DataOne: A Formula 1 Comprehensive Analysis
DSCI 590 : Data Visualization: Project Paper

Team members: Shruthi Senthilmani, Akash Patil

Abstract

Embarking on a journey through the analysis of Formula 1, our comprehensive project serves as an in-depth exploration of the sport's evolution from its inaugural year in 1950 to the present-day landscape of 2023. Armed with data sourced from diverse platforms, including the Fast F1 Python package, Kaggle, GitHub, and the Ergast API, our analytical endeavor aims to unravel the intricacies of Formula 1 through a multifaceted lens.

The project aims to analyze how drivers perform on different circuits, compare teammates' performance across seasons, and track constructors' points over time. We also explore factors like lap times of teams and drivers, circuits and the corners of every circuit, corner speeds of drivers in every circuit and tires impact. In our pursuit of unraveling Formula 1's secrets, we scrutinize podium rates, shedding light on the success stories of both teams and individual drivers. Exploring the global representation within the sport, we examine the triumphs of nationalities, offering a nuanced perspective on the diverse tapestry that constitutes Formula 1. This comprehensive project aims to highlight various nuances of Formula 1, employing a combination of data visualization techniques using matplotlib, seaborn, plotly, Altair to make interactive dashboards to uncover the crux behind the patterns and trends that define Formula 1.

1 Introduction

Formula One (F1) is a well-known international motorsport that has been captivating racing enthusiasts since its start in 1950. F1 has become famous for its high-speed races, cutting-edge cars, and iconic circuits, gathering a massive global following for seven decades. In the dynamic realm of Formula One, where speed, precision, and strategy converge, the integration of data-driven insights has become a pivotal aspect of unraveling the intricacies of the sport. Building upon the rich legacy that spans over seven decades, our project delves into the heart of Formula 1, leveraging a multifaceted approach to comprehensively analyze various facets of this high-stakes motorsport.

In this era of data-driven insights, even Formula One is not untouched by the power of data. Teams, fans, and analysts are increasingly relying on data to understand car and driver performance, as well as race strategies. To assist in this data exploration, we will be using the Formula 1 World Championship dataset available on Kaggle,GitHub, Ergast API and a robust Python package FastF1.

In this project, we aim to gain valuable insights into Formula 1 by comprehensively analyzing various aspects of the sport. Our goals revolve around understanding driver performance on different circuits, assessing how drivers fare against their teammates over a season, analyzing the points accumulated by constructors over time, and examining the impact of qualifying results on races. To achieve these goals, we will employ a range of visualization techniques, including interactive methods like Plotly and Altair, as well as traditional charts and graphs. Through these analyses, we seek to uncover patterns, weaknesses, and key trends within the world of Formula 1, shedding light on the factors that shape the outcomes of races and championships.

Our exploration begins as we transcend the race track to explore the global footprint of Formula 1, with a visual feast of different Grand Prix locations presented through Folium. Furthermore, our project delves into a meticulous examination of constructors' performances from 2021 to 2023, unraveling the evolving landscape of competition over time through stacked bar plots. We further chart the historical trajectory of championship victories by constructors since the inception of Formula

1 in 1950, encapsulating the enduring narratives of triumph and rivalry. Zooming into recent years, we intricately dissect the prowess of the top three constructors – Mercedes, Ferrari, and Red Bull – unraveling the subtle nuances that have shaped their trajectories in the past decade. Concurrently, we present a dynamic narrative of drivers' lap times and positions across circuits from 2019 to 2023, casting a spotlight on the evolving dynamics of individual and collective performance. This project also unveils the anatomy of each circuit, presenting a user-friendly interface with dropdowns that intricately detail corners and their significance in the racing narrative.

Qualifying results become a focal point, with bar plots delineating the drivers' performances relative to the leader in each race across seasons, offering insights into the nuances of pre-race dynamics.

Adding a temporal dimension, violin plots meticulously illustrate drivers' lap times across circuits, providing a nuanced understanding of performance variations. Joint plots bring the head-to-head comparisons of drivers within teams to life, unraveling the symbiotic dynamics that propel teams to success. This joint plot has a dropdown for every year starting from 1950-2023 and every team so we could compare two drivers and see if there is a Number 1 and Number 2 driver in a team!

Additionally, a boxplot that unpacks the distribution of lap times for constructors in every circuit in 2021, encapsulates the diverse challenges each track presents.

Finally, the project culminates, by taking the analysis into the live-action realm - an interactive animated plot gives a real-time visualization of the 2023 Las Vegas Grand Prix(the latest and first Grand Prix held in Las Vegas before the 2023 season ends), bringing the exhilaration of the race to the forefront.

Through this multifaceted exploration, our project illuminates the underlying patterns, strengths, and trends that define Formula 1, providing racing enthusiasts, teams, and analysts with a nuanced perspective on the factors shaping the outcomes of races and championships.

1.1 Motivation

Our project offers a unique approach to understanding Formula 1, a sport known for its complex engineering and ever-changing dynamics. Unlike traditional rulebooks, F1 lacks specific guidelines for race analysis. In F1, there is a lot of variation in different teams when it comes to engineering cars over different seasons and to understand the small changes in these requires in-depth knowledge about the mechanics of the car. This might be challenging for many people without prior knowledge. To overcome this hurdle, we use statistics and insights derived from team, driver, and circuit data. Through this, we aim to create interactive visualizations that enables individuals to grasp the subtle intricacies of the sport before even tuning in to watch a race. This analysis serves as a valuable resource, providing information on drivers' performances, competitors' statistics over the years, team success across various circuits, and the evolving team dynamics that can make or break each session. Beyond this, we delve deeper into the sport's different sessions, such as the qualifying rounds, and whether or not these rounds impact the significance of securing a strong position in the race. We recognize that while a team's performance is heavily influenced by the engineering of their car, evaluating the relative strengths of two top teams is not as straightforward as tallying wins or points.

In essence, our motivation is grounded in the belief that everyone should have the opportunity to engage with and appreciate the complexities of Formula 1, irrespective of their prior knowledge or expertise. Through our project, we aspire to break down barriers to entry and also to foster a connection between people without prior knowledge and the world of Formula 1, one interactive visualization at a time, by introducing the world of F1 sport and then helping them to understand the historical dominance and success of teams.

1.2 Background

There is a plethora of existing data visualizations of f1 and some of them were truly inspiring to start this project.

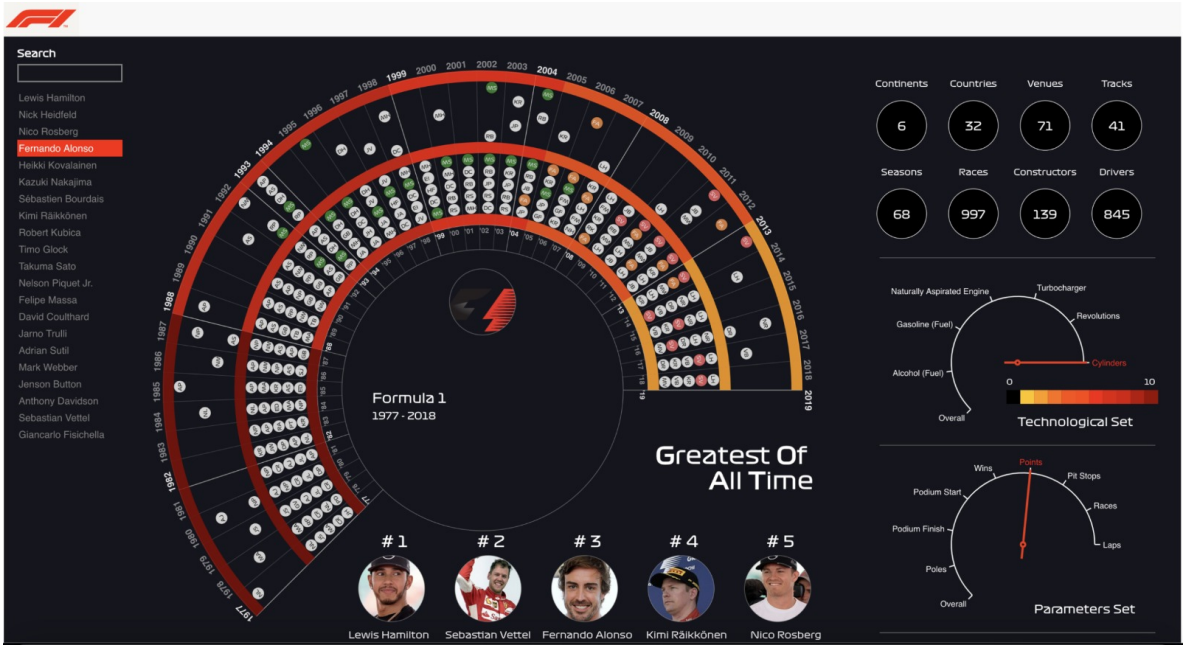


Figure 1: Interactive webpage

The Greatest of All Time (GOAT) Formula One data visualization by Jason J Paul[1]. The objective of this work was to find the greatest F1 driver of all time from 1977-2018 and he visualized this using an interactive web app featuring a design structure that included a main gauge. The main "gauge" is divided into two arcs. In the inner arc, circles with initials representing drivers represent the top 5 drivers for that particular year. In the outer arc, 5 driver circles spanning across a group of years (era) represent the top 5 drivers in that particular era. The colors used around the gauge represent the technological set defined in the dashboard. There is also a parameter set to decide the position of the drivers on the main gauge. The user can change the technological and parameter set to check for every driver.

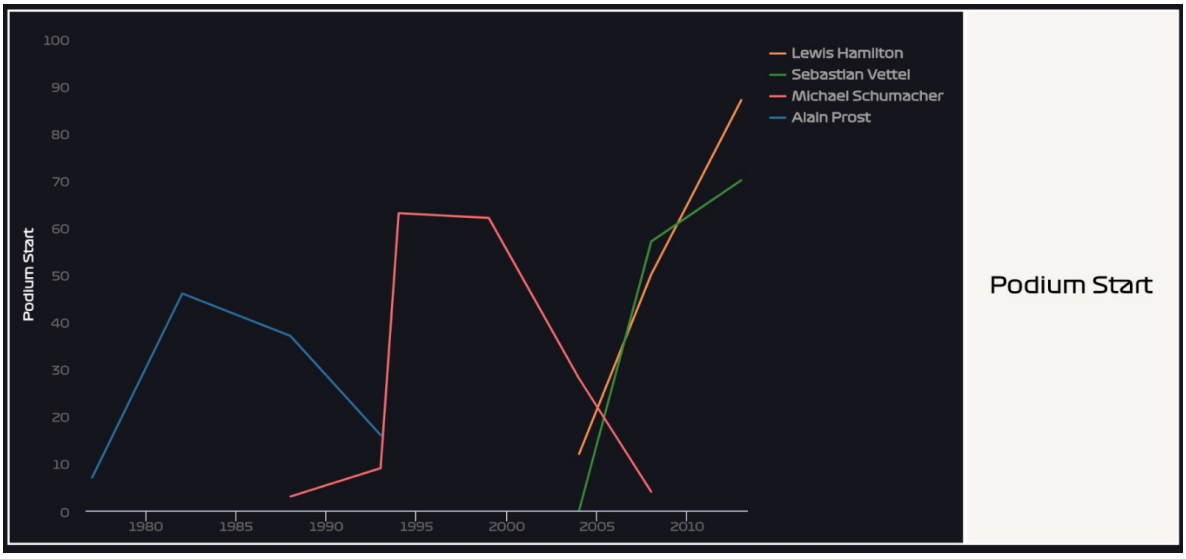


Figure 2: Line plot of podium start

There are other visualizations in the form of line charts for podium starts, podium finishes, laps, points etc. These line charts are difficult to understand as there is no specific information regarding the category(podium start). If there was sufficient information about the categories, this line chart

would make some sense. Another bad comparison is that Lewis Hamilton started around the year 2004 this is compared with Alain Prost, who retired in 1993. This comparison makes the visualization misleading as the lines are confounding because the Podium start value of 60 makes no sense.

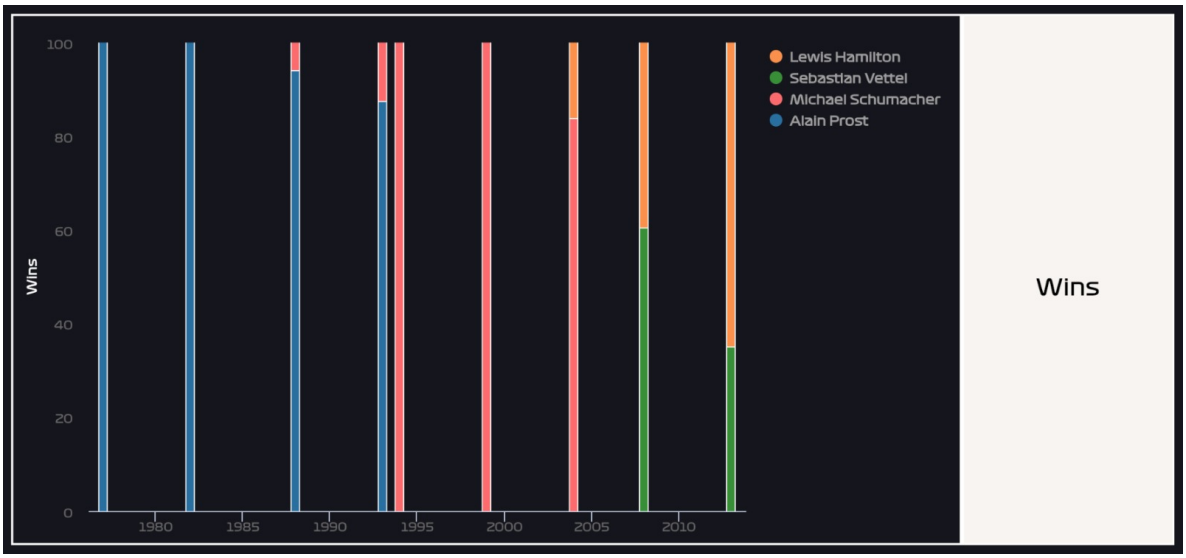


Figure 3: Stacked Bar chart

The stacked bar chart is supposed to display the dominance of a driver in a year among the selected set of drivers. It shows the proportion of races won by the driver in the year. This can be misleading as the y-axis is not represented as a percentage and it can be assumed as a number of wins which is incorrect. Another issue possessed by this stacked bar chart is the same as the line plot where all the drivers displayed did not race in the same year in order to compare them based on the proportion of their wins.

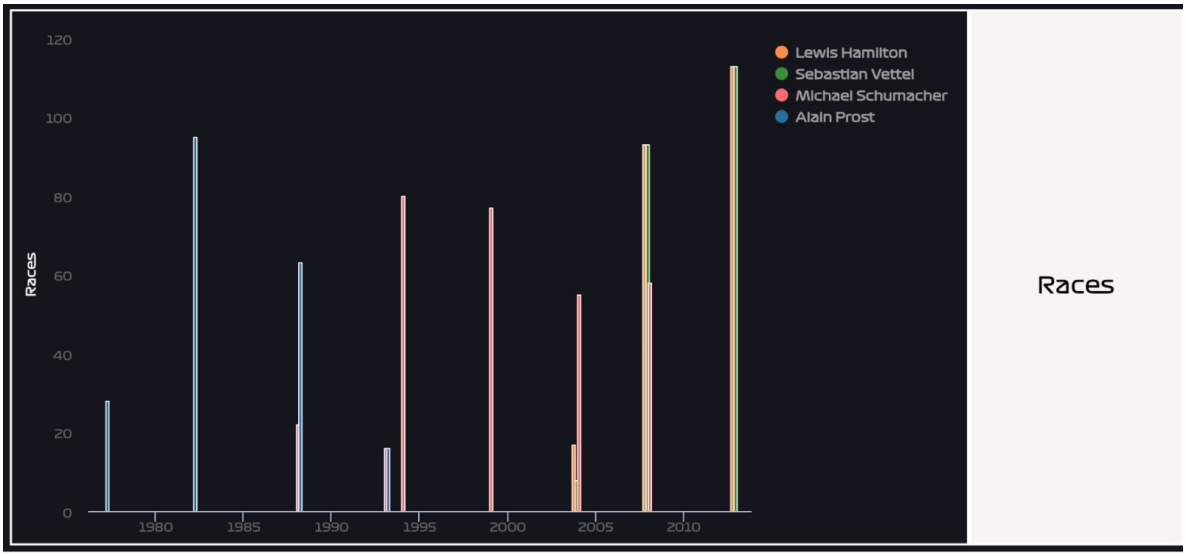


Figure 4: Bar chart

The bar chart above depicts the number of races a driver has participated in. First of all, the color choice is inappropriate as it is visually not appealing and hard to differentiate. The bars are overlapped which does not convey useful information. X axis is not labeled for all three graphs and the xticks selected are inappropriate.

This might be a good interactive visualization to brainstorm ideas. However indicating a few flaws,

we aim to provide an improved version of the analysis seen from this visualization.

Figure 5: f1-tempo

Another intriguing visualization would be [2]. Here we have an interactive dashboard that asks the user to enter lap chart details that include year, circuit, session, and driver names and it plots the lap times of the drivers selected. This can be helpful to visualize the time-to-time performance of selected drivers and compare it with other drivers. There is another option to show outliers of the data as well. Another interesting feature is that it uses telemetry data to show the speed, throttle, brake, gear, RPM, etc at every instance.

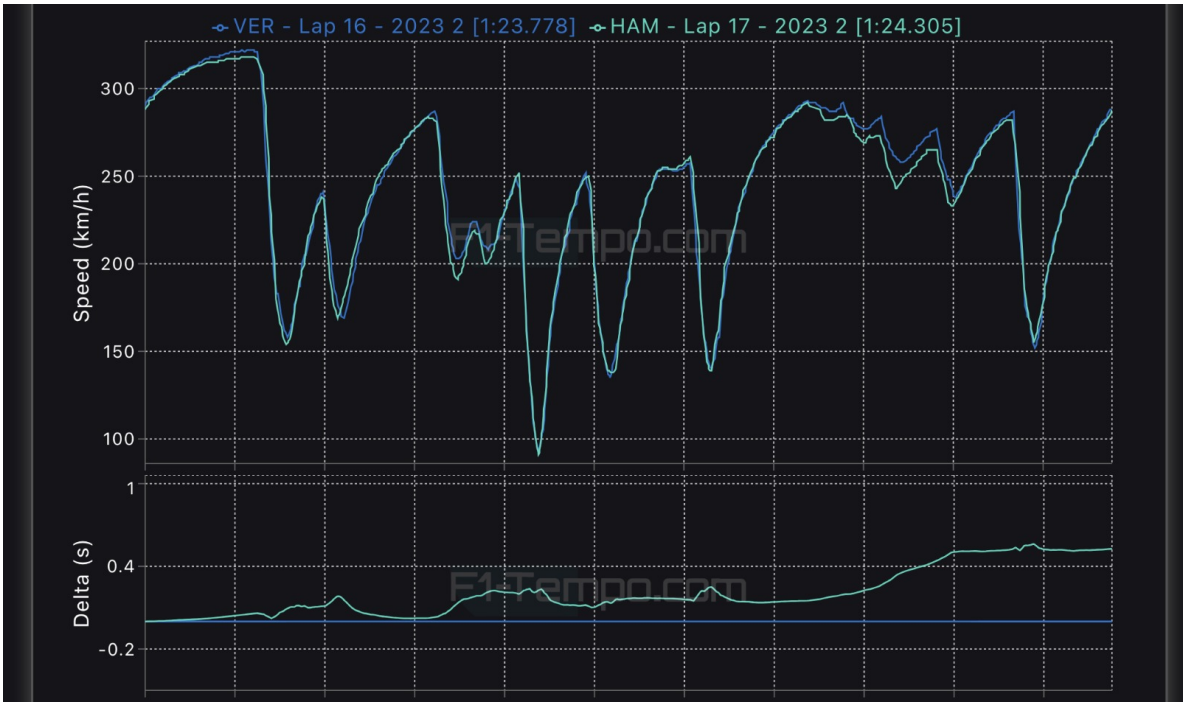


Figure 6: Telemetry information

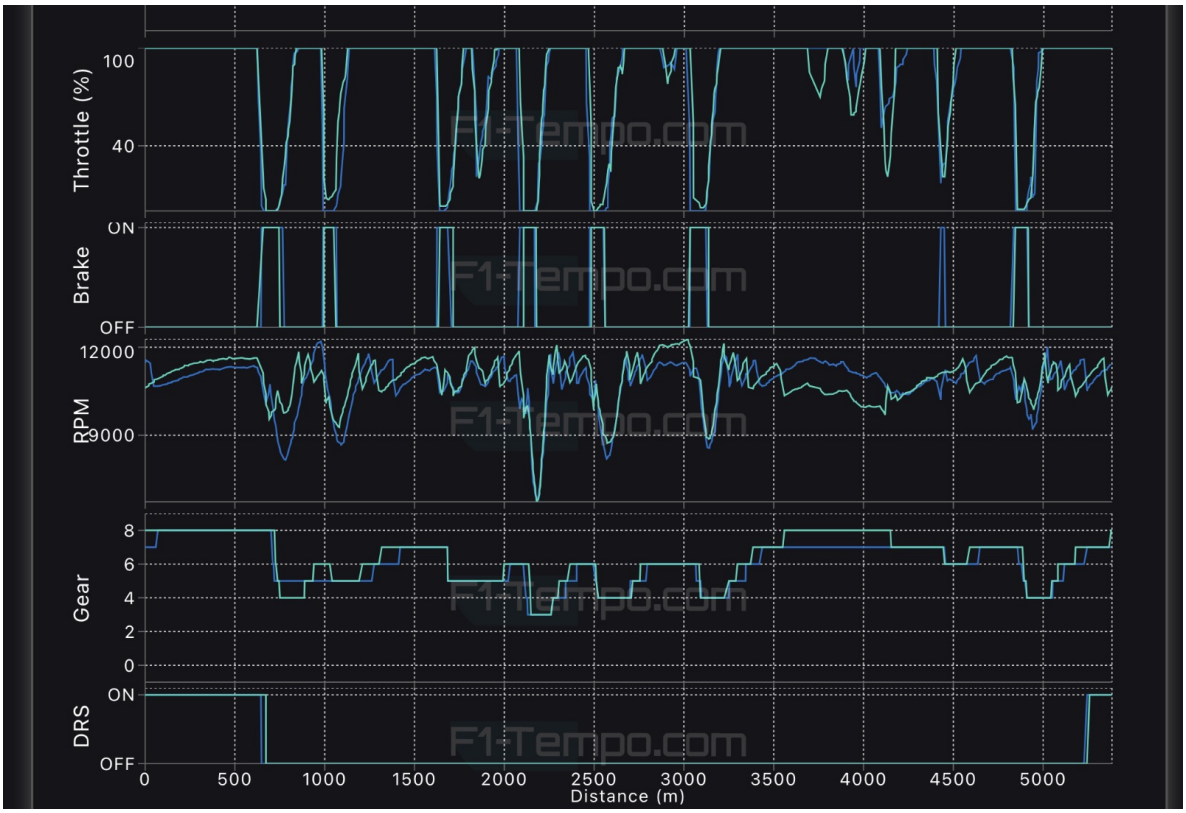


Figure 7: Telemetry information

The telemetry information charts convey useful information comparing the drivers' speed, breaking zones, throttles, and gear information in their fastest laps. As this information is given throughout the lap, it's easier to analyze who has the fastest lap in the race/qualifying.

Figure 8: Lap chart

In this above lap chart, we observe 6 drivers' lap information in 2023 at the Qatar Grand Prix and this can be useful to see how they progressed in the qualifying. One improvement could be using distinct colors to represent each line for the drivers as we can see white is used for RUS, PER, LEC and it is difficult to differentiate when plotted together.

1.3 Objectives

1.3.1 Initial expectations

Every sport has scientific and statistical information. Similarly, F1 also generated a lot of data that has many metrics related to drivers, constructors, circuits, weather, and team dynamics. We aimed to visualize the following:

1. **Analyse the driver's performance on different circuits**

   To analyze a driver's performance with the other drivers, we aim to plot the coordinates of the driver's position in every instance of time over the race.

2. **Analyse driver's performance over a season against their teammates**

   We aim to use different metrics such as lap times, speed of drivers across corners points gained in every race, and consistency in qualifying. We also plan to create radar charts to compare drivers' strengths in different areas.

3. **Analyse the constructor points over a period of time**

   The project will track and visualize the points accumulated by constructors over an extended period. This analysis will also illustrate the evolution of constructor performance, identifying trends, dominant eras, and shifts in the competitive landscape of Formula 1.

4. **Analyse the weather impact on the race**

   This will mainly focus on visualizing the weather in accordance with the tires and explain when the pit stops are taken. The duration of the pit stop can also affect the race times and lap

times. We can find the relation between speed and rainfall which directly affects the driver's performance.

5. **Analyse the circuits throughout the world**

   We will explore circuit-specific data, including track layouts, number of corners, and historical race results. Create interactive maps such as folium to show the location of F1 circuits around the world.

6. **Visualise the team pairings**

   We will explore the relationships within F1 teams, including driver pairings and team dynamics. Network graphs can help visualize these connections.

7. **Analyse the consistency of Finish/DNF**

   We will analyze the consistency of finishing races (Finish) versus not finishing (DNF) in Formula 1 is crucial to understanding a driver's reliability and performance over a season or career. This analysis can provide insights into a driver's ability to consistently complete races and compete for championship points.

8. **Analyse the engines used by constructors and success of nationalities in F1**

   We will analyze the engines used by constructors and the success of nationalities in Formula 1 can provide valuable insights into the dynamics of the sport. We aim to determine which engine manufacturers have been dominant in Formula 1 during different eras. We can visualize the number of championships, race wins, and podium finishes achieved by teams using specific engines.

1.3.2 Reformed Objectives

1. **Analyse the driver's performance on different circuits**

   There are many circuits(Locations of a race) and a driver may perform well in many of them. To analyze his performance with the other drivers, we aim to plot the coordinates of the driver's position in every instance of time over the race. This will help us to visualize the time lost/gained by the driver in the pits and also analyze the weakness of the driver in specific corners. For this, we plan to use an interactive method like Plotly, bokeh, Altair to show the positions on the circuits. We also plan to include bar charts for race wins, line charts for point standings, or even heatmaps to show lap time variations during a race.

2. **Analyse driver's performance over a season against their teammates**

   This visualization is unique because it helps us to determine the performance of the driver in the Driver's championship. We aim to use different metrics such as lap times, speed of drivers across corners points gained in every race, and qualifying results. For this, we plan to use various visualizations such as line graphs, bar charts, and stacked bar charts from Matplotlib, seaborn, and other visualization libraries to show head-to-head team statistics. We plan to use a joint plot to plot the teams(two drivers) and their positions in all the years from 1950-2023. One plot should be a density plot showing points gained by a driver and the positions he finished during every race. Another scatter plot which shows the exact position during the race and the points gained corrrespondingly.

3. **Analyse the constructor points over a period of time**

   This visualization gives the overall analysis of teams every season based on their budget and success rates. This is crucial as we make important conclusions based on the budget from the previous seasons as the more successful teams have more budgets. We plan to use stacked bar plots to visualize the team's points in every year and provide analysis. The project will track and visualize the points accumulated by constructors over an extended period. This analysis will also illustrate the evolution of constructor performance(Top teams), identifying trends, dominant eras, and shifts in the competitive landscape of Formula 1.

4. **Analyse the Lap of every race**

   This will mainly focus on visualizing the laps in accordance with the driver's position in every race. We plan to use violin plots to showcase drivers' lap times in every circuit across multiple years providing a nuanced understanding of performance variations, revealing not only the central tendencies but also the distribution and density of lap times. This facilitates the identification of outliers and trends, offering insights into driver consistency and adaptability to diverse track conditions. Additionally, the line plots depicting laps and positions of drivers from 2019 to 2023 in each circuit offer a dynamic visualization of race dynamics, enabling a comprehensive analysis of driver strategies, overtaking patterns, and the impact of position changes throughout a race. The boxplot analysis, specifically focusing on the distribution of lap times for constructors in each circuit in 2021, serves to highlight the variability and competitiveness among teams, shedding light on the performance disparities that contribute to the overall narrative of Formula 1. Together, these visualizations provide a holistic perspective on driver and team performances, contributing valuable insights to race strategy evaluation and team dynamics assessment.

5. **Provide overview of Qualifying**

   We plan to analyze the qualifying using a horizontal barplot detailing the qualifying times of drivers relative to the leader in every race across seasons. This visualization not only provides a snapshot of individual driver performances during the qualifying sessions but also offers a comparative gauge against the leader, illuminating the intensity of competition and the margins between drivers. By systematically presenting this information over multiple seasons, the plot will enable the identification of trends, improvements, or challenges faced by drivers, offering a dynamic perspective on their evolving competitiveness. The horizontal barplot encapsulates the essence of Formula 1's intense qualifying battles, fostering a deeper understanding of the sport and the reason why it is so important to start the race in higher positions.

We believe that these objectives as a whole with a few exploratory data analyses and visualizations would provide us with a comprehensive analysis and understanding of Formula One through the Data (DataOne)!

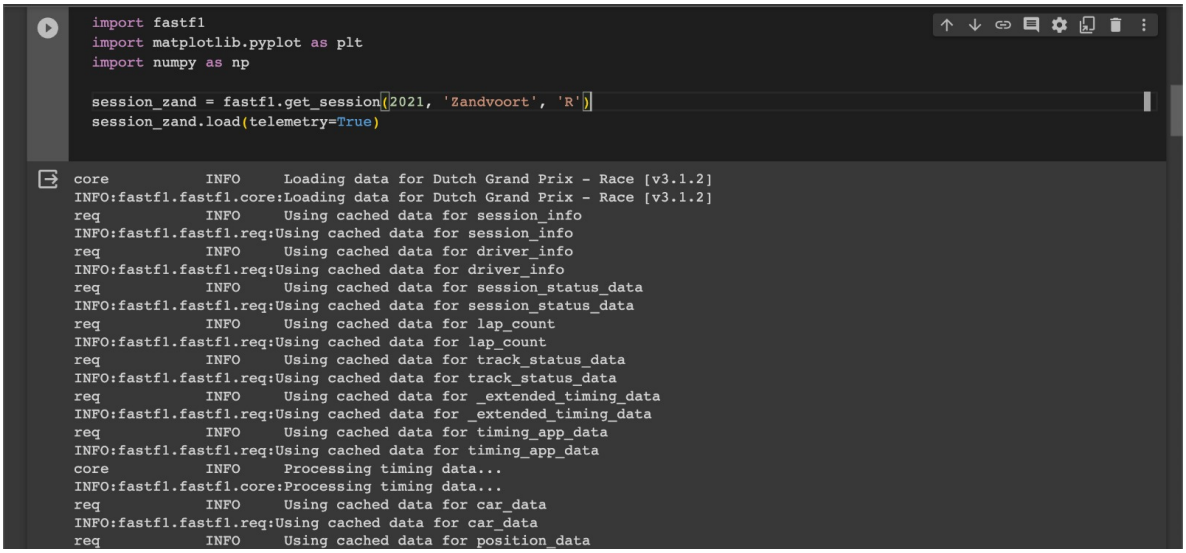
2 Process

2.1 Data and Analysis

The data has been collected from Kaggle, a Python package fastf1, and Ergast API. The data from Kaggle has 14 data files which include historical data related to drivers' constructors, circuits, pits, races, laps, etc. This data is from 1950-2023 and is essential for most of the visualization mentioned above. The data collected from fastf1, Ergast API contains telemetry data related to speed, car dynamics, location of the drivers on circuits during the race, etc. This is used for interactive visualizations of circuits and driver coordinates. This data is however only available starting 2018-2023. The data from GitHub[6] had useful information about the drivers and seasons and we used it to plot many of the visualizations. Hence we aimed to make the most intricate/ complex/interactive visualizations that do not require historical data using the fastf1 package as it contains the most recent detailed telemetry data.

There are 14 datasets from Kaggle which include the following: circuits, constructors, seasons, constructors standings, constructors results, results, sprint results, races, drivers, driver standings, lap times, qualifying, status etc.

There are many methods and functions in the fastf1 package. Some of them are illustrated below.



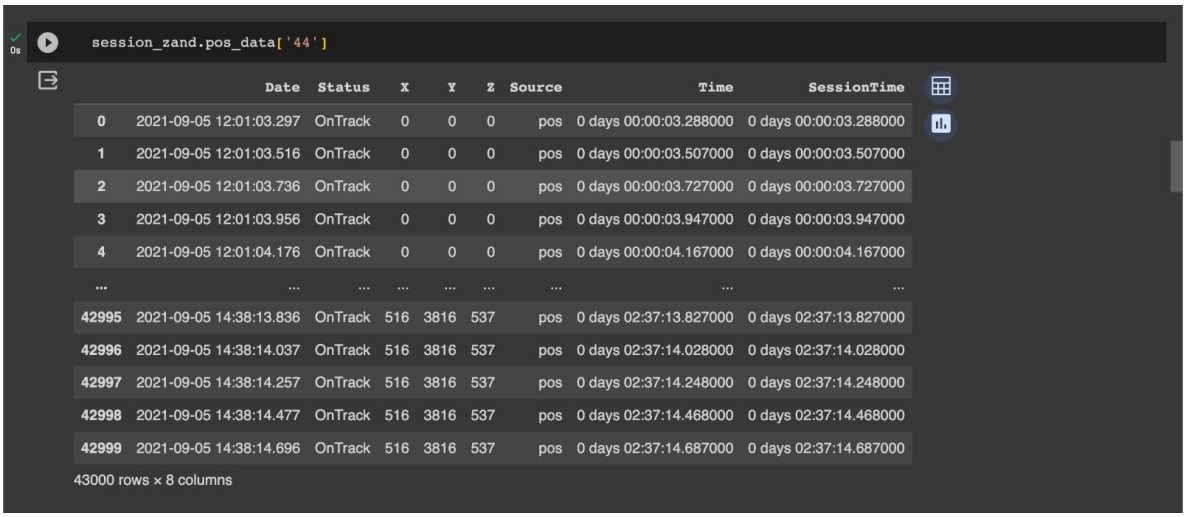
```
import fastf1
import matplotlib.pyplot as plt
import numpy as np

session_zand = fastf1.get_session(2021, 'Zandvoort', 'R')
session_zand.load(telemetry=True)
```

```
core           INFO     Loading data for Dutch Grand Prix - Race [v3.1.2]
INFO:fastf1.fastf1.core:Loading data for Dutch Grand Prix - Race [v3.1.2]
req            INFO     Using cached data for session_info
INFO:fastf1.fastf1.req:Using cached data for session_info
req            INFO     Using cached data for driver_info
INFO:fastf1.fastf1.req:Using cached data for driver_info
req            INFO     Using cached data for session_status_data
INFO:fastf1.fastf1.req:Using cached data for session_status_data
req            INFO     Using cached data for lap_count
INFO:fastf1.fastf1.req:Using cached data for lap_count
req            INFO     Using cached data for track_status_data
INFO:fastf1.fastf1.req:Using cached data for track_status_data
req            INFO     Using cached data for _extended_timing_data
INFO:fastf1.fastf1.req:Using cached data for _extended_timing_data
req            INFO     Using cached data for timing_app_data
INFO:fastf1.fastf1.req:Using cached data for timing_app_data
core           INFO     Processing timing data...
INFO:fastf1.fastf1.core:Processing timing data...
req            INFO     Using cached data for car_data
INFO:fastf1.fastf1.req:Using cached data for car_data
req            INFO     Using cached data for position_data
```

Figure 9: fastf1 loading

The above snippet shows data retrieval by calling the get session function in fastf1. The arguments passed are year, grand prix, and session type. The telemetry data is then loaded.



```
session_zand.pos_data['44']
```

	Date	Status	X	Y	Z	Source	Time	SessionTime
0	2021-09-05 12:01:03.297	OnTrack	0	0	0	pos	0 days 00:00:03.288000	0 days 00:00:03.288000
1	2021-09-05 12:01:03.516	OnTrack	0	0	0	pos	0 days 00:00:03.507000	0 days 00:00:03.507000
2	2021-09-05 12:01:03.736	OnTrack	0	0	0	pos	0 days 00:00:03.727000	0 days 00:00:03.727000
3	2021-09-05 12:01:03.956	OnTrack	0	0	0	pos	0 days 00:00:03.947000	0 days 00:00:03.947000
4	2021-09-05 12:01:04.176	OnTrack	0	0	0	pos	0 days 00:00:04.167000	0 days 00:00:04.167000
...	...	...	...	...	...	...	...	...
42995	2021-09-05 14:38:13.836	OnTrack	516	3816	537	pos	0 days 02:37:13.827000	0 days 02:37:13.827000
42996	2021-09-05 14:38:14.037	OnTrack	516	3816	537	pos	0 days 02:37:14.028000	0 days 02:37:14.028000
42997	2021-09-05 14:38:14.257	OnTrack	516	3816	537	pos	0 days 02:37:14.248000	0 days 02:37:14.248000
42998	2021-09-05 14:38:14.477	OnTrack	516	3816	537	pos	0 days 02:37:14.468000	0 days 02:37:14.468000
42999	2021-09-05 14:38:14.696	OnTrack	516	3816	537	pos	0 days 02:37:14.687000	0 days 02:37:14.687000

43000 rows × 8 columns

Figure 10: Position data from fastf1

The above code is the retrieval of position data for driver id 44, which is Hamilton on Zandvoort GP 2021. The position data includes X and Y coordinates, time which can be used to plot the driver's position on the circuit. This is shown below.

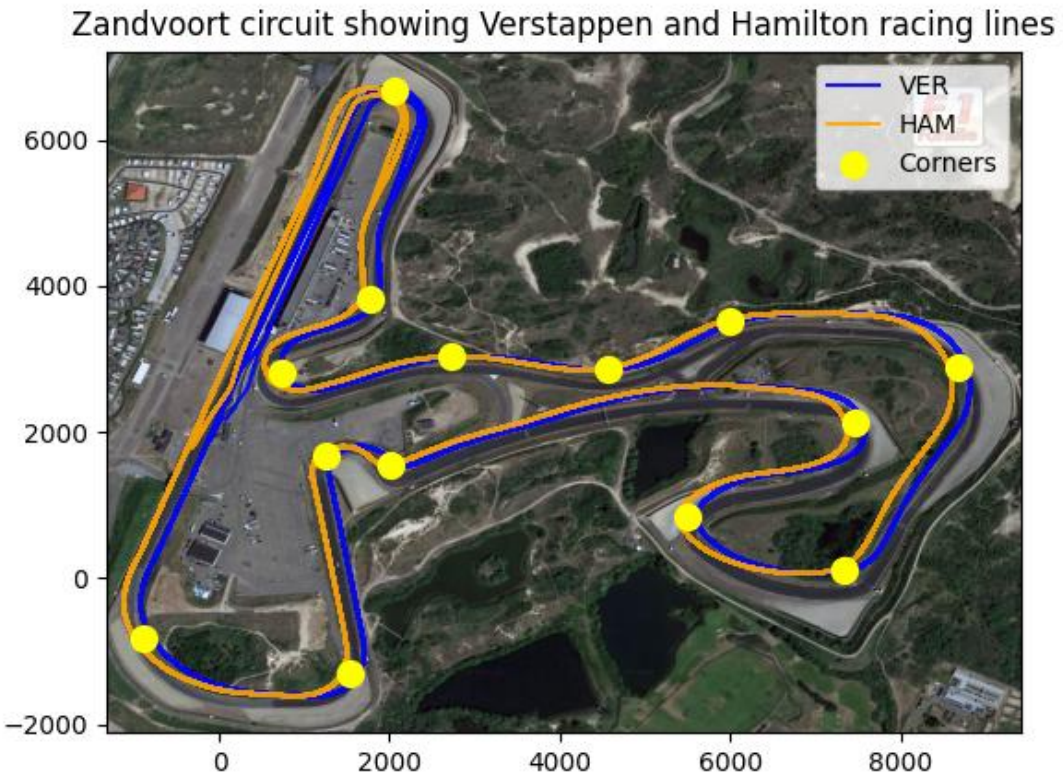


Figure 11: Zandvoort GP

The above image shows the track location(Zandvoort GP) and racing lines of two drivers, Verstappen and Hamilton in 2021. It also highlights the corners present in the circuit. This was further improved by making it interactive by showing the positions based on time(next section). With the corners marked in yellow, we can figure out the turning speed at the corners and whether or not it impacts the performance of the driver.

2.2 Candidate visualization methods

The diverse range of visualizations employed in this project reflects a strategic use of multiple plotting libraries, each chosen for its unique strengths and capabilities. Altair, with its declarative syntax, proves to be an excellent choice for creating intuitive and interactive visualizations, such as radial plots, line plot,s and barplots analyzing constructors' performance over time and lap data. Seaborn, known for its aesthetic appeal, is well-suited for crafting visually compelling plots like the violin plot, swarm plot, and joint plot illustrating driver lap times, allowing for a nuanced exploration of performance distributions. Matplotlib's interactive capabilities are useful in the live race visualization for the 2023 Las Vegas Grand Prix and the interactive dropdowns for circuits with corners, enhancing user engagement and exploration. It is also effectively employed in creating classic yet informative plots like bar plots showcasing constructors' progressions over time, qualifying overviews etc. The combined use of these libraries showcases a thoughtful approach to visualization, leveraging the strengths of each to convey a rich narrative of Formula 1's intricate dynamics, from historical trends to real-time race insights and nuanced performance analyses.

2.3 Experiments and Vision

2.3.1 Experiments

1. The interactive plot idea presented in Figure 11 (Zandvoort GP) is not possible because an image cannot be set as a background for an interactive plot so we decided to draw the outline of the map and represent all the 20 drivers.
2. We aimed to plot weather data initially but due to the limitations of data and quality of data, it was not possible to plot weather data as expected with the resources available as of now(although we want to explore this part in the future).
3. We also aimed to plot network graphs of all teams and teammates but due to a lack of data to specify the connections between teammates, we decided to make it our future work.
4. DNF data is not accurately available for all the races over time and there were inconsistencies in the data regarding a driver's DNF status.

2.3.2 Ideas and Prototypes

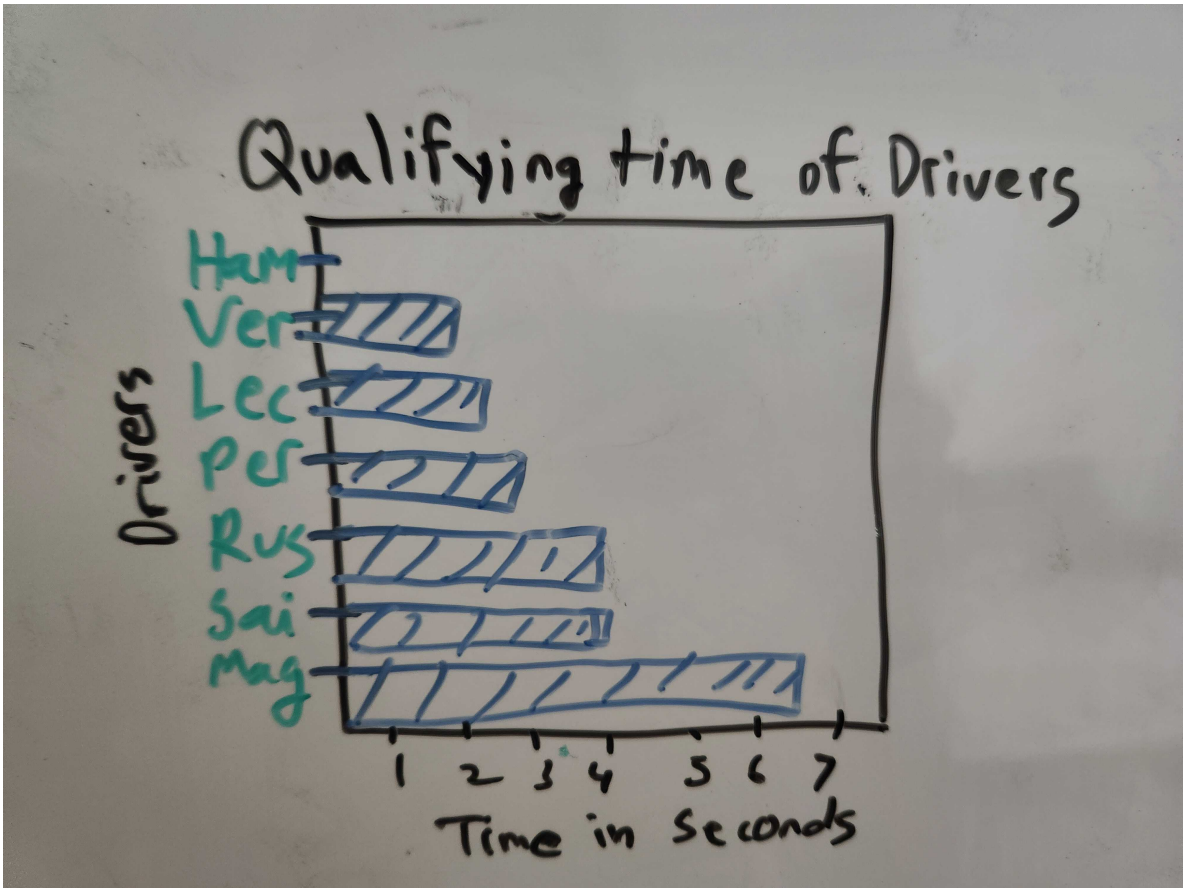


Figure 12: Barplot of qualifying times

The horizontal bar plot aims to display qualifying times for various drivers on a given circuit. The plot follows the Leader time style which means that the leader (fastest guy) has 0 gap(in seconds) to the leader (himself) and other drivers' lap time comparisons to the leader are plotted respectively.

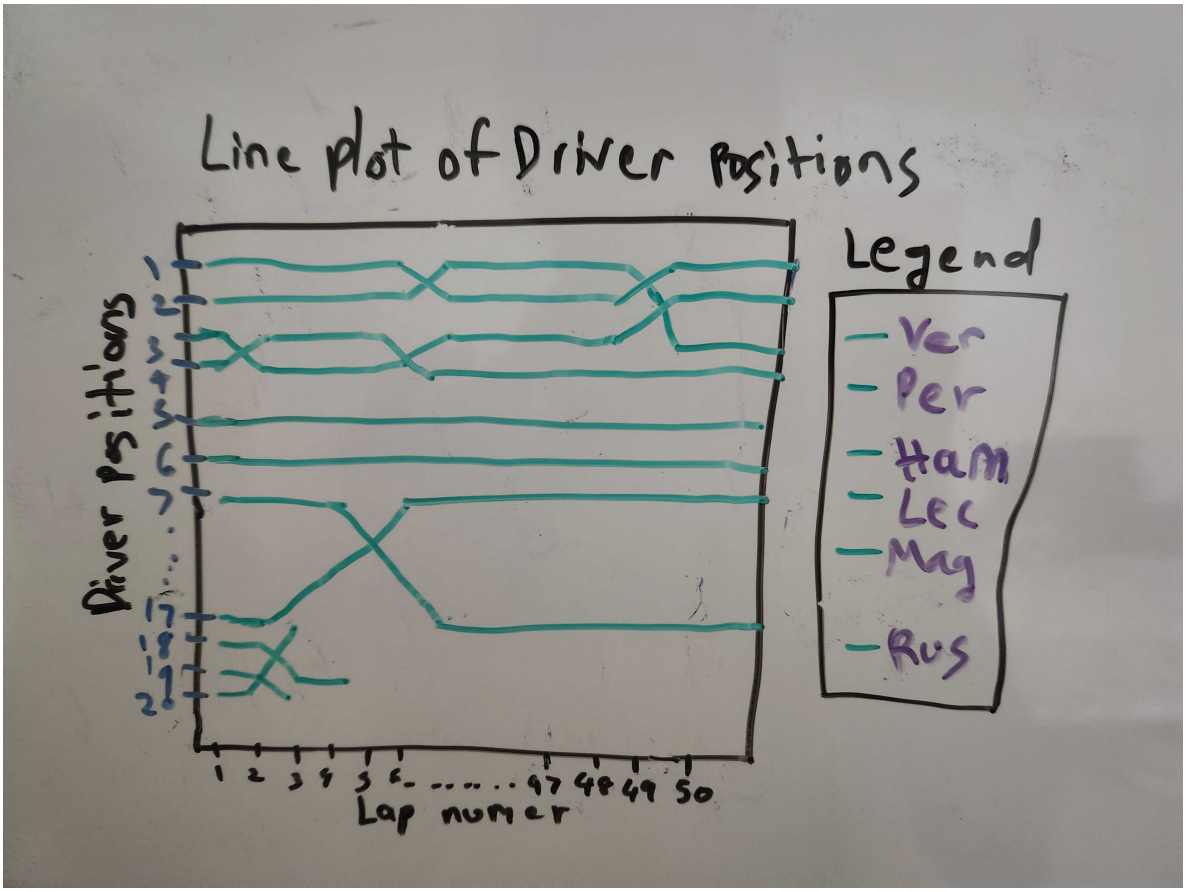


Figure 13: Line plot of driver positions in laps

The plot we aim to construct is an informative line plot, which shows the driver's position's progression throughout the race. The lines will represent drivers and how they gained and lost positions in the race. The lines that make it to last will represent drivers who completed the race at their given position and will be awarded respective points, while lines that don't reach the end represent disqualified drivers.

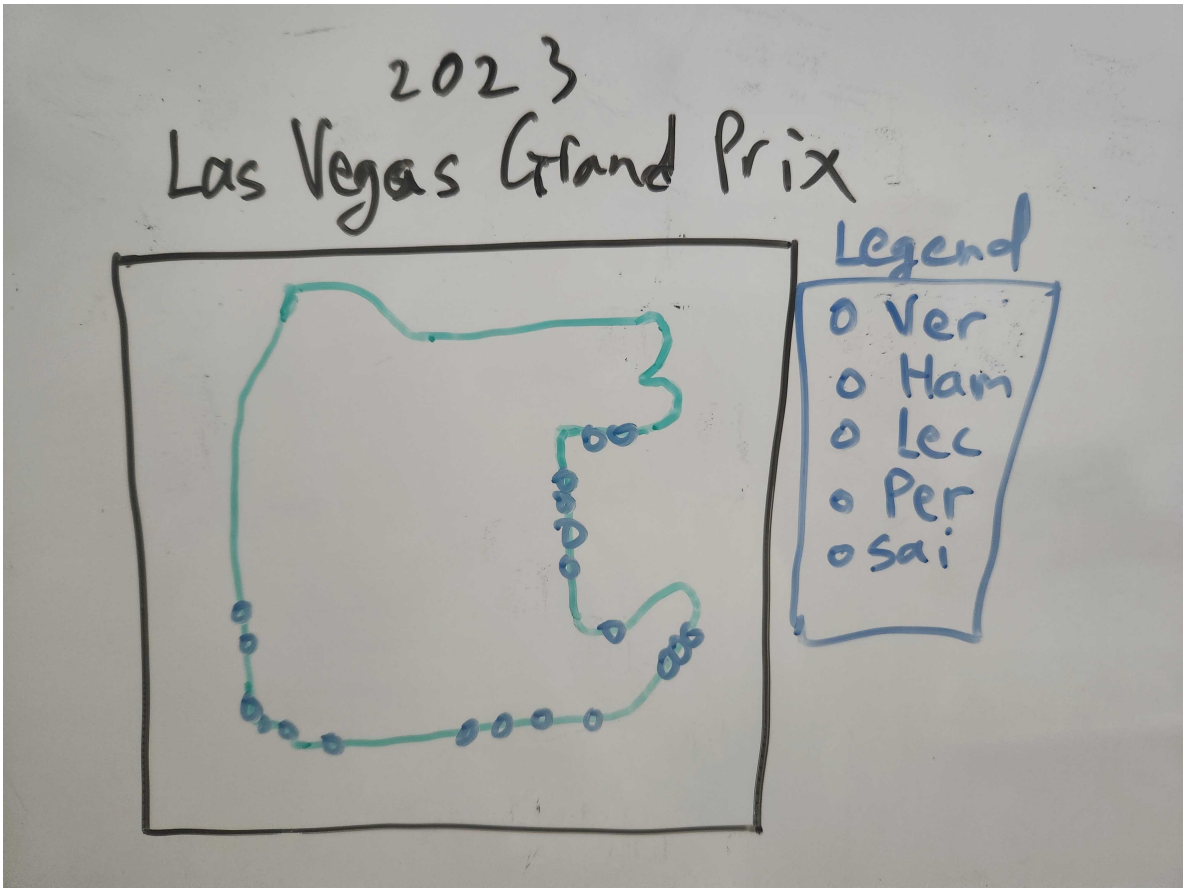


Figure 14: Interactive visualisation

The plot shows the layout for the Las Vegas 2023 Grand Prix, we aim to plot the layout of the circuit and draw drivers a points on top of the circuit. Then using the position data (x and y coordinates of the drivers) we aim to create an animated plot that shows the motion of the drivers on the circuit. This will hopefully result in a live Race simulation which is extremely pleasing to watch!

2.4 Contribution

The outlined objectives encompass a thorough and multifaceted exploration of Formula One, aiming to unravel the intricacies of driver and team performances across different circuits, seasons, and critical race components. By employing a diverse set of visualization techniques, including interactive methods like Plotly, bokeh, Altair, and established libraries such as Matplotlib and seaborn, this project offers a comprehensive understanding of Formula 1 dynamics. Analyzing drivers' positions over races provides insights into their strengths and weaknesses, enabling a nuanced examination of their performance in specific race conditions. The head-to-head analysis against teammates enhances our understanding of team dynamics, integrating various metrics for a holistic evaluation. Constructor points analysis sheds light on the competitive landscape, identifying dominant teams and tracking the evolution of their performance over time. The lap analysis, including violin plots and line plots, delves into the intricacies of race dynamics, showcasing performance variations and strategic nuances. Furthermore, the detailed overview of qualifying sessions through horizontal barplots offers a dynamic perspective on the intensity of competition and the significance of securing favorable starting positions. Together, these objectives weave a narrative that not only explores the rich history and evolving trends of Formula 1 but also sets the stage for a more profound understanding of the sport through the lens of comprehensive data analysis. In undertaking this project, we contribute not only to the field of

motorsport analytics but also provide a valuable resource for enthusiasts, analysts, new viewers and teams seeking a nuanced perspective on the captivating world of Formula 1.

3 Results and Insights

Overview of F1 Circuits in Countries

Figure 15: Circuits

Formula 1 is hosted across a dynamic array of circuits worldwide, each presenting a distinct challenge to drivers and teams. From iconic tracks like Monza and Silverstone to modern marvels like Yas Marina Circuit and Circuit of the Americas, the sport spans a diverse mix of venues. Whether winding through historic streets like Monaco or navigating the sweeping curves of Suzuka, the varying characteristics of these circuits contribute to the rich tapestry of Formula 1, creating a global spectacle that captivates racing enthusiasts across continents.

To learn more about the circuits, let's take a look at the corners of every F1 circuit.

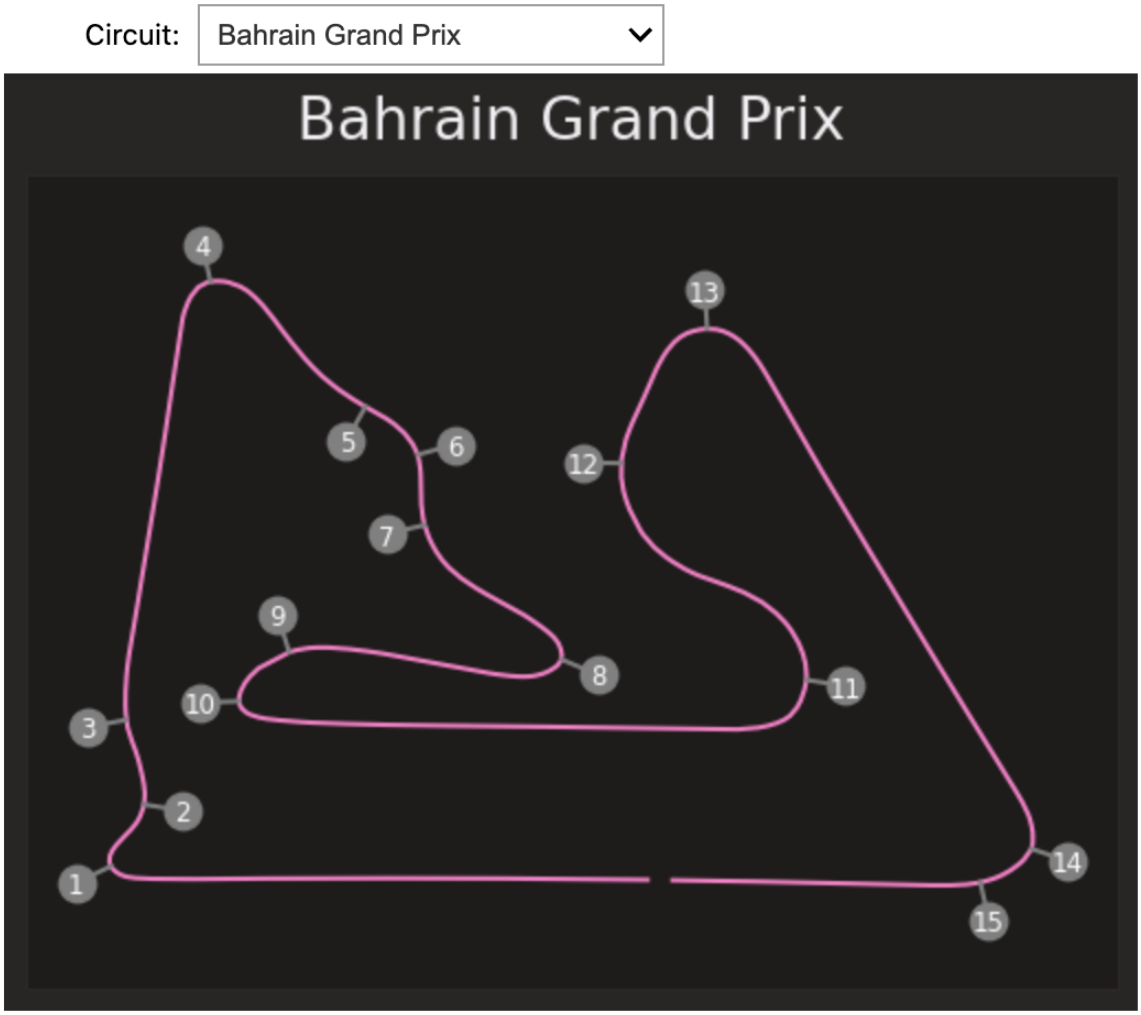


Figure 16: Corners on tracks

Corners in Formula 1 circuits are crucial because they play a significant role in shaping the characteristics of the track and influencing the overall racing experience. These turns, varying in radius and elevation, not only provide opportunities for overtaking but also significantly influence lap times and tire management. Beyond their technical importance, corners add spectacle and excitement, where thrilling moments such as overtakes and strategic moves often unfold, contributing to the overall thrill and uniqueness of each circuit.

Speed:

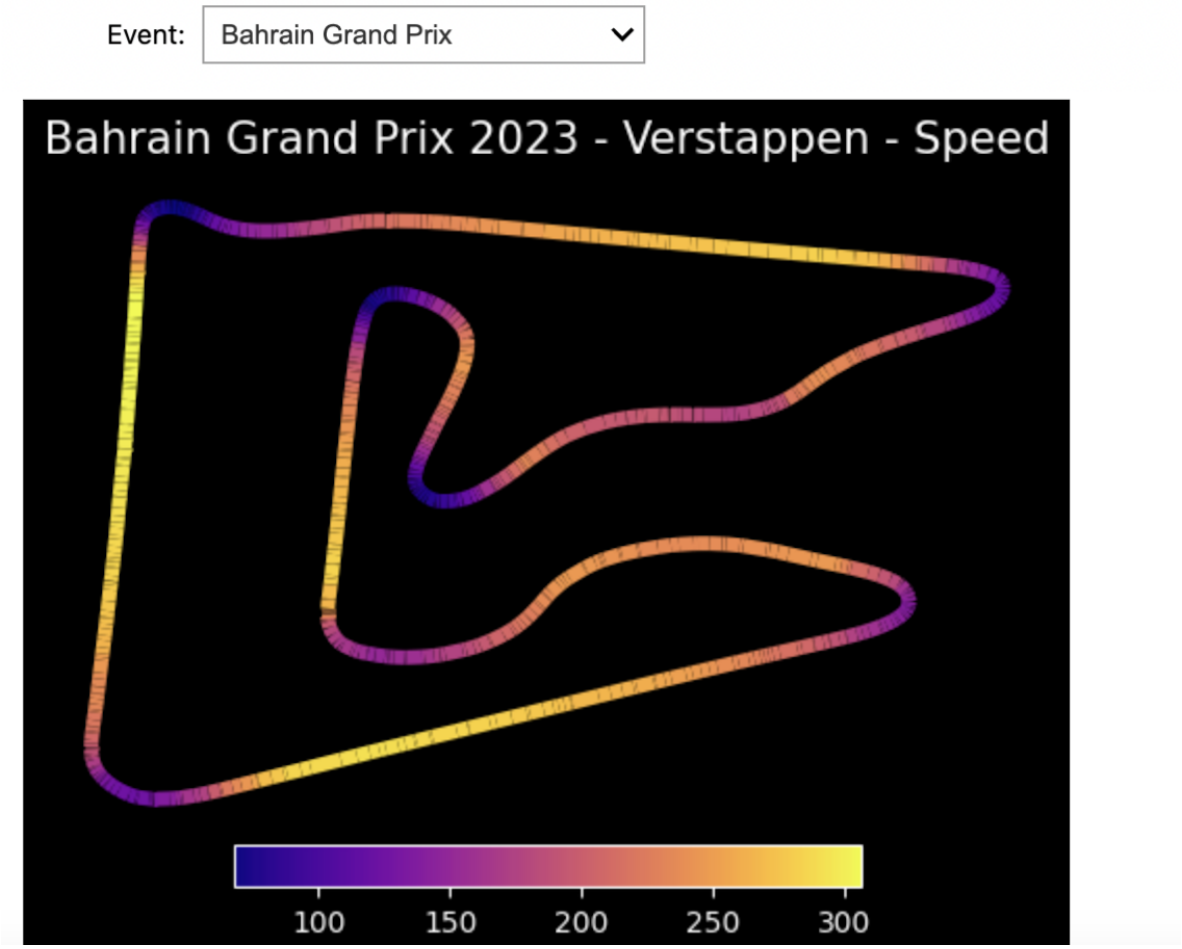


Figure 17: Speed heatmap in the circuit

The above track shows the circuit speed of Verstappen on Bahrain GP 2023. As we can observe, his speed is high on the straights of the circuit and in the corners, the speed drops drastically due to braking. This holds good for all the drivers in most of the circuits and shows us how important it is to not lose speed in the corners.

Nations participating in F1

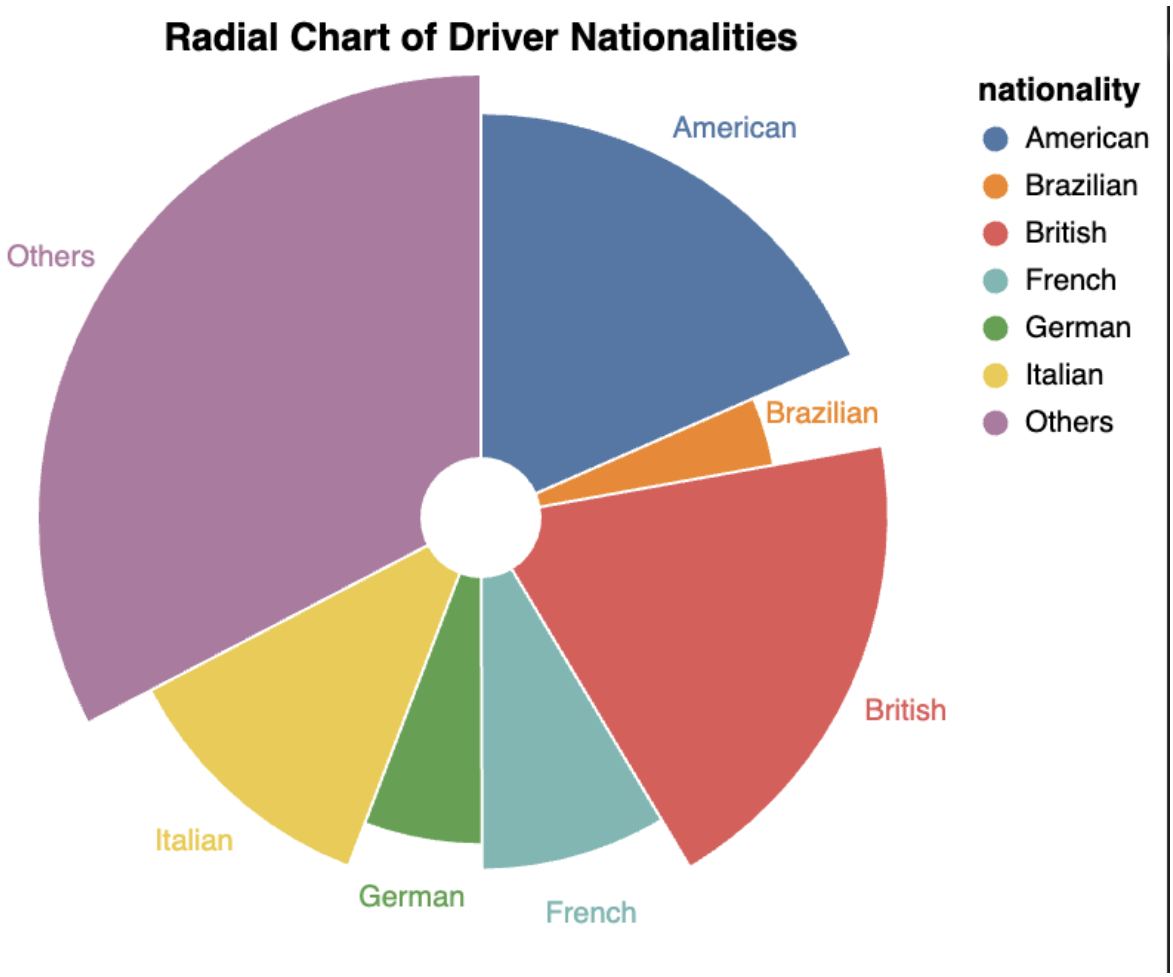


Figure 18: Radial chart of nationalities of the drivers

From the radial chart, we see the drivers are mostly from British and America(165, 158), and Germany has a few drivers. The rest of the countries as combined and it holds about 280 drivers.

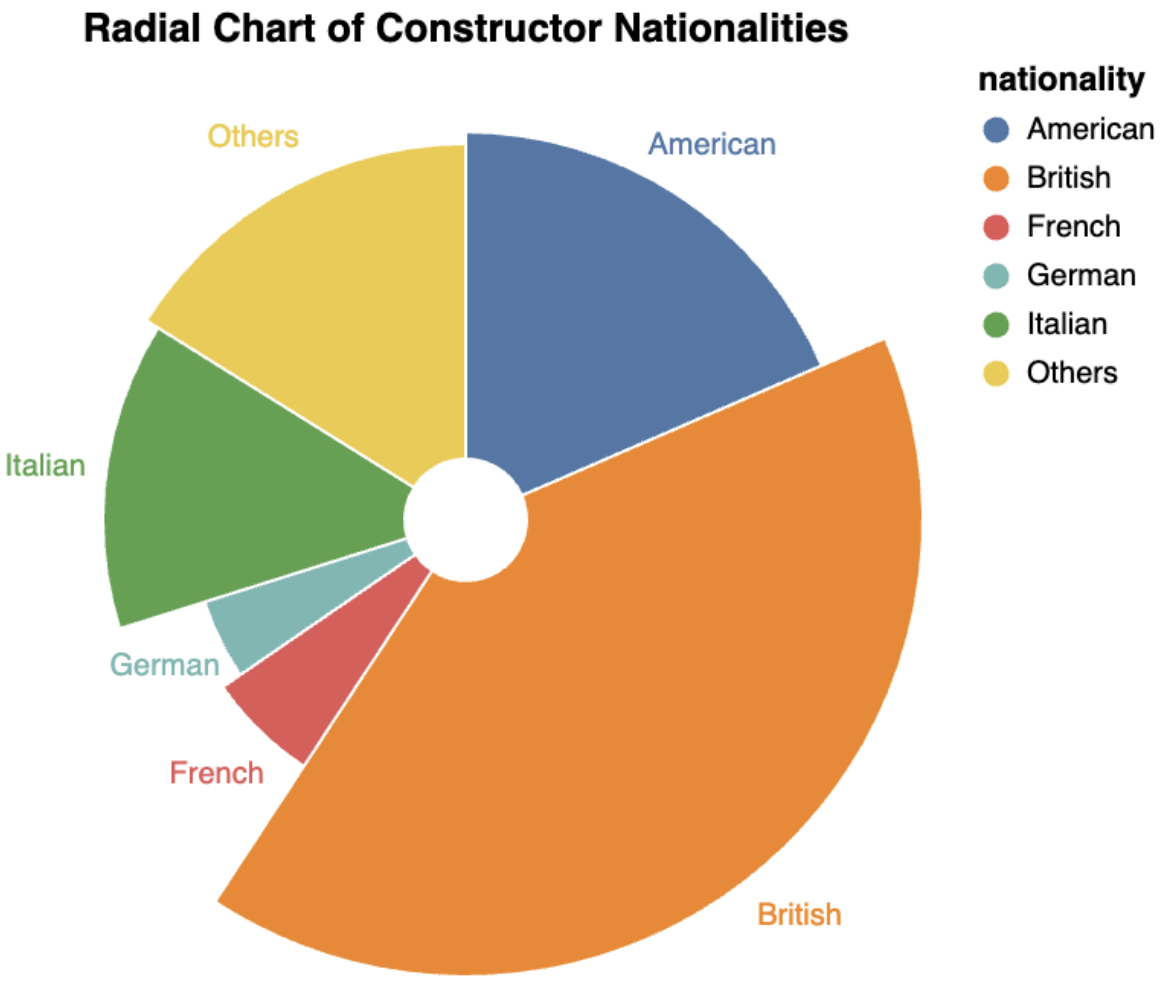


Figure 19: Radial chart of nationalities of the constructors

From the above radial chart, we see the constructor's nationalities and mostly they are British(86 teams), followed by Americans, Italians, French, and German. The others account for other 34 countries.

Champions

A closer look at the champions over the years-

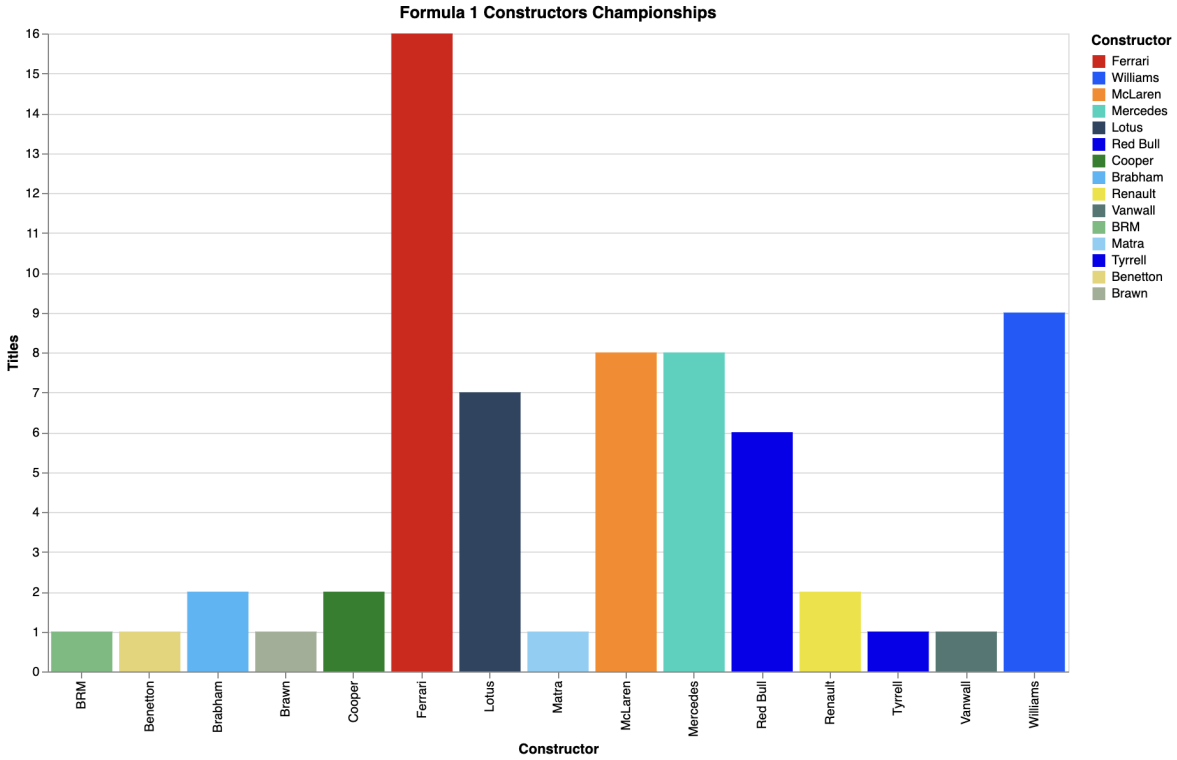


Figure 20: F1 Constructors Champions over time

Ferrari has been the Formula One World Constructors' champion for the maximum number of times(16). Followed by Williams(9 times), Mercedes and McLaren(8 times), and Lotus (7 times). Even though we do not find a few of these teams now, they have created numerous world records and have been one of the most successful in the past. Let us see if the statistics stay the same, even today?

Constructors over the recent years:

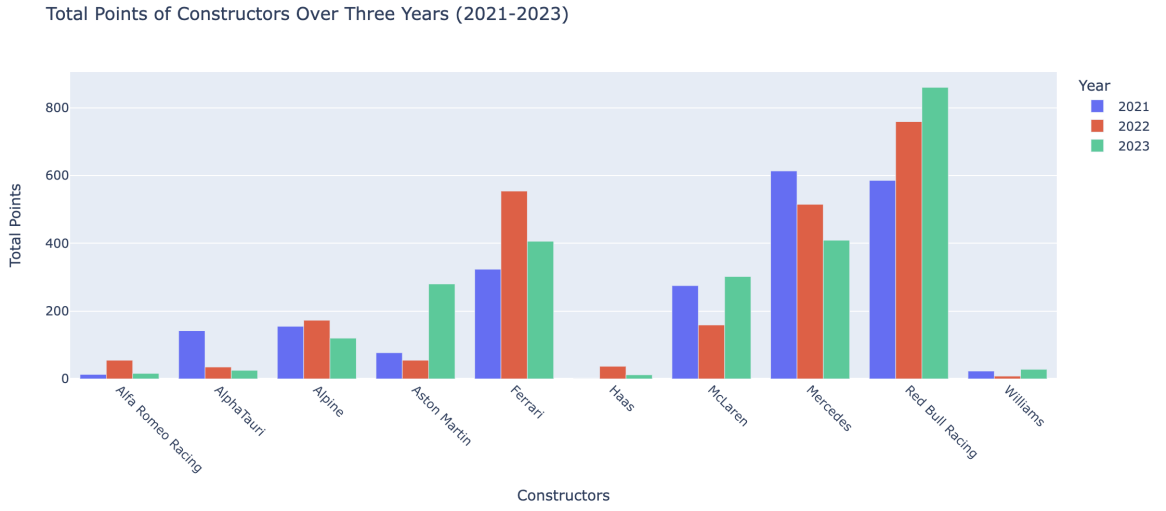


Figure 21: Total points of constructors over three years-2021-2023

From this interactive plot, we can see the points of all the teams over the past three years. Clearly, Redbull has improved a lot with respect to points over the years. Mercedes, who has been doing

well in 2021, have struggled to maintain their dominant performance from 2021 to 2023. Ferrari has predominately performed well in 2022 when compared to the other 2 years. Haas has scored some points around 2022 and 2023 but none in 2021. Aston Martin has significantly improved their score to 280 points this season, which nets them pretty close to the top 3 teams in 2023.

Let's look at the performance of the top 3 favourites - Ferrari, Mercedes, and Redbull

---

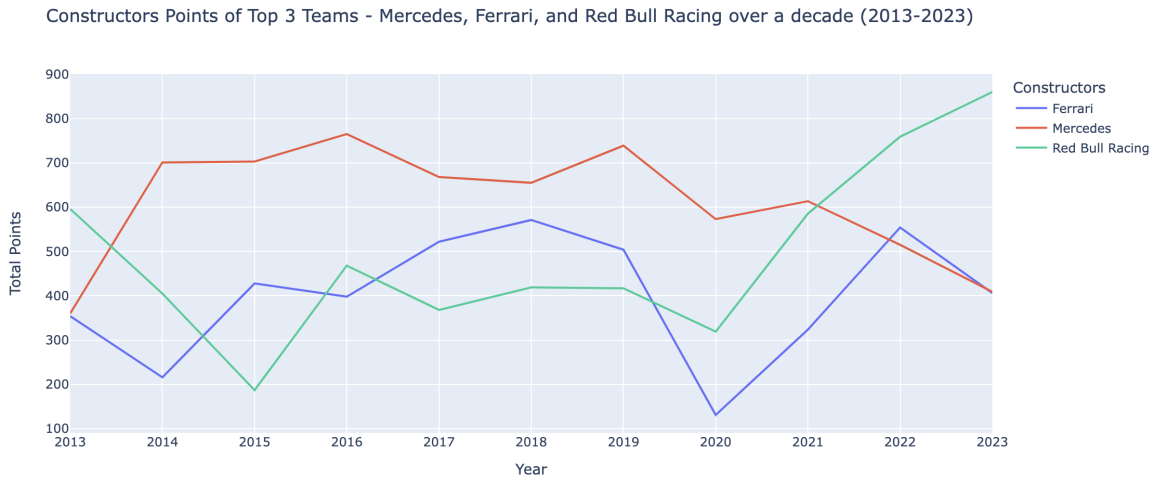


Figure 22: performance of top 3 teams - Ferrari, Mercedes and Redbull

From the top 3 constructors lineplot, we can observe in the past decade, RedBull Racing has improved its performance from 600 points to around 860 points by the end of the 2023 season. This is a significant lead that guarantees them the constructors championship. Whereas, Mercedes had almost ruled the decade by their gargantuan lead by scoring around 750 points each year which has started to decline in the past 2 years. Fans favorite team, Ferrari, being the most classic team of the F1 had a mediocre performance in the past decade when compared to the 2 newer teams.

Lets look at the Podiums of drivers in every team to understand how they compare with their teammates

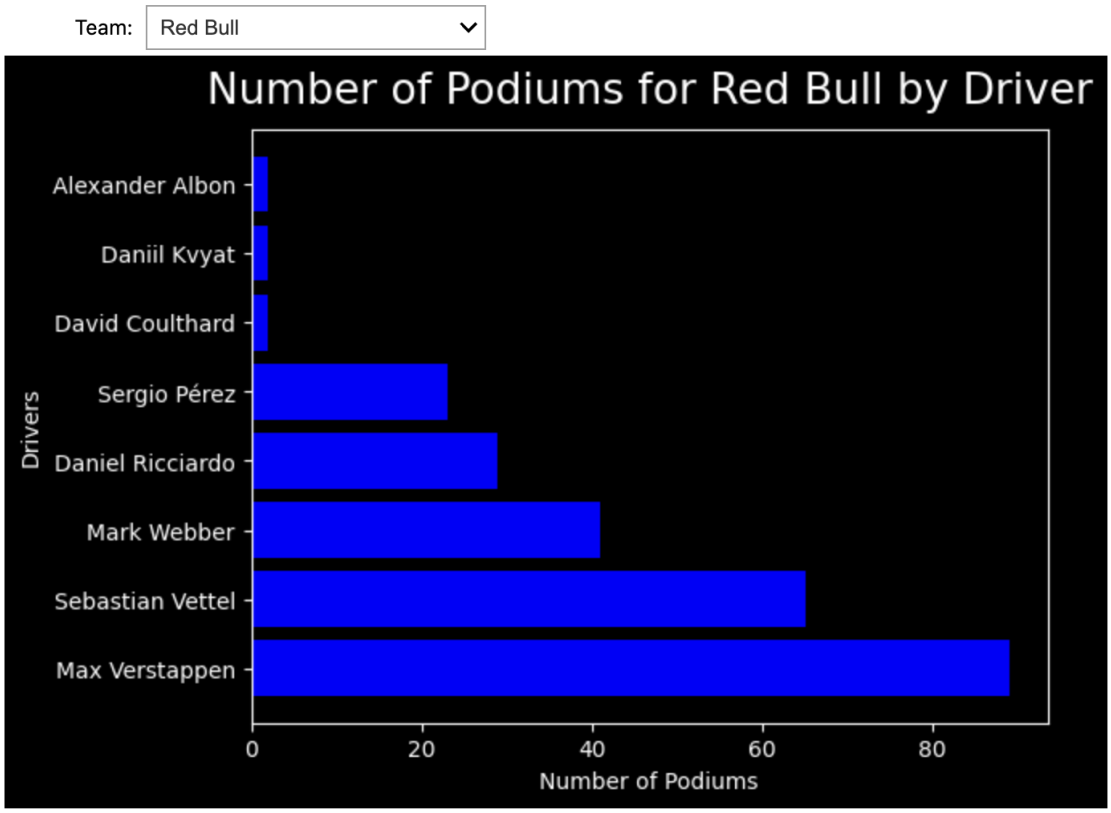


Figure 23: Podiums of drivers in every team

The above horizontal barplot represents the number of podiums by all the drivers of a team. Ferrari, Mercedes, and McLaren have many drivers who won podiums, whereas a relatively newer team RedBull has 8 drivers in the podium with Max Verstappen winning most podiums of 98, followed by Sebastian Vettel and Mark Webber. The dropdown gives the option to choose the teams as required.

We also observe that the new driver Perez, has been winning podiums almost less than half of Verstappen for Redbull but still lesser than what Daniel Ricciardo had won along with Max.

2021 Season

In the course of the season, a total of 22 races unfolded across 20 diverse countries, each contributing its unique flair to the Formula 1 calendar. Notably, Italy played host to two races, the first at Imola for the Emilia Romagna Grand Prix and the second at the legendary Temple of Speed in Monza.

The persistent global impact of the Covid-19 pandemic led to the exclusion of some familiar race venues well in advance of their scheduled dates. Among them were the revered circuits of Canada, Australia, Japan, and Singapore. In their absence, other notable tracks, including Imola, Portugal, and Turkey, stepped in to fill the void. Notably, the 2021 season saw the introduction of new circuits to the F1 lineup, with Qatar and Saudi Arabia hosting their inaugural Grand Prix events, adding fresh excitement and challenges to the racing calendar.

This season is selected for analysis as this was the most exciting season with many turnarounds with respect to drivers' championships(Hamilton and Verstappen battle to win the title) and constructors' championships.

Qualifying:

Qualifying time is crucial in Formula 1 as it determines a driver's starting position on the grid, profoundly influencing their race trajectory. Securing a favorable grid spot not only reduces the

challenges of navigating through traffic in the opening laps but also enhances overtaking opportunities. The strategic advantage gained from a strong qualifying performance extends to race tactics, pit stops, and overall team planning. Additionally, a higher starting position increases the likelihood of earning valuable championship points and enhances a team's reputation. In the fast-paced and competitive world of Formula 1, qualifying time serves as a pivotal factor in shaping race outcomes and a driver's overall success in the championship.

Season: 2021

Circuit: Abu Dhabi Grand Prix

```
<Figure size 640x480 with 0 Axes>
```

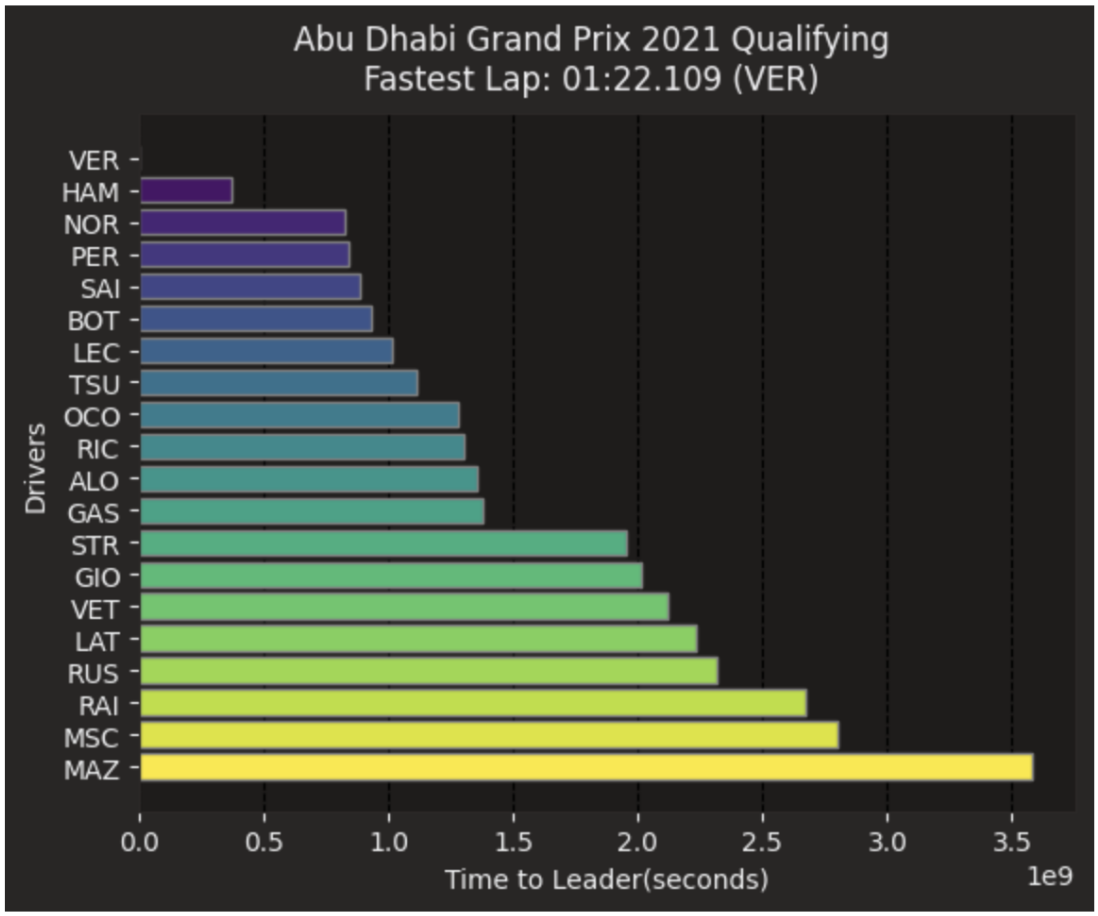


Figure 24: Barplot of qualifying times of every driver with the leader

The barplot is one of the simplest ways to show how the qualifying works on a larger scale. The Driver with the fastest lap leads the position on the grid, followed by other drivers who are closest to the leader in time. The x-axis represents the time in seconds to the leader and the y-axis represents the driver's names. The dropdown for the year and different circuits can be used to explore different qualifying circuits during different years.

Before the first race of the season in Bahrain, in the qualifying Verstappen had the pole position with the fastest lap of 01:28:997 minutes followed by Hamilton and Bottas who were around 0.4 and 0.7 seconds behind Verstappen's time.

Race

Points Progression of Top 10 Drivers over the season :

Here, let's take the drivers who finished in the top 10 of drivers standings

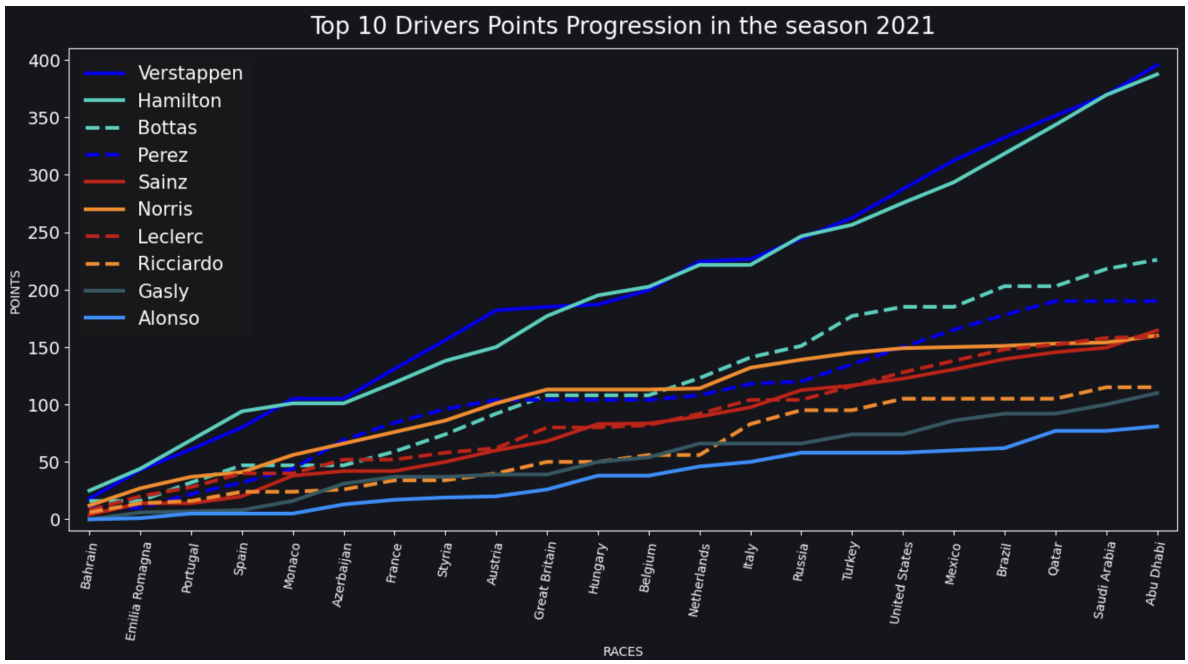


Figure 25: Points progression in the season 2021

Clearly, Hamilton and Verstappen are fighting the battle for the championship till the last race of the season which made it one of the memeorable seasons of F1. Though Verstappen was leading the season, Hamilton was no less with respect to points and clearly the two lines were far apart when compared to other drivers in the season progression. The last race was the deciding one where all the drama exploded leading to the win of a champion.

HAMILTON vs VERSTAPPEN :

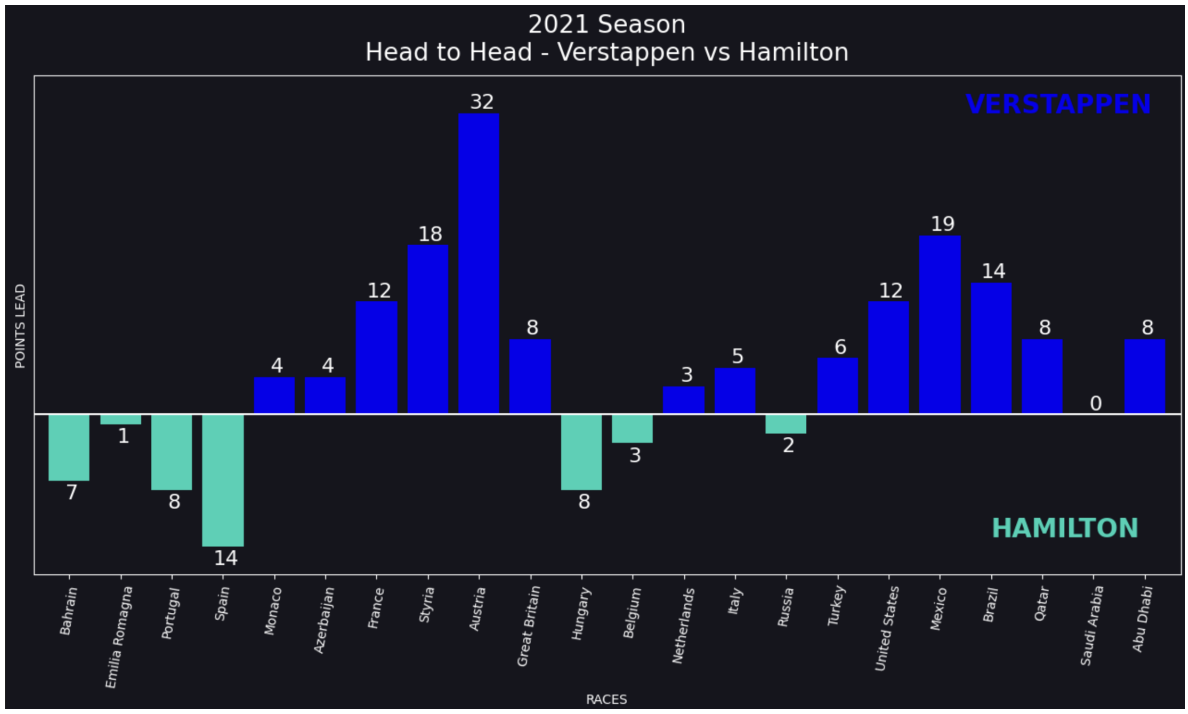


Figure 26: Hamilton vs Verstappen Points lead through the season

The above plot shows how close Hamilton and Verstappen were in this season to winning the world champion title. In the initial races, Hamilton was in the lead (we can see in the bars that represent the points), and later on Verstappen started scoring points in the pre-final Race in Saudi Arabia, both Verstappen and Hamilton were equal in points and the last race was the deciding one!

Laps distributions

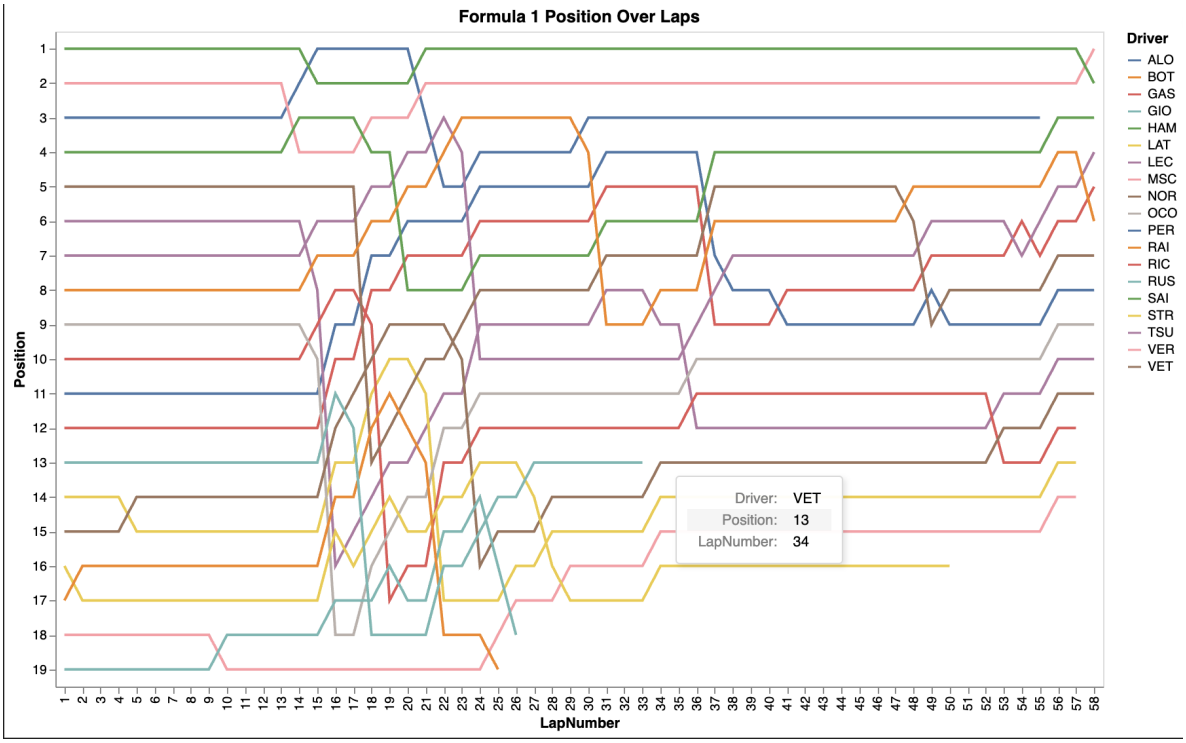


Figure 27: Line plot of all the laps of all the drivers in a race

From the telemetry position over laps data, and the x-axis represents the LapNumber and y-axis represents the starting Position grid of each Driver in the race. If we look at the Abu Dhabi Grand Prix we see that in Lap 1, all the drivers start at their position set at the qualifying in the previous plot and begin the race. Verstappen who was pole in qualifying started the race but Hamilton managed to overtake him in the first turn and hence we deduce that at the end of the first lap, the position of Hamilton is 1 and Verstappen is 2. This series continued until Lap 14 for Hamilton and then Perez overtook Hamilton to the pole for 5 laps before he regained his position back in Lap 21.

The random spikes indicate that maybe a crash took place(which caused a Red/Yellow flag) which led the driver to gain/lose positions. Another caveat to these spikes may be due to the pitstops of the drivers. There might be some cases where the line ends abruptly which indicates that the driver has retired his car in that lap. Here(Raikkonen) retired his car in lap 25 and Russel in lap 26.

Another important aspect of this particular race was that, it was the last Grand Prix of the season and both Hamilton and Verstappen were battling for the championship from the previous plot with the same number of points and Hamilton led the race in the first lap with Verstappen being the 2nd till the last two laps. But the championship turnaround occurred due to an incident with Latifi in Lp 50 where he crashed, causing a Yellow flag. This led to some decisions by the stewards regarding the positional dynamics of the race which concluded in Hamilton losing the lead of the race and Verstappen winning the World Driver Championship of the 2021 season.

This is an important plot that gives nuances of the race to find out what happened in the race, without actually watching it!

Year: 2021

Race: Monaco Grand Prix

```
<Figure size 640x480 with 0 Axes>

<Figure size 600x500 with 0 Axes>
```

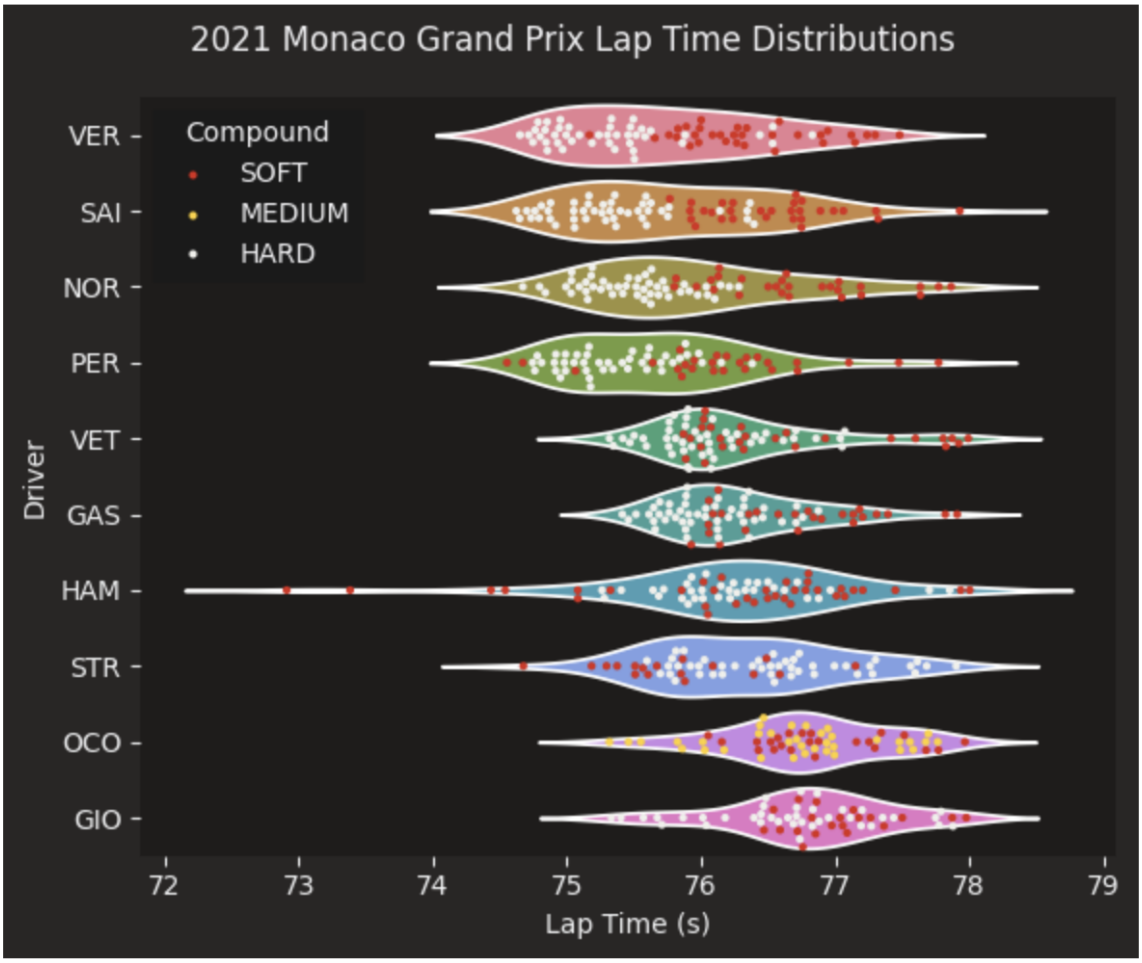


Figure 28: Violin Plot of lap time distributions of all drivers

From the above violin plot with the swarm plot, we can see the lap time distribution of every driver in every lap, for years 2019-23 in every circuit. Here is an example of the 2021, Monaco Grand Prix where we also have the tyre information used in these laps(among the points). We can observe that when the tires are soft, then the lap time is higher for most of the drivers in this race which means that soft tires might not be appropriate for this track or drivers started with the soft tires and ended their race on mediums/hards which reduced their lap times during the later stage as the fuel in the car is less during the end of the race(resulting in more speed). Medium tires are only used by Ocon in this race and other drivers use Hard and soft tires. When the drivers use hard tires, we can see that they cover more laps with hard tires than with soft tires(the number of white dots is more than red here). This plot tries to give the tire information(usage of tires in various laps) while also giving more insights about the distribution of lap times across different laps for every driver in the race.

Insights- Hamilton had performed consistently well and had a lesser lap time (72 seconds- 75 seconds) compared to other drivers with soft tires.

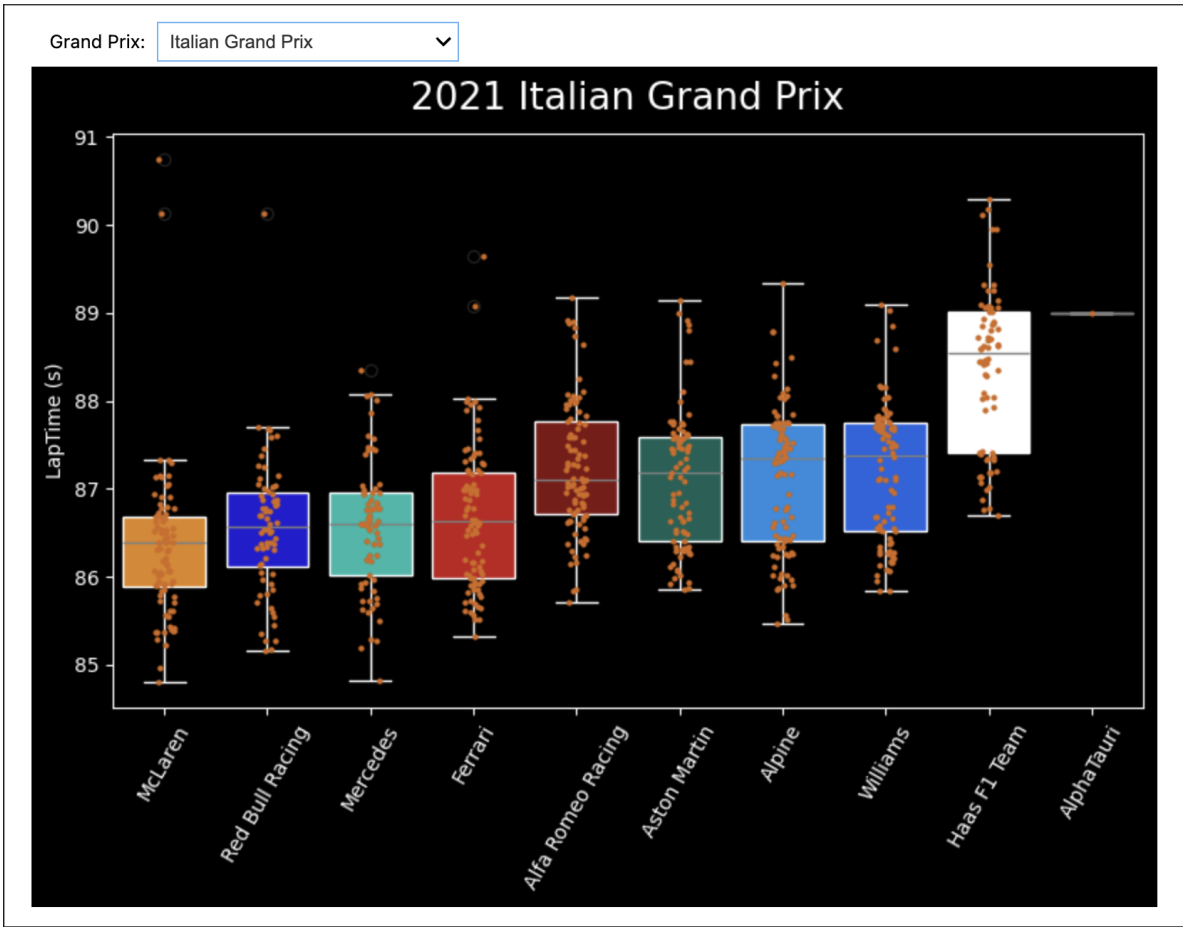


Figure 29: Boxplot of constructors lap times for every race in 2021

From the above boxplots, we can see the constructor's lap times(average of two drivers' laps) for every race in 2021. The example above shows the lap times in the Italian Grand Prix, where the circuit is pretty straight and does not have many corners(to analyze the corners please check the figure below). So McLaren has the fastest laps and we can deduce that they perform pretty well on straights compared to other top teams. AlphaTauri has been disqualified during their initial laps and hence we cannot see the lap times of both drivers.

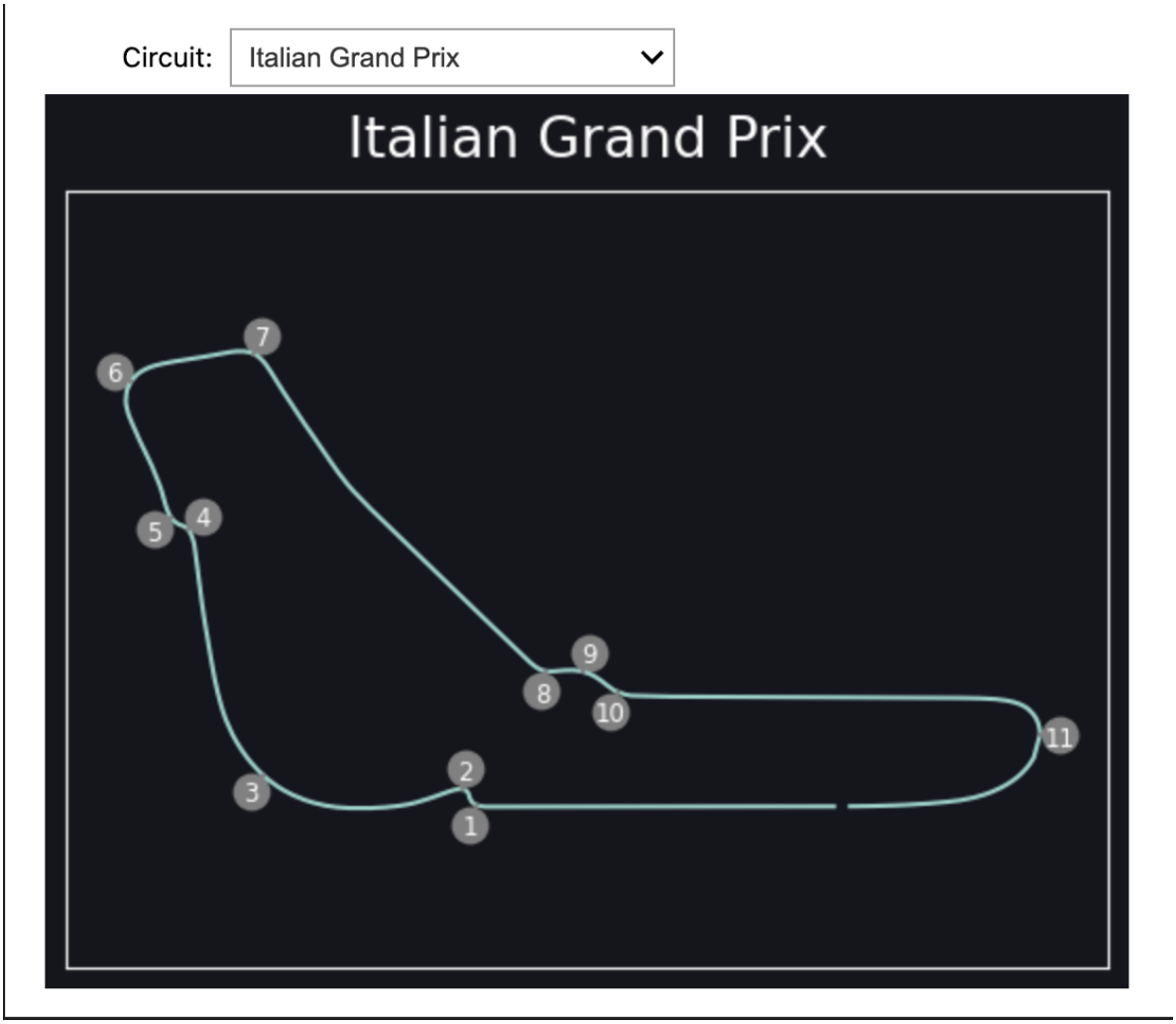


Figure 30: Corners plot for Italian GP

Here is a fun live race simulation of Las Vegas Grand Prix 2023!

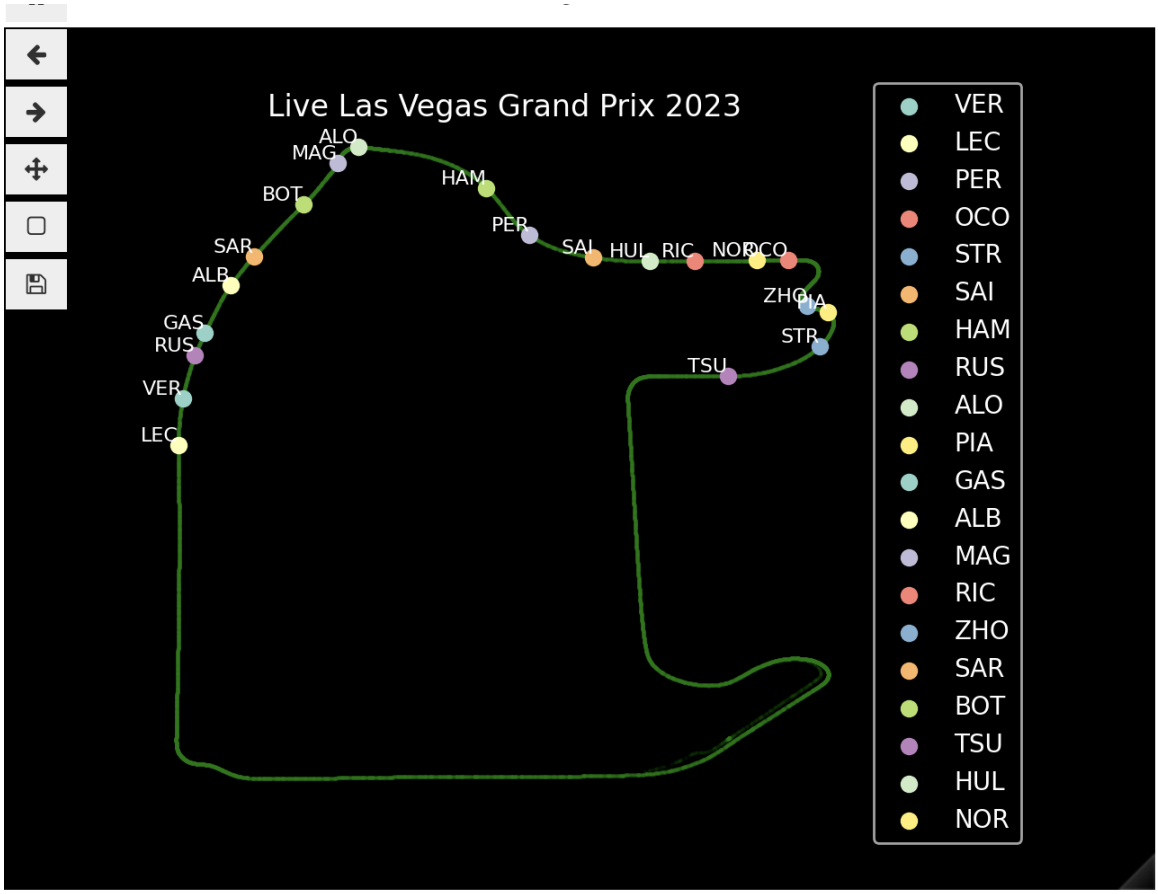


Figure 31: Live race simulation

3.1 Results and Conclusion

Circuit Analysis: The project extensively analyzed driver performance on different circuits. Visualization techniques like Folium were employed to present a visualization of different Grand Prix locations across the globe. Visualization of the corners in a race circuit and a speed heat map provided a brief overview about the importance of speed optimization during corners. Constructors' performances from 2021 to 2023 were examined through stacked bar plots, revealing the evolving competition landscape over time.

Nationality Analysis: Using Radial charts the different nationalities for teams and drivers which have participated in Formula one from 1950 to 2023 are shown.

Constructor and Championship Analysis: The historical trajectory of championship victories by constructors since 1950 was charted, capturing the long-term Champions.Top three constructors – Mercedes, Ferrari, and Red Bull – were compared head to head, uncovering highs and lows that shaped their trajectories in the past decade.

Driver Performance Analysis: Dynamic narratives of drivers' lap times and positions across circuits from 2019 to 2023 were presented, highlighting the evolving dynamics of individual and collective performance. An interactive interface detailed the anatomy of each circuit, focusing on corners and their significance in the racing narrative. Bar plots depicted qualifying results, offering insights into pre-race dynamics.

Temporal and Comparative Analysis: Violin plots illustrated drivers' lap times across circuits, providing a snapshot of the top 10 driver's speeds. Joint plots with drop downs for every year and team allowed for head-to-head comparisons of teammates.

2021 season analysis: The 2021 season regarded as an exciting season has been thoroughly explored. Detailed comparison of points and championship positions between Max Verstappen and Lewis Hamilton show how the rivalry shaped and give a better picture of the outcome.

Live-Action Visualization: An animated plot provided real-time-simulation visualization of the 2023 Las Vegas Grand Prix, adding a live-action dimension to the analysis.

3.2 Future Work

In our current undertaking, we had aspired to incorporate weather data into our analyses, recognizing its potential impact on Formula 1 races. However, the limitations in the availability and quality of the data at present prevented us from realizing this aspect as originally envisioned. Nevertheless, we acknowledge the importance of weather conditions in racing outcomes and, with the anticipation of improved data resources in the future, we aspire to explore and integrate weather-related insights into our analyses. Additionally, we had planned to delve into the intricate relationships within Formula 1 teams by plotting network graphs of all teams and their respective driver pairings. Regrettably, the lack of specific data detailing connections between teammates hindered the realization of this visualization in our current scope. As part of our future work, we intend to address this limitation by actively seeking and incorporating more comprehensive data, enabling us to unveil the intricate dynamics and collaborations within Formula 1 teams through network graph visualizations. We also plan to use a dashboard to represent all these visualisations in the future.

4 References

1. https://jasonjpaul.squarespace.com/
2. https://www.f1-tempo.com/
3. https://www.f1-fansite.com/f1-calendar/dutch-grand-prix-f1-race-info/
4. http://ergast.com/mrd/
5. https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020.
6. https://github.com/toUpperCase78/formula1-datasets/tree/master
7. https://www.formula1.com/
8. https://github.com/theOehrly/Fast-F1