Introduction

- why am I doing this, why look at tyres

- what do I want to look at (three questions)

- what did I use

- ergast api

- python packages

- what did I find (claims)

Methods:

- what did I use

- how did I perform my analysis per question

Results:

- what do the results say

- answer each of three questions

Discussion:

- input on what results show

- per three questions

- limitations

Conclusion:

- reiterate interest

- reiterate claims

- reiterate conclusions

- further investigation

Introduction

I wanted to look at tyre performance because there are a lot of factors that vary from team to team. I feel that in order to assess performance compared to other teams and understand yourself as well as your competitors, you need to look at the variables that are consistent amongst all teams.

**Why tyres?**

Tyres are an integral component of a car’s performance, as they are the basis of strategy due to its direct impact on the car’s pace. In Formula 1, Pirelli supplies the tyres that each team uses during a race weekend. There are a total of 5 different tyre options (compounds) that are scaled from more grip/fastest degradation (softer) to less grip/more durability (harder) (Fig 1). Drivers have to heat each tyre to its optimal temperature in order to get the most grip out of it. The harder compounds take longer to heat than the softer compounds. From these 5 compounds, 3 are allowed per race weekend, and it is up to the teams to decide which compounds to outfit the car with at different stages of the race, to put them in the best position to win or score points. For example, if a car is trailing another car by a few seconds in the earlier stage of a race, then they may elect to pit earlier for a harder compound, to get a head start on warming the tyres up and making up ground while the car ahead of them continues to degrade their current tyres.

The purpose of this article is to investigate three main questions regarding tyres in Formula 1:

1. Tyre usage – How long was each tyre used for?

2. How effective was each tire

3. When was each tyre used

To answer these questions, I utilized the ergast API to query lap time data from the 2020 Formula 1 season. The dataset included lap times from each driver in each race, as well as pit stop information. Additionally, I utilized the data published on Race-Fans.net to complete the dataset with information on what compounds each driver was on during the laps of each race. From this collection of data, I was able to utilize the timing information to categorize stints, stint length (km), total race length (km) which were imposed as constraints to compare the data on the same scale.

**Methods**

Formula 1 and motorsports in general, is at the forefront of technological development. Therefore, the industry is no stranger to big data, analytics, and machine learning. While the practices deployed may vary from team to team, data collection and analysis propel research, development, and evaluation in the sport. The cars racing in Formula 1 today are equipped with many sensors, ranging from those measuring tyre temperatures to those monitoring power unit output and characteristics. Drivers and teams can make use of this telemetry data to assess the car’s performance. This data can also provide insight to how the driver is performing at certain parts of the track (when they are on the accelerator/brake).

Because of the granularity of the data collected, much of it is close guarded. Teams go to great lengths to make sure that data regarding their car’s performance is secured to make sure their opponents can’t gain any insight on their strengths or weaknesses. Detailed telemetry data from the cars is not published or publicly accessible, therefore this is a limitation on the analysis that can be done from outside the industry.

However, I was able to make use of the Ergast API [1] which allowed me to collect lap time data from the 2020 season. This contained the lap timing data from each race for every driver. Pit stop information was also included. Additionally, I found that the weekly race recap articles published on race had information on the tyre compounds each driver was on each race, and the duration (in laps) that they ran them. By merging this data, along with circuit information, I was able to create a comprehensive dataset that would allow for this tyre analysis with respect to time and distance. Calculations were performed with existing data to determine stint length, total distance, and accumulated time.

I used a combination of DbBrowser/SQLite to compose the database, and used Python 3/Jupyter notebook to generate plots.

1.) How long was each tyre used for?

To determine how long each tyre compound was used for I split the data by compound. For each compound, I calculated the distance traveled for every stint of every race. I then grouped the data by constructor (team) to create the probability density plot which gives insight to how frequent the team was on each compound at a given distance (km).

2.) How effective was each tyre?

To determine a tyre’s effectiveness, I split the data by compound. Then, using the timing data I calculated the average lap time for every stint. I took the difference between the actual lap time and this stint average lap time to generate a time delta. I felt this was an appropriate approach as the delta was computed for each driver, each race, each stint and this was the closest I would get to weighting everything equally (comparing an early stint at Monza vs a late stint at Bahrain wouldn’t make sense).

I then plotted all deltas for each compound against stint distance in km. A higher positive delta reflects slower lap times (or lap times slower than the stint average) while negative deltas reflect faster lap times (lap times faster than stint average).

I then used Loess regression to generate a best fit curve. Loess regression essentially performs a weighted least squares calculation at different intervals to generate a prediction value. Then, the fitted curve gives more emphasis to values that are closest to the prediction value. This gives more weight to values that are closer together rather than having outliers skew the plot.

3.) When was each tyre used?

To determine when each tyre was used, I took the difference from the accumulated distance from each lap of each race and the total distance from each race. This results in the variable of distance remaining. Similar to observing the distance each tyre was used, I plotted the data grouped by constructor in a probability density plot.

**Results**