# MODELING THE OPTIMIZATION FOR FORMULA 1 RACING

Final Project Report
MATH 545 and MATH 618

Morgan Buterbaugh, Jaylee Lassinger, Josh Pettenó
Indiana University of Pennsylvania
May 2024

Title: Modeling the Optimization for Formula 1 Racing

Leveraging Formula 1 race data retrieved and stored within the Python library 'fastf1', we can construct simulation modules to predict optimal tyre strategies for the average Formula 1 team during a Grand Prix race. These modules incorporate data on tyre degradation, considering lap times as a function of tyre age (number of laps run), alongside data pertaining to pit stop time loss. By analyzing data related to the various tyre compounds and their corresponding lap times, degradation coefficients and baseline parameters for each compound are established. Subsequently, these values are implemented within mathematical models designed to estimate total race time and optimal tyre usage, including the number of pit stops and corresponding stint lengths.

TABLE OF CONTENTS

List of Figures

# CHAPTER 1

## INTRODUCTION

Formula 1, often abbreviated as F1, is considered the pinnacle of motorsport racing, known for its speed, technology, and glamour. It's not just a race; it's a blend of cutting-edge engineering, precision driving, and high-stakes competition that captivates over 430 million fans worldwide, surging as one of the most-watched sports in the world. It is also an extremely lucrative sport: in the United States alone, Formula 1 Group's revenue exceeded 2 billion U.S. dollars in 2021.

At its core, Formula 1 is now a series of 24 races held on circuits around the globe, with each race weekend consisting of practice sessions, qualifying laps, and the main event - the Grand Prix race itself. There are usually 3 free practice sessions, 1 qualifying session, and the main race. The Grand Prix is a 3-day event, in which usually 5 sessions take place. During the free practice sessions, the cars are fine-tuned aerodynamically, and data is collected to have the best set-up and strategy for the main race. The qualifying session determines the starting position of each car for the main race, in which the drivers compete to get to the finish line in front of the others. During some events, the amount of free practice sessions is reduced to one, and two of the sessions are being replaced with a sprint race and its qualifying session. A sprint race works exactly like a normal race, but the amount of laps is reduced to one-third of the main race. These events are not as common, but are nonetheless important, as the sprint race still yields points to the first 8 drivers, although the amount is greatly reduced with respect to the main event. These races are the culmination of years of development by teams, engineers, mechanics, manufacturers, and strategists, as well as the skill and daring of the drivers working together to achieve the best team performance.

Teams play a crucial role in Formula 1, with 10 teams consisting of 2 drivers and a group of engineers, mechanics, and strategists. The teams are in a constant battle for supremacy, both on and off the track, as they strive to develop faster cars, outsmart their rivals, and optimize the fastest race time. Points are given to the first 10 drivers who complete the race, and the points are distributed unequally and proportionally between the top 10. The driver that wins the race will get the most points (25), the tenth driver will receive just one point. For each team, getting to the highest place in the constructors' leaderboard is fundamental. Being in the highest seeds at the end of the season coincides with significantly higher cash prizes. Money is essential in this highly technological sport, and every part of the game needs to be optimized. Race strategy and the right use of data are therefore extremely important for the success, or even just the survival of a Formula 1 team. This is why the role of a data scientist is one of the most important among the team. Great strategists can analyze data and make accurate predictions for the race, leading to better yielding choices. This means that data scientists compete to get the best mathematical models. It is therefore a great environment to test our data science skill sets and mathematical knowledge.

The cars in Formula 1 are designed for ultimate speed and performance. They are built to be aerodynamic, lightweight, and incredibly powerful, capable of reaching speeds over 350 kilometers per hour. The heart of an F1 car are its engine and chassis. However, the intricate strategies surrounding tyre management is one of the biggest factors to consider. These high-performance cars run on specialized tyres designed to provide optimum grip, durability, and performance across a variety of track conditions.

Each racing team has a crew of race strategists who analyze data and try to predict and optimize the race time. The analysts use information gathered before the race to come up with mathematical models. They then use these models to make an accurate prediction for the race development as accurately as possible. Finally, using real-time data monitoring during the race, they are able to foresee and quickly react to continuously changing circumstances.

**Background Information**

In Formula 1, there are 5 different tyre compounds provided by the official tyre supplier, Pirelli: C1, C2, C3, C4, and C5. Each compound has distinct characteristics that affect a car's handling, speed, and longevity. The compounds are ordered from the hardest compound, to the softest. Although there are 5 different compounds, only 3 are chosen and given to teams at each race: the softest one of the lot will be regarded as the Soft compound, signaled by a red band on the sides. The hardest compound will be the Hard tyre, signaled with a white band on the side, while Medium tyres are yellow and are intermediate with respect to the other two. Soft compound tyres offer the highest level of grip but tend to wear out more quickly. Thus, the degradation of the tyre is high, while the base time for this tyre compound is low. They are typically used for qualifying sessions and maximum performance. They provide excellent grip around corners, allowing drivers to push the limits of their car. Medium compound tyres strike a balance between grip and durability. They offer decent performance while lasting longer than soft tyres. The degradation for the medium compound is less than that of the soft compound and the base time for this tyre compound is higher. Medium tyres are often the preferred choice for longer stints during the race when teams aim to extend their tyre life without sacrificing too much speed. Hard compound tyres are the most durable but offer less grip compared to soft and medium compounds. The degradation for this compound is the lowest, whereas the base time is the highest. They are primarily used during the later stages of a race or when track conditions are particularly abrasive. Hard tyres allow drivers to maintain consistent lap times over longer distances but may lack the outright speed of the softer compounds. Therefore, the compounds are ordered from quickest to slowest and from least to most durable, respectively.

In addition to these three primary compounds, Pirelli also produces intermediate and wet tyre compounds for use in wet weather conditions. These tyres feature special tread patterns and compounds designed to disperse water and provide grip on a wet track surface, ensuring safety during rain-affected races. For this study, we are only focusing on the dry tyre compounds, excluding the two special compounds. It is important to note that during each Grand Prix, a tyre compound change is mandatory, meaning the car must use at least two different tyre compounds for the race. The driver is allowed to change between multiple tyre compounds, as long as all four tyres are the same compound. Teams must analyze factors such as track temperature, track surface, tyre wear, and degradation to determine the optimal tyre strategy for each race. Making the right call on tyre selection can mean the difference between victory and defeat.

The choice of tyre compound and the timing of tyre changes, known as pit stops, play a crucial role in a team's race strategy. The time for a pit stop starts as soon as the car exits the track all the way until the car re-enters the track. Thus, the number of stints is equal to the number of pit stops + 1. Hence if the driver took one pit stop during the race, there would be two stints for that driver's Grand Prix.

**Problem Description**

The objective of our study is to come up with an equation that is able to determine the total amount of time for a driver to finish a race, and then minimize the output value for that function. Our goal is to come up with the optimal strategy before the race. Since it is a competition influenced by other teams' choices, plans likely change as the race develops. The team's objective is to finish first, not to finish as quickly as possible, thus the real-life decisions made in the garages might differ from what could be the best decision regarding race time. Nonetheless, time minimization is a good indicator of what the best decision could be based on ideal conditions.

To build our model, we have to understand the determining factors for race time:

- Track Specifications: Every race track is different. The layout, the number of turns, the type of asphalt, and the geographical location are all factors that determine how quickly a driver can complete a race.
- Weather conditions: Race cars are heavily affected by weather conditions. The slight change in grip that is given by shifting track temperature and the presence of water can heavily affect the drivers' lap time. The wind intensity and direction are also impactful, as they can increase the aerodynamic load, as well as increase the car's instability through different sectors of the track.
- Race developments: Formula 1 races are generally unpredictable, and many factors might impact race time. For example, safety cars are a tool for race officials to slow down the drivers when needed. They are usually deployed when an accident occurs, and the race track needs to be cleaned before drivers can safely return to compete at high speeds. Also, driver battles for the position usually increase lap time and tyre wear, therefore increasing overall race time.

- Tyre and pit-stop strategy: This is arguably the most important decision for a strategist. Tyre compounds in Formula 1 differ in performance and durability. It is important to minimize the number of pit stops in the race, as well as the appropriate tyre choice for each circuit to optimize race time. Concepts like undercut and overcut are also important factors to consider when it comes to the ideal time to change tyres, in the optics of maintaining racing position against rivaling teams and drivers.
- Car performance: Not all cars are the same. All cars are designed and produced by the different teams. Some parts of the car can be bought from other teams, but there are parts of the car that have to be designed from scratch, as per FIA rules (International Automobile Federation). Thus, some teams can perform better than others with their specific car build.
- Driver skills: Just like cars, drivers are also very different. Each driver has a set of specific strengths and weaknesses, and their driving style might be more suitable for one type of circuit compared to a different one. Driver mistakes also increase race time.
- Fuel calibration: Cars are heavily affected by weight. Excessive fuel in the tank can cause the car to slow down. Thus it is important to calculate the appropriate amount of fuel for the race. The maximum fuel flow has been set by FIA rules, so it has become a less relevant factor. Even so, fuel is something to take into consideration when exploring data, as the cars get lighter each lap, which thus impacts lap time.

**Literature Review**

Formula 1 analysis and exploration is not a new concept. Each Formula 1 team has hundreds of workers who are continuously crunching numbers and analyzing each race. Tyres, fuel, load, and speed, are just a few elements the analysts focused on, with the goal in mind of optimizing the best race time. While the analysts of each team have more information and better technology, many others have examined other areas of Formula 1 using a few different approaches.

The implementation of neural networks has been used to solve similar problems. A group of three professors worked on tyre prediction for the different Formula 1 tyre compounds [10]. A neural network model was created to approximate the tyre performance. The network attempted to make predictions on the condition of the tyre after each lap and determine whether or not the tyre would be too worn out for another lap, meaning it would slow the car down. For their neural network, they decided to use lap times, track features, current tyre compound (in the form of expected tyre life), and weather. Once the network was set up, it was then trained and evaluated. The parameters of the neural network went through the process of backpropagation, utilizing the adaptive gradient. They used K-Fold Cross Validation from SKLearn in Python for the validation of their neural network.

There are a few resources that explain how to get the data and create the equations that are needed to achieve the optimization result. One is written by Kerberz Engineering [5], it explains the simulator and the objective function that they used in order to find the optimal expected race time. The parameters that were used were the number of pit stops, tyre compounds for every stint, and the stints' length in laps. These were used to then calculate the expected race time and each time was then compared to the optimal solution in order to see the effect that the different information for the parameters had on the race time. The equation that was used was that lap time = base lap time + fuel correction + tyre correction + random disturbance. Their program could also provide expected tyre wear per lap for every wheel and for each compound, it also could provide maximum stints for every compound. They also were able to add additional parameters for other aspects that could affect race time on the track including safety cars and other people on the track. This website provided useful information and an idea of what the objective function should look like and what factors affect this equation.

There are also research competitions and online forums for optimizing different aspects of Formula 1. These offer some insights into useful optimization problems for circuit data, lap time, pit stops [7] [8], and tyre degradation [6].

# CHAPTER 2

# DATA ANALYSIS

The data we used for this study came from the 'fastf1' library. This Python package contains session results, timing data, and telemetry from all Formula 1 events since 1950 [4]. We will focus only on the current era of the sport, as different eras bring different rules and specifications, which then alter the consistency of the data. From the problem statement, it is easy to see that there are many factors, some of them being hard to predict, that can prevent us from finding an accurate or mathematically describable model. Hence, we opted to simplify our model by making specific choices, and assumptions about a race.

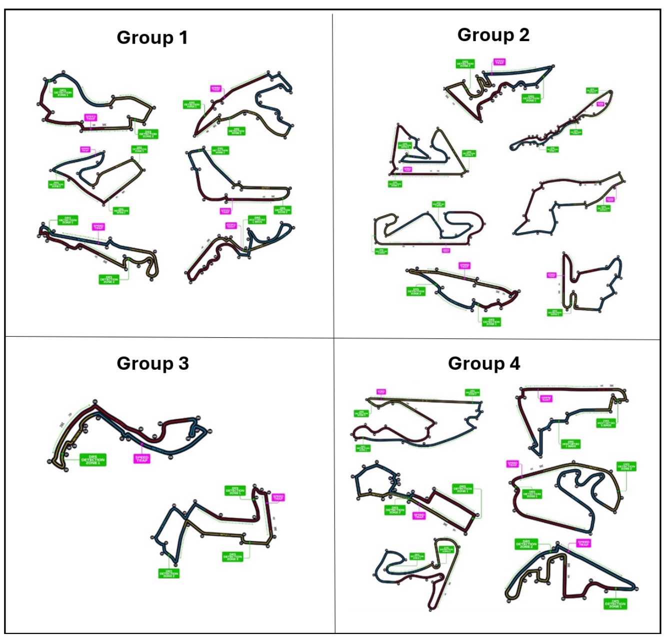
For this study, we will only be focusing on the 57-lap Bahrain track in Sakhir. We chose to focus on the F1 track in Bahrain for multiple reasons. The most important one being all of the pre-season testing for each team has been done here in this current era, leaving us with more data than we would have for other tracks. Since this is the first F1 track of the season, the race is held in February each year. The weather conditions are consistent at this location, dry and warm, and it is a high degradation circuit, meaning that it is easier to model linearly.

We will be focusing on pre-race strategy only and use all the drivers' and teams' data to get an average estimate for lap time, tyre compound age. We elected to use the data from all 3 testing days and 3 free practice sessions. This included the lap time, the number of laps, the time going into the pit and the time coming out, and the compound of the tyre. We only used the data records for ideal track conditions and accurate lap time. We neglected the possibility of race interruptions, such as an accident or a safety car. Moreover, fuel consumption is negligible since the data does not allow fuel correction. This is because we chose to use only data coming from practice and test sessions. The teams know how much fuel they introduce, but no one outside of them and the FIA officials know the exact amounts during practice. While estimates can be done for a race, the same cannot be done in practice, as teams can come in and out however and whenever they like. Thus, our target becomes getting to the finish line as quickly as possible. The only remaining decision left for us to find out is the number of pit stops and the type of tyre compound to use. To do so, we have to collect data properly so that we can build proper degradation and performance models for each tyre compound, and the parameters that affect them. In this way, we can then fit functions that can describe the tyre behavior accurately.

## Exploratory Data

To get familiar with how 'fastF1' worked, as well as to gather useful knowledge, we decided to see if different tracks had similar features so that our strategy applied in Bahrain could also work for other tracks too. There are many factors that could be included to identify specific characteristics of a track, from average track temperature and asphalt 'abrasion' coefficient to the number of turns with respect to the speed at which they are taken, which can be even better explained by the quantity and duration of lateral g-forces are applied to the car, to the highest speed that it can be reached by the car, the average speed, and the percentage of time the driver can push at full throttle, during an entire lap. The 'fastF1' library allows corner classification, but for time constraints, as well as a lack of knowledge of the library tools, we decided to explore the data through the last 3 factors, which were easier data to gather through the telemetry data of the race. Through a K-Means algorithm, and the use of the Elbow method to determine that all the circuits were better represented by 4 clusters, we discovered that according to the algorithm, the Bahrain circuit was in the same group as the tracks in Texas, Saudi Arabia, Spain, Hungary, Canada, and Imola. That is a group of very fast circuits, with a few slower turns, that diminish overall the average speed, and the amount of time drivers can go full throttle around the course. The results of this 4 K-Means cluster can be seen in Figure 2.1. There is not enough data to say that similar strategies can apply to these races, as mechanical grip levels are quite different in these locations, but it allows us to gather information about track-layout similarities.

Figure 2.1. *4 K-Means Clustered F1 track groups.*




For all three years we looked at Bahrain, we had to load in the respective testing days and free practice sessions. Once we had this loaded in we built a data frame that housed all the data, dropping all unknown tyre compounds and compounds that did not have know testing data.

To visualize the data set, scatter plots were created to find the distribution of the three tyre compounds in relation to the tyre life and lap time (found in Appendix - A). Since much of the data overlapped, a box-and-whisker plot was created to help better visualize the data. Looking at the box-whisker plot we can see that it is easier now to categorize each tyre compound. We want to use the data that is in the IQR for each compound.

For 2022, we see that there are a lot of outliers, especially for both the medium and hard tyres, which we consider removing. These can be explained by either slow laps, which the driver may be taking to warm up the tyres, or to recharge the battery, or in-laps and out-laps, which are laps done to come in or out of the pits respectively. Other outliers, such as fast laps, are the laps the driver took to practice for the qualifying round, which is not relevant or indicative for the race pace. Although it is unusual for qualifying runs to happen on harder tyre compounds, it is a possible occurrence, especially in testing. Thus, we remove the outliers from both the medium and hard compound tyres. After the outliers were removed, we were left with a scatter plot that has a group of soft tyres in both the slow and fast lap times for the first few laps. This can be explained by the slow laps taken to recharge, and the fast laps taken to simulate qualification laps. Since these are not indicative of race pace, we choose to remove these points from our data set. Through the process of observation, we found the best cut-off time to be data above 110 seconds and below 94 seconds. This same exact process was used for both 2023 and 2024 as well, with the cut-off time for the data being above 110 seconds and below 92 seconds and above 110 seconds and below 91 seconds respectively.

Once all the data was cleaned, we were able to train the data using a linear regression model for each tyre compound, using both the number of laps per compound and the lap time. Using the packages from the 'SKLearn' library in Python, we divided the whole data set into 3 compounds, soft, medium, and hard, and used a linear regression algorithm to find the predicted lap time equation. For each tyre compound, we trained a linear regression model to understand the relationship between tyre life and lap time. This involved using the number of laps per compound as the independent variable and lap time as the dependent variable. Our model aims to find the equation that predicts lap time based on the number of laps the tyre has been used.

## CHAPTER 3

## RESULTS FROM DATA ANALYSIS

After running the linear regression model on the three tyre compounds for 2022-2024 Bahrain, we found the succeeding data.

### Bahrain 2022

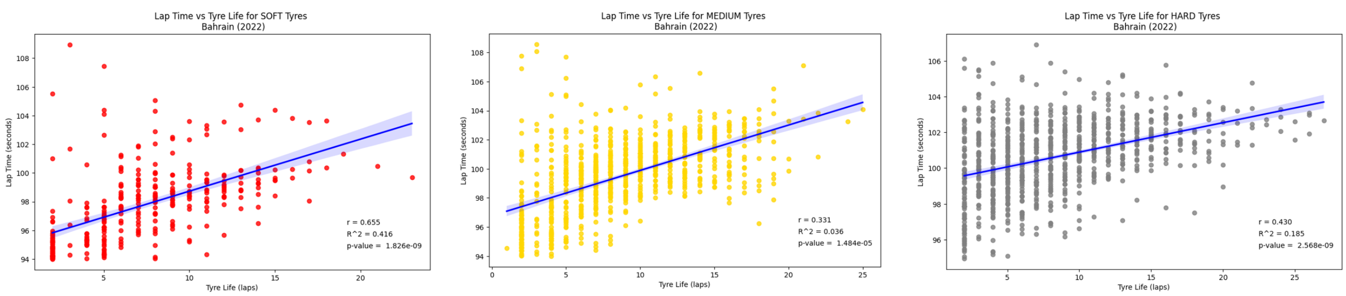
The linear regression equations for the soft, medium, and hard compound tyres used during the three testing days and three free-practice sessions can be seen below. Figure 3.1 displays the linear regression line plotted on each of the respective tyre compound scatter plots.

$$\text{LapTime} = 0.38 \times l_s + 95.09$$

$$\text{LapTime} = 0.30 \times l_m + 96.99$$

$$\text{LapTime} = 0.17 \times l_h + 99.17$$

Figure 3.1. *Linear regression scatter plots for soft, medium, and hard tyre compounds for Baharain 2022.*




### Bahrain 2023

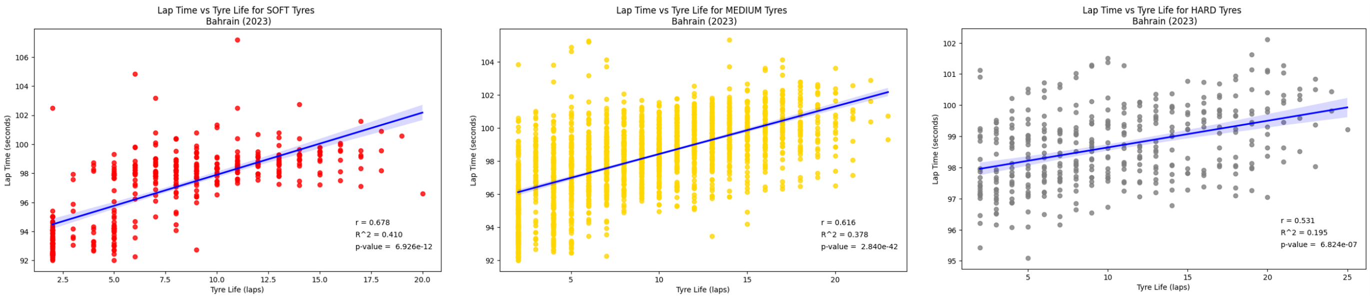
The linear regression equations for the soft, medium, and hard compound tyres used during the three testing days and three free-practice sessions can be seen below. Figure 3.2 displays the linear regression line plotted on each of the respective tyre compound scatter plots.

$$\text{LapTime} = 0.43 \times l_s + 93.66$$

$$\text{LapTime} = 0.23 \times l_m + 95.24$$

$$\text{LapTime} = 0.05 \times l_h + 97.05$$

Figure 3.2. *Linear regression scatter plots for soft, medium, and hard tyre compounds for Baharain 2023.*




**Bahrain 2023**

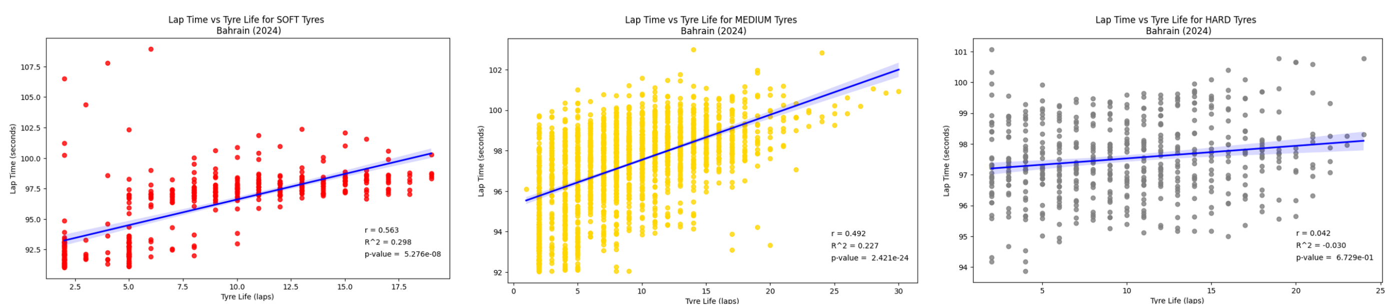
The linear regression equations for the soft, medium, and hard compound tyres used during the three testing days and three free-practice sessions can be seen below. Figure 3.1 displays the linear regression line plotted on each of the respective tyre compound scatter plots.

$$\text{LapTime} = 0.40 \times l_s + 92.55$$

$$\text{LapTime} = 0.23 \times l_m + 95.21$$

$$\text{LapTime} = 0.04 \times l_h + 97.12$$

Figure 3.3. *Linear regression scatter plots for soft, medium, and hard tyre compounds for Baharain 2024.*



# CHAPTER 4

## Objective Function and Constraints

The objective function that we are trying to minimize is designed to calculate the time it takes a formula one driver to complete a race. We have our function:

$$\mu_{ps} \times n_{ps} + b_s \times l_s + \frac{p_s}{2} \times \sum_{i=0}^{ns_{new}} (l_{s_{new_i}})^2 + b_m \times l_m + \frac{p_m}{2} \times \sum_{i=0}^{nm_{new}} (l_{m_{new_i}})^2 + b_h \times l_h + \frac{p_h}{2} \times \sum_{i=0}^{nh_{new}} (l_{h_{new_i}})^2$$

$$+ p_s \times A_s \times \sum_{i=0}^{ns_{used}} (l_{s_{used_i}}) + \frac{p_s}{2} \times \sum_{i=0}^{ns_{used}} (l_{s_{used_i}})^2 + p_m \times A_m \times \sum_{i=0}^{nm_{used}} (l_{m_{used_i}}) + \frac{p_m}{2} \times \sum_{i=0}^{nm_{used}} (l_{m_{used_i}})^2$$

$$+ p_h \times A_h \times \sum_{i=0}^{nh_{used}} (l_{h_{used_i}}) + \frac{p_h}{2} \times \sum_{i=0}^{nh_{used}} (l_{h_{used_i}})^2$$

.

The constraints that are used are the following:

$$\sum_{i=0}^{ns_{new}} l_{s_{new_i}} + \sum_{i=0}^{ns_{used}} l_{s_{used_i}} = l_s$$

$$\sum_{i=0}^{nm_{new}} l_{m_{new_i}} + \sum_{i=0}^{nm_{used}} l_{m_{used_i}} = l_m$$

$$\sum_{i=0}^{nh_{new}} l_{h_{new_i}} + \sum_{i=0}^{nh_{used}} l_{h_{used_i}} = l_h$$

$$l_s + l_m + l_h = l_{total}$$

$$\sum_{i=0}^{ns_{new}} Round(\frac{l_{s_{new_i}}}{l_{s_{new_i}} + 0.1}) + \sum_{i=0}^{nm_{new}} Round(\frac{l_{m_{new_i}}}{l_{m_{new_i}} + 0.1}) + \sum_{i=0}^{nh_{new}} Round(\frac{l_{h_{new_i}}}{l_{h_{new_i}} + 0.1}) +$$

$$\sum_{i=0}^{ns_{used}} Round(\frac{l_{s_{used_i}}}{l_{s_{used_i}} + 0.1}) + \sum_{i=0}^{nm_{used}} Round(\frac{l_{m_{used_i}}}{l_{m_{used_i}} + 0.1}) + \sum_{i=0}^{nh_{used}} Round(\frac{l_{h_{used_i}}}{l_{h_{used_i}} + 0.1}) = n_{ps} + 1$$

All variables with regards to laps in a stint must be a non-negative integer.

In order to achieve the objective function and constraints that are shown above, there are numerous variables and mathematical concepts that need to be discussed in order to make sense of how the objective function is used to minimize race time. To start we want to focus on the objective function, so looking at the first part of it $\mu_{ps} \times n_{ps}$, is being used to calculate the total time lost to pit stops. This is the time that the F1 car goes in for the pit stop to the time they are back into the race, not just the time it takes to change a tyre or any other tasks that might be done during this time. The variable $\mu_{ps}$ represents the average time lost to pit stops, while $n_{ps}$ represents the number of pit stops that are taken during the race. We chose to manually enter the number of pit stops for each iteration that we ran. We chose the number of pit stops to be equal to 1, 2, or 3, since drivers will typically not take anymore than three during a race and the minimum amount of pit stops that a driver can take is one since they have to change tyre compounds at least once.

The next section of the objective function that we want to look at is $b_s \times l_s + \frac{p_s}{2} \times \sum_{i=0}^{ns_{new}} (l_{s_{new_i}})^2$. We are going to explain this part of the equation in terms of the soft tyre compound, but it can be seen that the same part is added for medium and hard compound tyres as well. To figure out how much time should be added for using a new soft compound tyre, we need to think of taking the integral, or finding the area underneath of a line. The equation of the line that we are finding the area under is $y = p_s \times l_s + b_s$, where $p_s$ is the time penalty if the driver decided to use a soft compound tyre also known as the degradation coefficient. The variable $l_s$ represents the number of laps that are going to be driven on the soft compound tyre and lastly $b_s$ is the base time for using that compound tyre. The graph represents a stint in the race, thus there is a minimum of two stints and a maximum of 4 stints in our specific problem, one more than the number of pit stops. If we think of a graph of a line, we know that if we draw a line down at a specific $l$ value, the area underneath the curve can be broken into a rectangle and a triangle. The area of the rectangle is going to be equal to $b_s \times l_s$ and the area of the triangle is going to be equal to $\frac{1}{2} \times p_s \times (l_s)^2$. In the objective function, each $l_{s_{new_i}}$ represents the number of laps on a soft compound tyre in a specified stint. We need the summation since there could be more than one stint that a driver is driving on a soft compound tyre and this would need to be multiplied by $\frac{p_s}{2}$ as well. This only works for new tyres since for used tyres the graph would not start at $l = 0$, yet at some decided number of laps that have already been driven on the tyre. This leads into the next section of the objective function which is $p_s \times A_s \times \sum_{i=0}^{ns_{used}} (l_{s_{used_i}}) + \frac{p_s}{2} \times \sum_{i=0}^{ns_{used}} (l_{s_{used_i}})^2$.

Again, we are going to explain this part of the equation in terms of the soft tyre compound, but it can be seen that the same part is added for medium and hard compound tyres as well. In order to see how this is calculating the time that needs to be added when a driver decides to use a used tyre, we need to think of the same line $y = p_s \times l_s + b_s$, but instead of starting at $l_s = 0$, we start at $l_s = A$. Therefore, we are trying to find the area underneath the line, or the integral from $A$ to $l + A$. This can be broken into three different sections, two rectangles and a triangle. The bottom rectangle would be the same as previously, it is equal to $b_s \times l_s$, but this was already added once to the equation so we do not have to add it again since $l_s$ is equal to the total laps on soft compound tyres. The area of the second rectangle is equal to $p_s \times A \times l$ and the area of the triangle is equal to $\frac{1}{2} \times p_s \times (l_s)^2$. For these two equations, the summation must be used again since we are breaking the representation of the variables into stints, so this would be the number of laps on a soft compound tyre for a specific stint. The $l_{s_{used_i}}$ variable is equivalent between the two summations. From here we are able to add all of these pieces together in order to have one objective equation that is used to minimize race time by looking at which tyre compound should be used for how many stints and how many laps for each stint.

For the constraints, the first three state that the number of laps that are driven on a tyre compound in the different stints have to be equal to the total number of laps that are driven on that compound. The following constraint is saying that the total laps driven on each tyre compound needs to be equal to the total number of laps in the race. We then have the constraint that is rounding the laps on a new or used tyre, by dividing the variable that we are discussing by itself plus 0.1. When entering this into LINGO, our chosen software to solve this problem, each summation rounds to either 0 or 1. This is telling LINGO that the number of variables that are allowed to have values are one more than the number of pit stops, which is equal to the number of stints that we have in the race. Remember that our $l_i$ variables are in terms of laps per stint on a tyre compound (new or used), which is why that is the amount of variables that are allowed to have values, making the remaining variables equal to 0. The final constraint just states that all of the variables that are in terms of laps in a stint have to be non-negative integers since the racer must complete an integer number of laps.

CHAPTER 5

# RESULTS FROM OPTIMIZATION

## Bahrain 2022

For the year 2022, the values that we had for our variables were the following:

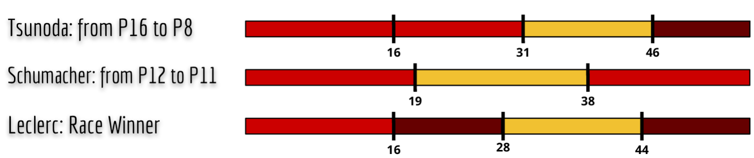
$$b_s = 95.09, p_s = 0.38, b_m = 96.88, p_m = 0.30, b_h = 99.17, p_h = 0.17, \mu_{ps} = 25.15, l_{total} = 57,$$

$$A_s = 4, A_m = 3, A_h = 2$$

We ran numerous iterations of the code with these values for the variables. For the number of new soft compound tyres available there are either 1 or 2, for medium compound tyres there is 1 new tyre, and lastly for hard compound tyres there is 1 new tyre available. There were used tyres of each of the compounds available as well. The top three results that we received were the following:

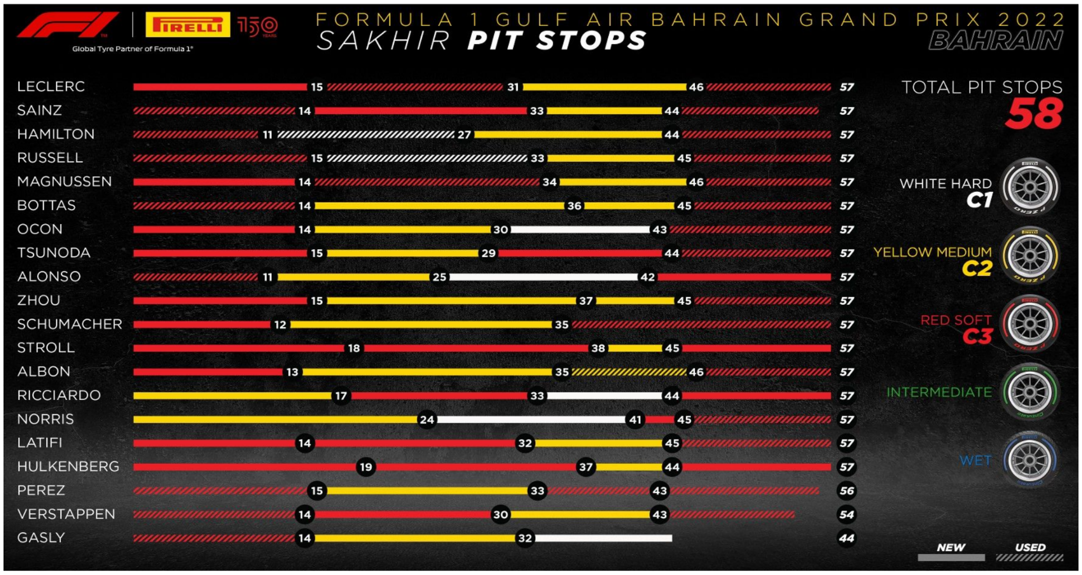
1. 3 Pit Stops, Race Time = 5687.28, New Soft = 16, New Soft = 15, New Medium=15, Used Soft = 11
2. 2 Pit Stops, Race Time = 5695.77, New Soft = 19, New Medium = 19, New Soft = 19
3. 3 Pit Stops, Race Time = 5708.73, New Soft = 16, New Medium = 16, Used Soft = 12, Used Soft = 13

Figure 5.1. *Results found from optimization for the Baharain 2022 Grand Prix.*




For the year 2022, our results are pretty accurate. Figure 5.1 displays our results, and Figure 5.2 displays the actual tyre compound usage for the Grand Prix. Our best strategy was strikingly similar to the one used by Yuki Tsunoda. The Japanese driver was arguably one of the best that day, as he overtook 8 drivers, placing 8th after starting from position 16. It has to be said that per regulations, only the backmarkers of the race were likely to be able to do the first tyre strategy, as the people that go further in the qualifying session have to use more soft tyres, thus losing some strategy flexibility in exchange for better starting positioning. Nonetheless, a great job by Tsunoda.

Figure 5.2. *Actual tyre compound usage for the Baharain 2022 Grand Prix [1].*




The second best option was adopted by Mick Schumacher, the son of Formula 1 legend Michael Schumacher. Although he only finished 11th when he started 12th, he was spun out at the very first lap of the race, having then to recover the time loss to get back on track. The tyre strategy and driver skill allowed loss minimization in a highly compromised race.

Finally, the third best strategy, which equals the best one, but with just older tyres, is very similar to the one chosen by Charles Leclerc, the race winner.

It has to be said that at lap 44, a yellow flag and therefore a safety car, allowed cars to change tyres with minimal penalty loss, but since there were many cars changing tyres just before that occurrence, it means that teams believed that the optimal window to change tyres was around the time predicted by the algorithm.

**Bahrain 2023**

For the year 2023, the values that we had for our variables were the following:

$$b_s = 93.66, p_s = 0.43, b_m = 95.24, p_m = 0.23, b_h = 97.05, p_h = 0.05, \mu_{ps} = 25.15, l_{total} = 57,$$

$$A_s = 4, A_h = 4$$

There is no value for $A_m$ because there are no medium compound tyres, used or new, for this race. For the year 2023, the number of new soft compound tyres available is either 0, 1, or

2, and for hard compound tyres there are 1 or 2 new tyres available. There were used tyres of soft and hard tyre compounds. The top three results that we received were the following:

1. 2 Pit Stops, Race Time = 5590.875 seconds, New Soft = 11, New Soft = 10, New Hard = 36

2. 1 Pit Stop, Race Time = 5597.665, New Soft = 13, New Hard = 44

3. 3 Pit Stops, Race Time = 5603.195, New Soft = 11, New Soft = 11, New Hard = 28, Used Soft = 7

Figure 5.1 displays our results, and Figure 5.2 displays the actual tyre compound usage for the Grand Prix.

Figure 5.3. *Results found from optimization for the Baharain 2023 Grand Prix.*

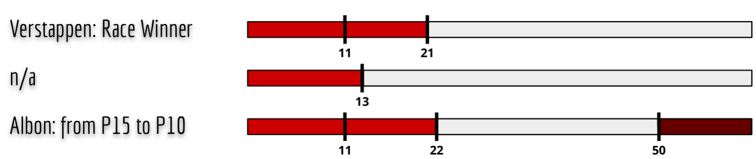
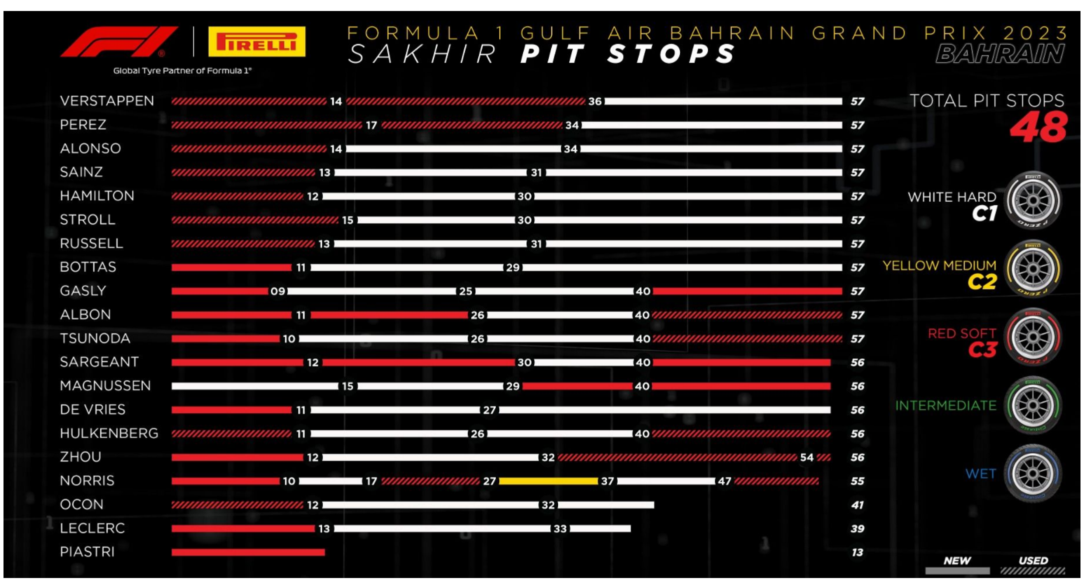



Figure 5.4. *Actual tyre compound usage for the Baharain 2023 Grand Prix [2].*



**Bahrain 2024**

For the year 2024, the values that we had for our variables were the following:

$$b_s = 92.55, p_s = 0.40, b_m = 95.21, p_m = 0.23, b_h = 97.12, p_h = 0.04, \mu_{ps} = 24.72, l_{total} = 57,$$

$$A_s = 4$$

There are no values for $A_m$ or $A_h$ because there are no used or new medium compound tyres for this race and there are no used hard compound tyres for the race either. For the year 2024, the number of new soft compound tyres available are 1 or 2 and for new hard compound tyres there are either 1 or 2. The top three results that we received were the following:

1. 2 Pit Stops, Race Time = 5552.54, New Soft = 14, New Soft = 14, New Hard = 29

2. 3 Pit Stops, Race Time = 5557.9, New Soft = 13, New Soft = 14, New Hard = 21, Used Soft = 9

3. 1 Pit Stop, Race Time = 5572.26, New Soft = 16, New Hard = 41

The race in the year 2024 was the only one that did not have an interruption during the race such as a wreck leading to a safety car or virtual safety car. Thus, we were able to compare our results to the results from that race. The comparisons were the following:

1. +47.798 seconds from 1st Place, Placed 6th

2. +53.158 seconds from 1st Place, Place 8th

3. +67.518 seconds from 1st Place, Place 9th

We can see that through our optimization process we were able to place in the top 10 for each of our top three results.

Figure 5.1 displays our results, and Figure 5.2 displays the actual tyre compound usage for the Grand Prix.

Figure 5.5. *Results found from optimization for the Baharain 2024 Grand Prix.*

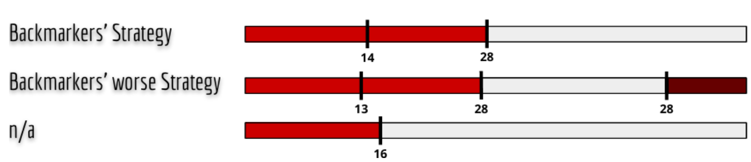
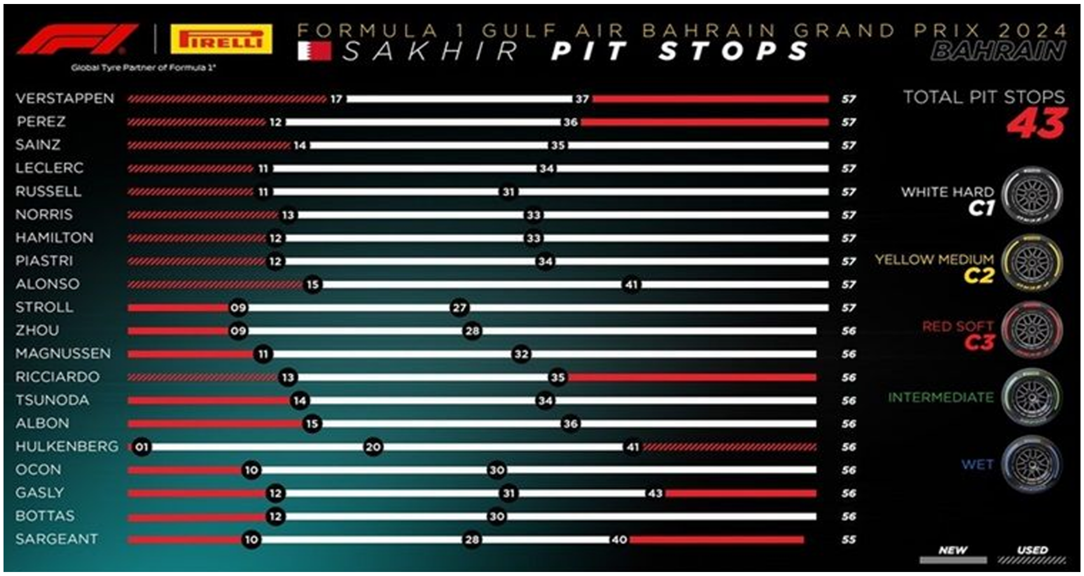


Figure 5.6. *Actual tyre compound usage for the Baharain 2024 Grand Prix [3].*




For the years 2023 and 2024, the hard tyre was more widely adopted, and that meant that longer stints were chosen and predicted. Longer tyre stints resulted being the weakness of our model, as teams opted for shorter stints with respect to what we predicted. We believe that this happens because data regarding hard tyres are usually utilised during longer stint sessions. This means that cars drive around the track for long period of times, burning plenty of gas, and therefore decreasing the weight of the car lap after lap. While the effect is not as visible for Soft an Medium compounds, as their tyre lives are significantly shorter, the same can't be said for Hard tyres. The lighter car might mitigate the loss in performance that the tyres have during long lap stints, and that results in a much more gentle degradation coefficient for the hard tyre. In order prevent that, more work on the data should be carried out to compensate for the time improvement due to gas consumption.

# CHAPTER 6

# CONCLUSION

Our findings demonstrate that a relatively simplified approach to tyre strategy modeling can achieve favorable results. This is particularly noteworthy considering that the majority of data science tools employed were primarily involved in data cleaning, and the linear regression model utilized is one of the most fundamental and readily interpreted machine learning algorithms. However, the model's consistency remains a concern, rendering it unreliable for real-world application, especially given the diverse track conditions and specifications encountered throughout a Formula 1 season. Further data pre-processing techniques are necessary to address issues arising from car weight fluctuations during a race. Additionally, incorporating a broader range of factors influencing tyre strategy, such as track temperature, the probability of safety car interventions, and rival team decisions, may necessitate the adoption of non-linear learning methods like Convolutional Neural Networks (CNNs). These algorithms typically exhibit greater flexibility, potentially leading to improved versatility across various race scenarios.

## Future Development

One major thing that could be used to develop our model in the future is implementing nonlinear functions to represent the tyre wear. With our assumptions that are discussed above we removed a few factors that include race disruptions, fuel-correction, game theory components, weather conditions, etc, which would have an effect on the total race time as well as how the tyres are degrading. With further development of the model we could use these factors as a way to more accurately depict what happens in the actual race. Lastly, we could apply machine learning prediction methods as well as consider genetic algorithms to optimize race-line given different tyre compounds.

REFERENCES


1. *Bahrain 2022*. Formula One World Championship Limited, https://www.formula1.com/en/racing/2022/Bahrain.html, 2022.

2. *Bahrain 2023*. Formula One World Championship Limited, https://www.formula1.com/en/racing/2023/Bahrain.html, 2023.

3. *Bahrain 2024*. Formula One World Championship Limited, https://www.formula1.com/en/racing/2024/Bahrain.html, 2024.

4. *fastf1 3.3.5*. Python Software Foundation, https://pypi.org/project/fastf1/, 2024.

5. *Pre-race strategy optimisation in formula 1*. Keberz Engineering, https://www.keberz.com/post/race-strategies-in-formula-1, 2023.

6. *Why Hamilton lost out to George in Miami and Tyre Degradation Analysis*. Tracing Insights, https://tracinginsights.substack.com/p/why-hamilton-lost-out-to-george-in, 2022.

7. C. Beaume, *Faster than Hamilton! Optimizing Formula 1 Strategies*. 2022.

8. C. Beaume, *Formula 1 Strategy Competition*. https://cbeaume.com/en/teaching$_f$1*.html*, 2024.

9. S. Pontin, *AI Based Race Strategy Assistant and Car Data Monitor*, Master's thesis, Luleå Tekniska Universitet, 2023.

10. S. N. Ramesh, F. Shiwa, and R. Singhal, *F1 Tire Prediction*. Cornell, https://people.ece.cornell.edu/land/courses/ece5760/FinalProjects/f2020/sn438 fs383 rs872/sn438 fs383 rs872/index.html, 2020.

# Appendix A
# Graphs and Related Plots

Bahrain 2022:

Figure 6.1. *Scatter plot of data before cleaning (2022).*

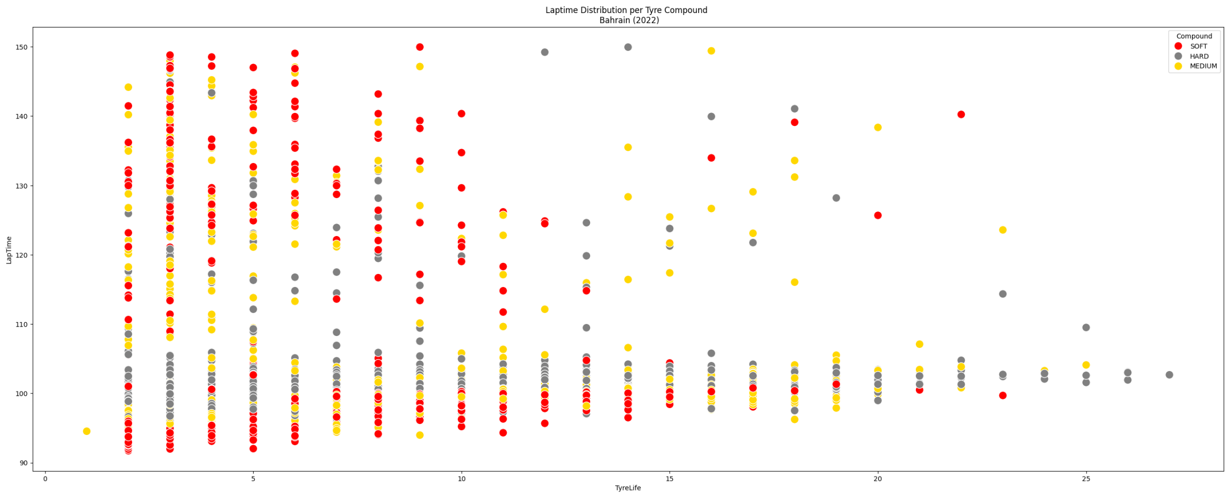



Figure 6.2. *Box-and-Whisker plot of data before cleaning (2022).*

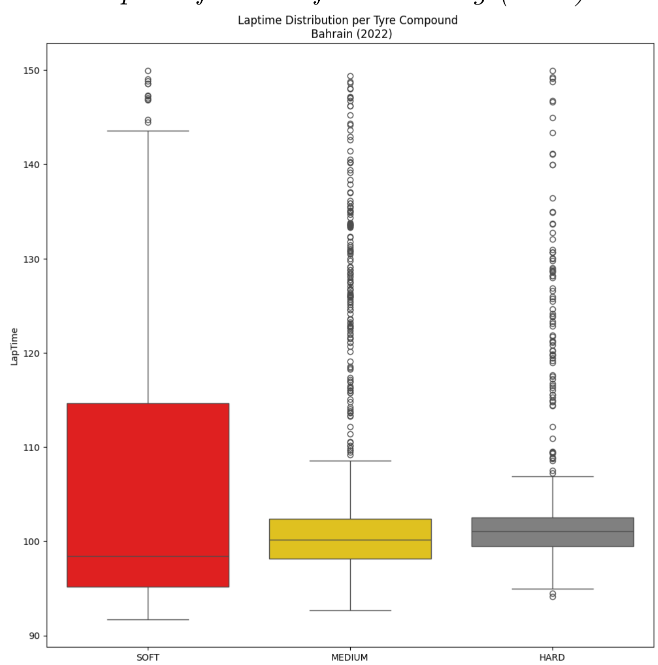


Figure 6.3. *Scatter plot of data after removing outliers (2022).*

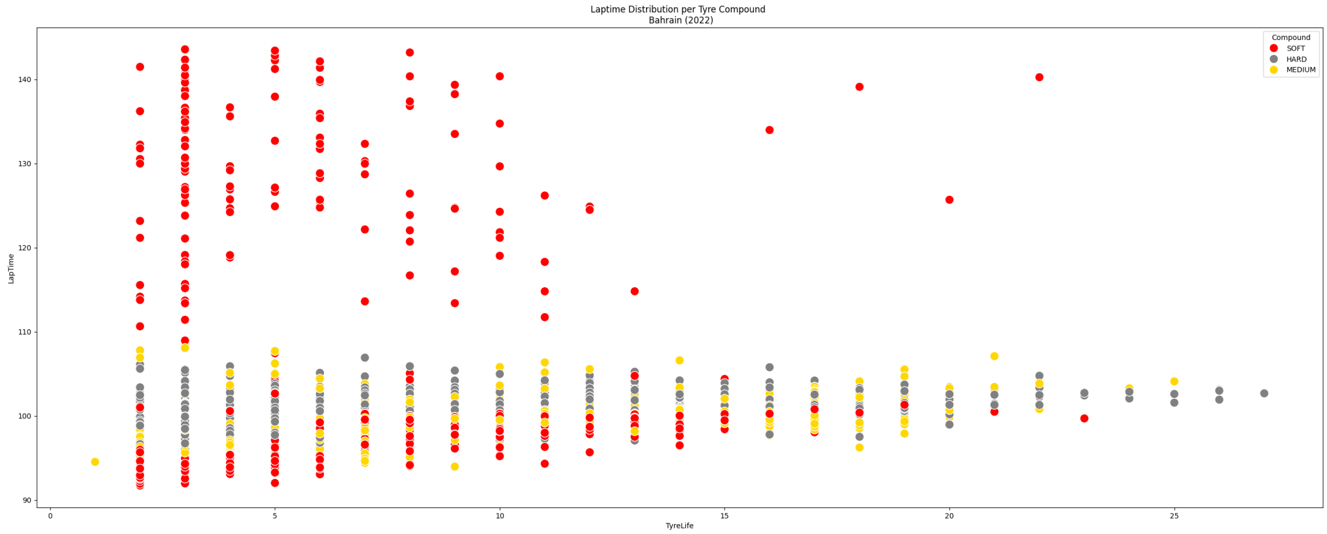
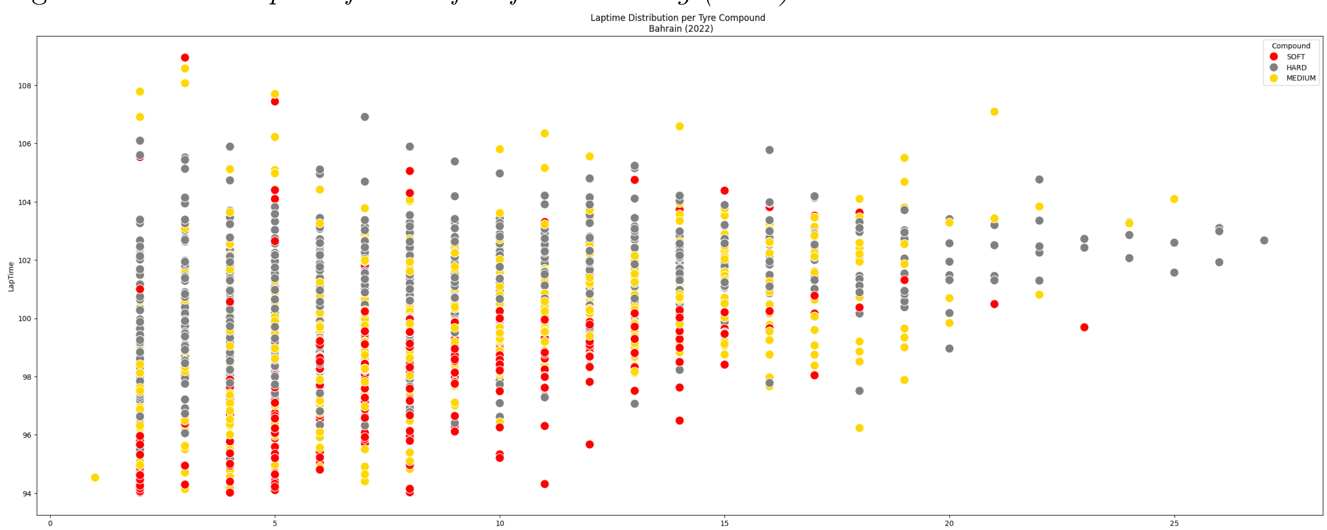



Figure 6.4. *Scatter plot of data after final cleaning (2022).*



Bahrain 2023:

Figure 6.5. *Scatter plot of data before cleaning (2023).*

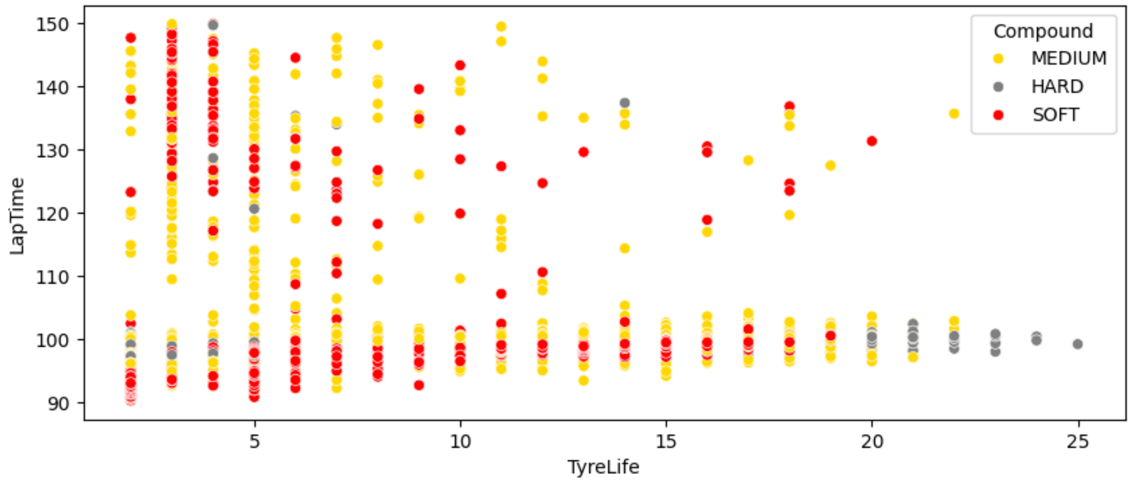



Figure 6.6. *Box-and-Whisker plot of data before cleaning (2023).*

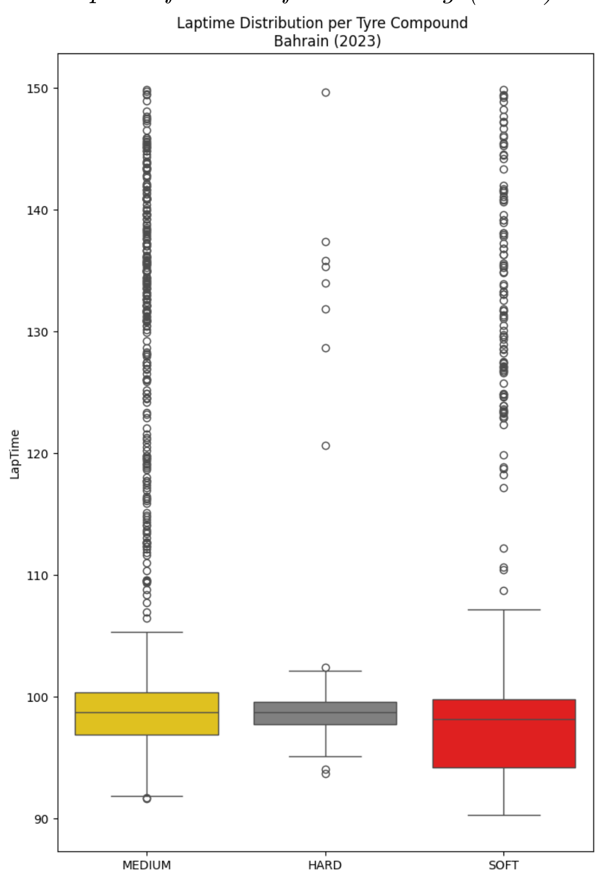
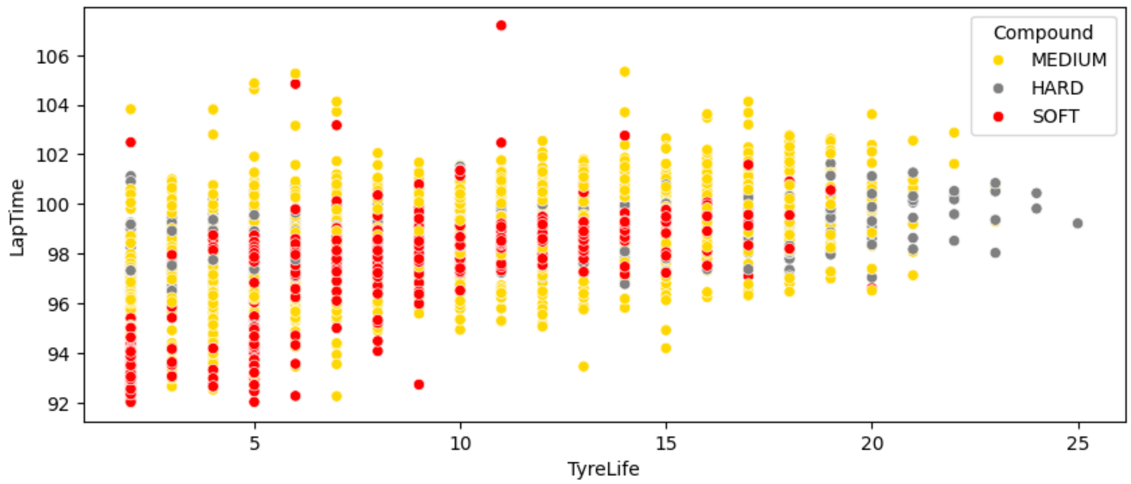


Figure 6.7. *Scatter plot of data after final cleaning (2023).*




Bahrain 2024:

Figure 6.8. *Scatter plot of data before cleaning (2024).*

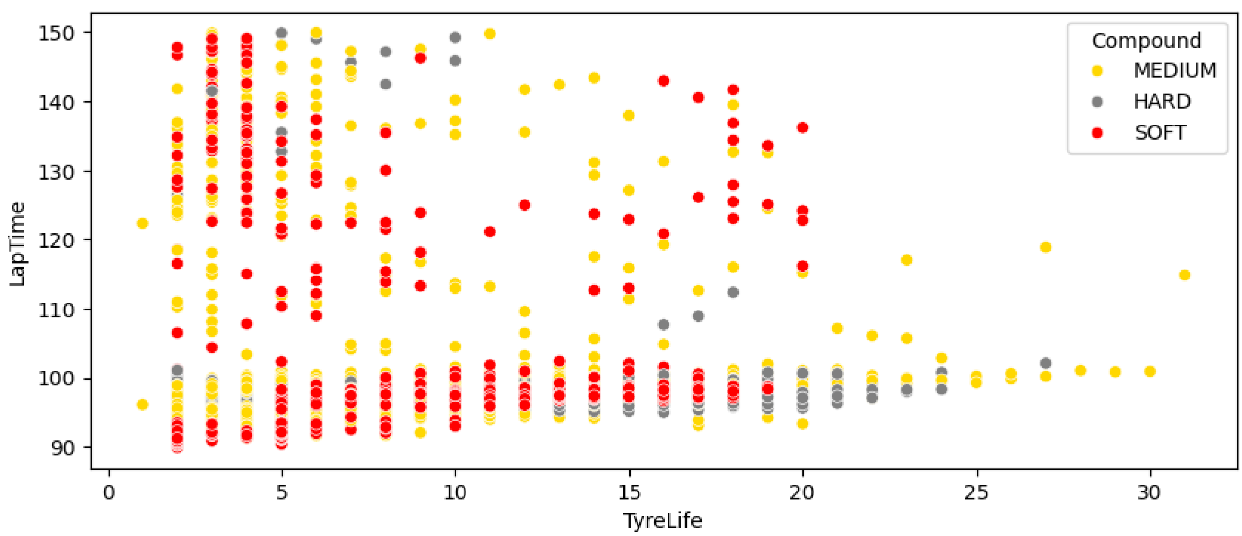


Figure 6.9. *Box-and-Whisker plot of data before cleaning (2024).*

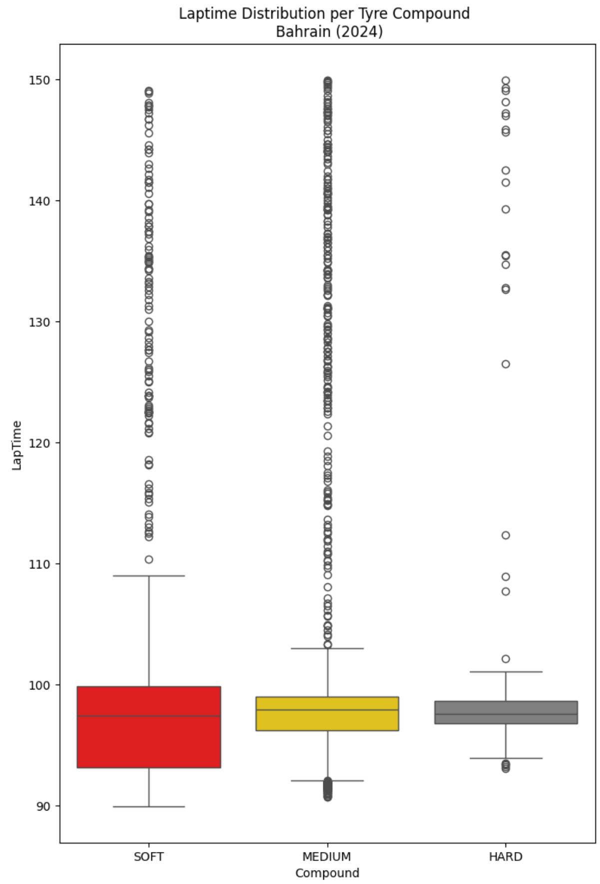
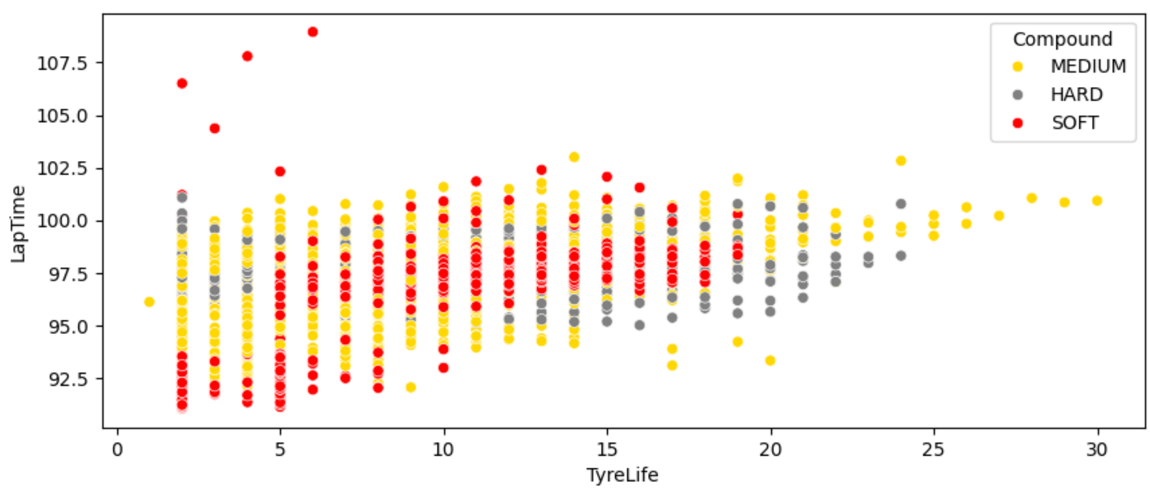



Figure 6.10. *Scatter plot of data after final cleaning (2024).*



Appendix B
Code

Listing 6.1. Python Code for K-Means Model on all F1 tracks for 2022 qualifications.

```
!pip install fastf1
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import fastf1 as f1
import scipy.stats as stats


q_bar = f1.get_session(2022, 'Bahrain', 'Q')
q_bar.load()

q_jed = f1.get_session(2022, 'Saudi', 'Q')
q_jed.load()

q_aust = f1.get_session(2022, 'Australia', 'Q')
q_aust.load()

q_imo = f1.get_session(2022, 'Imola', 'Q')
q_imo.load()

q_mia = f1.get_session(2022, 'Miami', 'Q')
q_mia.load()

q_spa = f1.get_session(2022, 'Spain', 'Q')
q_spa.load()

q_mon = f1.get_session(2022, 'Monaco', 'Q')
q_mon.load()

q_bak = f1.get_session(2022, 'Baku', 'Q')
q_bak.load()

q_can = f1.get_session(2022, 'Canada', 'Q')
q_can.load()

q_uk = f1.get_session(2022, 'Silverstone', 'Q')
q_uk.load()

q_aus = f1.get_session(2022, 'Austria', 'Q')
q_aus.load()

q_fra = f1.get_session(2022, 'France', 'Q')
q_fra.load()

q_hun = f1.get_session(2022, 'Hungary', 'Q')
q_hun.load()

q_bel = f1.get_session(2022, 'Belgium', 'Q')
q_bel.load()

q_ned = f1.get_session(2022, 'Netherlands', 'Q')
q_ned.load()

q_ita = f1.get_session(2022, 'Monza', 'Q')
q_ita.load()

q_sin = f1.get_session(2022, 'Singapore', 'Q')
q_sin.load()

q_jap = f1.get_session(2022, 'Japan', 'Q')
q_jap.load()
```

```python
q_usa = f1.get_session(2022, 'Austin', 'Q')
q_usa.load()

q_mex = f1.get_session(2022, 'Mexico', 'Q')
q_mex.load()

q_bra = f1.get_session(2022, 'Brazil', 'Q')
q_bra.load()

q_abu = f1.get_session(2022, 'Abu Dhabi', 'Q')
q_abu.load()


quali = [q_bar, q_jed, q_aust, q_imo, q_mia, q_spa, q_mon, q_bak, q_can, q_uk,
         q_aus, q_fra, q_hun, q_bel, q_ned, q_ita, q_sin, q_jap, q_usa, q_mex,
         q_bra, q_abu]


avg_speed = []
top_speed = []
throttle = []

for i in range(len(quali)):

  q = quali[i].laps.pick_fastest().get_telemetry()

  throttle.append((q[q.Throttle >= 95.0].Throttle.count()/q.Throttle.count()).round(2))
  avg_speed.append(q.Speed.mean().round(2))
  top_speed.append(q.Speed.max())


event = ['Bahrain', 'Saudi Arabia', 'Australia', 'Emilia Romagna', 'Miami', 'Spain', 'Monaco', 'Azerbaijan
    ', 'Canada', 'Britain',
         'Austria', 'France', 'Hungary', 'Belgium', 'Netherlands', 'Monza', 'Singapore', 'Japan', 'Texas',
              'Mexico',
         'Brazil', 'UAE']

nary = {'Event': event, 'Top Speed': top_speed, 'AVG Speed': avg_speed, 'Full Throttle %' : throttle}

df = pd.DataFrame(nary)
df = df[df.Event != 'Britain']
df


from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

wcss = []

for i in range(1, 11):
  kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
  kmeans.fit(df.iloc[:, 1:])
  wcss.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()


kmeans = KMeans(n_clusters=4, init='k-means++', max_iter=300, n_init=10, random_state=0)
y_pred = kmeans.fit_predict(df.iloc[:, 1:])

df['Cluster'] = kmeans.labels_


group1 = df[df.Cluster == 0]
group2 = df[df.Cluster == 1]
group3 = df[df.Cluster == 2]
```

```
group4 = df[df.Cluster == 3]

group1.mean(numeric_only=True).round(2)
group2.mean(numeric_only=True).round(2)
group3.mean(numeric_only=True).round(2)
group4.mean(numeric_only=True).round(2)
group2
```

Listing 6.2. Python Code for Finding Pre-Race Data in Bahrain. (*Used same code for all three years, changing the year and specific ranges of time to omit.*)

```
!pip install fastf1
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import fastf1 as f1
import scipy.stats as stats

##Loading Data
test1 = f1.get_testing_session(2022, 2, 1)
test1.load()

test2 = f1.get_testing_session(2022, 2, 2)
test2.load()

test3 = f1.get_testing_session(2022, 2, 3)
test3.load()

fp1 = f1.get_session(2022, 'Bahrain', 'FP1')
fp1.load()

fp2 = f1.get_session(2022, 'Bahrain', 'FP2')
fp2.load()

fp3 = f1.get_session(2022, 'Bahrain', 'FP3')
fp3.load()

##Building the DataFrame
df = pd.concat([test1.laps, test2.laps, test3.laps, fp1.laps, fp2.laps, fp3.laps])
df.LapTime = df.LapTime.dt.total_seconds()
df = df.reset_index()
df = df[df.IsAccurate == True]
df = df[df.TrackStatus == '1']
df = df[df['Compound'] != 'UNKNOWN']
df = df[df['Compound'] != 'TEST_UNKNOWN']

colors = {'SOFT': 'red', 'MEDIUM': 'gold', 'HARD': 'grey'}
plt.figure(figsize = (32,12))
plt.title('Laptime Distribution per Tyre Compound \n Bahrain (2022)')
sns.scatterplot(data = df, x = 'TyreLife', y = 'LapTime', s=150, hue = 'Compound', palette = colors)

plt.figure(figsize = (12,12))
plt.title('Laptime Distribution per Tyre Compound \n Bahrain (2022)')
sns.boxplot(data = df, x = 'Compound', y = 'LapTime', order = ['SOFT', 'MEDIUM', 'HARD'], palette = colors
    )

##Filtering Data
# SOFT
sq1 = df[df.Compound == 'SOFT'].LapTime.quantile(0.25)
sq3 = df[df.Compound == 'SOFT'].LapTime.quantile(0.75)
sIQR = sq3-sq1
s_ub = sq3 + 1.5*sIQR
s_lb = sq1 - 1.5*sIQR

# MEDIUM
mq1 = df[df.Compound == 'MEDIUM'].LapTime.quantile(0.25)
mq3 = df[df.Compound == 'MEDIUM'].LapTime.quantile(0.75)
mIQR = mq3-mq1
```

```
m_ub = mq3 + 1.5*mIQR
m_lb = mq1 - 1.5*mIQR

# HARD
hq1 = df[df.Compound == 'HARD'].LapTime.quantile(0.25)
hq3 = df[df.Compound == 'HARD'].LapTime.quantile(0.75)
hIQR = hq3-hq1
h_ub = hq3 + 1.5*hIQR
h_lb = hq1 - 1.5*hIQR

lb_cat = {'SOFT': s_lb, 'MEDIUM': m_lb, 'HARD': h_lb}
ub_cat = {'SOFT': s_ub, 'MEDIUM': m_ub, 'HARD': h_ub}

def filter(row):
  return (row['LapTime'] >= lb_cat[row['Compound']]) and (row['LapTime'] <= ub_cat[row['Compound']])

df = df[df.apply(filter, axis = 1)]

colors = {'SOFT': 'red', 'MEDIUM': 'gold', 'HARD': 'grey'}
plt.figure(figsize = (32,12))
plt.title('Laptime Distribution per Tyre Compound \n Bahrain (2022)')
sns.scatterplot(data = df, x = 'TyreLife', y = 'LapTime', s=150, hue = 'Compound', palette = colors)

df = df[df.LapTime < 110]
df = df[df.LapTime > 94]

colors = {'SOFT': 'red', 'MEDIUM': 'gold', 'HARD': 'grey'}
plt.figure(figsize = (32,12))
plt.title('Laptime Distribution per Tyre Compound \n Bahrain (2022)')
sns.scatterplot(data = df, x = 'TyreLife', y = 'LapTime', s=150, hue = 'Compound', palette = colors)

##Linear Regression
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

##Division of Data Based on Tyre Compound
df_s = df[df.Compound == 'SOFT'][['TyreLife', 'LapTime']]
df_m = df[df.Compound =='MEDIUM'][['TyreLife', 'LapTime']]
df_h = df[df.Compound == 'HARD'][['TyreLife', 'LapTime']]

xs_train, xs_test, ys_train, ys_test = train_test_split(df_s.TyreLife, df_s.LapTime, test_size = 0.2)
xm_train, xm_test, ym_train, ym_test = train_test_split(df_m.TyreLife, df_m.LapTime, test_size = 0.2)
xh_train, xh_test, yh_train, yh_test = train_test_split(df_h.TyreLife, df_h.LapTime, test_size = 0.2)

xs_train = xs_train.values.reshape(-1, 1)
xs_test = xs_test.values.reshape(-1, 1)
xm_train = xm_train.values.reshape(-1, 1)
xm_test = xm_test.values.reshape(-1, 1)
xh_train = xh_train.values.reshape(-1, 1)
xh_test = xh_test.values.reshape(-1, 1)

ys_train = ys_train.values.reshape(-1, 1)
ys_test = ys_test.values.reshape(-1, 1)
ym_train = ym_train.values.reshape(-1, 1)
ym_test = ym_test.values.reshape(-1, 1)
yh_train = yh_train.values.reshape(-1, 1)
yh_test = yh_test.values.reshape(-1, 1)


##LR for Soft Tyres
reg_s = linear_model.LinearRegression()
reg_s.fit(df_s[['TyreLife']], df_s['LapTime'])
pred_s = reg_s.predict(xs_test)

# print("Coefficients: \n", reg_s.coef_)
# print("Mean Squared Error: %.2f" % mean_squared_error(ys_test, pred_s))
# print("R-Squared: %.2f" % r2_score(ys_test, pred_s))
# print(f"Training score: {reg_s.score(xs_train, ys_train)*100}%")
# print(f"Testing score: {reg_s.score(ys_test, pred_s)*100}%")
```

```
print("Equation: LapTime = {:.2f} * TyreLife + {:.2f}".format(reg_s.coef_[0], reg_s.intercept_))

r_s, p_s = stats.pearsonr(xs_test.ravel(), ys_test.ravel())
r_sq_s = reg_s.score(xs_test, ys_test)

print("Pearson correlation coefficient (r):", r_s)
print("Pearson correlation coefficient squared (R^2):", r_sq_s)
print("P-value:", p_s)

plt.figure(figsize = (10,6))
sns.regplot(data = df_s, x = 'TyreLife', y = 'LapTime', scatter_kws={'color':'red'}, line_kws={'color':'
    blue'})
plt.xlabel('Tyre Life (laps)')
plt.ylabel('Lap Time (seconds)')
plt.title('Lap Time vs Tyre Life for SOFT Tyres \n Bahrain (2022)')

r_s, p_s = stats.pearsonr(xs_test.ravel(), ys_test.ravel())
r_sq_s = reg_s.score(xs_test, ys_test)

plt.text(0.79, 0.2, f'r = {r_s:.3f}', horizontalalignment='left', verticalalignment='center', transform=
    plt.gca().transAxes)
plt.text(0.79, 0.15, f'R^2 = {r_sq_s:.3f}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.text(0.79, 0.1, f'p-value = {p_s: .3e}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.show()

##LR for Medium Tyres
reg_m = linear_model.LinearRegression()
reg_m.fit(xm_train, ym_train)
pred_m = reg_m.predict(xm_test)

# print("Coefficients: \n", pred_m.coef_)
print("Mean Squared Error: %.2f" % mean_squared_error(ym_test, pred_m))
print("R-Squared: %.2f" % r2_score(ym_test, pred_m))
print(f"Training score: {reg_m.score(xm_train, ym_train)*100}%")
print(f"Testing score: {reg_m.score(ym_test, pred_m)*100}%")
print("Equation: LapTime = {:.2f} * TyreLife + {:.2f}".format(reg_m.coef_[0][0], reg_m.intercept_[0]))

r_m, p_m = stats.pearsonr(xm_test.ravel(), ym_test.ravel())
r_sq_m = reg_m.score(xm_test, ym_test)

print("Pearson correlation coefficient (r):", r_m)
print("Pearson correlation coefficient squated (R^2):", r_sq_m)
print("P-value:", p_m)

plt.figure(figsize = (10,6))
sns.regplot(data = df_m, x = 'TyreLife', y = 'LapTime', scatter_kws={'color':'gold'}, line_kws={'color':'
    blue'})
plt.xlabel('Tyre Life (laps)')
plt.ylabel('Lap Time (seconds)')
plt.title('Lap Time vs Tyre Life for MEDIUM Tyres \n Bahrain (2022)')

r_m, p_m = stats.pearsonr(xm_test.ravel(), ym_test.ravel())
r_sq_m = reg_m.score(xm_test, ym_test)

plt.text(0.79, 0.2, f'r = {r_m:.3f}', horizontalalignment='left', verticalalignment='center', transform=
    plt.gca().transAxes)
plt.text(0.79, 0.15, f'R^2 = {r_sq_m:.3f}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.text(0.79, 0.1, f'p-value = {p_m: .3e}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.show()

##LR for Hard Tyres
reg_h = linear_model.LinearRegression()
reg_h.fit(xh_train, yh_train)
pred_h = reg_h.predict(xh_test)

# print("Coefficients: \n", reg_h.coef_)
```

```
print("Mean Squared Error: %.2f" % mean_squared_error(yh_test, pred_h))
print("R-Squared: %.2f" % r2_score(yh_test, pred_h))
print(f"Training score: {reg_h.score(xh_train, yh_train)*100}%")
print(f"Testing score: {reg_h.score(yh_test, pred_h)*100}%")
print("Equation: LapTime = {:.2f} * TyreLife + {:.2f}".format(reg_h.coef_[0][0], reg_h.intercept_[0]))

r_h, p_h = stats.pearsonr(xh_test.ravel(), yh_test.ravel())
r_sq_h = reg_h.score(xh_test, yh_test)

print("Pearson correlation coefficient (r):", r_h)
print("Pearson correlation coefficient squated (R^2):", r_sq_h)
print("P-value:", p_h)

plt.figure(figsize = (10,6))
sns.regplot(data = df_h, x = 'TyreLife', y = 'LapTime', scatter_kws={'color':'grey'}, line_kws={'color':'
    blue'})
plt.xlabel('Tyre Life (laps)')
plt.ylabel('Lap Time (seconds)')
plt.title('Lap Time vs Tyre Life for HARD Tyres \n Bahrain (2022)')

r_h, p_h = stats.pearsonr(xh_test.ravel(), yh_test.ravel())
r_sq_h = reg_h.score(xh_test, yh_test)

plt.text(0.79, 0.2, f'r = {r_h:.3f}', horizontalalignment='left', verticalalignment='center', transform=
    plt.gca().transAxes)
plt.text(0.79, 0.15, f'R^2 = {r_sq_h:.3f}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.text(0.79, 0.1, f'p-value = {p_h: .3e}', horizontalalignment='left', verticalalignment='center',
    transform=plt.gca().transAxes)
plt.show()

##Mean PitStop Time Estimation
race = f1.get_session(2022, 'Bahrain', 'R')
race.load()

pits = race.laps[['PitInTime', 'PitOutTime']]
pits = pits.reset_index()
pits = pits.dropna(how = 'all')
pits['PitOutTime'] = pits['PitOutTime'].shift(-1)
pits = pits.dropna()
pits['PitTime'] = (pits.PitOutTime - pits.PitInTime).dt.total_seconds()

q1 = pits.PitTime.quantile(0.25)
q3 = pits.PitTime.quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - (1.5 * iqr)
upper_bound = q3 + (1.5 * iqr)

pits = pits[(pits.PitTime >= lower_bound) & (pits.PitTime <= upper_bound)]
print(pits.PitTime.mean())
```

Listing 6.3. LINGO Optimization (*Used same code for all three years, changing values and variables as needed.*)

```
!3 pit stops, 2022, 2 new soft, 1 new medium, 1 new hard;
Model:
Min = 25.15*3 + 95.09*Ls + (0.38/2)*Lso*Lso + (0.38/2)*Lst*Lst + 96.88*Lm + (0.30/2)*Lmo*Lmo + 99.17*Lh
    +(0.17/2)*Lho*Lho +
 (0.38)*4*Lsuo + (0.38/2)*Lsuo*Lsuo + (0.30)*3*Lmuo + (0.30/2)*Lmuo*Lmuo + (0.30)*3*Lmut + (0.30/2)*Lmut*
    Lmut +
 (0.17)*2*Lhuo + (0.17/2)*Lhuo*Lhuo + (0.17)*2*Lhut + (0.17/2)*Lhut*Lhut;
Lso + Lst + Lsuo = Ls;
Lmo + Lmuo + Lmut = Lm;
Lho + Lhuo + Lhut =  Lh;
Lso + Lst + Lsuo + Lmo + Lmuo + Lmut + Lho + Lhuo + Lhut = 57;
@ROUND(Lso/(Lso+0.1),0) + @ROUND(Lsuo/(Lsuo+0.1),0) + + @ROUND(Lst/(Lst+0.1),0) + @ROUND(Lmo/(Lmo+0.1),0)
 + @ROUND(Lmuo/(Lmuo+0.1),0) + @ROUND(Lmut/(Lmut+0.1),0)+ @ROUND(Lho/(Lho+0.1),0) + @ROUND(Lhuo/(Lhuo+0.1)
    ,0) + @ROUND(Lhut/(Lhut+0.1),0) = 4;
@GIN(Lso);
```

```
@GIN(Lsuo);
@GIN(Lst);
@GIN(Lmo);
@GIN(Lmuo);
@GIN(Lmut);
@GIN(Lho);
@GIN(Lhuo);
@GIN(Lhut);
```