  |
| qualifying.csv |  |
| race.csv |  |
| results.csv |  |
| seasons.csv |  |
| sprint\_results.csv |  |
| status.csv |  |

# App Overview

Describe the main functionality of your application. Explain what the user can do and what information the application will display in response. Explain any main design decisions you made e.g. why did you chose particular inputs and outputs? You should illustrate this section with screenshots of your application.

DO THIS LAST

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.

## Project Structure

FILE STRUCTURE

LIBRARIES USED

GENERAL CODE STRUCTURE

## App Structure

The application is split into 3 tabs, each accessible via a so-called hamburger menu located next to the title.

1. **Season** -
2. **Driver** -
3. **Constructor** -

## Interface

Several user interface *(UI)* elements were incorporated to allow users to customize the output of the visualisations.

The Season tab contains the most UI, with 3 drop down menus:

1. **Championship** – Allowing selection between either the drivers or constructors championship.
2. **Season** – Allowing selection of F1 season by year, ranging across the dataset from 1950 *(the inaugural F1 season)* to 2023 *(the most recent complete season)*.
3. **Feature** – Allowing selection of the feature to be analysed. This can be points, race wins, podiums *(referring to a top 3 race finish for a driver/team)*, and finishes *(referring to a completed race where the driver/team does not crash, become lapped, or otherwise not finish normally).*

These can be seen in Figure 1 below:

A screenshot of a notebook

Description automatically generated

Figure 1: Season Tab Input UI

TALK ABOUT DRIVER AND CONSTRUCTOR TABS

## Output

GRAPHS

# Insights

Explain what have you learned while writing this application. Describe the main challenges you faced and how you overcame them e.g. challenges with obtaining the data, challenges with writing the application. Did your application turn out as you expected? Is there anything you would change if you were to start anew? Are there any extensions to the application that you would like to have made if you had more time?

## Learnings

## Challenges

Several areas of this project present

## Comparison to Initial Goals

## Future Changes

# Conclusion

## Project Summary

## Insights

Give a brief summary of your project and the insights gained from it.