Formula 1: Data Visualization

Diogo Dória[1], **Mariana Paulino**[1]
University of Coimbra
[1]Student, Department of Informatics Engineering,
Faculty of Sciences and Technology
uc2020246139@student.uc.pt; uc2020190448@student.uc.pt

***Abstract*—** Formula 1 racing began in 1950 and is the world's most prestigious motor racing competition, as well as the world's most popular annual sporting series. The FIA Formula One World Championship usually runs from March to December spanning around 20 races in 20 countries across four continents. This is also one of the sport competitions that produces more data than others, and where analysis models are continuously being developed. The objective of this study is to demonstrate the application of different visualization models in the performance analysis of Red Bull drivers during the last four seasons and conclude why some have better performance than others.

***Keywords*—** Formula 1, Data Visualization, Data Analytics, plotly, Dash.

I. Introduction

The FIA Formula One World Championship is the top automotive competition since its inaugural season back in 1950. A Formula One season consists of a series of races, known as Grand Prix, which take place worldwide on both purpose-built circuits and closed public roads. A classification scheme, based on a points system is used at each Grand Prix to determine two Annual World Championships: one for the drivers, tshe other for constructors.[1]

The objective of this paper is to study Max Verstappen's performance alongside his teammates since 2019, in order to understand why Sergio Perez, as teammate, led Verstappen to achieve the title in 2021 and 2022. We will also demonstrate why Pierre Gasly and Alex Albon weren't able to do the same thing as Perez, situation that led to their demotion to second line Teams [7].

We'll be comparing Gasly's and Albon's performance with Verstappen's, studying the meaning of the Gasly demotion for Albon's promotion [2][3]. In 2020, we'll analyse Sergio Perez's and Alex Albon's performance, and validate for Pierre Gasly' signs of improvement [4][5]. Finally in 2021 and 2022 we will compare Sergio Perez and Max Verstappen performances [6]. These abovementioned comparisons about performance and consistency, comprises analysing every season's Grand Prix, showcase position gained/lost, lap times, pitstops that influence in the race strategies, fastest lap telemetry, points scored, race results, and championship standings.

From the literature reviewed, we were inspired by the models used for the visualization of times and points. For the lap times and seasons points we performed a line scatter plot and for the final positions we used a bar graph.

For our application, we wanted to create a user-friendly dashboard, easy to select the season, the race and type of visualization, if it is about Lap times, Final Race Position, Season Points or Championship Standings.

The next section explores the different analyses made with historical Formula 1 data and how this paper relates to it. On section 3 we will describe the dataset provided, alongside the tools used for data preparation and implementation of the models. We will focus on the methodology followed, the Setup and how the visualization to our methods were conducted. Section 4 presents the design requirements, the model chosen and the implementation of our application. Finally, in last section we will perform a critical reflection of our work, and we will present our conclusions and suggestion of improvements.

II. Related Work

In this section we investigate visualizations of various Formula 1 data analysis, proposed by various authors using *plotly* and *sklearn* libraries.

Reference [8] is a tutorial which describes how to use Formula One data for analysis, covering data capture, cleaning and analysis.

As a continuity of the previous tutorial [8], Reference [9] introduces several potential analyses using F1'1 historical data.

Reference [13] proposes an explorative analysis of F1 to break down some historical data on circuits, drivers, and racing data to understand what makes them the best or the worst.

Reference [14] is focused on the 2021 season, comparing only two drivers, in this case, Lewis Hamilton and Max Verstappen. This being one of the most scratched championships recently. Its analysis includes Lap Analysis, Race Results, Pitstop Analysis, Conversions and Fastest Lap telemetry.

Reference [15] proposed a website, called F1 Tempo, which allows to graph telemetry data from the FastF1 python package. We used this as inspiration for our application.

III. **Description**

In this section, a description of the data available and their sources is given in A- Data description and in B- Framework and methodology is where various frameworks are presented and how they relate to the data abovementioned.

A. Data collection and sources

In our study, we will be using public domain Formula 1 Worlds championship (1950 – 2022) dataset from Kaggle [10]. Which is compiled from Ergast Developer API.[15]

The Kaggle dataset consists of 14 csv files:

- Circuits: where F1 races are held
- Constructor Results: Race results of the constructor's championship
- Constructor Standings: Final standings of the constructor's championship
- Constructors: Information on all current and previous constructors
- Driver Standings: Final standings of the driver's championship
- Drivers: Information on all current and previous drivers
- Lap Times: Lap times of all completed laps by all drivers in all events
- Pit Stops: All pit stops made, when and duration (pit in -pit out)
- Qualifying: Results of all qualifying sessions, including the separate Q1, Q2 and Q3 sessions.
- Races: Races where F1 raced.
- Results: Both constructor and driver
- Season: Seasons of F1
- Results of F1 sprint races
- Status of various statuses

As additional data to our study, we will use the data provided by FastF1 python package [11][12]. This package allows to access and analyse Formula 1 results, schedules timing data and telemetry. The module is designed around *Pandas*, *Numpy* and *Matplotlib*, which makes it easy to use, and at same time offers several possibilities for exploring data analysis and visualization.

B. Frameworks and Methodology

Our work starts by reviewing and preparing the data. During this process, we created *Panda* data frames from joining the different datasets available in *CVS* format. We have cleaned the data, and identify which attributes were the more relevant for our analysis and visualization models.

Our first data frame (*df*) is a merge of lap times with some attributes from driver's *CSV*, such as driverid, code and driverRef. Additionally, we merged the resulting data frame with races attributes such as racesID, name (race name), date and year.

Some attributes had to be created based on the existing ones, such as lap times (df_laps).

To analyse driver's position in the championship, we needed to create a table which measure is the total points earned by a driver in a Grand Prix season. In order to achieve this, we created a data frame (Season Points) by merging drivers with driver's standings and other races attributes.

With the data obtained, it is possible to see how many points each driver scored in each Grand Prix.

The next step was to have an overview of driver position after each Grand Prix. For this we created a data frame called *df* which consist of merging drivers and other races attributes. We had to drop duplicates and solve some missing values ("N" stands for *did not finish the race*), missing values stands to 0 points.

Our study focuses in comparing lap times, final race position and championship points scored, by each Grand Prix.

With our data ready, we grouped it by seasons and the drivers object of our study (Max Verstappen, Sergio Perez, Pierre Gasly and Alex Albon.

Using *plotly* Graph Object, we implemented line scatter plots of lap times and points scored for each race and bar graphs for final race position and championship standings.

Further details about design and implementation are given on the next section 4.

IV. **Design and Implementation**

This section focuses about design requirements, design options, setups, and application's implementation details.

A. Design

We have considered two visualization models *Scatter Plots and Bar Charts*, for which we created specific layouts:

- Lap Times
- Race Results

Two filters have been added using dropdown menus:

- Season

- Race

Scatter Plots [17]

Lap Times – To visualize this information in a line scatter plot, we choose a layout with a more defined axis and a soft grid background.

To build this graph, we have selected from the data frame, the following information:

- For x axis – Lap
- For y axis – Lap duration in seconds [16]

Each marker represents the time that took the driver to complete a lap (x axis). Each driver has a different color, facilitation its identification.

For the x axis, times are displayed with a 5-lap interval. Figure (1) can be found after section 6.

In case of that a driver retires or takes much time to complete a lap due to a vehicle malfunction, the data is out the range of the other drives, constituting an outlier. This situation impacts the visualization of the graph.

In order to address this situation, we tried to implement a solution which limited the data to a specified time range. We couldn't implement this solution. Figure (2) can be found after section 6.

Points Scored - For the points scored during the season, we have chosen another layout for this type of results, for which we have maintained the defined axis and the soft grid background.

To build this graph, we have selected from the race results data frame, the following information:

- For x axis – Race name and date
- For y axis – Final race position

Each marker represents the total points that the driver finished each Grand Prix.

Each driver has a different color easing the identification. Figure (3) can be found after section 6.

Bar Charts [18]

Championship Standings – In this chart, we show the final championship standings for all drivers.

To build this graph, we have chosen Race Results layout, for which we have maintained the defined axis and the soft grid background.

We have selected from the race results data frame, the following information:

- For x axis – Race name and date
- For y axis – Final race position

As we are showing all drivers, we assigned a neutral color. Figure (4) can be found after section 6.

Final Race Position - For these results of this session, a grouped bar cart has been chosen to display the selected driver's final position of each Grand Prix.

Using the layout results, maintaining the defined axis and the soft grid background, we have built a graph with the following axis:

- For x axis – Race name and date
- For y axis – Final race position

based on the information we have extracted from the Position Results data frame.

Each driver has a different color, facilitation its identification. Figure (5) can be found after section 6.

B. Implementation

The application that was built, is a site web with the following areas/pages:

- Home – where it is presented a description about the drivers, F1 and this project.
- Races – where the information about the laps time of each race for each 4 pilots is is presented. Figure (7)
- Season Standings – where it is possible to visualize the score of points, according to the season chosen. Figure (6)
- Race Standings – where it is possible to visualize the final position of each race, according to the season chosen. Figure (8, 9)
- Navigation Menu – has links to each abovementioned page, and a link to the official F1 site.

Then the option of that radio has the list of races of the year associated and the visualization is a result of the user's choice.

V. Conclusion and Reflection

The goal of this study was to implement a dashboard where the user can analyze lap times, final race position and driver's standings. With our visualizations it is possible to tell which driver has the best performance by season and race. In this case we can confirm that Max Verstappen was always consistent, about his teammates we can tell that Pierre Gasly demotion form Red Bull primary team was a mistake indeed. Although it is possible to confirm that Sergio Perez is a great fit for the team which helped Max Verstappen achieve two World Driver Championships.

Finally, we think that sorting the races by date in some visualization and solving the situation about the lap times outlier would improve the effectiveness of our study.

VI. References

[1] Formula 1: https://en.wikipedia.org/wiki/Formula_One

[2] Analysis Why Red Bull made their latest blockbuster driver swap, 2019: https://www.formula1.com/en/latest/article.analysis-why-red-bull-made-their-latest-blockbuster-driver-swap.5OSpVBQ1sdbxO5VEj80ykZ.html

[3] Albon vs Gasly – The Road to Red Bull's driver swap, 2019: https://www.formula1.com/en/latest/article.albon-vs-gasly-the-road-to-red-bulls-driver-swap. 6sAlWq2IHA6D93c60BwHJi.html

[4] Helmut Marko 'really happy with my results' says Gasly – but possibility of Red Bull return 'not my call': https://www.formula1.com/en/latest/article.helmut-marko-really-happy-with-my-results-says- gasly-but-possibility-of-red.6IP6CQyfOTbzxU1txq8cAS.html

[5] Perez to partner Verstappen at Red Bull in 2021, as Albon becomes reserve driver: https://www.formula1.com/en/latest/article.breaking-perez-to-partner-verstappen-at-red-bull-i n-2021-as-albon-becomes.21qHfmHAyfzAjVHT3PfVBd.html

[6] Verstappen dubs Perez an amazing human being as he credits mexican's heroics: https://www.formula1.com/en/latest/article.verstappen-dubs-perez-an-amazing-human-being- as-he-credits-mexicans-heroics.1czF8vWLDo1PSkKzuDStco.html

[7] Analysis: Why Red Bull and Sergio Perez extended their marriage until the end of 2024: https://www.formula1.com/en/latest/article.analysis-why-red-bull-and-sergio-perez-extended- their-marriage-until-the-end.4qaSjHCbvKAAK2EWIQtYrk.html

[8] Getting Started with F1 statics, Leo van der Meulen, 2022: https://blog.devgenius.io/getting-started-with-ergast-f1-statistics-and-python-5112279d743a

[9] Analyzing F1 Statistics – Part 1, Leo van der Meulen, 2022: https://blog.devgenius.io/analyzing-f1-statistics-part-i-a526d15b6fc8

[10] Formula 1 World Championship (1950 – 2022): https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?resour ce=download&select=circuits.csv

[11] Fast F1 Documentation : https://theoehrly.github.io/Fast-F1/

[12] Fast F1 Examples Gallery: https://theoehrly.github.io/Fast-F1/examples_gallery/index.html

[13] Formula 1 a visual explorative analysis: https://www.kaggle.com/code/akhilreddy9554/formula-1-a-visual-explorative-analysis

[14] F1 Tempo: https://www.f1-tempo.com

[15] Ergast API: http://ergast.com/mrd/

[16] Formula 1 interval meaning: https://www.radiotimes.com/tv/sport/formula-1/formula-one-interval-meaning/

[17] Plotly Scatter plots - https://plotly.com/python/line-and-scatter/

[18] Plotly Bar charts - https://plotly.com/python/bar-charts/

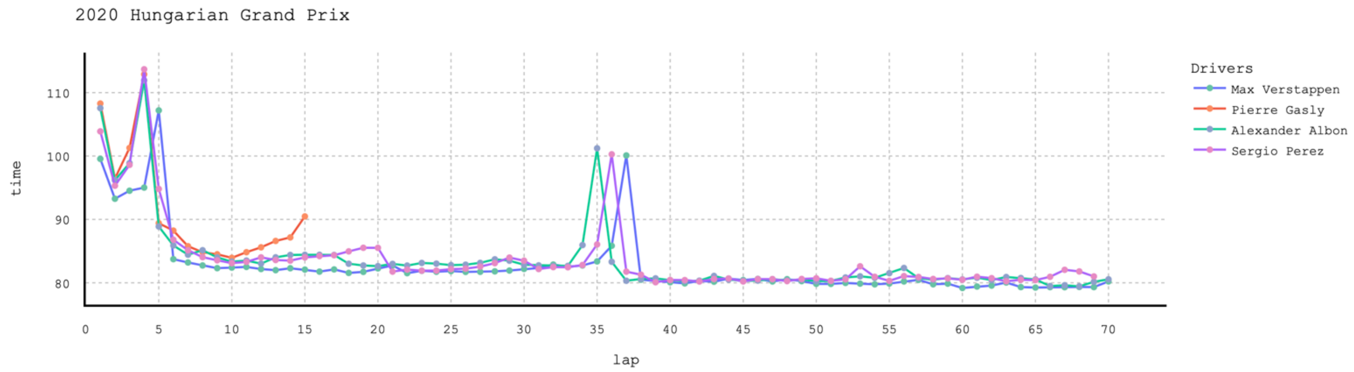


Fig. (1) – Lap Times

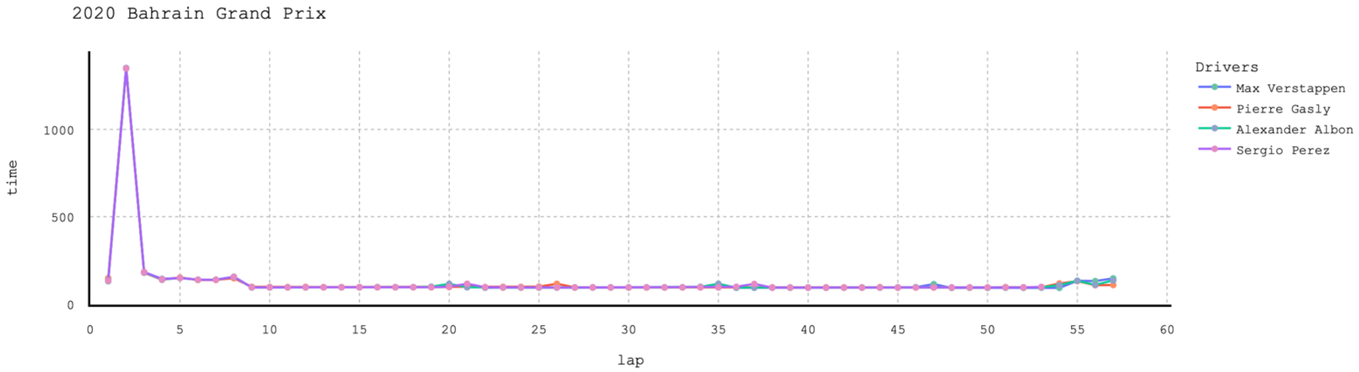


Fig. (2) – Lap times 2020 Bahrain Grand Prix accident

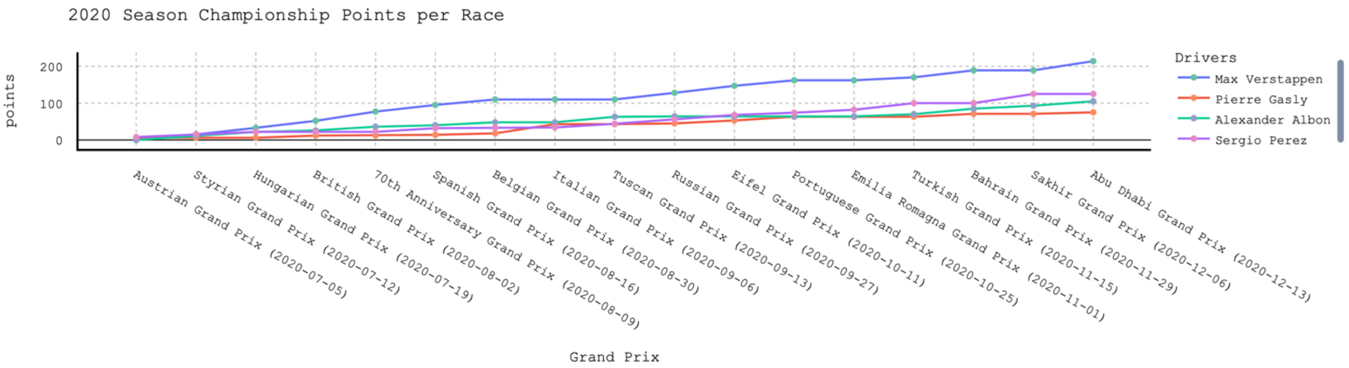


Fig. (3) – Season Points

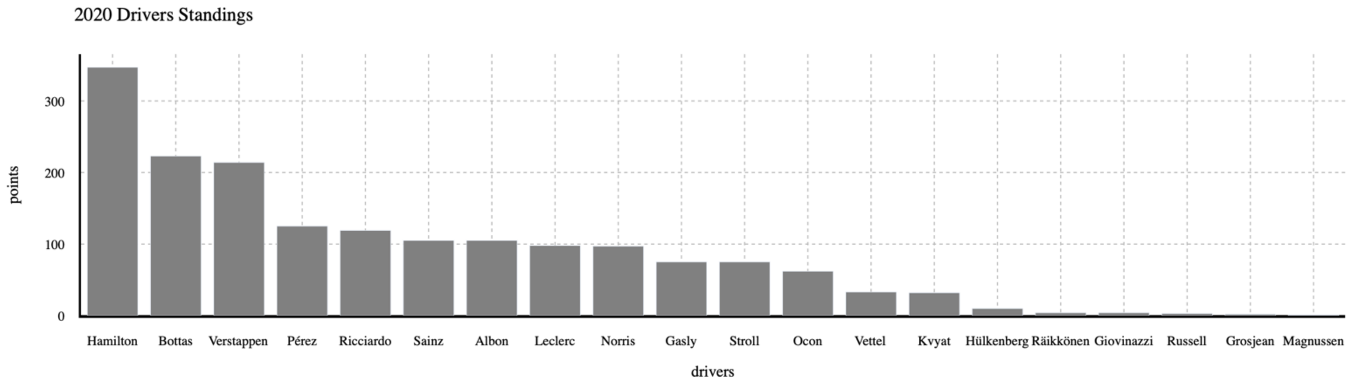


Fig. (4) – Championship Standings

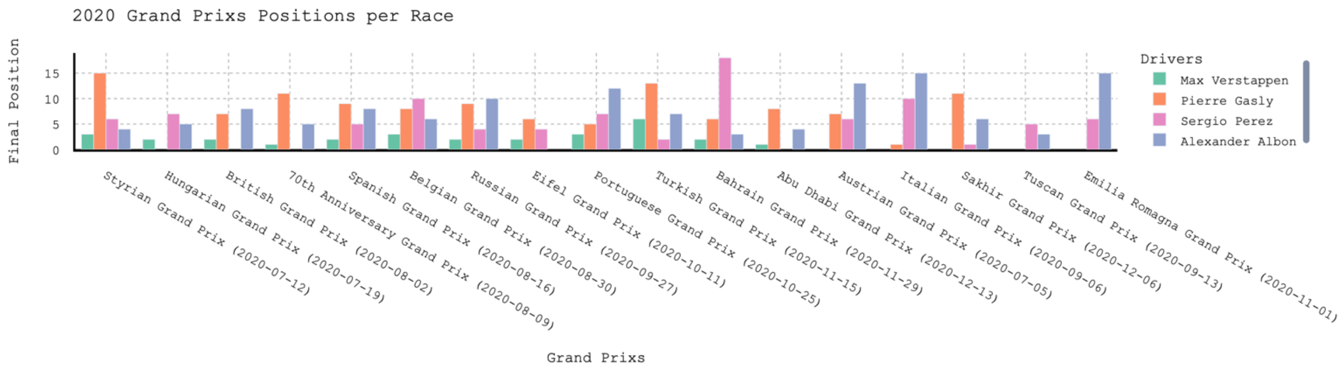
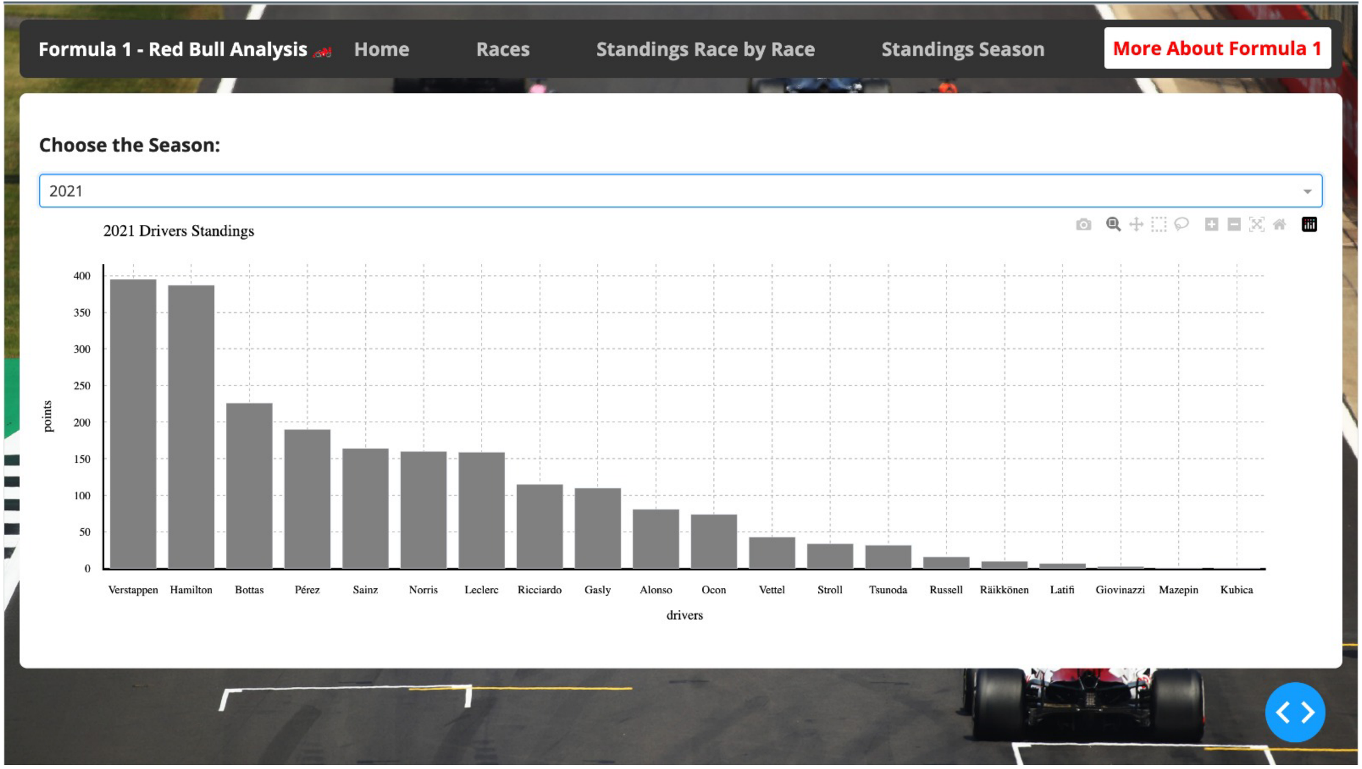


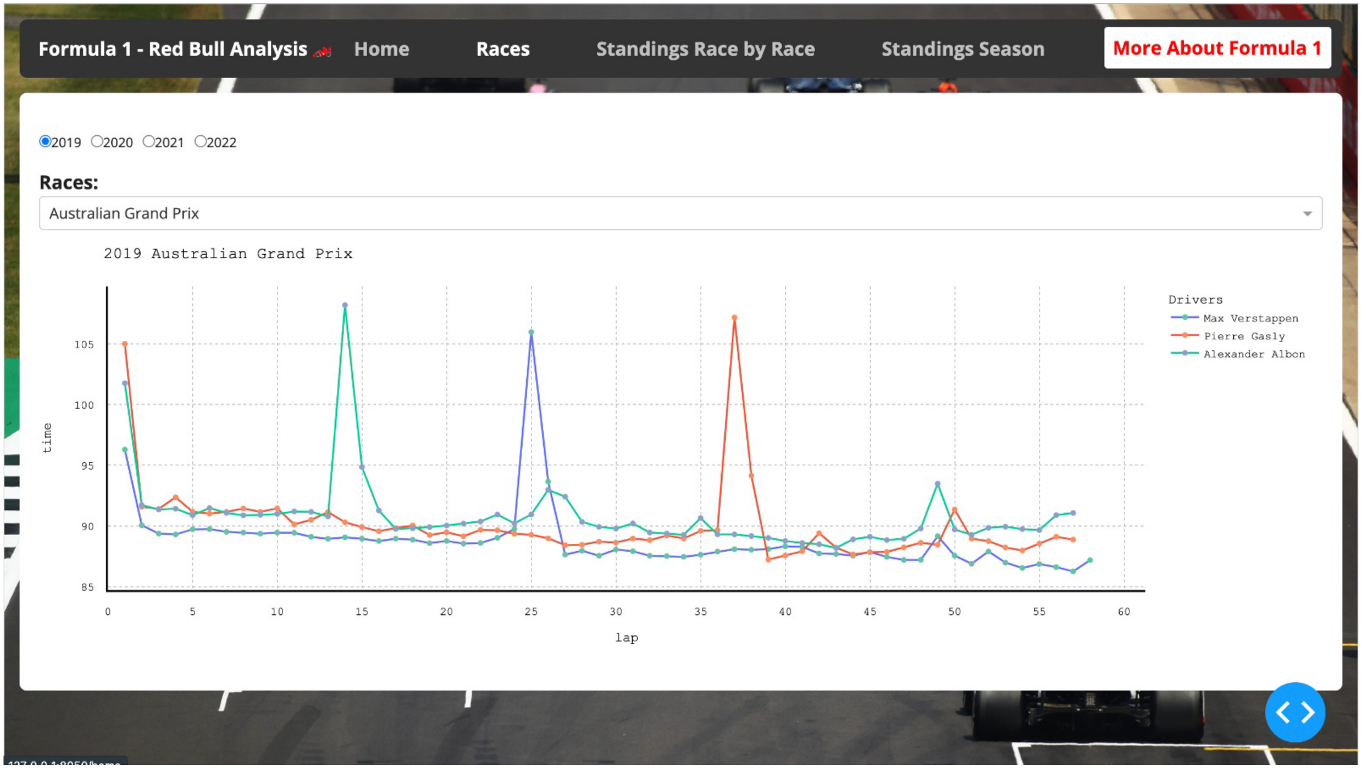
Fig. (5) – Final Position of each Grand Prix



Choose the Season:

2021

Fig. (6) – Championship Standings application

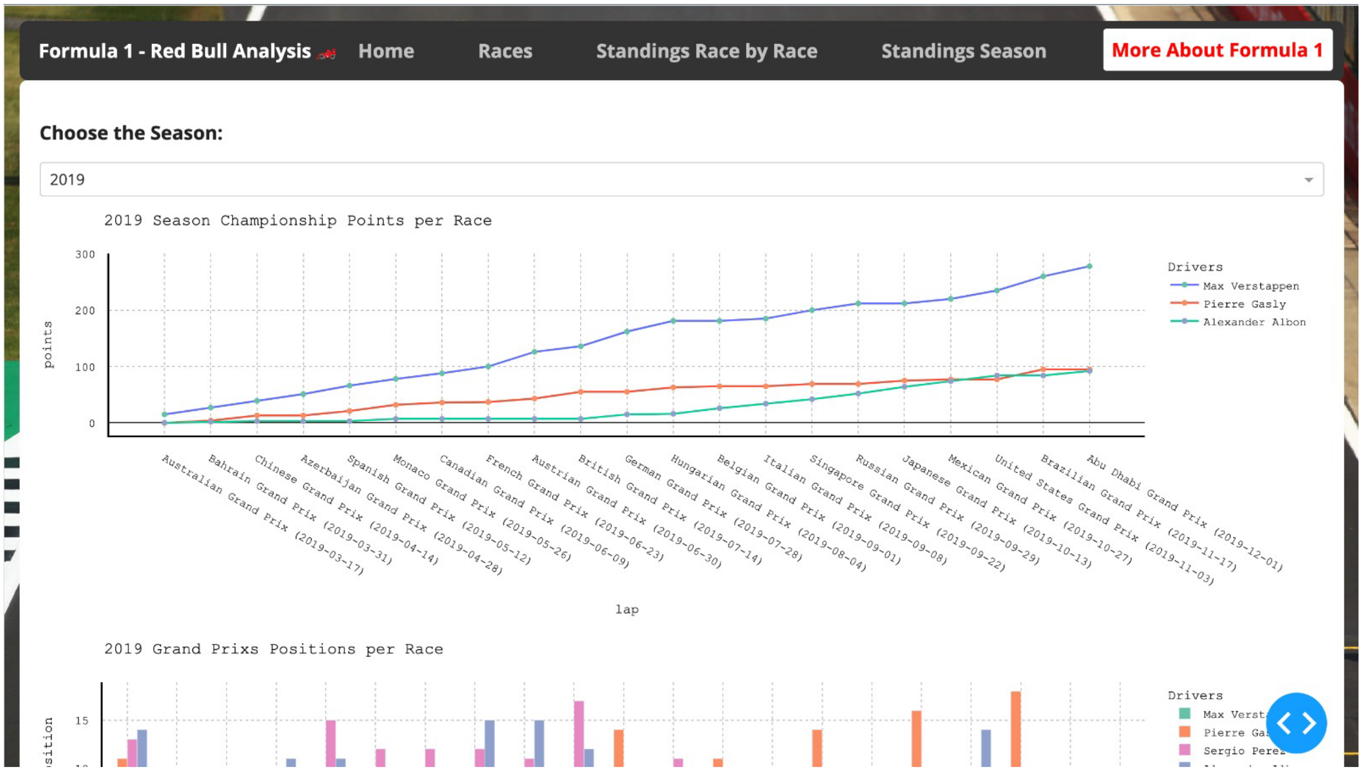


● 2019 ○ 2020 ○ 2021 ○ 2022

Races:

Australian Grand Prix

Fig. (7) – Races application


Formula 1 - Red Bull Analysis
Home
Races
Standings Race by Race
Standings Season
More About Formula 1

Choose the Season:

2019

2019 Season Championship Points per Race
points
300
200
100
0
Australian Grand Prix (2019-03-17)
Bahrain Grand Prix (2019-03-31)
Chinese Grand Prix (2019-04-14)
Azerbaijan Grand Prix (2019-04-28)
Spanish Grand Prix (2019-05-12)
Monaco Grand Prix (2019-05-26)
Canadian Grand Prix (2019-06-09)
French Grand Prix (2019-06-23)
Austrian Grand Prix (2019-06-30)
British Grand Prix (2019-07-14)
German Grand Prix (2019-07-28)
Hungarian Grand Prix (2019-08-04)
Belgian Grand Prix (2019-09-01)
Italian Grand Prix (2019-09-08)
Singapore Grand Prix (2019-09-22)
Russian Grand Prix (2019-09-29)
Japanese Grand Prix (2019-10-13)
Mexican Grand Prix (2019-10-27)
United States Grand Prix (2019-11-03)
Brazilian Grand Prix (2019-11-17)
Abu Dhabi Grand Prix (2019-12-01)
lap
Drivers
Max Verstappen
Pierre Gasly
Alexander Albon
2019 Grand Prixs Positions per Race
sition
15
Drivers
Max Verst
Pierre Ga
Sergio Perez

Fig. (8) –Season and Race Results application

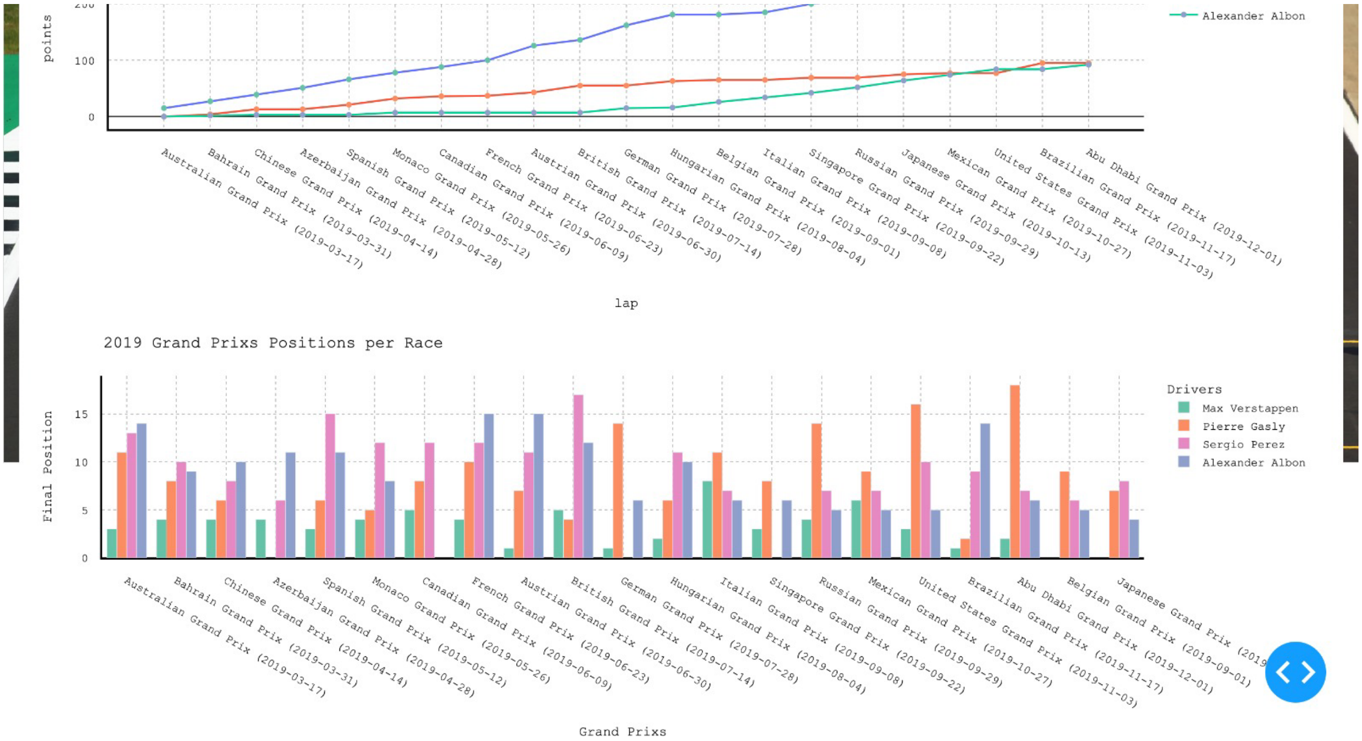

200
points
100
0
Alexander Albon
Australian Grand Prix (2019-03-17)
Bahrain Grand Prix (2019-03-31)
Chinese Grand Prix (2019-04-14)
Azerbaijan Grand Prix (2019-04-28)
Spanish Grand Prix (2019-05-12)
Monaco Grand Prix (2019-05-26)
Canadian Grand Prix (2019-06-09)
French Grand Prix (2019-06-23)
Austrian Grand Prix (2019-06-30)
British Grand Prix (2019-07-14)
German Grand Prix (2019-07-28)
Hungarian Grand Prix (2019-08-04)
Belgian Grand Prix (2019-09-01)
Italian Grand Prix (2019-09-08)
Singapore Grand Prix (2019-09-22)
Russian Grand Prix (2019-09-29)
Japanese Grand Prix (2019-10-13)
Mexican Grand Prix (2019-10-27)
United States Grand Prix (2019-11-03)
Brazilian Grand Prix (2019-11-17)
Abu Dhabi Grand Prix (2019-12-01)
lap
2019 Grand Prixs Positions per Race
Final Position
15
10
5
0
Drivers
Max Verstappen
Pierre Gasly
Sergio Perez
Alexander Albon
Australian Grand Prix (2019-03-17)
Bahrain Grand Prix (2019-03-31)
Chinese Grand Prix (2019-04-14)
Azerbaijan Grand Prix (2019-04-28)
Spanish Grand Prix (2019-05-12)
Monaco Grand Prix (2019-05-26)
Canadian Grand Prix (2019-06-09)
French Grand Prix (2019-06-23)
Austrian Grand Prix (2019-06-30)
British Grand Prix (2019-07-14)
German Grand Prix (2019-07-28)
Hungarian Grand Prix (2019-08-04)
Italian Grand Prix (2019-09-08)
Singapore Grand Prix (2019-09-22)
Russian Grand Prix (2019-09-29)
Mexican Grand Prix (2019-10-27)
United States Grand Prix (2019-11-03)
Brazilian Grand Prix (2019-11-17)
Abu Dhabi Grand Prix (2019-12-01)
Belgian Grand Prix (2019-09-01)
Japanese Grand Prix (2019
Grand Prixs

Fig. (9) –Season and Race Results application