

# Project Proposal

Federico Perotto, 1701738

## Abstract

The idea is to create a dashboard for Formula 1 races that provides insight into drivers' statistics over the race, aiding in identifying factors that contributed to the race result, such as strategy choices, change in weather conditions, or outlier performances.

## 1 Dataset

The dataset would be obtained by through the FastF1 API: [docs.fastf1.dev/](https://docs.fastf1.dev/). FastF1 provides lap timing, car telemetry, position, tyre data, weather data, the event schedule and session results, mostly through pandas DataFrame objects. The data provided isn't perfect, nor do I intend to use all of it, so after some preprocessing to clean it up, it can be saved to a csv file (rather than relying on an API call and repeating the preprocessing at runtime). As for the AS index, since each race has a different amount of laps and a different time duration, it's hard to provide a precise number even after preprocessing, but as a lowerbound consider that the race with the least amount of laps this year has a total of 815 laps and 31 dimensions (columns), which even after the preprocessing should leave me with more than the required amount (especially if we also consider information not included in the Laps dataframe such as weather data and telemetry).

## 2 Specify the general idea (Analytics part, Visual Part)

The main focus of the dashboard is to gain insight and compare the performances of the drivers in a race. The user will have to choose a race from a dropdown menu, and for each race there will be the following visualizations:

- A linechart (or connected scatterplot) that visualizes the progression of the race by displaying the time gaps between the drivers over the laps of the race. This way it should be easy to visualize a good performance as a line that gets progressively closer to the drivers ahead (or leaves other drivers behind). This chart will also include other information such as weather data, track status (safety car, yellow flags, ...), tyre compounds used and lap times.
- A bar chart to visualize the tyre strategy of every driver. This visualization should allow for an immediate comparison of the strategies chosen, as well as their implications in the race (an early pitstop that forced a driver to stay on a tyre compound for longer than intended, a gamble in a compound choice, etc).
- A parallel coordinates plot to compare several individual performance metrics for each driver, such as lap time consistency, average speed, pit stops duration, etc
- A scatterplot for the dimensionality reduction, which ideally will cluster drivers that performed similarly.

The user can then interact with each of these visualizations to refine the analysis of a driver's performance or to compare the performance of two or more drivers.

## 3 Specify the intended user

- Race commentators, pundits and content creators, who can use the dashboard to verify their initial intuition, analyze a race and then formulate opinions and content (articles, videos, interviews) that have data to support their claims.
- Formula 1 enthusiasts and fans, who can go back and analyze the performance of the drivers to a greater extent than what is possible during the live broadcast (which often can only focus on a handful of drivers).

## 4 Specify the used analytics

The main analytics will be in the dimensionality reduction section, but the project will also compute other measures that aren't readily available in the dataset, such as the performance metrics in the parallel coordinates plot, and the time deltas between drivers in the lineplot.

## 5 Specify the various characteristics in relation to the Visual Analytics cycle

The goal is to have all the visualization update after every interaction with the system. The user will be able to decide which drivers to focus on, as well as which part of the race to focus on, and the plots will update accordingly. For example the user can brush over the lineplot to select a specific interval of laps to look at, or they can select a tyre stint from the bar plot, and the other visualization will update to include only those laps.

## 6 A mockup of the user interface

