1 Introduction

In the high-speed, high-stakes world of Formula 1, the difference between victory and defeat often lies in the fractions of a second. This realm of roaring engines and lightning-fast reflexes may seem an unlikely place for data science to make its mark, yet this is exactly where it thrives. Just like a well-oiled engine, the gears of this cutting-edge discipline hum in harmony with the rhythm of F1. From race strategy to car design and driver performance, data science is the invisible co-pilot, charting the path to the podium. In this project, we delve into how data science applications are revolutionizing the world of Formula 1, making it a sleek, digitally driven sport.

“Being second is to be the first of the ones who lose.”

* Ayrton Senna
  1. Formula-1

In Formula 1, drivers control high-speed vehicles to be the first to complete a set number of laps, thus securing victory. However, it's not merely about racing; in Formula 1 teams of [drivers, mechanics, engineers?] work together to integrate technological innovation, strategic planning, and driving skills. Each season, teams participate in a sequence of races, known as Grands Prix, held on various circuits worldwide. These races occur over a weekend, commencing with two practice sessions on Friday, an additional one on Saturday, and a qualifying round. The final race, the true test of endurance and speed, takes place on Sunday.

During Friday's practice session, drivers familiarize themselves with their vehicles to mitigate issues such as understeering or unstable handling. Their performance during Saturday's qualifying rounds is crucial as it determines their starting position for the Grand Prix. A superior performance leads to an advantageous starting position. A lower starting position, on the other hand, imposes a significant challenge, compelling the driver to overtake up to 19 other drivers, each traveling at speeds exceeding 200mph!

Qualifying consists of three segments: Q1, Q2, and Q3, each a time-bound knockout round. In Q1, all 20 drivers have 18 minutes to record their fastest lap. The slowest five are eliminated, securing the final five slots on the grid. Q2 follows, lasting 15 minutes, with the five slowest drivers once again eliminated, determining grid positions 11 to 15. The top drivers advance to Q3, the top-ten shootout, lasting 12 minutes. Here, the fastest driver secures the coveted pole position, followed by the second-fastest starting second, and so forth. Qualifying is a tactical play that influences the race. The top ten qualifiers must commence the race on the tires used to clock their fastest lap in Q2. This rule forces teams to strike a balance between a quicker qualifying time and an optimal race strategy. Meanwhile, those who didn't advance to Q3 have the liberty to choose their tires, occasionally allowing them to deploy innovative strategies to compensate for their grid position deficit.

The top 10 finishers in each race earn points. The driver securing the first position receives 25 points, the second-place finisher 18 points, the third-place finisher 15 points, and so on, down to the tenth-place driver who is awarded 1 point. Points are accumulated by drivers and teams throughout the season, and the one amassing the most points claims the championship.

Pit stops in F1 are precise and efficient. In mere seconds, mechanics refuel the car, replace tires, adjust aerodynamic elements, and return the car to the race. Pit stops are integral to the team's strategy. The timing and frequency of stops can dictate the race's trajectory, offsetting the time lost in the pit against the advantage of fresh tires.

Weather forecasting, vehicle health diagnostics, competitor analysis, real-time decision-making - all aspects of F1 hinge on data. Teams analyze gigabytes of telemetry data relayed from sensors on the cars to the pit wall and back to their home bases. This data-centric methodology equips teams with the insights necessary to make pivotal decisions, from vehicle setup to race strategy.

Now that we've understood the format and the rules, here's a secret - a big part of winning isn't just about having the fastest car or being the best driver. It's also about outsmarting your competitors. Enter data science - a realm that approaches the speed limit cautiously, helmet securely fastened, and a smirk on its face. But more on that in the chapters to come! For in F1, much like in data science, speed matters, but strategy takes the chequered flag!

1. Data

Formula-1 racing is a complex sport, here success is determined by a multitude of factors. Our project aims to explore the intricate relationship between weather conditions, pit-stop strategies, safety car deployments, and the underlying data structure that enables this analysis.

* 1. Describing our data

2.2.1 Schedule Data

Schedule data includes information about the Formula-1 events:

* - RoundNumber: Number of the event.
* - Country: The country hosting the event.
* - Location: Location of the Event.
* - Official Event Name: Name of the event.
* - EventDate: Date of the event.
* - EveentName: Event Name.
* - Event Format: Format of the event.
* - Session1, Session2, …: Details of the different sessions.

{space for schedule data table}

2.2.2 Session Data

Session data provides information about the drivers, teams, and their performance:

- DriverNumber: Driver’s number.

- BroadcasrName: Driver’s broadcasr name.

- Abbreviation: Driver’s abbreviation.

- DriverId: Unique team ID.

- TeamName: Team Name.

- TeamColor: Team color code.

- TeamId: Unique team ID.

- FirstName: Driver’s first name.

- FullName: Driver’s full name.

- Position: Driver’s position in the race.

- Status: Driver’s status in the race.

- Points: Points earned in the race.

{Space for session data table}

2.2.3 Car Data

The car data provides detailed information about various parameters related to the car’s performance during a race. It includes:

- Date: Timestamp of the data.

- RPM: Revolutions per minute of the engine.

- Speed: Speed of the car in km/h.

- nGear: Current gear.

- Throttle: Throttle position (0 to 100).

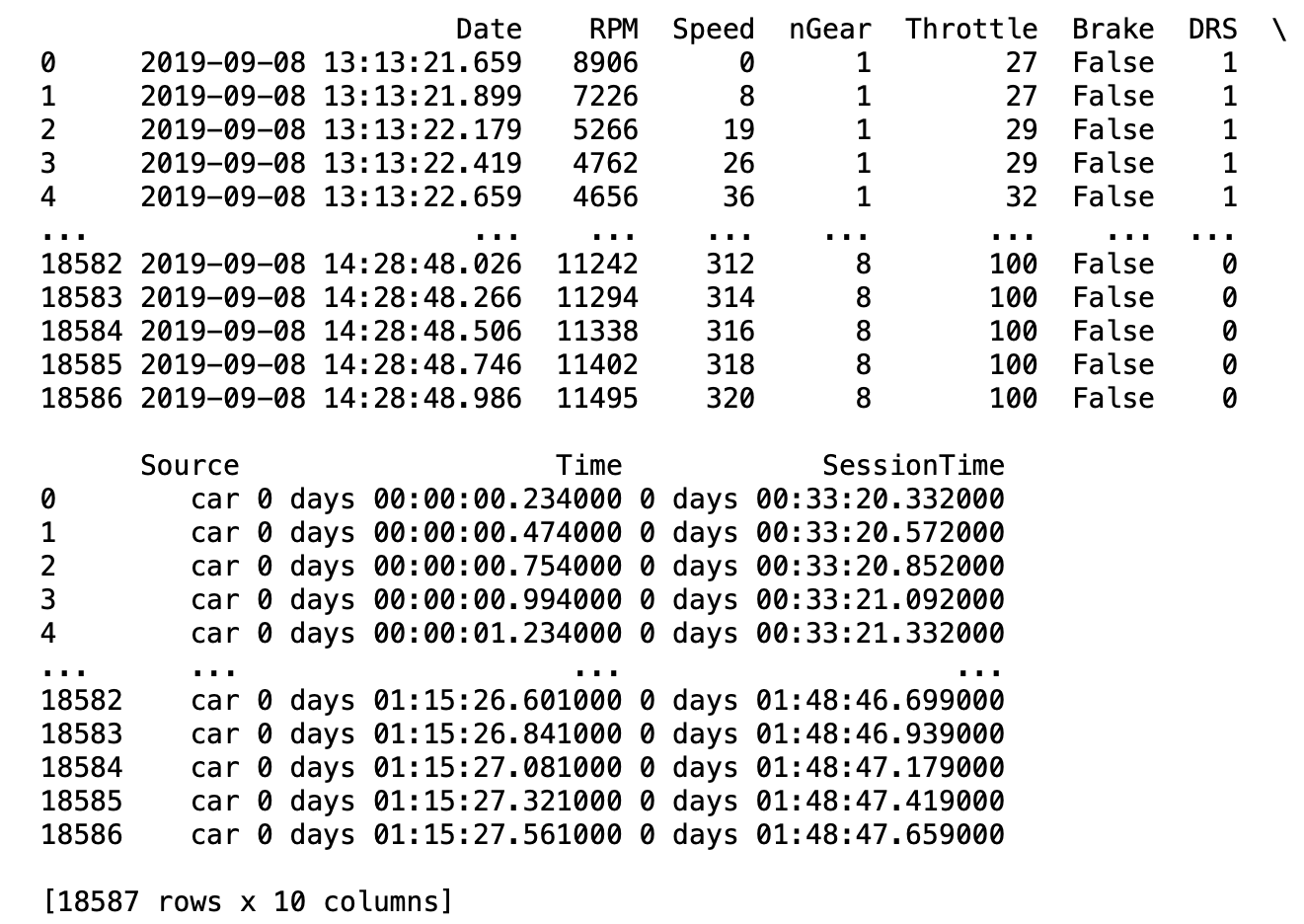
- Brake: Brake status (True/False).

- DRS: Drag Reduction System status.

- Source: Data source.

- Time: Time since the start of the session.

- SessionTime: Time since the start of the race.



* + 1. Stint Data (Tire Data)

Stint refers to the portion of a race between pit stops. Drivers divide each race into several “stints,” over which the performance of a car can vary due to tire condition, fuel load, and other factors. The tire data provides information about the sints, including:

* Driver: The abbreviation for the driver’s name.
* Stint: The stint number within the race.
* Compound: The type of tire compound used in the stint (e.g., SOFT, MEDIUM, HARD).
* StintLength: The number of laps completed in the stint.

A chart of different colored rectangular shapes

Description automatically generatedA screenshot of a computer screen

Description automatically generated

Soft = Red, Medium = Yellow, Hard = White

* + 1. Weather Data

Weather data includes information about the weather conditions during the race:

[---working on it---]

* 1. Fun Interpretations from our data

I will use violin plot and swarm plot; Violin plot will show the distribution of lap times for each driver, displaying the density of different time values. It is called a violin plot because of its symmetrical shape. This allows us to see not only common lap times (like median) but also how the lap times spread. With this we can also compare the performance of different drivers, we can see who was more consistent, who had outliers, etc. The width of the plot at different values indicates the density pf the data at that point. A wider section means more lap times around that value, and a narrow section means fewer lap times (if a driver’s plot is consistently narrow, it may indicate very consistent lap times. Also if there are any usual or unexpected lap times they may show up as narrow “spikes” in the plot).

