Analysing the Race Strategies of Charles Leclerc and Sergio Perez
In the 2022 Abu Dhabi Formula 1 Grand Prix

COSC3000 – Data Visualisation Report

Johanes Steven - 47282726

I. Introduction

- **Dataset**

The dataset used in this project was taken from Kaggle, titled 'Formula 1 World Championship (1950-2023)'. The dataset contains several .csv files, each containing information regarding every annual season of Formula 1 that has happened. The dataset that we are most concerned about in this project will be the drivers' lap times during the final race of the 2022 season, which is the 2022 Abu Dhabi Grand Prix.

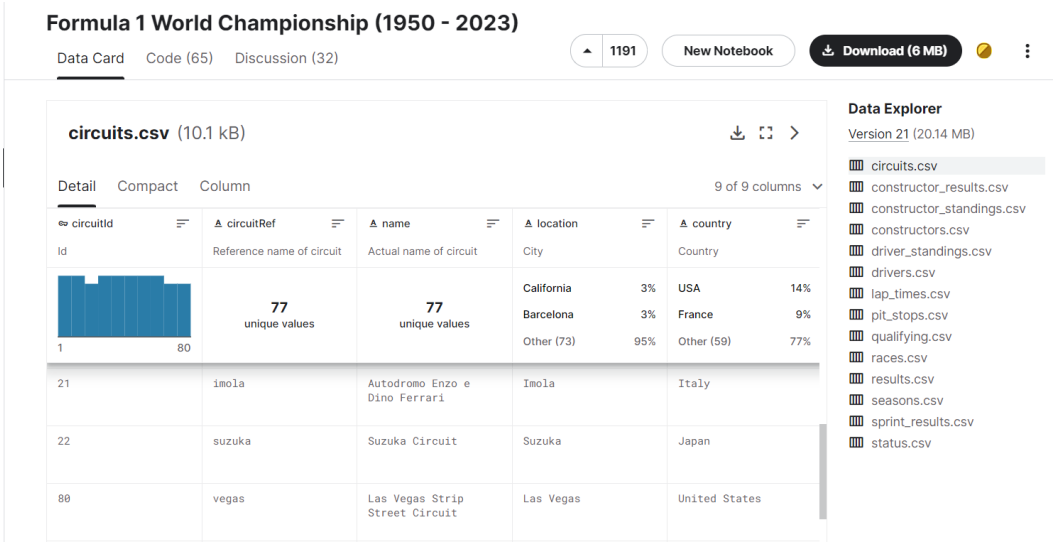

Figure 1.1. Kaggle Page for the Dataset

- **Project Aims**

This project aims to visualise and analyse the tyre strategies utilised by Ferrari's Charles Leclerc and Red Bull Racing's Sergio Perez. During the 2022 Abu Dhabi Grand Prix, Leclerc pitted once, changing his tyres from the Medium tyres to the fresher, albeit 'slower' Hard tyres. Perez pitted twice, taking an additional Hard tyre at the end of a Medium-Hard-Hard tyre strategy.

What is happening here is that Ferrari is gambling that their driver Leclerc would be able to drive a car on heavily used tyres all the way to the end of the race. They will not lose an additional approximate 22 seconds needed for a tyre change in the pit stop, but the driver will not be able to lap quickly as the drivers with fresher tyres. This gamble ultimately worked, as Leclerc was able to finish in second place, ahead of Perez who would end up in third.

This project will then go through factors that could have happened to analyse just why Perez was unable to finish ahead of Leclerc in what many people would believe in the faster Red Bull car, as proven by Max Verstappen in the other Red Bull car who ended up qualifying and finishing the race in first place.

- **Motivation**

Race strategy is an important part of Formula 1, as it could end up influencing where a driver finishes the race. In Formula 1, only the top-10 drivers will gain points, with first place netting 25 points, second place 18 points, third 15 points, all the way to tenth place, which will net the driver a single point. Every single point counts in Formula 1, and each team will fight tooth-and-nail for them. This is because of the cash prize at the end of the season, where hundreds of millions of dollars are on the line.

For additional context, both Alfa Romeo and Aston Martin finished the season with 55 points, but Alfa Romeo finished ahead in sixth due to them getting better results in the race, which results in the tiebreaker lying in favour of them instead of Aston Martin. This resulted in Alfa Romeo gaining an approximate $10-12 million extra prize money. Therefore, a single point could be the deciding factor between winning millions of dollars more than your rivals or not, which is why teams put so much emphasis on the importance of race strategy.

Race strategies allow drivers with a slower car to fight with those with faster ones, as shown by the 2022 Abu Dhabi Grand Prix, where Leclerc in the slower Ferrari was able to stay ahead of Perez in the Red Bull. As such, this project will aim to dig deeper on why Ferrari's one pit stop strategy was better than Red Bull's two-pit stop strategy. By understanding why certain strategies work better than others, it will be possible to learn from past mistakes, so that future race strategies will yield more ideal results, which will net the driver and the team more points, and more prize money as a result.

II. Methodology

- **Data Cleaning**

The data cleaning was done using Jupyter Notebook, a very well-known software for data science purposes. By using Pandas, I was able to put the relevant .csv files into dataframes, which I will be able to manipulate using Python codes. From all the provided data, I want to pull out all the lap time information for the 2022 Abu Dhabi Grand Prix, which is easily done using Python.

```
In [2]: import pandas as pd

        lap_times = pd.read_csv('lap_times.csv')
        drivers = pd.read_csv('drivers.csv')

        abudhabi = lap_times.loc[lap_times['raceId'] == 1096]

        abudhabi
```

Out[2]:

	raceId	driverId	lap	position	time	milliseconds
537004	1096	830	1	1	1:32.198	92198
537005	1096	830	2	1	1:30.541	90541
537006	1096	830	3	1	1:29.968	89968
537007	1096	830	4	1	1:30.368	90368
537008	1096	830	5	1	1:30.502	90502
...	...	...	...	...	...	...
538116	1096	822	53	16	1:32.998	92998
538117	1096	822	54	16	1:32.995	92995
538118	1096	822	55	16	1:31.236	91236
538119	1096	822	56	15	1:30.566	90566
538120	1096	822	57	15	1:30.743	90743

1117 rows × 6 columns

Figure 2.1. 2022 Abu Dhabi GP 2022 Lap Times

What I did next was to link the driverId field to their actual names, so that I know which driver put in which lap times. Below is the resulting dataframe:

```
In [3]: match_drivers = drivers[drivers['driverId'].isin(abudhabi['driverId'])][[ 'forename', 'surname', 'numbe
        match_drivers = match_drivers.merge(abudhabi[['driverId', 'lap', 'position', 'time', 'milliseconds']],
        abudhabi2022 = match_drivers.sort_values(['lap', 'position'], ascending=True)
        abudhabi2022[:10]
```

Out[3]:

	forename	surname	number	code	driverId	lap	position	time	milliseconds
427	Max	Verstappen	33	VER	830	1	1	1:32.198	92198
197	Sergio	Pérez	11	PER	815	1	2	1:33.251	93251
659	Charles	Leclerc	16	LEC	844	1	3	1:34.192	94192
0	Lewis	Hamilton	44	HAM	1	1	4	1:34.864	94864
485	Carlos	Sainz	55	SAI	832	1	5	1:35.436	95436
717	Lando	Norris	4	NOR	846	1	6	1:35.908	95908
775	George	Russell	63	RUS	847	1	7	1:36.506	96506
543	Esteban	Ocon	31	OCO	839	1	8	1:36.964	96964
82	Sebastian	Vettel	5	VET	20	1	9	1:37.455	97455
55	Fernando	Alonso	14	ALO	4	1	10	1:38.193	98193

Figure 2.2. Drivers' Lap Times Information

This dataframe will serve as the main source for our visualisation, as such I will put it in a separate .csv file which is also possible using Python.

```
In [4]: abudhabi2022.to_csv('AbuDhabi2022LapTimes.csv', encoding='utf-8', index=False)
```

Figure 2.3. Outputting Dataframe to a .csv File

- **Visualisation**

To visualise the data, I used Tableau, a visual analytics platform that has a lot of data visualisation tools which will be helpful for this project. Tableau allows its users to import data sources and work with them directly with a user-friendly UI. It also allows for smaller-scaled data manipulation so that we do not have to go back to Jupyter to create a new .csv every time.

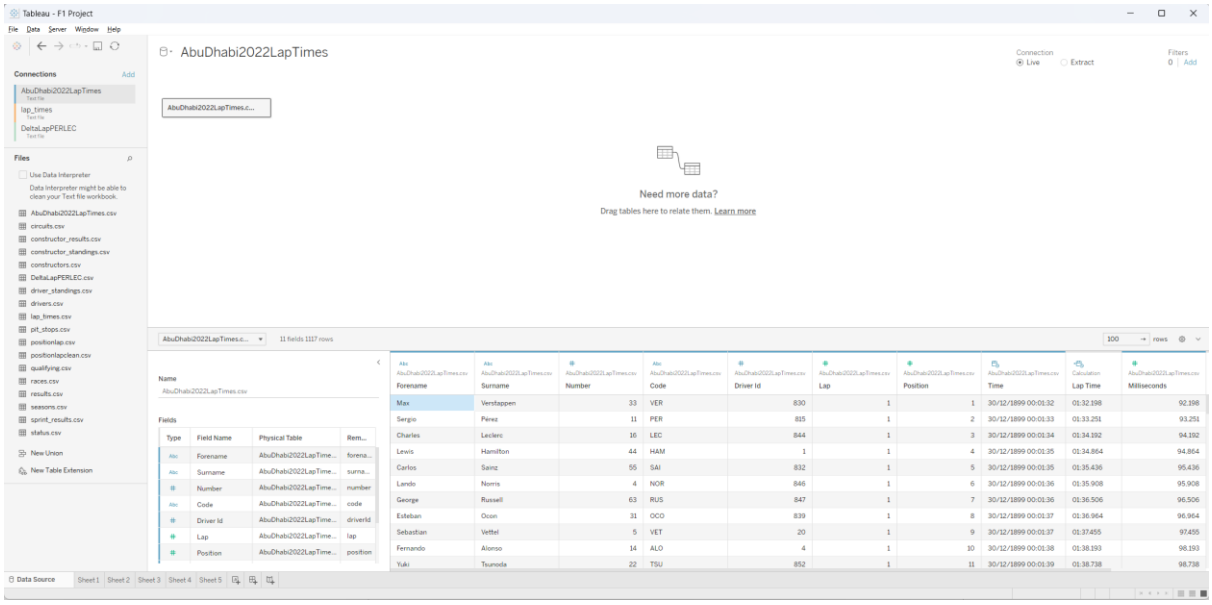

Figure 2.4. Working with .csv Files in Tableau

When visualising the data, it is also important to take note of the design. The design I want to incorporate needs to be clean, showing only the important details, but also not leaving too much out. I also made sure the labels for the graphs are big enough to be read clearly, but not too big that it blocks the graph. Finally, Tableau provides a great colour palette to make sure my plots are colourblind-friendly, which I will incorporate into all my visualisations.

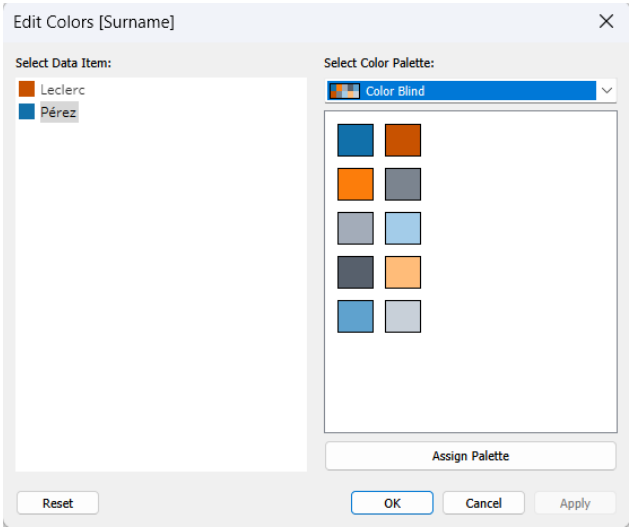

Figure 2.5. Tableau's Colour-Blind Palette

III. Visualisation and Analysis

- **Race Overview**

First visualisation I want to create is to show the general rundown of the race itself, from who finished where, how many laps were completed, etc. The below graph shows my first attempt at that.

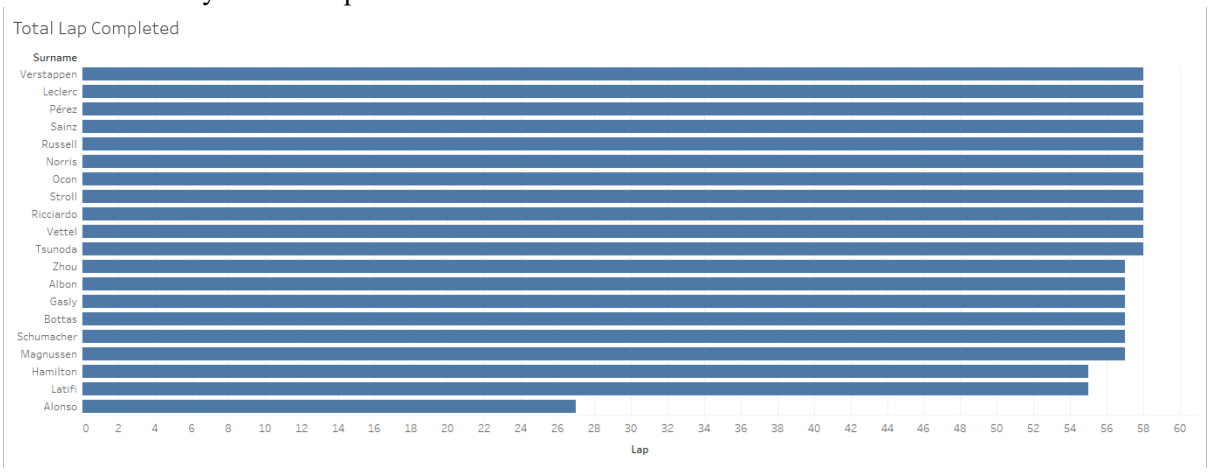

Figure 3.1. General Race Overview 1

The figure above is sorted based on the finishing order, with Verstappen finishing first, completing all 58 laps, while Alonso was classified last because he only completed 27 laps. This graph does not fully distinguish between finishers and non-finishers, which can be quickly resolved by adding colours to the graph.

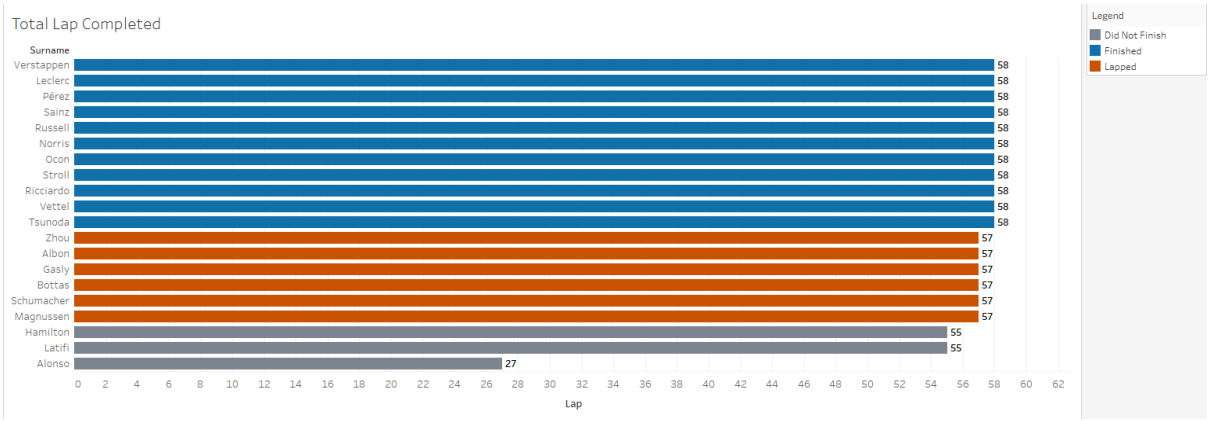

Figure 3.2. General Race Overview 2

By adding distinguishing colours, more clarity is provided to properly differentiate between finishers, lapped finishers (managed to reach the finish line but was lapped), and those who did not finish (did not reach the finish line due to issues, etc.). I made sure the colours are colourblind-friendly, and I also added labels to the graph to show the lap count of the drivers.

- **Lap Times of Leclerc vs. Perez**

Now that the general rundown of the race is clearer, I can more specifically analyse the lap times of our two drivers of interest. My first attempt to do so is by creating boxplots of both drivers, to compare the distribution of lap times between two drivers to get an insight of which driver in general put in faster lap times.

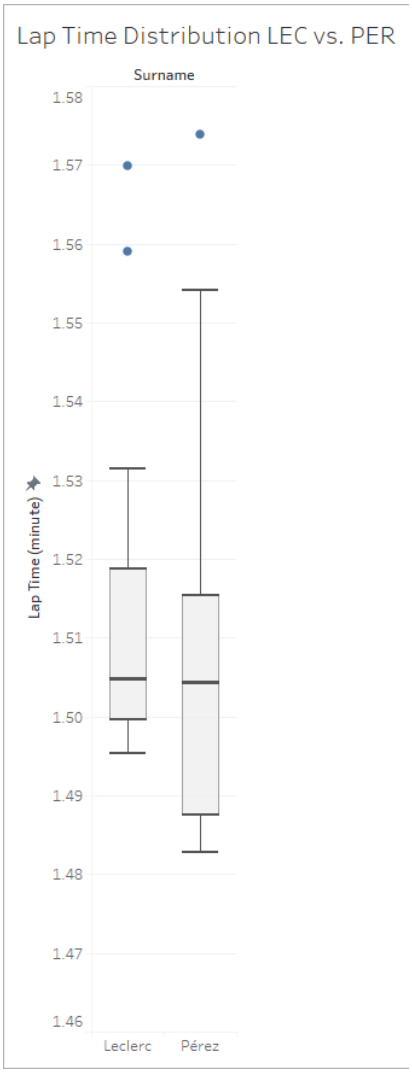

Figure 3.3. Lap Time Distribution

This plot shows that Perez in general put in faster lap times, but does not provide any context as to why and how he was able to do so. Many factors could come into play, such as better tyres, faster car, less traffic, so on and so forth. Thus, a better visualisation that can tell a better race story is needed.

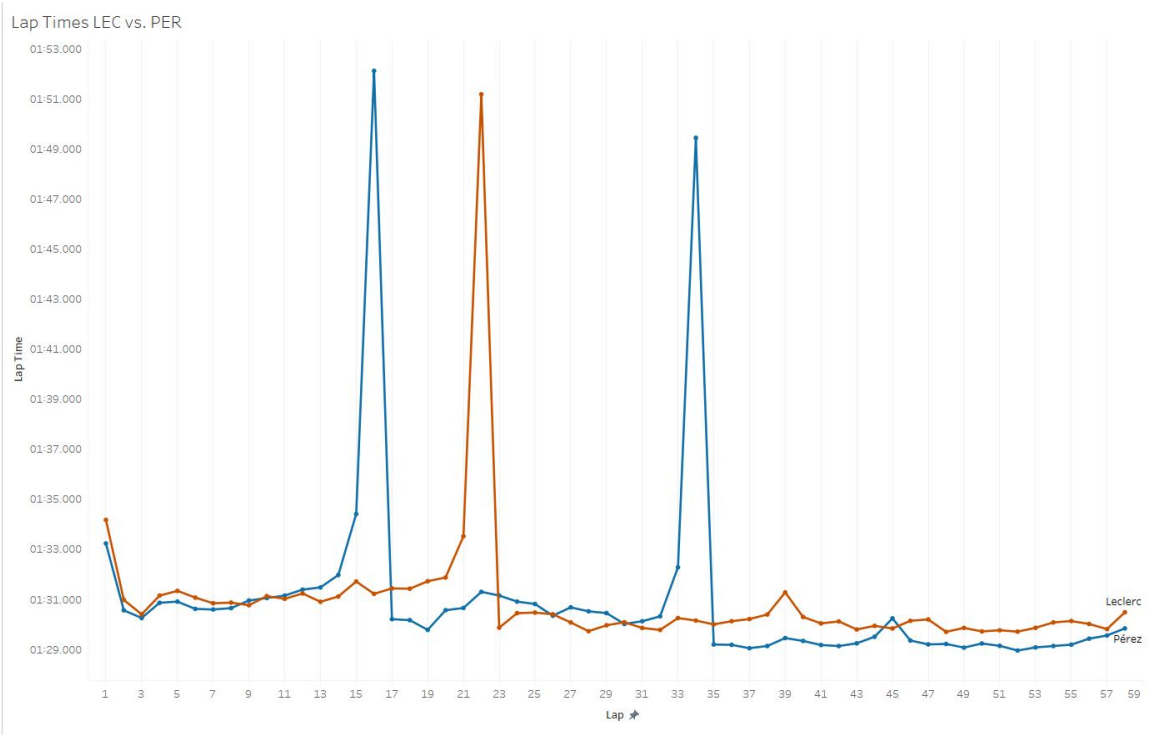

Figure 3.4. Lap-by-Lap Time Visualisation

The graph above tells a better story of how the race between these two commenced. The red line shows Leclerc's lap time, while the blue line shows Perez's. In general, we see how the trend for lap time is decreasing as the lap counter increases. This can be attributed to more onboard fuel being burned as the race progresses, meaning less weight on car, which leads to quicker lap times.

On the first lap, Perez manages to put in faster lap times than Leclerc until about the ninth lap. Perez soon pitted after that, as denoted by the massive spike in lap time during the 16th lap. This information tells me that he was having issues with managing his tyres, thus forcing the team to pit him early as he was losing time to Leclerc. After pitting, he was able to put in massively quicker lap times than Leclerc from laps 17-21. Ferrari responded to this by pitting Leclerc soon after on lap 22. Both then managed to put in similar lap times, trading quicker lap times from laps 23-32, when Perez seems to lose time to Leclerc once again. Keep in mind that we now know that Perez was having issues with managing his tyres, thus his team deciding to pit him once again so that he can have fresher tyres to chase down Leclerc all the way to the end of the race.

This plan that Red Bull came up with seems to work, at least until around laps 42-45, when he seems to lap much slower than what he was able to do in previous laps, even somehow lapping slower than Leclerc who was on very old tyres in lap 45. Thus, from this lap-by-lap visualisation, we are able to know that something must have happened during laps 42-45 that made him lap slower than what he could have done, losing him precious seconds that could have let him overtake Leclerc at the end of the race.

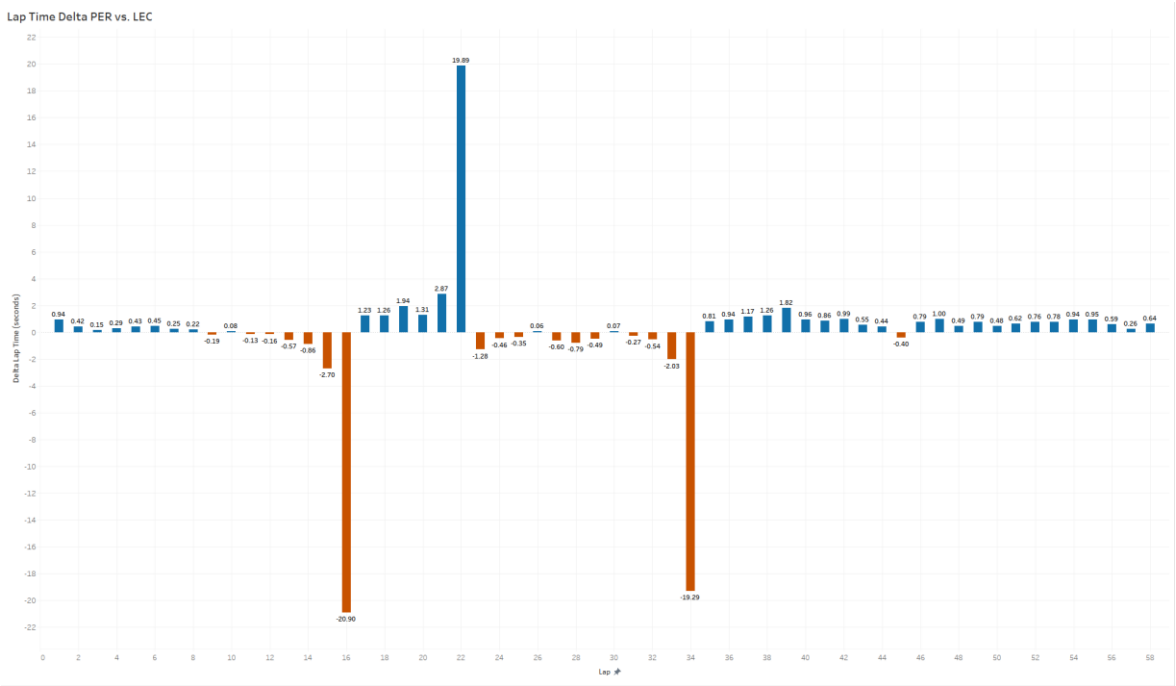

Figure 3.5. Lap-by-Lap Time Difference Visualisation

The above visualisation shows the delta lap times between Perez and Leclerc, in reference to Perez. Blue bars mean that Perez is gaining time during that lap, while red bars mean that Perez is losing time during that lap. This plot shows that after Perez's second pitstop, he was catching Leclerc at a very quick rate, lapping almost a second quicker than his rival. From laps 35-39, he was lapping quicker and quicker, until a sudden decline of about 0.9 seconds in lap time gained in the 40th lap. This decline continued, until lap 46, at which point he was starting to gain time again, though not as fast as before. Yet again, this visualisation points toward the fact that something happened to Perez at around laps 40-45 which made him lose potential time gained on Leclerc.

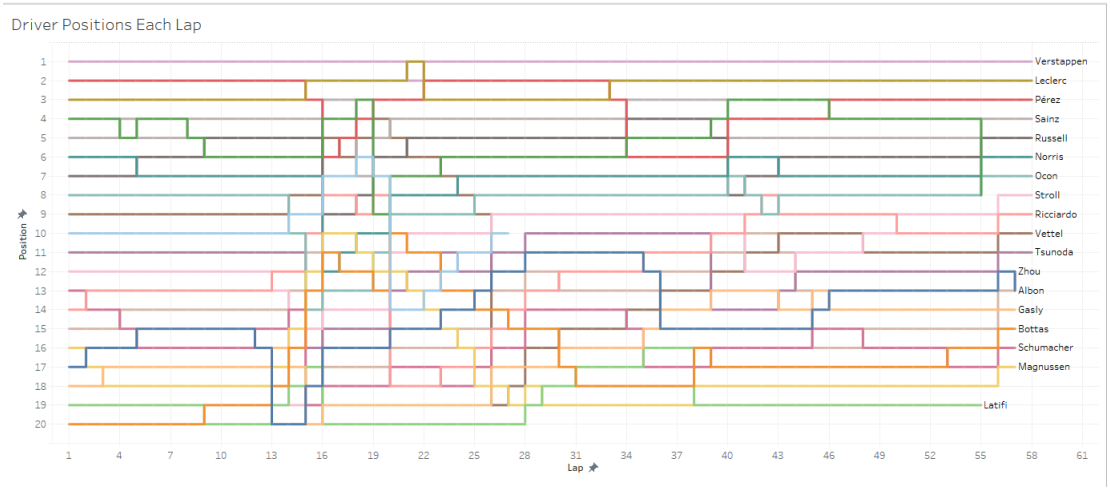

Figure 3.6. Lap-by-Lap Position Tracker 1

To investigate what happened, I plotted a lap-by-lap position tracker, which shows me where every driver was positioned at the start of each lap. However, the plot above looks too noisy, making it hard to find the information that we want. This is fixed

by only highlighting drivers who are around Perez and Leclerc i.e., drivers who could have affected both of their race.

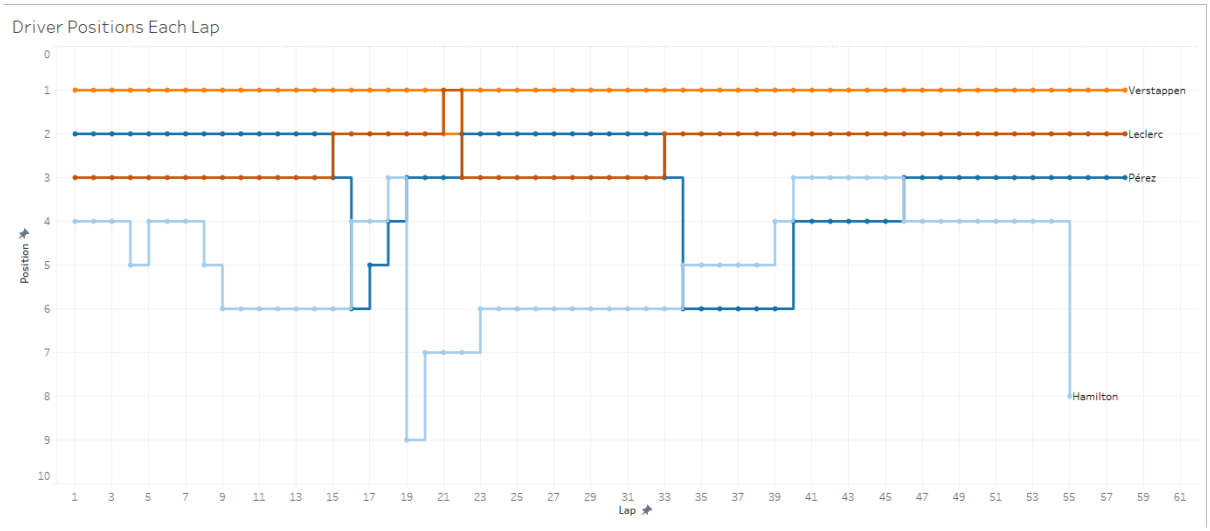

Figure 3.7. Enhanced Lap-by-Lap Position Tracker

Here, I was able to get a clearer picture of what happened. Perez lost a lot of positions on lap 34 because of his second pit stop and ended up behind Hamilton on track. He was then stuck behind Hamilton, until lap 46 when he overtook him. It is clear now that getting caught up behind Hamilton, struggling to overtake him was what costed Perez massive amount of lap time that he could have used to catch up to Leclerc at the end of the race.

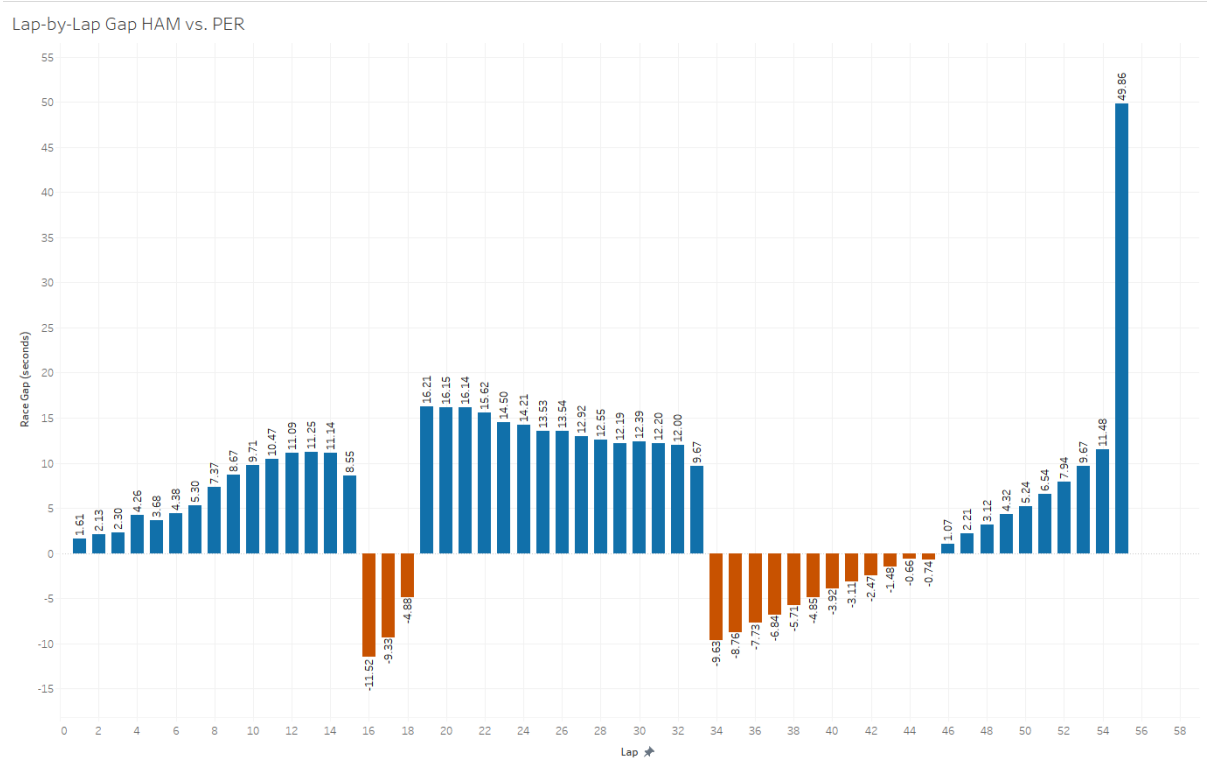

Figure 3.8. Lap-by-Lap Race Gap of Perez vs. Hamilton

To gain a sense at how Hamilton managed to affect Perez's race, I plotted the race gap per lap of Perez and Hamilton. Blue shows how much Perez is leading Hamilton by, and red shows how much Perez is trailing Hamilton by. From the chart alone, we can see how much faster Perez is for most of the race, leading Hamilton by more than 10 seconds at certain points of the race. However, despite being faster than him, Hamilton manages to stay ahead of him for longer than he would have liked, as shown by the plot above at laps 44-45, when he was within one second gap of Hamilton but failing to overtake him. This is also reflected that by the time the overtake was done, Perez manages to pull away from Hamilton by more than one second a lap, which is extremely quick in the context of Formula 1. By being stuck behind Hamilton for about 2 laps, it can then be approximated that Perez lost around 2 seconds of lap time, which could be crucial to catch up to Leclerc.

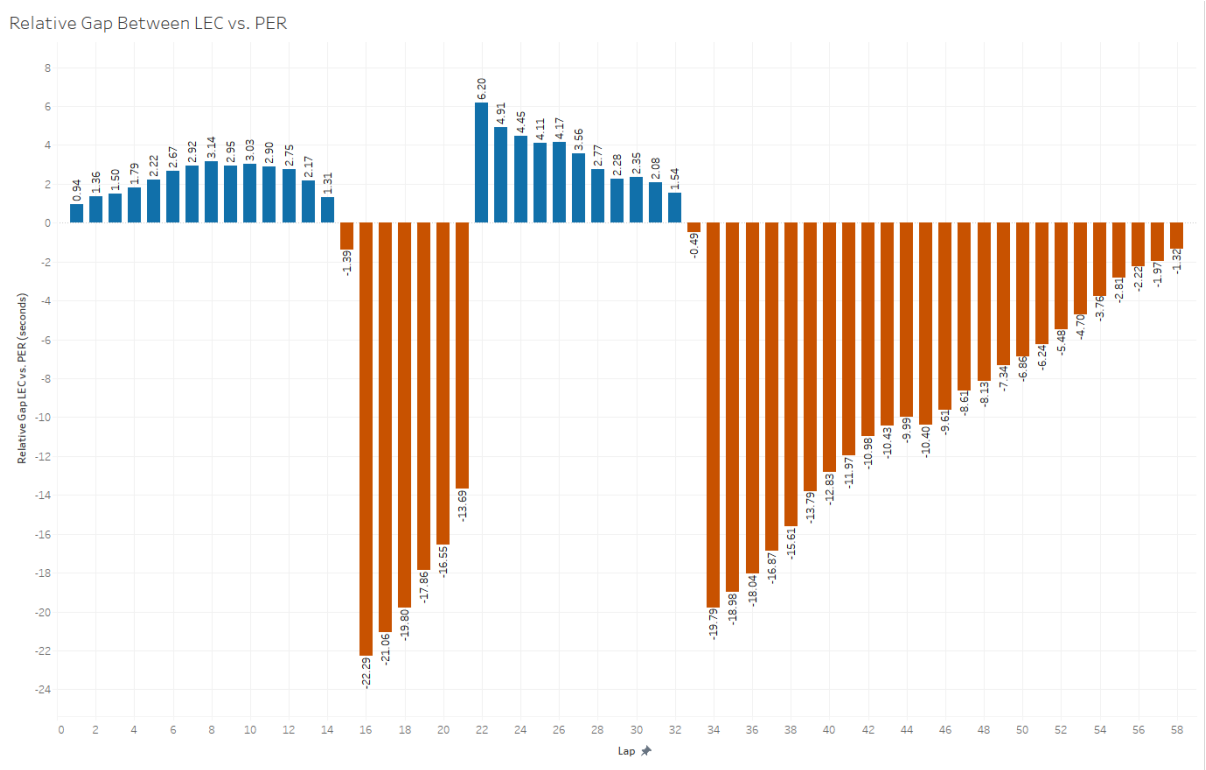

Figure 3.9. Lap-by-Lap Race Gap of Perez vs. Leclerc

The plot above shows a similar information as Figure 3.8., except Perez is now compared to Leclerc. Blue bars show how much Perez is leading Leclerc by, red bars show how much Leclerc is leading Perez by. One main takeaway from this plot is that we see how Leclerc is gaining time on lap 45, as mentioned in the previous part. The other main takeaway is how Leclerc was only 1.32 seconds ahead of Perez by the time they finished. As the previous plot tells us, Perez lost about 2 seconds by getting stuck behind Hamilton. Those 2 seconds would have given him the chance to overtake Leclerc and finish second in the race.

IV. Conclusion

As such, using the data that was available and visualising it, I was able to analyse the race strategies employed by both Ferrari and Red Bull Racing towards their respective drivers, Charles Leclerc and Sergio Perez. By first using Jupyter Notebook to prepare the dataset, I was then able to use visualisation tools provided by Tableau to visualise the data in a way that is easier to analyse and draw a conclusion from. The visualisation involves making comparisons of lap times between Leclerc and Perez, and the gap between them at every lap. From there, I was able to identify that another driver was involved in influencing the result of the race for these two, which is Hamilton.

The main findings from the data can be listed as such:

- Pitting for new tyres allow drivers to go much faster than other drivers on older tyres, at the cost of around 20 seconds of pit stop.
- The strategy employed by Red Bull to pit Perez twice was not the main reason why Perez lost to Leclerc, getting stuck behind Hamilton was.
- Race strategies that involve an extra pit stop for fresher tyres should consider drivers around them that they would need to overtake, not just the race gap between the driver and their main rival.

- For future considerations:

Formula 1 is a complex sport, with many underlying factors that can affect how quick a car can go during a lap. Things that I did not consider deeply during this project when judging why Perez was not able to catch Leclerc include engine modes, onboard fuel loads, and possible mechanical issues. This is because those factors were not included in the Kaggle dataset, but Formula 1 does provide a subscription service where users are able to listen to onboard radio between drivers and their race engineers. This radio conversation could bring certain insights that are not able to be gained from data alone, such as when drivers were told to push, when they are told to hold back to save their tyres, issues that exist in the car, etc. This insight could bring more interesting insights if incorporated to the analysis.

Design-wise, for the visualisations, I would like to learn more deeply as to how to use Tableau better, because I recognize that my visualisations are not the most creative. This is my first time using Tableau, and I know that there are still many features that I could probably have used in the visualisations to make things clearer for readers in the future.

All in all, I have been a long-time fan of Formula 1, but this project was the first time I tried to delve deeper into its data, digging deeper to uncover the reasons behind why certain things happened. I thought this project has been a great experience to learn more about the sport I love and it was interesting learning to use a new tool in Tableau too. It was fun learning how to design good visualisations, and I though I did a good job at visualising all the information that I found to be relevant to my project aims. I took into account legibility, colour-blindness, and other such considerations that I never thought of before learning design from this course.

V. References

Vopani (2023) *Formula 1 World Championship (1950 - 2023)*, *Kaggle*. Available at: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020 (Accessed: April 20, 2023).