ISYE 412 Final Project

F1 Pit Stop Analysis

Najla Alkhathlan, Avery Finck, Zhaoqi Liang, Praharshith Jamalapuramn

I. Introduction: background of the dataset

The dataset chosen is called “Formula 1 World Championship (1950-2024)”. It includes 14 (.csv) files detailing many topics related to Formula 1. The different csv files include info on races, drivers, constructors, qualifying, circuits, lap times, pit stops, seasons, sprint results, and status. Each separate data file has a large number of attributes of all types, including nominal, binary, ordinal, numeric. Many of the files can be joined through an attribute like “circuitid” or “raceid”. It has a plethora of data that will be useful when conducting our analysis. The dataset encompasses comprehensive records covering everything from races, drivers, and constructors, to qualifying rounds, circuits, and championship outcomes spanning from 1950 to the current 2024 season. Moreover, the data is meticulously updated and curated, ensuring high reliability for predictive analytics and historical trend analysis.

Due to the high volume of data, for the sake of ease and simpler analysis, we have shortened the scope of years to the last two decades, i.e., from 2004 to 2024 in our analysis.

II. Objective: why do you analyze this dataset and what kind of questions that you would like to answer from the dataset.

The primary objective of analyzing this dataset was to address the question: "What defines a good pit stop strategy?" This inquiry holds interest not only for Formula 1 enthusiasts but also for drivers, team members, constructors, and team owners, where an effective pit stop strategy could be the key to winning or losing a race. Given the broad nature of this question, four specific questions were formulated to guide the analysis toward a conclusion. These include: "Do the duration of pit stops impact race outcomes?", "What are the key factors contributing to increasingly faster lap times across different circuits?", "Is there a connection between the number of pit stops, their timing, and the winner?", and "How have technological advancements in car design and engineering influenced the performance of drivers and constructors over the decades?" Although not all these questions directly address pit stop timing or duration, the analysis of these facets was integral in considering the overall pit stop strategies.

III. Your approaches to answer the identified questions and visualizations

Before analyzing each of the questions defined above, the data set was exported from Kaggle.com as a CSV file and cleaned using RStudio. It was then uploaded into Tableau software in order to analyze it.

A. Do duration of pit-stop times impact race outcomes?

To answer the question, “Do duration of pit-stop times impact race outcomes?”, first the files pit_stop.csv, results.csv, races.csv, and circuits.csv were loaded into Tableau and connected by driver_id, race_id or both. The first graph below shows a visualization of location of each circuit.

Location point was created using Longitude and Latitude attributes included in the circuits.csv file. Each color represents a different circuit and the size of the point shows the number of races we have data for over each circuit using count(circuit_id) from the races.csv file. The map can be seen below:

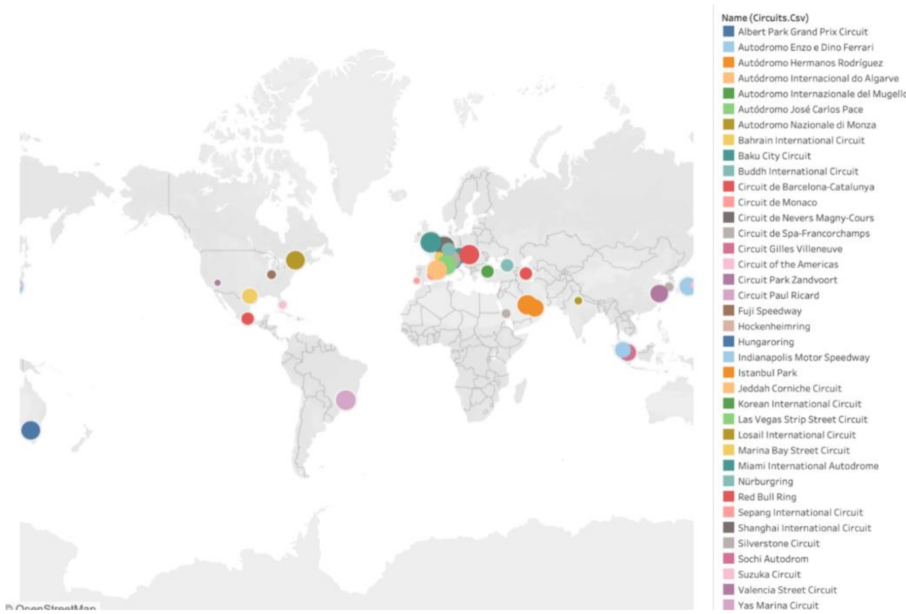

Figure 1: *Location of each circuit*

This graph does not lead to any conclusions, however, it is an interesting visualization for people who may not be familiar with F1 and gives them an understanding of F1 circuits, the amount of races there, and where in the world they are.

The next visualization created is a horizontal bar chart illustrating the average total pit stop time on each circuit. This can be interpreted as summing the total time that each driver spends in pitstop and taking the average of that. The visualization for each circuit can be seen in the chart below:

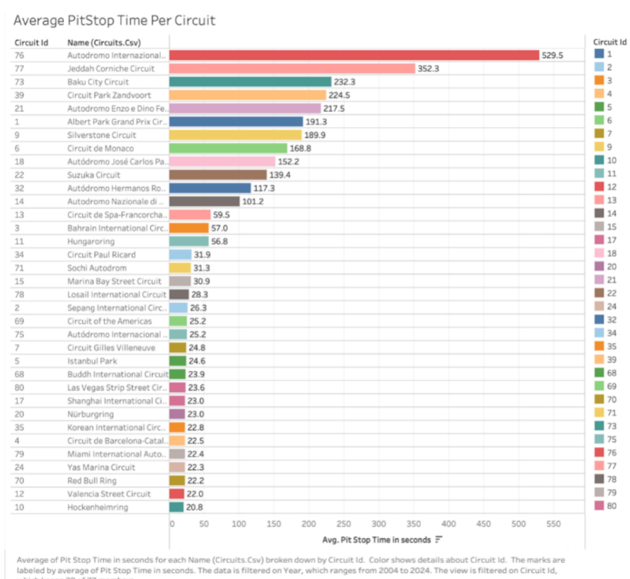

Figure 2: *Horizontal bar chart illustrating the average total pit stop time on each circuit*

The chart displays circuit ID, circuit name as well as average pit stop time in seconds to the right of the corresponding bar. It is sorted discerningly, the highest average pit stop time is at the top while the lowest is at the bottom. It can be observed that there is a discrepancy between the highest 15 average pit stop time and the lowest average pit stop time. This prompted the exploration into the 15 circuits with the highest average time versus the lowest.

The next graphs created were Pit Stop Time (duration) vs. Driver's Average (Finishing) Position Order. This was separated into two scatter plots, the top 15 circuits on the left and the bottom 20 circuits on the right, based on average pit stop time as discussed above. The x-axis is the total pitstop time of each driver in seconds on each different circuit and the y-axis is the inverted average order position. The dots are color coded on a green to orange scale. Green being a lower average finishing position order (better) and orange/red being higher (worse). The graphs can be viewed below:

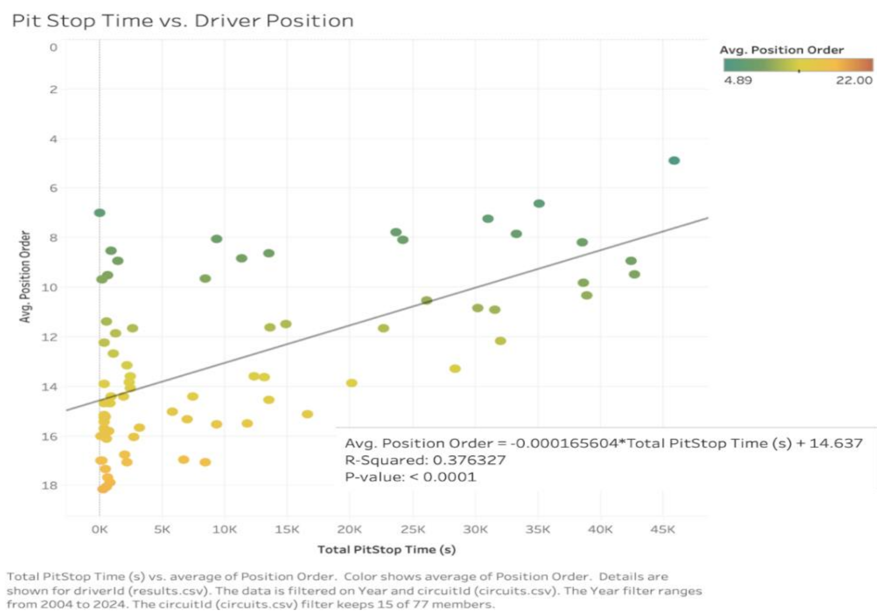

Figure 3: *Pit Stop Time (duration) vs. Driver's Average (Finishing) Position Order (a)*

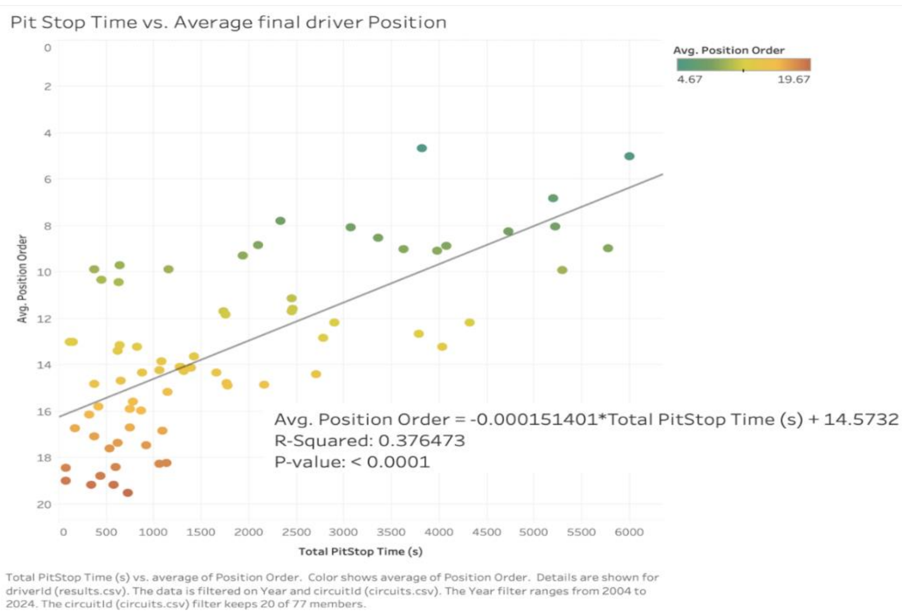

Figure 4: *Pit Stop Time (duration) vs. Driver's Average (Finishing) Position Order (a)*

As can be seen in the trendlines on each graph above there seems to be some positive correlations between higher total pitstop time and finishing early (lower average position order). It can also be observed that data is highly concentrated in the lower end of seconds. Meaning that people finish at all different positioning with relatively similar total pitstop time. However, when looking at the trend lines, both have an R-Squared of around 0.38. This is acceptable for social sciences, but is still on the lower end. This, along with a very low P-value leads to the conclusion the total pitstop duration does not explain a significant amount of variation in the average position order of a driver for all circuits.

In order to further confirm this conclusion correlation between total pit stop time and position order was added as a calculated variable. The visual representation of this can be seen below:

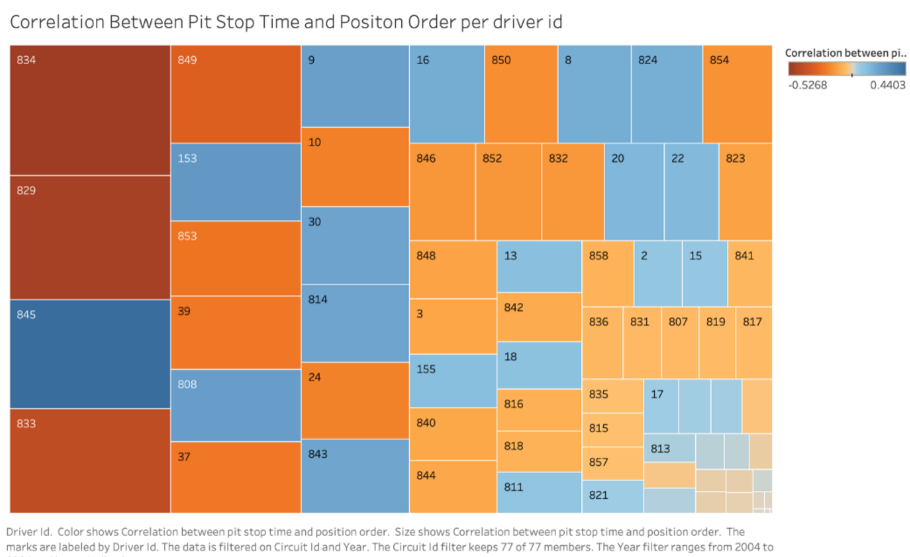

Figure 5: *Correlation between total pit stop time and position order*

This visualization is on an orange-blue color scale. The more negative theFigure 5o create the size of each Driver Id box. It can be seen from this graph that a majority of the drivers have negative correlation, in this case meaning the more pit stop time, the lower they finish in final position order. This is consistent with the scatter plots above, however, some drivers are positively correlated. Both position and negative correlations are nowhere near negative or positive ones. This again reaffirms that we cannot draw a definite conclusion on if pitstop duration affects finishing position.

B. Is there a connection between the number of pit stops, when it was taken, and the winner?

To effectively analyze the relationship among the number of pit stops, their timing, and race outcomes, data from various sources including pit_stop.csv, results.csv, races.csv, and circuits.csv was utilized. These datasets were integrated using race_id and driver_id as key connectors in Tableau for visualization purposes.

The initial analysis focused on exploring how the average number of pit stops correlates with drivers' finishing positions. As shown in Figure 6, a scatter plot was created where the average number of pit stops was plotted against position order, with winners highlighted in a distinct color. The analysis revealed that drivers within the top 10 positions generally took two pit stops. Contrary to the assumption that fewer pit stops could save time and enhance ranking, the data indicated that drivers with only one pit stop typically finished in lower positions.

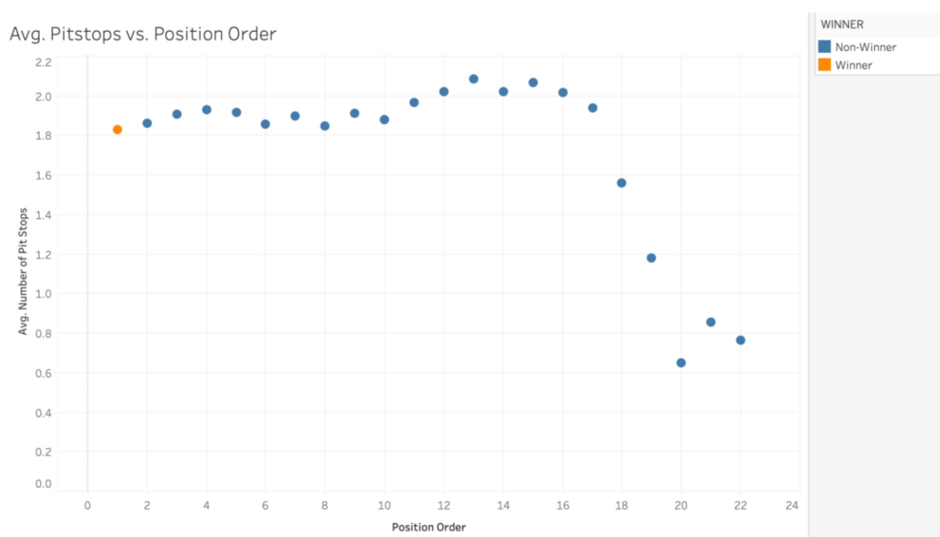

Figure 6: *Graph representing the connection between pit stops and position order*

Further analysis examined whether the number of pit stops is a decisive factor in winning races. Figure 7 illustrates this by mapping pit stop laps against position order, with race ID differentiated by color, focusing on recent races on circuits of similar length. The size of the markers indicated the winners, showing that most drivers follow a consistent pit stop strategy, yet the winner varies. For instance, at the British Grand Prix, the predominant strategy was two pit stops—around the 20th lap and near the 40th lap—highlighting that while strategic pit stops are crucial for a higher ranking, they do not solely determine the race winner.

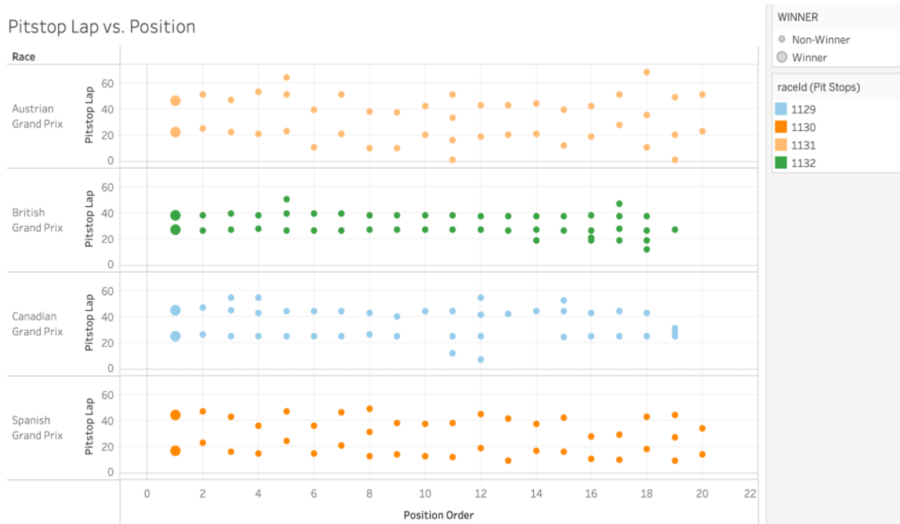

Figure 7: *Graph representing the connection between pit stop lap and position order*

Additional investigation assessed the link between the average number of pit stops and the circuit length, represented in a bar plot (Figure 8). Data on circuit length was sourced externally and appended to a spreadsheet that included race ID, circuit name, and pit stop data. The plot displayed circuit names against their respective lengths, with the average number of pit stops indicated by varying color intensities, where darker shades represented more frequent stops. This analysis concluded that circuit length does not significantly influence the number of pit stops. For example, despite the Belgian Grand Prix being one of the longest circuits, it did not necessitate more pit stops than the shorter Dutch Grand Prix, which had a higher average number of stops.

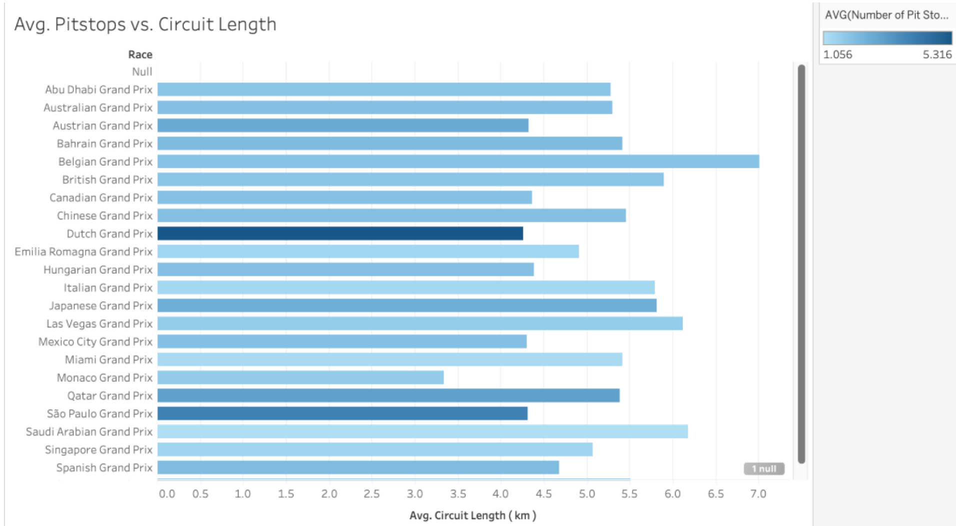

Figure 8: *Graph representing the relation between circuit length and average number of pit stops*

In conclusion, the detailed visualizations highlight the multifaceted nature of racing strategies, illustrating that while pit stop frequency and timing are crucial, they are just one of many variables that collectively determine the final race standings. This comprehensive analysis emphasizes the complexity of decision-making in motor sports, where strategic choices about pit stops must be balanced with other critical factors such as circuit characteristics and race conditions to optimize performance and achieve victory.

C. What are the key factors contributing to increasingly faster lap times across different circuits?

In order to summarize the key factors that contribute to faster lap time, data from various sources including lap_time.csv, results.csv, races.csv, drivers.csv, constructors.csv and circuits.csv were utilized. These data files are all linked in Tableau with lap_time.csv as the central file, making it convenient to create data visualizations.

This analysis explores the trends in lap times for various constructors over the years. Figure 9 depicts the average lap time (in milliseconds) plotted against the year, with individual lines representing different constructors. A notable trend is observed starting in the early 2000s, where there is a significant drop in lap times, particularly around 2001. The overall downward trend during this period is attributed to advancements in vehicle technology and improvements in strategy optimization. Following this steep decline, lap times continue to gradually decrease in the subsequent years. Interestingly, the earlier period (1995–2000) shows relatively consistent and higher lap times, which could be attributed to limitations in technology. For instance, in 2000, there were restrictions placed on the overall model design of F1 cars, which likely contributed to these spikes. In addition, all constructors have generally the same performance.

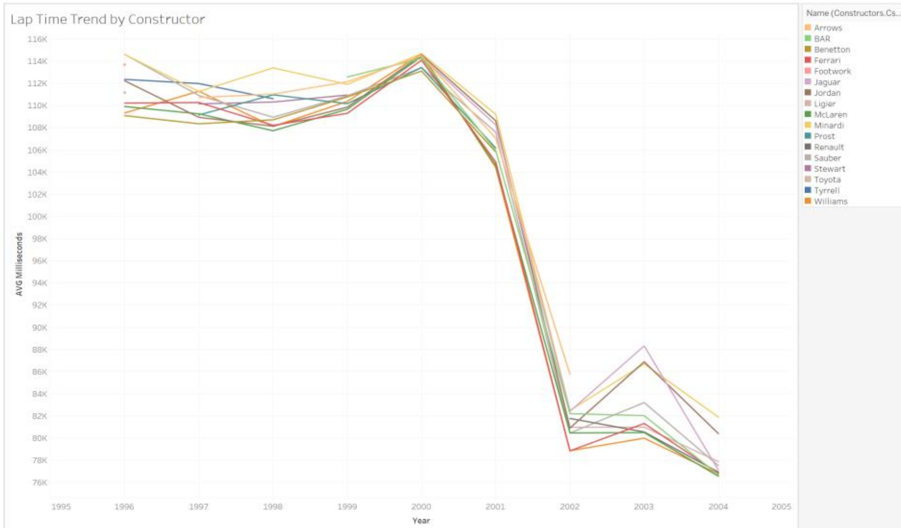

Figure 9: *Historical Lap Time Trends by Constructor (1995–2004)*

Figure 10 illustrates the average lap times (in seconds) for various drivers across different circuits. Most drivers exhibit consistent lap times between 75 to 200 seconds, reflecting similar performance across the majority of circuits. However, there is a significant outlier with a sharp spike in lap times at one circuit, reaching approximately 632 seconds. This spike could be attributed to unique circuit characteristics, such as an accident, adverse weather conditions, or the inclusion of non-racing laps. This outlier has little impact on the overall trend, so it can be concluded that the overall trend across circuits shows relative uniformity, indicating that differences in lap times are more likely influenced by circuit design and race conditions rather than driver performance.

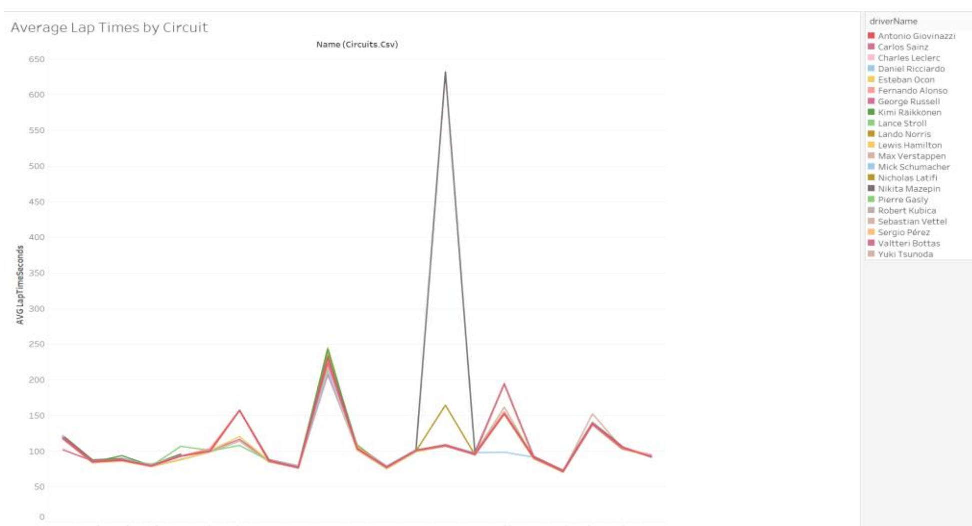

Figure 10: *Driver Performance Trends: Average Lap Times by Circuit*

The next visualization created is the average lap times of each driver for every lap, calculated by averaging multiple performances. As shown in Figure 11, each line represented in a different color, corresponds to a specific driver. When observed across laps, the lines generally show a downward trend, indicating improving lap times as the race progresses. Although there are a few noticeable spikes in the middle section of the chart, these peaks do not disrupt the overall downward trend. This downward trend aligns with common knowledge: as the race progresses, the fuel in the F1's tank decreases, reducing the overall weight of the car and allowing it to go faster. This visualization highlights the dynamics of lap-by-lap performance and the factors influencing it.

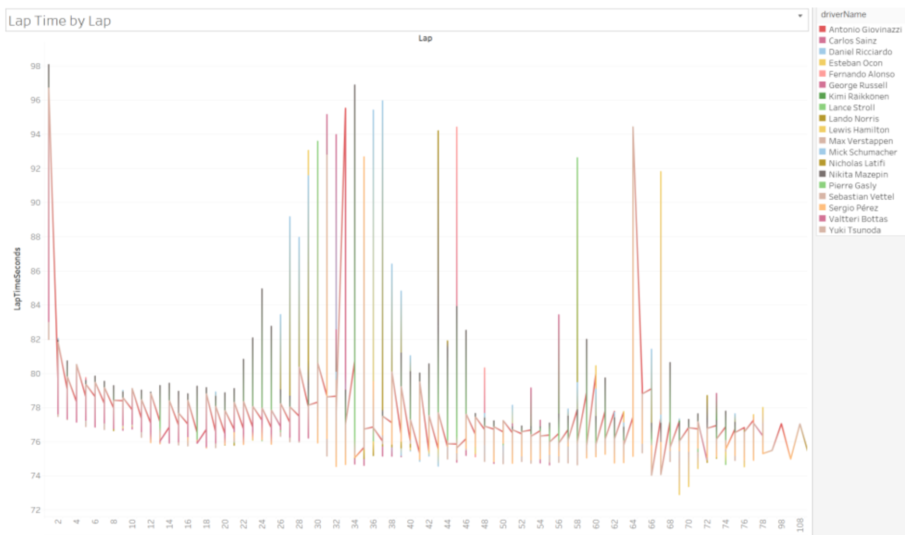

Figure 11: *Lap Time Trends by Lap for Each Driver*

Figure 12 builds on the previous analysis by examining when the fastest lap occurs during a race as a percentage of the total laps completed. From the visualization, it is evident that the fastest laps are most likely to happen between 70% and 80% of the race distance, with a more precise average of around 74%. This trend is largely influenced by two key factors: the cars are lighter at this stage of the race due to reduced fuel loads, which allows them to achieve better performance, and teams often adjust their strategies based on the race standings. For example, teams may prioritize setting the fastest lap to secure an additional championship point if the opportunity aligns with their position and goals during the race.

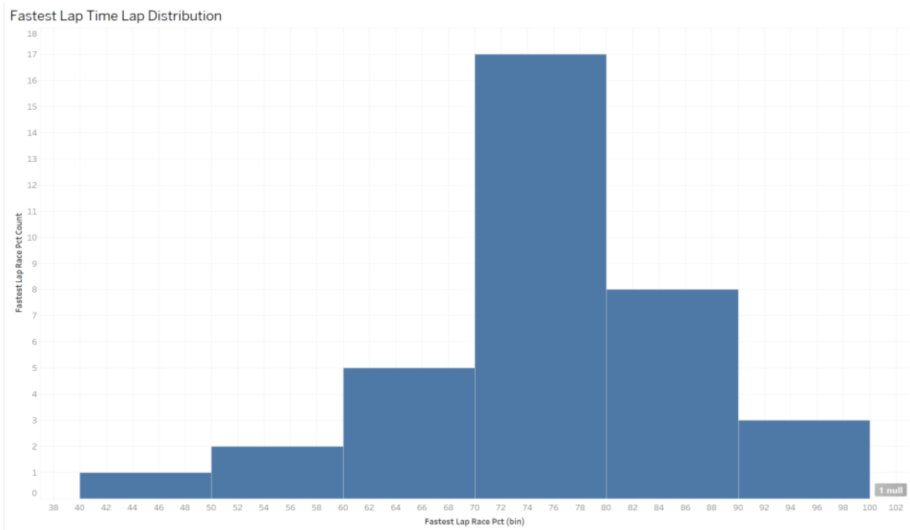

Figure 12: *Distribution of Fastest Lap Times as a Percentage of Total Laps by Circuit*

In summary, the factors influencing the fastest lap can be understood from the previous visualizations. First, on a broader time scale, both drivers and constructors exhibit a consistent overall trend in the reduction of fastest lap times as the years progress. However, the type of circuit plays a significant role in affecting lap times, with different tracks contributing to variations in performance. Due to lighter fuel loads as the race progresses and strategic decisions made by teams, within the course of a race, the fastest laps are typically recorded between 70% and 80% of the race's total distance.

D. How have technological advancements in car design and engineering impacted the performance of drivers and constructors over the decades?

This question focuses on the correlation between over performance of the teams with the changes in technology and car design. In order to tackle this, the approach taken was rather through correlative observations of different constructors who have been known to be leaders in innovation and dominating the F1 motorsport landscape. Figure 13 illustrates the data branches that have been considered in order to analyse the question.

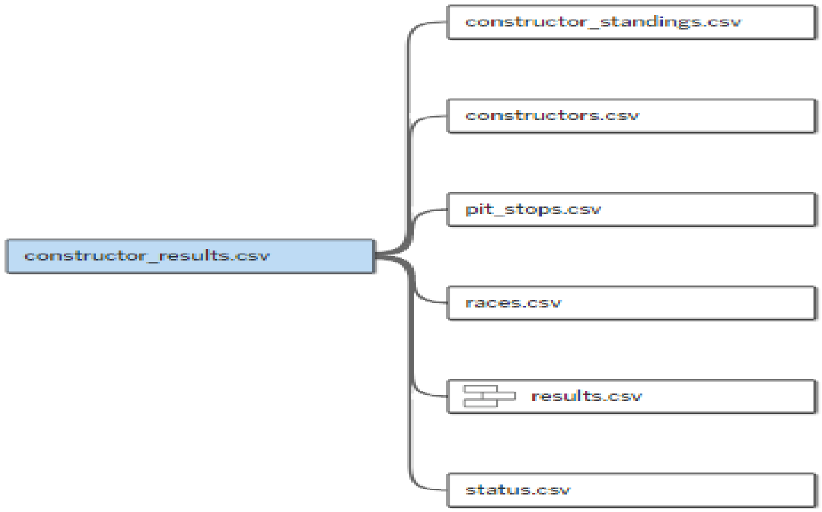

Figure 13: *Connections of datasets for the analysis of constructor performance*

The initial study was conducted on the trends of top constructors every season for the last two decades. It showed a promising trend of top constructors maintaining their dominance over others for consecutive seasons. Such as AMG Petronas Mercedes has been a pioneer of innovation and optimised car design for 7 seasons prior to 2021, where RedBull Racing has overtaken the former leaders with better car design and skillful drivers. Figure 14 & 15 maps the points scored by the teams over years which help draw the conclusion that the best overall performing car will most likely win the race as well as the constructor's championship.

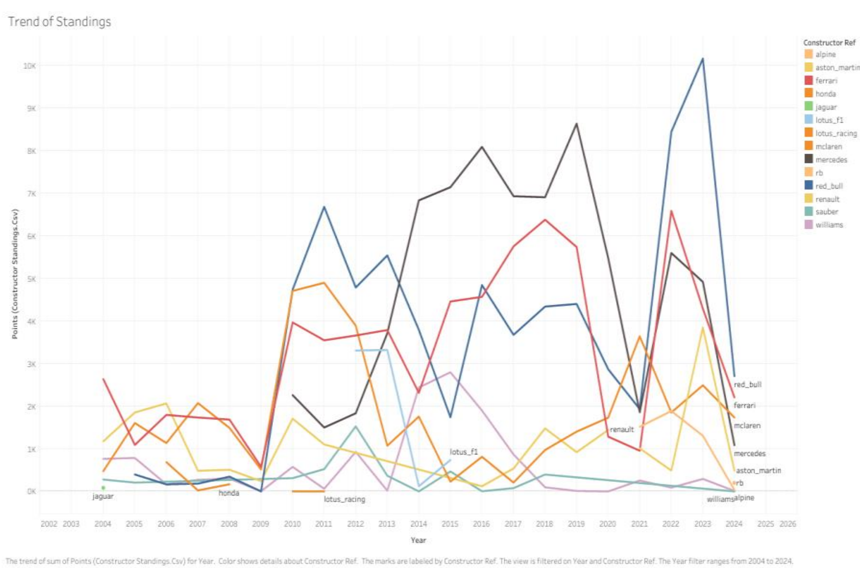

Figure 14: *Constructors Championship Trends: Over the last 2 decades*

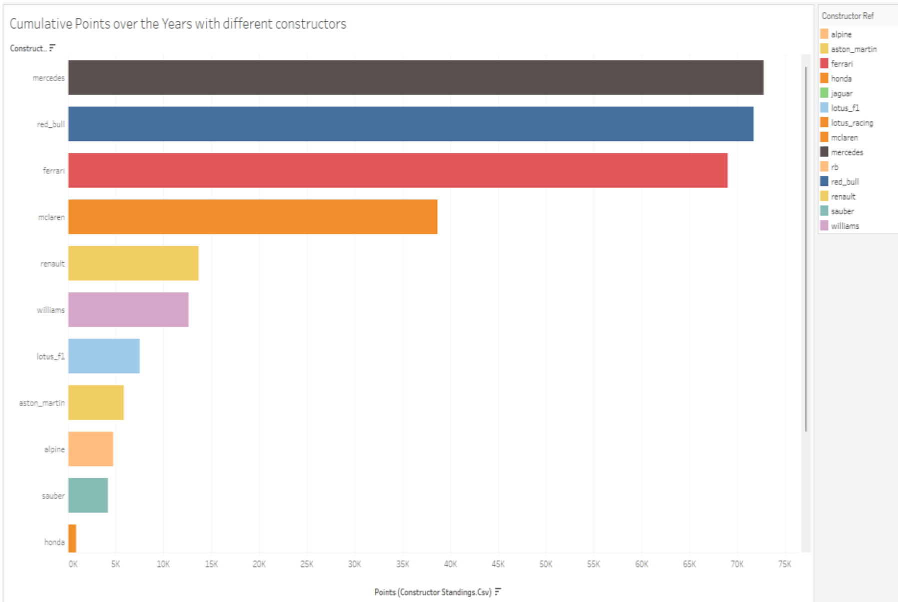

Figure 15: *Top Constructors: Points accumulated per team over the last 2 decades*

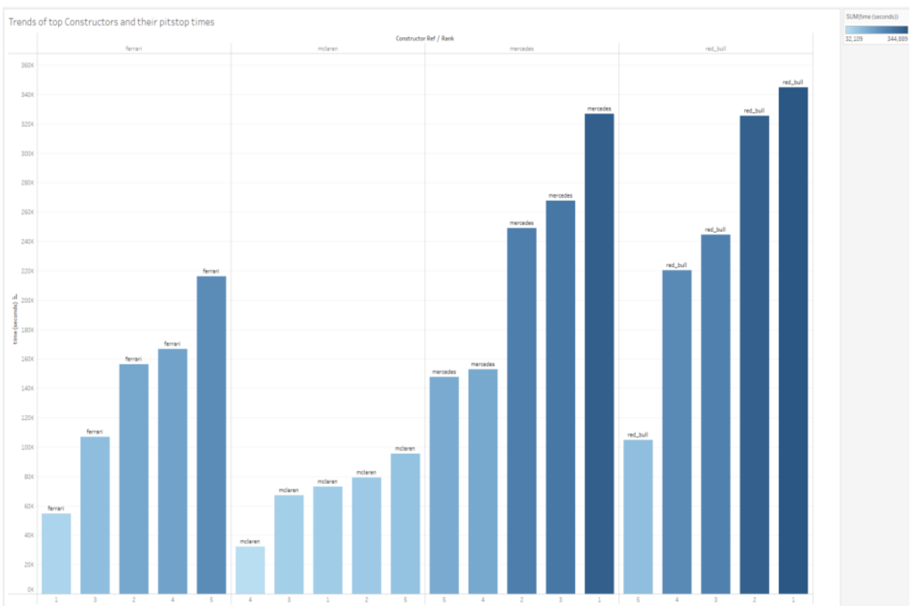

Figure 16: *Top 4 Constructors: Time spent in pitstop against their final positions in the grand prix*

For further understanding of the correlation between pitstop and team performance. The trend of top 4 constructors against their pitstop as well as their final position was observed. The initial hypothesis was that the faster the car, the faster the pitstop will be. However, a mixed set of observations were seen, where the hypothesis held true for teams such as Scuderia Ferrari and McLaren Racing, where the least cumulative amount of time spent in pitstops led to higher position at the end of the race. However, it was also observed that the inverse is just as true for the recently track dominating constructors such as RedBull Racing & AMG Petronas Mercedes. The more time they spent in the pitstop yielded them better results. Therefore, a conclusive understanding of the hypothesis was not observed. This could also be caused by the pitstop crew's varying performance that led to higher times, which were then compensated by the car's & driver's performance. Figure 16 delves into the same.

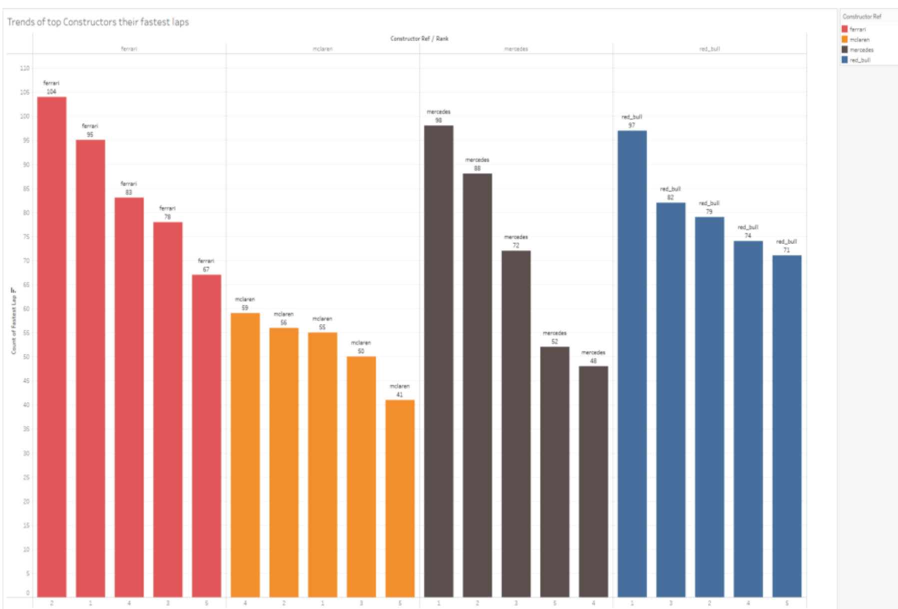

Figure 17: *Top 4 Constructors: The cumulative number of fastest laps*

In order to further analysis the performance statistics against the championship winning constructors, the number of fastest laps with the final position on the grid was mapped for the top 4 constructors. The hypothesis was that the fastest car is to always score the fastest lap and the highest place on the grid. It was observed that the top 2 dominant constructors did indeed have the highest number of fastest laps in the best grid position or often the race winning positions. However, it was also noted that the Scuderia Ferrari and McLaren Racing had a good number of fastest laps at different race positions by the end of the race. Which leads to speculations of presence of external factors at play that influence their performance at fastest laps, such as Tire conditions, track conditions, DNFs & Driver's performance. Figure 17 shows the trends of top constructors' fastest laps against their finishing positions.

Therefore, in conclusion, it can be understood that the teams that utilise technological advancements more efficiently are more likely to perform better and manage better pitstop times

& race statistics than their opponents. However, the presence of other factors such as Driver's Performance, Team Strategies, Pit Crew Performance, Tire Degradation Rate, Weather Conditions and Unforeseen accidents impact the outcome/performance of the constructors in significant measure that an optimized pitstop alone cannot assure victory.

IV. Summary and Comments

Overall, it is hard to say what defines a good pitstop strategy as it varies from circuit to circuit, race to race, constructor to constructor, and driver to driver. Some key factors that were not available in the dataset include Circuit conditions, weather, aerodynamic efficiency of the car and tire degradation rate. For the fastest lap time, it tends to decrease over the years and is closely related to strategy and F1 regulations. During a typical race, the fastest lap occurs around 74% of the race. It may be more beneficial to look at it from the perspective of one driver or team rather than the data as a whole. It is also beneficial to observe that the best pitstop strategy does not always equate to the best final position on the grid, as even the non-optimally performing cars with optimal pit crews can have the shortest pit stop, which was not noted in the dataset. This is a lesson learnt that shall be applied into future data analysis, in order to encourage out of box, or out of dataset thinking in order to understand the non-linear & human factors that are involved in the scenario.