Analyzing the Influence of Weather on Formula 1 Driver's Performance: An Exploratory Data Analysis from 2018 to 2023

UC Berkeley MIDS, W200: Python for Data Science, Prof. Gunnar Kleeman
Team members: Cat Weiss, Akin Akinlabi, Gia Nguyen
Github link: https://github.com/UC-Berkeley-I-School/Project2_Akinlabi_Weiss_Nguyen
December 13, 2023

1. Introduction

The popularity of Formula 1 in the United States (US) has exploded in recent years due to Netflix documentaries, increased sponsorships, and added races in the US[1]. The 2023 season saw record attendance thanks to the added Las Vegas Grand Prix[2], and Disney partnered with the sport to increase viewing opportunities to spectators across the world[3].

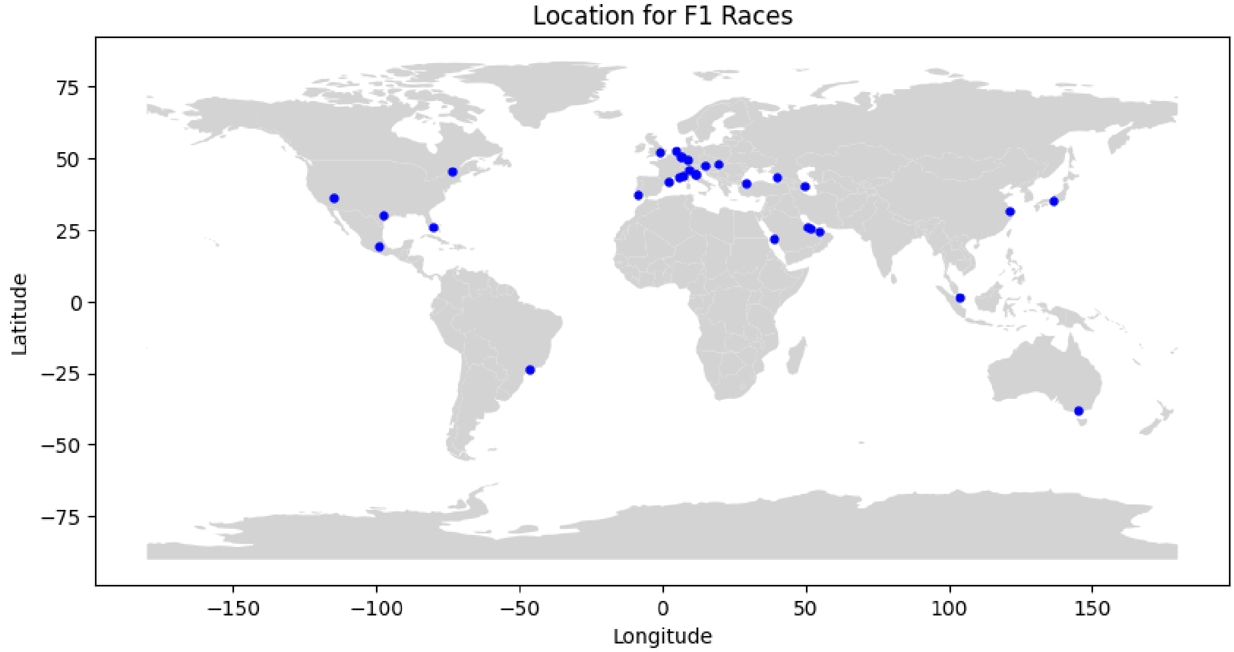


Figure 1. Map of the world showing locations where Formula 1 races occur.

While weather impacts many outdoor sports, there is a unique relationship between Formula 1 drivers and weather, as drivers and cars both have reputations for performing in certain weather conditions. Racing in extreme weather conditions significantly impacts both driver performance and the mechanical aspects of Formula 1 cars. In hot temperatures, drivers face the challenge of dehydration, with potential weight loss of up to 5%, and accelerated tire wear. The scorching heat intensifies these issues, making it a critical factor to manage during races. Conversely, racing in rain introduces a different set of challenges. Drivers contend with reduced visibility and increased risks, as colder temperatures affect brake performance. Cold weather not only hampers braking but also poses challenges in maintaining optimal tire temperatures. Formula 1 cars exhibit peak performance within specific temperature ranges, which becomes harder to achieve in colder conditions. Moreover, strong winds further complicate matters by impacting the handling and aerodynamics of the car[4]. In essence, each weather condition poses unique hurdles that drivers and teams must navigate for a successful race.

With the increased popularity of the sport, there have been many new locations for Formula 1 races, like the Jeddah circuit and the Las Vegas circuit, leading to differing weather conditions for each race. We were interested to explore an outside factor like weather on the effects of Formula 1 driver performance. We used historical weather data as well as historical Formula 1 data from the 2018-2023 seasons.

1.1. Definitions

Formula 1: Formula 1 is a motorsport that is the highest class of international racing for single-seater formula racing cars.

Teams: There are 10 Formula 1 teams per season, with a total of 14 teams in our dataset

Drivers: There are 20 Formula 1 drivers per race, 2 per team. There are a total of 25 drivers in our dataset.

DNF: Did Not Finish, a driver didn't finish the race due to either reliability issue(s), driver error or a racing incident. For the scope of our analysis, we won't distinguish between the three.

Driver Performance: Driver position at the end of the race. We analyzed both regular and sprint races.

Abnormal weather conditions: For our analysis, we defined abnormal weather conditions to be a race that included precipitation, wind greater than 25 miles per hour, temperature in the bottom ten percent of races, and temperature in the top 20 percent of races.

1.2. Focusing Questions

Using this data, we explored the following questions:

- How have abnormal weather conditions influenced driver standings at the end of each race from 2018 - 2023?
- Do specific drivers exhibit different DNF rates under varying weather conditions?
- What patterns emerge in driver's performances across different weather conditions?

We explored the difference between the normal and abnormal weather condition data for each of these questions. For the first question, we compared all drivers across all teams to each other, however for the second and third question, we compared drivers on the same team. This is due to the varying budgets of each team greatly affecting the car and limiting the driver's performance, as we only want to compare the driver performance to other drivers that are driving similar cars.

1.3. Datasets

The Formula 1 dataset was compiled using numerous datasets in the Ergast API[5] and is named as our "Race Dataset". The weather API from open-meteo[6] fetched the weather conditions on the day and location of the races based on the "Race Dataset." The results from the weather API are named as our "Weather Dataset." These two datasets were joined together to create a master dataset where we do our exploration on the impact of weather conditions on driver position in each Formula 1 race from 2018-2023.

1.4. Data Preprocessing

We used the Ergast API to gather relevant datasets to construct our "Race Dataset." Datasets we used were: circuits, constructor standings, constructors, driver standings, drivers, races, results, and status. Then, we merged them together based on different IDs and filtered it to only keep data from 2018-2023. We manually selected columns that we were interested in and renamed them for our purpose. After formatting the dataset, we performed data quality checks to ensure there were no missing values or data entry errors. We also renamed a few variables like the team names to keep it coherent. Upon further examination, we noticed that some drivers did not start the race at all due to disqualifications, illness, and withdrawals. We removed those entries from our analysis. We also removed teams and drivers who have less than 38 entries (minimum of 2 seasons) since we are only interested in drivers who have had the opportunity to race many times in our designed conditions. Lastly, we created a "F/DNF" column that is 0 if the driver finished the race and 1 if they did not for any given reason.

We fetched the weather data from the Open-Meteo Historical Weather API. Using the Race ID, latitude and longitude of the race location, and start time in UTC, we retrieved the hourly weather variables for the date and location of each race. The weather variables we retrieved from the API were the hourly precipitation in inches, hourly wind speed in miles per hour, and hourly temperature in fahrenheit. The maximum race time for a Formula 1 race is 3 hours, so we kept 3 hours worth of weather data for each race, starting at the race time and ending 3 hours later, with a total of 4 weather data points per race.

We defined the criteria for abnormal weather conditions as follows:

- wet conditions are any race where precipitation is greater than 0 inches
- windy conditions are any race where the wind speed was greater than 25 miles per hour
- cold conditions are the bottom 10 percent of races by temperature, this includes all races 62.4 degrees and below
- hot conditions are the top 20 percent of races by temperature, this includes all races 80.3 degrees and above

We included only the bottom 10 percent of the coldest races, as Formula 1 races are predominantly held in mild temperature climates. Conversely, we included the top 20 percent of the hottest races, reflecting the higher frequency of Formula 1 races in warmer environments. Our weather dataset was created with columns for windy_conditions, cold_conditions, hot_conditions, and wet_conditions, indicating the presence of any defined abnormal weather conditions.After developing both a weather dataset and a race dataset, we merged these based on race ID to form our primary dataset, which incorporated our categorical weather columns. In total, 125 races were aggregated, with 75 occurring in abnormal weather conditions and 50 in normal weather conditions.

2. Exploratory Data Analysis (EDA)

2.1. How have abnormal weather conditions influenced driver standings at the end of each race from 2018 - 2023?

After preprocessing the data, our objective is to assess the impact of abnormal weather conditions on driver performance at the end of each race. Calculating the delta between a driver's starting and finishing race positions provides a measure of weather effect. Notably, Lance Stroll and Daniil Kvyat consistently outperformed their peers in abnormal weather conditions, gaining over a position per race on average. Conversely, about a quarter of the drivers in the dataset faced challenges in abnormal weather, averaging a loss of over a position per race. Charles Leclerc and Nico Hülkenberg stood out as the only drivers that experienced a significant drop of two positions per race in abnormal conditions. Shifting focus to normal conditions, Kimi Räikkönen, Kvyat, Antonio Giovinazzi, and Zhou Guanyu emerged as top performers, gaining over a position per race. On the other end, Romain Grosjean faced difficulties in normal conditions, with an average loss of over three positions per race. Below is a bar chart illustrating average score change per condition.

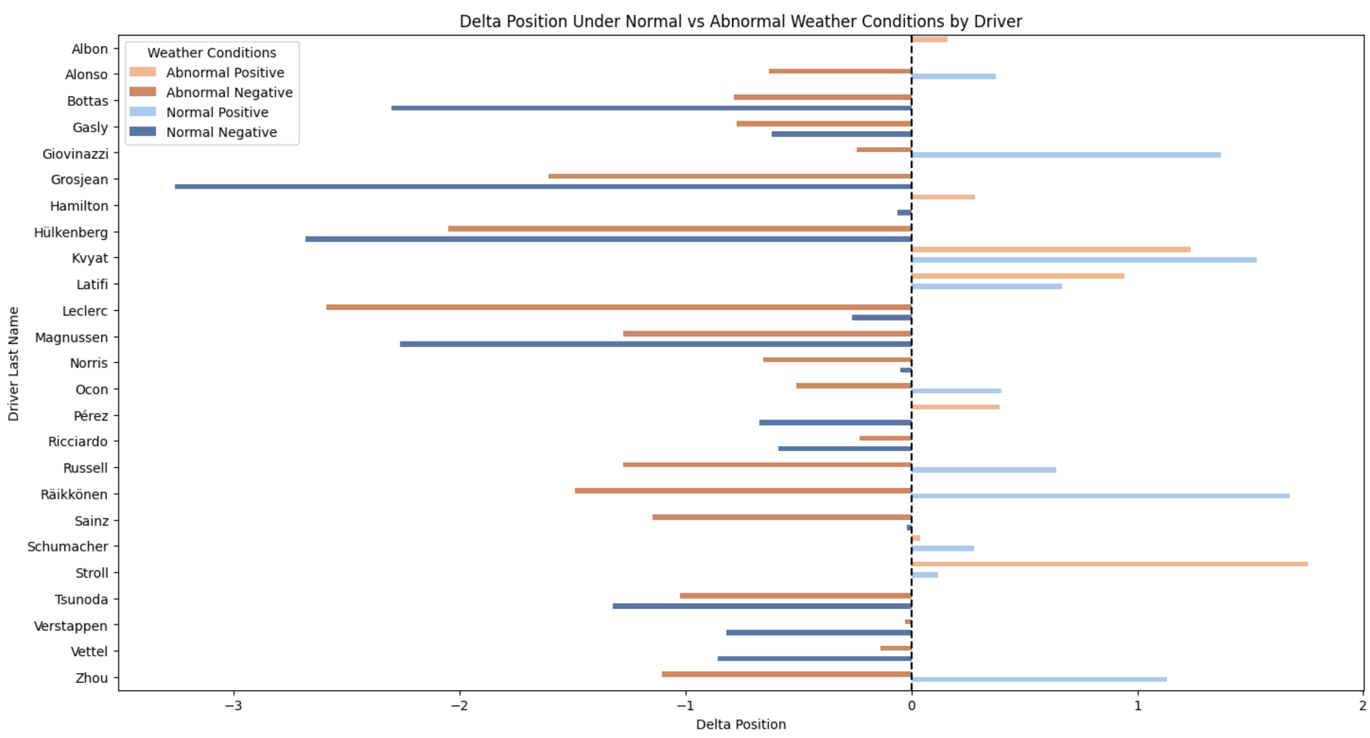


Figure 2. Showing average position change per condition for each driver. See appendix for table showing distribution.

In both normal and abnormal conditions, Kvyat, Nicholas Latifi, Mick Schumacher, and Stroll consistently gained positions on average, showcasing their versatility in varying weather conditions.

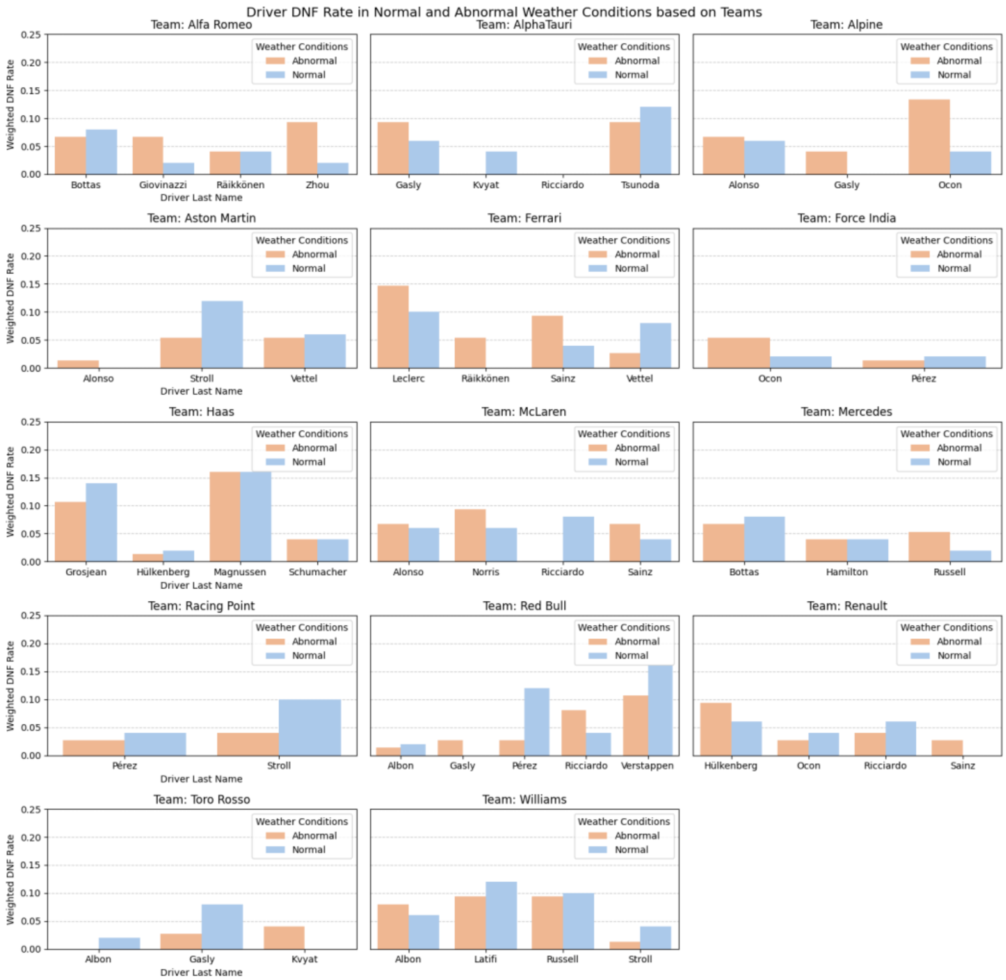
2.2. Do specific drivers exhibit different DNF rates under varying weather conditions?

Our next objective is to analyze drivers' and teams' finish rates in various weather conditions. To achieve this, we calculated each driver's Did Not Finish (DNF) rate in abnormal and normal conditions and normalized the rate based on the count of races in each weather category, ensuring values are comparable. The formula is represented as follows:

$$\text{Weighted Average} = \frac{\sum(\text{DNF Rate in Weather Condition} \times \#\text{ of Entries})}{\sum\text{Total Count of Races in Weather Condition}}$$

Most drivers exhibited a higher DNF rate in abnormal weather conditions compared to normal weather conditions. However, it's noteworthy that 15% of drivers had a DNF rate exceeding 10 percent in normal weather conditions, while only 10% of drivers faced a similar rate in abnormal weather conditions. Below is a chart showing driver's DNF rates in either abnormal or normal weather conditions by team.

Figure 3. Bar chart illustrating DNF rate for each driver by team. See appendix for table showing distribution.



Specifically, drivers like Kevin Magnussen, Leclerc, Grosjean, and Max Verstappen all faced DNF rates of 10% or higher in both normal and abnormal weather conditions. It's noteworthy that Grosjean and Magnussen, who raced for the Haas team, shared this trend. Surprisingly, 35% of drivers managed to maintain DNF rates below 5% in both weather conditions.

When comparing drivers within teams, notable patterns emerge. In abnormal weather conditions, drivers such as Esteban Ocon, Leclerc, Magnussen, and Verstappen exhibit higher DNF rates compared to their teammates. Similarly, in normal conditions, Leclerc, Stroll, Yuki Tsunoda, Magnussen, and Verstappen demonstrate higher DNF rates than their respective teammates. Leclerc and Verstappen, in particular, stand out with significantly higher DNF rates than their teammates in both weather conditions. Additionally, Ferrari drivers posted higher DNF rates in abnormal weather conditions, while Red Bull drivers exhibited the opposite trend in normal conditions.

Examining team-wise DNF rates in various weather conditions reveals interesting patterns. Notably, Haas, Red Bull, McLaren, Ferrari, and Williams consistently exhibit DNF rates exceeding 10% in both normal and abnormal weather conditions. On the contrary, Mercedes, Aston Martin, and historical teams such as Toro Rosso, Racing Point, and Force India maintain DNF rates below 10% in both categories. This observation prompts further exploration into the contributing factors behind the distinct performance of teams. Williams, Red Bull, and Haas, in particular, seem to face challenges leading to comparatively higher DNF rates. A detailed analysis of specific seasons or race types could unveil intriguing trends in team performance.

2.3. What patterns emerge in driver’s performances across different weather conditions?

Our last objective is to analyze the patterns that emerge in each driver’s performance across different weather conditions (hot, cold, wet, cold/wet, normal). We also preprocessed the data for windy, and all other windy combinations but there were not sufficient data points for analysis. Since different teams race in different cars and have different equipment, we are only comparing drivers from the same team in terms of performance to ensure comparability.

Under hot conditions, the drivers are exposed to many challenges such as engine overheating, tire problems, battery stress and visibility issues. They tend to cope with heat by using techniques like pre-race hydration and wearing cooling vests to perform optimally in the race. Fernando Alonso and Sebastian Vettel have been known to perform well in hotter races. Under cold conditions, drivers are known for their ability to manage tire temperatures. A notable driver for this weather condition is Lewis Hamilton. For wet conditions, there are a few notable drivers such as Max Verstappen, Sebastian Vettel, Fernando Alonso, and Lewis Hamilton. We will be focusing on a selection of teams and drivers to illustrate interesting findings.

Figure 4. Box Plots illustrating driver performance distribution in wet weather conditions.

Team Mercedes:

Lewis Hamilton and Valtteri Bottas have shown remarkable consistency in their performance across various weather conditions but there are some differences. Hamilton, known for his wet-weather driving skills, would be expected to perform strongly in wet conditions, which is reflected in higher median finishing positions (lower numerical values) and a smaller interquartile range (IQR), indicating consistency. Valtteri Bottas has a slightly wider IQR than

Hamilton, indicating more variability in his finishes. He performs the best in hot and wet conditions.

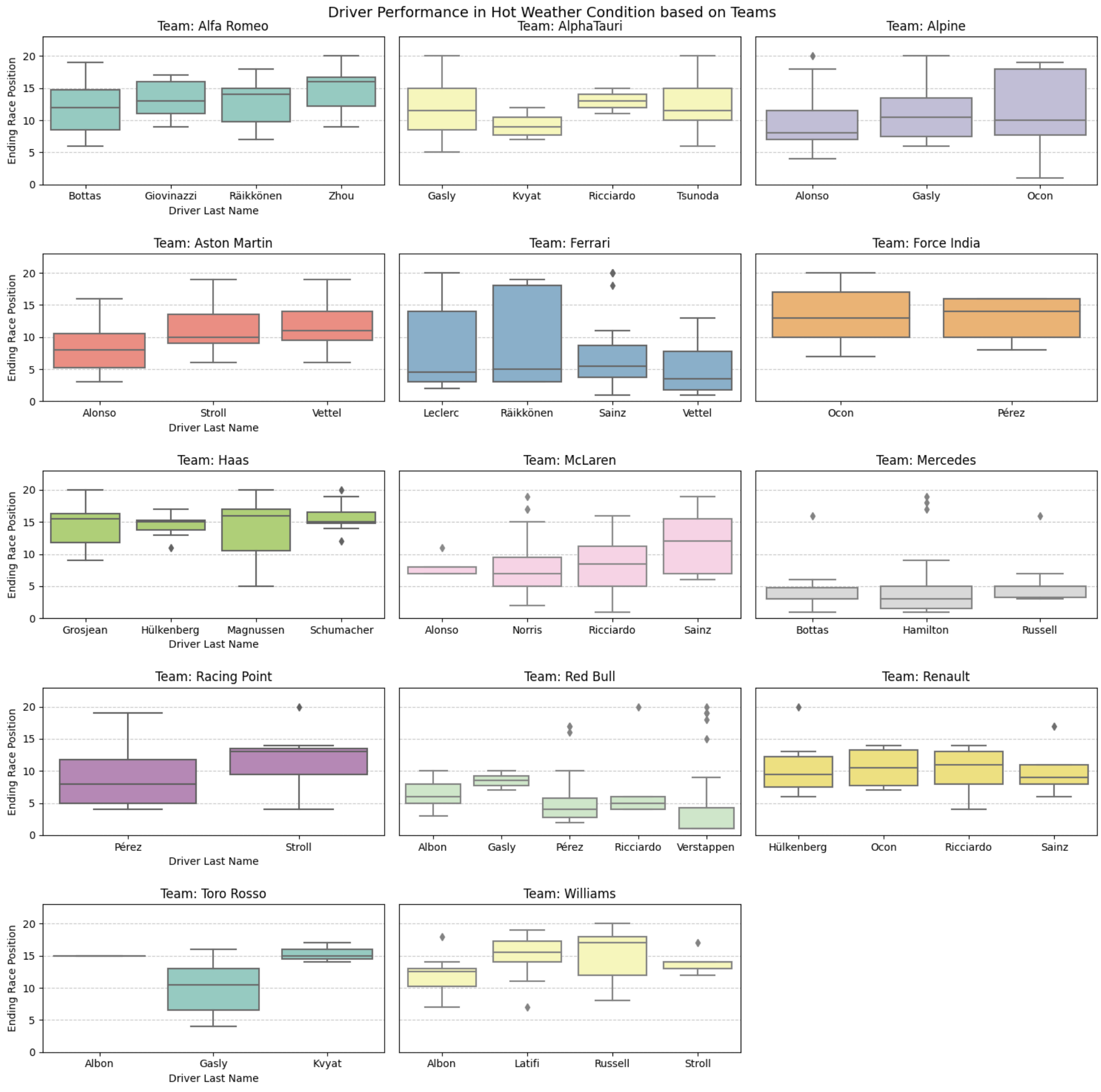


Figure 5. Box Plots illustrating driver performance distribution in hot weather conditions.

Team Red Bull:

Max Verstappen is renowned for his driving style in wet and cold/wet conditions. In normal and hot conditions, he tends to have more outliers and a wider IQR, indicating more variability in his finishes. Overall, his performance is consistently strong in all weather conditions. Daniel Ricciardo is also a well-known driver in this team. His performance is the best under hot weather conditions, and performs poorly in cold, wet, and cold/wet conditions.

Although it is expected that most drivers will perform better under normal conditions, it is interesting to see that such is not the case. Since most drivers have the most races under normal weather conditions, there will be more data points to work with, thereby increasing the variability of their finishes. Compared to abnormal weather conditions, where most drivers have only driven a few races, the finishing position for each race will heavily impact their IQR range. In conclusion, despite a few variabilities here and there, Mercedes drivers perform consistently well under all weather conditions that we examined. Please refer to the Appendix for box plots in other weather conditions.

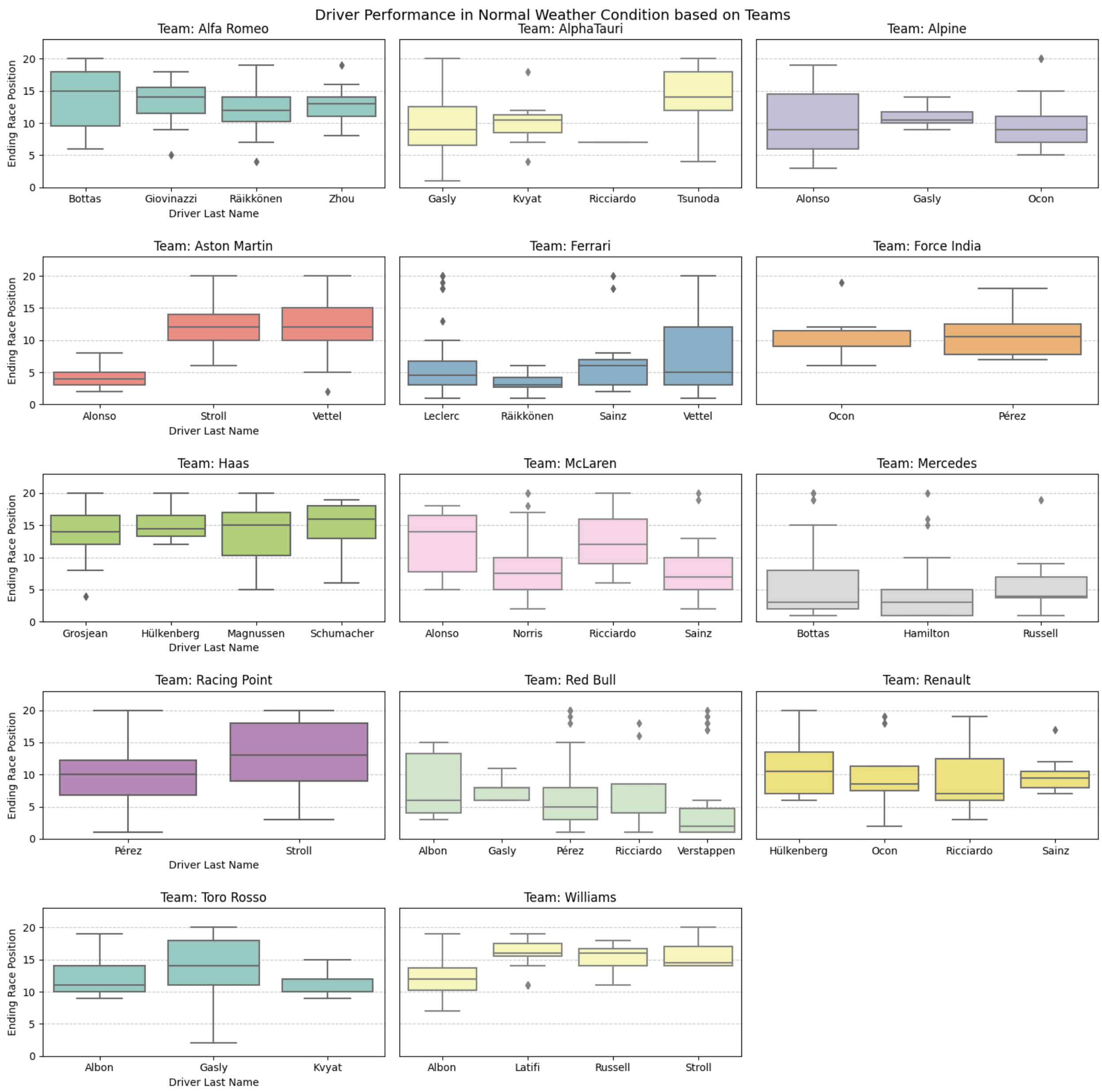


Figure 6. Box Plots illustrating driver performance distribution in normal weather conditions.

3. Conclusion

F1 drivers frequently contend with abnormal weather conditions, including extreme temperatures. Our analysis reveals that abnormal weather conditions contribute to a marginally higher volume of races and average DNF rate of 15%, compared to the 14% observed in normal weather conditions. Leclerc stands out as a driver who faced the second-highest DNF rate of 15% in abnormal weather conditions. Magnussen consistently holds the highest DNF rates across all weather conditions. Verstappen surprisingly exhibits one of the highest DNF rates in both normal and abnormal conditions, yet remains among the top performers. Hamilton demonstrates remarkable consistency with some of the lowest DNF rates in both weather conditions, particularly excelling in cold and wet conditions.

On the team front, Haas, Ferrari, and Williams consistently post higher DNF rates in both normal and abnormal conditions. Redbull and Haas stand out with the highest DNF rates in normal conditions, while Mercedes and Aston Martin showcase reliability with lower DNF rates across both weather conditions.

4. References

1. Driscoll, Finnegan. (September 12, 2022). Formula 1 in the US: What's behind its rise in popularity? [RTR Sports Marketing]
2. Brown, Maury. (March 29, 2023). Inside The Numbers That Show Formula 1's Popularity And Financial Growth. [Forbes]
3. Hall, Andy. (March 3, 2023). Formula 1 Presents Unique Opportunities for Disney Advertisers. [ESPN]
4. News Team. (2022). How Weather Affects Formula 1. [William Hill]
5. Ergast Developer API. n.d. API Documentation. [Ergast]
6. Open-Meteo. N.d. Free Weather API. [Open-Meteo]

5. Appendix

In this section of the report, readers will find a comprehensive collection of additional graphs and charts that dive deeper into the questions we are exploring. These visualizations have been created to provide more clarity and insights into different weather conditions and its effect on driver's performance.

	last_name	first_name	count
0	Bottas	Valtteri	125
1	Gasly	Pierre	125
2	Verstappen	Max	125
3	Sainz	Carlos	125
4	Hamilton	Lewis	123
5	Stroll	Lance	123
6	Pérez	Sergio	123
7	Ricciardo	Daniel	109
8	Russell	George	104
9	Norris	Lando	103
10	Leclerc	Charles	103
11	Ocon	Esteban	103
12	Vettel	Sebastian	100
13	Magnussen	Kevin	100
14	Alonso	Fernando	86
15	Albon	Alexander	81
16	Räikkönen	Kimi	79
17	Tsunoda	Yuki	66
18	Hülkenberg	Nico	63
19	Latifi	Nicholas	61
20	Giovinazzi	Antonio	60
21	Grosjean	Romain	56
22	Zhou	Guanyu	44
23	Schumacher	Mick	43
24	Kvyat	Daniil	38

	team_name	count
0	Red Bull	250
1	Ferrari	249
2	Mercedes	249
3	Haas	221
4	Alfa Romeo	206
5	McLaren	206
6	Williams	184
7	AlphaTauri	151
8	Alpine	132
9	Aston Martin	127
10	Renault	116
11	Racing Point	73
12	Toro Rosso	63
13	Force India	41

Table 1: Driver names and number of race entries

Table 2: Team names and number of race entries

	last_name	Normal	Abnormal
0	Leclerc	-0.261905	-2.590164
1	Hülkenberg	-2.681818	-2.048780
2	Grosjean	-3.260870	-1.606061
3	Räikkönen	1.676471	-1.488889
4	Russell	0.642857	-1.274194
5	Magnussen	-2.263158	-1.274194
6	Sainz	-0.020000	-1.146667
7	Zhou	1.133333	-1.103448
8	Tsunoda	-1.320000	-1.024390
9	Bottas	-2.300000	-0.786667
10	Gasly	-0.620000	-0.773333
11	Norris	-0.047619	-0.655738
12	Alonso	0.375000	-0.629630
13	Ocon	0.400000	-0.507937
14	Giovinazzi	1.370370	-0.242424
15	Ricciardo	-0.590909	-0.230769
16	Vettel	-0.857143	-0.137931
17	Verstappen	-0.820000	-0.026667
18	Schumacher	0.277778	0.040000
19	Albon	0.000000	0.160000
20	Hamilton	-0.061224	0.283784
21	Pérez	-0.673469	0.391892
22	Latifi	0.666667	0.941176
23	Kvyat	1.529412	1.238095
24	Stroll	0.120000	1.753425

Table 3: Drivers and their average position change in normal vs abnormal weather conditions

	team_name	dnf_abnormal	dnf_normal	Abnormal	Normal
0	Ferrari	0.161074	0.110000	0.160000	0.11
1	Haas	0.176471	0.211765	0.160000	0.18
2	Williams	0.192661	0.213333	0.140000	0.16
3	Alfa Romeo	0.162602	0.096386	0.133333	0.08
4	Red Bull	0.126667	0.170000	0.126667	0.17
5	Alpine	0.219512	0.100000	0.120000	0.05
6	McLaren	0.141667	0.139535	0.113333	0.12
7	AlphaTauri	0.155556	0.180328	0.093333	0.11
8	Renault	0.205882	0.166667	0.093333	0.08
9	Mercedes	0.080537	0.070000	0.080000	0.07
10	Aston Martin	0.112500	0.191489	0.060000	0.09
11	Force India	0.192308	0.133333	0.033333	0.02
12	Racing Point	0.125000	0.212121	0.033333	0.07
13	Toro Rosso	0.135135	0.192308	0.033333	0.05

Table 4: Teams & their raw DNF rates (dnf_normal, dnf_abnormal) and weighted DNF rates (Abnormal, Normal) in normal vs abnormal weather conditions sorted by Abnormal column

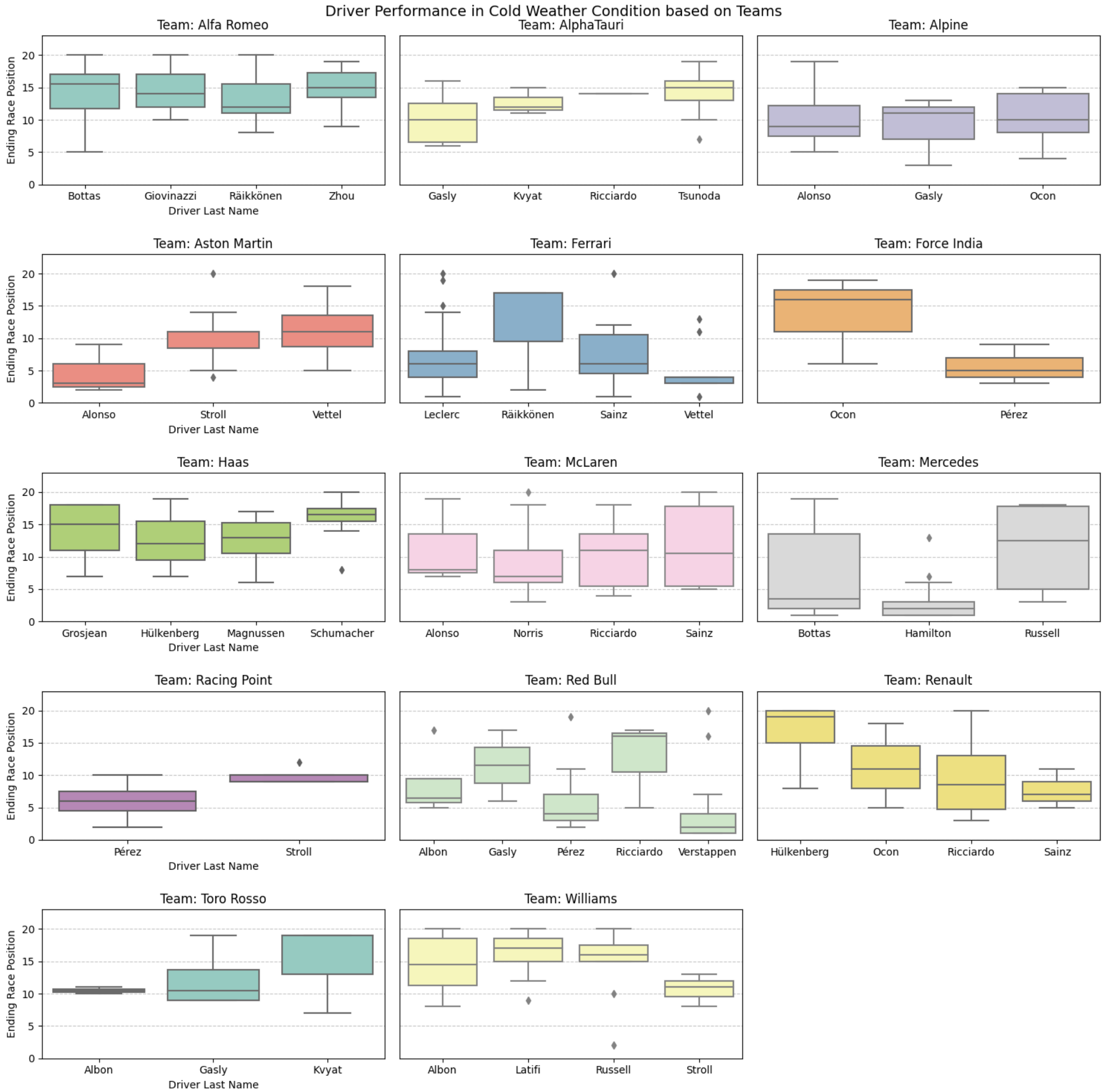


Figure 7. Box Plots illustrating driver performance distribution in cold weather conditions.

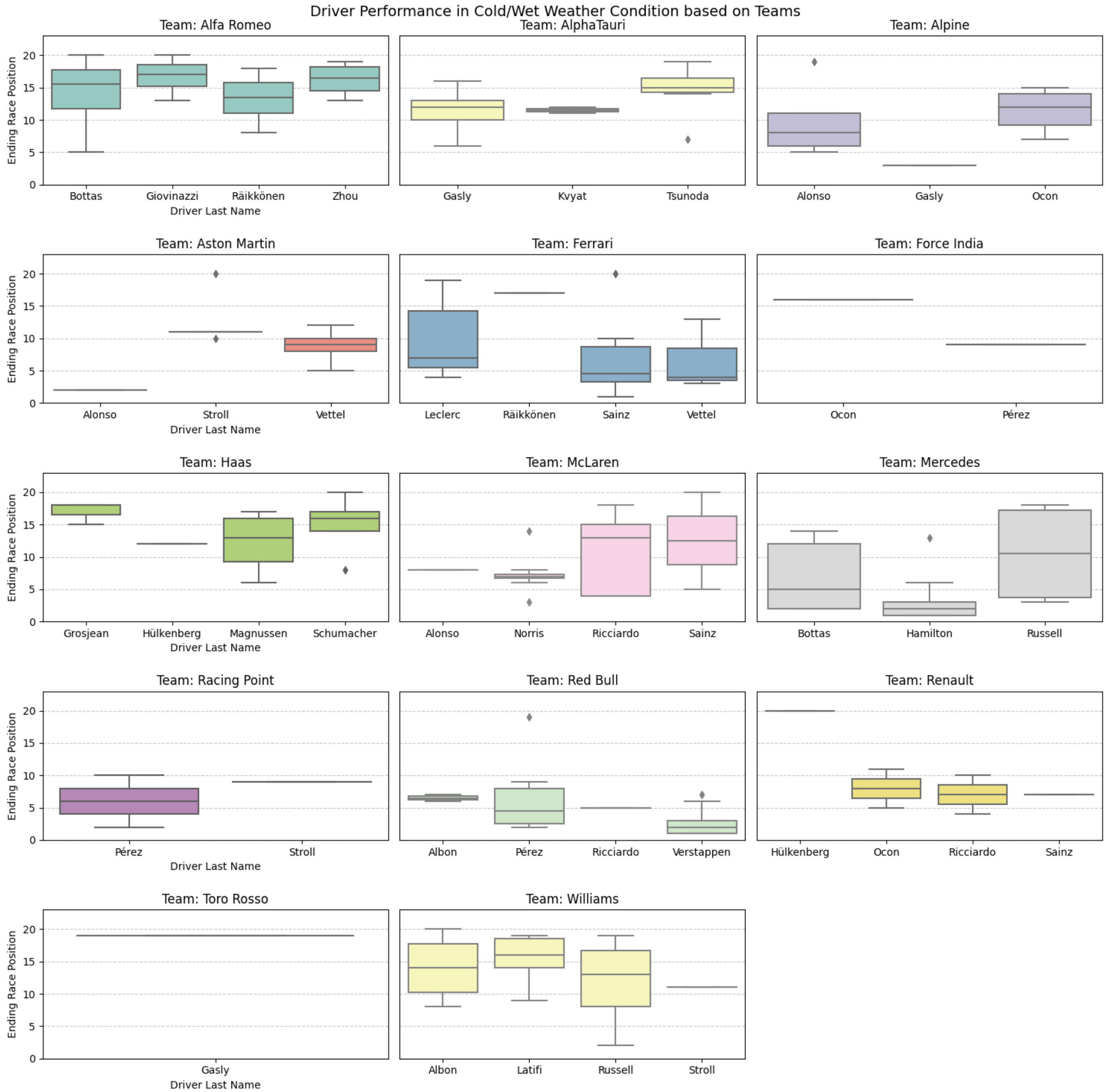


Figure 8. Box Plots illustrating driver performance distribution in cold/wet weather conditions.

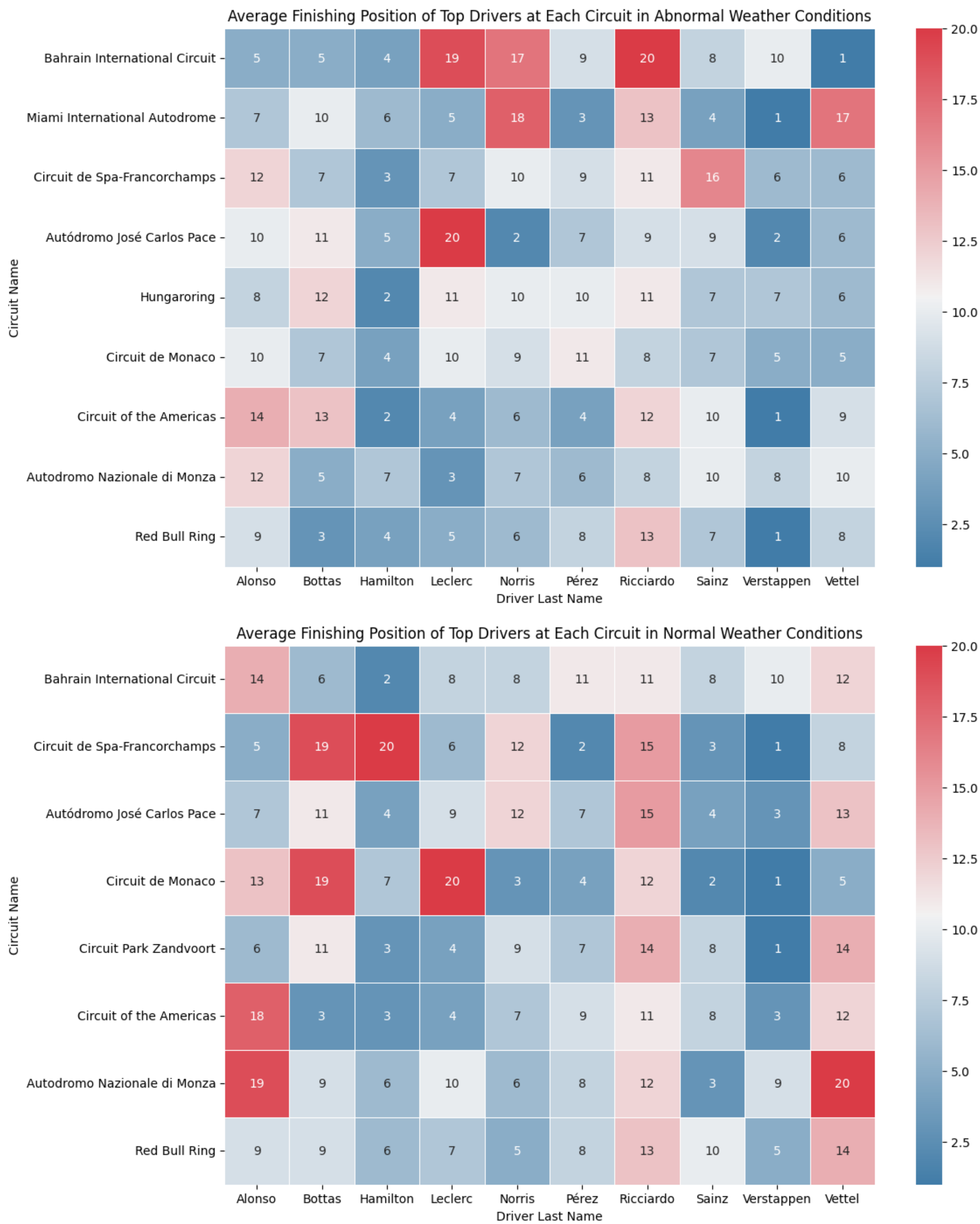


Figure 9. Heatmaps illustrating top 10 driver performance in top 10 circuits under normal vs abnormal weather conditions.