Visual Analytics Project Proposal

Angela Di Giampaolo
Sapienza University of Rome
Student ID: 2205621
digiampaolo.2205621@studenti.uniroma1.it

Bogdan Andrei Tutuianu
Sapienza University of Rome
Stuent ID: 2217481
tutuianu.2217481@studenti.unitoma1.it

I. Goal

The goal of this project is to develop a dashboard useful for analysis of Formula 1 Races.

II. Dataset

We chose a dataset from Fast F1 (https://docs.fastf1.dev/), that provides access to many data about Formula 1 races, with an interesting focus on car telemetry, weather and lap times information. Given the immensity of the dataset, that contains information about races from 2018 to 2025, we decided to focus only on Season 2025, so that the amount of data to manage is conform to the requirements of the project related to the AS Index. We retrieved the needed data for the analysis through FastF1 APIs, and we stored them into a single CSV file composed of 26692 rows and 25 columns. Each row describes laps performed by every single driver taking part to the race, and we have information about all the races held in 2025. More in detail, the information we stored is the following:

- Driver data (Name and Team)
- Lap data (Lap Number, Lap Time, Lap Start Time)
- Tyre data (Compound, Tyre Life)
- Data about driver speed and driver position during the lap
- Times registered by the driver in Track Sectors 1, 2 and 3
- Event Name, Country in which the event is held and Year
- Weather Data (Air Temperature, Humidity, Rainfall, Wind Speed, Wind Direction)
- Track temperature

The AS Index for the resulting dataset is $26692 \times 25 = 667300$. In order to perform a more feasible and manageable analysis, we considered the option consisting of allowing the users to choose the race (among the ones of Season 2025) for which they would like to visualize data and then represent them. In this way, the analyitics and the visualization parts will consider only a portion of the entire dataset for which the AS Index will be around the value 25000.

III. General Idea

The main idea of this project is to implement a dashboard for Formula 1 fans that allows users to visualize information about drivers' performance and strategies during a race. Thanks to this interface, users will be able to compare drivers' strategies and analyze how team decisions can impact on drivers' lap times, position and final results. It will be also possible to obtain derived information from the original dataset, so that the user can access to more complex insights.

IV. Visual Analytics Cycle

A. Analytics

Given the presence of several variables impacting on a race, we will perform a Principal Component Analysis and we will also apply a clustering algorithm in order to identify similarities and dissimilarities. Furthermore, we will also compute other statistics based on the elements selected by the user and we will represent them in some graphs. For example, we may compute average speed time for each sector or we may also perform mathematical operations in order to compute the tyres degradation. In this way, it will be possible to look inside single strategies and to see their overall impact on the race.

B. Visualization

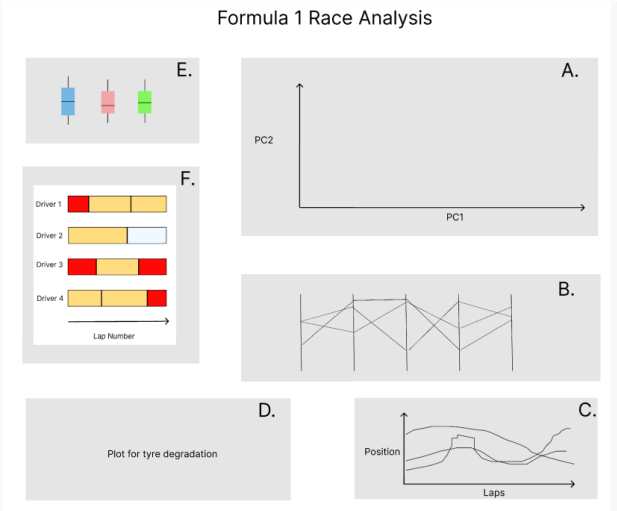

Fig. 1. Mockup of the user interface

Figure 1 represents a mockup of the interface. In particular, we will implement the following visualizations:

- A: a scatter plot representing the result of clustering, PCA, projection on two dimensions applied to the initial dataset

- B: a parallel coordinates representation that describes drivers' behavior considering their times in sectors and other performance elements
- C: evolution of drivers' position/times during laps
- D: visualization of tyre degradation
- E: box plot related to driver or team performances during the race
- F: bar chart describing the strategy adopted by the driver during laps

C. Interaction

First, the user will select some elements from plot A, where each point represents a single lap performed by a single driver during the race. Then, after the selection, all the other graphs (B, C, D, E, F) will change by computing and representing the respective information considering the drivers/lap selected by the user.