Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5833

Optimization of Formula 1 Racing Strategies: An Approach Based on Exploratory Analysis and Genetic Algorithms

Eduardo de Lira Brennand [**Serra dos Órgãos University Center (UNIFESO)** | ***eduardolirabrennand@gmail.com***]
Eugênio Silva [**Serra dos Órgãos University Center (UNIFESO)** | ***eugeniosilva@unifeso.edu.br***]
Gabriel Resende Machado [**Pontifical Catholic University of Rio de Janeiro (PUC-Rio)** | ***gabriel.machado@ecoa.puc-rio.br***]

Coordenação de Ciência da Computação, Prédio Alice Ribeiro Nunes Pereira, Centro Universitário Serra dos Órgãos (UNIFESO), Estrada Venceslau José de Medeiros, 1045, Prata, Teresópolis, RJ, 25976-345, Brasil.

Instituto ECOA, Edifício Padre Laércio Dias de Moura S.J., Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 22451-900, Brasil.

Received: 03 April 2025 • **Accepted:** 18 January 2026 • **Published:** 07 May 2026

Abstract Formula 1 (F1) is a motorsport that demands high technical and strategic expertise, where tactical decisions can significantly influence driver's performance and race results. This work addresses the challenge of optimizing pit stop strategies and proposes solutions for strategic decisions aimed at minimizing total race time, such as tire compound selection and optimal stint planning, through Exploratory Data Analysis (EDA) and Genetic Algorithms (GAs). The study relies on historical F1 race data obtained through the FastF1 dataset to examine variables such as tire degradation, lap times, and the impact of pit stops on drivers' final positions. Based on the insights from EDA, a GA model was developed to simulate different race strategies and identify the most effective ones, serving as a complementary tool to enhance strategic decision-making. The model offers data-driven insights that can support race strategists in refining and adapting their strategies based on real-time race conditions and expert judgment. The results indicate that the proposed methodology can support teams in designing more efficient strategies, leading to better performance across various circuits.

Keywords: Genetic Algorithms, Exploratory Data Analysis, Formula 1.

1 Introduction

Formula 1 (F1) is widely regarded as one of the most technologically advanced and strategically demanding motorsports in the world [Bonomi *et al.*, 2023]. The sport combines cutting-edge engineering with precise tactical decisions, where even the smallest adjustments in strategy can have a significant impact on race outcomes. Among the many challenges faced by F1 teams, the optimization of race strategy, particularly in terms of pit stop timing and tire compound selection, stands out as a critical factor in achieving competitive performance [Rondelli, 2022]. This paper addresses this challenge by proposing a computational approach that leverages Exploratory Data Analysis (EDA) and Genetic Algorithms (GAs) to optimize race strategies, aiming to minimize total race time while adhering to the complex regulations of F1.

In F1, race strategy is influenced by multiple factors, including tire degradation, track conditions and the timing of pit stops. Teams must make real-time decisions based on data collected from the cars' sensors, which provide insights into tire wear, lap times, and other performance metrics. However, the complexity of these decisions, combined with the dynamic nature of the race, makes it difficult to manually determine the optimal strategy. Computational methods, particularly those based on data analysis and optimization algorithms, offer a promising solution to this problem by systematically exploring the vast space of possible strategies and identifying the most effective ones [Linden, 2008].

Despite the potential of computational approaches, the application of such methods in F1 is often limited by the lack of public data, as well as the proprietary nature of team strategies [Bonomi *et al.*, 2023]. Most existing research in this area often relies on simplified models or simulations, which may not fully capture the intricacies of real-world race scenarios. This paper seeks to bridge this gap by utilizing publicly available historical F1 race data to conduct a detailed exploratory analysis and develop a straightforward yet effective GA-based optimization model. The proposed model focuses on key variables such as tire compound selection, pit stop timing, and lap times, with the goal of generating strategies that minimize total race time while complying with F1 regulations. By leveraging publicly accessible data, this approach ensures reproducibility and accessibility, allowing the methodology presented in this article to be replicated.

The use of GAs in this context is particularly advantageous due to their ability to handle complex, multi-variable optimization problems. In this study, each candidate solution was represented by a potential race strategy, encoded as a sequence of tire compounds and pit stop timings. The algorithm evaluated these strategies based on their simulated performance, using historical data to estimate lap times and tire degradation. Through iterative evolution, the GA was able to identify strategies that effectively balanced the trade-off between tire performance and pit stops, ultimately leading to improved race outcomes in most experiments.

While the proposed model shows strong potential for optimizing race strategies, it is important to highlight that AI-based solutions are not intended to replace the expertise of race strategists. Instead, the GA model serves as a strategic support tool, offering valuable insights that can guide teams in adjusting their decisions based on race conditions and their own tactical experience. Rather than providing a definitive solution, the model helps identify promising strategies, giving strategists a structured foundation to explore and refine their approach, ultimately enhancing overall race performance. In this context, the model supports the dynamic work of race strategists, since many decisions still need to be made in real time during the race. Unexpected events such as weather changes, on-track incidents, or shifts in competitor behavior often require immediate human judgment beyond what precomputed solutions can anticipate.

Furthermore, the generalist nature of the proposed model allows it to be adapted and applied to any Formula 1 team, as long as the full evolutionary process of the algorithm is carried out using the internal data available to each organization. Because the genetic algorithm relies heavily on the characteristics of the dataset used for training and evaluation, teams can integrate their own telemetry, historical race results, tire behavior patterns, and internal strategic considerations to generate customized solutions aligned with their operational realities. In this sense, the model is not constrained to a single dataset or team profile but instead provides a flexible computational framework capable of supporting diverse strategic environments across the sport.

The results of this study demonstrate the effectiveness of the proposed approach in optimizing race strategies for various circuits. The GA-generated strategies were then compared *a posteriori* to those used by actual drivers, showing competitive performance in terms of total race time. These findings highlight the potential of computational methods to support F1 teams in making more informed strategic decisions, ultimately enhancing their competitiveness on the track.

The remainder of this paper is organized as follows. Section 2 introduces key concepts related to F1 strategies, such as pit stops, stints, and tire compounds. Section 3 reviews related work on race strategy optimization. Section 4 outlines the research methodology, covering data collection, exploratory data analysis, and the development of the GA-based optimization model. Section 5 presents the experimental setup, results, and analysis, comparing GA-generated strategies to real-world data. Finally, Section 6 discusses the key findings of this study and suggests directions for future research.

2 Key Concepts

2.1 Pit Stops

Pit stops are a crucial aspect of motorsport, enabling teams to make adjustments that can significantly impact a car's performance during a race [Heine and Thraves, 2023]. In F1, pit stops are a critical component of race strategy, primarily used for tire changes and addressing minor mechanical issues. However, this process requires the car to stop in the pit box, causing a temporary loss of track position compared to competitors.

These stops take place in a designated area parallel to the circuit, known as the *pit lane*, where teams' garages are located. For safety reasons, F1 mandates speed limits between 60–80 km/h in the pit lane, increasing the overall duration of the stop. Although pit stops add time to a driver's total race time and can result in position losses, they also offer strategic advantages. Changing worn tires for a fresh set of specific compounds allows drivers to achieve faster lap times [Heilmeier *et al.*, 2020]. Thus, teams must carefully balance the time lost in pit stops with the potential performance gains in subsequent laps. The primary objective of this work is to optimize this trade-off to maximize race results.

2.2 Stints

The period between two consecutive pit stops is known as a *stint*, which represents the duration a driver remains on track using the same set of tires. In practical terms, a stint begins when a driver exits the pit lane with fresh tires and ends when the next pit stop occur. Stint length is a critical strategic decision influenced by factors such as tire degradation and track conditions.

2.3 Tire Compounds

Tire compound selection is one of the most decisive factors in F1 strategy, as it directly affects car performance and overall race time. Each tire compound offers unique characteristics in terms of grip, degradation, and durability. Currently, Pirelli is the exclusive tire supplier for F1 and provides five different types of tire compounds, as depicted by Figure 1:

- **Slick tires (*soft*, *medium*, *hard*):** it is represented by the first three models in Figure 1, from left to right. These tires are designed for dry track conditions, featuring a smooth surface that maximizes grip [Bonomi *et al.*, 2023]. Soft tires provide the highest speed but degrade quickly, making them ideal for short stints and fast lap times. Medium tires offer a balance between grip and longevity, while hard tires last the longest but provide the least grip.
- **Intermediate tires:** represented by the green-colored model, these tires are used when the track is wet but not fully soaked. They feature grooves in the tread to aid water dispersion, reducing the risk of aquaplaning while maintaining a reasonable level of grip. [Bonomi *et al.*, 2023].
- **Wet tires:** represented by the blue-colored model, these tires are designed for extreme wet conditions. They have deep grooves that help drain large amounts of water, preventing loss of control in highly wet environments [Bonomi *et al.*, 2023].

Choosing the right tire compound at the optimal time during a race is critical to minimizing lap times and maximizing overall performance. An effective pit stop strategy ensures that teams make tire changes at the most advantageous moments, ultimately influencing race results.

Figure 1. F1 tire compounds.

3 Related Works

The optimization of race strategies in F1 has garnered significant attention in both academic and professional circles, given the sport's reliance on data-driven decision-making. While F1 teams often employ proprietary methods and simulations to refine their strategies, academic research has explored various computational approaches to tackle this complex problem. These approaches range from traditional optimization techniques to more advanced Machine Learning and Evolutionary Algorithms.

[Bonomi *et al.*, 2023] present an evolutionary approach for optimizing race strategies in F1 using GAs. Their method generates optimized race plans by simulating various combinations of key variables, including tire compounds, number of pit stops, fuel load, and climatic conditions. The algorithm was evaluated through both simulation and real-world data provided by *Pirelli* engineers, demonstrating its practical viability compared to conventional strategies. However, the study is limited by a simplified representation of genetic operators, emphasizing only fuel load and tire compound variations, without mentioning their reliance on a constrained dataset that combines real measurements from *Pirelli* with simulation data derived from video-game environments. In contrast, this work exclusively leverages publicly available historical data, thereby enhancing reproducibility and ensuring an unbiased dataset. Furthermore, the genetic representation integrates both the sequence of tire compounds and the timing of pit stops, offering greater flexibility in strategy optimization.

Another research initiative is the study from [García Villalón, 2022], which employs Machine Learning (ML) to predict F1 race outcomes by conducting extensive EDA, feature engineering, and modeling. Using a variety of supervised and unsupervised learning techniques, including regression, classification, and Artificial Neural Networks (ANNs), the study analyzes diverse datasets covering circuits, constructors, drivers, lap times, pit stop occurrences, qualifying sessions, and final standings. While the work of [García Villalón, 2022] provides valuable insights into race result prediction, it does not address the strategic components of tire compound selection and pit stop execution, which are critical for optimizing overall performance. In contrast, this article specifically targets the optimization of pit stop strategies and tire usage, focusing on how these strategic variables impact total race time and performance.

Additionally, [Heilmeier *et al.*, 2020] propose a Virtual Strategy Engineer (VSE) that utilizes ANNs to support decision-making in F1 races. Their system automates decisions regarding pit stops and tire compound selection based on historical telemetry and race data from 2014 to 2019, incorporating factors such as yellow flag conditions, tire age, and race position. Although the VSE demonstrates robust performance in simulated environments, especially during critical events like safety car periods, it separates the decision of executing a pit stop from that of selecting the appropriate tire compound. This decoupling may reduce precision under rapidly changing race conditions, as it does not account for the integrated effect of multiple strategic decisions, such as the total number of pit stops and the full sequence of tire compounds. This work addresses these limitations by employing a GA that simultaneously optimizes the ordering of tire compounds and the timing of pit stops, thereby providing a more comprehensive framework for race strategy optimization. Table 1 specifies all differences between the related works and the relationship between them and this article.

This study advances the existing body of research by specifically applying GAs to optimize pit stop and tire strategies in competitive racing scenarios. By harnessing the adaptability and robustness of evolutionary computation, the proposed methodology seeks to enhance decision-making processes in real-time racing environments. The approach integrates real-time race dynamics with historical data patterns, enabling a more comprehensive and adaptive strategy formulation that responds effectively to the complexities of competitive racing.

4 Research Methodology

4.1 Technologies and Tools

The entire development of this study was conducted using the `Python` programming language, version 3.12.3. The libraries employed included `Pandas` and `NumPy` for data analysis and processing, `Matplotlib` and `Seaborn` for data visualization. It is worth mentioning the GA itself was implemented from scratch, relying only on a selection of the aforementioned Python libraries as foundational tools.

4.2 Data collection

The dataset used in this study was obtained through the `FastF1 Python` library [Oehrl and Schaefer, 2025], which provides access to historical F1 race data since the 2018 season. The library enables the extraction of raw race data, including telemetry information, lap times, tire compounds, and specific race events. `FastF1` seamlessly integrates with Pandas DataFrames, facilitating efficient data manipulation, filtering, aggregation, and visualization [McKinney *et al.*, 2010]. This integration allows for in-depth exploratory analysis, ensuring the extraction of relevant insights for the study.

4.3 Data processing

The initial phase involved several data preprocessing steps to ensure the analysis' accuracy, including attribute selection, data cleaning, and formatting. Key variables that directly

Table 1. Comparison of related works.

Study	Optimization Technique	Dataset Type	Methodological Approaches
[Bonomi *et al.*, 2023]	GAs	Simulation with real Pirelli data.	Simplified genetic operators; private dataset.
[García Villalón, 2022]	ML	Real-world race data.	No tire/pit stop; strategy optimization.
[Heilmeier *et al.*, 2020]	ANNs	Historical telemetry (2014-2019).	Decoupled decision-making for pit stop/tire selection
This work	GAs	Public historical data.	Comprehensive optimization; integrating tire sequence; pit stop timing.

impact race strategies, such as lap times, tire compounds, pit stop events, and track status, were selected. Rows with missing or null values in critical columns (e.g., tire compounds, lap times) were removed as they accounted for only 1.58% of the dataset, representing a negligible impact on the overall analysis. No imputation techniques were applied in order to preserve the dataset's authenticity and ensure maximum realism and reliability in the simulation. Laps under non-standard conditions (e.g., safety car periods) were also filtered out. Lap times were standardized to seconds for easier comparison between laps and drivers. The attributes that were selected for the analysis are organized in Table 2.

4.4 EDA Methodology

The EDA aimed to characterize key patterns in Formula 1 racing dynamics and to identify variables that meaningfully contribute to performance variation and strategic decision-making. In order to provide a statistical basis for the variables considered in this study, a Pearson correlation matrix was computed for all numerical attributes related to driving behavior during races. This analysis provided empirical support for variable selection and ensured interpretability for the subsequent GA-based strategy modeling. The resulting correlation structure is visualized in the heatmap shown in Figure 2, which summarizes the strengths and directions of the associations discussed below.

To complement the correlation-based examination of numerical variables, a set of focused exploratory analyses was conducted to further interpret how these relationships manifest in real race scenarios. Accordingly, the following subsections examine tire behavior, pit stop dynamics, and driver strategy patterns through visual and comparative analysis, providing domain-specific insight into the statistical associations highlighted in the heatmap.

- **Tire Performance:** the relationship between lap times and tire degradation was analyzed for different tire compounds (soft, medium, hard). Figure 3 illustrates how each compound's performance varies with wear, showing that soft tires offer faster lap times but degrade more quickly, while hard tires are more durable but slower;
- **Pit Stop Impact:** the impact of pit stops on lap times and final race positions was analyzed. Figure 4 illustrates how pit stops influenced the lap times of the 2022 Spanish *Grand Prix* winner. Typically, drivers tend to experience slower lap times before a pit stop and faster lap times immediately after changing tires;
- **Driver Strategies:** the tire and pit stop strategies of top-performing drivers were analyzed. Figure 5 highlights the strategies adopted by drivers who gained the most positions during the race, emphasizing the importance of tire choice and pit stop timing.

4.5 GA Implementation

A GA was designed to optimize race strategies, adhering to the principles of evolutionary computation. Each individual in the population represents a unique race strategy, modeled to align with F1 regulations and constraints. This evolutionary approach enables the exploration of a vast solution space to identify strategies that balance performance and resource management, with a primary focus on tire compound selection and pit stop timing.

The GA was trained using lap and tire data aggregated from all drivers, which implies that neither individual driver skill nor car performance was accounted for, given the fact that these attributes are inherently difficult to quantify. Consequently, the model prioritizes strategic decision-making rather than performance characteristics that cannot be reliably captured through publicly available data.

4.5.1 Fitness Function

The fitness function evaluates the total race time for a given strategy. Table 3 details its key components. The total race time is calculated using the Equation 1:

$$T_{\text{total}} = \sum_{i=1}^{n} t_i + P \cdot T_{\text{pit}} \tag{1}$$

where T_{total} represents the total race time; t_i corresponds to the lap time for lap i, calculated based on tire compound wear; n represents the total number of laps in the race; P corresponds to the number of pit stops made, and finally T_{pit} stands for the average time spent in a pit stop.

4.5.2 Selection

The selection process determines which individuals (strategies) are chosen to produce the next generation [Linden, 2008].

Table 2. Selected attributes for race strategy analysis.

Attribute Category	Attributes
Driver-specific information	Driver name and number; Team affiliation.
Performance metrics	Lap times; Personal best indicators; Track status conditions.
Tire management data	Compound tire types; Tire life tracking; Fresh tire indicators.
Race position information	Current position; Grid position; Final position; Racing status.
Engineered attributes	`PitStopBool`: Binary indicator for pit stop occurrence; `PitStops`: Cumulative count of pit stops.

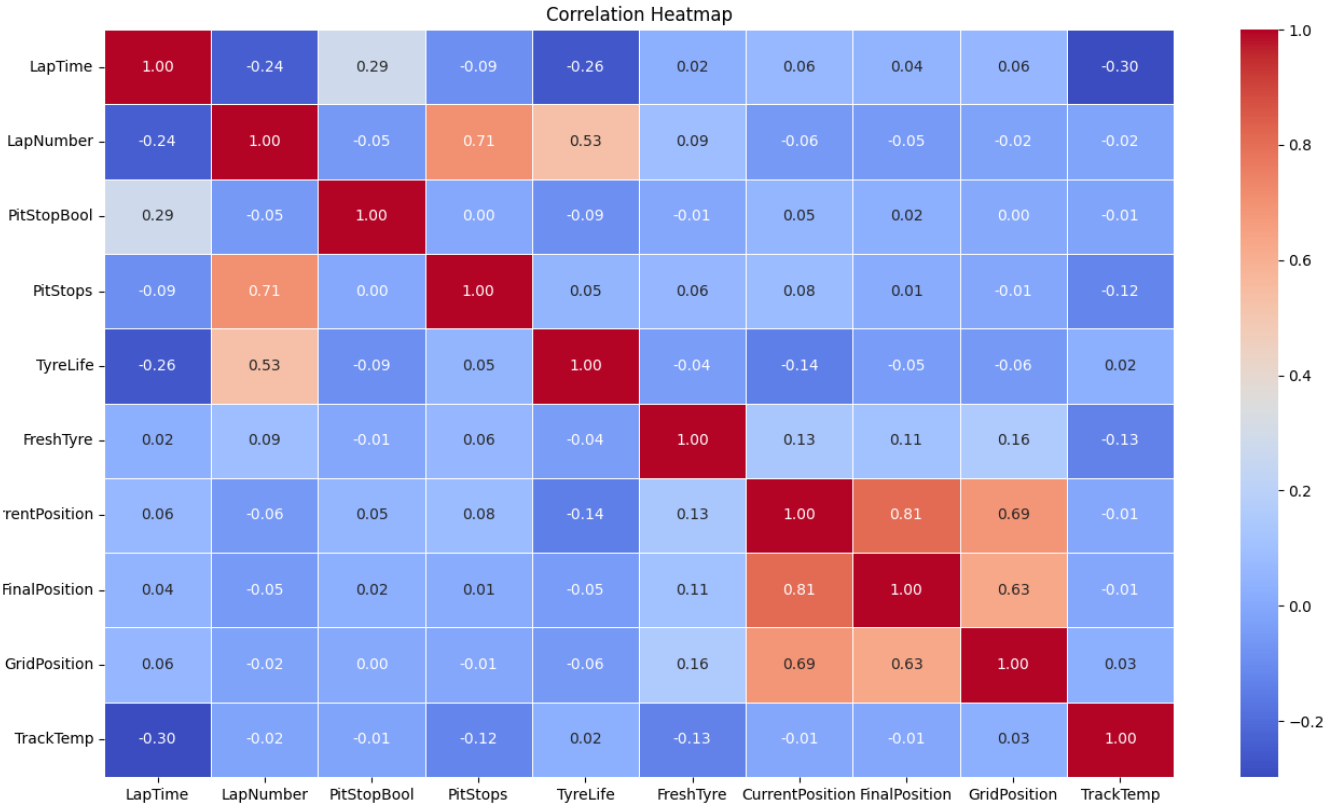

Figure 2. Heatmap of the correlation matrix for all numerical variables used in the EDA

Table 3. Key components of the fitness function.

Category	Components
Rule Compliance Verification	Validation of mandatory compound changes; Compliance check for minimum pit stops; Monitoring of maximum compound usage.
Performance Calculation	Lap time estimation based on tire compound; Consideration of tire degradation effects; Integration of pit stop times into total race time.
Penalty System	3600-second penalty for rule violations; Penalties for exceeding compound usage limits; Penalties for failing minimum pit stop requirements.

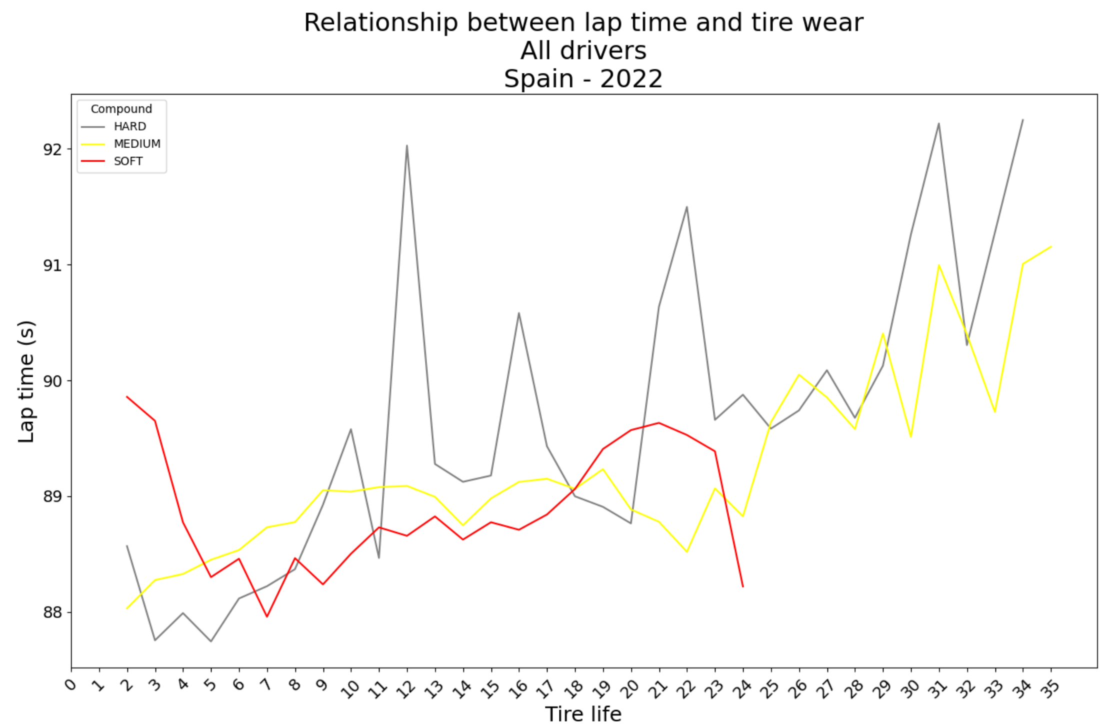

Figure 3. Relationship considering all drivers between lap time and tire wear in 2022 Spanish *Grand Prix* (GP).

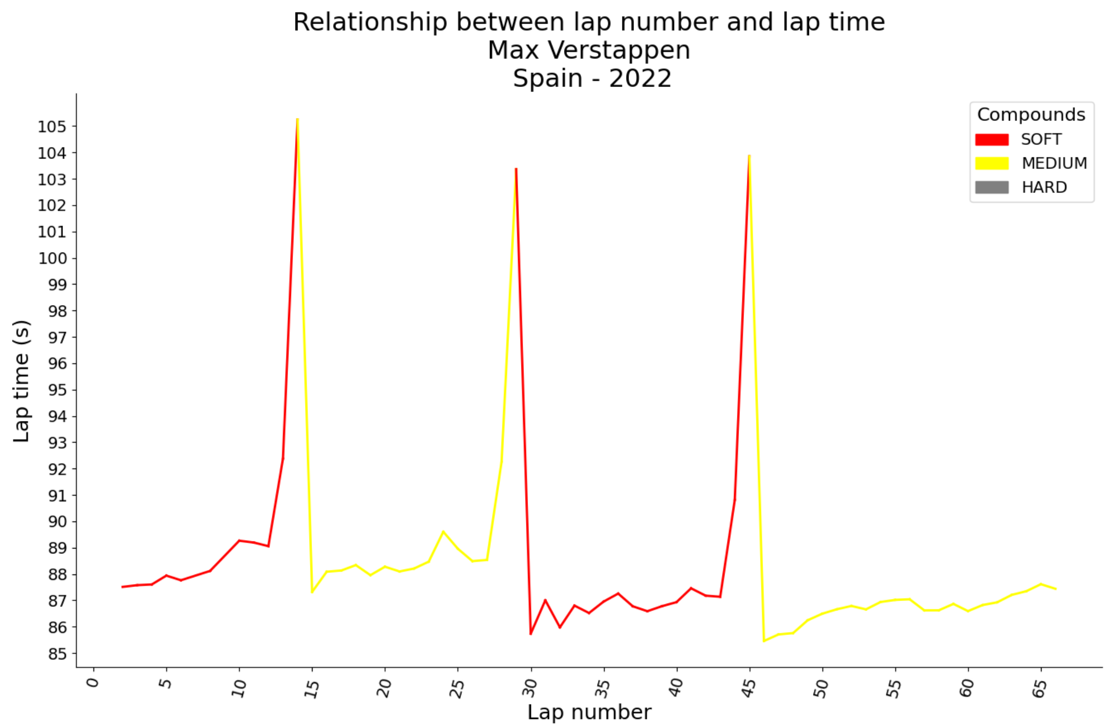

Figure 4. Laps of the winner Max Verstappen considering all his tire compounds used along the 2022 Spanish *Grand Prix*.

The roulette wheel selection method was used in this work, where each individual's probability of being selected is proportional to its fitness (i.e., lower total race times have higher probabilities). This method ensures that better performing strategies are more likely to be selected while still allowing less optimal strategies the opportunity to contribute to the gene pool.

To implement elitism, the top 10% of individuals (based on fitness) are automatically carried over to the next generation without modification. This ensures that the best solutions are preserved and not lost during the evolutionary process.

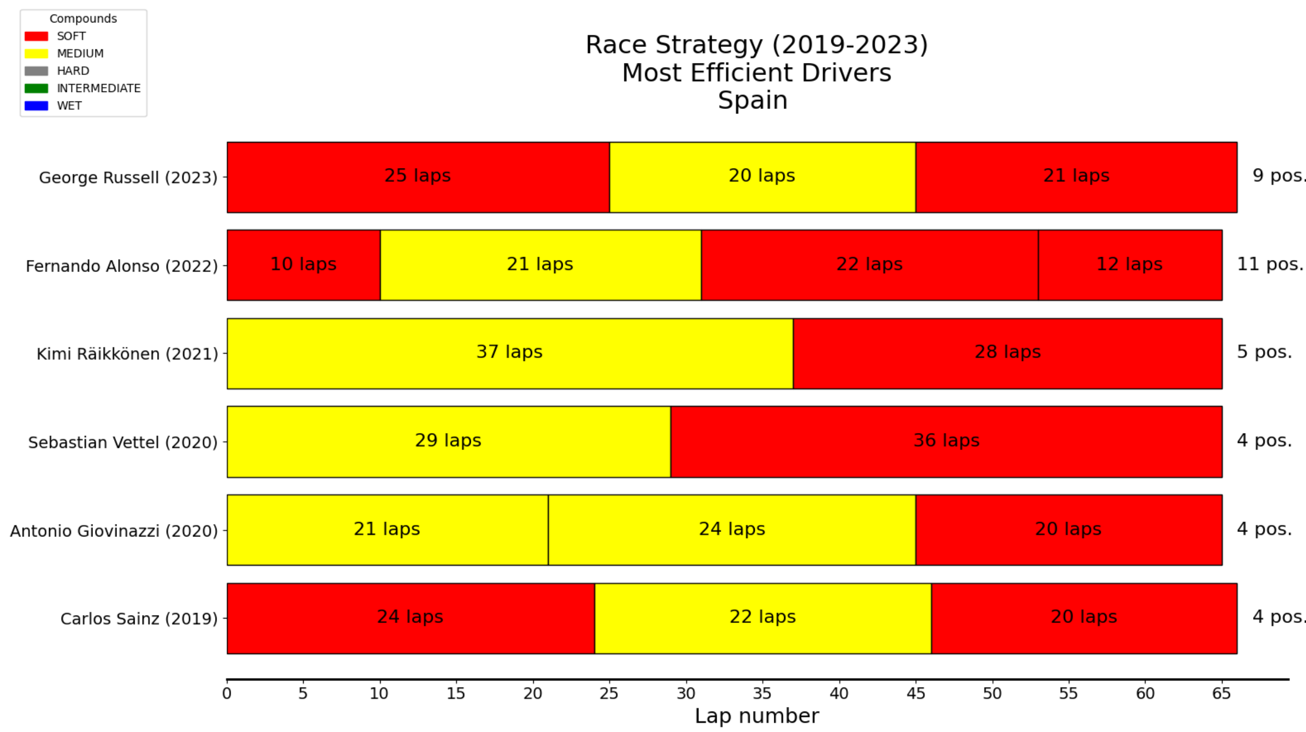

Figure 5. Race strategies adopted by the most efficient drivers during the Spanish *Grands Prix* from 2019 to 2023.

4.5.3 Crossover

The crossover operator combines genetic material from two parent individuals to generate offspring [Linden, 2008]. A one-point crossover method was applied separately to the `CompoundOrder` and `PitStops` components of the chromosome. For tire compounds, a random crossover point was selected within the sequence of both parents. The offspring inherits the tire sequence from the first parent up to this point and the remaining sequence from the second parent. A similar approach was used for pit stop timings, where the offspring inherits pit stop laps from the first parent before the crossover point and from the second parent afterward. To maintain chronological consistency, the pit stop sequence is then reordered if necessary.

4.5.4 Mutation

The mutation operator introduces random changes to the chromosome to maintain genetic diversity and explore new regions of the solution space. Three types of mutations were implemented, where each mutation type was defined to has a low probability of occurring (e.g., 1%) to balance exploration and exploitation.

- **Tire Compound Mutation:** A random tire compound in the `CompoundOrder` sequence is replaced with another valid compound (e.g., soft, medium, or hard). This allows the GA to explore different tire strategies, as presented by (a) in Figure 6;
- **Pit Stop Lap Mutation:** A random pit stop lap in the `PitStops` list is modified. This can involve changing the lap number, adding a new pit stop, or removing an existing one, as long as the minimum and maximum number of pit stops (1 and 5, respectively) are respected, as depicted by (b) in Figure 6;
- **Pit Stop Count Mutation:** The number of pit stops is increased or decreased by one, within the allowed limits. If the number of pit stops is increased, a new tire compound is randomly selected to ensure the consistency of the strategy. This mutation helps explore strategies with different numbers of pit stops, as indicated by (c) in Figure 6.

4.5.5 Algorithm Flow

The GA implemented in this study follows a structured flow to optimize race strategies for F1. The process begins with the initialization of the population and iteratively evolves through selection, crossover, mutation, and evaluation until a termination condition is met. Algorithm 1 provides a a formal explanation of the GA implemented in this study.

5 Results

5.1 Analysis and Discussion of EDA

The EDA focused on key variables influencing driver performance and strategic decision-making in F1, such as tire compounds, lap times, and final race positions.

The findings highlighted the distinct impact of each tire type on lap performance, which was found to be in consonance with the explanations related to Figure 1.

- **Soft Tires** provided the fastest lap times but degraded quickly, limiting their usability for extended stints;

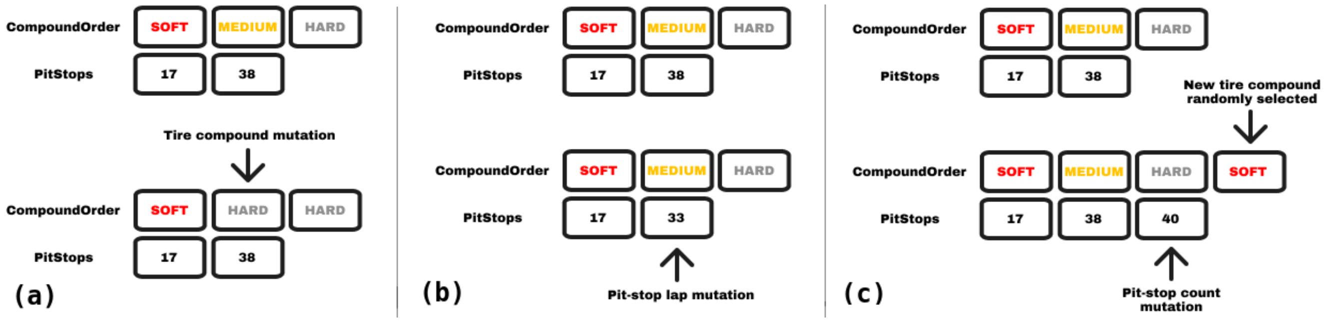

Figure 6. Mutation approaches: (a) tire compound; (b) pit stop lap; (c) pit stop count.

- **Medium Tires** balanced performance and longevity, making them a preferred choice for flexible pit stop strategies;
- **Hard Tires** offered greater durability but resulted in slower lap times.

The analysis revealed that excessive pit stops increased overall race time due to the additional time spent in the pit lane, whereas fewer stops allowed drivers to gain positions on track. Figure 7 illustrates this effect: Sebastian Vettel, who executed a one-stop strategy using medium and hard tires, achieved the highest position gains, whereas Pierre Gasly, with a three-stop strategy, experienced the most position losses. These results underscore the necessity of balancing pit stop frequency and tire selection to optimize race performance.

So as to complement this analysis, a correlation heatmap (Figure 2) was examined to summarize the statistical relationships among the numerical variables. Performance-related attributes displayed patterns aligned with racing aspects and stint evolution: `LapNumber` and `TrackTemp` showed moderate negative correlations with `LapTime` (–0.24 and –0.29), reflecting the effects of fuel burn-off, rubber buildup, and environmental conditions on race pace. The relationship between `TyreLife` and `LapTime` also revealed a nuanced dynamic, with a negative correlation (–0.26) arising from the combined influences of tire degradation and decreasing fuel mass over the course of a stint. Strategy-related variables behaved as expected, with `PitStopBool` correlating positively with `LapTime` (0.28) and `PitStops` correlating strongly with `LapNumber` (0.70), confirming the temporal structure of pit stop execution. Position-related variables (`CurrentPosition`, `FinalPosition` and `GridPosition` exhibited strong internal correlations, reflecting structural dependencies between starting position, in-race competitiveness, and final results.

The insights gained from the EDA provided a crucial foundation for developing a GA aimed at optimizing race strategies. By leveraging historical data patterns, the GA's hyperparameters were fine-tuned to simulate realistic race scenarios and seek the best combination of tire performance and pit stop frequency.

5.2 Discussion of GA Results

The GA demonstrated strong performance in identifying competitive race strategies across various circuits, achieving total race times comparable to or better than actual race results. For instance, in the 2023 Italian *Grand Prix*, the GA-derived strategy yielded a total race time lower than most of the drivers who completed the race, as shown in Table 4. Figure 8 compares the pit stop and tire strategies employed by real drivers with those generated by the GA, highlighting the algorithm's ability to optimize race strategies effectively.

It is important to note that only data from the years 2022 and 2023 were used in this analysis, as these are the years under the current regulations that began in 2022. During this period, the cars are highly similar, which facilitates standardization and ensures a more consistent evaluation of the GA's performance across different circuits.

To ensure consistency and comparability across all experiments, the GA was configured with a set of optimal parameters determined through a systematic parameter tuning process. This process involved evaluating a wide range of parameter combinations, including population size, mutation rate, crossover rate, number of generations, and elitism rate, to identify the configuration that consistently produced the best results. The optimal parameters identified through this process are presented below:

- **Population**: 50 individuals;
- **Mutation rate**: 1%;
- **Crossover rate**: 50%;
- **Elitism rate**: 10%;
- **Generations**: 300.

These parameters were selected because they consistently yielded the lowest total race times and the most stable performance across multiple test runs. By using the same configuration for all circuits, the study ensured that any variations in results could be attributed to the unique characteristics of each track rather than differences in the algorithm's setup.

To further validate the robustness and reliability of the GA, statistical experiments were conducted on three distinct circuits chosen randomly considering the 2022 and 2023 seasons: (i) Italy, (ii) Great Britain, (iii) and Spain. These circuits were chosen randomly to ensure an unbiased evaluation across different track characteristics. In addition to these experiments, a specific case study was performed using data from the 2024 Great Britain *Grand Prix*, a race that featured variable weather conditions and periods of rain. This analysis aimed to test the algorithm's ability to adapt its strategy under dynamic race environments and evaluate how effectively it responds to unpredictable climatic changes.

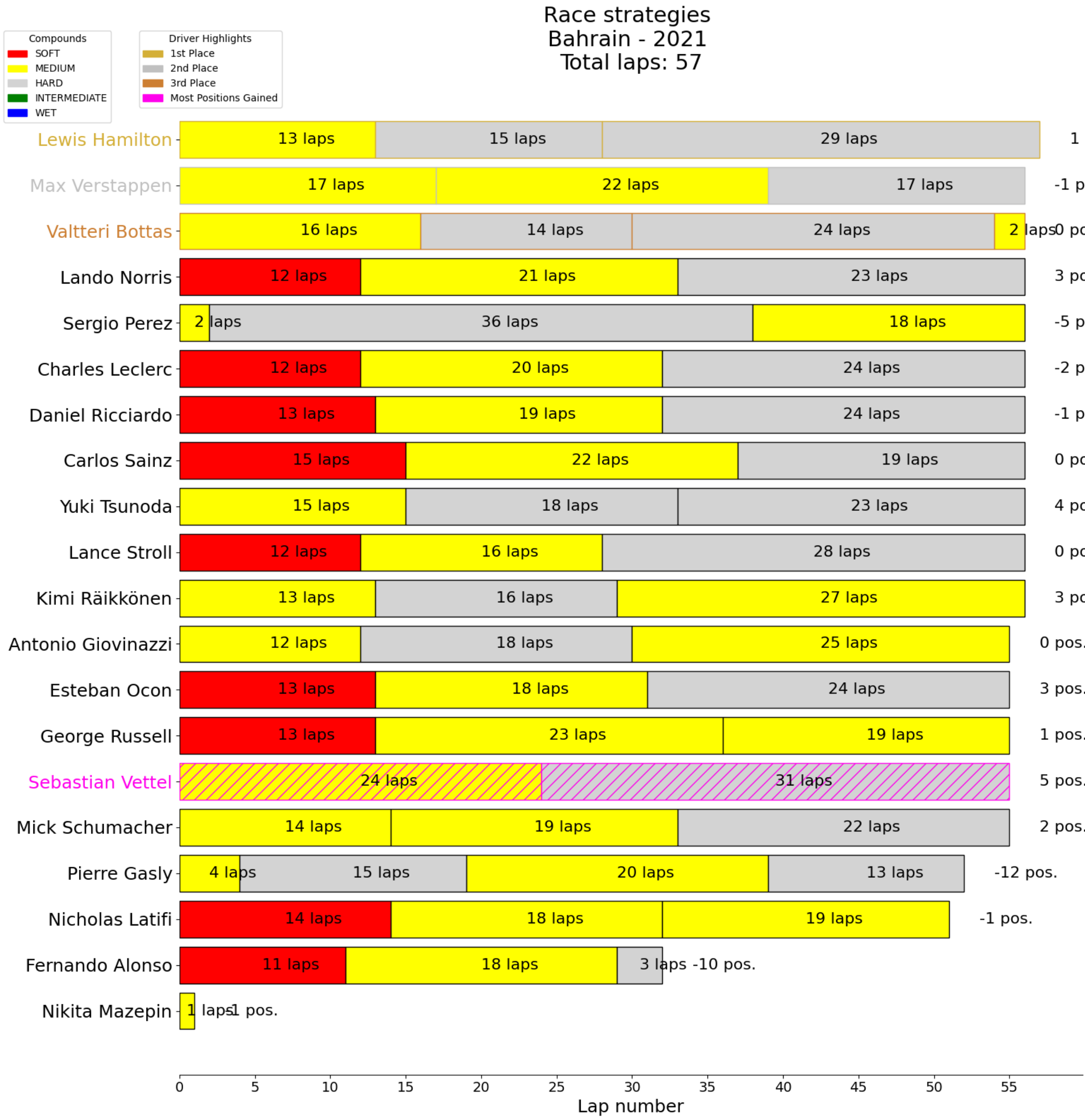

Figure 7. Race strategy for all drivers over 57 laps of the 2021 *Bahrain Grand Prix*.

Given that GA is a stochastic process, the results of each run may vary due to the probabilistic nature of the genetic operators. To assess the consistency of the results, the GA was executed 30 times for each circuit, using the same configuration parameters. In each execution, the best individual was recorded, then, the mean and standard deviation of the race times for each circuit were calculated, as shown in Table 5.

Table 6 shows the mean and standard deviation of the total race times for actual F1 races in 2022 and 2023, across the same circuits that were used by the GA. This data allows for an understanding of the overall consistency and variability in the race results for these specific years and circuits. Additionally, the table includes a comparison with the GA's final position based on its total race time, providing insight into how the GA's optimized strategies would have performed relative to actual race results. Furthermore, the table also presents the winner's total race time and the most efficient driver's total race time for each race, offering a benchmark to compare the GA's performance. The most efficient driver is defined as the one who gained the most positions during the race, highlighting their ability to recover and perform well despite challenges. It is important to note that some races exhibit a very high standard deviation in total race times. This is primarily due to instances where certain drivers faced issues that prevented them from finishing the race, resulting in unusually low total race times and, consequently, high standard deviations.

The mean represents the average performance of the strategies generated by the GA, while the standard deviation in-

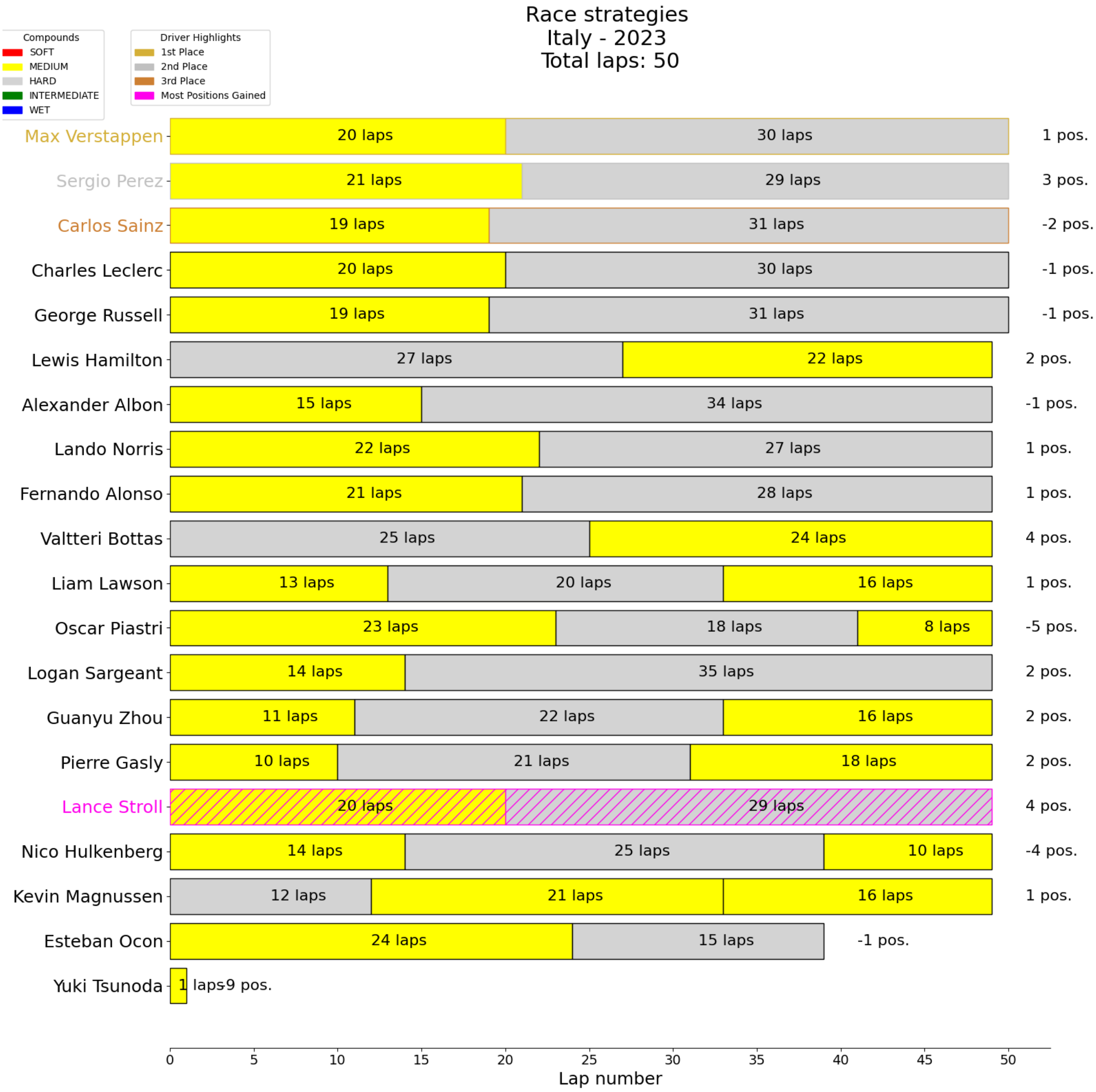

Figure 8. Drivers' strategies compared to GA's strategy regarding 2023 Italian GP over 50 laps.

dicates the degree of dispersion in the results. A smaller standard deviation suggests that the GA is producing more consistent and predictable solutions, whereas a larger standard deviation may indicate that the results are more sensitive to variations in the search space.

These results underscore the importance of evaluating the reliability of GA outcomes across different scenarios. For an evolution-based algorithm, a smaller standard deviation is desirable, as it indicates that the optimization process is converging toward consistent solutions and that the search space is being explored efficiently. In the context of F1 race strategies, this implies that the GA is capable of identifying robust solutions that can be applied with greater confidence during an actual race.

The results obtained for Great Britain demonstrate that the GA was able to adapt well to the specific characteristics of this circuit, resulting in more predictable strategies. In contrast, the higher standard deviation observed for Spain suggests that the search space for this circuit may be more complex or sensitive to small variations in input parameters. This opens the door for further adjustments to the crossover and mutation operators, or even a more detailed exploration of the strategic parameters specific to this circuit.

Overall, the GA's ability to generate competitive strategies across multiple circuits reinforces its potential as a valuable tool for optimizing race strategies in F1. However, the complexity of real-world race variables, such as dynamic track conditions and competitor behavior, may lead to some deviations between algorithmic and practical outcomes. These findings highlight the adaptability of the GA to different rac-

Algorithm 1: GA for F1 Race Strategy Optimization

Input: N : Population size (number of individuals per generation);
G_{max} : Maximum number of generations;
p_c : Crossover rate (probability of recombination);
p_m : Mutation rate (probability of mutation per gene);
p_e : Elitism rate (percentage of top individuals preserved);
T_{total} : Total individuals generated ($N \times G_{max}$);
$f(x)$: Objective function (total race time minimization)
Output: x_{best} : Best pit stop strategy found

$P_0 \leftarrow$ InitializePopulation(N);
$t \leftarrow 0$;
$total\_individuals \leftarrow 0$;
While $t < G_{max}$ ***and*** $total\_individuals < T_{total}$ ***and !StoppingCriterion()*** **do**
  For $i \leftarrow 1$ ***to*** N **do**
    $fitness_i \leftarrow f(x_i)$; `// Evaluate each individual's race time`
  end
  $n_e \leftarrow \lfloor N \times p_e \rfloor$; `// Number of elite individuals`
  $P_{elite} \leftarrow$ SelectElite(P_t, n_e);
  $P_{offspring} \leftarrow P_{elite}$;
  $num\_offspring \leftarrow n_e$;
  $P_{parents} \leftarrow$ SelectParents($P_t \setminus P_{elite}$);
  While $|P_{offspring}| < N$ **do**
    $(p_1, p_2) \leftarrow$ SelectParentPair($P_{parents}$);
    If *random()* $< p_c$ **then**
      $(c_1, c_2) \leftarrow$ Crossover(p_1, p_2);
      $num\_offspring \leftarrow num\_offspring + 2$;
    end
    else
      $(c_1, c_2) \leftarrow (p_1, p_2)$;
      $num\_offspring \leftarrow num\_offspring + 2$;
    end
    $c \leftarrow$ SelectRandomOffspring();
    For *each gene g in c* **do**
      If *random()* $< p_m$ **then**
        $c \leftarrow$ Mutate(c, g);
      end
    end
    $P_{offspring} \leftarrow P_{offspring} \cup \{c\}$;
  end
  $P_{t+1} \leftarrow P_{offspring}$;
  $t \leftarrow t + 1$;
  $total\_individuals \leftarrow total\_individuals + num\_offspring$;
  $x_{best} \leftarrow$ GetBestSolution(P_t);
end
return x_{best}

ing conditions while also pointing to areas where further refinement may be necessary to enhance its performance in more challenging scenarios.

Table 4. Total Race Time: Drivers vs. GA in 2023 Italian GP.

Driver	Total Race Time (*s*)
Genetic Algorithm	**4300,38**
Max Verstappen	4333.61
Sergio Perez	4341.30
Carlos Sainz	4345.29
Charles Leclerc	4345.73
George Russell	4352.43
Lewis Hamilton	4374.58
Alexander Albon	4379.40
Lando Norris	4379.77
Fernando Alonso	4380.46
Valtteri Bottas	4398.96
Logan Sargeant	4404.40
Oscar Piastri	4405.25
Liam Lawson	4406.27
Lance Stroll	4414.66
Guanyu Zhou	4414.92
Pierre Gasly	4415.24
Nico Hulkenberg	4425.12
Kevin Magnussen	4434.64
Esteban Ocon	4440.58

5.3 Case Study: Silverstone 2024 (Wet Conditions)

Since the current model does not explicitly account for race interruptions such as safety car periods or sudden weather changes, an additional experiment was conducted to evaluate the GA's adaptability under less predictable conditions. To this end, the algorithm was tested on the 2024 British *Grand Prix* at Silverstone — a race that featured a mid-race rain period. The rain interval (laps 18 to 38) was defined according to official race data, allowing the GA to adapt its strategy within this context. This setup enabled us to observe how the algorithm adjusts tire selection and pit stop timing when weather transitions occur.

The strategy generated by the GA is illustrated in Figure 9. The algorithm identified a three-stint configuration composed of hard and intermediate tires. The first stint lasted 19 laps on hard tires under dry conditions, followed by a 19-lap stint on intermediate tires during the rainy phase, and a final 14-lap stint again on hard tires after track drying.

The alignment between the intermediate tire stint and the rain window demonstrates that the GA was able to consistently adapt its recommendations when externally informed of the weather transition. Even without explicit predictive modeling of climatic dynamics, the algorithm converged to a realistic and technically coherent wet–dry strategy. The strategy found by the GA resulted in a total race time of 4867.76 seconds, faster than the actual winner's time of 4947.05 seconds at the 2024 British *Grand Prix*, placing the GA-derived strategy in first position. These results reinforce both the robustness and the optimization capability of the proposed approach under non-ideal race conditions, while confirming its role as a strategic support tool rather than a real-time decision-making system.

Table 5. Mean (μ) and standard deviation (σ) of total race times generated by the GA compared with all drivers' total times, across 30 runs.

Circuit	Year	GA's Total Time: $\mu \pm \sigma$ (s)	All Drivers' Total Time: $\mu \pm \sigma$ (s)
Great Britain	2022	4856.933 ± 23.320	4476.002 ± 921.965
Great Britain	2023	4749.151 ± 0.094	4754.631 ± 1045.073
Italy	2022	3908.444 ± 1.915	4413.743 ± 998.075
Italy	2023	4300.381 ± 3.137	4278.211 ± 195.590
Spain	2022	5814.582 ± 8.858	5539.404 ± 1040.499
Spain	2023	5207.152 ± 11.571	5320.421 ± 24.651

Table 6. Times for actual F1 races in Great Britain, Italy, and Spain during the 2022 and 2023 seasons, along with the winner's total race time, the most efficient driver's total race time, and the final position of the GA based on the average race time over 30 runs.

Circuit	Year	Winner Time (s)	Efficient Driver Time (s)	GA Final Position
Great Britain	2022	4433.172	4939.114	3rd
Great Britain	2023	4938.467	5129.820	1st
Italy	2022	4827.511	4832.891	1st
Italy	2023	4333.618	4326.883	1st
Spain	2022	5840.475	5847.322	1st
Spain	2023	5277.940	5310.329	1st

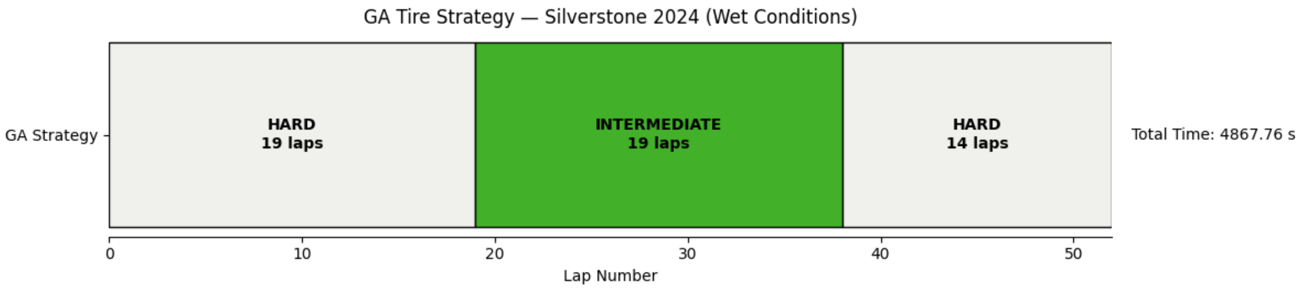

Figure 9. GA-generated tire strategy for the 2024 British *Grand Prix* at Silverstone under wet conditions.

6 Final Considerations

The EDA techniques used in this study were instrumental in gathering detailed insights into tire behavior and the impact of pit stop frequency on race performance. The strategic use of different tire compounds was shown to be crucial for optimizing lap times and the number of required pit stops. These historical data insights provided a strong foundation for configuring the GA with realistic performance constraints. The GA implementation, primarily guided by insights from EDA, demonstrated effectiveness in optimizing race strategies for F1, particularly in tire selection and pit stop timing. It is important to note that the GA model was trained using race times from all drivers, including those who retired due to mechanical failures or accidents. Retaining this data influenced the average race time, shifting it toward lower values and increasing the stringency of the GA's fitness evaluation. This, in turn, enhanced the model's robustness in adversarial scenarios, where identifying competitive strategies is crucial. The results showed by Tables 5 and 6 point out the GA model frequently achieved race times comparable to or better than those observed in actual races, underscoring the potential of evolutionary computation in the dynamic context of F1 strategy optimization.

Furthermore, it is important to emphasize that the proposed GA model is intended to serve as a support tool rather than a definitive solution for race strategy planning. The model provides a structured framework for exploring different strategic options, enabling race strategists to incorporate data-driven insights into their decision-making process. By combining computational efficiency with human expertise, the model allows for more informed adjustments based on real-time race conditions and strategic priorities. This highlights the potential of AI-based approaches not as replacements for human judgment, but as valuable instruments for enhancing the precision and adaptability of race strategies.

Despite the relatively simple chromosome representation and a limited set of variables, the experiments confirmed the effectiveness of GAs in optimizing strategies across different circuits. This suggests that even a straightforward model can generate meaningful insights and competitive strategies, highlighting the potential of AI-based approaches to enhance race planning and decision-making.

6.1 Future Work

To expand research on GAs in F1, future studies could explore more complex models incorporating additional variables, such as:

- **Dynamic Weather Conditions:** that affect tire performance and race outcomes;
- **Real-time Tire Wear Data:** to refine strategy adaptability.

- **Enhanced Pit Stop Modeling**: including pit-lane traffic as well as undercut and overcut effects — *i.e.*, early pit stop with fresh tires or delayed pit stop on older rubber — to improve strategy realism.

Advanced optimization techniques, such as hybrid algorithms combining GAs with ANNs, could also be implemented to improve performance prediction across diverse racing scenarios. Finally, a crucial challenge remains: the limited access to detailed F1 team data. While publicly available data provides a foundation for initial analyses, access to internal telemetry data, including precise tire temperature, aerodynamic drag, real-time fuel consumption, and specific tire degradation metrics, would enable the development of significantly more sophisticated and accurate models.

Declarations

Acknowledgements

The authors acknowledge the support provided by the Serra dos Órgãos University Center (UNIFESO) for this research. We also thank the developers of the Python FastF1 library Oehrl and Schaefer [2025], whose framework for accessing historical Formula 1 data was fundamental to the analysis and experimental setup presented in this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

Eduardo Brennand is the main author and was responsible for the development of the EDAs, the GA, as well as the execution and discussions of experiments. Gabriel Resende Machado contributed as the advisor professor and peer reviewer. Eugênio Silva contributed as the co-advisor professor and technical reviewer. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets and the source code generated along this study are available in this GitHub repository under MIT license.

References

Bonomi, A., Turri, E., and Iacca, G. (2023). Evolutionary f1 race strategy. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, pages 1925–1932. DOI: 10.1145/3583133.3596349.

García Villalón, M. J. (2022). Organización, análisis de datos e inteligencia artificial para la predicción de resultados en fórmula 1. Available at:https://hdl.handle.net/11441/142770.

Heilmeier, A., Thomaser, A., Graf, M., and Betz, J. (2020). Virtual strategy engineer: Using artificial neural networks for making race strategy decisions in circuit motorsport. *Applied Sciences*, 10(21):7805. DOI: 10.3390/app10217805.

Heine, O. F. C. and Thraves, C. (2023). On the optimization of pit stop strategies via dynamic programming. *Central European Journal of Operations Research*, 31(1):239–268. DOI: 10.1007/s10100-022-00806-4.

Linden, R. (2008). *Algoritmos genéticos (3a ediçao)*. Ciência Moderna. book.

McKinney, W. *et al*. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX. DOI: 10.25080/majora-92bf1922-00a.

Oehrl, A. and Schaefer, P. (2025). Fastf1: A python package for accessing and analyzing formula 1 data. Available at:https://docs.fastf1.dev/. Accessed: 2025-04-02.

Rondelli, M. (2022). The future of formula 1 racing: Neural networks to predict tyre strategy. Available at:https://amslaurea.unibo.it/id/eprint/27922/.