Formula 1: What is the Best Strategy?

CIS 4930: Data Mining Proposal

Edgar Guerrero

https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

I. Introduction

Monaco, a classic, a wonderous event for any Formula 1 fan. The energy in Monte Carlo is as electric as it is any year. However, as the race progresses, thunder clouds encompass the sky, and rain surmounts the track. As a team principal and as a driver you find yourself in a situation where you must consider the strategy going forward with this race. In a track like Monaco, every decision matters, and the risk-reward equation is critical. Lewis Hamilton, famous for his prowess and innate ability to overcome obstacles must work with his team to win the race, but their balance of the steps going forward is what determines the winner of the race. Whether the leader of a pack will remain at the front of the grid or surpassed.

Formula 1, also known as F1, is the pinnacle of Motorsport. The sport includes high-speed racing and constantly improving technology on different racetracks around the world. These races are a series of Grand Prix all proving to be different challenges for drivers.

This project seeks to delve into the factors that contribute to success in Grand Prix races. Understanding what metrics variable into success; team strategy, track/circuit features, weather conditions, and driver data. The goal is to extract data-driven conclusions from the wealth of this information to then creative predictive models that provide the best strategy going into a race. The problem that teams face is how they plan to approach a race as every race brings different challenges.

Moving forward, it's important to consider that Formula 1 circuits are designed to present varying challenges to teams and drivers, including the crucial aspect of tire management. Some tracks are notorious for their high tire degradation due to abrasive surfaces and demanding layouts, while others are relatively gentler on the tires. However, the actual degree of tire wear experienced during a race can be influenced by numerous factors such as weather conditions, track temperature, and specific tire compounds chosen by Pirelli for the event.

With this in mind, we see a diverse range of tire degradation across the various circuits. High degradation tracks like Circuit de Barcelona-Catalunya and Interlagos push tire management strategies to the forefront, demanding much from both teams and drivers. In contrast, circuits like Monte Carlo, known for its low degradation, offer a different type of challenge, focusing more on precision driving and less on tire wear.

The spectrum of tire degradation levels in F1 is a testament to the sport's complexity and the varied skills required to master different tracks. From the high degradation challenges of Silverstone and Suzuka to the more tire-friendly conditions at Sochi Autodrom, each circuit brings a unique set of variables into the strategic equation of Formula 1 racing. As the sport continues to evolve, understanding and adapting to these diverse conditions remains a key aspect of competition at the highest level of motor racing.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero
https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

`

II. Literature Survey

The realm of Formula 1 is a rich field of study for data analysts and strategists alike, with the sport's success heavily reliant on the meticulous interpretation of historical data. The strategic formulation for teams is deeply rooted in both historical performance metrics and individual driver analytics. This reliance on data is due to the direct impact it has on decision-making processes, from pit stop timing to tire selection, all of which can influence the outcome of a race. This also doesn't account for all the outside forces that may influence this even further.

The scholarly pursuit within this domain has yielded a plethora of articles and research papers that delve into various aspects of Formula 1 data analysis. For instance, the paper "A Data-Driven Analysis of Formula 1 Car Races Outcome" stands out by employing correlation-based models and Principal Component Analysis (PCA) to distill the datasets into insight. Veronica Nigro also researched and produced the "Formula 1 Race Predictor" represents a compelling application of machine learning within the domain of motorsport analytics. By leveraging historical datasets, this approach aims to forecast the winner of upcoming F1 Grand Prix events, highlighting the potential of predictive models in transforming vast arrays of data into competitive insights.

Such methods allow researchers to understand the complex relationships between different variables that dictate race performance and outcomes.

Race Factors	Description
Weather	Rain can dramatically alter tire choice and pit stop strategy.
Circuit Build	Different tracks have varying impacts on tire wear and car performance.
Tire Selection	Choosing the right tire compound for the race conditions and the car's handling characteristics.
Safety Car Deployments	Unplanned safety car periods can provide strategic opportunities for pit stops.
Car Performance	The car's characteristics can influence tire wear and fuel consumption.
Tire Degradation	How quickly a tire loses performance, impacting the number of pit-stops and the choice of tire compound.

III. Key Aspects of Formula 1's Season

Formula 1, sponsored by Pirelli, utilizes a range of tire compounds suitable for various racing conditions. These include Soft, Medium, and Hard tires, each designed for different levels of grip and durability. Soft tires offer the highest grip but wear out the fastest, making them ideal for shorter stints and qualifying laps. Medium tires strike a balance between durability and grip, suitable for a range of conditions. Hard tires, while offering less grip, are the most durable and are chosen for longer race stints.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero
https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

Intermediates (Inters) and Wet tires are designed for rainy conditions. Intermediates are used in light to moderate rain, providing better grip than dry-weather tires on wet tracks. Wet tires are for heavy rain, featuring deeper grooves to disperse larger amounts of water.

Teams' strategists play a crucial role, devising race plans based on historical data, circuit characteristics, and driver preferences. They must consider several factors in their strategy. Reference to table above.

The timing of pit stops is also critical. These stops are not just about tire changes; they also involve refueling and minor adjustments. The duration of a pit stop can vary, typically lasting around 2-3 seconds, but even minor delays can have significant race implications. Pit stop strategy often revolves around the number of stops (1, 2, or 3), influenced by the above factors.

In addition to the tactical considerations during the race, the events preceding the main event are pivotal for teams and drivers. Friday practice sessions offer an invaluable opportunity for teams to gather data on tire behavior, car setup, and track conditions. These sessions are critical for teams to fine-tune their cars' balance, engine settings, and determine the most suitable tire compounds in relation to the track's demands. It's a time for strategists to assess tire degradation patterns and adjust their race strategies accordingly.

Saturday's qualifying is a high-stakes precursor to the race, where drivers push their cars to the limit to secure the best possible grid position. The outcome of qualifying has a profound influence on race strategy, as starting further up the grid can provide a strategic advantage, allowing teams to potentially opt for fewer pit stops. Conversely, a lower grid position might compel a more aggressive strategy, with teams considering alternative tire choices or pit stop schedules to make up ground. The data collected during practice sessions is crucial here, informing decisions on tire allocation and setup configurations that can make the difference between pole position and a mid-grid start, setting up for Sunday's race.

In understanding F1 strategies, it's important to recognize the interplay between these elements. The relationship between tire degradation and circuit characteristics, for instance, can dictate whether a one-stop strategy is viable or if multiple stops are necessary. Additionally, the risk of human error during pit stops adds another layer of complexity to these decisions.

For a comprehensive understanding of F1 strategies, it's essential to appreciate the dynamic nature of each race, where factors like weather, safety car periods, and even the drivers' ability to manage their tires can lead to strategic pivots. These elements contribute to the evolving nature of race strategies and the ongoing efforts of teams to optimize their approach in the face of ever-changing conditions.

Out of the 317 Grand Prix Incidents Recorded since 1993, 42% of the incidents are spread throughout a mere 6 races. These races are Belgium, Brazil, Canada, Britain, Australia, and Singapore. These incidents are reflective of the harsh conditions the tracks incur. These Grand

Prix are characterized by High Tire Degradation, Heavy Rain/Winds, or a combination of both externalities. These circumstances can be taxing on the performance of the car and pose a challenge to the team's strategy going into a race.

IV. Methodology

The analysis of the Gradient Boosting Machine model, using data from the British Grand Prix, provides a compelling narrative about the nuances of Formula 1 race strategy. The model's findings, with a Mean Squared Error (MSE) of approximately 2.625, affirm its robustness in predicting race outcomes based on various strategic elements.

```
Mean Squared Error: 2.6254515272153265
         Feature  Importance
4  Strategy Used    0.366267
0         Driver    0.358447
3     #Laps Used    0.264026
2          Tyres    0.011261
1     Grand Prix    0.000000
```

Figure 1. Gradient Boosting Machine Model

Analyzing the British Grand Prix, feature importance metrics from the Gradient Boosting Machine model highlight key influences on driver standings. 'Strategy Used' tops the list, commanding a 36.63% importance score, accentuating pit stop strategy's decisive role in race performance. This factor encompasses not just the number of stops but also the sequence of tire usage. The 'Driver' attribute, contributing 35.84% to race outcomes, indicating that a driver's skill and race-time decisions are as critical as strategy.

The significance of tire wear is evident, with 'Number of Laps Used' attributing 26.40% to placement outcomes, underscoring tire management's vital place in strategy development. The way the model calculates tires records its influence at slightly over 1% but this is recording it as an individual and not the whole picture the way Strategy Use encompasses both together as a combination. The 'Grand Prix' itself adds an element of variability, affecting tire management strategies in accordance with weather and track conditions, especially pertinent in races with high degradation.

Figure 2. T-Test

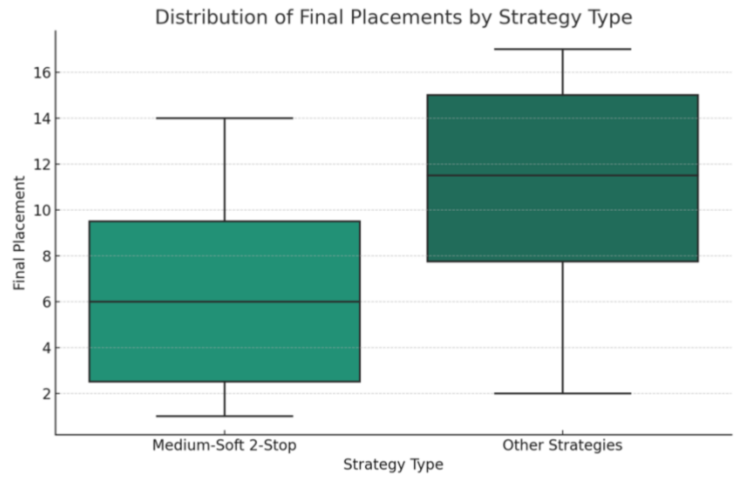
```
(-2.8054483042955205, 0.011285405772189002)
```

Statistical testing further confirms the advantage of a Medium-Soft 1-stop strategy, with a notable t-statistic and significant p-value indicating its positive impact on achieving lower (and thus better) average final placements.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero

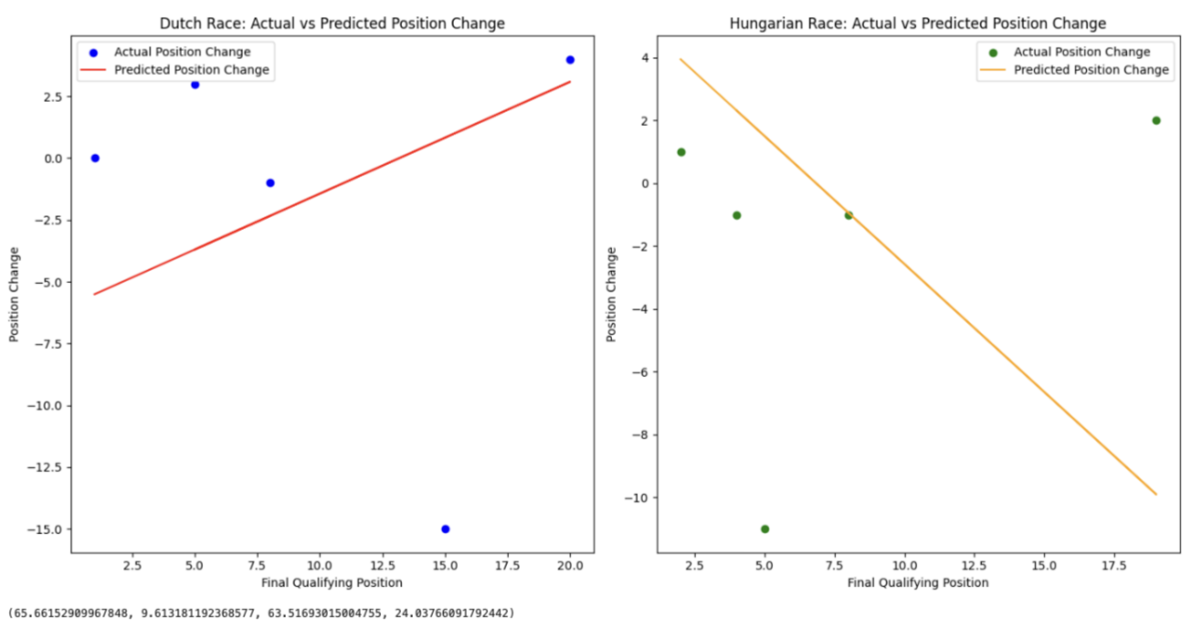
https://github.com/Edgar0618/FORMULA-1/tree/main **(Included ReadMe)**

Figure 3. Position Boxplot



For clarity, a box plot was created to show the results for the race placements comparing the 1-stop Medium-Soft strategy as opposed to all the others. This gave Redbull Racing and McLaren the edge over Ferrari who opted for the 2-stop and subsequently sacrificed many places despite similar performing cars. Despite the constructors fielding similarly performing cars, the outcomes varied remarkably, with time differences of 20 to 30 seconds, a significant margin in motorsports. This variation in race results cannot be solely explained by grid positions or driver skill; rather, it highlights the crucial role of strategy, particularly in tire selection and pit stop frequency. This underscores a data-driven strategy's value, highlighting how informed choices on tire management and pit stop timing are crucial differentiators in Formula 1's competitive theater.

Figure 5. Linear Regression



In Figure 4, the Hungarian Grand Prix demonstrated a significant level of variability in race outcomes, underscoring the critical role of strategic diversity in dry race conditions. With no rain in the forecast, teams had the freedom to explore a wider range of tire strategies and pit stop tactics. This led to an increased emphasis on overtaking, as teams could exploit a full spectrum of slick tire compounds to their advantage. In contrast, the Dutch Grand Prix, characterized by rainy weather, saw teams converging on strategies as they navigated the complexities of wet racing.

The importance of qualifying positions came to the forefront in races affected by weather, such as the Dutch Grand Prix. The unpredictable nature of rain could lock in track positions, making overtaking more challenging and elevating the significance of a strong qualifying performance. This contrast is reflected in the data, where the Dutch race exhibited a lower variance (approximately 9.61), indicating fewer shifts from grid positions. In contrast, the Hungarian race, unaffected by rain, displayed a higher variance (approximately 24.04), signifying a more fluid and strategy-dependent race dynamic.

Analyzing the numerical results, the higher variance in the Hungarian race underscores the impact of strategic flexibility. Teams and drivers could optimize their tire selections and pit strategies without rain constraints, resulting in a broader distribution of race outcomes. Conversely, the lower variance in the Dutch race highlights the criticality of qualifying in wet conditions, where strategic options are limited, and the potential for positional advancement during the race is reduced.

As a recommendation, teams struggling in rain conditions, such as McLaren, Haas, AlphaTauri, and Mercedes, should prioritize their focus on qualifying positions. Optimizing tire choices can be pivotal in maximizing position gains, especially in rain-affected races, where qualifying plays a pivotal role.

The mean squared error (MSE) for the Dutch race model is approximately 65.66, indicating the average squared difference between actual and predicted position changes. In contrast, the MSE for the Hungarian race model is approximately 63.52, showing a similar level of predictive accuracy. Variance, a measure of the spread of predictions, reveals further insights. The Dutch race has a variance of about 9.61, indicating relatively consistent predictions around the mean. On the other hand, the Hungarian race exhibits a higher variance of approximately 24.04, suggesting a wider range of predicted position changes and a greater degree of unpredictability.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero
https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

The accompanying graph visually illustrates the relationship between predicted and actual position changes for both races. It reinforces the notion that the Hungarian race is characterized by greater unpredictability, with a wider spread of points compared to the Dutch race.

Figure 5. Random Forest Model

The 2023 Monaco Grand Prix, characterized by low tire degradation and the complexity of the track, demonstrated the critical nature of strategic pit stops in Formula 1 racing. The race highlighted how teams like Aston Martin and Alpine, despite starting from similar grid positions and possessing similar car capabilities, saw divergent results due to their pit stop strategies. Aston Martin's early pit stop maneuver allowed them to gain a strategic edge, while Alpine's extended tire stint did not yield the expected advantage, particularly when rain introduced unpredictability to the race dynamics.

```
data = pd.read_csv(file_path)

combined_data = data.copy()
combined_data['2023 Driver and Pits'] = label_encoder.fit_transform(combined_data['2023 Driver and Pits'].astype(str))
combined_data['2023 Tyres'] = label_encoder.fit_transform(combined_data['2023 Tyres'].astype(str))
combined_data['2021 Driver and Pits'] = label_encoder.fit_transform(combined_data['2021 Driver and Pits'].astype(str))
combined_data['2021 Tyres'] = label_encoder.fit_transform(combined_data['2021 Tyres'].astype(str))
combined_data['2023 Race Time Wet'] = combined_data['2023 Race Time Wet'].apply(convert_time_to_seconds_v2)
combined_data['2021 Race Time Dry'] = combined_data['2021 Race Time Dry'].apply(convert_time_to_seconds_v2)
combined_data['Race Time'] = combined_data[['2023 Race Time Wet', '2021 Race Time Dry']].mean(axis=1, skipna=True)
combined_data.dropna(subset=['Race Time'], inplace=True)


features = combined_data[['2023 Driver and Pits', '2023 Tyres', '2021 Driver and Pits', '2021 Tyres', '2023 #Laps Used', '2021 #Laps Used']]
target = combined_data['Race Time']
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)
random_forest_model = RandomForestRegressor(n_estimators=100, random_state=42)
random_forest_model.fit(X_train, y_train)
y_pred = random_forest_model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
feature_importance = random_forest_model.feature_importances_

mse, feature_importance
```

```
(16679.363114600583,
 array([0.03343481, 0.03688504, 0.07793881, 0.04096524, 0.72232218,
        0.08845393]))
```

The Random Forest model's analysis of feature importance's indicates that the number of laps on which tires were used in 2023 ('2023 #Laps Used') was the most significant predictor of race times, pointing to tire strategy as a key factor in race outcomes. The decision to use a combination of hard, intermediate, and medium tires in a two-stop strategy appears to be optimal in such conditions.

In contrast, the 2021 race was largely influenced by car performance and driver ability, with most teams opting for a one-stop strategy, choosing a soft-hard tire combination. Strategy played a less critical role due to fewer variables affecting the race.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero
https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

The data from 2023 emphasizes the significance of a well-planned strategy, especially in adapting to changing conditions such as post-rain track drying. The selection of tires and the timing of pit stops, especially in a race where tire wear is not the primary concern, are shown to be crucial for achieving competitive race times. The difference in finish times between the winners of the 2021 and 2023 races further illustrates the evolving strategic challenges in Formula 1 and the impact of well-executed decisions.

Feature	Importance
2023 Driver and Pits	3.34%
2023 Tyres	3.69%
2021 Driver and Pits	7.79%
2021 Tyres	4.10%
2023 #Laps Used	72.23%
2021 #Laps Used	8.85%

VI. Conclusion

The analysis of Formula 1's 2023 Monaco Grand Prix has elucidated the profound impact of tire strategy and pit stop timing on race outcomes, particularly under the fluctuating conditions of the track. Insights from the Random Forest model reveal the Laps Used as a key determinant, underscoring the value of tire longevity and the strategic timing of pit stops. They reflect the necessity for teams to dynamically adjust their strategies to the unpredictable elements such as weather changes. The support for a two-stop strategy with a variety of tire compounds emerges as a preferred tactic in scenarios of low tire degradation, offering teams the flexibility to adapt to evolving race conditions.

Additionally, the Gradient Boosting model applied to the British Grand Prix data has been instrumental in highlighting the importance of various features that contribute to race outcomes. This model has allowed for a nuanced understanding of the interplay between different strategic variables. Moreover, a Linear Regression analysis comparing the Dutch and Hungarian Grand Prix demonstrated the variance in race outcomes due to qualifying positions, emphasizing the different strategic approaches necessitated by specific track characteristics. Comparing the 2023 and 2021 Monaco races through the lens of the Random Forest model has provided a clear narrative on the evolution of F1 race strategy. It contrasts the significant reliance on car performance and driver skill in 2021 with the more complex strategic environment in 2023 where data-driven decisions and telemetry played a pivotal role.

In essence, this multifaceted analysis, integrating telemetry data and machine learning models, has been pivotal in offering deeper strategic insights. The ability to leverage real-time car performance and track condition data refines strategies from reactive to proactive, enabling teams to swiftly navigate the ebb and flow of a race's tactical landscape. The combination of telemetry, Gradient Boosting, and Linear Regression models, along with the Random Forest analysis, underscores the increasing complexity and critical importance of strategic planning in Formula 1. The adept use of these analytical tools is essential for teams to make informed decisions that can lead to decisive advantages on the racetrack.

Formula 1: What is the Best Strategy?
CIS 4930: Data Mining Proposal
Edgar Guerrero
https://github.com/Edgar0618/FORMULA-1/tree/main (Included ReadMe)

References

1. F1CFA. "Formula 1 Tyres Statistics". Available at:

2. For 2023 Monaco GP: https://www.f1cfa.com/f1-tyres-statistics.asp?t=2023&gpn=Monaco
3. For 2021 Monaco GP: https://www.f1cfa.com/f1-tyres-statistics.asp?t=2021&gpn=Monaco
4. For 2023 Hungary GP: https://www.f1cfa.com/f1-tyres-statistics.asp?t=2023&gpn=Hungary
5. For 2023 Netherlands GP: https://www.f1cfa.com/f1-tyres-statistics.asp?t=2023&gpn=Netherlands

6. Nigro, V. (Date Published). "Formula 1 Race Predictor". Towards Data Science. Available at: https://towardsdatascience.com/formula-1-race-predictor-5d4bfae887da

7. Steveau5. "Formula 1 Big Data Analysis". GitHub repository. Available at: https://github.com/stevau5/Formula-1-Big-Data-Analysis/tree/master

8. Trotman, J. (Date Updated). "Formula 1 Race Events Datasets". Kaggle. Available at: https://www.kaggle.com/datasets/jtrotman/formula-1-race-events