Ca' Foscari
University
of Venice

Master's Degree in
Data Analytics for Business and Society

Final Thesis

The Importance of Strategy in Formula One Race Outcomes and the Impact of Regulatory Changes in the Competitive Balance of the Sport

Supervisor
Ch. Prof. Raffaele Pesenti

Graduand
Valentina Spolaor
Matriculation Number 863296

Academic Year
2021/2022

Abstract

Formula One is the highest and most competitive class of international motorsport racing sanctioned by the Fédération Internationale de l'Automobile. Its complexity and uncertainty is what brings athletes and spectators in its orbit. The championship's outcome of each team is heavily impacted by the driver-car abilities and the strategies adopted during qualifying and races. Hence, the purpose of this thesis is that of analysing the race-weekend sessions and their implication in the final race result.

Firstly, an analysis of qualifying session in terms of prediction and strategies formulations are proposed. Secondly, a race simulation was performed to formulate optimal strategies regarding the timing of pit-stops and the choice of tyre compound, comparing the results with the real-race outcome. Thirdly, an analysis on the competitive balance of the last 12 years is performed in order to understand the impact of the latest financial restriction regulation in the latest seasons (2021 and 2022).

Contents

List of Figures 2
List of Tables 3
List of Equations 4
Introduction 5
Chapter 1 8
1.1 Theoretical Background 11
1.1.1 Optimization problem 11
1.1.2 Monte Carlo Simulation 16
1.1.3 Regression and Classification Model Theory 17
1.2 Data Description and Preparation 22
1.3 Data Modelling and Results 28
Chapter 2 41
2.1 Theoretical Background 46
2.2 Data Description and Preparation 53
2.3 Results 56
Chapter 3 59
3.1 Theoretical Background 61
3.2 Data preparation and Results 63
Conclusions 67
Bibliography and Sitography 70

List of Figures

Figure 1 Conceptual Map of Models and Data needed for Qualifying Strategy simulations 9
Figure 2 Correlation matrix of independent variables used for prediction modelling 27

List of Tables

Table 1 – Results of Tyre Allocation Optimization Problem…………………………….. 29
Table2 – Monte Carlo Simulation Probabilities of Yellow and Red Flags in Qualifying Sessions …………………………………………………………………………….. 29
Table 3 – Monte Carlo Simulation Results on Probabilities of Drivers to Have Caused the Yellow or Red Flag…………………………………………………………….. 30
Table 4 – Bahrain Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 34
Table 5 – Saudi Arabian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………. 35
Table 6– Australian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………. 35
Table 7 – Miami Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 36
Table 8 – Monaco Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 36
Table 9 – Azerbaijan Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………. 37
Table 10 – French Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 38
Table 11 – Belgian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………. 38
Table 12 – Dutch Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 39
Table 13 – Italian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies ……………………………………………………………………………… 39
Table 14 – Prediction Of Cut-Off Timing Data And Pole Time …………………………… 40
Table 15 – Point Assignment During Formula One Race ………………………………….. 41
Table 16 – Average Starting Performance in Milliseconds, period 2019-2022 …….. 53
Table 17 – Probabilities of Teams for the Occurrence of Failure and the Fisk Parameters for the Pit-Stop Time Variability, period 2019-2022 ……………………….. 54
Table 18 – Probabilities of Drivers to Make an Accidents During the Race, period 2019-2022 ……………………………………………………………………………………… 54
Table 19 – Results of Strategy Simulation for 2022 Dry Races …………………………… 57
Table 20 – HHI and SD Measures Considering Drivers' and Teams' Championships…………………………………………………………………………………. 66

List of Equations

Equation 1 – Minimization Problem 12
Equation 2 – Optimal Solution for Minimization Problem 12
Equation 3 – Maximization Problem 13
Equation 4 – Optimal Solution for Maximization Problem 13
Equation 5 – Objective Function Formulation for Tyre Allocation Problem 13,15
Equation 6 – Binomial Distribution 17
Equation 7 – Expected Value of Binomial Distribution 17
Equation 8 – Variance of Binomial Distribution 17
Equation 9 – Underlying Mathematical Model of CART 19
Equation 10 – Objective Function to Be Optimized in CART 19
Equation 11 – Objective function with training loss and regularization 19
Equation 12 – Prediction value at step t 19
Equation 13 – Taylor expansion of the loss function up to second order 20
Equation 14 – Specific Objective Function at Step t 20
Equation 15 – Definition of Complexity within the Tree 20
Equation 16 – Definition of Complexity in Extreme Gradient Boosting 21
Equation 17 – Goodness of the Tree Structure 21
Equation 18 – Mean Squared Error 21
Equation 19 – Gini Index 21
Equation 20 – Sigmoid Function 22
Equation 21 – Lap Time Calculation 46
Equation 22 – Race Time Calculation 47
Equation 23 – Starting Time Performance 48
Equation 24 – Quadratic Polynomial Distribution of Lap Time 49
Equation 25 – Parameter of Beta Distribution Calculation ($\hat{\alpha}$) 49
Equation 26 – Parameter of Beta Distribution Calculation ($\hat{\beta}$) 49
Equation 27 – Objective Function of Optimum Stint Lengths 52
Equation 28 – Objective Function of Optimum Stint Lengths with Compounds and Stint Duration 52
Equation 29 – Linear Tyre Degradation Model 52
Equation 30 – Objective Function of Optimum Stint Lengths with Tyre Starting age 53
Equation 31 – Driver Starting Time Performance 53
Equation 32 – Standard Deviation of Winning Percentages 64
Equation 33 – Competitive Balance Calculation Based on Points Scored by each Driver 64
Equation 34 – Herfindahl-Hirschman Index (HHI) 65
Equation 35 – Herfindahl-Hirschman Index (HHI) Specification for Drivers 65

Introduction

Formula One is the highest and most competitive class of international motorsport racing for open-wheel single-seater formula racing cars that is sanctioned by the Fédération Internationale de l'Automobile (FIA). Thus making the FIA Formula One World Championship the premier class of four-wheels motorsport competition since 1950, year in which the first grand prix took place. [1]

Since the beginning, the Formula One World Championship has been synonym of finding the limits of both the driver and the cars and thus evolving till reaching the apex, both figuratively and practically speaking; as a matter of fact the apex is the term used to identify the part of the corner, of the race track, in which the driver is closest to the kerb, when it is reached correctly, it allows the driver to exit the corner as fast as the car allows him to, thus entering in full speed in the next part of the track.

The Formula One World Championship is yearlong competition, comprised of the following elements[1] :

- Ten teams and twenty drivers which compete among themselves, to win respectively the World Constructors' Championship and the World Drivers' Championship
- Sporting commission, race commission and race directors which are responsible for the control of every aspect of the weekend on-and-off track events.
- Twenty-two races, known as "Grand Prix" which take place in circuits that can be both built-for-the-purpose circuits and also city circuits (raced in closed-off city's streets).
- Each Grand Prix can have two formats, divided in the following sessions:
    - "Traditional" Format:
        1. Three Free Practices (denominated FP1, FP2, FP3) in which driers will familiarize with the circuits' conditions and will work

[1] For the purposes of the analysis which will follow in later chapters, the current 2022 season will be taken into consideration for supplying information about the championship organization and the agents involved.

together with the teams of the set-up pf the car and prepare qualifying and race strategies for the upcoming days.

2. A qualifying session divided in three parts (Q1, Q2, Q3) which will determine the starting grid for the race.
3. The race which determines the winners and awards championship points to the drivers which were successful in claiming the first ten positions and their teams.

- "Sprint" Format: which varies from the above by replacing FP2 with qualifying session and adds a sprint race that determines the starting grid for the race. [2]

For the purposes of the analysis that will be presented, only the traditional format has been taken into consideration.

Therefore, the goal of each team and driver are those of optimizing the Grand Prix results in order to claim the victory of the Championships. As a highly competitive motorsport class, a pivotal contribution to the final outcome of the season is given by the strategies adopted at each race. Each team and its drivers aim at finishing the race in the least time possible, in order to do so, they have to formulate strategies regarding which tyre compounds to be used during the race and when to perform the pit-stop.

As per regulatory demands, a pit-stop has to be performed by each driver in every race to which they take part. The pit-stop is the moment of the race in which the driver enters the pit lane (a special part of the circuit, which is directly connected to the main race track, in which the speed is limited and the drivers can stop in front of their garage) allowing their team-members to perform the tyre-change and eventually substituting any damaged part of the car and performing quick adjustment to the aerodynamic balancing of the car. Entering the pitlane implies a time-loss with respect to the race time, on average, of 20 seconds (depending on how long this portion of the circuit is) and could implicate a loss of positions during the race.

Hence, both the timing of pitstop during the race and the tyre compound sequences are essential parts in determining the difference between winning or losing a race. Through the various sessions that are available to drivers and teams, a multitude of data are collected regarding the best timing performance for each different tyre compounds supplied during the weekend and their wear which are determinant to formulate strategies. Data from the past races are also used in order to formulate solutions to potential issues regarding the race trend.

The purpose of this study is to delve deeper into the importance of the free practices data to predict and simulate a most likely scenario for qualifying session to have a better understanding at how the final result of this session is determined through the various free practices sittings. Moreover, as the most important part of the racing weekend is the race, as it's the only moment in which points are awarded, an analysis of various aspects impacting the race results will be performed, resulting in the formulation of race strategies that are viable to reach the desired outcome.

Furthermore, as Formula One represents one of the biggest sport events in the world, an analysis on the competitive balance in the last twelve years will be performed. Moreover, the impact of changing regulations will be observed, with more relevance given to the impact of the introduction of the budget cap in years 2021 and 2022 and its repercussion in the championships' results.

Chapter 1

This chapter will cover the theoretical background of the optimization and prediction models used to determine qualifying outcomes and their application to the dataset used for the analyses.

As free practices ranging from Friday to Saturday are essential to determine the setting and therefore the pace of both the qualifying and race sessions, different aspects have to be taken into consideration when determining the strategies for race but especially for qualifying sessions.

The qualifying session is of extreme importance for determining the race outcome. Indeed, starting the closest to the first position allows drivers to have clean air, an aspect this, that is fundamental for a better starting performance. By starting the in the grid's top positions, drivers will have less confusion surrounding them due to other cars that wants to overcome one another (especially in the midfield the battles at the race start could cause incidents and damages to cars that would, in turn, hinder the race outcome) and a better possibility at going through the first corner at the very least in the same position of the starting grid or even better. Of course, the starting performance depends heavily on both car qualities but especially on the driver’s readiness and penchant for a good start.

In order to give a complete view of the information that are needed by the team to formulate a clear path of strategies towards the best qualifying outcome possible, the conceptual map in *Figure 1* is proposed.

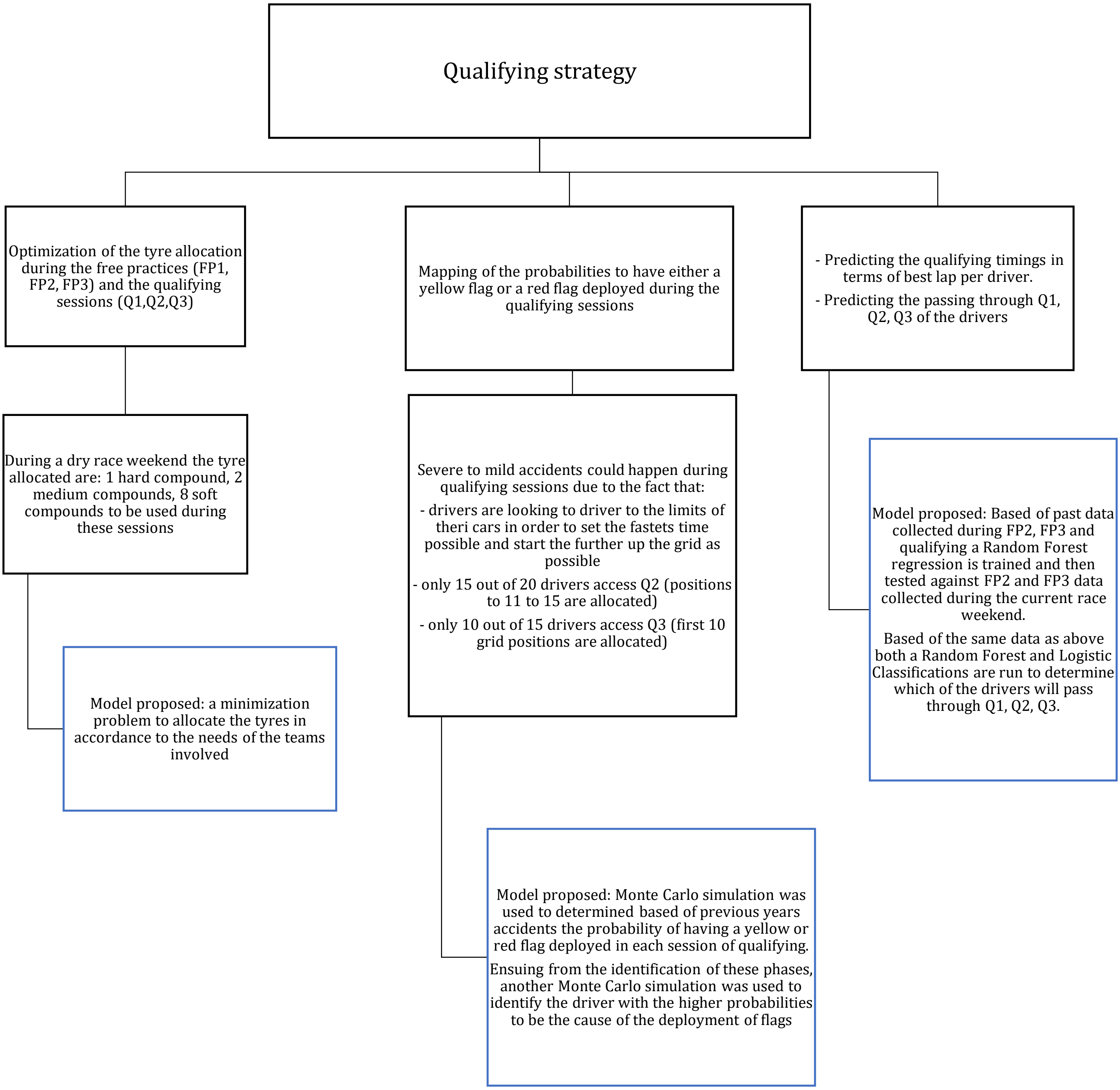

Figure 1 *Conceptual Map of Models and Data needed for Qualifying Strategy' simulations*

All these aspects are required to formulate the best "exit strategies" in all sessions of qualifying. With exit strategies meaning whether the teams will go through a one-exit strategy which is composed as:

- An out-lap: which is the lap performed when the driver exits the pit-lane, this lap is purposefully slow in order to allow the tyres to enter the correct temperature frame. Usually it's the lap that precedes the flying lap.
- A flying lap: it is the lap performed after the tyres have been heated up during the out-lap, this is the fastest lap that will be accounted for in terms of qualifying standings
- A Cooldown lap: is the lap performed by the driver after the flying lap, it allows the driver to cool the tyres after the extreme pressure and heat that they were exposed during the flying lap, in order to use them for another eventual flying lap. If the tyres are brought to the correct temperature frame the tyres could be good to make another flying lap at least as fast as the previous. This lap could be considered as a preparation lap that is performed to prepare in the best way possible the tyres for another flying lap.

These three laps are performed twice on only one tyre compound. This strategy is usually preferred when there is little to none tyre degradation allowing the tyre to be on the optimal window for more than a lap and, above all, when a yellow flag has a high probability to be deployed. This strategy allows the driver to perfect the lap when constrained by having only a fresh tyre in session.

Another exit strategy is a two-exit strategy which is composed of:

- An out-lap
- A flying lap
- A cooldown lap which is also the lap in which the driver enters the pitlane to change the tyres to a fresh set to perform another combination of the previous laps.

This strategy is usually deployed when there are two set available to perform the best lap so to have a fresh compound for each attempt. The timing of exiting of these two strategies is of vital importance. As a matter of fact, given the time constraint that varies accordingly to the sessions, 18 minutes are given for Q1, 15 minutes for Q2 and 12 minutes for Q3, [2] and taking into consideration the fact that the race track improves by each passing session, teams have to set the latest lap starting time in order to be able to perform at least two flying laps.

The latest starting time to perform either of these two combinations is calculated based on data collected during FP2 and FP3 and past races, which are the best indicators of the qualifying pace of each driver.

In the next sections the models will be proposed together with the data employed and the resulting outcome of their usage on these data.

1.1 Theoretical Background

1.1.1 Optimization model

The first aspect to be taken into consideration during the race weekend is the tyre compound issue. The term compound refers to the composition of the rubber of a specific tyre, the softer the compound the more grip it will provide thus allowing the driver to race faster at the expense of its durability. Hence, there being a trade-off between the velocity and the durability of the different tyre compounds. [3]

Every race weekend, the teams will be provided with 13 different sets of tyres, divided as follows:

- eight sets of soft compounds, this is the compound mostly used for qualifying as it offers the best performance in terms of minimizing the lap time;
- three sets of medium compounds, this is the compound with the best trade-off between performance and durability. It allows the driver to have good performance, which will be worse than the soft compound but better than the hard compound, and also a good durability, significantly higher than those of softer compound;

- two sets of hard compounds, these compound are best used during the race due to their high reliability and durability, while still allowing a good pace to the driver. [2]

Due to sporting regulations run by the FIA, of the abovementioned set of tyres, all teams are required to use one set of each medium and hard compound exclusively for the race, thus reducing to 1 hard compound and 2 medium compound the number of tyres available during the pre-race sessions. Moreover, it is mandatory that a set out of eight of soft compound, has to be used exclusively during the qualifying session Q3.[2]

Thus, a significant part of the weekend strategy relies upon the correct allocation of tyres during all different sessions. It is mandatory to use at least 2 set of tyres during each free practice session, these tyres are only going to be used during their respective session and will not be available for further use after the session has ended.[2]

Therefore, the tyre allocation issue requires an optimization problem to be solved in order to distribute in an optimal way the compounds available, as this will be a determinant for the race strategies and the car performance later during the race.

The definition of an optimization problem is specified by:

- A set E which elements are called solutions/decisions;
- A subset $F \subset E$, feasible set which elements are feasible solutions; whilst elements in $E \backslash F$ are named unfeasible/infeasible sets. The relationship $x \in F$ is called constraint.
- An objective function $f : E \rightarrow \mathbb{R}$ to be minimized or maximized depending on purpose of the problem.

In a minimization problem, the optimal value is $v = f(x^*)$ such that of Equation 1.

$$v = \min_{x \in F} f(x) \tag{1}$$

The optimal solution is given by Equation2, when each element $x^* \in F$ such that:

$$f(x^*) \leq f(y) \quad \forall y \in F \tag{2}$$

In a maximization problem, the optimal value is $v = f(x^*)$ such that of Equation 3.

$$v = \max_{x \in F} f(x) \tag{3}$$

The optimal solution is supplied by Equation 4, when each element $x^* \in F$ such that

$$f(x^*) \geq f(y) \quad \forall y \in F \tag{4}$$

The optimization problems can be continuous (either constrained if $F \subset R^n$ or unconstrained if $F = R^n$), discrete (either integer programming if $F \subseteq Z^n$ or binary programming if $F \subseteq \{0, 1\}^n$) or mixed.

For the purpose of providing a solution for the minimization problem at hand, the following system is provided in the case in which all sessions were declared dry:

- Sessions as a set $S = \{ FP1, FP2, FP3, Q1, Q2, Q3\}$
- Tyres as a set $T = \{H{:}\ 1, M{:}\ 2, S{:}\ 7\}$

Where, for each session two set of tyres have to be used, and all compounds have to be used before qualifying at least one time, as it is imperative for every team to have an analytical understanding of how the car performance behave with each different compound type.

Therefore, the decision regards the number and the compound of tyres to be used throughout the sessions.

The objective function is formulated as Equation 5.

$$\min \sum_{t \in T} \sum_{s \in S} w x_{ts} \tag{5}$$

It adds w if tyre T is used in session S, 0 otherwise. The w parameter is inserted due to the fact that, there is a preference of the teams to employ the hard compound during the FP2 session to perform a long run to measure the tyre wear and degradation that could occur during the race and a soft compound to be used to set flying laps in anticipation of the qualifying session that will take place the following day at the same hour. Whilst medium compounds are preferred for either FP1 or FP3 to test the car setting to be optimal.

The constraints, to which Equation 5 is subjected, are as follows, when taking into consideration top teams (in this analysis a top team in considered a team which consistently entered the Q3 session in the last 10 race weekends).

$$\sum_{t \in T} x_{ts} = 2 \quad \forall s \in S = \{\mathrm{FP1}, \mathrm{FP2}, \mathrm{FP3}, \mathrm{Q2}, \mathrm{Q3}\}$$

$$\sum_{t \in T} x_{ts} = 1 \quad \forall s \in S = \{\mathrm{Q1}\}$$

$$\text{Tyres} \quad \begin{cases} 1 \leq \sum_{t \in T} x_{ts} \leq 7 \;\; if \;\; t = S \\ 1 \leq \sum_{t \in T} x_{ts} \leq 2 \;\; if \;\; t = M \\ \sum_{t \in T} x_{ts} = 1 \;\; if \;\; t = H \end{cases}$$

The constraints, to which Equation 5 in subjected, are those presented below, for midfield teams (these are teams which didn't always finish in the top ten in the last 10 races).

$$\sum_{t \in T} x_{ts} = 2 \quad \forall s \in S = \{FP1, FP2, FP3, Q1, Q2\}$$

$$\sum_{t \in T} x_{ts} = 1 \quad \forall s \in S = \{Q3\}$$

$$\text{Tyres} \quad \begin{cases} 1 \leq \sum_{t \in T} x_{ts} \leq 7 \;\; if \;\; t = S \\ 1 \leq \sum_{t \in T} x_{ts} \leq 2 \;\; if \;\; t = M \\ \sum_{t \in T} x_{ts} = 1 \;\; if \;\; t = H \end{cases}$$

Thus solving this optimization problem, based on the team consistency allows us to have the best obtainable tyre allocation during the sessions previous to race.

This model could be generalized for wet sessions ,depending on which of these will be declared a wet session. In this case, for every session in which an intermediate tyre is used (intermediate tyre is a compound which is used exclusively if a session is declared wet, when the rainfall is moderate, otherwise, in case of heavy rainfall, wet tyre will be used) another intermediate tyre is given to each team to be used[2]. Thus the system and constraints are modified as follows.

For wet FP1 and FP2:

- Sessions as a set *S = { FP1, FP2, FP3, Q1, Q2, Q3}*
- Tyres as a set *T = {H: 1, M: 2, S: 7, I: 4}*

The minimization problem is that described by Equation 5.

$$min \sum_{t \in T} \sum_{s \in S} w x_{ts} \qquad (5)$$

In this case the weight is associated to the weather to which teams are subjected to, giving lower weight to intermediate tyres during the wet sessions with respect to the dry tyres (dry tyres can be used even in wet sessions at drivers and teams' own risk).

The constraints of Equation 5 are presented below.

$$\sum\nolimits_{\mathrm{t \in T}} \mathrm{x_{ts}} = 1 \quad \forall s \in S = \{FP1, FP2\} \quad \forall t \in T = \{I\}$$

$$\sum\nolimits_{\mathrm{t \in T}} \mathrm{x_{ts}} = 2 \quad \forall s \in S = \{FP3, Q1, Q2, Q3\} \quad \forall t \in T = \{H, M, S\}$$

$$\text{Tyres} \begin{cases} 1 \le \sum_{t \in T} x_{ts} \le 7 \;\; if \;\; t = S \\ 1 \le \sum_{t \in T} x_{ts} \le 2 \;\; if \;\; t = M \\ \sum_{t \in T} x_{ts} = 1 \;\; if \;\; t = H \\ 1 \le \sum_{t \in T} x_{ts} \le 4 \;\; if \;\; t = I \end{cases}$$

Instead, if the wet sessions are FP3 and Qualifying the system and constraints, for Equation 5, will be the following.

$$\sum_{t \in T} x_{ts} = 2 \quad \forall s \in S = \{FP3, Q1, Q2, Q3\} \ \forall t \in T = \{I\}$$
$$\sum_{t \in T} x_{ts} = 2 \quad \forall s \in S = \{FP1, FP2\} \ \forall t \in = \{H, M, S\}$$

Thus solving the model (5) depending on the weather of the session, yields the best allocation for the teams involved, indeed allowing them to formulate the best strategies.

1.1.2 Monte Carlo Simulation

An important aspect to be taken further into consideration during the qualifying session is the impact of yellow or red flags exhibit. As a matter of fact, drivers when trying to set the fastest lap, incur through various risks which can be as much rewarding as they can create great risk of making moderate to grave mistakes. It is deemed to be important to identify those circuits in which there is the higher probability of having an accident that could nullify the potential of the car to make the best time possible.

Yellow flags are exhibited when drivers make a moderate mistake ranging from exiting the track lines going in escape tracks (parts outside of the track allowing drivers to exit the track and avoid serious consequences) to hitting the barriers (safety barriers that are found all around race track to prevent the cars to exit it due to severe accidents) and then being able to enter the pits.

Whilst yellow flags can be exhibited and then taken out in mere seconds up to a minute, red flags exposure signal severe accidents involving drivers hitting the barriers and most likely destroying a part of the car. This prevents the drivers to re-enter the pits, thus abandoning the car inside the track, making it a danger for other drivers. When the red flag is exposed the session immediately stops, it is then resumed later on, when the track results to be clear of any debris.

In order to forecast the possibility of having a yellow or red flag throughout the three qualifying sessions, Q1, Q2, Q3, a Monte Carlo simulation (MCS) based on past data was performed to identify in which session of which race it will be more likely for a flag to be exposed. This is an important aspect to be accounted for in qualifying strategy because, if the track is identified to have a potential of this exposure, the drivers will try to put the fastest lap in the first outing preventing eventual problems nullifying their being able to perform another flying lap to set the fastest time possible.

MCS are methods in which statistical sampling is employed to approximate solutions to quantitative problems. In MCS, rather than estimating random quantities in a deterministic manner, random quantities are employed to provide estimates of deterministic quantities to make probabilistic assessments on the range of possible solution values.

The binomial distribution, Equation 6, was chosen to fit the data.

$$Y \sim Bin(k, p) \tag{6}$$

Where k are the trials and the parameter $p \epsilon (0,1)$ is the probability of success at each trial. The expected value and variance are calculated, respectively, in Equation 7 and Equation 8.

$$E[Y] = kp \tag{7}$$

$$[Y] = kp(1-p) \tag{8}$$

1.1.3 Regression and Classification Model Theory

In regard of the prediction of the qualifying timings, different regression models were employed while taking into consideration the past data gathered from 2019 to 2021.

Moreover, to deepen the understanding of how the qualifying grid is determined, classification models where used to predict which of the team-driver pairing was going to make through each session of qualifying.

Therefore, the models employed for predicting the qualifying best lap time of each driver are the machine learning (ML) techniques of the Extreme Gradient Boosting algorithm and the Random Forest algorithm.

Machine learning is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values[4]. It refers to a system's ability to acquire and integrate knowledge through large-scale observations and to improve and extend itself by learning new knowledge rather than by being programmed with that knowledge (Shapiro, 1992). These techniques organize existing knowledge and acquire new knowledge by intelligently recording and reasoning about data.[5]

There are four basic approaches to ML:

- supervised learning: algorithms are supplied with labelled training data and define the variables they want the algorithm to assess for correlations. Both input and output variables are defined.
- unsupervised learning: algorithm that train on unlabelled data, scanning through the dataset looking to significant connections.
- semi-supervised learning: this approach involves the mix of the techniques aforementioned. the algorithm could be fed with labelled data but being free to explore the data on its own to develop connections.
- reinforcement learning: typically used to teach a machine to complete a multi-step process for which there are clearly defined rules. An algorithm is programmed to complete a task and give it positive or negative cues as it works out how to complete a task.[5]

The onward focus is put on either gradient boosting or random forest algorithms. Indeed, gradient boosting is a machine learning technique which is used for both classification and regression problems and produces a prediction model that is composed of an ensemble of weak prediction models, which, combined, enhance the performance of the technique as a whole. This algorithm is used to find any nonlinear relationships between the target and features used for the model, it is especially useful when dealing with missing values, outliers and high cardinality categorical features[5]. It involves three elements: a loss function to be optimized; a weak learner to make predictions and an additive model to add weak learners to minimize the loss function.

The algorithm that was adopted for regression modelling is the Extreme Gradient Boosting which is a decision three ensemble that consists of classifications and regression trees (CART). In CART, a real score is associated with each of the leaves, which yields richer interpretations that go beyond classification. The underlying mathematical model in the form of Equation 9. [6]

$$\hat{y}_i = \sum_{k=1}^{K} f_k\,(x_i), \quad f_k \in \mathcal{F} \tag{9}$$

where K is the number of trees, f_k is a function in the functional space F, and F is the set of all possible CARTs. The objective function to be optimized is given by Equation 10.

$$obj(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \omega(f_k) \tag{10}$$

Where $\omega(f_k)$ is the complexity of the tree f_k. Indeed, Tree boosting and random forests are quite similar, the difference rises when in training the models.

The objective function which always contains both the training loss and regularization, is formulated in Equation 11.

$$obj = \sum_{i}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \omega(f_i) \tag{11}$$

In order to learn the model (11), additive strategy is used where it is fixed while learning adding one tree at a time. The prediction value at step t as $\hat{y}_i l t_1$, in presented in Equation 12.

$$\hat{y}_i^{(t)} = \sum_{i=1}^{n} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \qquad (12)$$

In order to determine which trees will be selected the Taylor expansion of the loss function up to second order is used, following Equation 13.

$$obj^{(t)} = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) + \omega(f_t) + \text{constant} \qquad (13)$$

Where:

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$$

By removing the constant, the specific objective (12) at step t (which is the optimization goal of the following tree) results in Equation 14.

$$\sum_{i=1}^{n} [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \omega(f_t) \qquad (14)$$

As the objective function depends exclusively on g_i and h_i, every loss function can be optimized using the solver taking g_i and h_i as inputs.

Regarding the regularization term, the complexity of the tree $\omega(f_k)$ has to be defined, the definition of the tree being *f(x)* presented in Equation 15.

$$f_t(x) = w_{q(x)}, w \in R^T, q: R^d \rightarrow \{1,2, \dots, T\} \qquad (15)$$

Where w is the vector of score on leaves, q is a function assigning each data point to the corresponding leaf, T the number of leaves.

In Extreme gradient boosting the complexity is defined as Equation 16.

$$\omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \qquad (16)$$

The goodness of the tree structure is thus measured in Equation 17.

$$G_j = \sum_{i \in I_j} g_i \ and\ \ H_j = \sum_{i \in I_j} h_i$$

$$obj^{(t)} = \sum\nolimits_{j=1}^{T} \left[G_j w_j + \frac{1}{2}\left(H_j + \lambda\right) w_j^2\right] + \gamma T \qquad (17)$$

Where $I_j = \{I \mid q(x_i) = j\}$ is the set of indexes of data points assigned to j-th leaf.[6]

The Random Forest Algorithm is composed of different decision trees, each with the same nodes, but using different data that leads to different leaves. It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees.[7]

To solve regression problems the mean squared error (MSE), Equation 18, is used to allow the data to branch out in each node.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (f_i - y_i)^2 \qquad (18)$$

Where N is the number of data points, f_i is the value returned by the model and y_i is the actual value for data point i. The MSE is used for calculating the distance of each node from the predicted actual value thus helping in deciding the better decision branch.

In regards to classification problems, Gini index is employed on the decision-making regarding the nodes of a decision tree branch. The Gini index, Equation 19, is a synthetic indicator capturing the level of inequality for a given variable and population, where perfect equality is 0 and 1 is extreme inequality.[7]

$$Gini = 1 - \sum\nolimits_{i=1}^{C} (p_i)^2 \qquad (19)$$

Through this formula, the Gini of each branch is determined using class and probability, thus allowing to determine which branches are most likely to occur. pi represents the relative frequency of the class you are observing in the dataset and c represents the number of classes.

Moreover, another machine learning technique which will be adopted as a classification algorithm is the Logistic Regression. This is based on the probability for a sample to belong to a class. Furthermore, the probabilities must be continuous and bounded between (0, 1). It results to be dependent on a threshold function to make a decision that is called Sigmoid or Logistic function.

The Sigmoid function, Equation 20, $(\oint = 1/(1+exp^{-z}))$ (20), takes real number values as input and transforms them into values in the range [0,1] with an intercept $\oint(Z) = 0.5$. In addition to this, the cross-entropy is commonly used to quantify the difference between two probability distributions. [8]

1.2 Data Description and Preparation

The data that was used for analyzing the problems listed in the previous sections were taken from the `FastF1` python package that contains data relative to all weekend sessions going from FP1 through the race, for the purposes of the analyses that were performed, the data from 2019 to 2022 (current championship) were analysed.

In this section, the data gathered during Free Practices and Qualifying are taken into consideration, whilst in the following chapter those of the race will be analysed.

For each race performed in years 2019 through 2022, the following data are available lap-wise:

- Lap Time set by the driver;
- Driver who set the lap;
- The lap number at which the lap time was set;

- Stint number (corresponds to the period of time in which the drivers exits the pits to then enter them again after a certain number of laps);
- Session time in which the pit-in or pit-out occurred;
- Sector time (each track is divided in three sectors, which are portions of the track at the end of which, the time that the driver took to make it through it, is recorded and then summed up to give the total lap time);
- The speed recorded where the speed trap in each sector is put;
- Compound used for the lap;
- Tyre life, which corresponds to the laps driven on the tyre;
- Fresh tyre which is set to 1 if the tyre is mounted for the first time and 0 otherwise;
- Team for which the driver races;
- Track status, and indicator informing of whether the track in clear, there are yellow or red flags exposed, or if there is a virtual safety car(VSC) or a safety car (SC). These last two devices are deployed during the sessions and race in case of accidents and their meaning will be explored in the following chapter;
- Temperatures regarding air and track;
- Pressure and humidity in session.

In order to be able to better understand data, the lap time was transformed in milliseconds, the fresh tyre indicator was deleted as it was not accurate to the real situation on track, information regarding probabilities of yellow and red flags were added by scraping web information in the Formula One website.

Since the 2022 Grand Prix calendar has race tracks that were used also in the years before, in order to solve the optimization problem regarding the qualifying latest lap starting time, the following variables were taken into consideration for each of the race tracks in which in previous years at least a race was performed:

- `Driver`: variable in which the abbreviation of the driver surname is stored, thus making it a three-digit categorical variable
- `Team`: categorical variable in which the name of the team is contained;
- `Lap time`: numerical variable storing the time in milliseconds that the driver took for performing an entire lap across the circuit. The distribution of data in right-skewed for each year from 2019 to 2022;

- `Sector Time`: numerical variable storing the time in milliseconds referred to the time that each driver, in a selected lap, took to complete the sector, there are three variable related to this data: Sector 1, Sector 2 and Sector 3 timing;
- `Pit-out time`: numerical variable storing the time of the session in which the driver exited the pit-lane (the format being `hh:mm:ss.000`);
- `Pit-in time`: numerical variable storing the time of the session in which the driver entered the pit-lane(the format being `hh:mm:ss.000`);
- `Q session` in which the time was set: three binary variables corresponding to Q1, Q2, Q3 that are set to 1 if the lap was completed in the relative session or to 0 otherwise;
- `Best Lap`: binary variable set to 1 if the lap was the fastest one in the session, 0 otherwise
- `Flying lap`: binary variable set to 1if the lap was a flying lap in the session, 0 otherwise. To determine which laps were flying ones or not, the 107% rule was used. Having determined the fastest lap of the session, as per mandatory requirements of the FIA regulations, a driver which has as best lap of qualifying session a time which is higher than the 107% of the fastest lap set in session in disqualified. Thus if the lap time was lower than this threshold it was deemed as a flying lap;
- `Out-lap`: binary variable set to 1 if the lap was started from the exiting of the pitlane, 0 otherwise;
- `Cooldown lap`: binary variable set to 1 if the lap was performed over the 107% rule following a flying lap;
- `Time lost`: time in milliseconds occurred to perform a lap.

The data needed for the Monte Carlo simulation were collected manually by scraping information through the Formula One website and integrating the information already provided by the `FastF1` package. The data imputed in the simulation are:

- `Event`: categorical variable storing the Grand Prix event name;
- `Yellow flag` in Q1/Q2/Q3: three binary variables taking 1 as a value if at least a yellow flag was exposed during the qualifying session, 0 otherwise;
- `Red flag` in Q1/Q2/Q3: three binary variables taking 1 as a value if at least a red flag was exposed during the qualifying session, 0 otherwise.

Finally, in order to perform the prediction and classification models, the dataset which was imputed was comprised of the following variables measured in a lap-wise fashion:

- `Event`: categorical variable storing the Grand Prix event name, which for the purposed of modelling was one-hot-encoded before its use as independent variable, as it had more than 30 different event names;
- `Year`: numerical value indicating the year in which the event occurred;
- `Driver`: variable in which the abbreviation of the driver surname is stored, thus making it a three-digit categorical variable;
- `Team`: categorical variable in which the name of the team is contained. As teams have a tendency to change names across different seasons due to change in the ownership or management, the name of each team was updated to the current season name. Then the variable was label-encoded as it contained only 11 different values;
- `Lap time FP2/FP3`: two numerical variable storing the fastest lap set in each of these two sessions . FP2 and FP3 where the only free practices taken into consideration due to the fact that FP1 is the least representative session in comparison with the track conditions during both race and qualifying timing sessions. The team will usually employ the whole session to test the aerodynamics, eventually new parts to trial are adopted and mainly working on the general balance of the car. Instead, FP2 is the most representative, track conditions-wise, of the three free practice sessions due to the fact that takes place at roughly the same time at qualifying and race will take place in subsequent days , hence during this session the teams are focused much more tightly on qualifying and race preparation. Moreover, the FP3 session is mostly used to confirm the data and balancing brough forth during the FP2. [9] This implying that, in order to predict the qualifying time, the best data to take into consideration are those of FP2 and FP3 both;
- `Tyre life` information for FP2 and FP3: two numerical variables storing the number of laps that were done with the tyre at the moment of the fastest lap timing;
- The type of `compound` used to make the lap in FP2/FP3: two categorical variables listing 0 if hard, 1 if medium and 2 if soft. There has to be taken into consideration

the fact that till last year the drivers who qualified for Q3 had to start the race with the compound used to set the fastest lap in Q2, thus the presence of medium compound in previous years, especially in Q2. From this year, onward, the teams can freely decide which compound to mount during the race independently of the ones used during Q2 [2];

- `Air` and `Track temperature` for both FP2/FP3: two numerical variables for each Free Practice, stating respectively, the air temperature and track temperature that were present during the lap;
- A variable of the `optimal lap` was created combining the best sector times set during through all free practices. This lap is very close to the timing that the drivers will have during qualifying, and it is the best indicator of driver ability that we can get. As a matter of fact, due to different aspects, the fastest lap set by a driver is not always the one comprised of the absolute fastest sectors set during FPs due to the fact that, micro mistakes can always occur when driving to the limit;
- `Optimal sectors` were stated as three numerical variables, encasing the three values used to compute the optimal lap;
- `Lap time` regarding the fastest lap set during the qualifying session, which is the dependent variable of the regression and classification models;
- `Q1, Q2, Q3` binary variables taking the value 1 if the time was set in that Q and 0 otherwise. These three variables will be the dependent variables of the classification models.

A correlation matrix was provided to preliminarily look into the relationships among the different variables at hand, the result of which is in Figure 2.

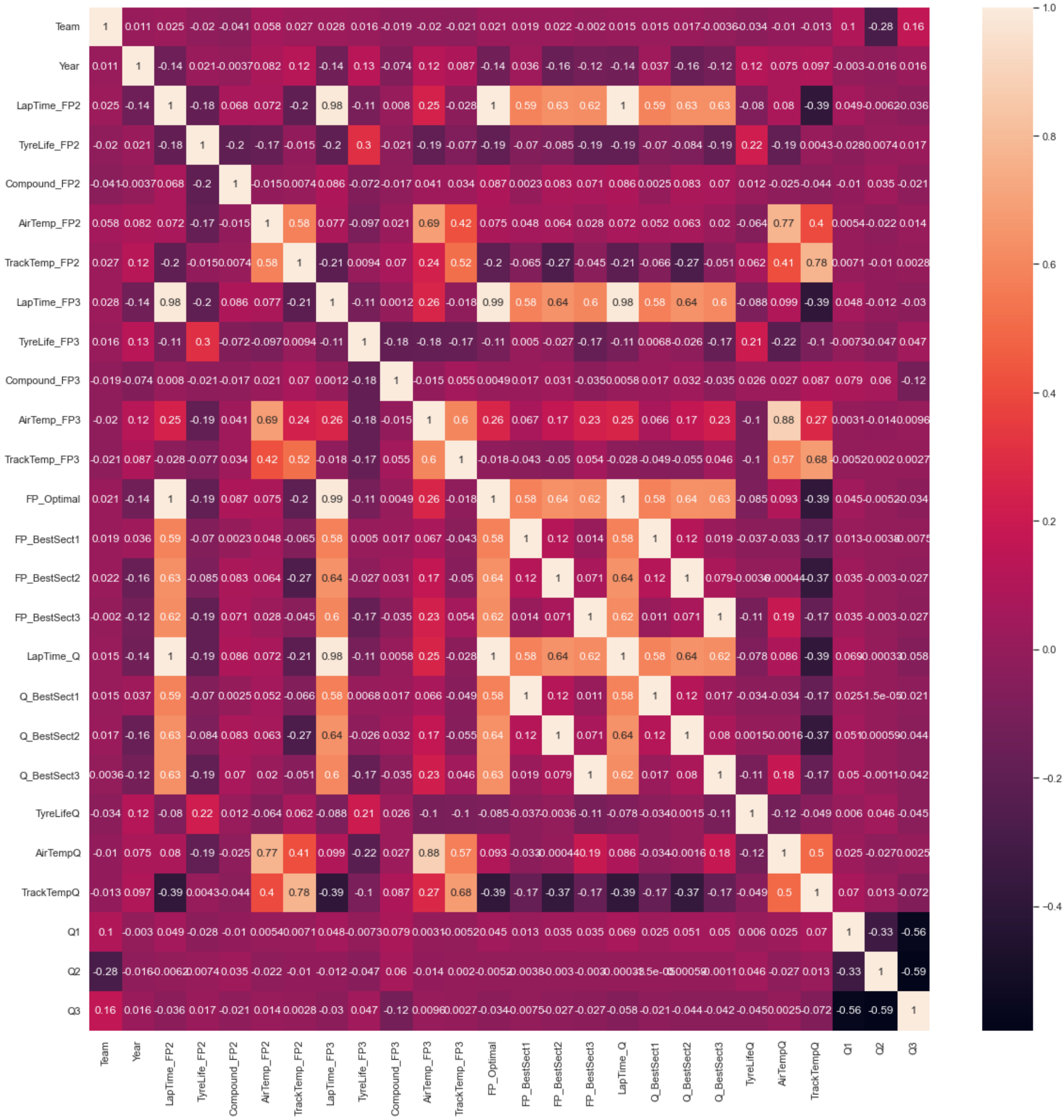

Figure 2 *Correlation matrix of independent variables used for prediction modelling*

From Figure 2 we can infer the relationships among variables that were expected:

- Track temperature in each session in dataset (FP2, FP3, Q) is negatively correlated to the lap time of the relative session, as a matter of fact the higher the temperature the slower the lap timing will be. Of course, it is positively related to Air temperature, given that it is the air temperature that impact is the track one.

- The lap time set at qualifying is highly positively correlated (values close to 1 if not 1) to FP2, FP3 and optimal lap time, meaning that they are a good indicator of how fast the lap time in qualifying session will be set.
- Tyre life, irrespective of the session, is negatively related to the lap time, suggesting that as the laps on the tyre increase the lap time will be lower. This is consistent to the fact that as the tyre reaches the optimal temperature through consecutive usage, the grip increases leading to more velocity and thus faster lap times.
- Of course the sector timings and the lap timings are positively related as they are a consequence of each other.
- Then Q1, Q2, Q3 are all negatively correlated to each other, this could be due to the fact that as only the best 15, then 10 drivers are able to make the cut to go through the next session, the fastest drivers will reach the faster time in subsequent to Q1 sessions thus the negative relation among these three variables.

1.3 Data Modelling and Results

By setting the various datasets that will be used to compile and assess the models listed in the first two sections of this chapter, the final models and their solutions will thereby be provided below.

Starting from the optimization problem regarding the tyre allocation, for the purposes of these analysis only the dry weekends will be considered as the variability in conditions are more stable, thus the results that came from minimizing the model proposed in section 1.1.1 are as follows:

- in FP1 two soft tyres will be used,
- in FP2 a hard tyre and a soft tyre will be used, and finally,
- in FP3 two medium tyres will be used.

This result is consistent to the fact that the softer compound is used during FP1 which is the session in which all teams trial on various setting and parts in order to prepare the car for the following sessions. During FP2, it is customary to try a long run with the hardest compound so to understand whether the car setting is consistent with what the race pace is predicted to be, giving valuable information regarding tyre wear and degradation that could appear during the race. Then for the last part of this session, a soft tyre will be used to try to set fast lap times, en vis of the upcoming qualifying the next day. During the FP3 the car is again tested to see the trend with the medium compound that will be used during the race.

Regarding the part of the optimization problem, which is related to qualifying, the resulting allocation of tyres is presented in Table 1.

***Table 1** Results of Tyre Allocation Optimization Problem*

Range of Team	Q1	Q2	Q3
Top-Team	1	2	2
Midfield Team	2	2	1

Note that regardless of the fact that a driver will reach the last session of qualifying, a set of soft tyres will have to not be used prior to the Q3 session, this is mandatory for all drivers on the grid.

By compiling the Monte Carlo simulation over the past data form 2019 to 2021, the results are reported Table 2. It can be noted that, for certain Grand Prixs the probability of having either a yellow flag or a red flag is very high, due to the fact that there are circuits with a narrow track coupled with a low number of escape routes, causing the increase in the accident probability.

***Table 2** Monte Carlo Simulation Probabilities of Yellow and Red Flags in Qualifying Sessions*

Type of Flag exposed	Qualifying Session	Predicted Races affected	Prob(flag in session)
Yellow	Q1	Bahrain Grand Prix	0.67
		Azerbaijan Grand Prix	0.67
		Belgian Grand Prix	0.67
		Mexico City Grand Prix	0.67
		Abu Dhabi Grand Prix	1
		British Grand Prix	0.33
		French Grand Prix	0.33
	Q2	Azerbaijan Grand Prix	0.67

		Italian Grand Prix	0.67
		Canadian Grand Prix	0.33
		Saudi Arabian Grand Prix	0.33
		Abu Dhabi Grand Prix	0.33
	Q3	Saudi Arabian Grand Prix	0.33
		Italian Grand Prix	0.33
		Mexico City Grand Prix	0.33
Red	Q1	Azerbaijan Grand Prix	0.67
		French Grand Prix	0.33
		Belgian Grand Prix	0.33
		Italian Grand Prix	0.33
	Q2	Azerbaijan Grand Prix	0.67
		Bahrain Grand Prix	0.33
	Q3	Monaco Grand Prix	0.33

Moreover, as yellow and red flags are caused by drivers mistakes, the probability for each driver to be the cause of these flags were calculated, the results of which are shown in Table 3.

***Table 3** Monte Carlo Simulation Results on Probabilities of Drivers to have caused the yellow or red flag*

Type of Flag exposed	Qualifying Session	Driver involved	Prob(flag|causing flag)
Yellow	Q1	Hamilton L.	0.029
		Stroll L.	0.029
		Hulkenberg N.	0.029
		Ricciardo D.	0.029
		Vettel S.	0.088
		Leclerc C.	0.029
		Perez S.	0.059
		Albon A.	0.059
		Magnussen K.	0.029
		Verstappen M.	0.059
		Russel G.	0.029
		Latifi N.	0.059
		Sainz C.	0.088
		Alonso F.	0.029
		Schumacher M.	0.029
		Tsunoda Y.	0.029
	Q2	Magnussen K.	0.2
		Leclerc C.	0.07
		Ocon E.	0.13
		Perez S.	0.13
		Ricciardo D.	0.07
		Vettel S.	0.07
		Sainz C.	0.07

		Alonso F.	0.07
	Q3	Hulkenberg N.	0.11
		Bottas V.	0.22
		Ocon E.	0.11
		Hamilton L.	0.11
		Gasly P.	0.11
		Verstappen M.	0.11
Red	Q1	Perez S.	0.07
		Albon A.	0.07
		Tsunoda Y.	0.14
		Schumacher M.	0.14
		Stroll L.	0.14
		Norris L.	0.07
	Q2	Leclerc C.	0.25
		Bottas V.	0.25
		Latifi N.	0.5
		Sainz C.	0.5
		Ricciardo D.	0.25
		Russel G.	0.25
	Q3	Leclerc C.	0.2
		Sainz C.	0.2
		Tsunoda Y.	0.2
		Norris L.	0.2

At last, the regression models where compiled with the information reported in the section above.

In order to obtain the best training regression model, the training dataset was divided in training and validation sets, with respectively 75% of randomly chosen data and 25% of randomly chosen data.

Through the usage of the `sklearn` method `GridSearchCV` which is used for implementing a "fit" and a "score" method, performing an exhaustive search over specified parameter values for an estimator, the best Random Forest Regression model was identified with the following characteristics:

- `Number of trees to be trained`: 50 and
- `Maximum depth of a tree`: 5

Yielding a mean absolute error of 1200.2 and an R-squared score of 0.93 when tested over the validation set.

This model was then trained on the whole dataset and tested over the test set build with the first race of 2022 results of FP2 and FP3, resulting in a substantial reducing in terms of mean absolute error, amounting to 375.96 and an R-squared score of 0.735.

Whilst using the same train and validation split, the best trained Extreme Gradient Boosting model with, as characteristics:

- `Objective function`: regression with squared errors;
- `Learning rate` of 0.1
- `Maximum depth` of 15
- `Number of trees to be trained`: 100
- `Alpha`: 10

Resulting is a mean squared error of 3548.768 and an R-squared score of 0.9155 when tested over the validation set. Whilst yielding a mean squared error of 1053.01 and an R-squared score of 0.730 for the testing over the test set of the first race of 2022 season.

Thus both models are able to correctly predict, respectively, 0.735 and 0.730 of the qualifying lap time future data, having in the Random Forest technique the best prediction.

Moreover, as the qualifying session in which the time was set is very important in determining the grid position for the race, classifications models where run with Q1, Q2, Q3 variables as dependent ones, the models employed where once again the Random Forest and the Logistic model.

The best Random Forest classification model trained to predict Q1 outcome is that with the following characteristics:

- `Number of estimators`: 400
- `Maximum depth`:10
- `Criterion`: Gini
- and a `Min Max scaler` (it transforms features by scaling each feature to a given range, in this case 0,1) applied to both dependent and independent variables.

The model resulted, when tested against the validation set in an accuracy score of 0.866 and a ROC AUC score of 0.837. Then it was tested against the data gathered from the first grand prix, the accuracy score was of 0.80.

Then a Logistic model was trained to predict the Q1 outcome with the inverse of regularization strength equal to 1 was tested against the validation set resulting in and accuracy score of 0.80 and a ROC AUC score of 0.833. When tested against the test set the accuracy score amounted to 0.80. Thus, based on the accuracy score of the test set, both classification models can be used interchangeably.

In order to predict Q2 the best Random Forest classification model had the same characteristics as before, with a ROC AUC score of 0.688 and when tested against the validation set the accuracy score was of 0.742 whilst tested against the test set it resulted in an accuracy score of 0.75.

Moreover, a Logistic model was trained to predict the same Q2 outcome with the inverse of regularization strength equal to 0.1 was tested against the validation set resulting in and accuracy score of 0.76 and a ROC AUC score of 0.708. When tested against the test set the accuracy score result gave 0.80. Indeed, based on the accuracy score of the test set, the model that was able to better capture the data trend is the Logistic one.

As of the last two classifications models to predict the Q3 outcome, the Random Forest one had as `number of estimators` a value of 500, a `maximum depth` of 50 with `min max scaler` applied to both dependent and independent variables. The result tested against the validation sets where of 0.853 for the ROC AUC score and 0.78 as accuracy score, while tested on the test set it yielded an accuracy score of 0.80. In addition to this, the Logistic model was trained with the inverse of regularization strength equal to 1 which, in turn resulted in a ROC AUC score of 0.823 with an accuracy score of 0.75 when tested against the validation set and of 0.75 when tested against the test set. Therefore, the best model to classify Q3 outcomes is the Random Forest one.

Thanks to these models, the qualifying prediction to calculate the latest starting time to perform either of the two exit strategies yielded the results in Table 4, for each driver for the first race of the year. Note that the two rows that are selected, correspond to the predicted maximum time to access the following qualifying session. Thus for those drivers that had a higher predicted timing the exit strategies for the two following session were not calculated. Indeed, with this prediction overview, those teams who have their drivers in the elimination zones, could opt for the realization of strategy number 2, consisting of two flying laps performed on two fresh sets allowing the driver to have more than one chance to set the fastest time possible. This results to be consistent with the tyre allocation optimization results.

Table 4 *Bahrain Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
Leclerc C. (LEC)	90799	241	9.19	7.07	6.19	4.07	3.19	1.07
Verstappen M. (VER)	90799	118	9.15	7.06	6.15	4.06	3.15	1.06
Sainz C. (SAI)	90970	283	8.81	6.50	5.81	3.50	2.81	0.50
Perez S. (PER)	91371	450	8.76	6.41	5.76	3.41	2.76	0.41
Russell G. (RUS)	91392	-824	9.07	6.84	6.07	3.84	3.07	0.84
Hamilton L. (HAM)	91498	260	9.16	6.92	6.16	3.92	3.16	0.92
Bottas V. (BOT)	91787	227	8.90	6.72	5.90	3.72	2.90	0.72
Alonso F. (ALO)	91902	-293	8.21	5.53	5.21	2.53	2.21	0.21
Magnussen K. (MAG)	91916	108	9.22	7.13	6.22	4.13	3.22	1.13
Schumacher M. (MSC)	**91966**	-32	8.98	6.78	5.98	3.78	2.98	0.78
Norris L. (NOR)	92026	18	8.47	5.83	5.47	2.83		
Gasly P. (GAS)	92416	78	8.69	6.20	5.69	3.20		
Hulkenberg N. (HUL)	92464	-313	8.70	6.29	5.70	3.29		
Ocon E. (OCO)	92588	806	8.73	6.44	5.73	3.44		
Zhou G. (ZHO)	**92755**	-788	8.84	6.53	5.84	3.53		
Stroll L. (STR)	92755	-277	8.29	6.07				
Tsunoda Y. (TSU)	92780	30	9.00	6.83				
Ricciardo D. (RIC)	93477	532	8.64	6.30				
Latifi N. (LAT)	93503	-131	8.83	6.44				
Albon A. (ALB)	93527	863	8.77	6.44				

The calculation of the latest starting time can be essential in the performance of the optimal exit strategy. As the track conditions will increase in driveability, allowing the drivers to extract form their car an ever better performance, exiting further down the session time, could give the driver the possibility to set his fastest time in the best conditions that the track could offer.

Moreover, as the aim of all drivers and teams is to start in a position which is closest to the starting line, consistently with the optimal tyre allocation problem solutions and the predicted timing, the strategy number two is majorly taken into consideration by those team-driver couplings that are in the midfield where having two chances at setting the fastest time results in classifying better than predicted.

In the Tables 4 to 13, the prediction for all races of the season are portrayed. It is important to point out the fact that, for numerous reasons, not all the drivers take part to the qualifying sessions or are incapacitated, by accidents, to perform at least a fast lap. Hence the difference in the number of drivers that can be seen in the Tables form 4 to 13. As previously said, only the qualifying sessions that were performed in dry conditions and on weekends with the traditional format are taken into consideration.

Table 5 *Saudi Arabian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
LEC	88757.3	532.33	9.42	7.20	6.42	4.20	3.42	1.20
VER	88757.3	296.33	9.42	7.20	6.42	4.20	3.42	1.20
PER	88781.5	581.51	9.42	7.20	6.42	4.20	3.42	1.20
SAI	88928.5	526.49	9.40	7.18	6.40	4.18	3.40	1.18
ALO	88928.5	-218.51	9.40	7.18	6.40	4.18	3.40	1.18
BOT	88928.5	-254.51	9.40	7.18	6.40	4.18	3.40	1.18
OCO	88991.6	-76.44	9.40	7.17	6.40	4.17	3.40	1.17
GAS	88991.6	-262.44	9.40	7.17	6.40	4.17	3.40	1.17
HAM	89006.4	-1336.62	9.40	7.17	6.40	4.17	3.40	1.17
MAG	**89055.7**	-532.28	9.39	7.16	6.39	4.16	3.39	1.16
MSC	89348.4	-571.61	9.36	7.13	6.36	4.13		
NOR	89440.7	-210.34	9.35	7.12	6.35	4.12		
RUS	89654.4	550.38	9.33	7.09	6.33	4.09		
RIC	89792.2	19.19	9.32	7.08	6.32	4.08		
ZHO	**89792.2**	-26.81	9.32	7.08	6.32	4.08		
STR	89792.2	-1216.81	9.32	7.08	6.32	4.08		
TSU	89900	Didn’t Start						
HUL	89943.1	-599.89	9.31	7.06				
ALB	90118.5	-373.53	9.29	7.04				
LAT	90623.4	-1193.62	9.24	6.97				

Table 6 *Australian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
LEC	77721.40	-146.60	10.48	8.53	7.48	5.53	4.14	2.20
VER	78421.40	267.40	10.44	8.48	7.44	5.48	4.10	2.15
RUS	78643.82	-289.18	10.38	8.41	7.38	5.41	4.05	2.07
SAI	78721.40	-686.60	10.35	8.36	7.35	5.36	4.01	2.03
NOR	78872.09	169.09	10.39	8.42	7.39	5.42	4.05	2.09
PER	78873.94	633.94	10.42	8.46	7.42	5.46	4.08	2.13
HAM	78983.41	158.41	10.37	8.40	7.37	5.40	4.04	2.07

RIC	78983.41	-48.59	10.36	8.39	7.36	5.39	4.03	2.05
OCO	79255.38	194.38	10.35	8.37	7.35	5.37	4.02	2.04
GAS	**79284.87**	58.87	10.34	8.36	7.34	5.36	4.01	2.03
ALO	79384.59	569.59	10.36	8.39	7.36	5.39		
TSU	79643.82	219.82	10.32	8.33	7.32	5.33		
MSC	80547.03	82.03	10.22	8.21	7.22	5.21		
ALB	80547.03	412.03	10.24	8.24	7.24	5.24		
MAG	**80556.05**	302.05	10.23	8.23	7.23	5.23		
ZHO	80656.05	501.05	10.23	8.23				
VET	80906.32	-242.68	15.30	15.30				
LAT	81581.16	209.16	10.13	8.09				

Table 7 *Miami Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
LEC	88655.92	-140.08	9.43	7.21	6.43	4.21	3.43	1.21
PER	88655.92	-380.08	9.43	7.21	6.43	4.21	3.43	1.21
RUS	88655.92	-1517.08	9.43	7.21	6.43	4.21	3.43	1.21
VER	88659.33	-331.67	9.43	7.21	6.43	4.21	3.43	1.21
HAM	88733.81	-891.19	9.42	7.20	6.42	4.20	3.42	1.20
SAI	89104.33	118.33	9.39	7.16	6.39	4.16	3.39	1.16
TSU	89604.33	-327.67	9.34	7.10	6.34	4.10	3.34	1.10
ALO	89757.99	-402.01	9.32	7.08	6.32	4.08	3.32	1.08
GAS	89900.65	210.65	9.31	7.06	6.31	4.06	3.31	1.06
BOT	**89960.12**	270.12	9.30	7.05	6.30	4.05	3.30	1.05
RIC	89985.63	-324.37	9.30	7.05	6.30	4.05		
NOR	90034.97	284.97	9.30	7.05	6.30	4.05		
STR	90034.97	-641.03	9.30	7.05	6.30	4.05		
MSC	90385.63	-37.37	9.26	7.00	6.26	4.00		
VET	**90604.33**	390.33	9.24	6.98	6.24	3.98		
MAG	90757.99	-508.01	9.23					
ALB	90858.36	-407.64	9.22					
ZHO	90964.40	-55.60	9.21					
LAT	91992.39	667.39	9.11					

Table 8 *Monaco Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
LEC	71328.9	-47.13	11.10	9.32	8.10	6.32	5.10	3.32
SAI	71328.9	-272.13	11.10	9.32	8.10	6.32	5.10	3.32
PER	71328.9	-300.13	11.10	9.32	8.10	6.32	5.10	3.32
VER	71328.9	-337.13	11.10	9.32	8.10	6.32	5.10	3.32
NOR	71355.1	-493.90	11.10	9.32	8.10	6.32	5.10	3.32
RUS	71668.3	-443.68	11.07	9.28	8.07	6.28	5.07	3.28
ALO	71886.2	-360.77	11.05	9.25	8.05	6.25	5.05	3.25

TSU	72030.7	-766.31	11.04	9.24	8.04	6.24	5.04	3.24
VET	72183.2	-548.84	11.02	9.22	8.02	6.22	5.02	3.22
HAM	**72343**	-216.54	11.01	9.20	8.01	6.20	5.01	3.20
MAG	72478.2	-442.83	10.99	9.18	7.99	6.18		
BOT	72939	30.05	10.95	9.13	7.95	6.13		
MSC	72939	-141.95	10.95	9.13	7.95	6.13		
RIC	73002.5	38.49	10.94	9.12	7.94	6.12		
ALB	**73062**	-549.35	10.94	9.11	7.94	6.11		
OCO	73138.3	91.31	10.93					
GAS	73379.4	-280.57	10.91					
STR	73994.3	316.28	10.85					
ZHO	74568.2	-1037.82	10.79					
LAT	74815.1	412.08	10.77					

Table 9 *Azerbaijan Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
PER	101192.43	-448.57	8.22	5.69	5.22	2.69	2.22	-0.31
LEC	101221.32	-137.68	8.22	5.68	5.22	2.68	2.22	-0.32
VER	101921.32	215.32	8.15	5.60	5.15	2.60	2.15	-0.40
SAI	102047.07	233.07	8.14	5.58	5.14	2.58	2.14	-0.42
RUS	102676.59	-35.41	8.07	5.51	5.07	2.51	2.07	-0.49
GAS	102745.7	-99.30	8.07	5.50	5.07	2.50	2.07	-0.50
ALO	102781.47	-391.53	8.06	5.49	5.06	2.49	2.06	-0.51
VET	102932.23	-158.77	8.05	5.48	5.05	2.48	2.05	-0.52
TSU	102943.1	-112.90	8.05	5.48	5.05	2.48	2.05	-0.52
HAM	**103171**	247.34	8.03	5.45	5.03	2.45	2.03	-0.55
NOR	103476.59	78.59	8.00	5.41	5.00	2.41		
ZHO	103671.34	-118.66	7.98	5.39	4.98	2.39		
OCO	103689.89	104.89	7.98	5.38	4.98	2.38		
RIC	103762.49	188.49	7.97	5.38	4.97	2.38		
ALB	**104201**	-517.71	7.93	5.32	4.93	2.32		
BOT	104422.37	-21.63	7.91	5.30				
MAG	104684.18	41.18	7.88	5.26				
MSC	105401.29	-373.71	7.81	5.18				
LAT	105507.38	140.38	7.80	5.16				
STR	105671.34	300.34	7.79	5.14				

Table 10 French Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
VER	91249.1	73.13	9.10	6.75	6.10	3.75	3.10	0.75
LEC	91564.1	692.09	9.07	6.71	6.07	3.71	3.07	0.71
HAM	91930.1	165.14	9.04	6.66	6.04	3.66	3.04	0.66
PER	91947.3	612.27	9.04	6.66	6.04	3.66	3.04	0.66
RUS	92219.8	88.75	9.01	6.63	6.01	3.63	3.01	0.63
ALO	92451.2	-100.76	8.99	6.60	5.99	3.60	2.99	0.60
NOR	92490.5	458.52	8.98	6.59	5.98	3.59	2.98	0.59
RIC	92904.8	-17.22	8.94	6.54	5.94	3.54	2.94	0.54
TSU	92948.1	168.11	8.94	6.54	5.94	3.54	2.94	0.54
BOT	**92948.1**	-103.89	8.94	6.54	5.94	3.54	2.94	0.54
OCO	93098.2	50.22	8.92	6.52	5.92	3.52		
STR	93098.2	-340.78	8.92	6.52	5.92	3.52		
VET	93459.4	183.36	8.89	6.47	5.89	3.47		
ALB	93533.9	110.86	8.88	6.46	5.88	3.46		
GAS	**93630.2**	191.25	8.87	6.45	5.87	3.45		
ZHO	93748.1	74.11	8.86	6.44				
LAT	93910.7	116.68	8.84	6.42				

Table 11 Belgian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
VER	103745	79.91	7.88	5.20	4.88	2.20	1.88	-0.80
PER	103745	-717.09	7.88	5.20	4.88	2.20	1.88	-0.80
SAI	104372	75.29	7.82	5.13	4.82	2.13	1.82	-0.87
LEC	104372	-180.71	7.82	5.13	4.82	2.13	1.82	-0.87
OCO	105569	388.78	7.71	4.98	4.71	1.98	1.71	-1.02
ALO	105569	200.78	7.71	4.98	4.71	1.98	1.71	-1.02
HAM	105609	105.77	7.70	4.97	4.70	1.97	1.70	-1.03
RUS	105909	132.77	7.67	4.94	4.67	1.94	1.67	-1.06
RIC	105929	161.77	7.67	4.94	4.67	1.94	1.67	-1.06
ALB	**105969**	131.78	7.67	4.93	4.67	1.93	1.67	-1.07
MSC	106409	-1309.23	7.63	4.88	4.63	1.88		
VET	106469	124.78	7.62	4.87	4.62	1.87		
ZHO	106489	403.77	7.62	4.87	4.62	1.87		
NOR	106534	355.68	7.61	4.86	4.61	1.86		
LAT	**106569**	167.78	7.61	4.86	4.61	1.86		
STR	106609	-2.23	7.61	4.85				
MAG	106609	51.77	7.61	4.85				
TSU	106609	-83.23	7.61	4.85				

GAS	106619	791.77	7.60	4.85				
BOT	107869	2.78	7.48	4.70				

Table 12 *Dutch Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
VER	70400.93	58.93	11.19	9.43	8.19	6.43	5.19	3.43
LEC	70551.014	188.01	11.18	9.42	8.18	6.42	5.18	3.42
SAI	70727.405	293.40	11.16	9.39	8.16	6.39	5.16	3.39
HAM	71128.872	480.87	11.12	9.35	8.12	6.35	5.12	3.35
RUS	71227.405	80.40	11.11	9.33	8.11	6.33	5.11	3.33
PER	71289.243	212.24	11.11	9.33	8.11	6.33	5.11	3.33
NOR	71328.872	154.87	11.10	9.32	8.10	6.32	5.10	3.32
ALO	71328.872	-284.13	11.10	9.32	8.10	6.32	5.10	3.32
MSC	71353.197	-88.80	11.10	9.32	8.10	6.32	5.10	3.32
OCO	**71355.1**	-249.90	11.10	9.32	8.10	6.32	5.10	3.32
GAS	71379.43	-132.57	11.10	9.32	8.10	6.32		
ZHO	71467.295	-236.70	11.09	9.30	8.09	6.30		
ALB	71553.197	-248.80	11.08	9.29	8.08	6.29		
RIC	71942.442	-138.56	11.05	9.25	8.05	6.25		
TSU	**72042.4**	-513.56	11.04	9.23	8.04	6.23		
MAG	72275.464	-43.54	11.01	9.21				
BOT	72280.466	319.47	11.01	9.21				
VET	72353.197	-37.80	11.01	9.20				
LAT	72733.603	-619.40	10.97	9.15				

Table 13 *Italian Grand Prix Qualifying Predictions and Latest Starting Times for Exit Strategies*

Driver	Predicted Qualifying Time (ms)	Difference Predicted - Actual (ms)	Strategy 1 Latest Exit Time Q1 (min)	Strategy 2 Latest Exit Time Q1 (min)	Strategy 1 Latest Exit Time Q2 (min)	Strategy 2 Latest Exit Time Q2 (min)	Strategy 1 Latest Exit Time Q3 (min)	Strategy 2 Latest Exit Time Q3 (min)
LEC	80484.38	323.38	10.15	8.07	7.15	5.07	4.15	2.07
SAI	80532.98	103.98	10.15	8.07	7.15	5.07	4.15	2.07
VER	80541.90	235.90	10.15	8.07	7.15	5.07	4.15	2.07
PER	80666.34	-539.66	10.14	8.05	7.14	5.05	4.14	2.05
RIC	81554.06	-370.94	10.05	7.94	7.05	4.94	4.05	1.94
RUS	81666.34	124.34	10.04	7.93	7.04	4.93	4.04	1.93
NOR	81667.34	83.34	10.04	7.93	7.04	4.93	4.04	1.93
HAM	81707.17	183.17	10.03	7.92	7.03	4.92	4.03	1.92
BOT	81780.92	-454.08	10.03	7.91	7.03	4.91	4.03	1.91
MAG	82047.20	-592.80	10.00	7.88	7.00	4.88	4.00	1.88
VET	82122.30	-513.70	9.99	7.87	6.99	4.87		
OCO	82322.46	192.46	9.97	7.85	6.97	4.85		
ALO	82373.06	512.06	9.97	7.84	6.97	4.84		
STR	82387.98	-360.02	9.97	7.84	6.97	4.84		

GAS	82554.06	-93.94	9.95	7.82	6.95	4.82		
ZHO	82554.06	-22.94	9.95	7.82				
LAT	82895.44	308.44	9.92	7.78				
MSC	83387.98	382.98	9.87	7.72				

For the last four races (to this date, 26/09/2022, there are 5 races to be performed one of which is a sprint type of weekend, therefore out of scope of this research), the dataset used to make the predictions was updated with the data gathered in the previous races of the current season, thus yielding the result presented in Table 14. The data results in Table 14 regard the prediction of the cut-off time for both Q1 and Q2 and the prediction for the pole in terms of milliseconds. Knowing the cut-off time could enable the teams to formulate alternative strategies given the timings set by their drivers during the free practices' sessions.

Table 14 *Prediction of cut-off timing data and pole time*

Race	Q1 Predicted cut-off time (ms)	Q2 Predicted cut-off time (ms)	Pole predicted time (ms)
Singapore Grand Prix	97979	97683	97357
COTA Grand Prix	97089	95874	93512
Mexico City Grand Prix	78594	78233	76188
Abu Dhabi Grand Prix	83124	82597	82399

Chapter 2

The main event of the racing weekend is, indeed, the race, in which both teams and drivers will be awarded points for the championships, given that they finish in the top 10 positions. Points are also awarded during the sprint race in those weekends in which this format is applied. As it was stated beforehand, the race weekends in which a sprint is scheduled are excluded from our analysis because the organization of the sessions is extremely different and does not allow to undergo the analyses proposed in the previous chapter. Moreover, as there is no possibility of making a pit stop for changing tyres or damaged car parts, because, by regulation, the outcome of this sprint race (that is going to determine the starting grid of the race) is heavily dependent on drivers' and cars' capabilities, it does not involve any pit-stop strategy. The only decision to take is choosing the best compound for the race, which will, in turn, be dependent on the tyre supplier projection for the sprint race. [2]

The race ending points are awarded based on the position in which the driver is when crossing the finish line. For the Drivers' championship the points are given following Table 15. [2]

***Table 15** Point Assignment during Formula One race*

1ST	2ND	3RD	4TH	5TH	6TH	7TH	8TH	9TH	10TH
25	18	15	12	10	8	6	4	2	1

In addition to this, the driver that performs the fastest lap and arrives withing the first 10 positions gets an additional point [2]. The points awarded for the teams' championship is the sum of the points gained individually by the teams' drivers. Thus, it is important, for both drivers and teams to gain as many points as it is possible as the championship (both drivers' and teams') could be determined even by 1 point of difference.

In this type of competition the race strategy is made of two main aspects: pit-stops and tyre compounds. By doing a pit-stop, the driver enters the pit lane at a speed of maximum

[2] During the sprint race points are awarded to the first 8 finishers, for 1st to 8th the distribution of points will be: 1st – 8 points; 2nd – 7 points; 3rd – 6 points; 4th - 5 points; 5th – 4 points; 6th – 3 points; - 7th – 2 points; 8th – 1 point.

80 km/h. Then, when reaching the his team's garage the car is stilled to change the tyres for a matter of seconds (during the current season the fastest standstill moment to change tyres has been of 1.9 seconds), this portion of the pit-stop can be timed, on average, of 2.5 seconds. Of course, there is variability due to how fast the mechanics' crew that is charged with changing tyres will perform it, thus allowing for great variability for this aspect. Once the tyres are changed, the car will go through the rest of the pit lane and then enter the race track once again.

As it can be understood from the above descriptions, the issue with the pit stop is the fact that the car which is entering the pits will go significantly slower than those who will continue on the race track, thus implying a position loss when the pitted car re-enters the race track. Doing at least a pit-stop in a race is mandatory for each team, but the timing with which it is done and the number of pit-stops entirely depends on the strategy that the single team will adopt during the race. [10]

Indeed, concerning the strategies of the pit-stops we have to consider three main possibilities:

1. Covering the pit-stop strategy on the driver that precedes, meaning that the follower will follow exactly the pit-stop timing of the driver ahead. This could be done in the eventuality that the optimal time-window (that is usually about three to five laps) that gives the best estimated result, is equal for the two drivers. This is usually adopted to preserve the position in which a driver is, in the case in which there is no possibility of overcoming the opponent.
2. Undercut pit-stop strategy: this strategy could be adopted by either the driver ahead (which, of course, could even be the leader of the race) or by the following driver. This strategy relies on the fact that, if carefully timed, when pitting earlier than the opponent, the driver will have fresh tyres and clean air in order to heathen the tyres correctly and coincidentally driving as fast as possible to gain advantage time with respect to the opponent. So that, when the driver will perform the pit-stop the first to stop will have built a time gap that could allow him an advantage through the race. This strategy is usually adopted when the track is heavily impacting tyre degradation and, more importantly, to perform faster times in clean air (this is when the drivers has no one in front of them for at least two to three seconds thus allowing him to bring the tyres in the optimal temperature of

employment while still performing the fastest laps in order to gain advantage over the cars which are in front). The issue with this strategy is that if the drivers pits too early in comparison to his opponent he will not gain the desired outcome. Indeed, ss the tyres will wear before the opponent, thus having significantly worn tyres at the end of the race in comparison to the opponent's fresher ones. Hence, this strategy is optimal when the rival pits at most 3 laps later than the one who performed the undercut.

3. Overcut pit-stop strategy: this strategy is adopted in tracks in which tyre degradation is minimum. Thus, the driver will perform the pit-stop at the latest time as possible, implying having clear air when other opponents will pit, and consequently allowing him to speeding up the pace in order to exploit the chance given. Due to the fact that the track is not affecting heavily the tyres performance, this strategy allows the driver to exploit major tyre performances till the maximum time limit of the tyres.

The mostly used strategy is the undercut one, which has a significant advantage for followers when they find themselves stuck behind a driver that they are not able to overcome, but not slow enough to be solidly behind the leader, thus wearing off the tyres faster due to turbulence caused by close contact of the two cars. This strategy is also advantageous when looking after performing the fastest timing, because the driver will most likely find clean air with a fresh tyre, allowing him the optimal condition to set a time that could be the one awarding the surplus point.

Regarding the overcut strategy, the advantage is clear in tracks where it is not easy to overcome an opponent and the tyres do not degrade fast, thus allowing optimal performance for the duration of the stint (a stint is the driver's time on track between two pit-stops). The most clamoured example of this strategy was the one adopted by Williams on the driver Alexander Albon during this year Australian grand prix, in which they opted to use the hard compound till the penultimate lap (making the stint last for 57 laps out of 58) then changing the tyres to come out in tenth position to claim the first point of the season for the team. This is, by all means, an extreme case of overcutting but shows how, in certain contexts, it is better to lengthen to the maximum possibilities the tyre stint in order to allow the driver to gain positions by not losing time on pitting in the earlier stages of the race.

Furthermore, whilst deciding the correct timeframe to pit is of vital importance, when considering this aspect each team has to take into consideration the tyre compound which they have available and the performance they have estimated for each different compound, derived from the timings set during FP2 and FP3 sessions. As a matter of fact these two sessions are vital for both qualifying pace and race pace due to the fact that they're approached, by the teams, as a simulation of the race thus making the drivers do a long stint to simulate race conditions and obtain data on tyre wear and degradation.

Thus, to optimize the strategy adopted during the race, the teams have to plan the tyre usage, meaning that they have to choose the tyre compound to start with and which will be the ones to follow that. Due to the fact that, by regulatory demands, every driver has to change the tyre compound at least one time during the race, the type of compound with which they start is of vital importance for the overall strategy, especially when considering races in which only a pit-stop will be performed.

Usually, teams choose the softer compound as the starting one, when they want to be set in an "aggressive" strategy. Due to the fact that, the softer compound enters the correct usage temperature faster than any of the other three compounds, it permits to have a better starting performance to the driver thus potentially allowing him to overcome as many opponents as he can or gaining a safe time gap to the follower in the early stages of the race. After the employment of the soft compound, the employment of either the hard or medium compounds is used for longer stints, depending on weather and track conditions.

If the team chooses for a "conservative" strategy, the first compound to be use at race start, will be the medium one, allowing them to have less starting performance, but to develop more pace and speed during each passing lap, having the possibility to run a longer stint than those on the softer compound. The most conservative strategy is that of employing hard compound at the starting of the race. This strategy his highly risky as heating up this compound from a standstill is always very difficult, implying potential position loss during the earliest stages of the race, but allowing the driver to have clean air when his opponents will perform the pit, which in every case listed above, will happen consistently earlier than a pit-stop on hard tyres will be. This strategy can be adopted to favour the usage of the softer compound in the final stages of the race thus allowing the driver to be significantly faster than his opponents due to the coupling of less car weight, (caused by fuel consummation), and fresher and faster tyres.

Having said all this, deciding in advance the possible optimal strategies, depending on various race occurrences, is very important in order to have viable alternatives in cases that are deemed most likely and having in return a clear vision for the race trend, not only for the team's drivers but also for all other opponents. A further important thing to take into consideration is the fact that, when making strategies, teams have to take into consideration that their drivers could make mistakes, thus triggering the deployment of either yellow flags or virtual safety car or safety car in case of a mild accident, but also a red flag in case of a severe accident.

The deployment of yellow flags, like in qualifying, is triggered when a driver makes a mild mistakes such as going wide in a turn and exiting the track for a limited period of time for then returning into the track with no problem after the error. When yellow flags are deployed no overtaking is allowed and a decrease of speed of all racers is mandatory in order to not cause further problems.

Rather, the usage of virtual safety car is employed when a driver makes a mistake and is not able to come back on track due to any damages occurred to the car. Thus if the damaged car is in a place with easy access for the stewards to safely an timely remove the car, in order to prevent any further accident, all drivers have to decrease speed up to 40% of race speed and maintain the positions (no overtaking manoeuvres are allowed) in which they were when the virtual safety car phase was deployed. This phase will end at the lap in which the car will be removed. During this phase it is advantageous to make a pit-stop as the time-loss for entering the pit is reduced to 30% of the usual time-loss when in normal conditions, thus allowing the driver to lose less time and be in a favourable position with fresh tyres when returning on track.

Then, when the damaged car is not easily accessible for its removal, the safety car phase will be deployed. In this situation a physical car will be leading the race in front of the actual leader, decreasing the speed of about 65% of usual race speed, allowing those who perform the pit stop to lose only half the time with respect to the time loss occurring when in normal race conditions, entailing an even greater advantage with respect to the virtual safety car phase. The safety car phase will end during the laps following the safe removal of the damaged car.

These two phases allows drivers to preserve the tyres, thus prolonging their lives, due to the fact that they will not be able to push to the limit thus saving the tires and potentially allowing them longer stints.

The last eventuality concerns the exposure of the red flag. In this case the race is immediately stopped, due to a sever incident that can not be dealt with by only deploying the safety car. This possibility has various outcomes on race ending and strategies, as a matter of fact, the suspension of race, does not prevent the time to flow, due to regulatory demands races have to be terminated within a temporary timeframe which is decided by the race commission at the starting of the year. Thus a suspension for an extended period of time could prevent the resuming of the race till the completion of all laps instead going by the time distance to the maximum allowed timing for a race.[2]

2.1 Theoretical Background

In order to take into consideration all these different aspects, simulate and compare the effects of various race strategies, race simulations are used. The most widely employed methodology is that of discretizing the race lap-wise, entailing that the first steps is that of calculating the expected lap times t_{lap} of each driver for each race lap l. in order to obtain these timings, the calculation is presented in Equation 21[10].

$$t_{lap}(l) = t_{base} + t_{tire}(a_{tire}, c_{tire}) + t_{fuel}(l) + t_{car} + t_{driver} + t_{grid}(l, p_g) + t_{pit,in\text{-}lap/out\text{-}lap}(l) \qquad (21)$$

Where:

- t_{base} is the lap time that the fastest car-driver combination can achieve in the race in conditions similar to the qualifying ones, where the fuel level is at minimum and the tyres are fresh. This in turn, takes into consideration the characteristics of the track.

- t_{tire} is added to the former time to account for the effect of tire degradation. This value is dependent on the age of the tire and its compound, respectively (a_{tire}, c_{tire}).
- In addition to these measurements, t_{fuel} is added as a measurement that takes into consideration the time that is lost due to the carried fuel mass.
- t_{car} and t_{driver} for the car and driver abilities are added, as these are significant for the fact that there are differences among cars. As there is no predetermined designed to follow, sometimes, such as this years, the design of cares can differ greatly from a team to another thus creating slight to substantial differences in performance. Moreover, not all drivers are equal, and even competing in the same car there could be a driver whom adapts better to that car because it is in line with his guiding style thus making for differenced in abilities of the drivers.
- t_{grid} is added to account for the time that is lost at the race start (dependent on grid position p_g)
- at last $t_{pit,in\text{-}lap/out\text{-}lap}$ are considered for the time that is lost in pit stops.[10]

Ensuing from this calculation, consecutive lap times of a driver are then summed up to obtain his race times t_{race} at the end of every lap is given by Equation 22.

$$t_{race}(l) = \sum_{i=1} t_{lap}(l) \tag{22}$$

This measurement is the focal point of the simulation which will be performed to simulate the race trends.

The lap-wise discretization enables the strategy engineer to quickly compare the results of many different strategies before and during a race, even in respect of including probabilistic events (accidents, failures, deployment of safety car etc) that could ensure a stabler race strategy in terms of robustness against unforeseen events. [10] Monte Carlo Simulation is employed "*to study properties of systems with components behaving in a random fashion*", this is applied by the implementation of realistic models for probabilistic effects then simulating races to determine the estimated distribution of rank position which is the variable of interest of the strategy engineer.[10]

The first race aspect to take into consideration, is the drivers starting performance, which indeed, as the name suggests, it refers to each driver's performance at race start. Drivers can either be good starters or bad starters, thus to be able to incorporate this aspect to the simulation, an average starter performance in calculate by means of Equation 23.

$$t_s = \frac{\sqrt{2(p_g p_s)}}{a_{avg}} + t_r \tag{23}$$

where:

- the parameters p_s and t_r stand for the (virtual) position of the start line and the reaction time of a human driver. They shift the origin so that a driver who would start directly on the start line would only have to overcome his reaction time. p_s is set 0.8 because the start line is located only slightly in front of the pole position. Therefore, the distance to the pole starter is significantly smaller than the usual distance between two grid positions. For the reaction time tr we use 0.2 s. [10]
- The average acceleration during the race start a_{avg} is set equal to 11.2 ms^{-2} (this is the evaluation for 2019 start speed, as the data regarding this year were not available this value was inputted to the function). [10]

Having this as a reference curve, then the difference from the real performance of each racer to the respective data point were calculated, these deviations are then used to compute the mean and standard deviation of a driver-specific Gauss distribution used to model the starting performance. [10] Thanks to the data of each race of 2022 the driver-specific starting performance was calculated by subtracting to the time at the end of the lap the starting time in session, the actual time set by the driver calculated from starting line to end of the lap.

Another pivotal aspect to be taken into consideration is the fact that none of the drivers can exactly repeat the lap time set at the previous lap, as a matter of fact the lap times during the race are all scattered around the mean value. Therefore following Heilmer work [10] quadratic polynomials, of Equation 24, are fitted to the real lap times t_{lap} of each stint excluding noise laps (these laps are the first two laps of the race, laps in which a grave mistake by the driver was made and laps including a passage in-and-out of pitlane) as these could affect the distribution in a mostly wrong fashion.

$$t_{lap,poly}(l) = k_2\, l_2 + k_1\, l + k_0 \tag{24}$$

Then the deviations of the laps from those calculated by the polynomials were shown to be approximately normally distributed. Thus, to model the lap time variability, a Gaussian distribution with mean zero a driver-specific standard deviation was used[10]. Moreover, variations in pit-stops timing were modelled by the means of a Fisk distribution with parameters shape, location and scale.

In the event that the driver is not able to finish the race, the cause could be either an accident caused by the same driver (such as a collision with another driver, a heavy contact with protective barriers etc) or by a technical failure of the car. These events have to be considered during strategy compiling as they can determine either yellow flags, Safety Car or Virtual Safety Car deployments and in the most severe cases a red flag.

In order to include these to the modelling of race simulation, a driver and car probability of accident and failures, respectively, was calculated by means of, Bayesian inference in which the prior distribution was set as Beta distribution and the Bernoulli distribution was employed as likelihood function. The parameters of the Beta distribution were calculated as the Equation 25 and Equation 26. [10]

$$\hat{\alpha} = \left(\frac{1-\hat{\mu}}{\hat{\sigma}^2} - \frac{1}{\hat{\mu}} \right) \hat{\mu}^2 \tag{25}$$

$$\hat{\beta} = \hat{\alpha} \left(\frac{1}{\hat{\mu}} - 1 \right) \tag{26}$$

$\hat{\mu}$ and $\hat{\sigma}$ stand for sample mean and standard deviation of the prior distribution which are determined by using the total accident fraction per driver, and the total failure fraction per team. Season-specific driver- and team- posterior distributions are calculated taking into account the corresponding accident and failure fractions within the particular season.

As stated in the previous section, the deployment of yellow flags or Safety Car (SC) or Virtual Safety Car (VSC) are a significant variability to be included while modelling the race simulation thus, whilst the effect of yellow flags is negligible when taking into consideration the lap time, the effect of the other two alternatives is going to heavily impact the race lap timing (by 1.6*$t_{lap,min}$ and by 1.4*$t_{lap,min}$ respectively), making the gap between the drivers either completely disappear or reduce gravelly.

The starting of either a VSC or a SC is simulated race-time-wise (in terms of seconds passed for the start of the race to the deployment of either phase) this is due to the fact that drivers could find themselves in different lap numbers in that specific period of time (in each race there are cars which are lapped thus making them being behind a lap from race leaders). [10]

In order to have a robust simulation, following Heilmer et al. considerations[10], the these phases and retirements are determined before starting the simulation. Due to the fact that, by observing past data, the VSC is usually deployed after a car failure as the driver could be able to bring the car to an exit point easily accessible by the stewards to remove the car. In turn, the SC is, in most cases, deployed when accidents happen. Hence, firstly SC phases are determined and accidents are derived, then failures are determined and VSC phases are derived. At last these measurements are converted in race progress timing[10].

The SC phases are fixed at first because they have a significant impact on race strategy, and therefore their probability of occurrence should be on conditional probability. The quantity of SC phases for a race is chosen between zero and three, whereby empirical probabilities $P_{sc,quant}$ is calculated according to 2019-2022 real data. Then the starting of each phase in determined (6 sections with their individual probabilities are defined: first lap, ≤ 20%, ≤ 40%, ≤ 60%, ≤ 80%, ≤ 100%) and their duration is determined between 2 to 8 laps[10].

The start of an SC phase is further modified by a uniform distribution *U(0, 1)* to include the fact that it does not start precisely at the point laps are completed. The last control is performed to prevent SC phases to overlap during the simulation. Due to the fact that each SC is determined by an accident the simulation chooses the driver that retires based on their specific accident probabilities [10].

Then, for all the drivers that remain in race, the simulation determines which of them will suffer a failure based on the probabilities calculated in a season-specific setting. By assuming that for each failure a VSC will be deployed (this is a simplistic assumption as it could be employed also for mild accidents), the related probability will be calculated by dividing the number of VSC phases in the period over the number of failure in the period 2019-2022. . The start of the failure is sampled from a uniform distribution $U(0, n_{laps})$, the VSC duration is chosen in the range between one and four laps, with empirical probabilities $P_{vsc,duration}$, and modified by a uniform distribution $U(0, 1)$ as with the start of SC phases. After these determinations, the race progress is converted from lap-wise to time-wise[10].

Furthermore, by using the Virtual Strategy Engineer (VSE) developed by Heilmer et al [11], relationships between inputs (lap time, tire age etc.) and output (chosen race strategy) were modelled. The strategy decision in VSE is split in two parts: pit stop decision and the tyre compound choice. Hence, a first neural network will determine whether the driver should make a pit stop and if the pit-stop is deemed as a choice, the second neural network will determine which tyre compound should be fitted. In this way, the tyre choice will not influence the timing of the pit-stop[11].

The VSE can be used in the simulation to determine the timing of intended pit-stops and the tyre compound choice at each stop. By adding these two neural networks to the race simulation, two aspects have to be taken into consideration: determining all possible combinations of tyre compound for a chosen number of pit-stops and determining the optimal stint lengths for all compounds involved.

Thus, in order to determine both aspects, Heilmer et al [11], proposes to opting out all unforeseen events and drivers' interactions to maintain a somewhat pure measurement, not affected by external factors. Indeed, for determining all possible tyres combinations, giving that all drivers have to use two different compounds per race, and assuming that tyre degradation is independent of changing track conditions and fuel mass (which accounts for 13% of the total car mass thus not influencing the result in a significant way), in regards of the last aspect, as the fuel mass decreases the acceleration increases maintaining the tyre degradation similar, for each driver, 22 possibilities are available.

Considering the second objective, by definition, optimum stint lengths are those with the minimum race duration with respect to the set, assuming the same elements opting out as in the above-mentioned aspect, the race duration of a given set is dependent only on the time loss affected by tyre degradation, therefore it is dependent on the stint length. A mixed-integer quadratic optimization is solved to determine the optimal stint lengths, this was the model chosen due to it being the most robust model. The objective function is found in Equation 27 [11].

$$min\, t_{race}(l_{tot}) \mathrel{\widehat{=}} \min \sum_{l=1}^{l_{tot}} t_{tire}(l) \tag{27}$$

With l_{tot} being equal to the total lap number of the race. Then, the race is split into N fixed stints with index i for each stint, the optimization variables are: c_i (compounds used in stint) and α_i (the stint duration), thus formulating the model as Equation 28.

$$\begin{gathered} \underset{[\alpha_1 \ldots \alpha_N]}{min} \sum_{i=1}^{N} \sum_{\alpha=1}^{\alpha_i} t_{tire}(a, c_i) \\ \text{subject to} \sum_{i=1}^{N} \alpha_i = l_{tot} \quad \alpha_i \in \mathbb{N}^{+} \;\; \forall\, 1 \le i \le N \end{gathered} \tag{28}$$

As the time of a compound is dependent on its age at the start of the stint and the overall performance of the compound, which differs for each compound specification following a linear tyre degradation model, the model is then modified as Equation 29.

$$\sum_{i=1}^{N} \sum_{a=1}^{\alpha_i} t_{tire}\,(a, c_i) = \sum_{i=1}^{N} \left(k_0(c_i) \cdot \alpha_i + \; k_1(c_i) \sum_{a=1}^{\alpha_i} a \right) \tag{29}$$

Then the Gaussian sum is used to rewrite the coefficients of the linear degradation model. Moreover, it is important to consider the fact that the tyres may not be fresh at the start of the stint, especially when taking into consideration the soft compound which could have been used during the qualifying session, a starting age $a_{s,i}$ is added thus resulting in Equation 30.

$$\sum_{i=1}^{N}\sum_{a=1}^{\alpha_i} t_{tire}\left(a, c_i, a_{s,i}\right) = \sum_{i=1}^{N}\left(k_{0,i}\left(\alpha_i + a_{s,i}\right) + k_{1,i}\left(\frac{1}{2}\left(\alpha_i + a_{s,i}\right)^2 + \frac{1}{2}\left(\alpha_i + a_{s,i}\right)^2\right)\right) - \left(k_{0,i} \cdot a_{s,i} + k_{1,i}\left(\frac{1}{2}a_{s,i}^2 + \frac{1}{2}a_{s,i}\right)\right) \qquad (30)$$

Therefore, having proposed these modelling instances, in the next section the data used to obtain the results required is exposed.

2.2 Data Description and Preparation

The data used to solve the models proposed in the above section, are collected for the `FastF1` Python package, Ergast API platform and considering the track parameters, information was given by either scraping the FIA Formula One or Formula 1 web pages.

Thus, considering the calculation of the starting performance of each driver, the Equation 31 was applied for those races that happened prior to 26/09/2022.

$$\textit{Driver starting performance} = [(\textit{Session time at the end of first lap (s)} - \textit{session time at the beginning of first lap(s)}) - \textit{lap time(s)}] \qquad (31)$$

Then for the races that have yet to be performed a mean of the previous races' results was taken resulting in Table 16.

Table 16 *Average Starting performance in milliseconds, period 2019-2022*

Driver	Average Starting performance (ms)
Albon A.	260
Alonso F.	242
Bottas V.	234

Hamilton L.	225
Gasly P.	232
Leclerc C.	250
Latifi N.	220
Magnussen K.	271
Hulkenberg N.	264
Russell G.	222
Verstappen M.	229
Norris L.	219
Zhou G.	267
Vettel S.	237
Stroll L.	240
Ocon E.	247
Perez S.	219
Sainz C.	239
Schumacher M.	240
Tsunoda Y.	220

Moreover, the team-specific parametrization for failure probability $P_{failure}$ and pit stop duration variability $t_{pit,var}$ for all teams of the 2022 season is that of Table 17.

Table 17 *Probabilities of teams for the occurrence of failure and the Fisk parameters for the pit-stop time variability, period 2019-2022*

Team	$P_{failure}$ (2022)	$t_{pit,var}$ in s (F (shape, loc, scale))
AlfaRomeo	0.242	[2.631, 0.843, 0.563]
Ferrari	0.165	[2.076, -0.152, 0.944]
HaasF1Team	0.19	[8.613, -0.716, 2.413]
McLaren	0.088	[3.638, -0.342, 1.097]
Mercedes	0.062	[0.319, 0.336, 0.352]
AstonMartin	0.088	[4.759, -0.260, 1.522]
RedBull	0.165	[1.612, 0.074, 0.334]
Alpine	0.113	[2.849, 0.291, 1.651]
AlphaTauri	0.19	[3.848, -0.473, 1.362]
Williams	0.113	[5.785, 0.255, 1.861]

Then the driver-specific parametrization for accident probability $P_{accident}$ for all drivers in 2022 season were calculated, resulting in Table 18.

Table 18 *Probabilities of drivers to make an accident during the race, period 2019-2022*

Driver	Paccident (2022)
Yuki Tsunoda	0.054
Alexander Albon	0.076
Carlos Sainz Jnr	0.076
Charles Leclerc	0.054
Daniel Ricciardo	0.033
Kevin Magnussen	0.054
George Russell	0.054
Mick Schumacher	0.054
Guanyu Zhou	0.054
Lance Stroll	0.033
Lando Norris	0.054
Lewis Hamilton	0.054
Max Verstappen	0.033
Esteban Ocon	0.033
Pierre Gasly	0.054
Nicholas Latifi	0.076
Fernando Alonso	0.054
Sebastian Vettel	0.033
Sergio Perez	0.054
Valtteri Bottas	0.054
Nico Hulkenberg	0.033

In order to determine the abilities/capabilities of the car and driver which are dependent on the season that is taken into consideration, for the year 2022, with respect to the calculation of the abilities of the driver is done by averaging the difference in seconds from the winner of the race with respect to the driver taken into consideration. Then, for calculating the abilities in terms of lap times in races of the car, the average of the lap time per team, averaged to the difference with respect with the fastest team in each race.

2.3 Results

The races that have been taken into consideration for the analysis, have been those already performed. The model proposed in the theoretical section was applied to make a comparison between the strategy adopted in race by the top 3 finishers and the optimal race calculated by the model.

Thus, in Table 19 the results for the races that have already taken place are presented. It has to be highlighted the fact that, maintaining the disruptures brought forth by the mechanical failures that heavily impacted the results of races like Bahrain Grand Prix, Azerbaijan Grand Prix, Australian Grand Prix and Austrian Grand Prix in which, had the cars with failures, been able to terminate the race the results would have been substantially different. [3]

Moreover, as it can be seen in Table 19, the result of 3 Grand Prixs: British Grand Prix, French Grand Prix and Dutch Grand Prix; would have been different if the strategy chosen would have been the one suggested by this model.

These are prime examples of the fact that, by choosing a determined strategy, the outcome of the race could vary significantly, impacting the championship differently than what occurred in reality. Abiding to optimal strategies in F1 is as determinant as having a fast car-and-driver pairing. As a matter of fact, even with the fastest car, given that the level of competition is very high, it will not automatically imply winning the races as it is the combination of the adoption of a good strategy and fast car-and-driver pairing that results in winning.

Strategy is, indeed, a clear determinant of the final results in both Qualifying sessions and Race. Ensuing from the analyses of Chapter 1 and 2, it is clear how the data gathered during the Free Practices, as much as, past data, are essential into formulate viable strategies for a number of possible scenarios, in order to have a clearer view of what implications changing the course of exit-timings, pit-stops and tyres' compounds will ultimately change the final result of the entire race weekend. Thus, the formulation of

[3] The races analysed in this chapter are those occurred in complete dry conditions, due to the less variability that affects the wet or mixed weather races. The Spain Grand Prix has not been taken into consideration due to the lack of accurate data for some of the drivers involved in top positions.

optimal race strategies is a turning point in determining who wins and who loses races and subsequently the championship.

Table 19 *Results of Strategy Simulation for 2022 dry races.*[4]

Race	Driver	Final Position (Real)	Startegy adopted (Real)	Final Position (Simulation)	Strategy adopted (Simulation)
Bahrain Grand Prix	LEC	1	Soft (15) - Soft (16) - Medium (15) - Soft (11)	1	Soft (15) - Soft (16) - Medium (15) - Soft (11)
	SAI	2	Soft (14) - Soft (19) - Medium (11) - Soft (13)	2	Soft (15) - Soft (17) - Medium (12) - Soft (13)
	HAM	3	Soft (14) - Hard (16) - Medium (17) - Soft(13)	3	Soft (15) - Hard (14) - Medium (18) - Soft(13)
Saudi Arabian Grand Prix	VER	1	Soft (16) - Medium (34)	1	Soft (16) - Medium (34)
	LEC	2	Soft (16) - Medium (34)	2	Soft (15) - Medium (35)
	SAI	3	Soft (16) - Medium (34)	3	Soft (17) - Medium (33)
Australian Grand Prix	LEC	1	Medium(22) - Hard (36)	1	Medium(22) - Hard (36)
	PER	2	Medium(20) - Hard (38)	2	Medium(24) - Hard (34)
	RUS	3	Medium(23) - Hard (35)	3	Medium(22) - Hard (36)
Miami Grand Prix	VER	1	Medium (26) - Hard (31)	1	Medium (26) - Hard (31)
	LEC	2	Medium (24) - Hard (33)	2	Medium (24) - Hard (33)
	SAI	3	Medium (27) - Hard (30)	3	Medium (25) - Hard (32)
Azerbaijan Grand Prix	VER	1	Soft (18) - Medium (15) - Medium (18)	1	Soft (15) - Medium (18) - Medium (18)
	PER	2	Soft (16) - Medium (17) - Medium (18)	2	Soft (15) - Medium (19) - Medium (17)
	RUS	3	Soft (9) - Medium (24) - Medium (18)	3	Soft (12) - Medium (24) - Medium (15)
Canada Grand Prix	VER	1	Soft (9) - Medium (34) - Medium(21)	1	Soft (12) - Medium (34) - Medium(28)
	SAI	2	Soft (20) - Medium (29) - Medium(21)	2	Soft (22) - Medium (27) - Medium(21)
	HAM	3	Soft (10) - Medium (33) - Medium(27)	3	Soft (9) - Medium (35) - Medium(26)
British Grand Prix	SAI	1	Medium (20) - Hard (19) - Soft (13)	2	Medium (20) - Hard (19) - Soft (13)
	PER	2	Medium (5) - Hard (34) - Soft (13)	3	Medium (15) - Hard (24) - Soft (13)
	HAM	3	Medium (33) - Hard (6) - Soft (13)	4	Medium (29) - Hard (10) - Soft (13)
	LEC	4	Medium (25) - Hard (27)	1	Medium (25) - Hard (19) - Soft (13)
France Grand Prix	VER	1	Medium (16) - Soft (37)	1	Medium (16) - Soft (37)
	HAM	2	Medium (18) - Soft (35)	2	Medium (18) - Soft (35)
	RUS	3	Medium (18) - Soft (35)	4	Medium (18) - Soft (35)
	PER	4	Medium (18) - Soft (35)	3	Medium (16) - Soft (37)
Hungary Grand Prix	VER	1	Soft (16)- Medium (22) - Medium (32)	1	Soft (16)- Medium (22) - Medium (32)
	HAM	2	Medium (19)- Medium (32) - Soft (19)	2	Soft (16)- Medium (32) - Medium (19)

[4] In the columns "Strategy adopted" either real or simulated, the numbers between brackets represent the number of laps driven on that tyre compound.

	RUS	3	Soft (16)- Medium (23) - Medium (31)	3	Soft (16)- Medium (25) - Medium (29)
Belgium Grand Prix	VER	1	Soft (15) - Medium (15) - Medium (14)	1	Soft (15) - Medium (15) - Medium (14)
	PER	2	Medium (14) - Medium (13) - Hard (17)	2	Medium (13) - Medium (12) - Hard (19)
	SAI	3	Soft (11) - Medium (14) - Hard (19)	3	Soft (12) - Medium (11) - Hard (20)
Dutch Grand Prix	VER	1	Soft (18) - Medium (30) - Hard (8) - Soft (16)	2	Soft (18) - Medium (30) - Hard (8) - Soft (16)
	RUS	2	Medium (31) - Hard (17) - Medium (9) - Soft (15)	3	Medium (31) - Hard (26) - Soft (15)
	LEC	3	Soft (17) - Medium (28) - Hard (11) - Soft (16)	4	Soft (17) - Medium (28) - Hard (11) - Soft (16)
	HAM	4	Medium (29) - Hard (19) - Medium (24)	1	Medium (29) - Hard (28) - Soft (15)
Italian Grand Prix	VER	1	Soft (25) - Medium (23) - Soft (5)	1	Soft (25) - Medium (23) - Soft (5)
	LEC	2	Soft (12) - Medium (21) - Soft (15) - Soft (5)	2	Soft (21) - Medium (27) - Soft (5)
	RUS	3	Soft (23) - Hard (24) - Soft (6)	3	Soft (23) - Hard (24) - Soft (6)

Chapter 3

The Formula One Championship, represents one of the biggest yearlong sporting event in the racing car world . Indeed, its total revenue of 2021 is estimated to be approximately 2.14 billion US$, ranking only behind Football World Cup and Olympic Games in terms of live television viewers [12].

Ensuing from the disastrous impact that the pandemic had on the 2020 season operating income, that amounted to a loss of $ 386 million, in 2021, the operating profit amounted to $92 million. This recovery allowed Liberty Media Group (the commercial rights holder of Formula One) to be able to increase the total prize money, allotted to the teams, up to $1.068 billion last year [12].

Indeed, teams' revenues is determined not only by sponsorship, but also by the above mentioned price money. Of the total season revenues, 47% of it is given back to the competing teams as follows:

- 23.7% of the total amount is divided in equal parts among the 10 teams;
- 23.7% of the total amount is divided according to the final constructors' standings;
- 2.5% of the total amount is given back to Ferrari by means of a contract that ties the team with the Formula One championship;
- the remaining amount is split among the historical teams.

Thus, given the revenue allocation, the importance of winning races and/or finishing in the first 10 cars rewarded with points, not only determines the constructors' championship winner but also the revenue allocation to the teams involved.

In the last twenty years, it has been observed a lack in competitive balance among the teams involved, as first Ferrari (from 1999 to 2004) then Red Bull (from 2010 to 2013) to finish with Mercedes (from 2014 to 2021) won an extensive number of constructors' titles. During the last eight years only Mercedes was able to win the constructors' championship, also winning 7 out of 8 drivers' championships, being defeated only last year by the Red Bull driver, Max Verstappen. [13]

Due to this uncompetitive balance and to the fact that, as proposed by Rottemberg (1956), the spectators derive great utility from observing contests with uncertain outcomes [14], the regulatory framework surrounding the sport has been amended numerous times, impacting the sporting event results, as much as the budget spent by teams in the yearlong seasons. As a matter of fact, ensuing from the grave impact that the pandemic had on teams, the regulating body of Formula One, FIA, decided to put a restraint on the budget available to each team in the season.

Starting from 2021, a progressive limiting of budget (called budget cap) has been sanctioned by FIA, in order to allow smaller independent teams to be able to compete in a more balanced way with the top three teams (Mercedes, Ferrari and RedBull) that have monopolized the championships in the last two decades. Thus, the rationale behind the enforcing of the budget cap has been that of trying to equalize the spending of the various teams, giving smaller teams the chance to fight for the positions that counts instead of only being spectators to the championship's battle concerning only those teams that had greater economic means to achieve better results.

Moreover, starting from 2022, the introduction of a technical regulation changed the car design and its components in a drastic way, with respect to the previous years. As a matter of fact, more limitations were applied on car design, its dimensions and the overall parts it is made of. Once again, the rationale under this change was that of optimizing car performance and fuel expenditure, together with restricting teams in the car composition as to allow for the title battle to be more choral rather than made up by one or two constructors.

Thus, in this chapter an analysis on the competitive balance of the seasons ranging from 2010 to 2022 will be performed. This was the selected period, given the fact that, the analysis will take into consideration not only the winning percentage but also the points' allocation across the season. Indeed, before 2010 the points awarded were different form the scheme that is implemented nowadays, preventing an analysis on an equal field for those seasons.

3.1 Theoretical Background

Competitive balance has been viewed as a major factor in sports economics as it is an important determinant of the demand for sporting events and thus of the sport's overall success. It reflects the outcome's uncertainty of professional sporting events, as it is a measure of "parity", of the overall balance of a sport's championship with respect to the competitive closeness of all participants in the contest.

Indeed, preserving the uncertainty of the outcome is deemed as a key determinant of a sport financial health, due to the fact that perennially unbalanced contests would eventually cause fan interest to wane and thus industry revenues to fall [15].

Thus, the concepts of competitive balance and uncertainty of outcome are the most relevant reasons as to why spectators decide to attend sporting events finding them entertaining and interesting. In sports economics, the hypothesis that the uncertainty of outcome influences the utility for consumers, driving the willingness to pay and attend sporting events, has played an important role since 1956, due to the fact that as a contest is more uncertain in its outcome, the more the demand of the spectator would arise thus increasing the revenues of the sporting events. The degree of the outcome uncertainty depends on the disparity between competitors the bigger the difference between strong and weak competitors the higher the competitive imbalance. [15]

Formula One is not considered a team sport, as its more of a hybrid team-individual contest, in which the driver (individual associated with success) depends on the rest of the team to optimize car performances through the setting and the strategy decisions involving when to perform the pit-stops and which type of tyre compound to fit. [15]

Moreover, differently to the other motor racing series, each team is responsible of for the design and manufacturing of their own cars, thus increasing the unbalance of the competition, as a consequence of having different cars, powered by different engines and driven by drivers with different abilities making the driver-car pairing a fundamental focal point of the championship title fight.[15]

From the study of Krauskopf, Langen and Bunger (2010) [16], the conclusion that, a too high level of competitive balance is undesirable in this highly unbalanced contest. Indeed, it is seen as desirable, a season-long fight among few top contenders rather than a more comprehensive fight among most of the driver-team pairings. This result is supported by Budzinski and Pawlowski (2014) giving evidence of the fact that it is sufficient for the championship to be balanced by a narrow oligopoly of contenders than among all competitors involved [17].

Delving deeper in the impact of regulatory changes on sport events' outcomes, a study by Mastromarco and Runkel (2009), investigates the relationship between regulation changes and the level of competitive balance in Formula One. Regulatory changes are brough forth and implemented before the starting of the season. Varying reasons, ranging from safety measures, reducing costs, increasing the relevance of the sport worldwide, balancing the racing results have been at the base of changes regarding technical, monetary and other aspects of the sport [18].

In this study, a two-stage model is proposed, where, in the first stage the FIA decides whether to change the rules. In doing this, it maximizes the broadcasting revenue (that is positively correlated to the fan interest). As the competitive balance is understood as the uncertainty of the winner of that specific F1 season, this determination is performed in the second stage of the model. In this stage, the teams with varying range of capabilities convert the effort in performance to compete for the prize money, thus teams expend effort to improve its performance, increasing, in turn, the probability of winning the championship [18].

The result of this research supports the theory that the expected number of rule changes at the beginning of a season is high if the competitive balance of the previous season has been low. In addition to this, the regulatory changes brought forth at the beginning of the season exert a significant positive impact on the competitive balance during the same season [18]. These aspects can be seen in the comparison between the 2020 and 2021 seasons, where for the first time a budget cap severely cutting the teams' expenditure was implemented, resulting a in tight fight for the drivers' title that culminated in the title assignment in the last race of the season.

When technical regulations changes the car design drastically, there might be the possibility for one or more teams to prevail upon the others, as it has been the case in 2010 with Red Bull dominance for the subsequent four years, or in the case of the turbohybrid era (seasons ranging from 2014 to 2021) which have seen the dominance of Mercedes, winning consistently the drivers' and constructors' titles with no real contented with the exception of the 2021 season in which Red Bull was able to win the drivers' championship.

In 2022, together with the budget cap regulation, a new technical regulation established the revolution of the car concept and design, thus allowing teams to develop ex-novo their cars, giving a possibility at levelling the field by potentially increasing the competitive balance. As the season it is still ongoing, it is yet to be determined if the regulatory changes have had the impact that was hoped upon, but a certainty has emerged, once again we have the top three teams fighting to win races in a similar fashion to the previous seasons. The only thing that seems to have changed is the strength balance of these teams, resulting in the RedBull being consistently stronger than both Ferrari and Mercedes.

3.2 Data Presentation and Results

In order to perform an analysis on the competitive balance during the last decade, the following indexes have been taken into consideration to measure the competitive distance of teams and drivers in a within-season perspective that regards the outcome uncertainty of the evolution of the points standings.

At first, within-season competitive balance was calculated by means of the standard deviation of the winning percentage driver-wise and constructor-wise. As Formula One championship is comprised of two championships, one regarding the drivers and the other regarding the constructors, it is important to take into consideration both aspects on title fight. Indeed, even with the dominance of a team over all others, there could be a tight fight for the drivers' championship title among the drivers of the same team, thus making the competition among them an extremely balanced one, as they are driving in the same car.

Moreover, as consistency is key to win championships, there could occur that a constructor is in the championship fight due to the combined ability of both drivers to optimize the results as a team thus bringing valuable points to the team.

The most common form of quantifying the competitive balance is through the calculation of the standard deviation (Equation 33) of the winning percentages of both drivers' and teams' championships of the season[16].

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(winperc_i - averagewin)^2}{n}} \tag{33}$$

Where the $winperc_i$ is the winning percentage of team *i*, the *averagewin* the average winning percentage across season and *n* is the number of races in the season. The larger the standard deviation, the less the competitive balance is in the season.

This measure was also applied to the season points' scoring. Winning races is of course, important, but it has been clear that the key point to win a championship is also consistency across the season, this implies that when winning is not possible the driver and the team have to optimize the result in order to lose, with respect to the winning-awarding points, the less points possible. Thus, the competitive balance was also quantified in terms of points gained though the season, Equation 34.

$$\sqrt{\frac{\sum_{i=1}^{n}(points_i - maxpoints)^2}{n}} \tag{34}$$

Where $points_i$ is the points scored by the driver, *maxpoints* is the maximum achievable points in season and n is the number of races.

Then, the Herfindahl-Hirschman Index (HHI) was used to calculate the concentration of wins across teams. This index is mainly used to characterize the distribution of a variable of interest by measuring its degree of concentration across units, such as firms, teams or households.[19]

In measuring competitive balance, the HHI has been applied to the distribution of wins across teams in a particular season. Depken (1999) interprets a team's share as the number of wins by the team in a season as a proportion of total wins in the competition[19], Equation 35.

$$HHI = \sum_{i=1}^{n} (w_i \ / \sum_{i=1}^{n} w_i)^2 \qquad (35)$$

Where n equals the number of teams in the league and w_i is the number of wins for team i. As for the standard deviation measure, an increase in HHI represents a decrease in the season competitive balance.

Moreover, a further specification of HHI was used as indicated in Equation 36.

$$HHI = \sum_{i=1}^{n} (p_i \ / \sum_{i=1}^{n} p_t)^2 \qquad (35)$$

Where p_i are the points scored by driver i of team t and p_t are the points scored by team t over the season.

To calculate the mentioned indexes, data regarding the seasons from 2010 to the current one were taken into consideration in the form of drivers' placements across seasons and overall teams' placements across season. It can be seen that, taking into consideration HHI, there has been an increase in the competitive imbalance reaching its maximum in years 2013 and 2020 where the winner and his first contented, both in terms of season winning percentage and in terms of points scores, observed a clear difference of more than 100 points and more than the double of winning percentage. Whilst taking into consideration the standard deviation, the trend is not as clear but it loosely follows the HHI results. This could be due to the fact that considering the winning percentage is not as good as a fit as HHI for this type of data.

Whilst in years 2012 and 2017 we have a somewhat more balanced competition among the first two contenders. As far as this year as come, it is clear that the competitive imbalance will reach another minimum point at the end of the season. These results are reported in Table 20.

Moreover, if teams are taken into consideration, the year in which the competitive balance has reached its lowest point is 2016. This year has been characterized by the dominance of a team among all others, that was able to win, with its drivers, 19 out of 21 races. Meanwhile, the most balanced seasons in terms of competitiveness have been those of 2010 and 2012, where, the top-two-contenders' wins were more evened out than the following years.

It can be observed that the impact of the regulatory changes in 2021 was somewhat positive, reverting the trend in the current season as there is a clear difference in terms of both winning percentages and points achieved. In particular if we observe the first 2 positions where, there is the widest difference since 2010. Moreover, the fact that only 2 teams were available to achieve a win is detrimental for the sport's overall competitive balance.

Thus whilst in 2021 a clear positive effect on the increase of competitive balance could be observed, in 2022 the effect is negative, increasing the difference among the teams participating in the championships. As a matter of fact, the gap between the first three teams and the rest of the grid increased sensibly in this year, despite the purpose of the budget cap regulations, that were put in place to level the field of car-driver performances.

Table 20 *HHI and SD measures considering Drivers' and Teams' Championship*

YEAR	HHI (drivers)	SD (drivers)	HHI (teams)	SD (teams)
2010	2188.37	0.085	3628.8	0.119
2011	3905.82	0.138	5013.9	0.146
2012	1650	0.165	2425	0.090
2013	4958.45	0.155	5069.3	0.147
2014	4293.63	0.143	7340.7	0.183
2015	4016.62	0.137	7340.7	0.183
2016	4149.66	0.133	8231.3	0.350
2017	3000	0.113	4930.7	0.131
2018	3514.74	0.159	4325	0.120
2019	3424.04	0.118	6075	0.148
2020	4532.87	0.156	6055	0.171
2021	3471.07	0.117	3843	0.127
2022	5866.67	0.156	6088	0.180

Conclusions

As one of the most competitive and highly performing championship in the world, Formula One is a complex "machine" composed of regulations, races, different teams and drivers with a penchant for entertaining a close-quarter battle for the championships titles. Year after year, 10 teams and their 20 drivers fight to win, first and foremost, the championships (both the drivers' and teams' ones) and secondly, to assert their being the fastest car-driver combination alive.

In order to better understand the way through which certain results are attained by certain teams, analyses were performed on the entirety of the race weekend which is comprised of three free practices, a qualifying session and a race (when considering the traditional format). By starting with optimizing the allocation of tyres, going through simulating accidents and flag exposures, ending with predicting the qualifying classification and hence compiling the better exit strategies, the part preceding the race was examined.

Due to the extreme importance that the starting grid position has on the end result of the race, each team has to achieve the better position possible for each of their drivers, in order to have an increased chance to make a good race and not be involved in the starting battles that are mostly found in the midfield of the grid.

Through the usage of Monte Carlo simulation on the probability of yellow or red flags to be exposed, a most likely scenario on where these events would most likely happen was detected. This is vital to predict in advance, as it is important for exit strategies during the three sessions of qualifying, hence having relevance for the final result. Then the drivers whom were most likely to be the cause of those flags' exposures were identified through the use of data gathered from 2019 to 2021.

Moreover, in order to have a realistic qualifying simulation, the qualifying timings for each drivers were predicted based on data accumulated during the free practices (FP2 and FP3). Ensuing from this prediction and past data on pit-exit and cool-down lap timing, two strategies for each driver were proposed, according to the number of tyres the team choose to use and to the general velocity of the driver-car pairing. Moreover, the predicted Q1 and Q2 cut-offs were predicted, also for the races that have yet to be performed.

After having ascertained the starting grid, a race simulation was performed to propose optimal race strategies in terms of number and timing of pit-stops and type of compound to use and when. This simulation was performed on past data of the current season's races, to understand the impact of the adoption of optimal strategies on final race results and therefore championships' standings.

As it was expected, the adoption of optimal strategies, in some cases in which the cars battling for the win were extremely balanced, produced a final podium different than that that actually occurred. This is consistent with the literature. In the measure in which, given the high volatility faced by teams during races, the difference between choosing an optimal strategy rather than a suboptimal strategy could determine the final result in a way such that, not always the fastest car wins, especially when the competitors are close by in terms of overall performance.

The determination of the grid starting positions in qualifying, coupled with a possible uncertainty of the race outcome, is one of many aspects, together with the spectacularity of the events as a whole, that attracts millions of spectators to the F1 circus. The last part of this writing was dedicated to the competitive balance aspect of the sport.

Competitive balance is a measure that determines how much far apart are the competitors in a contest by means of past and current results. This is deemed to be one of the many important aspects of the F1 Championship, as it can be seen by data analysed, there is some fluctuation from a more competitive year to a less competitive balanced year. The issue of competition is of vital importance to attract spectators. For this reason, many regulatory changes have been performed by FIA during the last two decades, with the aim to level the playing field of all teams, allowing competitive balance to increase.

Indeed, the impacts of the most stringent regulations on the budget used by each team during the season were taken into consideration to see the actual impact that these measures had on the championships' races on years 2021 and 2022. While in 2021 an increase of competitive balance could be observed, the effect that this regulation, coupled with a complete redesign of the car have had on this year's season, is that of considerably decreasing the competitive balance of the sport as a whole, given that, in the first 15 races of the year, only 2 teams were able to score at least a win. Of these two teams one driver has outperformed the other in terms of winning races, thus tracing a clear line between the top teams and all the other that were not able to keep up with the increasing performance showed through this period.

Therefore, while investigating the purely sporting part of qualifying and racing, the vital importance of correct strategy formulation was highlighted and proved. In regards of the level of competitive balance of the sport as a whole, F1 has still important steps to take in order to allow more competition into the races, hence more uncertainty which is, for the most part, what attracts people to the sport.

Bibliography and Sitography

[1] Fédération Internationale de l'Automobile, "2022 FIA Formula One Championship", [Online]. Available : https://www.fia.com/events/fia-formula-one-world-championship/season-2022/2022-fia-formula-one-world-championship. [Accessed on Feb. 3, 2022].

[2] Fédération Internationale de l'Automobile, "2022 Formula One Sporting Regulations", Issue 8, July 19th 2022 , [Online]. Available: https://www.fia.com/sites/default/files/fia_2022_formula_1_sporting_regulations_-_issue_7_-_2022-06-29.pdf [Accessed on July 30, 2022].

[3] E. Gupta, "How Do F1 Tyres Work?". July 29, 2020. Medium.com. [Online]. https://medium.com/projectf1/how-do-f1-tyres-work-ac19ced7ff61. [Accessed on May 5, 2022].

[4] E. Burns, "Machine Learning". Techtarget.com. [Online]. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML. [Accessed on June 5, 2022].

[5] B. Park Woolf, *Building Intelligent Interactive Tutors*. Morgan Kaufmann, 2009, pp. 403-450. Doi: https://doi.org/10.1016/B978-0-12-373594-2.00019-8.

[6] "Introduction to Boosted Trees". xgboost.io [Online]. https://xgboost.readthedocs.io/en/stable/tutorials/model.html. [Accessed on June 5, 2022].

[7] M. Schott, "Random Forest Algorithm for Machine Learning". April 25, 2019. Medium.com. [Online]. https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb. [Accessed on June 5, 2022].

[8] G. Ershad, "Logistic Regression with Mathematics", Medium.com. [Online]. https://pub.towardsai.net/logistic-regression-with-mathematics-ac69bafdc72 .

[Accessed on June 5, 2022].

[9] Red Bull Racing, “Bulls’ Guide To: Free Practice”. [Online]. https://www.redbullracing.com/int-en/bulls-guide-to-free-practice [Accessed on March 5, 2022].

[10] A. Heilmeier, M. Graf, J. Betz, M. Lienkamp, “Application of Monte Carlo Methods to Consider Probabilistic Effects in a Race Simulation for Circuit Motorsport”. 2020. *Applied Sciences*. [Online]. Doi: https://doi.org/10.3390/app10124229.

[11] A. Heilmeier, A. Thomaser , M. Graf, J. Betz, “Virtual Strategy Engineer: Using Artificial Neural Networks for Making Race Strategy Decisions in Circuit Motorsport”. 2020. *Applied Sciences*. [Online]. Doi: https://doi.org/10.3390/app10217805.

[12] Liberty Media Group, “Lbierty Media Corporation Reports Fourth Quarter and year End 2021 Financial Results”. [Online]. Available: https://www.libertymedia.com/investors/news-events/press-releases/detail/461/liberty-media-corporation-reports-fourth-quarter-and-year.

[13] O. Budzinski, A. Feddersen, “Measuring Competitive Balance in Formula One Racing”. 2019. [Online]. Doi: http://dx.doi.org/10.2139/ssrn.3357687.

[14] S. Rottemberg, “The baseball player’s labour market”. 1956. *Journal of Political Economy*, 64(3), pp. 242-258.

[15] C. Judde, R. Booth, R. Brooks, “Second Place Is First of the Losers: An Analysis of Competitive Balance in Formula One”. 2013. *Journal of Sports Economics*, 14(4), pp. 411–439. Doi: https://doi.org/10.1177/1527002513496009.

[16] T. Krauskopf, M. Langen, B. Bünger, “The Search for Optimal Competitive Balance in Formula One”. 2010. *CAWM Discussion Paper, No. 38*. [Online]. Available: https://www.econstor.eu/bitstream/10419/51362/1/672457962.pdf. [Accessed on August 15, 2022].

[17] O. Budzinski, T. Pawlowski, “The behavioural economics of competitive balance: Implications for league policy and championship management”. 2014. *Ilmenau Economics Discussion Papers, No. 89* [Online]. Available: https://www.econstor.eu/bitstream/10419/103420/1/796714266.pdf. [Accessed on August 15, 2022].

[18] C. Mastromarco, M. Runkel, "Rule changes and competitive balance in Formula One motor racing". 2009. *Applied Economics*, pp. 3003-3014.
Doi: https://doi.org/10.1080/00036840701349182.

[19] P.D. Owen, M. Ryan, C.R. Weatherston, "Measuring Competitive Balance in Professional Team Sports Using the Herfindahl-Hirschman Index". [2007] Rev Ind Organ 31, pp. 289–302. Doi: https://doi.org/10.1007/s11151-008-9157-0