FORMULA 1

Speed, Strategy, Statistics

Unraveling the Data Behind Formula 1 Racing

Malihe Alavi

Sep 2024

Agenda

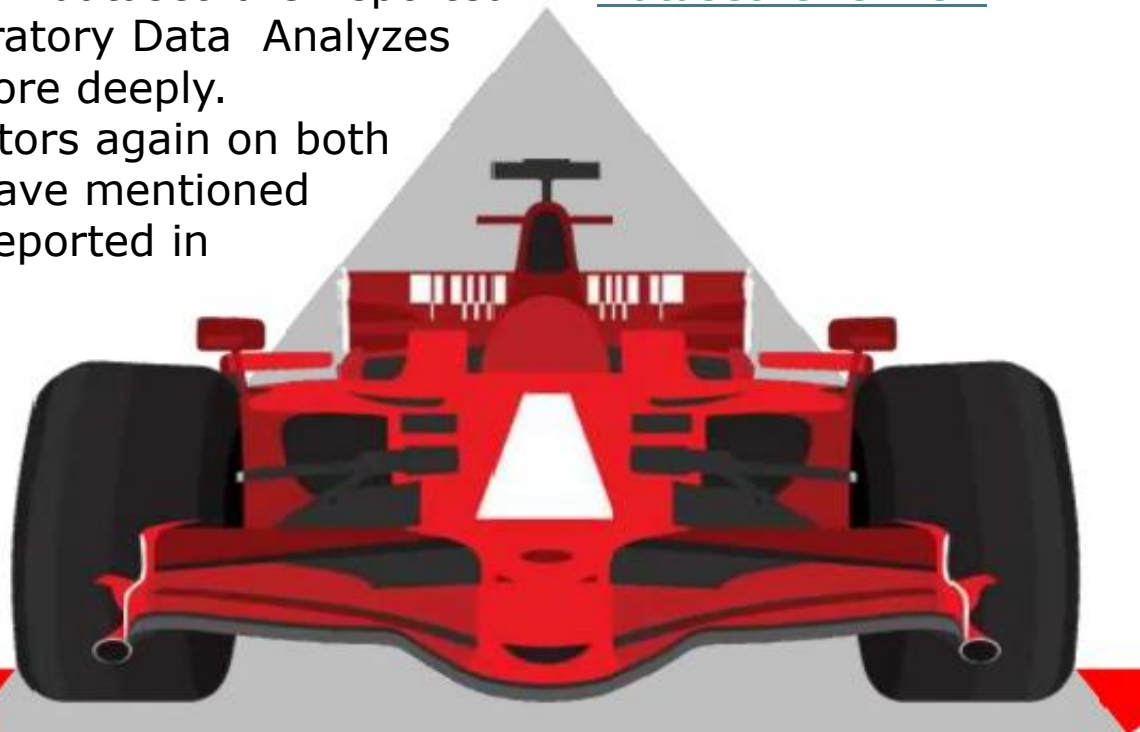
Introduction

Introduction

Formula 1, the pinnacle of motorsport, is not only a thrilling spectacle but also a rich source of data ripe for analysis. By examining datasets related to Laps status, weather, results and other race factors, we can uncover insights into the factors that influence race outcomes and team strategies. This analysis allows us to delve deeper into the environmental and technical aspects that affect performance on the track, providing a comprehensive understanding of how these variables interplay in the high-stakes world of F1 racing.

In this work, first, we have analyzed 2024-data and then merged it with Singapore-races to analyze it better and deeper. We first analyzed datasets of 2024 races one by one to understand their statistics and extracted any possible info from them. The analysis on each dataset are reported in Dataset Overview section. In Third section we have reported the results of Exploratory Data Analyzes on both merged dataset and 2024 datasets to explore them more deeply.

Forth Section focuses on all relationships between different factors again on both Datasets, merged one and 2024 data. In the fifth section we have mentioned final tips and conclusion. Key Findings and Appendix are also reported in second and sixth sections respectively.



Dataset Overview

Datasets

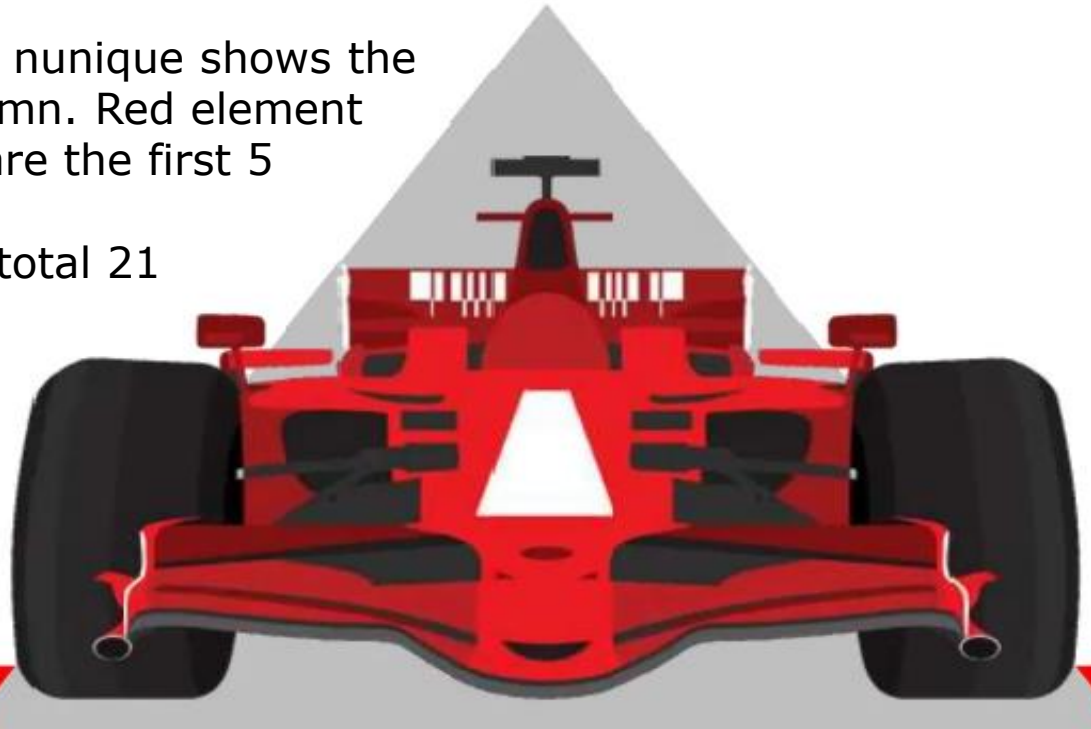
Car Data 2024

Car dataset has 6,064,741 records and 12 columns.

```
car_data_2024 :  (6064741, 12)
EventName :  |nnull: 0  |nunique: 15  | object |['Bahrain Grand Prix', 'Bahrain Grand Prix', 'Bahrain Grand Prix', 'Bahrain Grand Prix', 'Bahrain Grand Prix']
DriverName :  |nnull: 0  |nunique: 21  | object |['VER', 'VER', 'VER', 'VER', 'VER']
Date :  |nnull: 0  |nunique: 325127  | object |['2024-03-02 15:03:42.390', '2024-03-02 15:03:42.630', '2024-03-02 15:03:42.870', '2024-03-02 15:03:43.110', '2024-03-02 15:03:43.270']
RPM :  |nnull: 0  |nunique: 10870  | int64 |[9963, 9755, 8495, 6815, 5695]
Speed :  |nnull: 0  |nunique: 357  | int64 |[0, 0, 3, 11, 17]
nGear :  |nnull: 0  |nunique: 110  | int64 |[1, 1, 1, 1, 1]
Throttle :  |nnull: 0  |nunique: 100  | int64 |[15, 15, 15, 15, 15]
Brake :  |nnull: 0  |nunique: 2  | bool |[True, True, False, False, False]
DRS :  |nnull: 0  |nunique: 11  | int64 |[1, 1, 1, 1, 1]
Source :  |nnull: 0  |nunique: 1  | object |['car', 'car', 'car', 'car', 'car']
Time :  |nnull: 0  |nunique: 315731  | object |['0 days 00:00:00.048000', '0 days 00:00:00.288000', '0 days 00:00:00.528000', '0 days 00:00:00.768000', '0 days 00:00:00.928000']
SessionTime :  |nnull: 0  |nunique: 319116  | object |['0 days 00:59:59.959000', '0 days 01:00:00.199000', '0 days 01:00:00.439000', '0 days 01:00:00.679000', '0 days 01:00:00.839000']
```

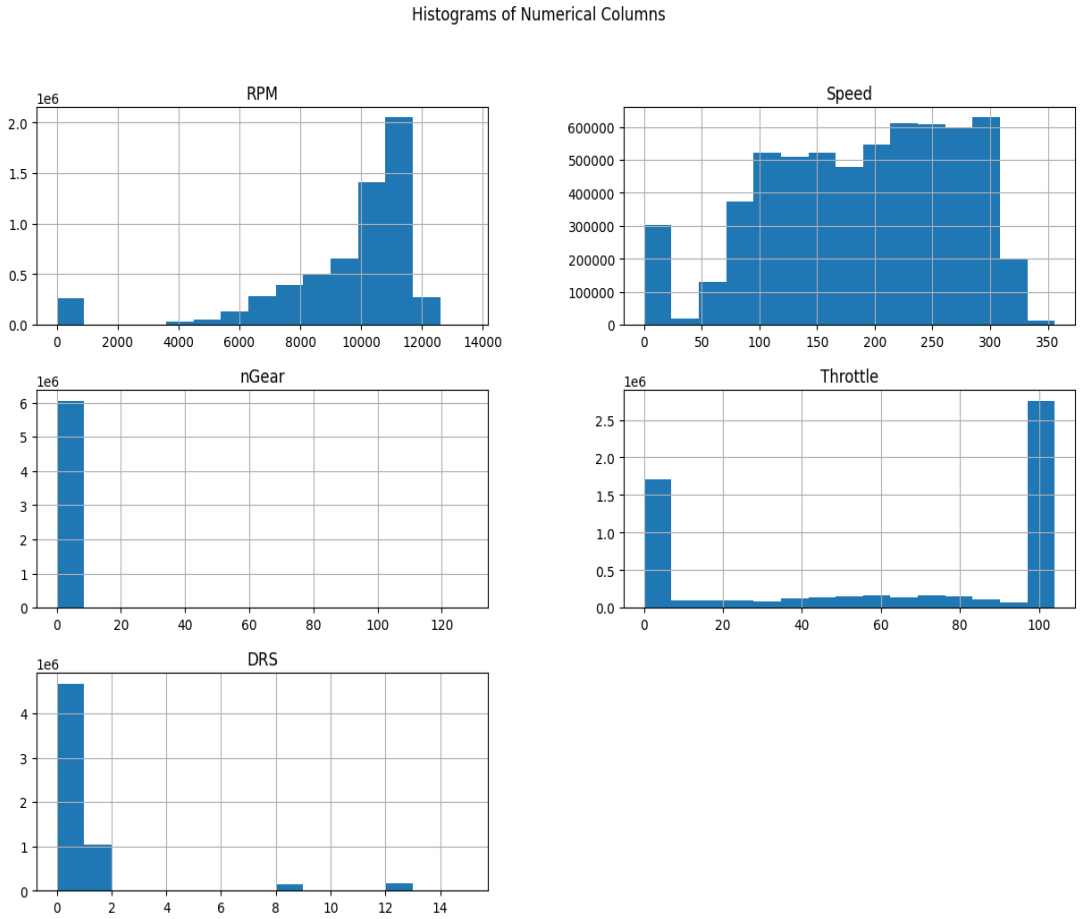
The above image shows columns info of Car dataset. nnull and nunique shows the number of null values and unique values of corresponding column. Red element indicates the data-type of each column and last blue element are the first 5 records of each column.

As it can be seen, dataset contains the info of 15 Events, with total 21 drivers participating in these events. The dataset is really rich without any null values. Source column has just single value (nunique) which indicates that all information are gathered from the car.



Datasets

Car Data 2024



The figure shows histogram of 2024 car-data numeric columns.

RPM refers to the number of times the engine's crankshaft makes one full rotation every minute. In F1, managing RPM is vital for optimizing acceleration and power delivery. From the histogram plot, RPM has peaked at values between 10,000 and 12,000.

The **Speed** of cars varies between 0 to 350 km/h how ever it's value is mostly between 100 to 300 km/h.

nGear indicates the current gear that the car is in at any given moment. It has 110 unique values from 0 to 128 in dataset, however histogram plot indicates nGear values was always less than 10 in all events and cars.

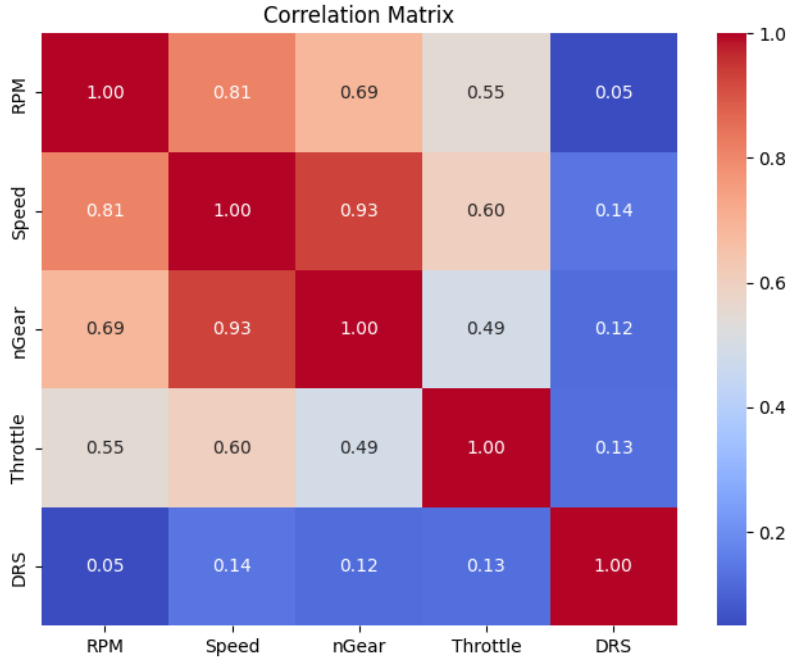
Throttle measures how much the driver is pressing the throttle pedal, which directly influences engine power output. The histogram has two prominent peaks, indicating a bimodal distribution. One peak is near 0% throttle, and the other is near 100% throttle. The peak near 0% throttle suggests that there are many instances where the throttle is not engaged or minimally engaged, possibly indicating periods of braking or coasting. The peak near 100% throttle indicates that the throttle is frequently fully engaged, likely during acceleration or on straights where maximum speed is required.

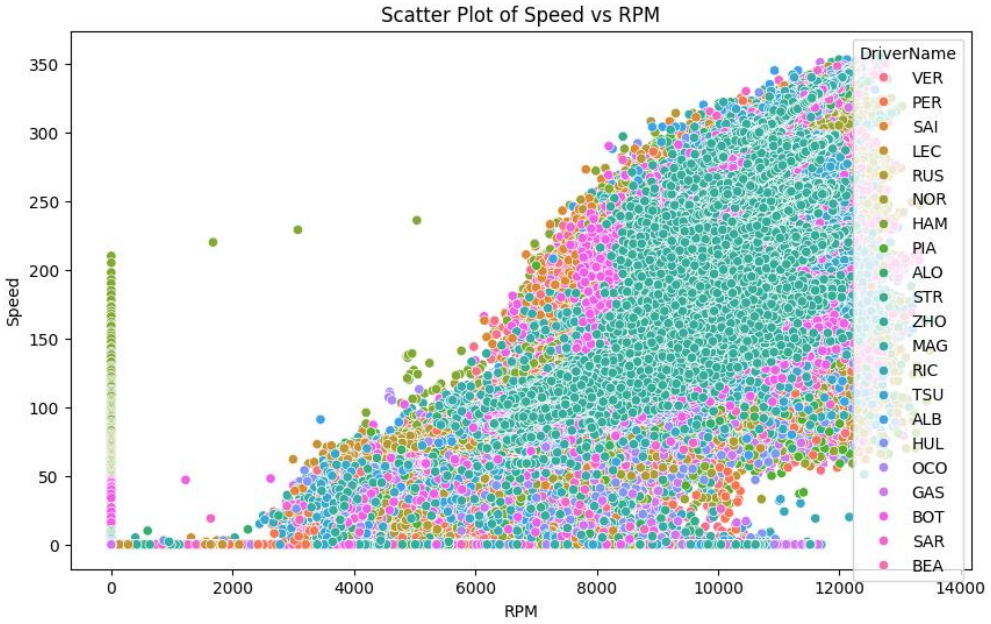
DRS is a system designed to reduce aerodynamic drag on the car, thereby increasing speed under certain conditions. The peak at 0 suggests that the DRS is not activated for the majority of the time. This is expected, as DRS can only be used in specific zones on the track during races and under certain conditions.

Datasets Car Data 2024

The right Figure shows Pearson correlation between numeric features of car dataset. The following insights can be derived from it.

- **Speed**, **RPM**, and **nGear** are closely related. As expected in a mechanical system like a car, these features are interdependent.
- Throttle has a moderate impact on both RPM and Speed, which is logical as the throttle controls the engine's power output.
- DRS (Drag Reduction System) shows little correlation with other features, indicating it might be used independently to reduce drag and increase speed without directly affecting RPM or throttle significantly.





The left figure displays a scatter plot of speed and RPM for each driver, illustrating their correlation from another perspective. As shown, these two features increase uniformly, indicating a high correlation, as mentioned earlier.

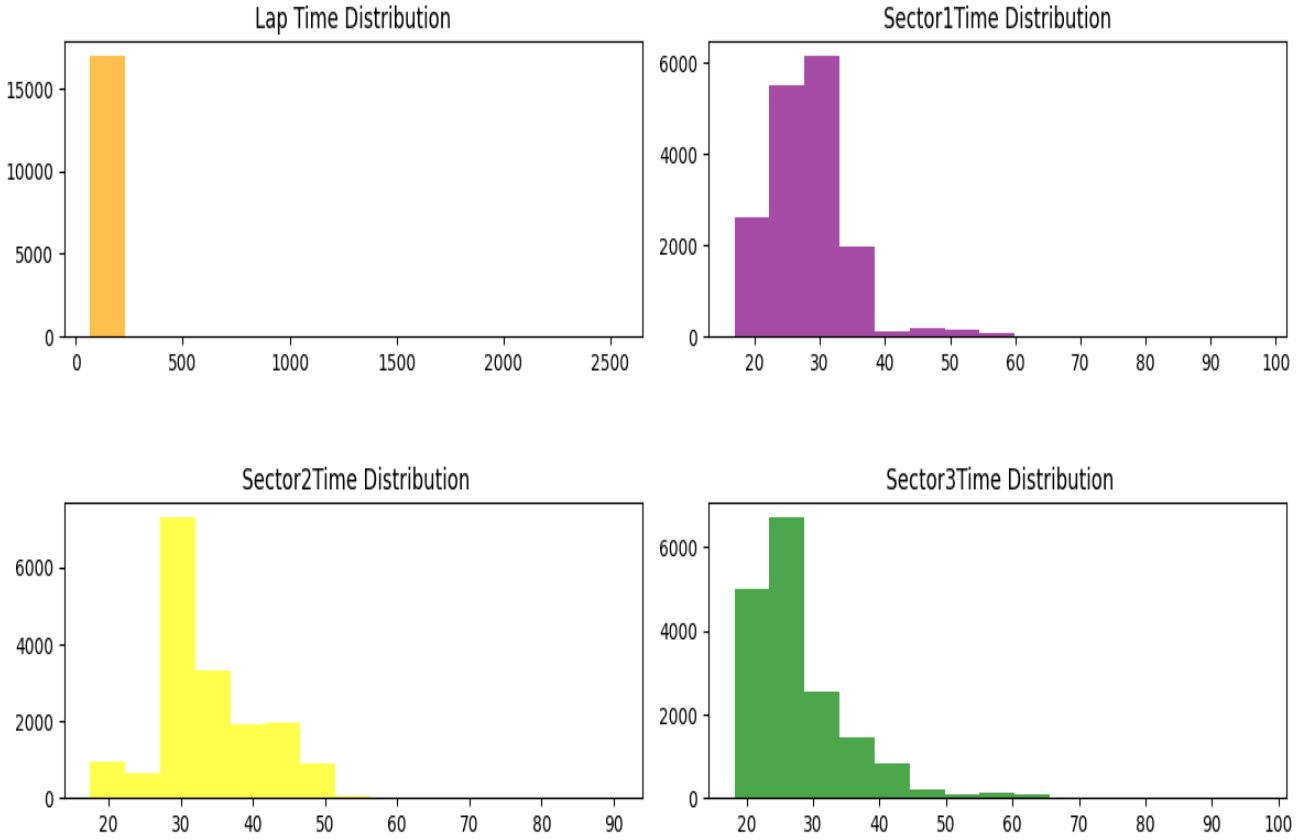
Datasets Lap 2024

This image shows columns statistics of lap-2024 dataset. Next to each column name , nnull and nunique shows the number of null values and unique values of corresponding column. Red element indicates the data-type of each column and last blue element shows the list of all unique values of the column.
This Lap data contains **17123** records with **32** features. Time columns originally are in time-delta format however they had string values in dataset which were preprocessed to be converted to time-delta and then add new columns for each of them which shows corresponding time in seconds (last columns).

96.76% values of PitOutTime and **96.75**% values of PitInTime are null. Also SpeedI1, SpeedFL and SpeedST columns have **16.1**% , **3.3**% and **7.5**% null values respectively. Other columns are populated relatively complete. Based on the image, there are **10 Teams**, **5** type of tire **compound**, **5 stints** , and the maximum number of **laps** is **78**.

Datasets

Lap 2024

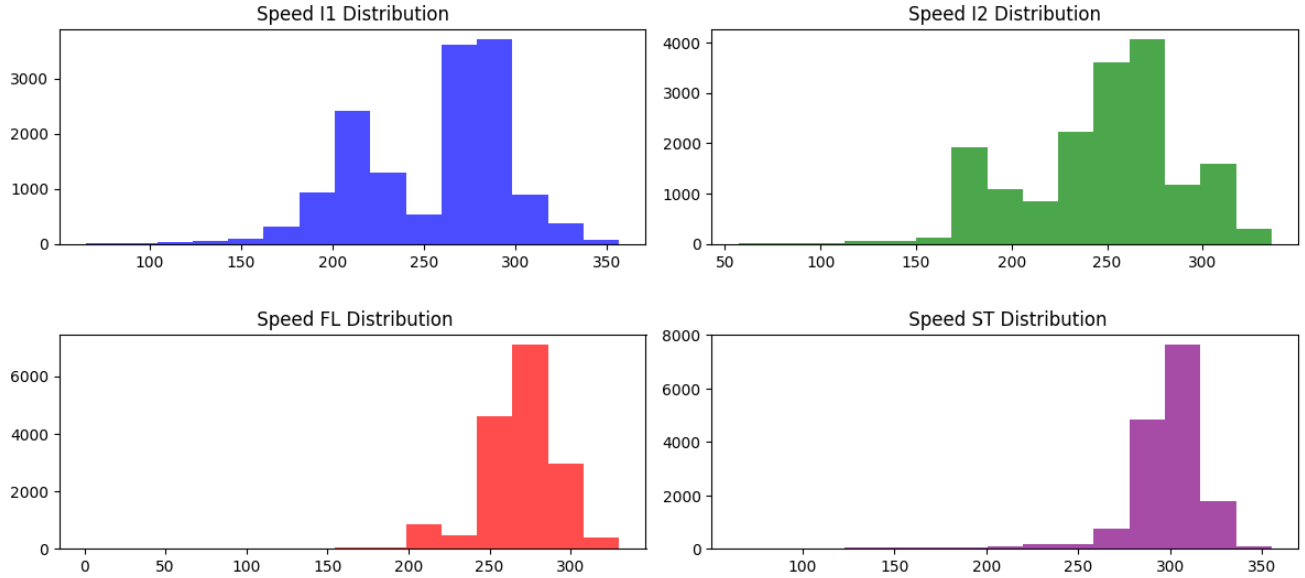


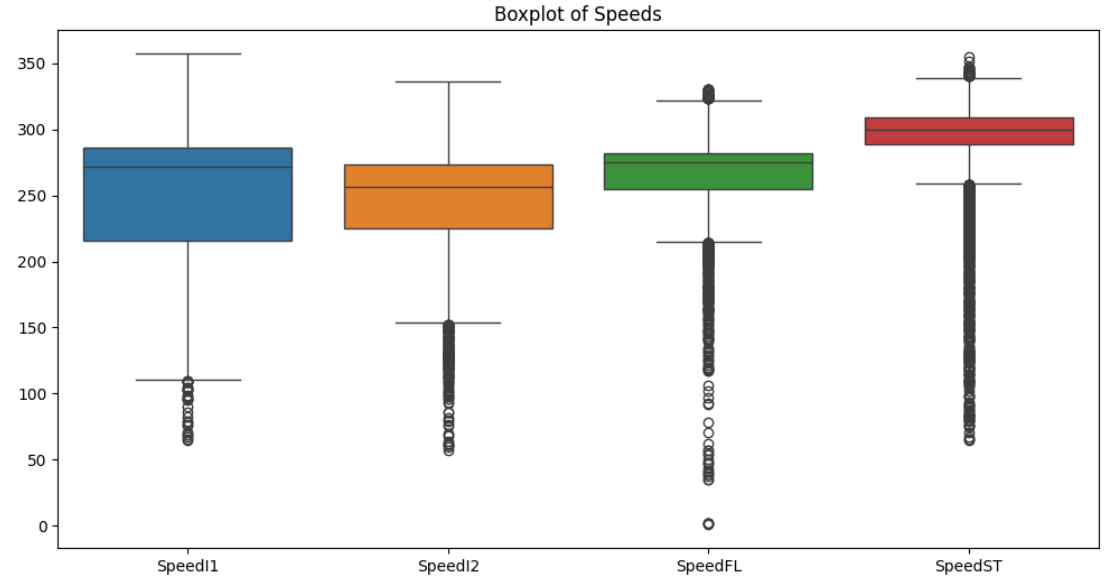
This histogram plot shows the distribution of lap-time. sector1Time , sector2Time and sector3Time. The x-axis shows time in seconds y-axis is the count of each bins. Lap-time histogram indicates all the laps takes about 50-300 seconds. Sector1Time usually take 20 to 40 seconds , Sector2Time 20 to 50 seconds where it peaks at 30 to 35 seconds and Sector3Time takes 20 to 45 seconds.

It makes sense since each lap is divided into 3 sectors and lap-time is the sum up of these sectors-time and timing distribution prove this facts.

Another point is that it seems sector2time takes abit more time in average than the two others because it’s peak is near 30 seconds and after, while the others are maximized in around 20 to 30 seconds.

Datasets Lap 2024

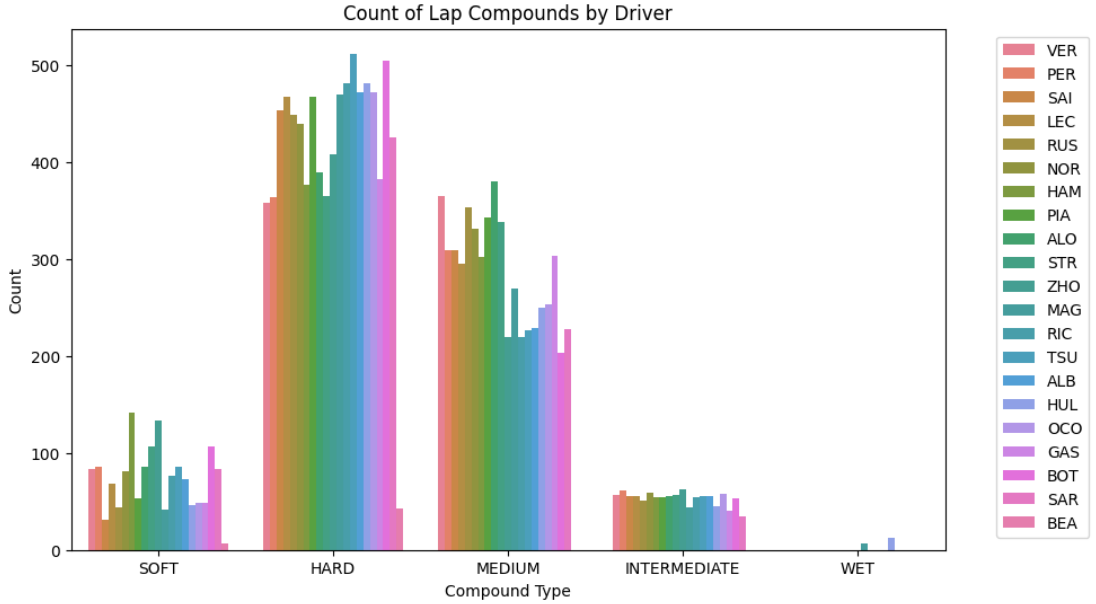


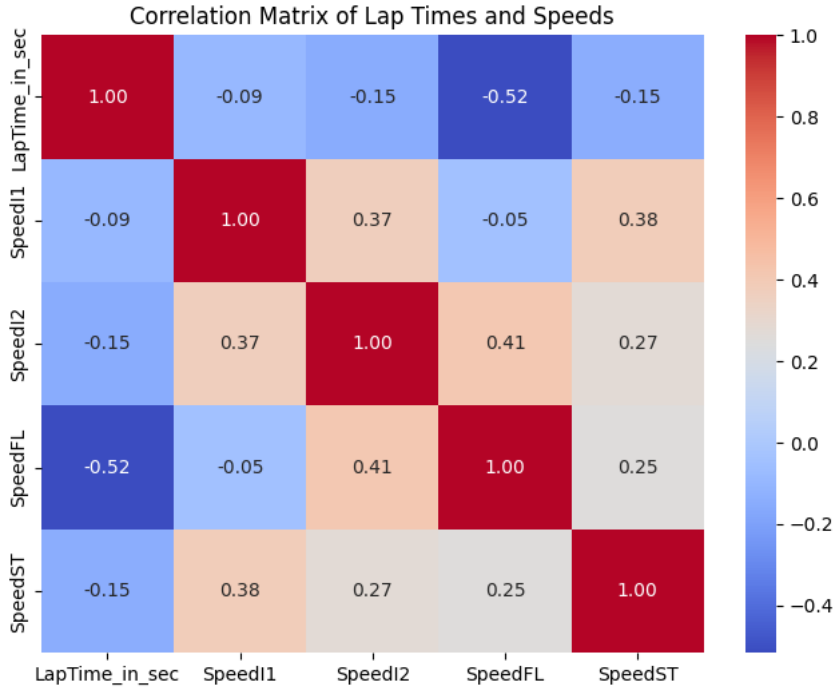


Left figure is the histogram of 4 speed values recorded in lap dataset and the right box plot shows these values in another perspective. Based on these two Figures, Most Speed I1 are clustered between 250 km/h and 300 km/h. Similar to Speed I1, speeds I2 are mostly between 250 km/h and 300 km/h. The majority of Speed FL at the finish line are concentrated between 250 km/h and 300 km/h. The Speed ST on the longest straight are higher, predominantly between 300 km/h and 350 km/h. Overall, the graphs suggest that the longest straight (Speed ST) allows for the highest speeds, while other sectors show similar distributions with slightly lower peak speeds.

Datasets Lap 2024

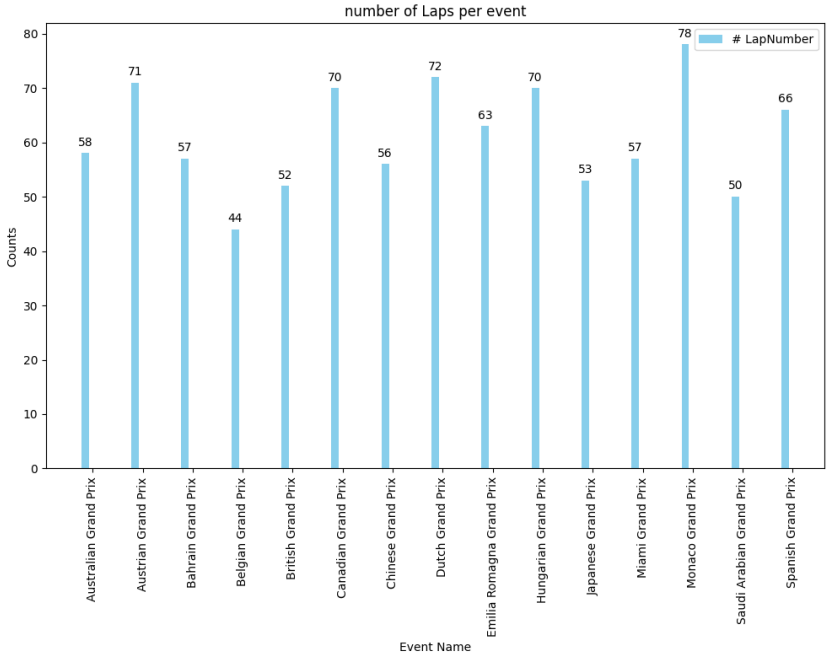
he count plot visualizes the usage of different tire compounds by each driver. The plot indicates that Hard and Medium are the most frequent used compound across most driver, and Wet is the least frequent used. Soft and Intermediate are also less used.
The distribution indicates a heavier reliance on harder compounds, possibly for durability.
The low usage of SOFT and WET compounds may reflect strategic choices depending on track conditions or race strategy.

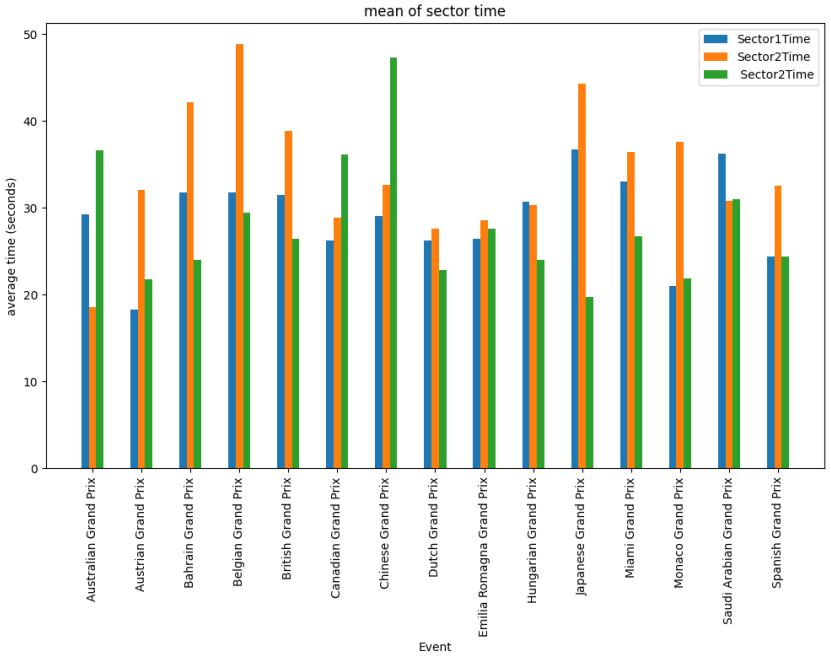


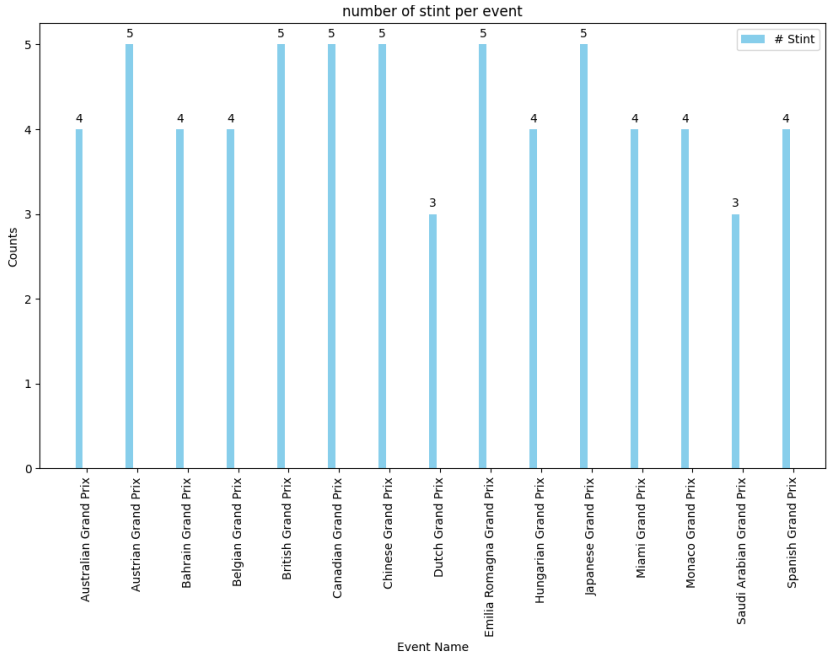


Based on the correlation matrix, speed features have negative moderate correlation between with lapTime. and speedFL with **-0.52** is the most correlated feature. It means that have a high speed in finsh line influences on lap-Time decrease more than others.
the four speed features are moderately correlated with each other.

Datasets Lap 2024

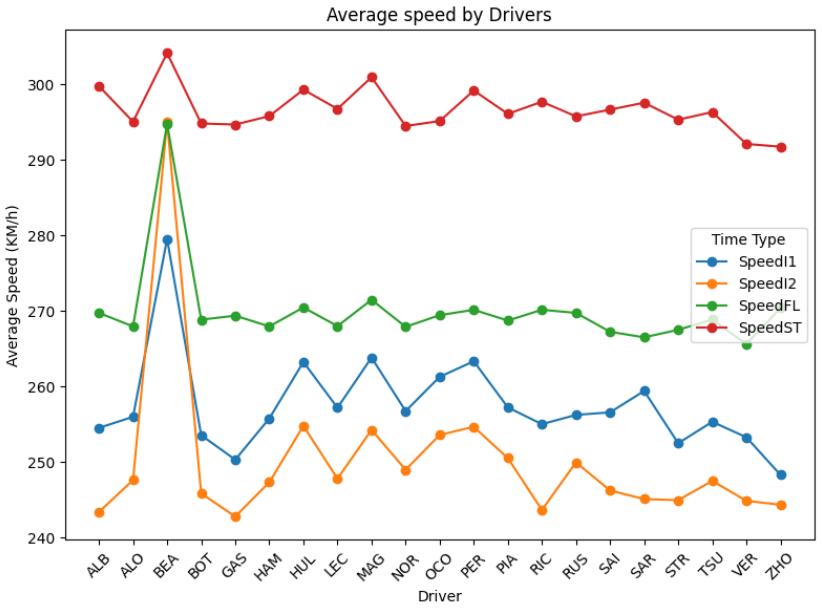




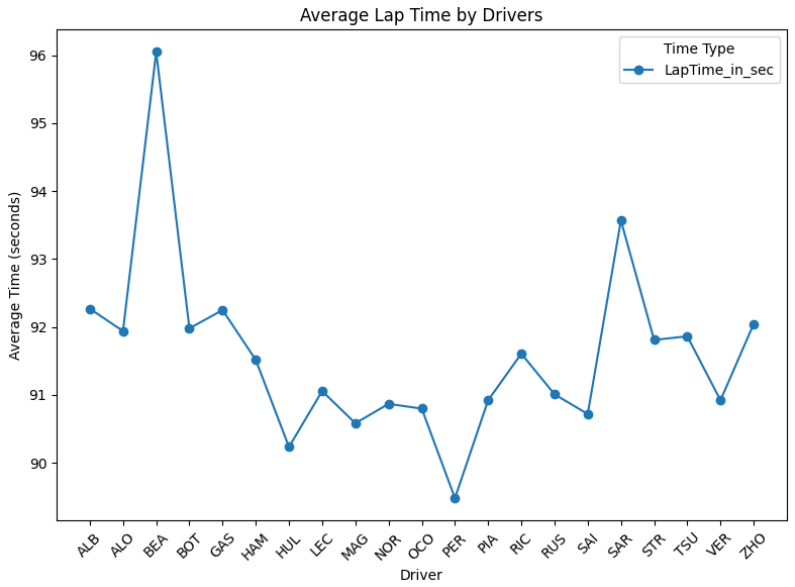


The left figure is the number of laps for each events. Monaco Grand Prix has the maximum number of laps among all event with 78 laps and Belgian Grand Prix has the least number of laps with 44 laps.
The middle Figure illustrates the average time of each sectors for all events. Austrian, Bahrain, British, Japanese , Monaco , Miami , Spanish and Belgian events have spent much higher time in sector2 rather than other two sectors. Australian , Canadian and Chinese events however have spent more time in sector 1. Saudi Arabian event on the other hand is the only event which drivers have spent more time in sector1 in average.
The right Figure illustrate the number of stints per event . Dutch and Saudi Arabian are the only events with 3 stints. The others have either 4 or 5 Stints

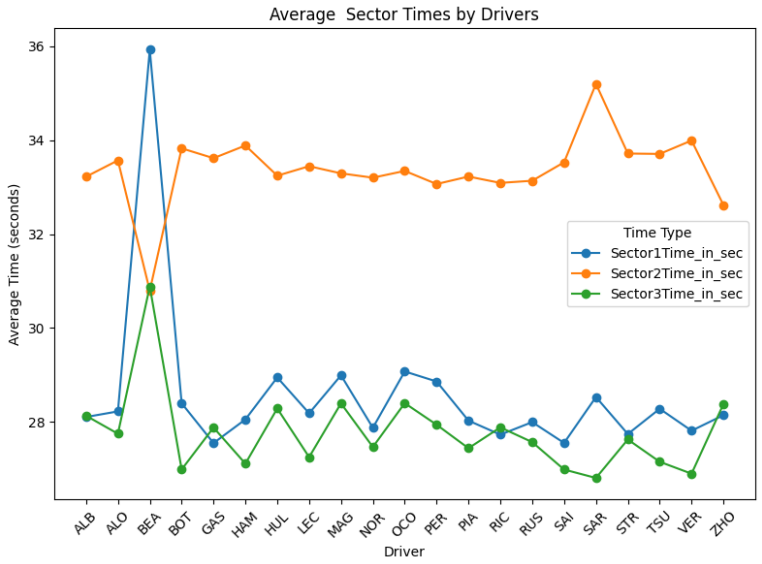
Datasets Lap 2024



The above figure shows average Speed (KM/h) per driver. First, one can seen that speed on longest straight is much higher than other speeds and sector2speed is the lowest speed. second, BEA had much higher speeds in all 4 speed types than others

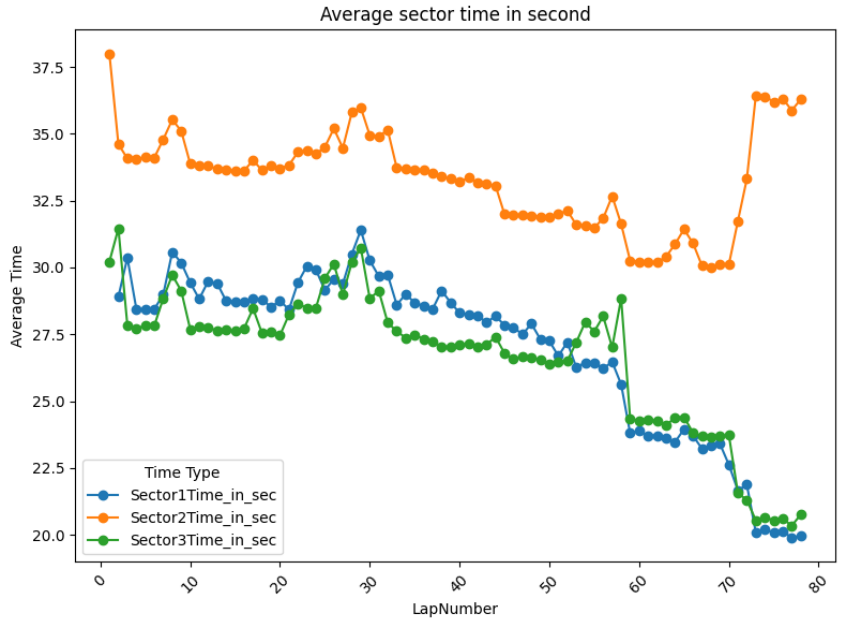


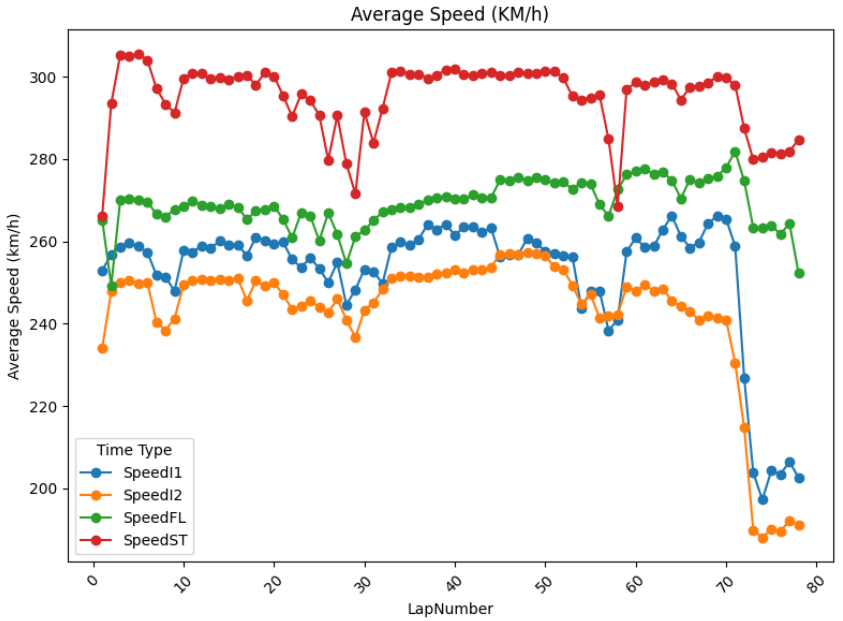
The above figure shows average lap time per driver. Based on this , BEA has spent the most time in laps which PER has spent the lowest time in Laps in average.
It's weird because BEA had higher speed relative to others nonetheless he spent more time in laps !!!

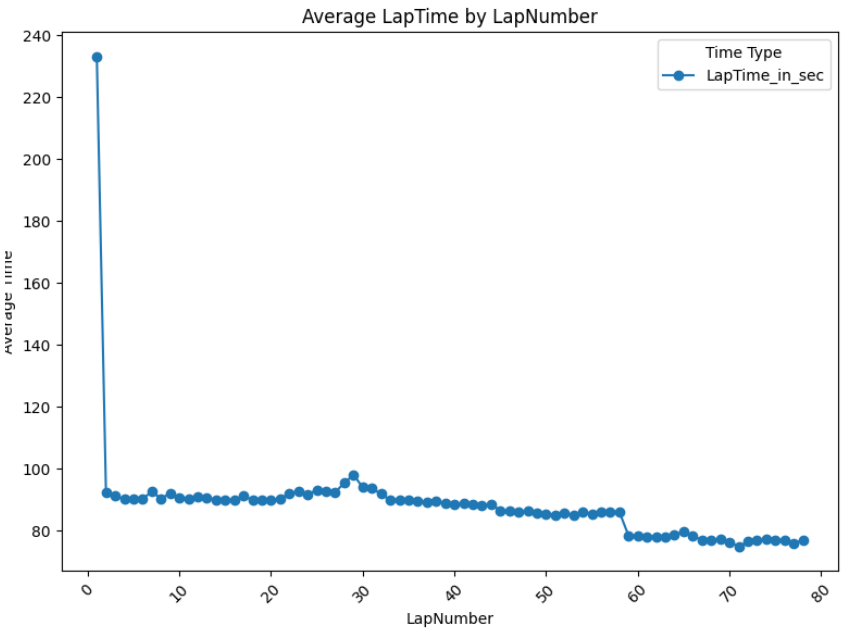


The above figure shows average sector time per driver. In all events As it can be seen, except BEA which has spent much more time in Sector1, the other drivers have spent considerable more time in Sector 2

Datasets Lap 2024







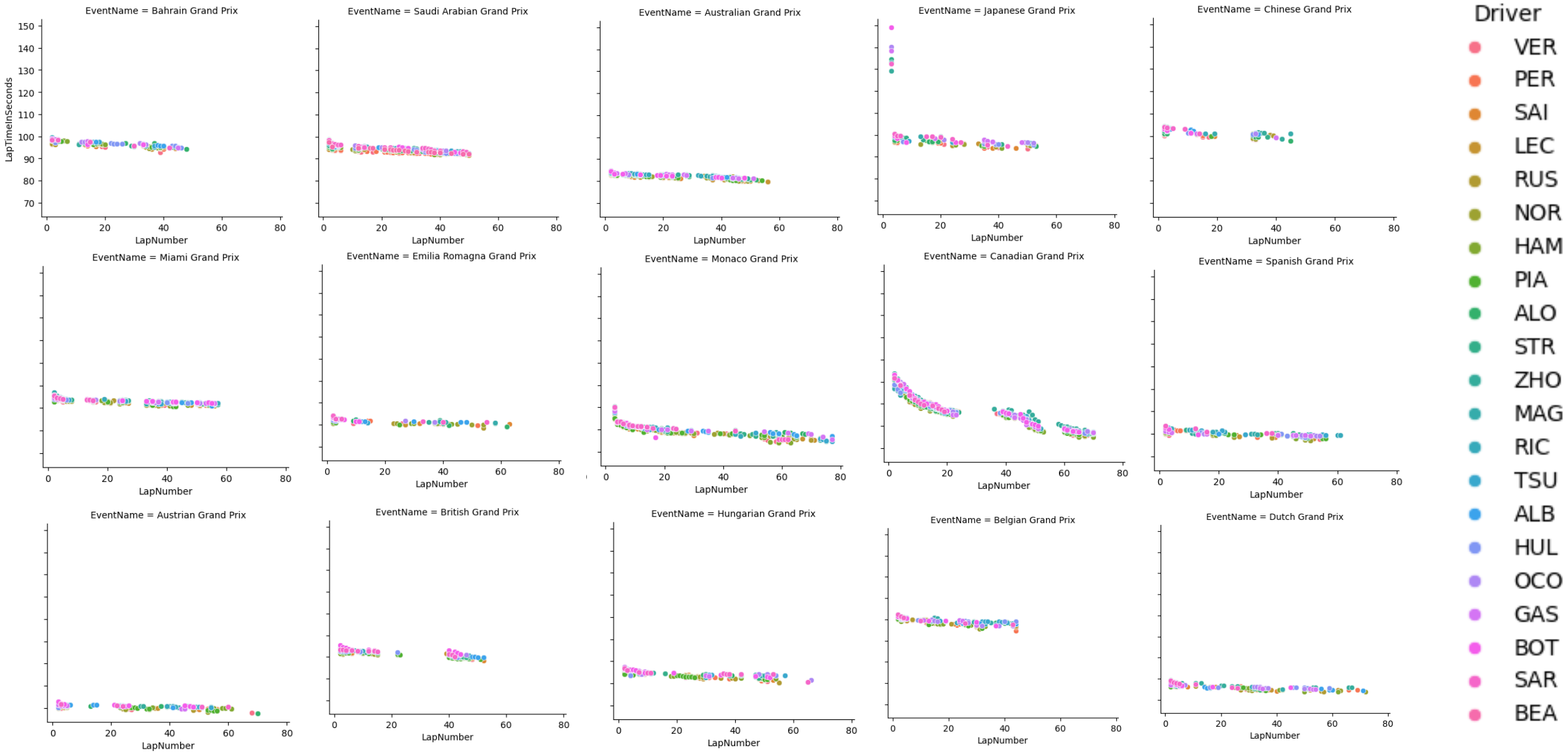
These three figures illustrates various factors based on lap number. The left figure shows the average sector time in each lap number. Based on this we can conclude that as drivers reached to the final laps average time in sector1 and 3 declines, but there is no specific pattern in sector2time.

Middle figure shows average speed per lap. As can be seen speed declines sharply around lap 70 which can have several reason including : Tire Degradation, Fuel Management, Strategic Positioning , Safety Car or Caution Periods or Navigating through traffic or lapping slower cars. The right figure also illustrate average lap-time per lap. It declines uniformly and very slightly over time as lap number grows.

The figures of next slide also illustrate lap-time per lap-number for each event separately. From those figures also one can conclude that lap-time decreases slightly over time.

Datasets

Lap 2024

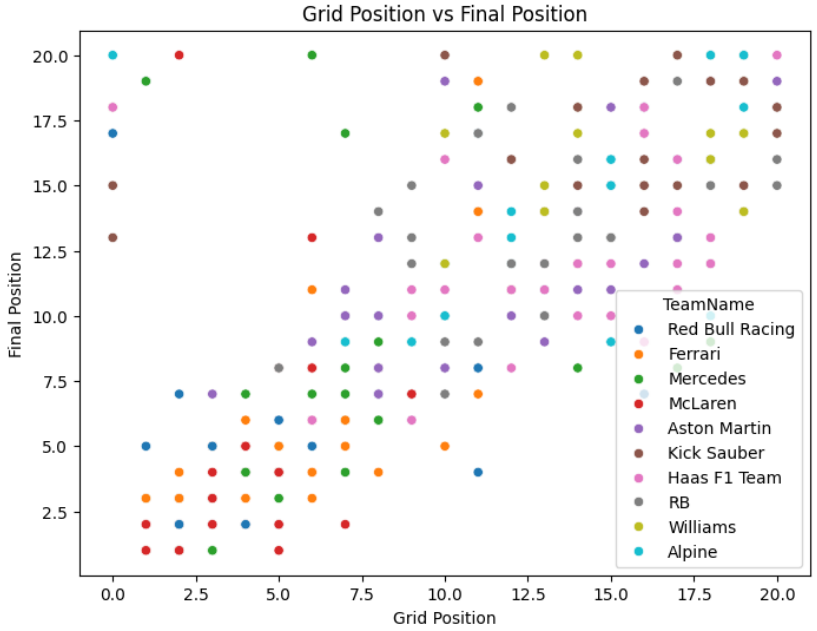


Datasets

Results 2024

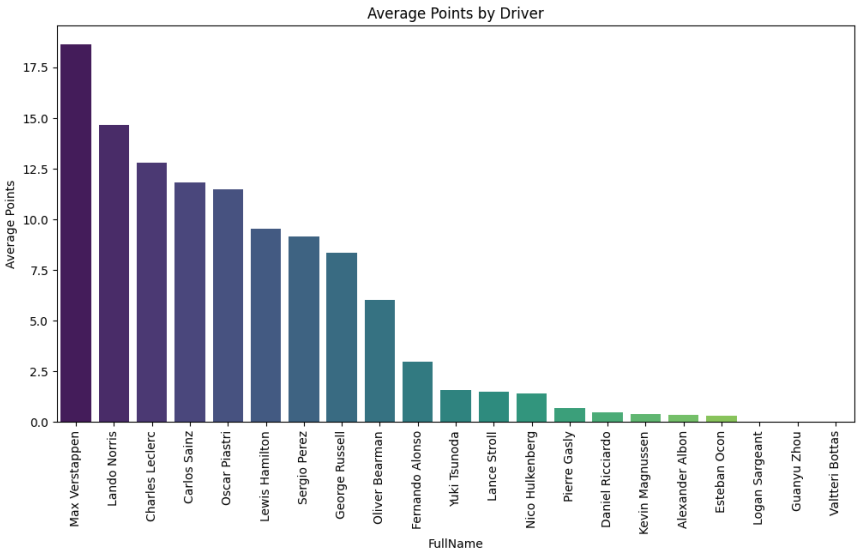
```
result_2024 :  (299, 22)
EventName :  |nnull: 0  |nunique: 15  | object |['Bahrain Grand Prix', 'Saudi Arabian Grand Prix', 'Australian Grand Prix', 'Japanese Grand Prix', 'Chinese Grand Prix', 'Miami Grand Pri
DriverNumber :  |nnull: 0  |nunique: 21  | int64 |[1, 11, 55, 16, 63, 4, 44, 81, 14, 18, 24, 20, 3, 22, 23, 27, 31, 10, 77, 2, 38]
BroadcastName :  |nnull: 0  |nunique: 21  | object |['M VERSTAPPEN', 'S PEREZ', 'C SAINZ', 'C LECLERC', 'G RUSSELL', 'L NORRIS', 'L HAMILTON', 'O PIASTRI', 'F ALONSO', 'L STROLL', 'G ZH
Abbreviation :  |nnull: 0  |nunique: 21  | object |['VER', 'PER', 'SAI', 'LEC', 'RUS', 'NOR', 'HAM', 'PIA', 'ALO', 'STR', 'ZHO', 'MAG', 'RIC', 'TSU', 'ALB', 'HUL', 'OCO', 'GAS', 'BOT',
DriverId :  |nnull: 0  |nunique: 21  | object |['max_verstappen', 'perez', 'sainz', 'leclerc', 'russell', 'norris', 'hamilton', 'piastri', 'alonso', 'stroll', 'zhou', 'kevin_magnussen',
TeamName :  |nnull: 0  |nunique: 10  | object |['Red Bull Racing', 'Ferrari', 'Mercedes', 'McLaren', 'Aston Martin', 'Kick Sauber', 'Haas F1 Team', 'RB', 'Williams', 'Alpine']
TeamColor :  |nnull: 0  |nunique: 19  | object |['3671C6', 'e8002d', '27f4d2', 'ff8000', '229971', '52e252', 'b6babd', '6692ff', '64c4ff', 'ff87bc', 'E80020', 'FF8000', '3671C6', '6692F
TeamId :  |nnull: 0  |nunique: 10  | object |['red_bull', 'ferrari', 'mercedes', 'mclaren', 'aston_martin', 'sauber', 'haas', 'rb', 'williams', 'alpine']
FirstName :  |nnull: 0  |nunique: 21  | object |['Max', 'Sergio', 'Carlos', 'Charles', 'George', 'Lando', 'Lewis', 'Oscar', 'Fernando', 'Lance', 'Guanyu', 'Kevin', 'Daniel', 'Yuki', 'Al
LastName :  |nnull: 0  |nunique: 21  | object |['Verstappen', 'Perez', 'Sainz', 'Leclerc', 'Russell', 'Norris', 'Hamilton', 'Piastri', 'Alonso', 'Stroll', 'Zhou', 'Magnussen', 'Ricciard
FullName :  |nnull: 0  |nunique: 21  | object |['Max Verstappen', 'Sergio Perez', 'Carlos Sainz', 'Charles Leclerc', 'George Russell', 'Lando Norris', 'Lewis Hamilton', 'Oscar Piastri',
HeadshotUrl :  |nnull: 1  |nunique: 20  | object |['https://media.formula1.com/d_driver_fallback_image.png/content/dam/fom-website/drivers/M/MAXVER01_Max_Verstappen/maxver01.png.transfo
CountryCode :  |nnull: 0  |nunique: 15  | object |['NED', 'MEX', 'ESP', 'MON', 'GBR', 'AUS', 'CAN', 'CHN', 'DEN', 'JPN', 'THA', 'GER', 'FRA', 'FIN', 'USA']
Position :  |nnull: 0  |nunique: 20  | float64 |[1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0, 18.0, 19.0, 20.0]
ClassifiedPosition :  |nnull: 0  |nunique: 23  | object |['1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', 'R', 'W', 'D']
GridPosition :  |nnull: 0  |nunique: 21  | float64 |[1.0, 5.0, 4.0, 2.0, 3.0, 7.0, 9.0, 8.0, 6.0, 12.0, 17.0, 15.0, 14.0, 11.0, 13.0, 10.0, 19.0, 20.0, 16.0, 18.0, 0.0]
Q1 :  |nnull: 299  |nunique: 0  | float64 |[nan]
Q2 :  |nnull: 299  |nunique: 0  | float64 |[nan]
Q3 :  |nnull: 299  |nunique: 0  | float64 |[nan]
Time :  |nnull: 120  |nunique: 179  | object |['0 days 01:31:44.742000', '0 days 00:00:22.457000', '0 days 00:00:25.110000', '0 days 00:00:39.669000', '0 days 00:00:46.788000', '0 days
Status :  |nnull: 0  |nunique: 13  | object |['Finished', '+1 Lap', '+2 Laps', 'Accident', 'Gearbox', 'Engine', 'Brakes', 'Retired', 'Collision', '+7 Laps', 'Water pressure', 'Hydraulic
Points :  |nnull: 0  |nunique: 18  | float64 |[26.0, 18.0, 15.0, 12.0, 10.0, 8.0, 6.0, 4.0, 2.0, 1.0, 0.0, 25.0, 16.0, 19.0, 7.0, 13.0, 11.0, 5.0]
```

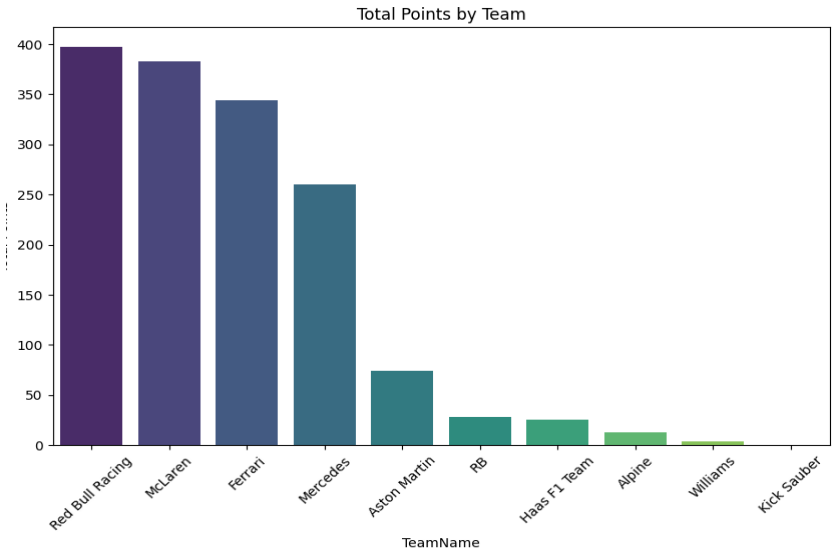
Result-2024 data contains 299 rows with 22 features. Three Q1, Q2 and Q3 features are empty without any records. Other features except Time are fully populated.
There are 13 unique values of Status which shows drivers status on a race. Points also determines the total point a driver earns in each race and has 18 values.

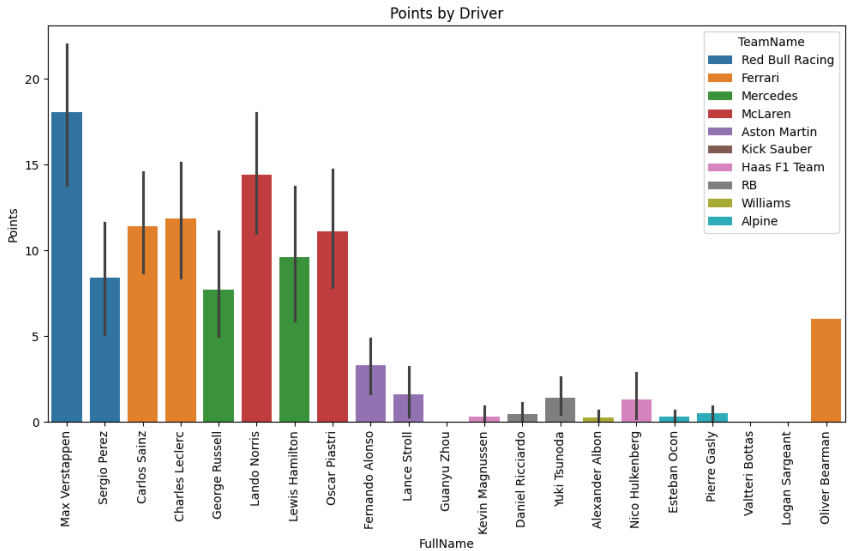


This plot illustrates the correlation between drivers Grid Position in the beginning of each race and their final position at the last of race. Based on the Scatter plot, there is high correlation between these two factors , which mans starting position highly influences on the final result in a race.
The plot shows the correlation for 10 Teams which are initialized with different colors.

Datasets Results 2024





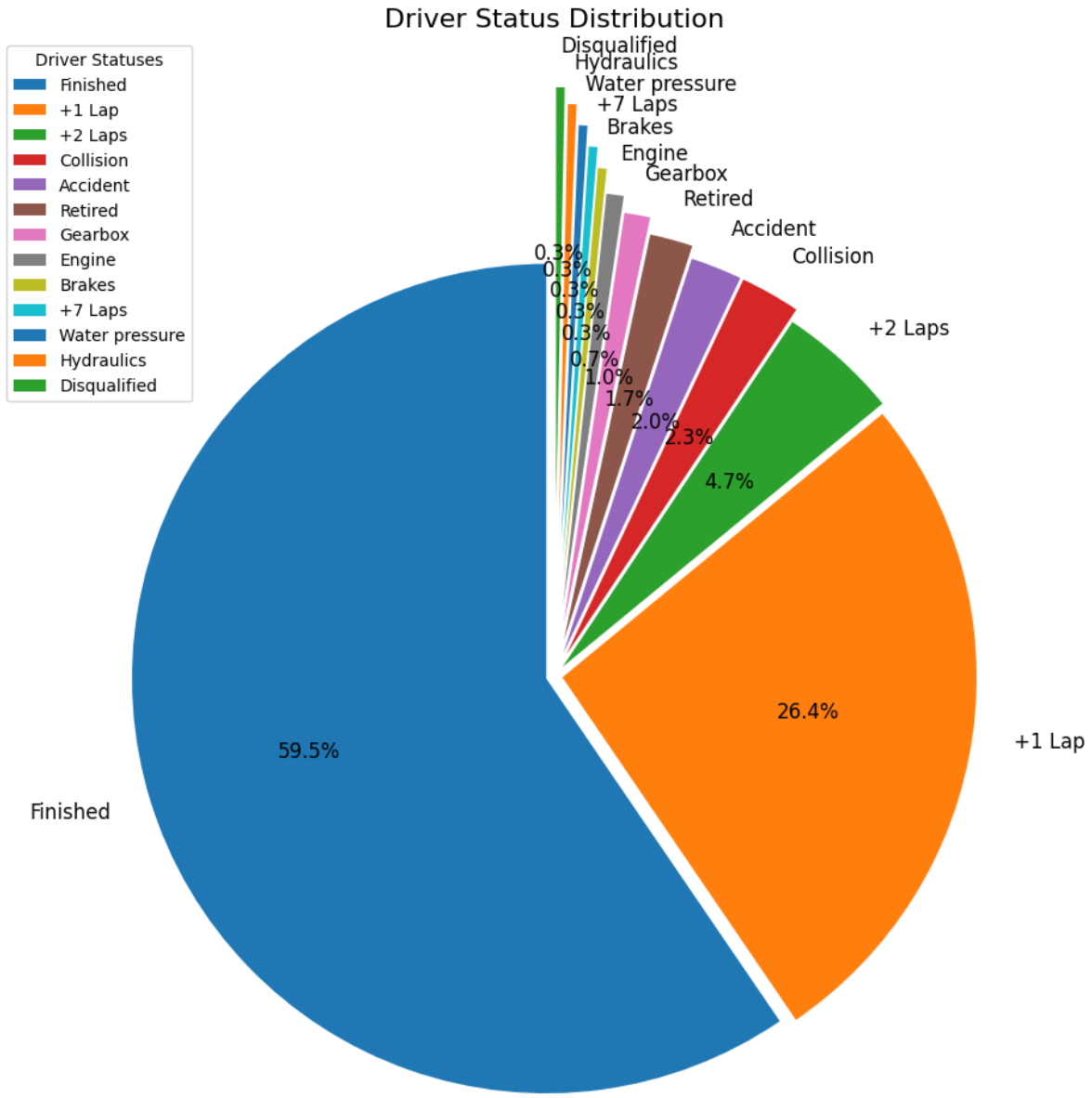


the bar charts provides a visual representation of the points scored by each Formula 1 driver, categorized by their respective teams. The distribution shows a sharp decline in points after the top few drivers, indicating a significant performance gap.

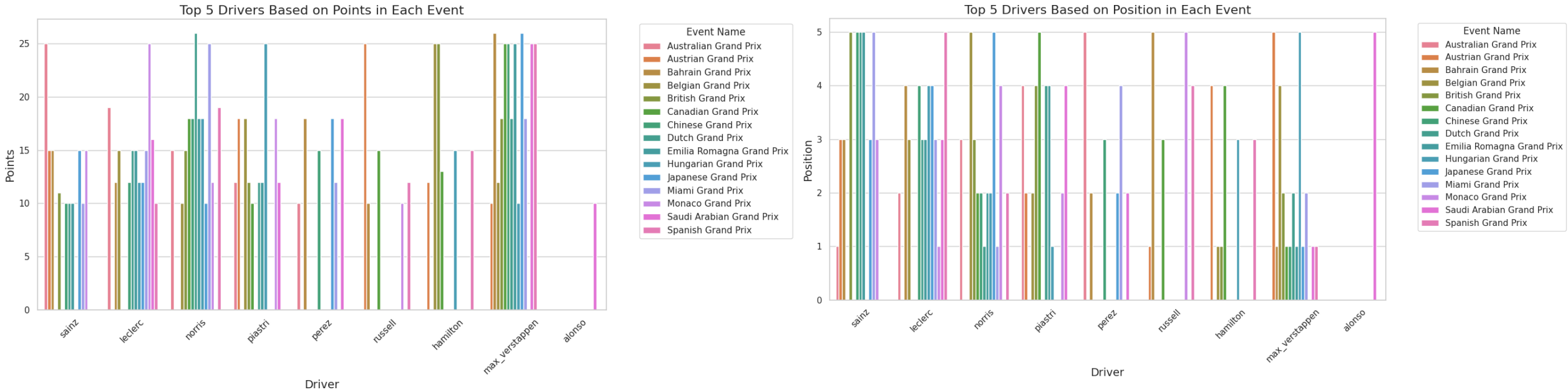
- Max Verstappen from Red Bull Racing leads with the highest points, indicating strong performance across events.
- Drivers from Ferrari and Mercedes, like Carlos Sainz, Charles Leclerc, George Russell, and Lewis Hamilton, also have significant points, showcasing competitive performance.
- Red Bull Racing, Ferrari, and Mercedes appear to be the strongest teams, with multiple drivers scoring high points.
- McLaren, represented by Lando Norris and Oscar Piastri, also shows competitive performance but slightly behind the top teams.
- Aston Martin, with drivers like Fernando Alonso and Lance Stroll, sits in the midfield, indicating a moderate performance.
- Teams like Kick Sauber and Alpine have lower points, reflecting challenges in competing with the top teams
- Drivers from Williams, Haas F1 Team, and RB are at the lower end of the points spectrum, suggesting they might be struggling in the championship.

Datasets Results 2024

The Pie chart provides information about drivers status distribution in all race. Finishing the race is the dominant status that indicates 59.5% of times drivers finish the race.
The second most common status is '+1 Lap,' which shows that 26.4% of drivers have completed at least one lap. According to the chart, only 4.7% of the time do drivers complete the second lap, while 4.3% of the time, accidents or collisions occur during the races.
Additionally, drivers retire from the race 1.7% of the time. Other statuses, such as gearbox issues, +7 laps, and brake problems, account for less than 1% each
In summary, the pie chart reveals that the majority of drivers (59.5%) finish races, while a significant portion (26.4%) complete at least one lap. However, incidents such as accidents and retirements remain relatively low, highlighting overall race completion rates.



Datasets Results 2024



The above bar charts comparing the top 5 Formula 1 drivers based on points and positions across various events.

Left Chart:

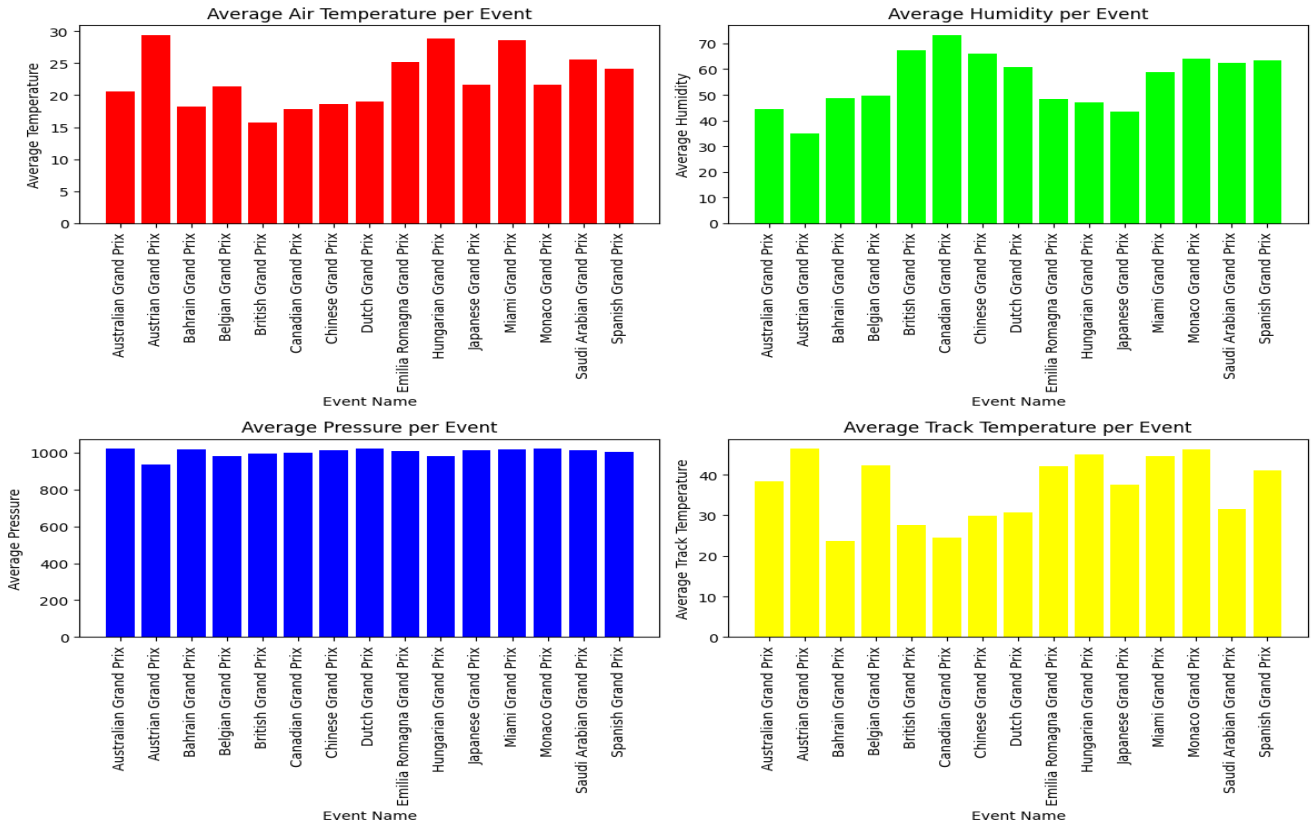
1. Max Verstappen stands out with consistently high points across many events, indicating strong and reliable performance.
2. Lewis Hamilton and Sergio Perez also show high points in multiple events, reflecting their competitiveness.
3. There is noticeable variability in points across events for drivers like Charles Leclerc and George Russell, suggesting fluctuating performances.
4. Some events, like the Belgian and British Grand Prix, have higher point distributions, possibly indicating races where certain drivers excelled.

Right Chart:

1. Max Verstappen frequently achieves top positions (1st or 2nd), which aligns with his high points.
2. Carlos Sainz and Lando Norris also secure top positions occasionally, showcasing strong performances in certain races.
3. Fernando Alonso and Oscar Piastri show varying positions, indicating less consistent performance compared to Max Verstappen.
4. Certain drivers perform exceptionally well in specific events, as seen with spikes in top positions for races like the Monaco and Spanish Grand Prix.c

Datasets Weather 2024

```
weather_2024 :  (2288, 9)
EventName :  |nnull: 0  |nunique: 15  | object |['Bahrain Grand Prix', 'Saudi Arabian Grand Prix', 'Australian Grand Prix', 'Japanese G
Time :  |nnull: 0  |nunique: 2288  | object |['0 days 00:00:14.093000', '0 days 00:01:14.084000', '0 days 00:02:14.093000', '0 days 00:
AirTemp :  |nnull: 0  |nunique: 141  | float64 |[18.9, 18.8, 18.7, 18.6, 18.5, 18.4, 18.3, 18.2, 18.1, 18.0, 17.9, 17.8, 17.7, 17.6, 25
Humidity :  |nnull: 0  |nunique: 49  | float64 |[46.0, 45.0, 47.0, 48.0, 49.0, 50.0, 51.0, 67.0, 65.0, 66.0, 64.0, 63.0, 61.0, 62.0, 66
Pressure :  |nnull: 0  |nunique: 123  | float64 |[1017.1, 1017.0, 1016.9, 1017.2, 1017.3, 1017.4, 1017.5, 1012.0, 1012.1, 1012.2, 1012.
Rainfall :  |nnull: 0  |nunique: 2  | bool |[False, True]
TrackTemp :  |nnull: 0  |nunique: 292  | float64 |[26.5, 26.2, 26.1, 26.0, 25.8, 25.7, 25.6, 25.5, 25.4, 25.3, 25.0, 24.9, 24.7, 24.6,
WindDirection :  |nnull: 0  |nunique: 346  | int64 |[162, 55, 85, 178, 56, 155, 15, 51, 147, 84, 64, 26, 143, 36, 216, 37, 27, 48, 140,
WindSpeed :  |nnull: 0  |nunique: 60  | float64 |[0.9, 1.0, 1.1, 1.5, 0.8, 1.6, 1.3, 1.4, 0.6, 0.5, 1.2, 0.4, 0.7, 1.8, 2.0, 2.1, 0.0,
```



Weather dataset contains 2288 rows with 9 columns all fully populated. This dataset reports the weather of each events.

Based on the below bar chart :

1.Average Air Temperature per Event:

- Temperatures vary across events, with some like the Bahrain Grand Prix showing higher averages.
- The Australian, Miami, and Saudi Arabian Grand Prix also have relatively high air temperatures.

2.Average Humidity per Event:

- Humidity is notably high at the Chinese and Japanese Grand Prix.
- Events like the Australian and Austrian Grand Prix have lower humidity levels.

3.Average Pressure per Event:

- Pressure remains relatively consistent across all events, with slight variations.
- There is no significant deviation, indicating stable atmospheric conditions.

4.Average Track Temperature per Event:

- Track temperatures are highest at the Bahrain and Saudi Arabian Grand Prix, paralleling higher air temperatures.
- Lower track temperatures are observed at events like the British and Belgian Grand Prix.

Key Findings

Key Findings

1. Dataset Composition: The analysis utilized a comprehensive dataset with over 6 million records pertaining to car performance, lap times, and weather conditions across various events in the 2024 Formula 1 season.
2. Speed and RPM Correlation: There is a strong correlation between speed, RPM, and gear (nGear), indicating that as RPM increases, so does the speed of the car. Throttle position also moderately affects both RPM and speed.
3. Lap Time Distribution: The average lap time varies significantly across different tracks, with most laps taking between 50 to 300 seconds. Sector times indicate that sector 2 generally takes longer than sectors 1 and 3.
4. Tire Compound Usage: Hard and medium tire compounds were predominantly used by drivers, while soft and wet compounds were less frequently utilized, suggesting strategic choices based on track conditions.
5. Tire Compound Performance: Analysis shows that tire compound choice significantly affects lap times. Hard and medium compounds generally result in slower lap times compared to soft compounds, which provide better grip and faster speeds. However, the durability of hard compounds allows for longer stints, making them strategically advantageous in certain race conditions.
6. Stint Duration Impact: The duration of stints has a direct correlation with lap times. Longer stints tend to lead to increased lap times due to tire degradation. Drivers often experience a notable drop in performance after approximately 20 laps on the same set of tires, necessitating pit stops for fresh tires to maintain competitive lap times.
7. Lap Time Trends Across Stints: The data indicates that drivers who utilized shorter stints with more frequent pit stops achieved more consistent lap times. In contrast, those who opted for longer stints often faced variability in their lap times due to tire wear and changing track conditions, highlighting the importance of strategic pit stop planning.
8. Pit Stop Trends: Most drivers made between 2 to 4 pit stops during races, with a few outliers making only one or up to five stops. This indicates a general trend towards fewer stops for certain drivers.
9. Driver Performance Metrics: The analysis highlighted that Max Verstappen consistently scored the highest points across events, while other drivers like Lewis Hamilton and Sergio Perez also performed well but with more variability in their results.
10. Impact of Starting Position: There is a significant correlation between a driver's starting grid position and their final race position, emphasizing the importance of qualifying performance.
11. Weather Conditions: Variations in weather conditions, such as air temperature and humidity, were noted to influence race dynamics. Events like the Bahrain Grand Prix had higher average temperatures compared to others.
12. Sector Speed Analysis: Speeds recorded in different sectors displayed distinct patterns, with the highest speeds occurring on the longest straights, while sector speeds varied more significantly across different drivers.
13. Accident Rates: The analysis indicated low rates of retirements and accidents during races, with finishing the race being the most common outcome for drivers.

Exploratory Data Analysis

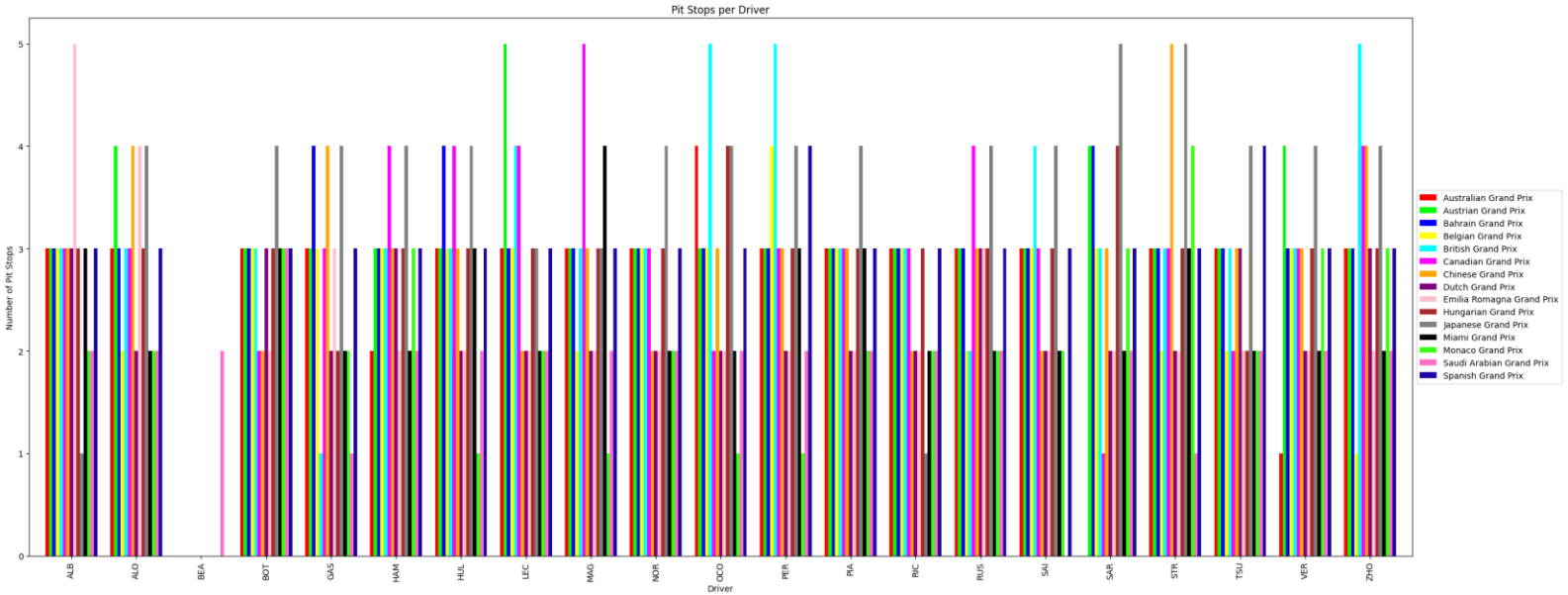
EDA Races 2024 Total Number of Pit Stops

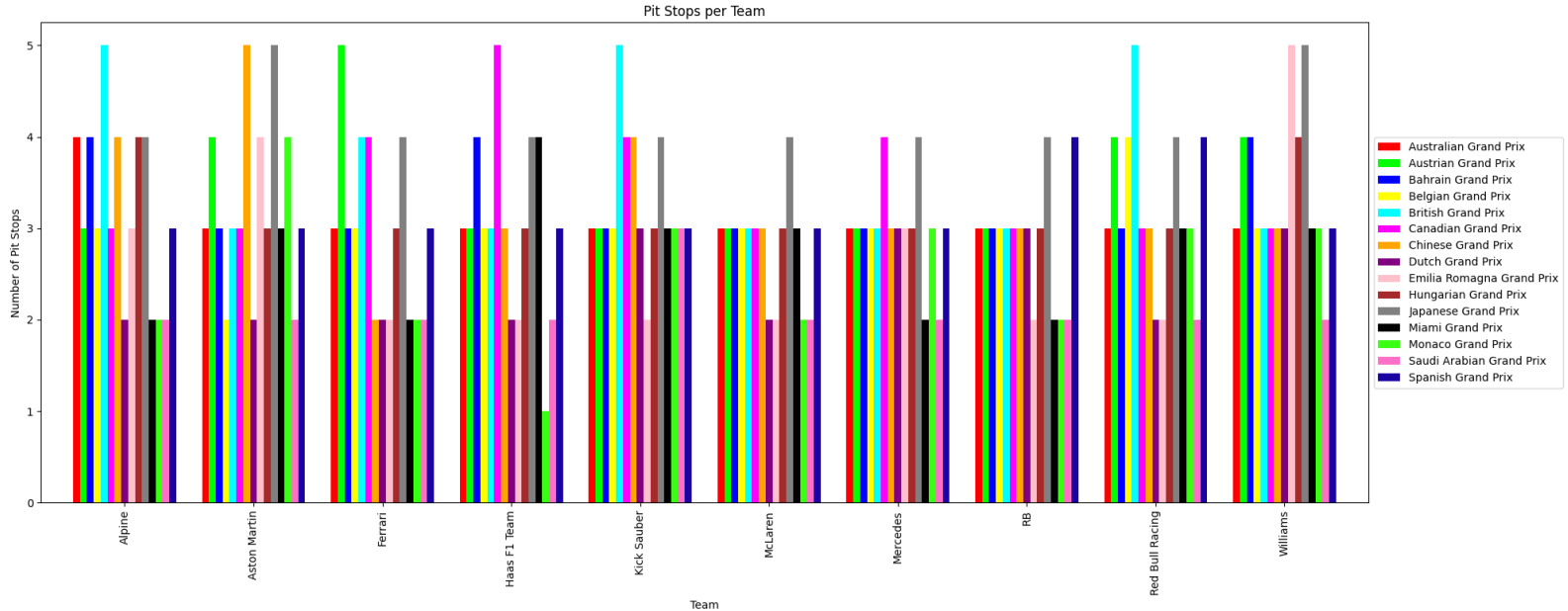
For each driver and team in each race of the 2024 F1 season, determine the total number of pit stops made. Identify any patterns or trends in pit stop frequency.

These two bar charts illustrate the number of pit stops for each driver and team per 2024 events . The number of pit-stops is calculated by **Stint** column in lap dataset that each record shows which stint each driver is in. (Stint is the duration between two consecutive Pit-Stops and can be used to calculate the number of Pit-Stops)

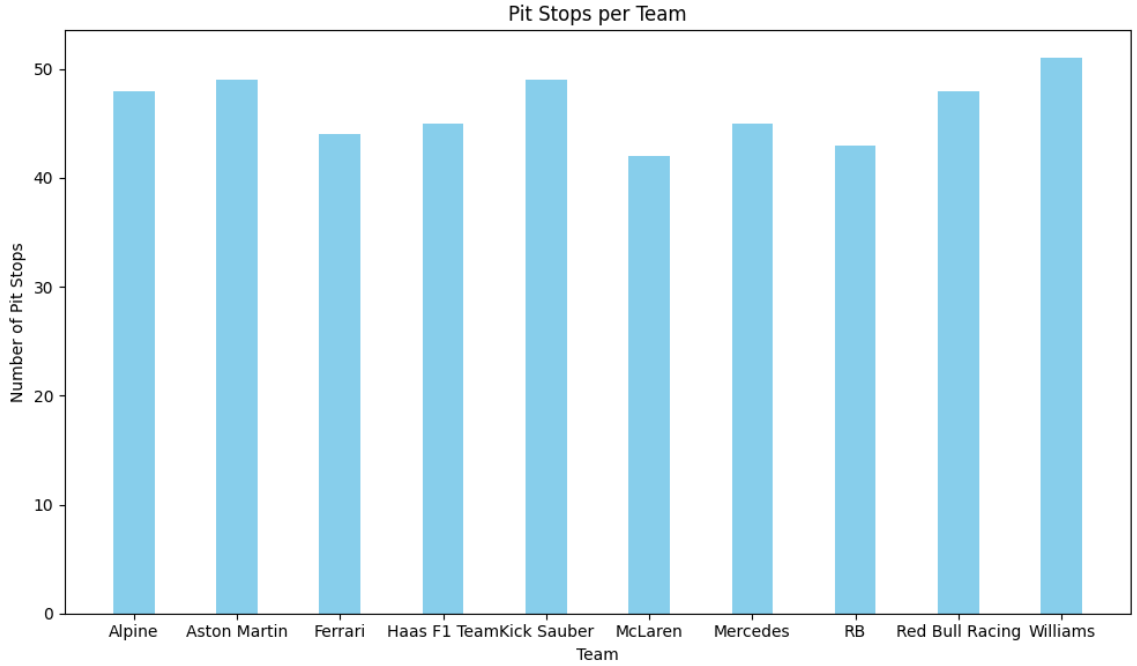
From the above chart which is driver-based, there are a few drivers who stops in 5 pits, more precisely only 9 drivers have stops in 5 pits an all events. However more drivers had 4 Pit-Stops in all races (about 40 time). Most drivers had 2 or 3 Pit Stops in each event. Less drivers had also 1 Pit-Stop and 3 Pit Stops is more frequent.

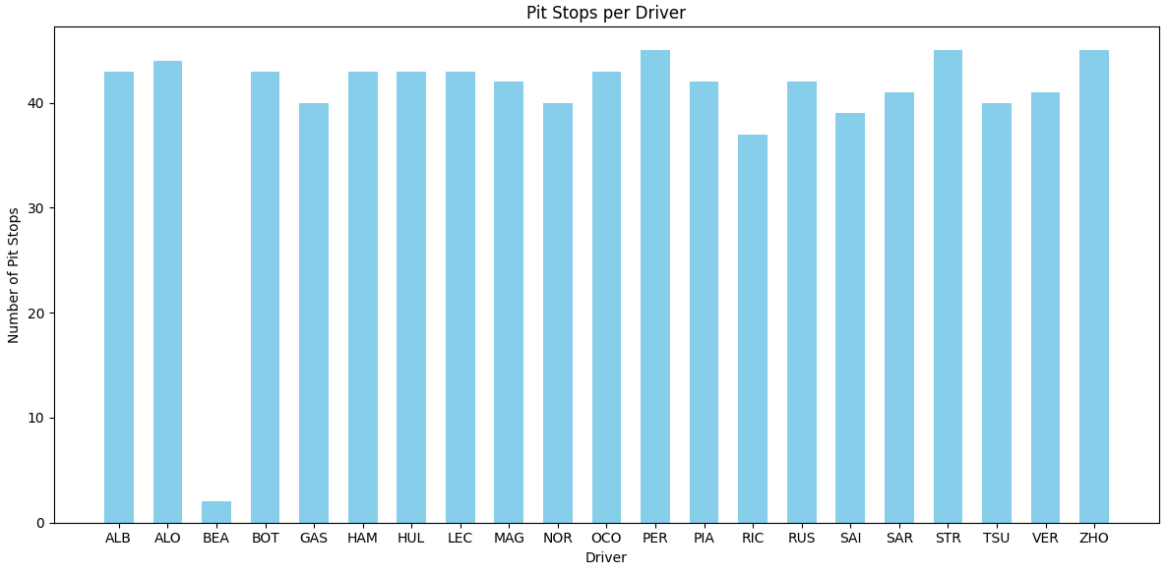
The below charts shows Pit-Stops based on team. It indicates that 9 teams had 5 Pit-Stop during all races but more team had 4 Pit-Stop. However teams with 3 Pit-Stop are much more which is the most frequent number of Pit-Stop for each team in events. Only one team had 1 Pit-Stop (Haas F1 Team in Monaco Grand Prix)





EDA Races 2024 Total Number of Pit Stops





For each driver and team in each race of the 2024 F1 season, determine the total number of pit stops made.

The below charts illustrate total number of Pit-Stops per driver and team over all 2024 events. All of drivers except BEA had near 40 Pit-Stops just a few less or more that indicates their strategy in pit stops are relatively the same. BEA is outlier and his total number of Pit-Stops is less than 5 which mean he doesn't believe in changing tires at all :)
Among the other drivers RIC has the fewest Pit-Stops around 36.
By looking at the teams, on the other hands, all of them have between 40 to 50 Pit-Stops which indicates Pit-Stops's same strategy for teams.

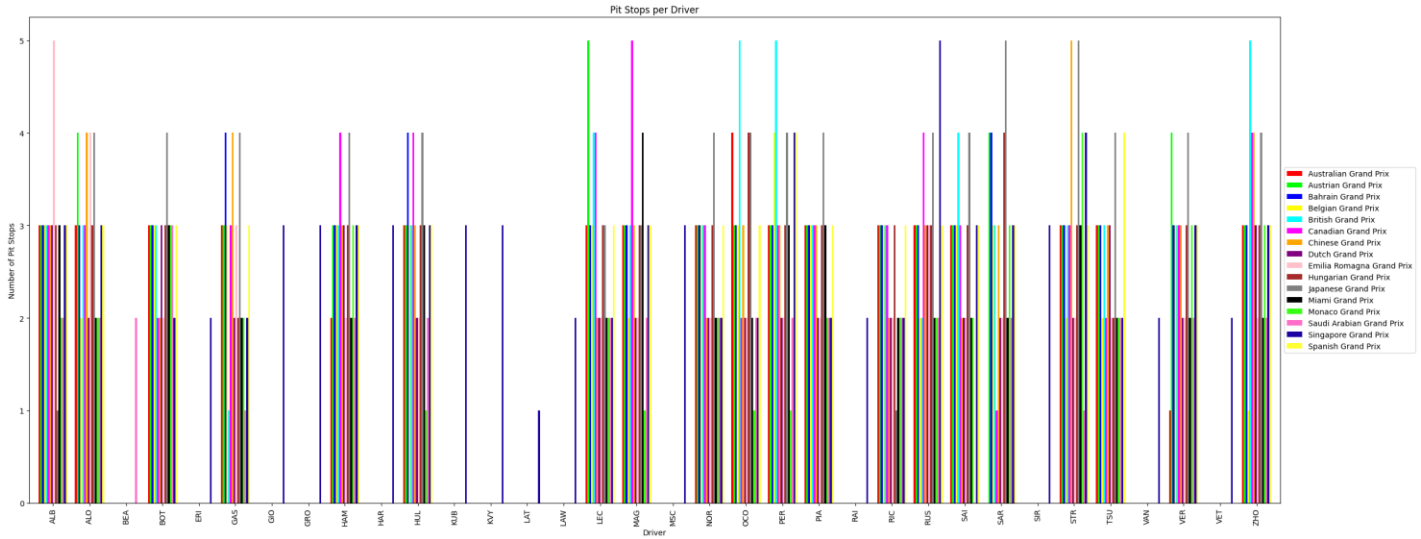
EDA > Merged Data > Total Number of Pit Stops

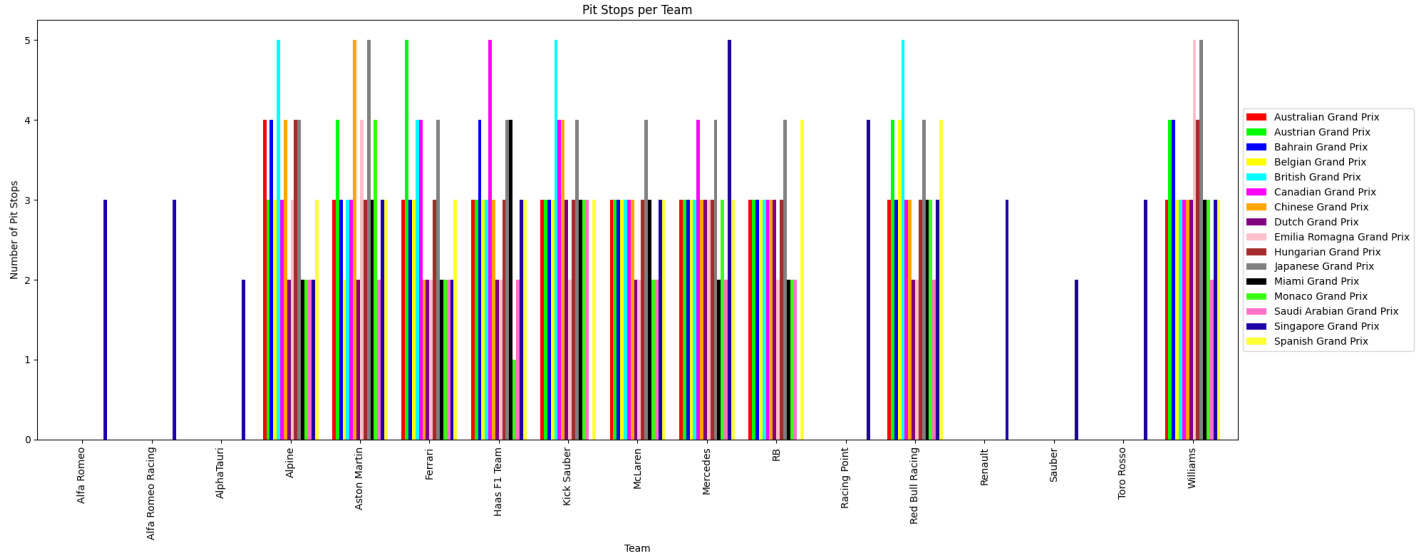
For each driver and team in each race of the 2024 F1 season, determine the total number of pit stops made. Identify any patterns or trends in pit stop frequency.

These two bar charts illustrate the number of pit stops for each driver and team for merged 2024 and Singapore events .

From the above chart which is **driver-based**, there are a few drivers who stops in 1 or 5 pits. Most drivers had 2 or 3 Pit Stops in each event. Also one can seen that some drivers only have participated at Singapore GP.

The below charts shows Pit-Stops based on **team**. Seven **teams** have participated only in Singapore GP with 2 and 3 Pit-Stops during the race. Among the other teams, a few had 1 Pit-Stops and the most frequent number is 2 and 3. 4 Pit-Stops is less frequent.

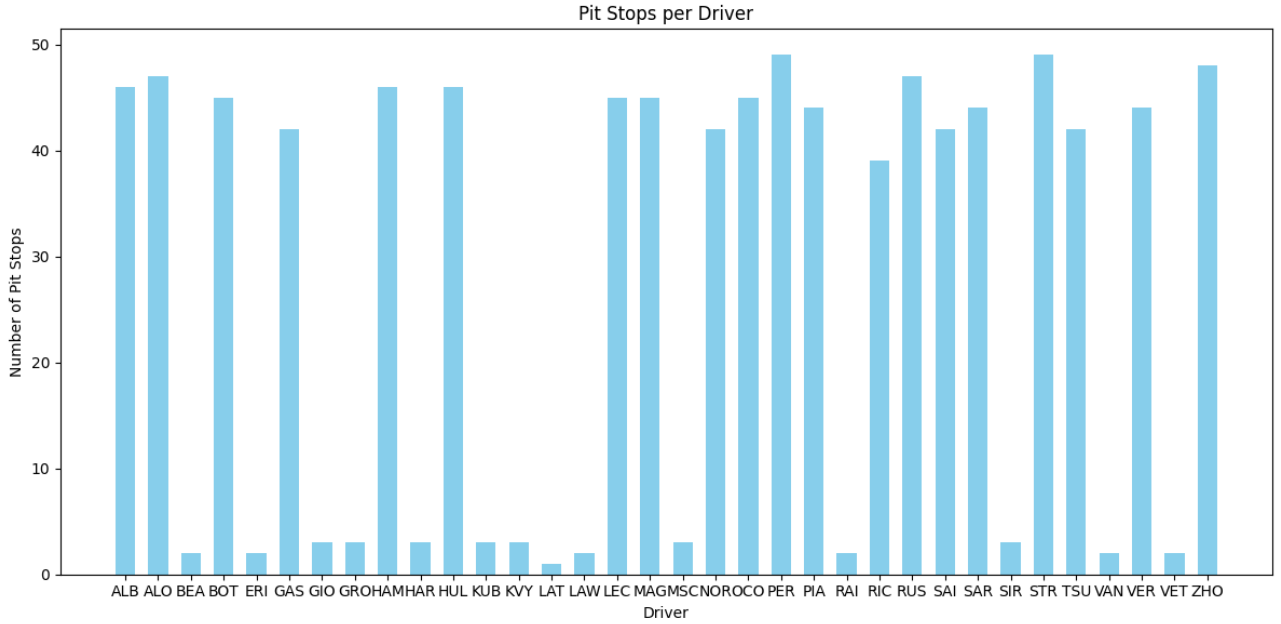


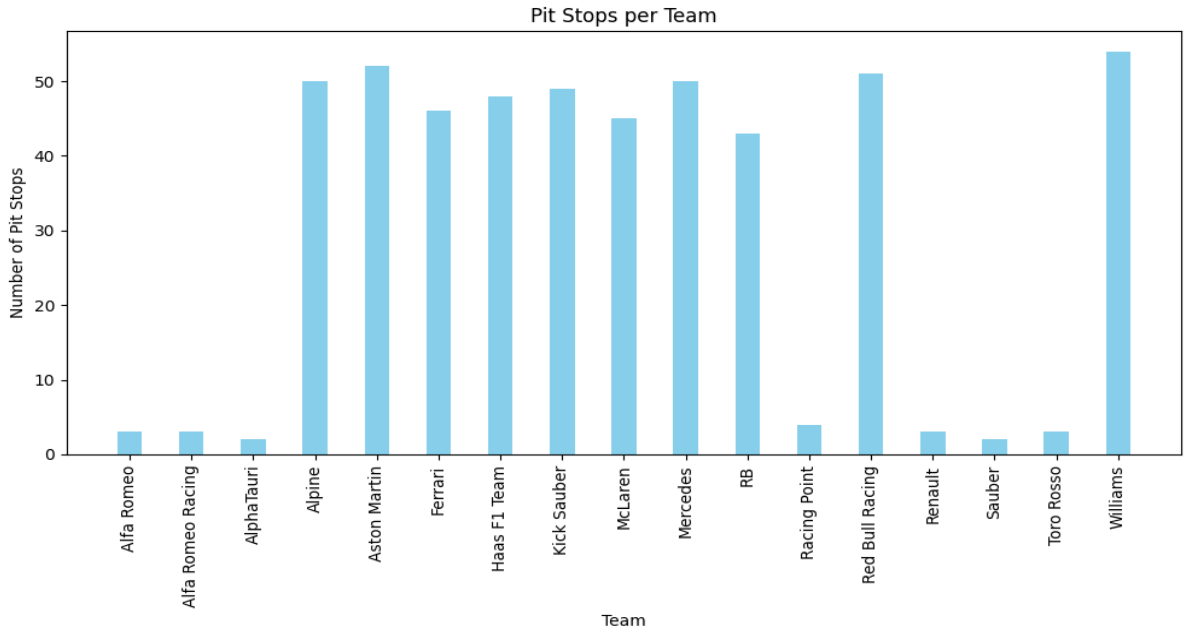


EDA > Merged Data > Total Number of Pit Stops

For each driver and team in each race of the 2024 F1 season, determine the total number of pit stops made.

These charts illustrate total number of Pit-Stops per driver and team over all **Merged** events. Since some drivers only attended in Singapore GP the total number of Pit-Stops for them is less than 10 which actually shows their Pitstops only in Singapore GP. It also implies for the teams. Seven teams stopped in pits less than 5 times and only attended in Singapore GP as we saw in the previous chart.
The other drivers and teams total Pit-Stops is between 40 to 50.

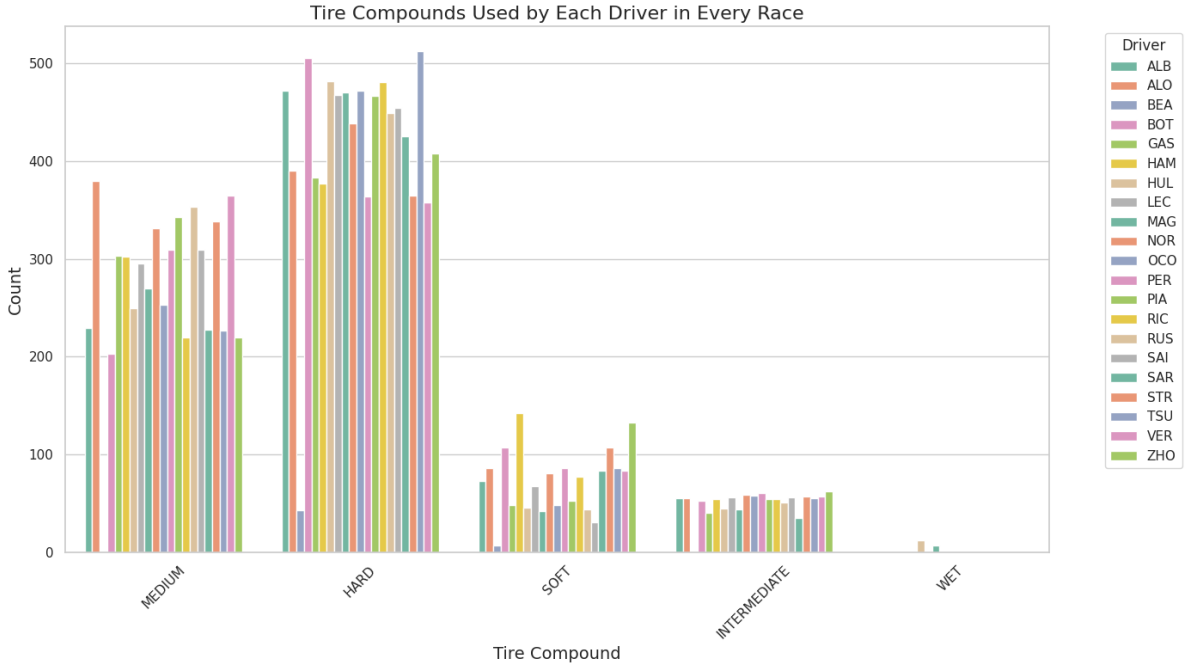


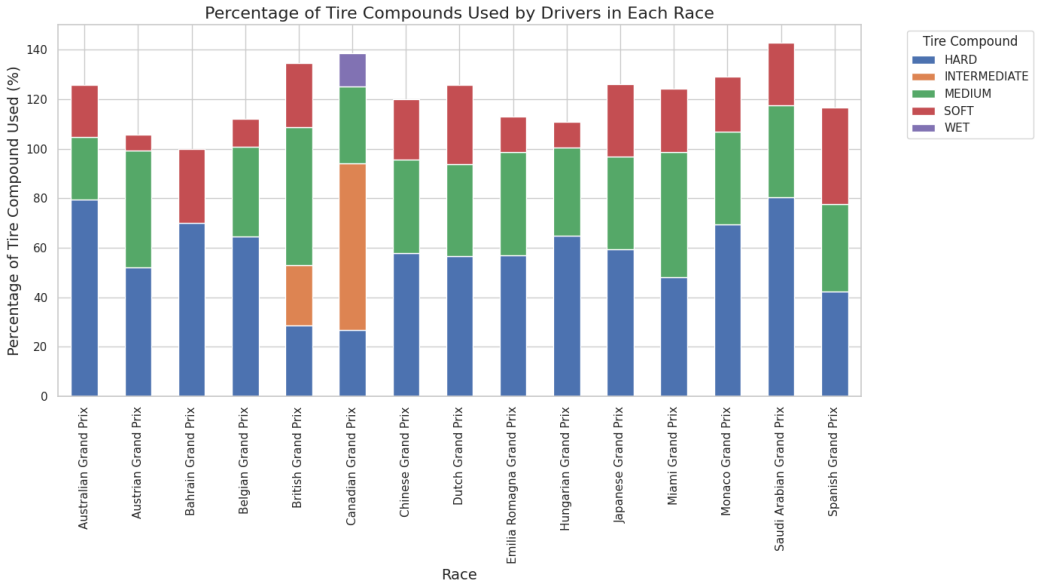


EDA Races 2024 Tire Compounds Used During the Race

Identify the tire compounds used by each driver in every race. Create visualizations to compare the tire compound strategies across different teams and drivers..

This chart illustrate total number of particular compound which has been used by each driver in **2024** events. (This data is obtained by 1- grouping the data on driver and Stints to gather used compound per stint by each driver , and then 2-another group-by on drivers and sum up all same compound).
Hard compound is the most used compounds among all drivers it has been used 400 times in average. The second dominant compound is Medium which is used 250 times by each driver in average. Soft and intermediate are less used about 50 times in average. And wet compound used by only 2 drivers which shows it rarely is used and weather wasn't rainy most of the time



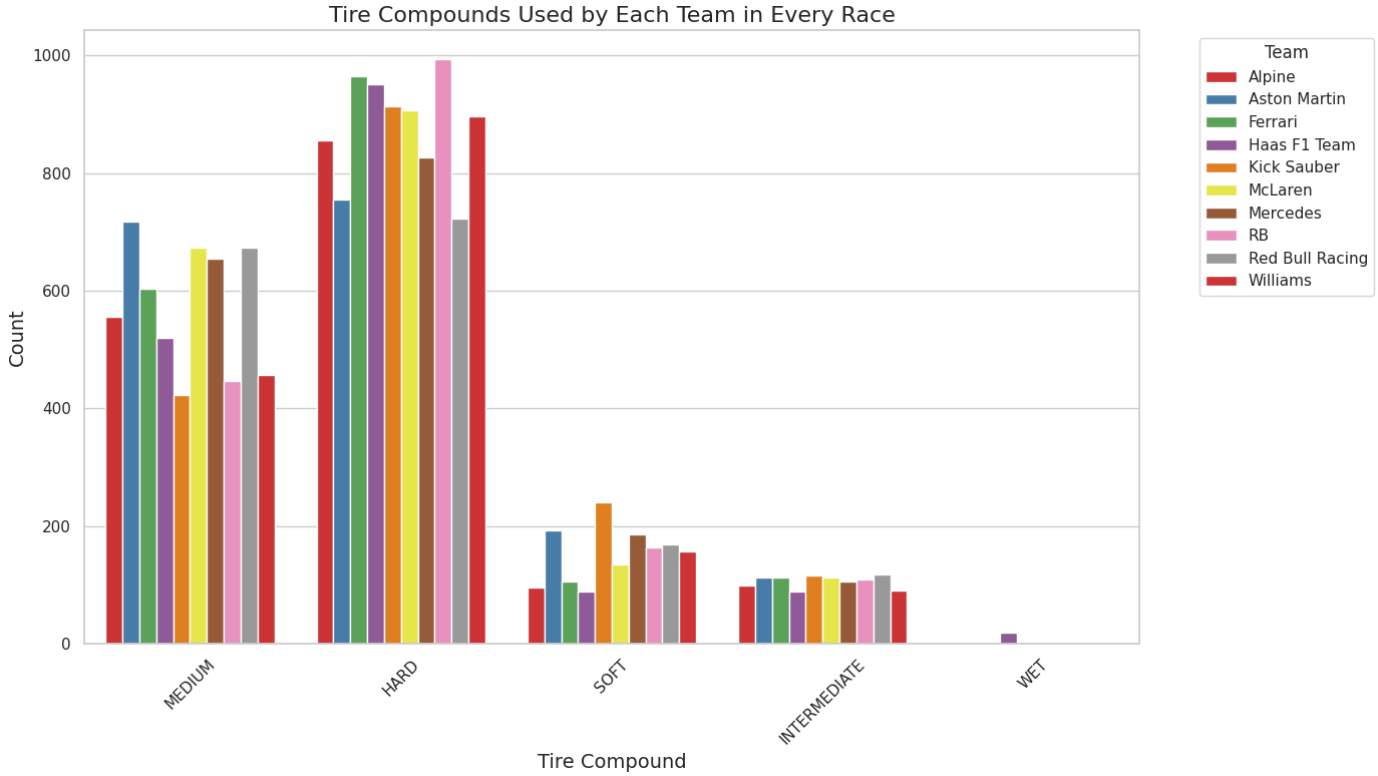


This left chart shows the percentage of each tire compound used by drivers in each race. The blue and green parts of each bar shows hard and medium compounds which are dominant compounds in all races.
In only two races (Britich and Canadian) intermediate compound have been noticeably used. These is because of the weather condition in England and Canada which is mostly rainy with too humidity. Soft compound is used in all races but less hard and medium. Wet is only used in Canadian GP that indicates weather was rainy during race time

EDA Races 2024 Tire Compounds Used During the Race

Identify the tire compounds used by each driver in every race. Create visualizations to compare the tire compound strategies across different teams and drivers..

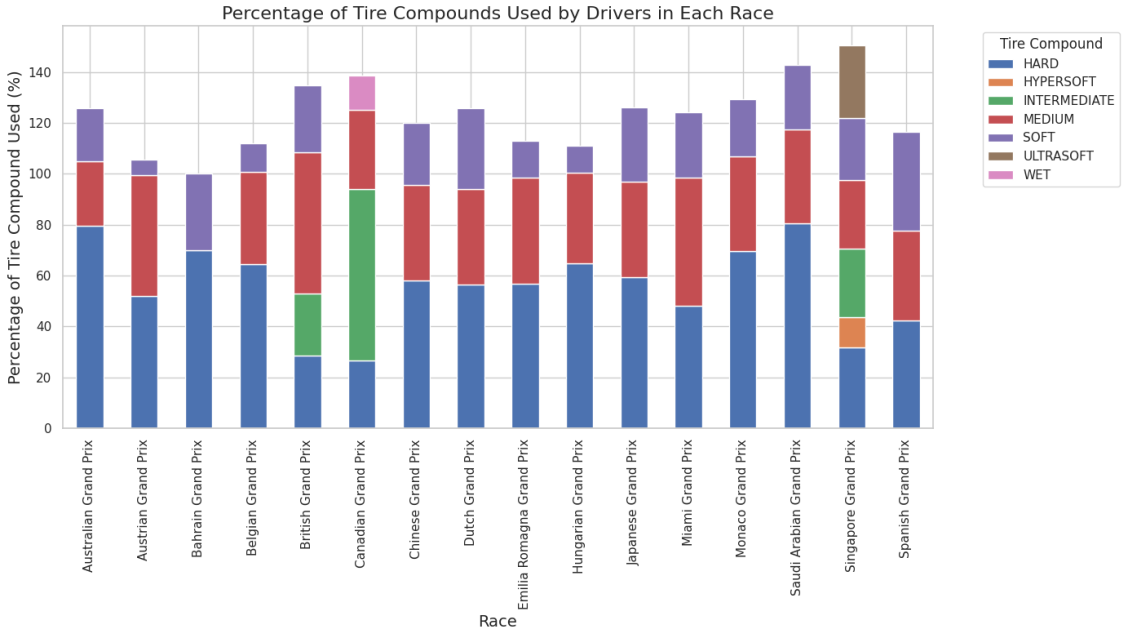
This chart illustrate total number of particular compound which has been used by each team in **2024** events. (This data is obtained by 1- grouping the data on team and Stints to gather used compound per stint by each team , and then 2-another group-by on drivers and sum up all same compound).
It illustrates that hard is the most frequent used compound by each team more than 800 times and after that medium is in the second place.
Both soft and intermediate are used less than 200 times by teams and wet has been used only once by Haas f1 team.

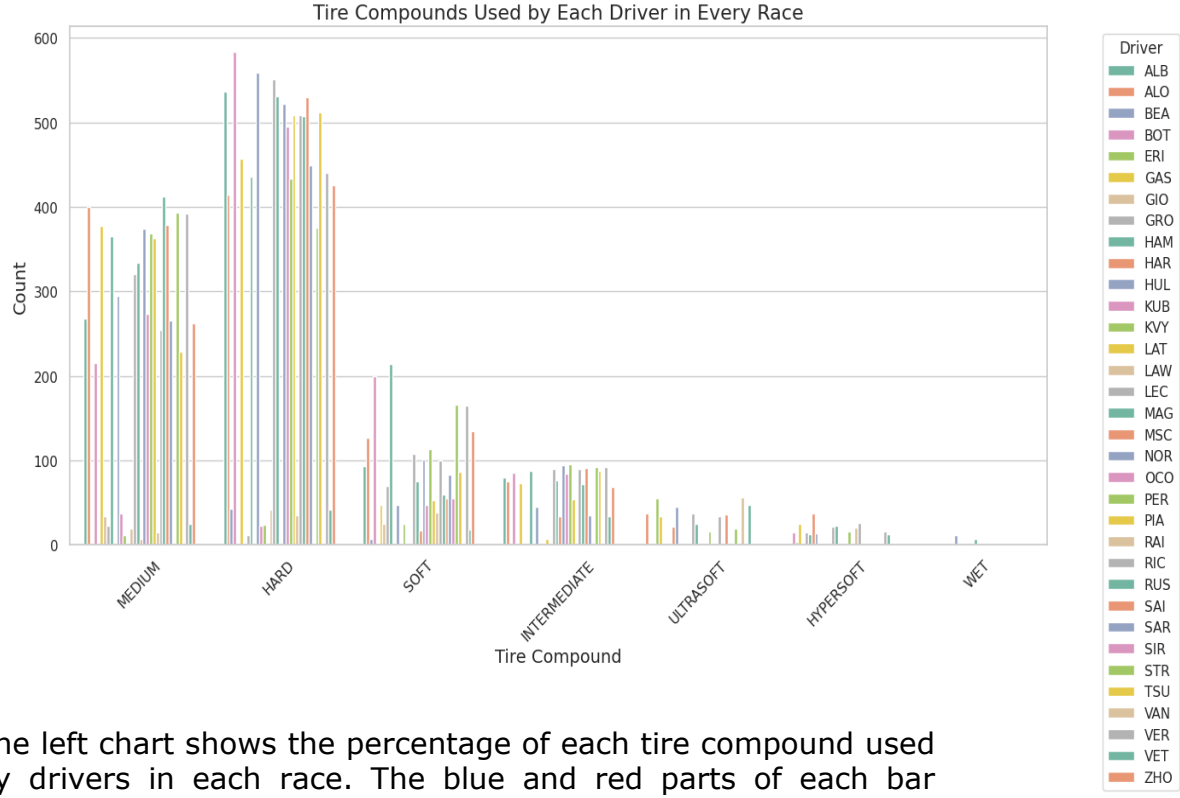


EDA Merged Data Tire Compounds Used During the Race

Identify the tire compounds used by each driver in every race. Create visualizations to compare the tire compound strategies across different teams and drivers..

Right chart illustrate total number of particular compound which has been used by each driver in **merged** events.
Two **ultrasoft** and **hypersoft** compound are added in Sinagpore GP. Like previous, Hard compound is the most used compounds among all drivers it has been used more than 400 times by drivers. The second dominant compound is Medium which is used 300 times by each driver in average. Soft, intermediate , ultra-soft and hyper-soft stands in next places in order. However both ultra-soft and hyper-soft are used in only Singapore GP and their low used number is because of that.



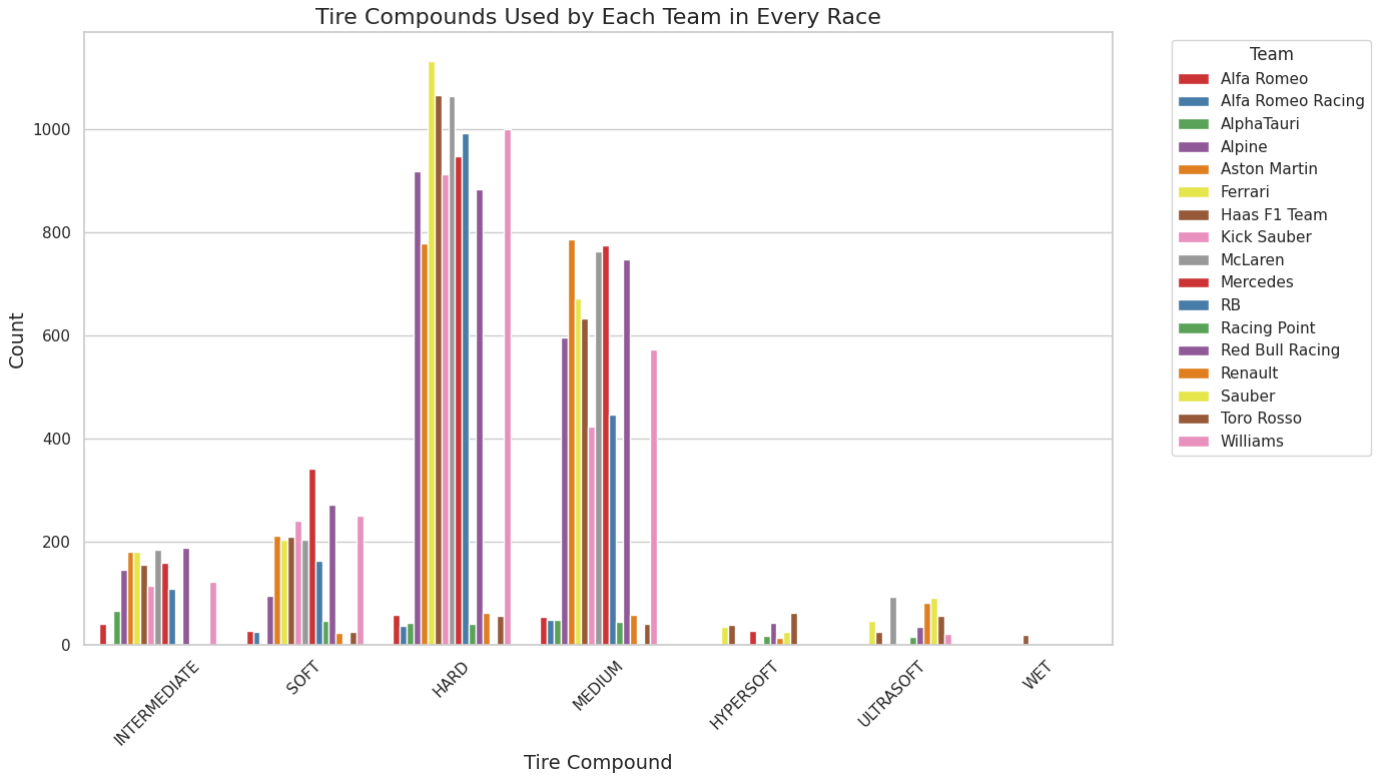


The left chart shows the percentage of each tire compound used by drivers in each race. The blue and red parts of each bar shows hard and medium compounds which are dominant compounds in all races.
The two ultra-soft were used as much as hard one in all Singapore GP which is noticeable. Except of hyper-soft The compounds were relatively equally used in Singapore grand prix.
hyper-soft is the least used compound in Singapore GP.

EDA

Merged Data

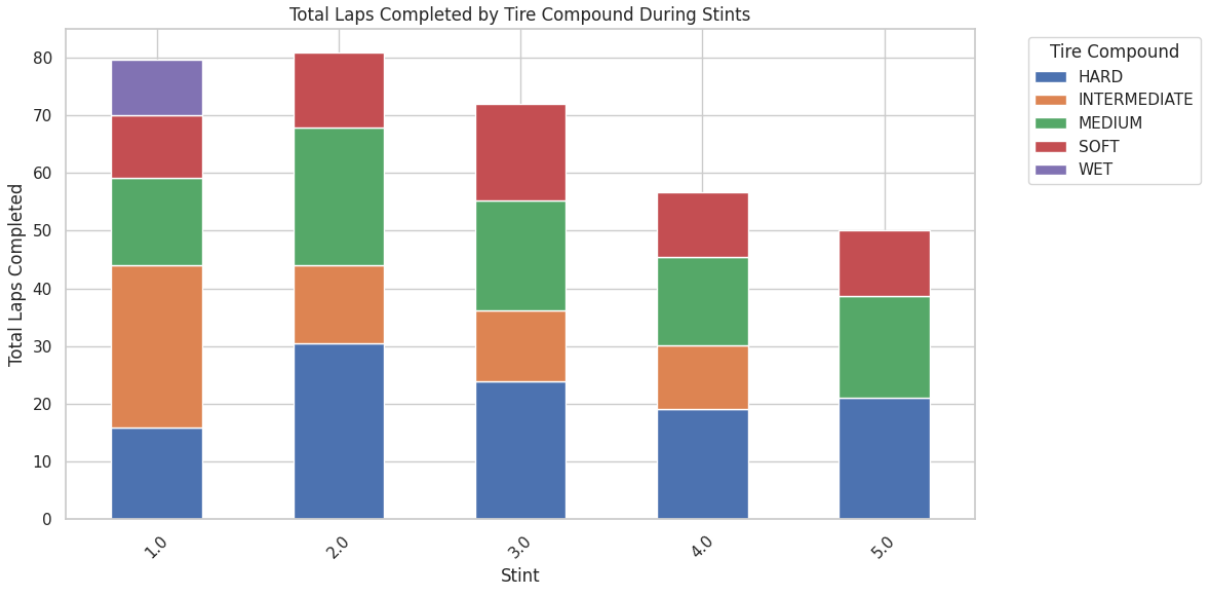
Tire Compounds Used During the Race

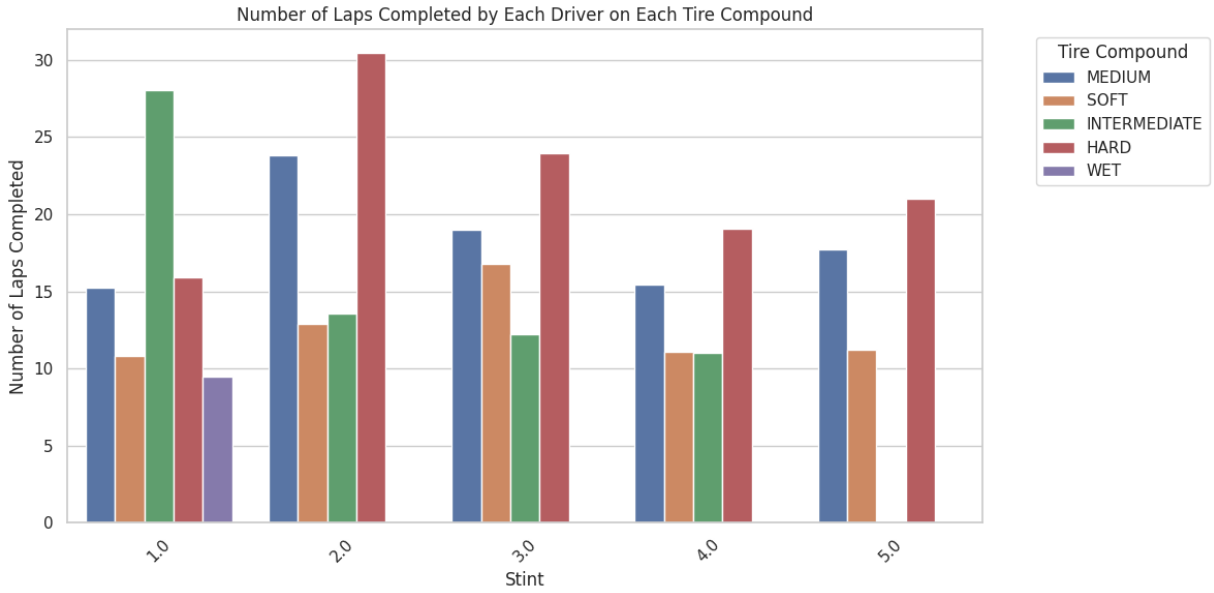


Identify the tire compounds used by each driver in every race. Create visualizations to compare the tire compound strategies across different teams and drivers..

This chart illustrate total number of particular compound which has been used by each team in **merged** events.
The two hyper-soft and ultra-soft compounds were used less than 100 times by 8 and 9 teams respectively.
Hard have were used more than 900 time in average and medium were used more than 600 times

EDA Races 2024 Number of Laps Completed on Each Tire Compound

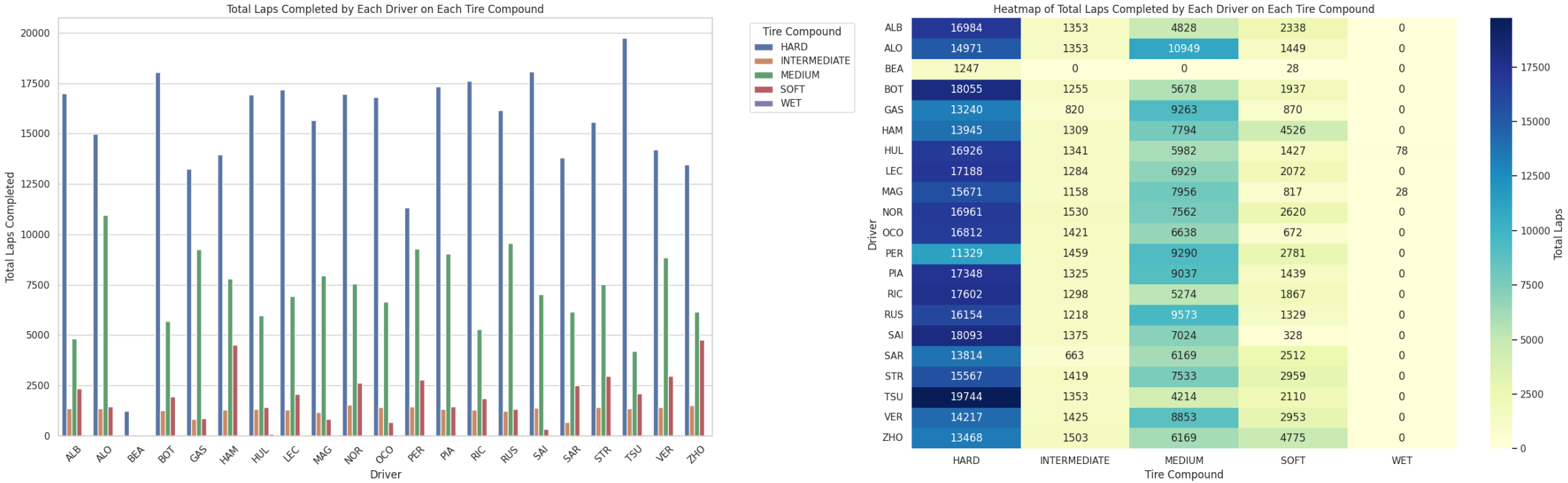




Analyze the number of laps completed by each driver on each tire compound during their stints. Use visuals to illustrate the relationship between stint length and tire compound usage.

Each of the bars in the left figure illustrates the stint number and all compounds that are used on this stint. In first stint intermediate compound could reach to 28 laps which is more than the other compounds. Medium and hard each has reached 15 laps. Soft reached 10 and wet 9 laps. However in the next stints (2 to 5) hard reached the most number of laps and medium stands after it. Only in second stint hard could reach more than 30 laps which is the most number of laps among all stints on one compound

EDA Races 2024 Number of Laps Completed on Each Tire Compound

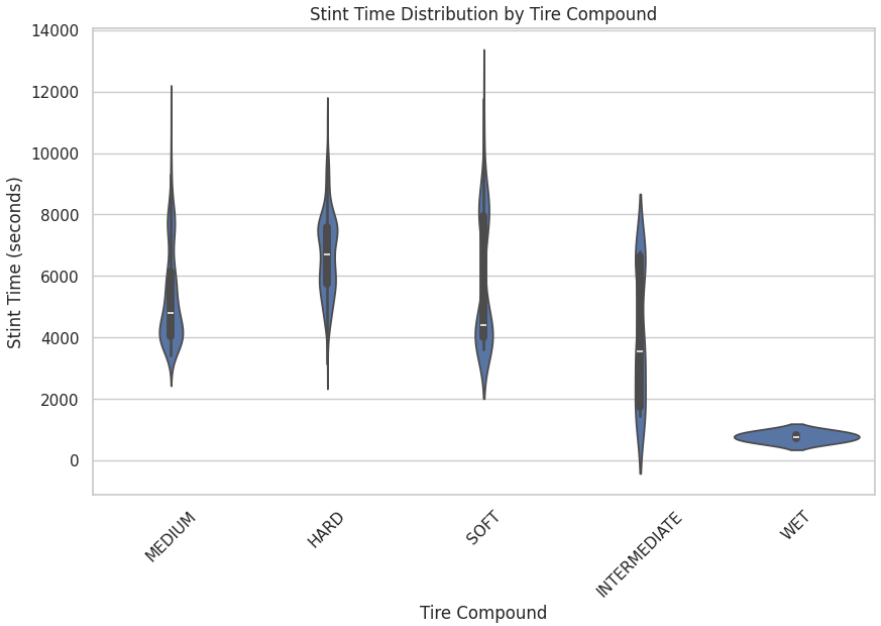


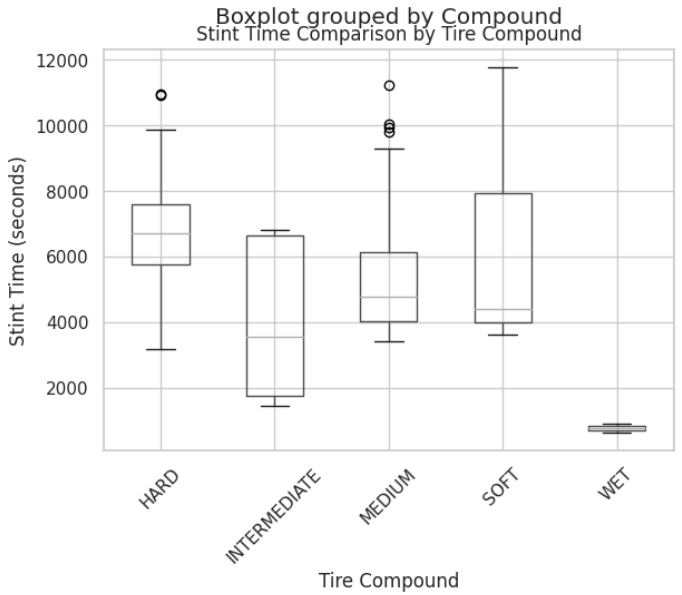
Analyze the number of laps completed by each driver on each tire compound during their stints. Use visuals to illustrate the relationship between stint length and tire compound usage.

the left figure shows the number of completed laps on each compounds per driver in all 2024 races. Most of drivers have used hard for most of the laps and the gap between hard and medium in completed laps is large and noticeable. Nearly 15000 laps are completed by each driver in hard compound. While about 7000 laps are completed by each driver with medium compound as the second most frequent compound.

Right figure shows the exact number of completed laps by each driver on each compound. TSU has completed the most number of laps with 19744 on hard compound. ALO has completed the most number of laps with 10949 on medium compound. HAM has completed the most number of laps with 4526 on soft compound.

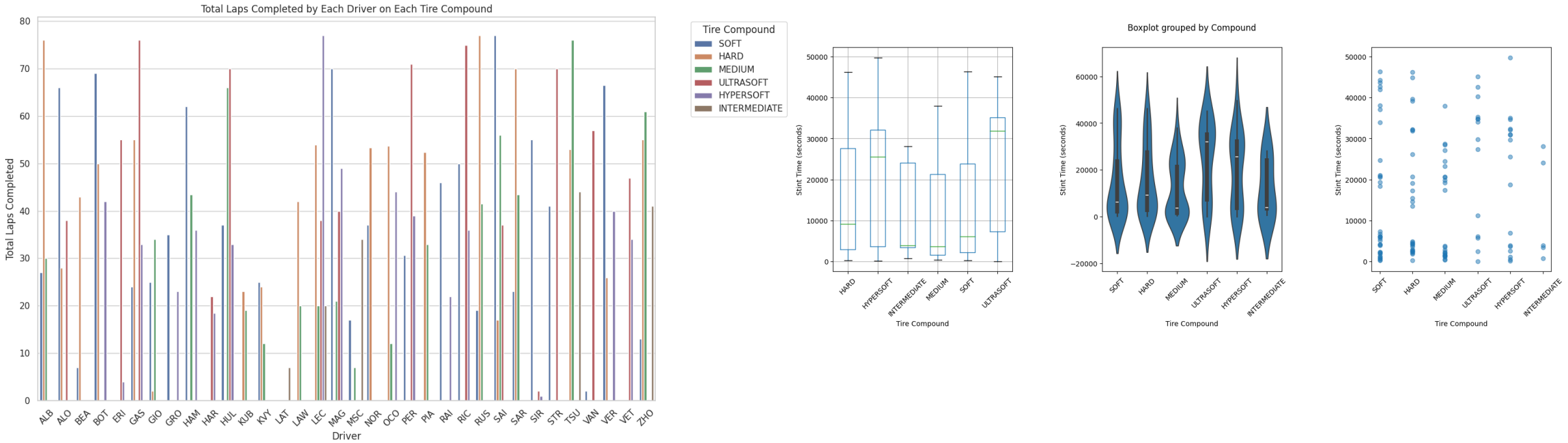
EDA Races 2024 Number of Laps Completed on Each Tire Compound





Use visuals to illustrate the relationship between stint length and tire compound usage.
Based on the above box-plot which illustrate stint time on each tire compounds, in average 2000 seconds are spent on hard compound, 4000 seconds are spent on intermediate compound, 2500 seconds are spent on medium compound and 2500 seconds are spent on soft compound

EDA Merged Data Number of Laps Completed on Each Tire Compound



Analyze the number of laps completed by each driver on each tire compound during their stints. Use visuals to illustrate the relationship between stint length and tire compound usage.

the left figure shows the number of completed laps on each compounds per driver in **merged** races. Despite the 2024 races in merged data drivers used various compound as their dominant compound to complete laps. Hard, hyper-soft, ultra-soft , soft and medium are used by drivers which each driver have reach his most lap number by these compounds.

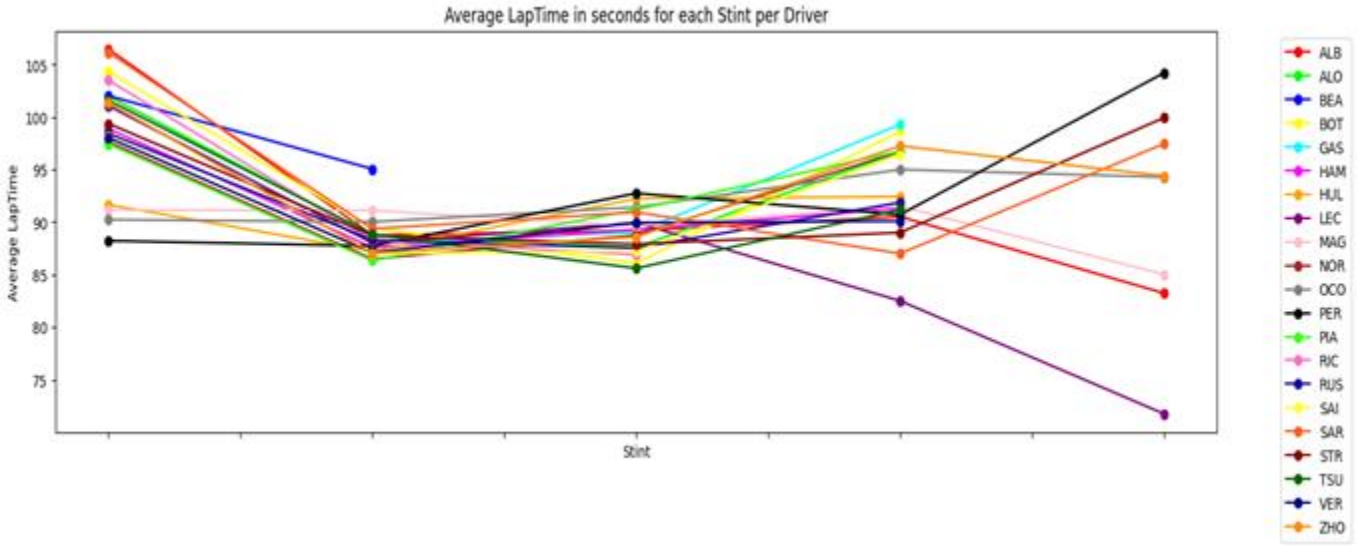
Right figure illustrate stint time on each tire compounds, in average more than 20,000 seconds are spent on hard compound, around 30,000 seconds are spent on hyper-soft compound, nearly 20000 seconds are spent on both intermediate and medium and soft compound, less than 30000 seconds are spent on ultrasoft compound.

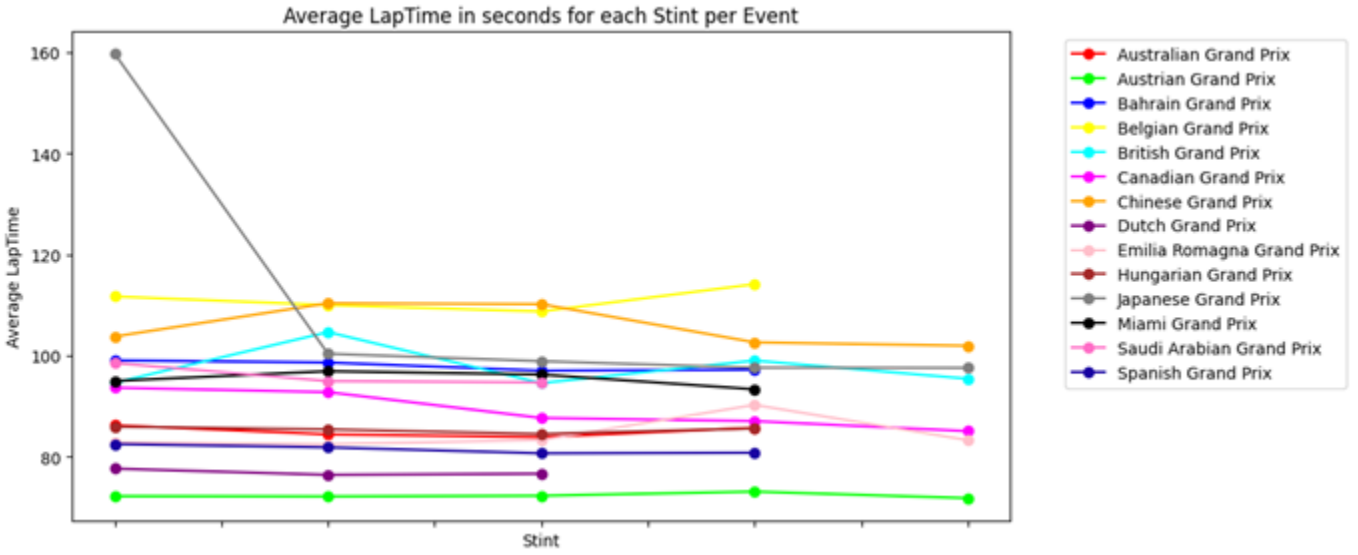
EDA Races 2024 Average Lap Time per Stint and Delta Time per Tire Compound

Calculate the average lap time for each stint. Visualize the changes in lap times within each stint to understand how tire performance evolved over the course of a race.

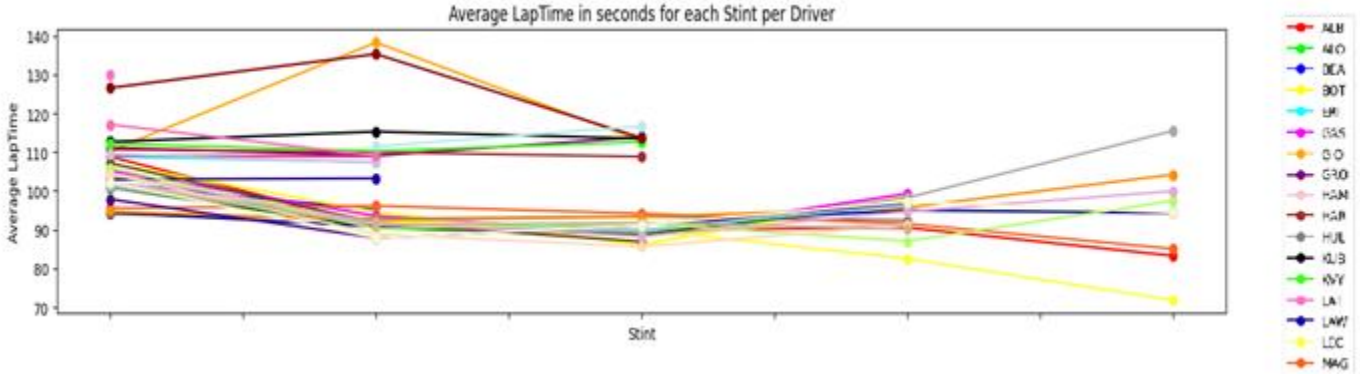
Top figure illustrate average lap time in each stint per driver where bottom figure shows per 2024 event. Totally most drivers have spent much time in first stint and one can say that second and third Stints drives have spent the same time on that. However in the last two stints(4 and 5) drivers have choses different strategies. in the last Stint specially , some drivers decided to go as fast as they can while some of them reduced their speed.

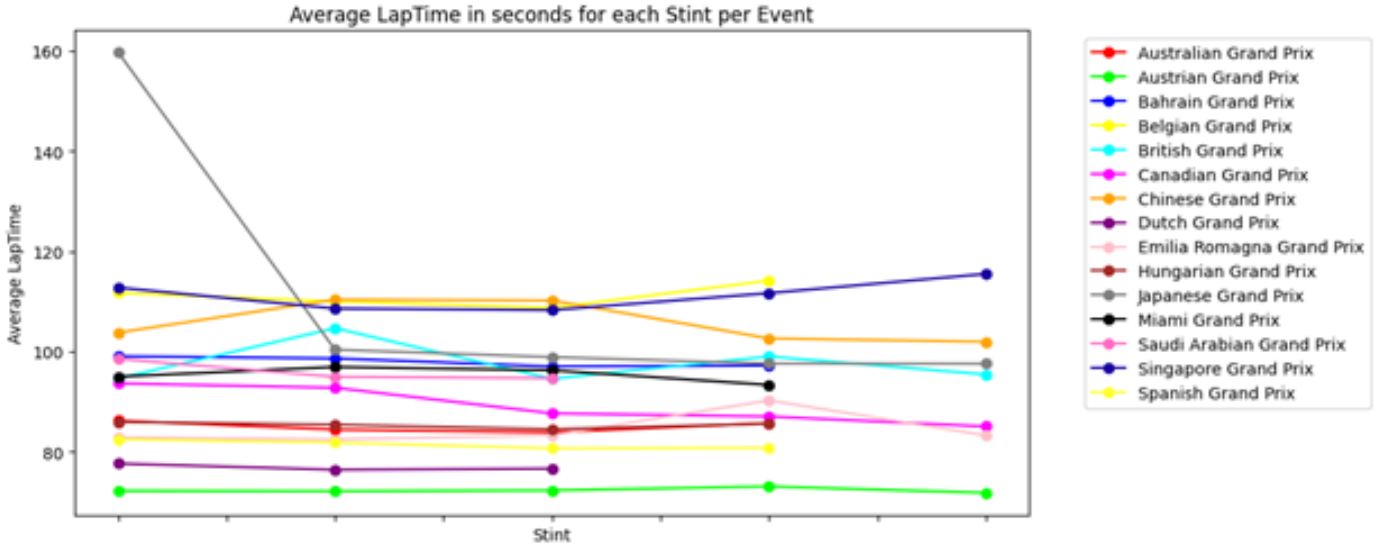
The bottom figure shows average lap time in each stint per event. One can say that generally in each events time in different stints are almost similar without any sharp changes. However in Japanese event first drivers spent too much time in first stint (maybe a car accident or unpredictable event happend) but after that they have drove their car much faster in a unform trend





EDA > Merged Data > Average Lap Time per Stint and Delta Time per Tire Compound





Calculate the average lap time for each stint. Visualize the changes in lap times within each stint to understand how tire performance evolved over the course of a race.

Top figure illustrate average lap time in each stint per driver where bottom figure shows per event.

Drivers starts with different strategies in the first stint of race with various speed. They spent various time in first stint. Unlike 2024 data which drivers spent similar time in second stint, this plot remarks that drivers spent various time in second Stint. In third and fourth stint they drove more in similar speeds. However in the last stint they have chosen different strategies and the time they spent in last stint varies from 70 seconds to 110 seconds in laps.

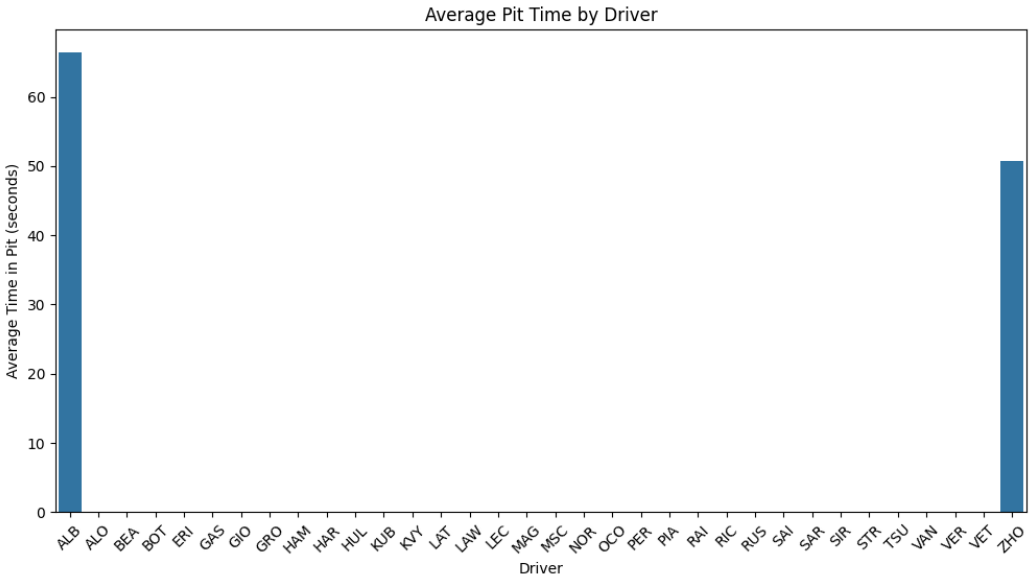
The bottom figure shows average lap time in each stint per event. One can say that generally in each events time in different stints are almost similar without any sharp changes. However in Japanese event first drivers spent too much time in first stint (maybe a car accident or unpredictable event happend) but after that they have run their car much faster in a unform trend

EDA Races 2024 Time Spent in Pits

Measure the total time spent in the pits by each driver during the races. Create visual representations to show how pit time varies across drivers and teams.

The figure shows the total time each driver spent in the pits in all races which is calculated by (PitsInTimeInSeconds - PitsOutTimeInSeconds). As mentioned earlier, more than 96% of these two columns are null in the dataset and we calculate pit-time only by the remaining 4% records. Which couldn't be valid data since when we filtered the data on 'IsAccurate' column to calculate only on these ones there remains nothing from these two columns.

Anyway , the only results achieved by the 4% of data are for ALB and ZOO which shows ALB has spent more than 65 seconds and ZOO has spent near 50 seconds in pits in all events

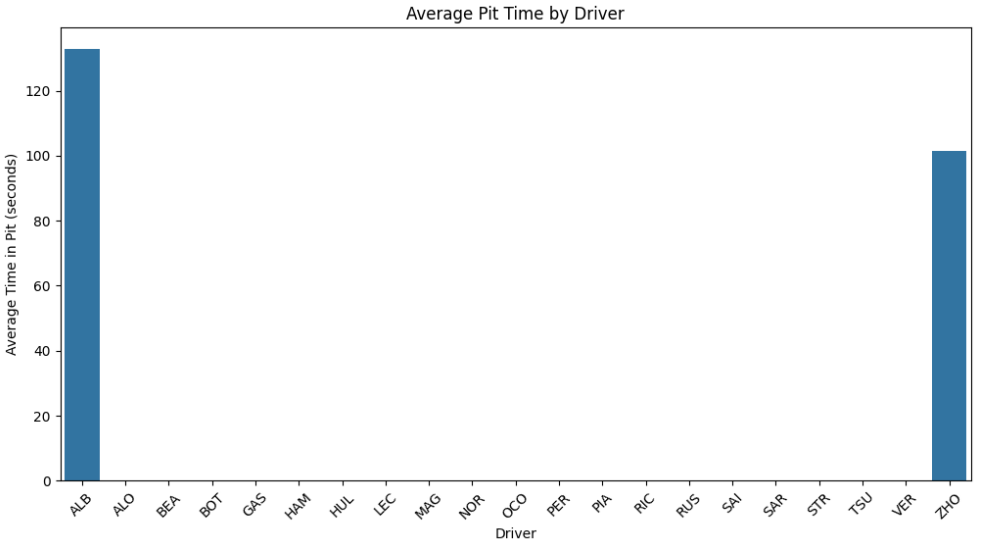


EDA > Merged Data > Time Spent in Pits

Measure the total time spent in the pits by each driver during the races. Create visual representations to show how pit time varies across drivers and teams.

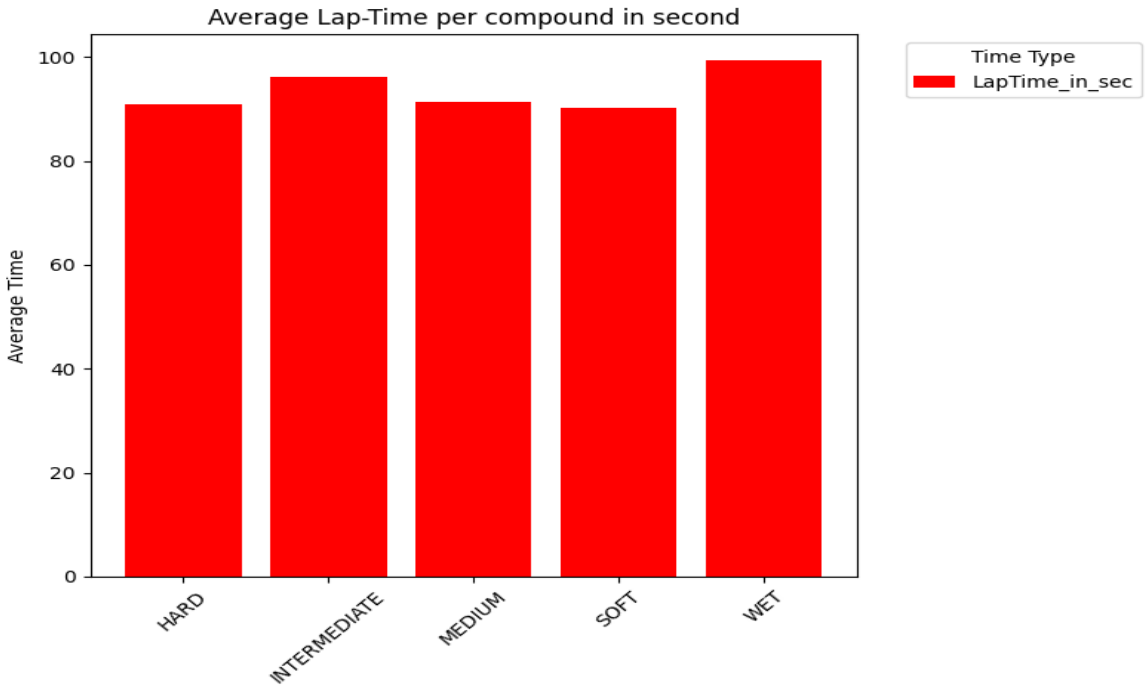
The figure shows the total time each driver spent in the pits in all races which is calculated by (PitsInTimeInSeconds - PitsOutTimeInSeconds). As mentioned earlier, more than 96% of these two columns are null in the dataset and we calculate pit-time only by the remaining 4% records. Which couldn't be valid data since when we filtered the data on 'IsAccurate' column to calculate only on these ones there remains nothing from these two columns.

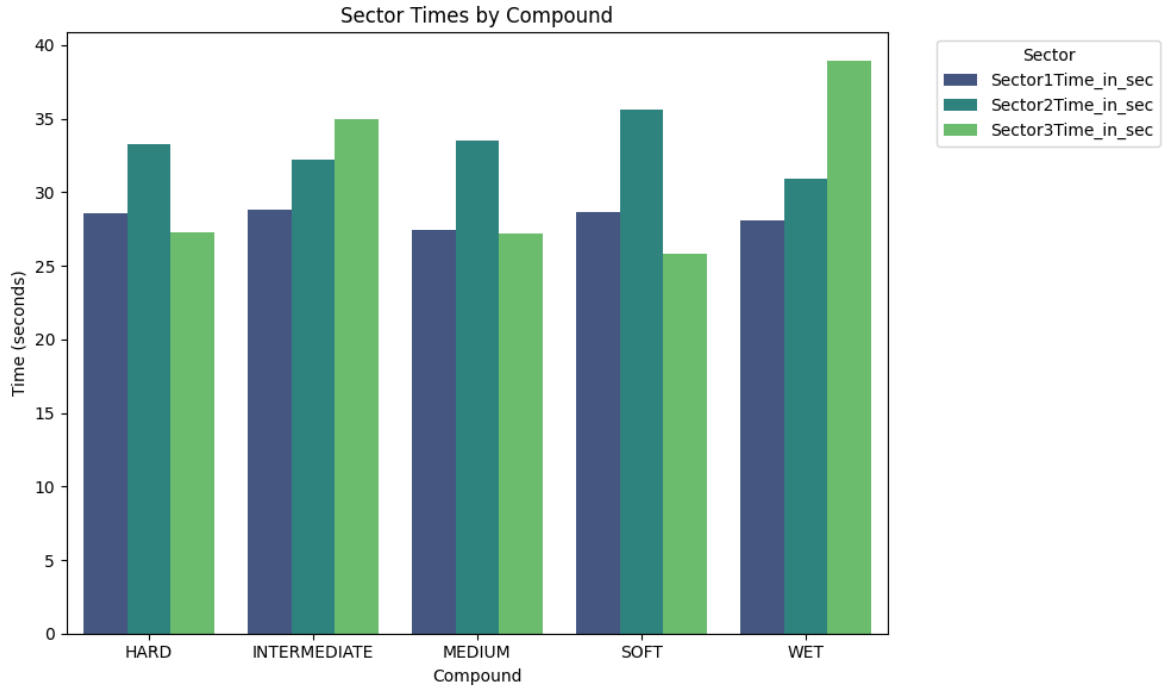
Anyway , the only results achieved by the 4% of data are for ALB and ZOO which shows ALB has spent more than 120 seconds and ZOO has spent near 100 seconds in pits in all events



Relationships

Relationships > Races 2024 > Tire Compound Choice vs Lap Time



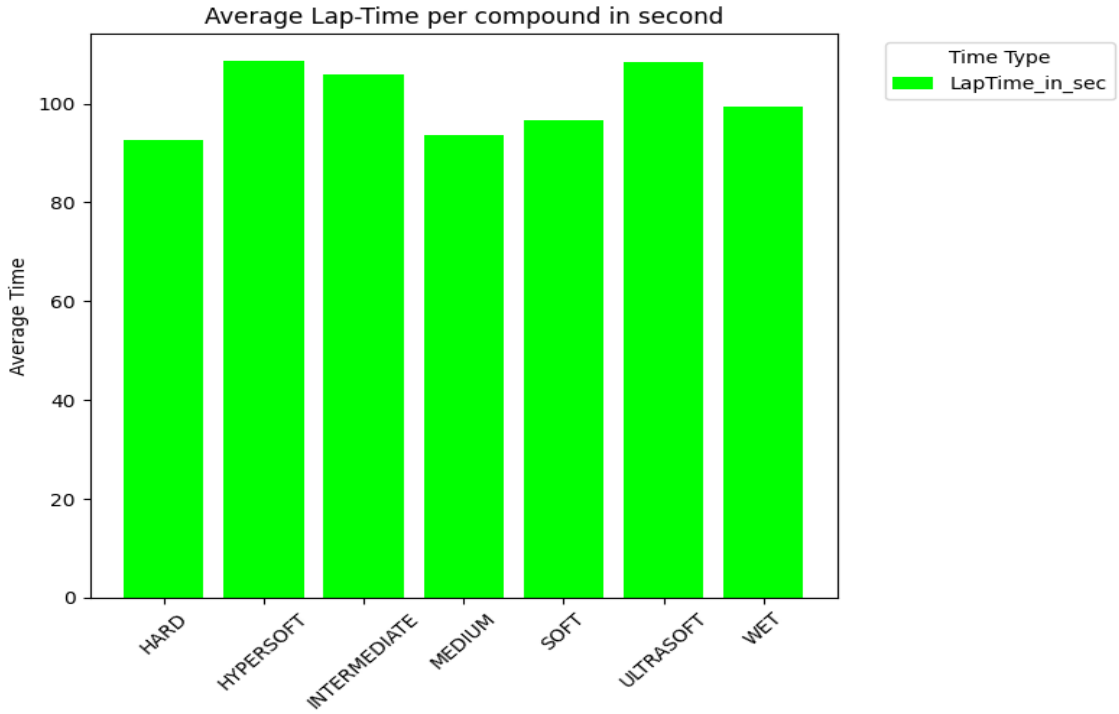


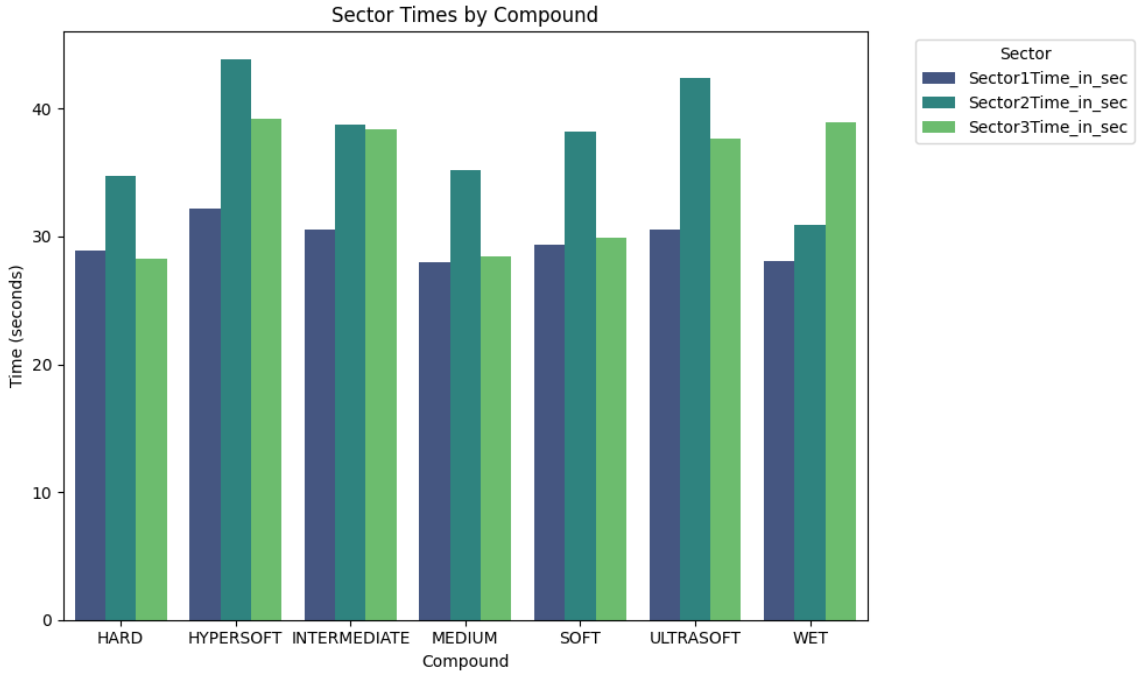
Investigate how different tire compounds (e.g., Soft, Medium, Hard) correlate with lap times throughout the race.
Based on these two image which shows average lap-time spent on each compound one conclude that:

- **HARD**: Times are relatively lower specially in sector 3
- **INTERMEDIATE**: Times are generally higher, particularly in Sector 3.
- **MEDIUM**: Shows a balanced performance, with times lower than hard in 2 and higher than hard, soft and intermediate in sector 3 .
- **SOFT**: Shows a balanced performance, with times lower than others in sector 3 and higher than other in sector 2 .
- **WET**: Has the highest times, indicating slower performance in all sectors.

- HARD and MEDIUM compounds are more consistent in longer stints due to durability but are slower in terms of outright lap time.

Relationships > Merged Data > Tire Compound Choice vs Lap Time



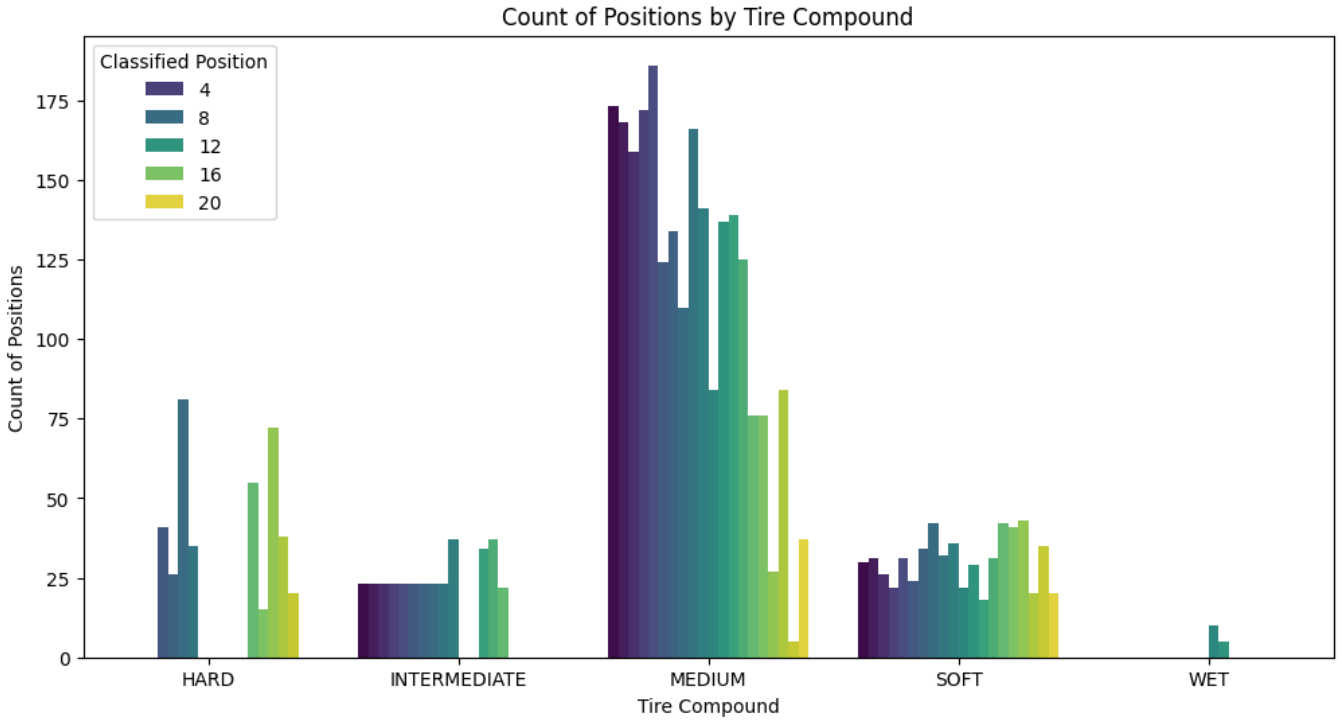


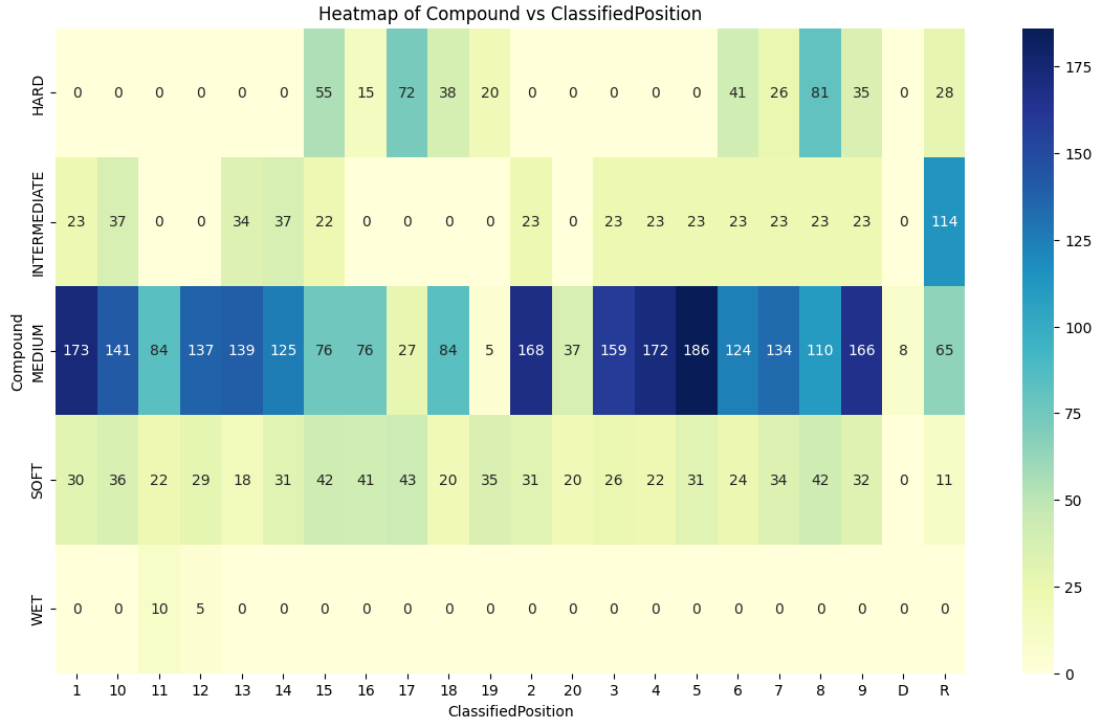
Investigate how different tire compounds (e.g., Soft, Medium, Hard) correlate with lap times throughout the race.
Based on these two image which shows average lap-time spent on each compound one conclude that:

- **HARD**: Times are relatively higher across all sectors compared to softer compounds.
- **HYPERSOFT**: Has the shortest time in Sector 1 but higher times in Sectors 2 and 3.
- **INTERMEDIATE**: Times are generally higher, particularly in Sector 2.
- **MEDIUM**: Shows a balanced performance, with times lower than HARD and higher than SOFT.
- **SOFT**: Generally shows better times than MEDIUM, especially in Sectors 2 and 3.
- **ULTRASOFT**: Performs well, especially in Sector 1, but times increase in Sectors 2 and 3.
- **WET**: Has the highest times, indicating slower performance in all sectors.

In dry conditions, softer compounds such as HYPERSOFT and ULTRASOFT perform better in initial stints but may degrade quickly.
HARD and MEDIUM compounds might be more consistent in longer stints due to durability but are slower in terms of outright lap time.

Relationships > Races 2024 > Starting Tire Type vs Final Classification





Explore how starting the race on different tire compounds (Hard, Soft, Medium) impacts the final classification

The above images shows the number of drivers achieved each position on a particular compound in the begging of the race. Based on the images :

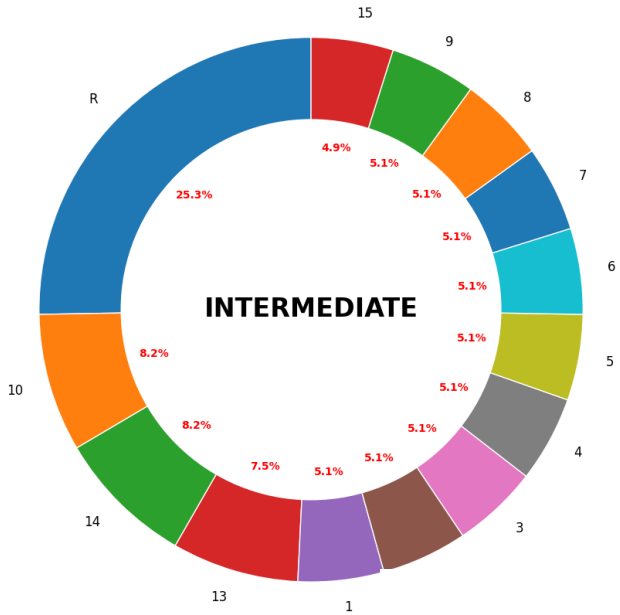
Hard Compound: Not frequently used for top positions and More common in mid to lower positions (15th and above).

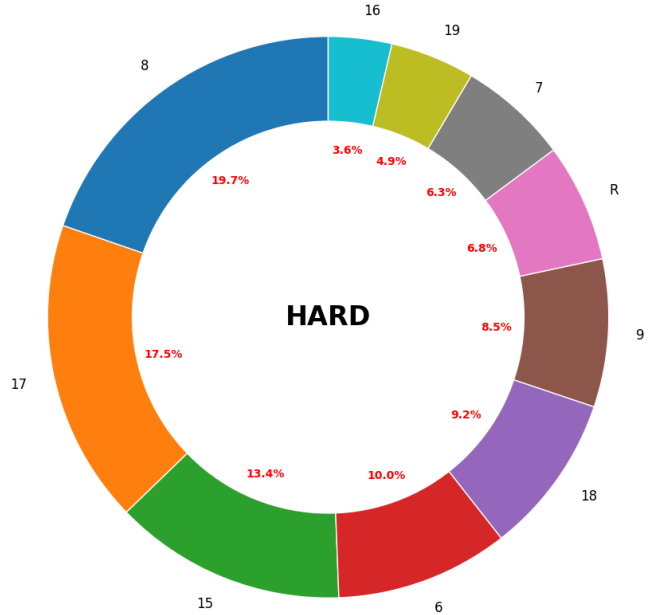
Medium Compound: Most frequently used across all position. Significant number of top ten finishes, especially from 3rd to 10th have used it.

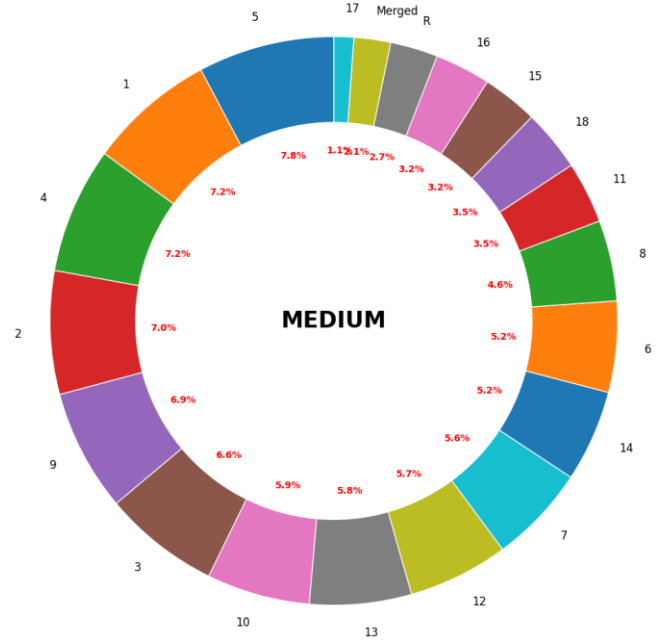
Soft Compound: Also used frequently, but less than Medium and achieves good results in top ten finishes, with some wins and podiums. It Offes competitive results, useful for aggressive strategies, especially when track position is critical.

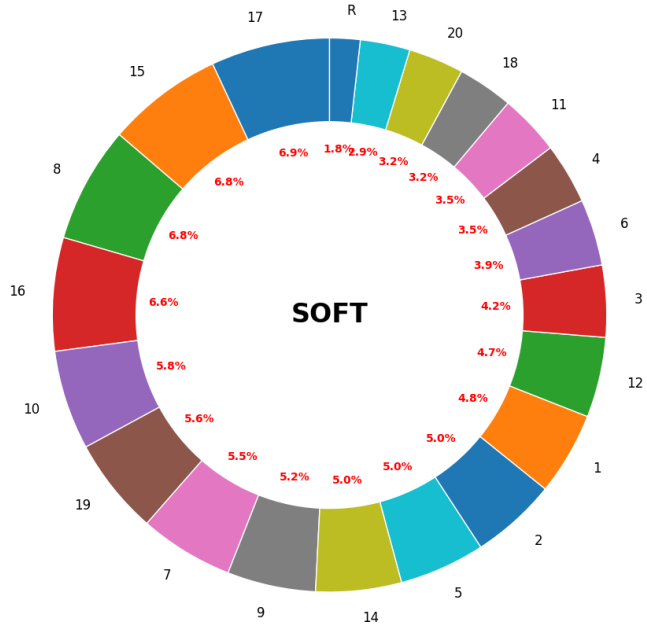
Next slide shows pie chart per compound and the percentage of each final position achieved by drivers

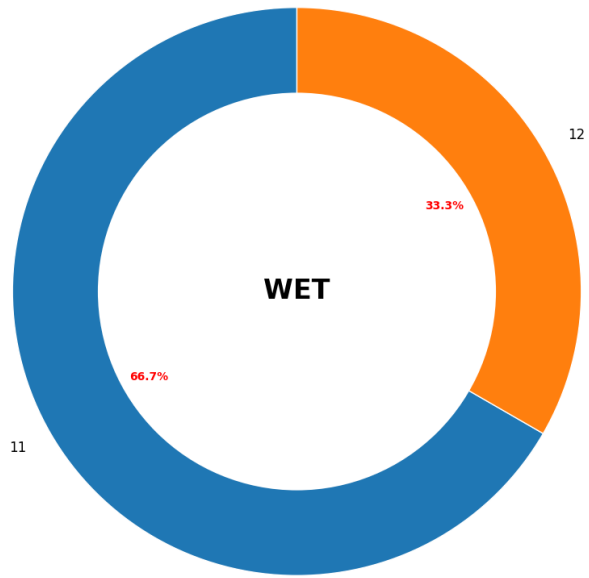
Relationships > Races 2024 > Starting Tire Type vs Final Classification



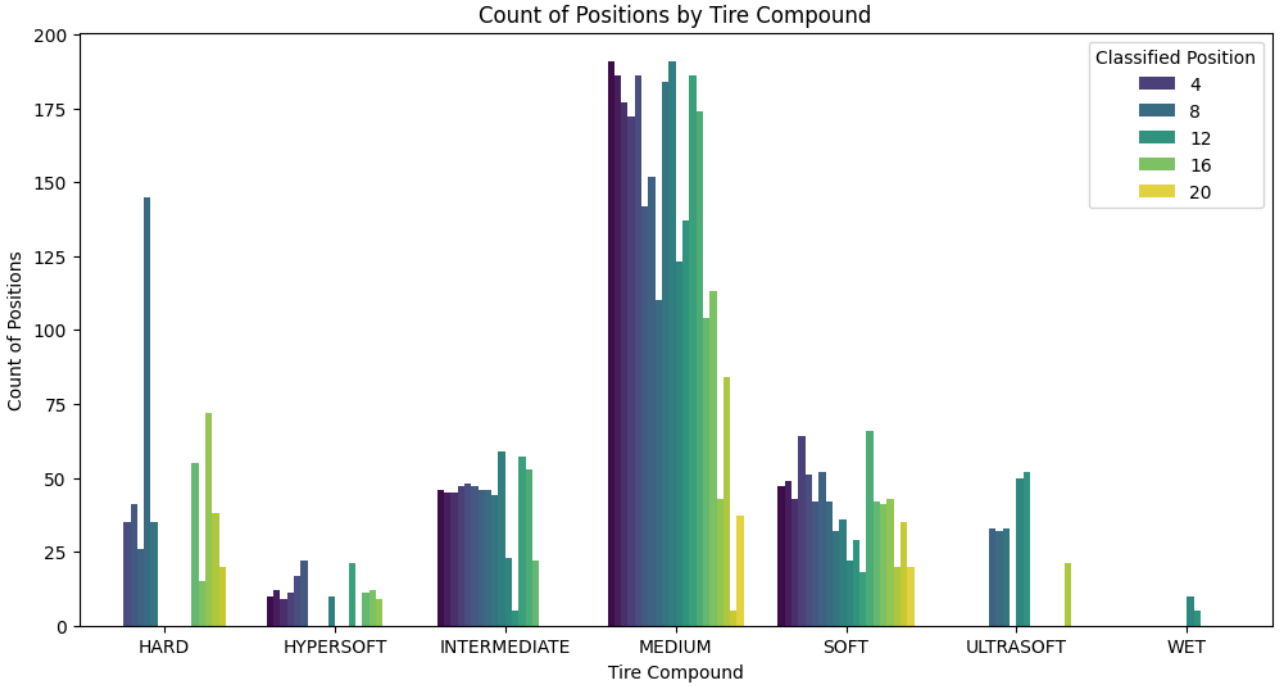


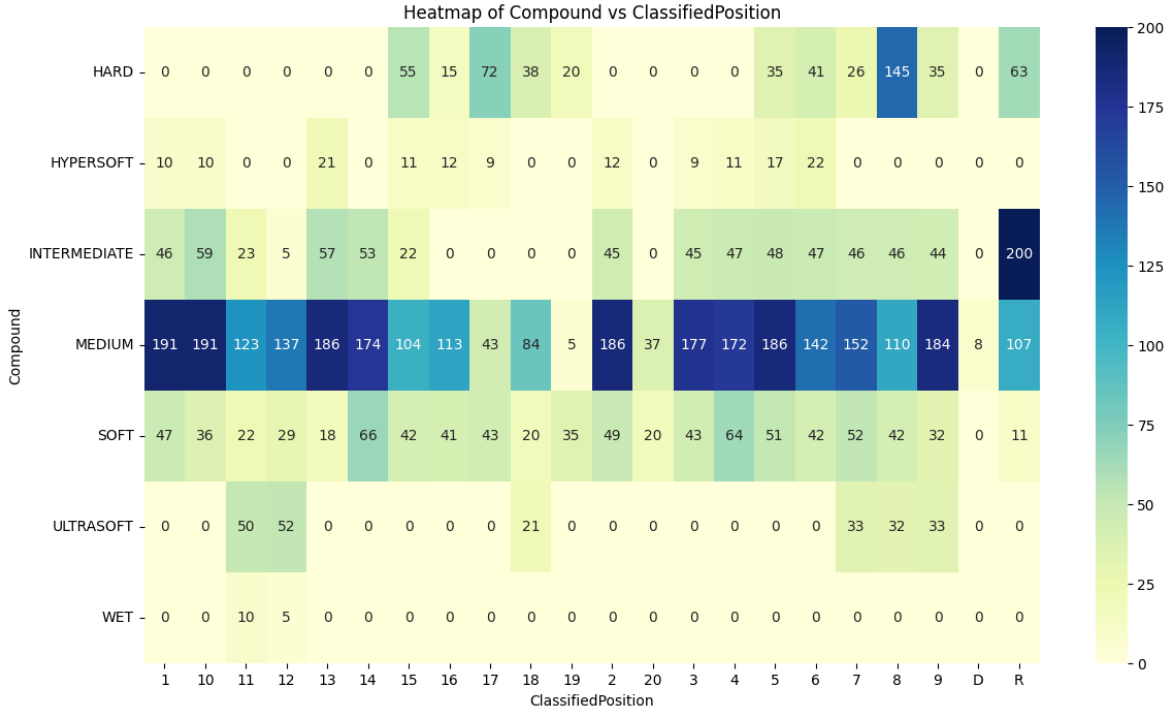






Relationships > Merged Data > Starting Tire Type vs Final Classification





Explore how starting the race on different tire compounds (Hard, Soft, Medium) impacts the final classification

The above images shows the number of drivers achieved each position on a particular compound in the begging of the race. Based on the images :

Medium Compound: Most frequently used across all position. Significant number of top ten finishes, especially from 3rd to 10th have used it.

Soft and intermediate Compound: Also used frequently, but less than Medium and achieves good results in top ten finishes, with some wins and podiums.

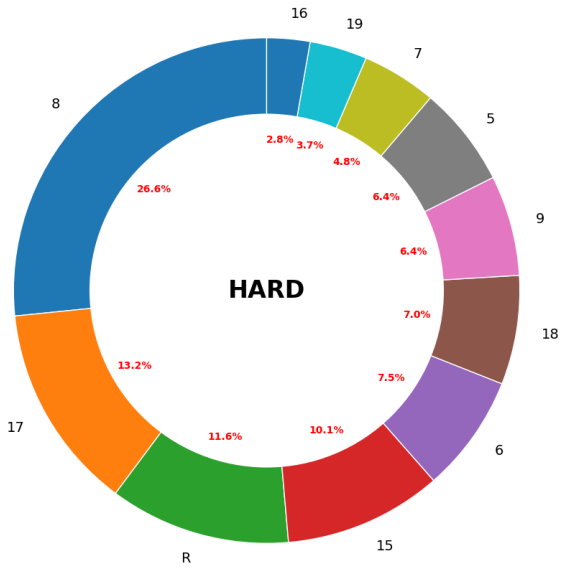
Hard Compound : used by some of top ten finishers and also from 16th to 20th position. This compound are more strategic for endurance and minimizing pit stops.

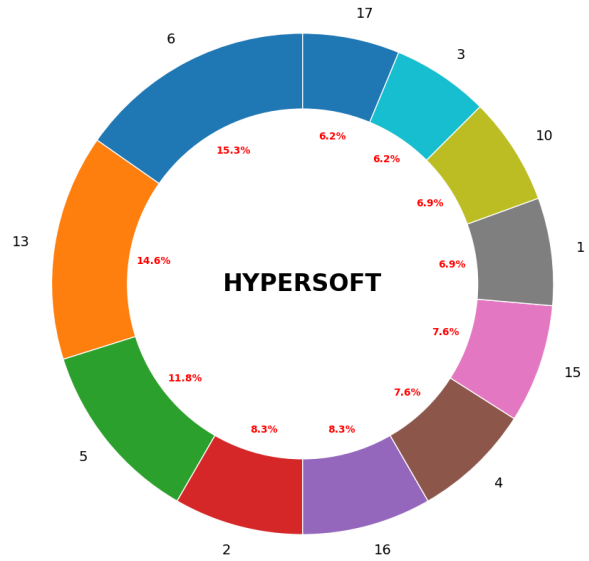
Hyper-soft : mostly used by drivers in Singapore GP and top 10 finishers have used it as their first compound

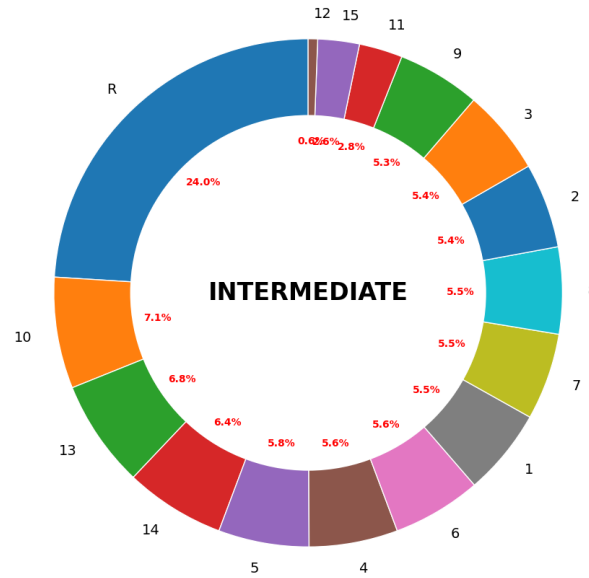
Utra-soft : not used by top drivers and the one who used it achieved 7th to 18th

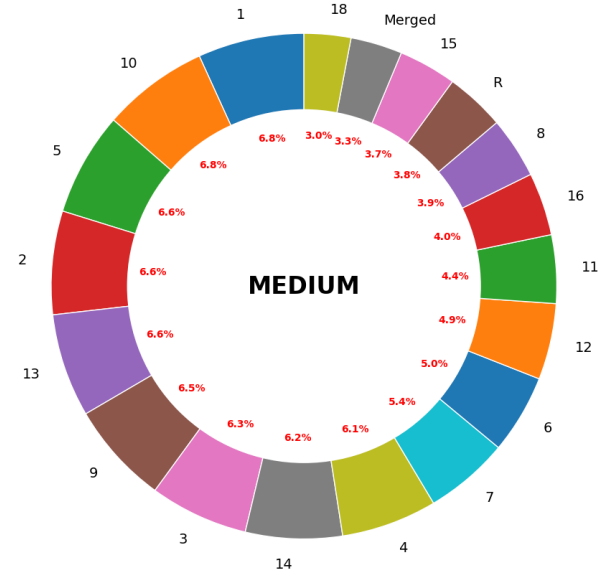
Next slide shows pie chart per compound and the percentage of each final position achieved by drivers

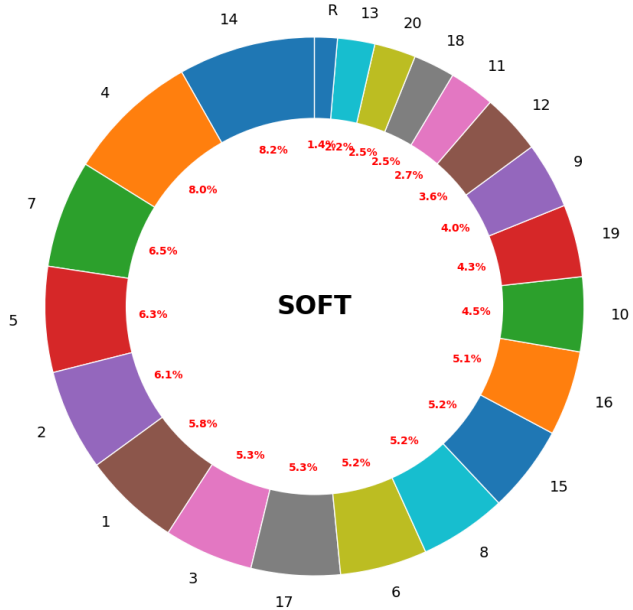
Relationships > Merged Data > Starting Tire Type vs Final Classification

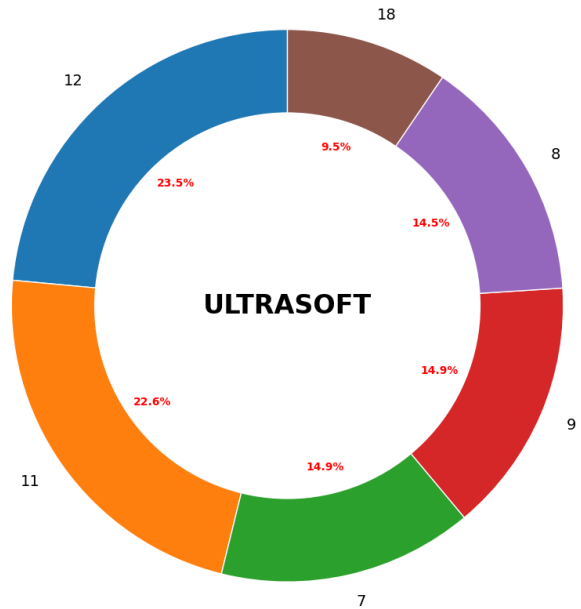


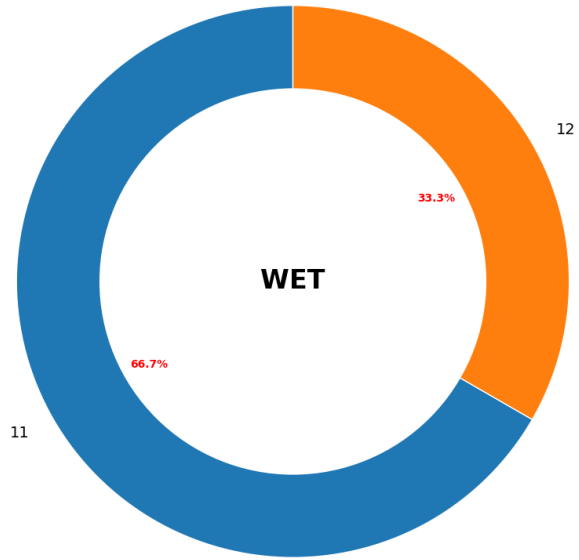












Relationships > Races 2024 > Number of Laps on a Compound vs Delta Time

Task: Explore the relationship between the number of laps driven on a particular tire compound and the delta time (change in lap times) over those laps.

The image shows three bar charts that relate to tire compounds used in racing. green bar shows lap number , yellow bars shows average lap time and red bars shows average time

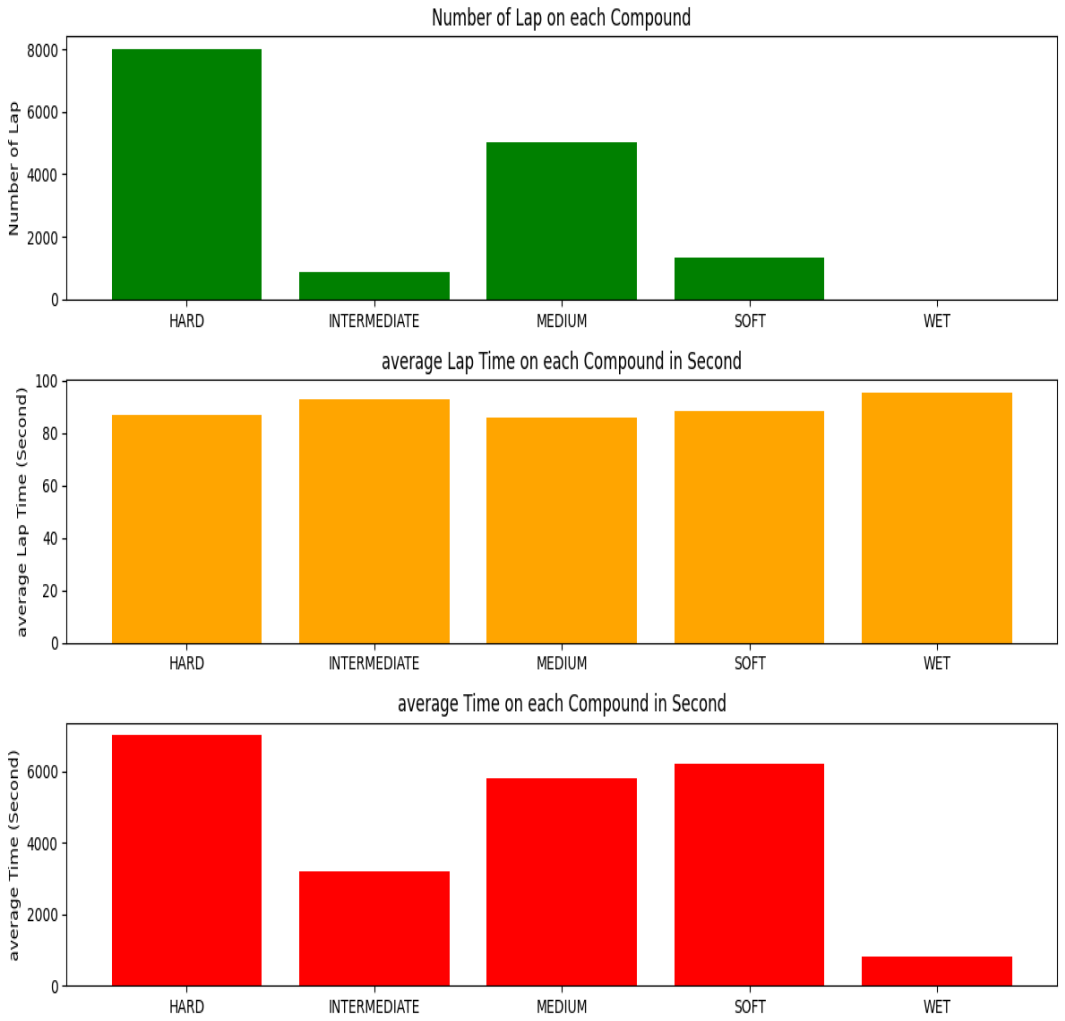
Based on this image hard has the highest number of laps which indicates durability, and medium after that has about 4000 of laps but soft and intermediate have fewer lap number.

But all these compounds have relatively the same lap time (80 – 100 sec)but hard with a bit lower lap-time and wet with highest average lap time that indicates slower pace.

In general HARD tires are used for longer stints (more laps), but not necessarily for the fastest laps. HYPERSOFT tires, while allowing faster individual laps, are used less often.

NTERMEDIATE, MEDIUM, and SOFT tires offer a balance between durability and speed, showing moderate lap times and usage.

ULTRASOFT and WET tires are likely used in specific conditions, leading to fewer laps and specific performance characteristics.



Relationships > Merged Data > Number of Laps on a Compound vs Delta Time

Task: Explore the relationship between the number of laps driven on a particular tire compound and the delta time (change in lap times) over those laps.

The image shows three bar charts that relate to tire compounds used in racing. green bar shows lap number , yellow bars shows average lap time and red bars shows average time

Number of Laps HARD: Highest number of laps, indicating durability.
•MEDIUM: Moderate number of laps.
•INTERMEDIATE and SOFT: Fewer laps than MEDIUM.
•HYPERSOFT, ULTRASOFT, WET: Very few laps, suggesting limited use.

Average Lap Time
•HARD: Moderate average lap time.
•HYPERSOFT: Highest average lap time, indicating slower pace.
•INTERMEDIATE, MEDIUM, SOFT: Lower average lap times compared to HYPERSOFT.

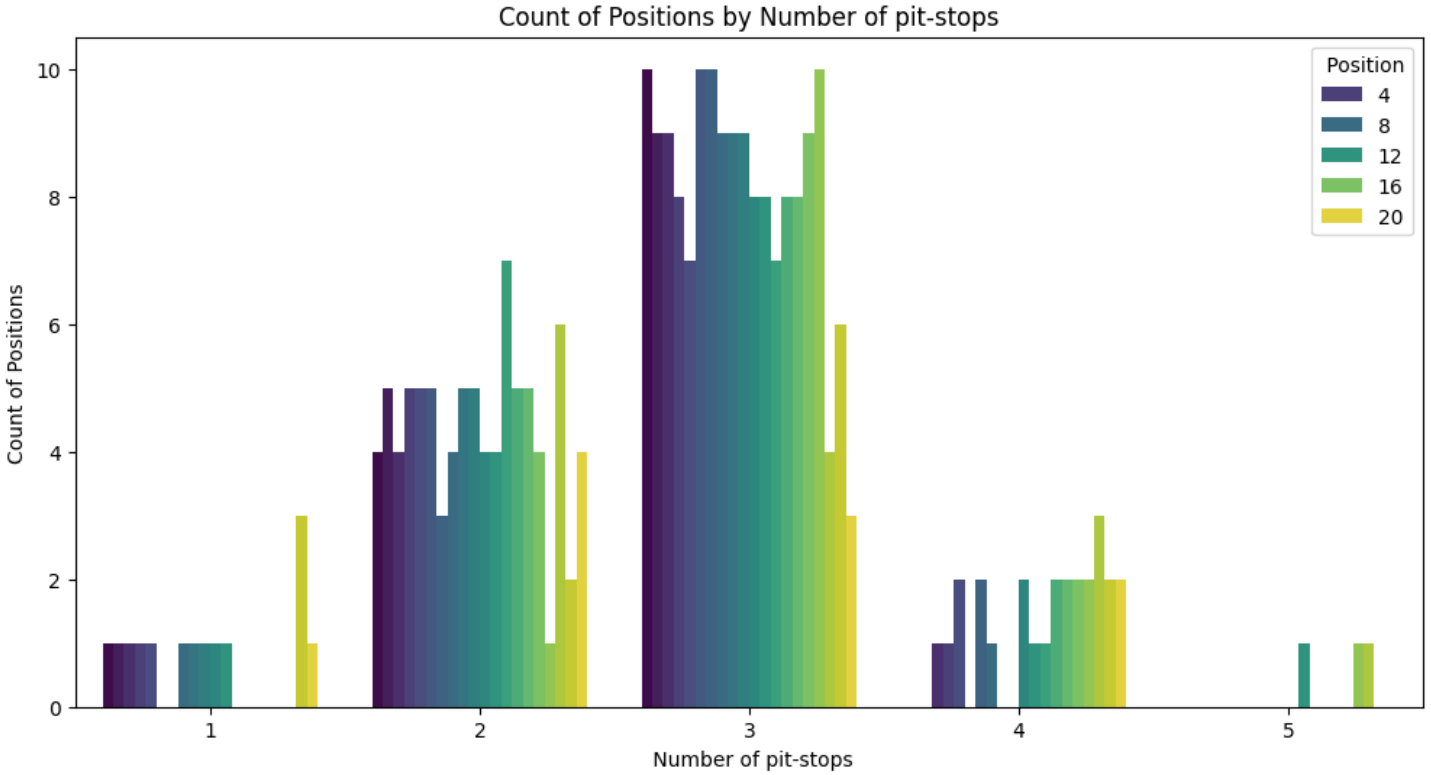
Average Time on Each Compound
HARD: Highest total time, correlating with the highest number of laps.
•HYPERSOFT: Lower total time, reflecting fewer laps.
•INTERMEDIATE, MEDIUM, SOFT: Moderate total times.
•ULTRASOFT, WET: Lower total times, consistent with fewer laps.

In general HARD tires are used for longer stints (more laps), but not necessarily for the fastest laps. HYPERSOFT tires, while allowing faster individual laps, are used less often.
NTERMEDIATE, MEDIUM, and SOFT tires offer a balance between durability and speed, showing moderate lap times and usage.
ULTRASOFT and WET tires are likely used in specific conditions, leading to fewer laps and specific performance characteristics.

Relationships > Races 2024 > Number of Stops vs Final Position

Analyze the relationship between the number of pit stops a driver makes and their final position in the race.

The image illustrates the relationship between the number of pit stops and the final race positions of drivers. based on the image , The 2 and 3 pit stop strategies are the most flexible, allowing for competitive positioning. These two pits stops are widely used by drivers with various positions.
Fewer pit stops can lead to better positions but may not always be feasible due to tire wear or race conditions.
More pit stops might be a response to unexpected events, leading to poorer positions.

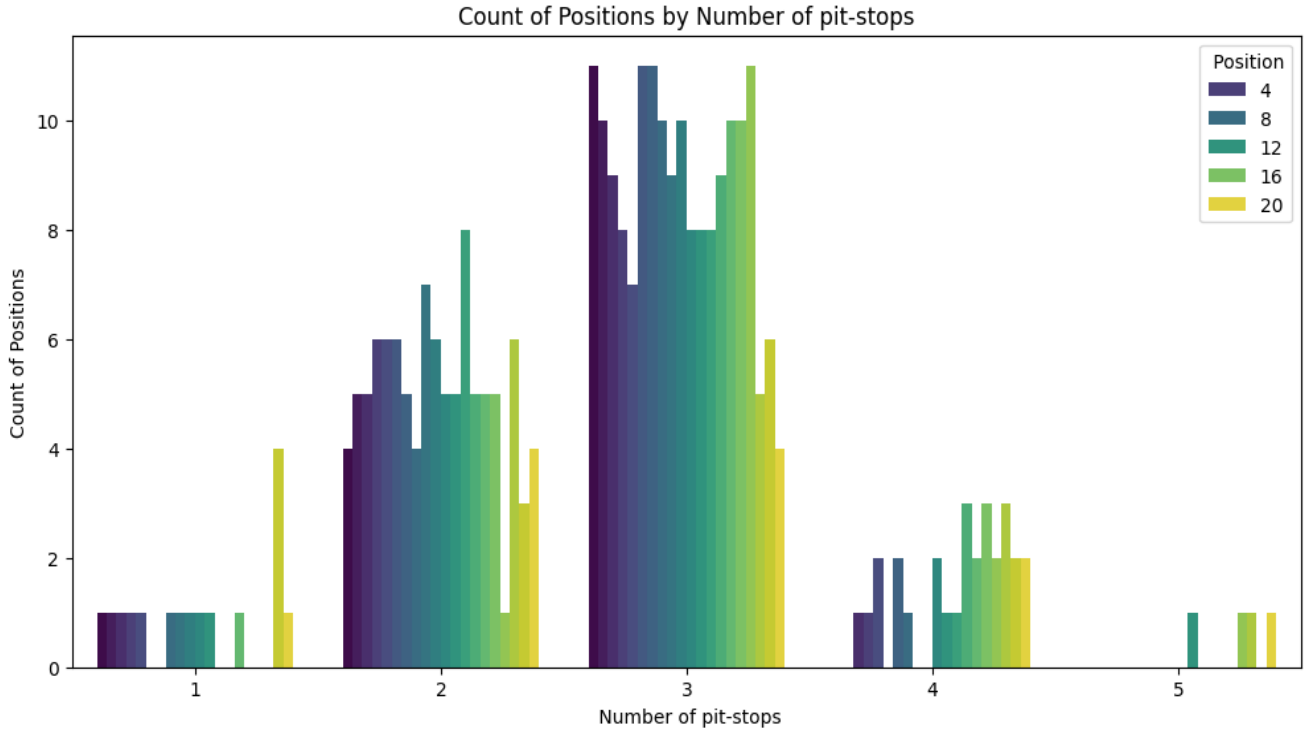


Relationships > Merged Data > Number of Stops vs Final Position

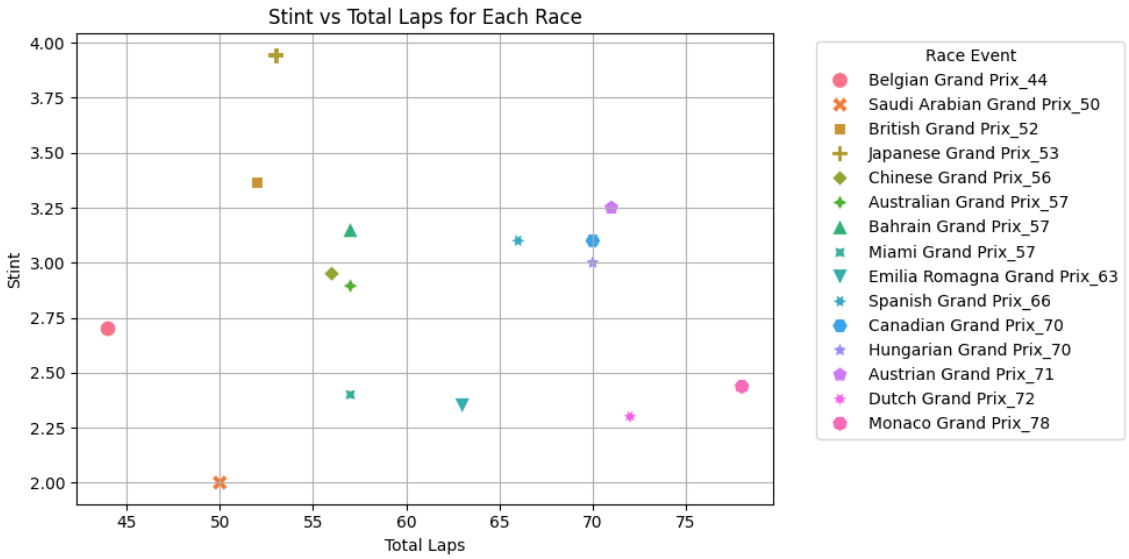
Analyze the relationship between the number of pit stops a driver makes and their final position in the race.

The image illustrates the relationship between the number of pit stops and the final race positions of drivers. based on the image , The 2 and 3 pit stop strategies are the most flexible, allowing for competitive positioning. These two pits stops are widely used by drivers with various positions however 3 pit stop strategy Slightly more concentration towards middle positions (12 to 16).

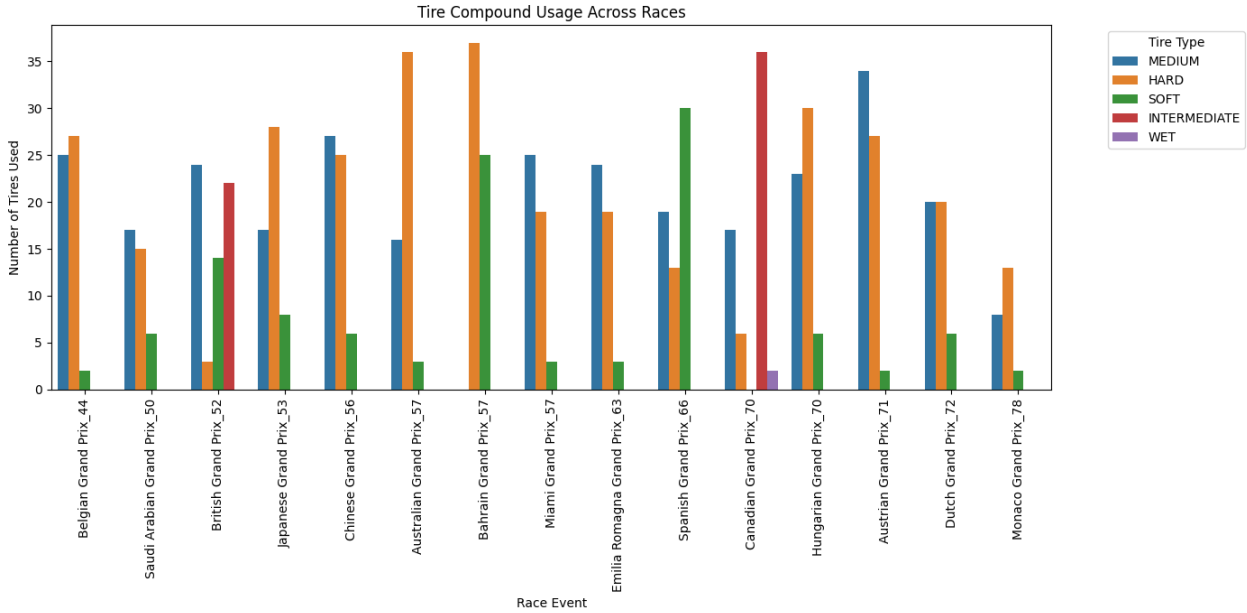
Fewer pit stops can lead to better positions but may not always be feasible due to tire wear or race conditions.
More pit stops might be a response to unexpected events, leading to poorer positions.



Relationships > Races 2024 > Race Length vs Strategy

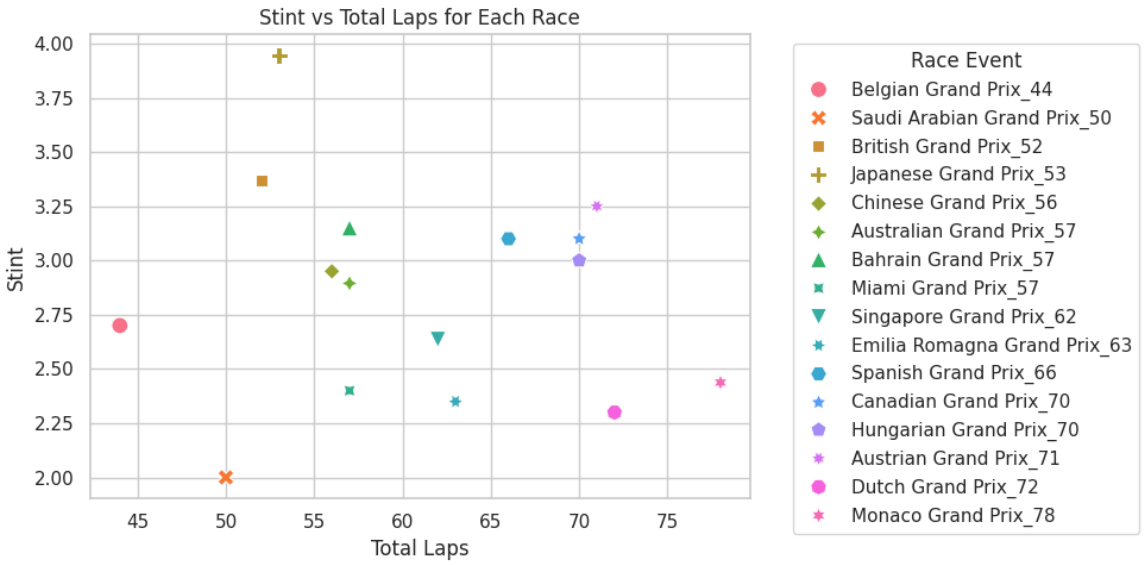


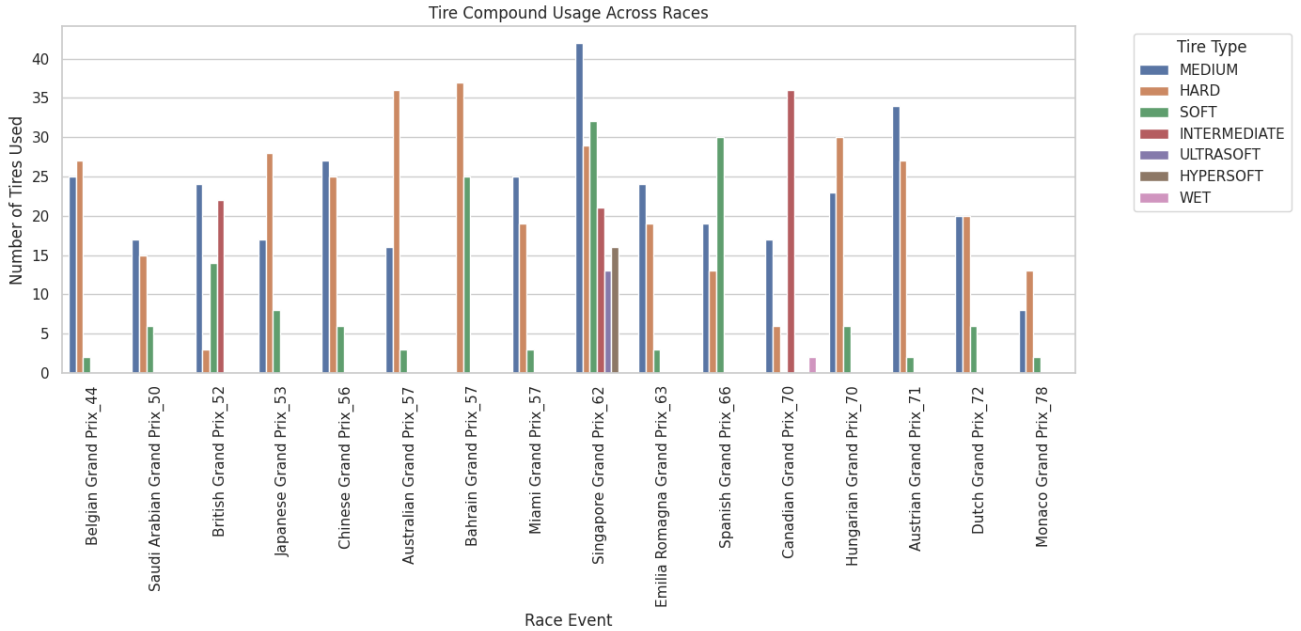
Based on the scatter plot, races with a higher total lap count (e.g., 70 laps) tend to have more stints. This suggests that teams may opt for multiple pit stops to manage tire wear and maintain optimal performance throughout the race. Races with fewer laps (e.g., 44 laps) generally have fewer stints, indicating that teams might choose more durable tire compounds to minimize pit stops.



The above bar plot shows the relationship between the compound across all races. Th x-axis shows the race name with its total lap number which are written in ascending orders .
Comparing these races indicates that for races with a higher total lap count, teams likely opted for more durable tire compounds such as Hard or Medium. These compounds wear out slower, reducing the need for frequent pit stops. The goal in longer races would be to minimize the number of pit stops, as each stop can cost valuable time. Durable tires help achieve this by lasting longer on the track.
In races with fewer laps, teams might choose softer compounds like Soft or Ultrasoft. These tires provide better grip and performance but wear out faster. Since the race is shorter, the increased wear is less of an issue.

Relationships > Merged Data > Race Length vs Strategy

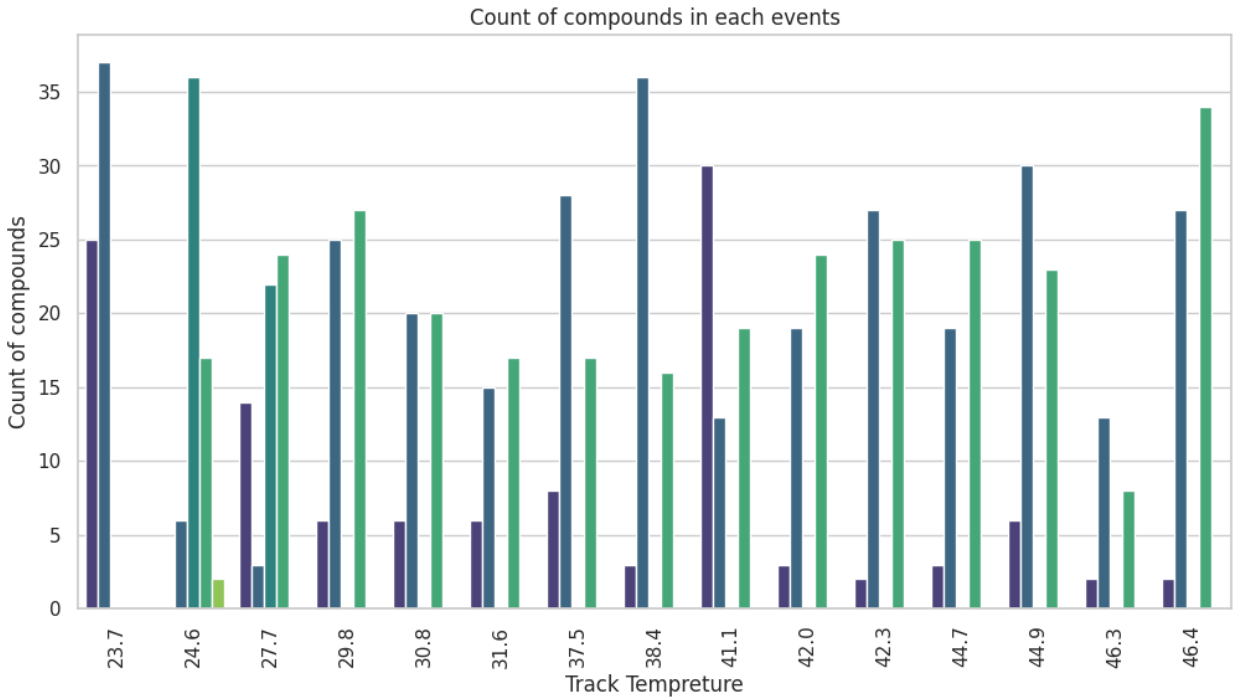




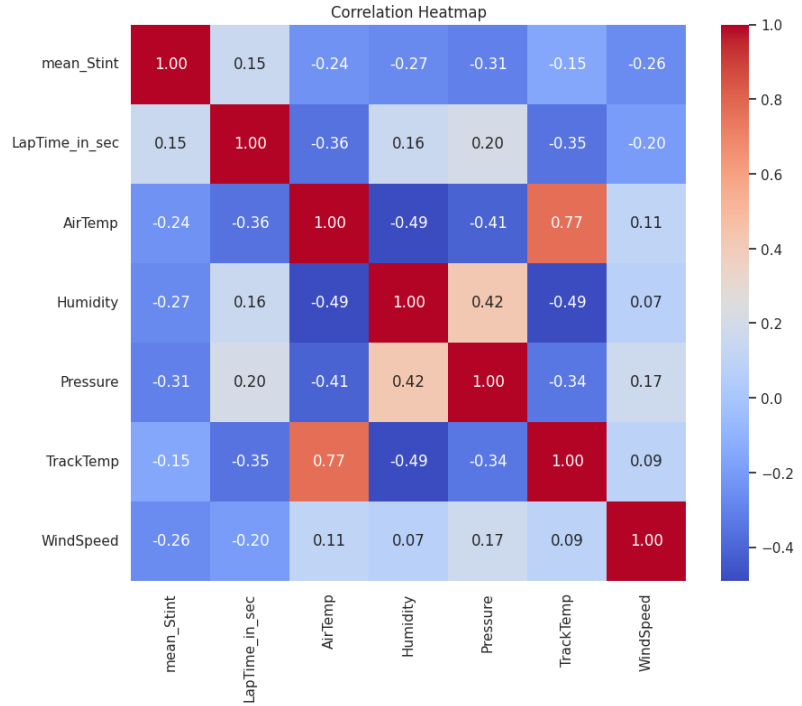
The scatter plot shows the relationship between the total number of laps in each race and the average number of stints (pit stops) teams and drivers used.
There is a general trend towards more stints in mid-length races, indicating a balance between tire performance and pit stop time.
Longer races do not always correlate with significantly more pit stops, likely due to strategic tire choices aimed at extending stint lengths.
Teams tailor their strategies based on race-specific factors beyond just the total laps, including track conditions and tire durability.

The above bar plot shows the relationship between the compound across all races. Th x-axis shows the race name with its total lap number which are written in ascending orders .
Comparing these races indicates For races with a higher total lap count, teams likely opted for more durable tire compounds such as Hard or Medium. These compounds wear out slower, reducing the need for frequent pit stops.
In races with fewer laps, teams might choose softer compounds like Soft or Ultrasoft. These tires provide better grip and performance but wear out faster. Since the race is shorter, the increased wear is less of an issue. Teams can afford to be more aggressive with their tire choices, knowing that the race will end before the tires degrade significantly.

Relationships | Races 2024 | Weather and Strategy

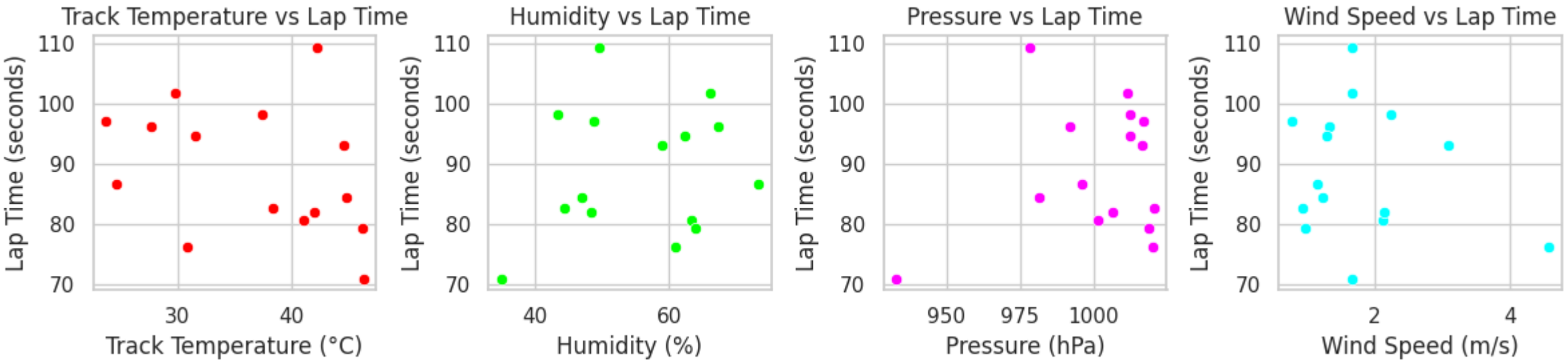


The chart illustrates the relationship between track temperature and the choice of tire compounds used by drivers. Here's a breakdown:
Soft Compound is Most frequently used in moderate track temperatures (around 38.4°C and 41.1°C). And Less common in very low or very high temperatures.
Hard Compound: Is widely used across a broad range of temperatures, particularly in higher temperatures (above 44.7°C). Also it Shows versatility and preference in hotter conditions.
intermediate Compound: this has selected mainly in mid-range temperatures (27.7°C to 31.6°C) , Usage decreases as temperatures rise above 38.4°C.
Medium Compound: it is popular in cooler temperatures (23.7°C to 31.6°C).
Wet Compound: Predominantly used in the lowest track temperature (23.7°C).



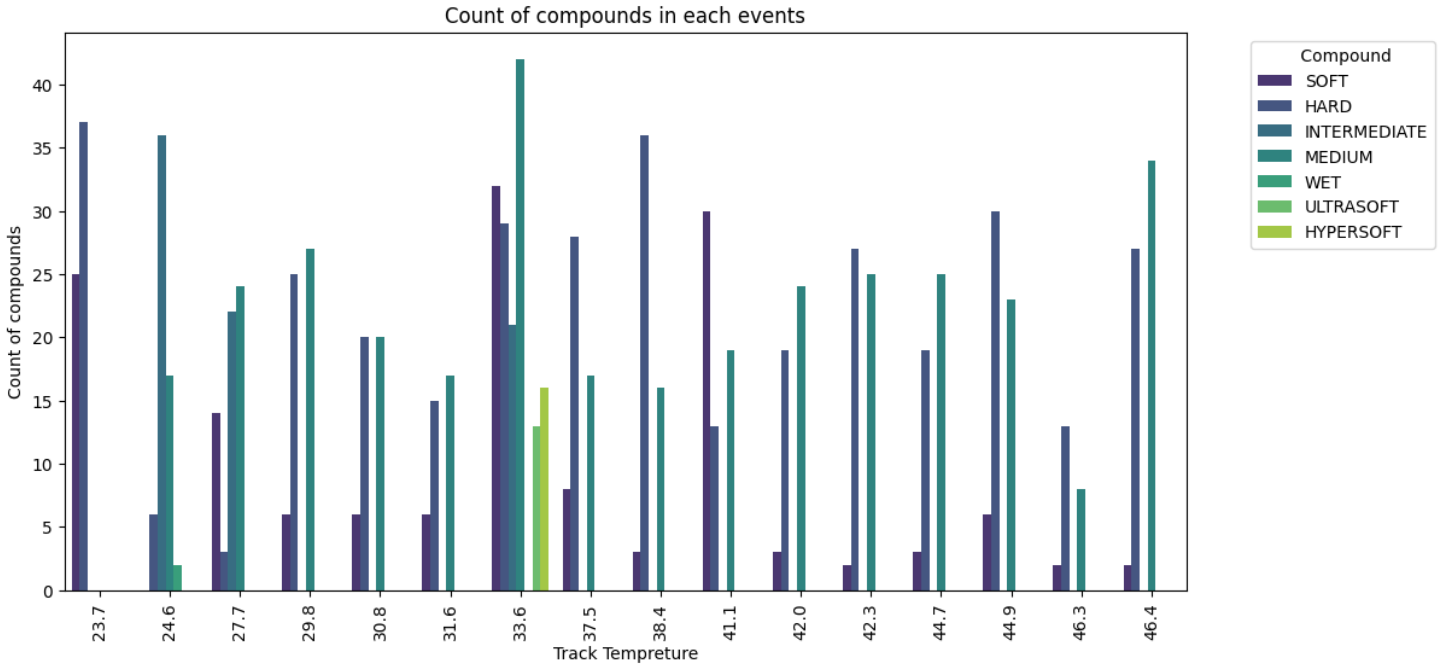
The correlation heatmap revealing how weather conditions impact performance. Key findings include a moderate negative correlation between lap times and both track and air temperatures, indicating that higher temperatures generally lead to faster lap times, while humidity tends to slow them down slightly. Track temperature shows a strong positive correlation with air temperature, and a moderate negative correlation with pressure. Other correlations are weaker, such as the minimal impact of wind speed on lap times and temperature. This data offers valuable insights for teams to optimize strategies and car setups based on varying weather conditions.

Relationships > Races 2024 > Weather and Strategy

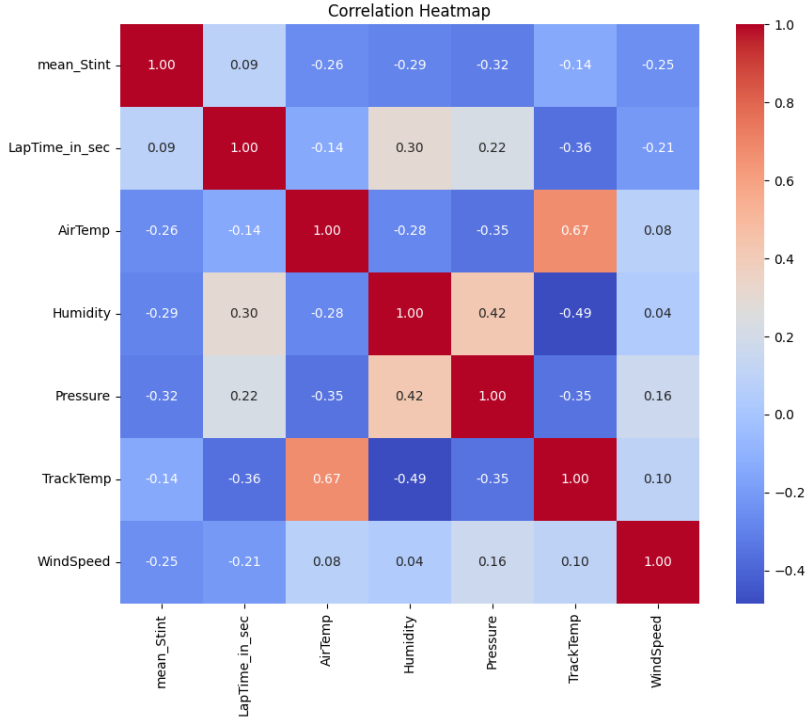


temperature : As track temperature increases, lap times tend to decrease slightly, indicating that warmer track conditions might provide better grip and faster lap times.
humidity :Higher humidity levels seem to correlate with slightly longer lap times, possibly due to the impact on engine performance and tire grip.
pressure : There is a slight trend where higher pressure might be associated with marginally faster lap times, potentially due to better air density aiding engine performance.
wind speed :Increased wind speed appears to have a mixed effect on lap times, possibly due to varying impacts on car aerodynamics and stability.

Relationships > Merged Data > Weather and Strategy

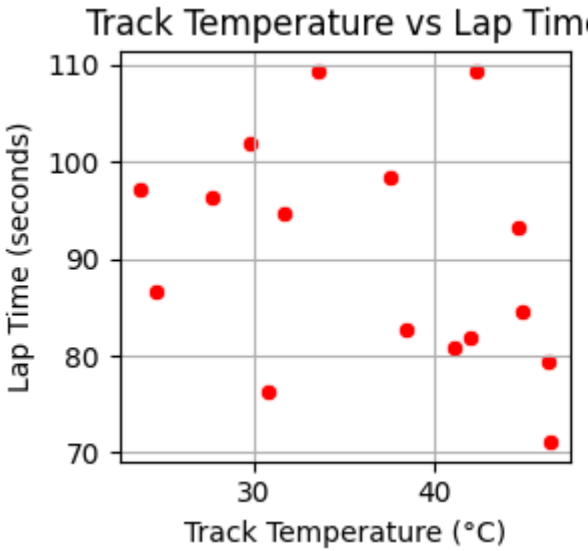


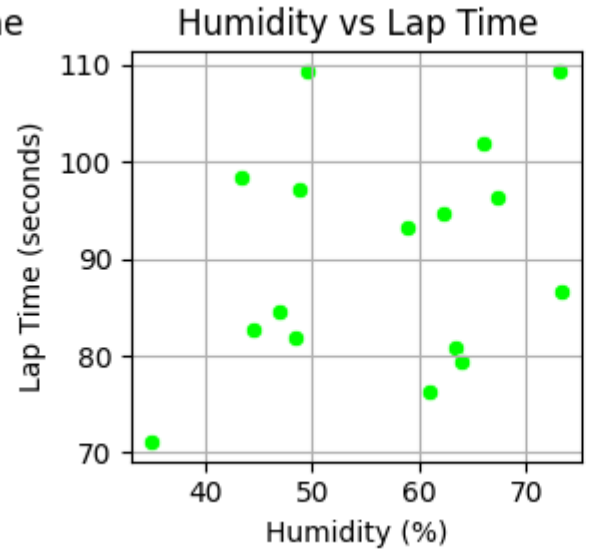
The bar graph illustrates the relation between track temperature and tire compound selection in Formula 1 racing, covering a temperature range from 23.7°C to 46°C. The analysis reveals that soft, hard and medium compounds are the most commonly selected across a broad range of temperatures, particularly thriving in the mid-range of 33.6°C to 44.9°C. In contrast, Intermediate and wet compounds are favored in cooler conditions below 30°C, while ultrasoft and hyper-soft compounds are used less frequently, typically for specific race strategies or qualifying sessions.

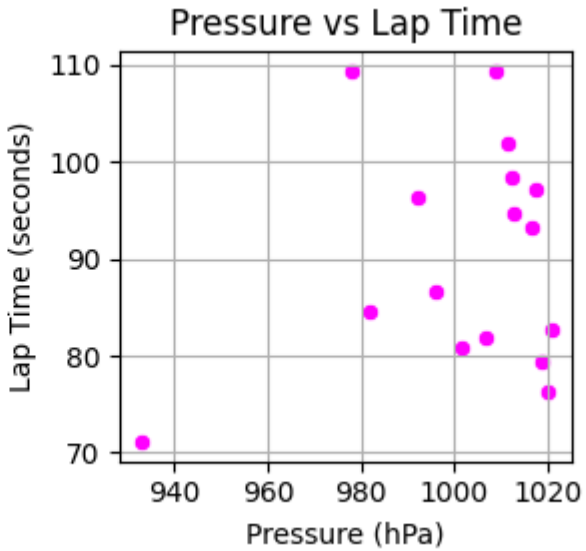


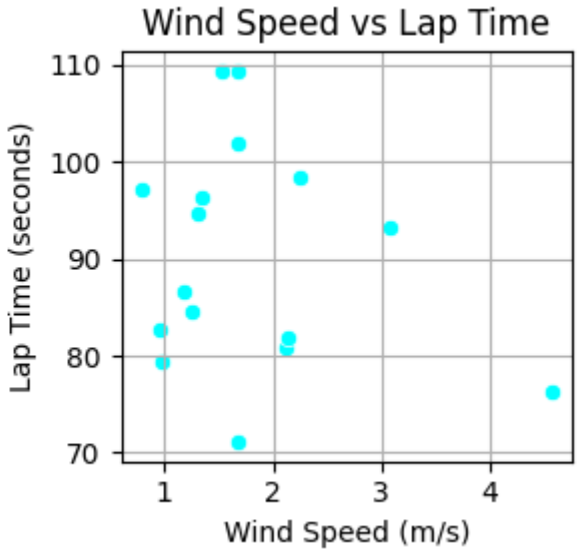
The correlation heatmap revealing how weather conditions impact performance. Key findings include a moderate negative correlation between lap times and both track and air temperatures, indicating that higher temperatures generally lead to faster lap times, while humidity tends to slow them down slightly. Track temperature shows a strong positive correlation with air temperature, and a moderate negative correlation with pressure. Other correlations are weaker, such as the minimal impact of wind speed on lap times and temperature. This data offers valuable insights for teams to optimize strategies and car setups based on varying weather conditions.

Relationships > Merged Data > Weather and Strategy







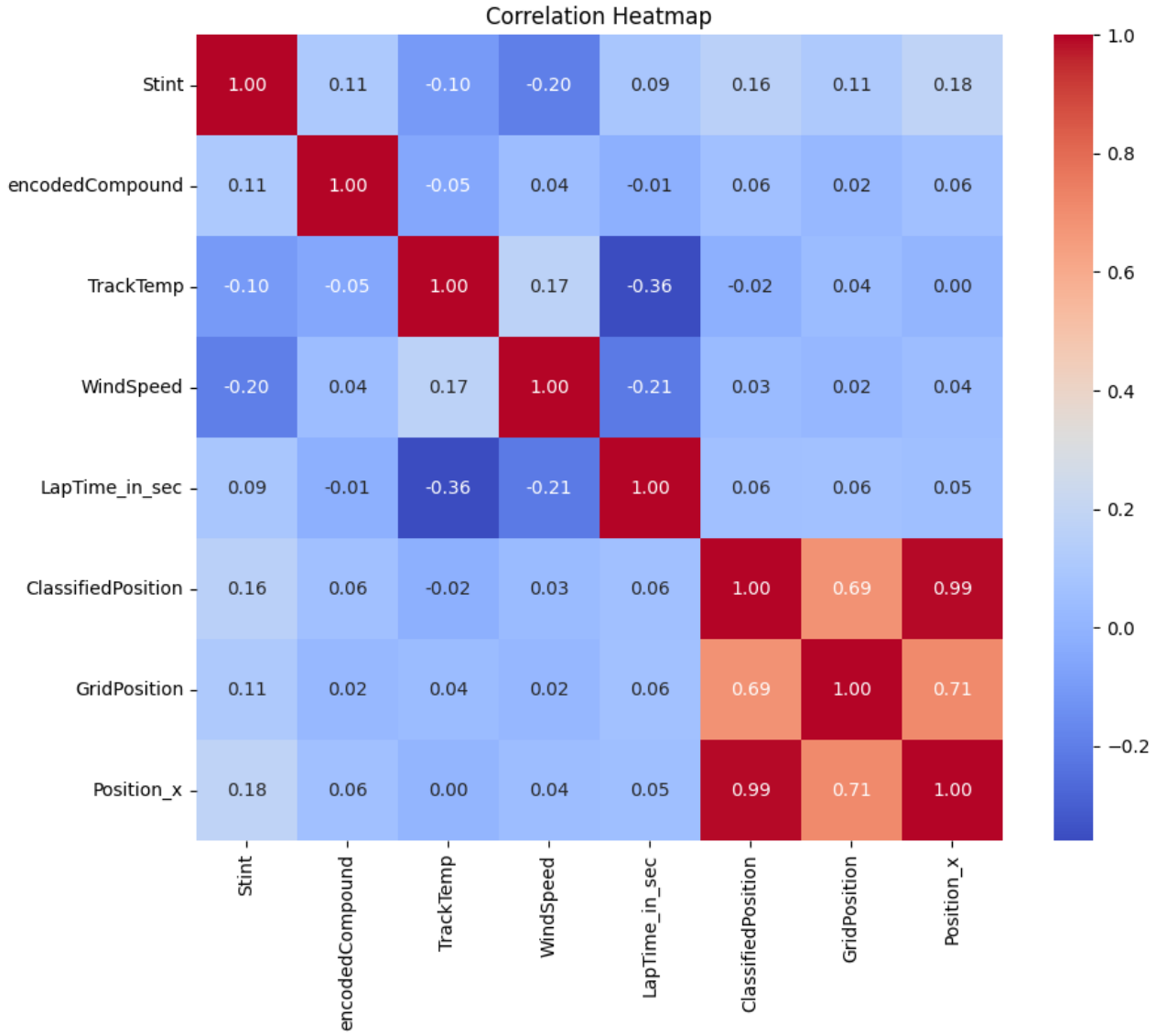


Bonus Question

Bonus Question 2024 Data

this image shows the impact of various factors in formula 1 races on final position. it is based on drivers in each event. encodedCompound shows the number of each tire compound a driver used in each races. Based on this image

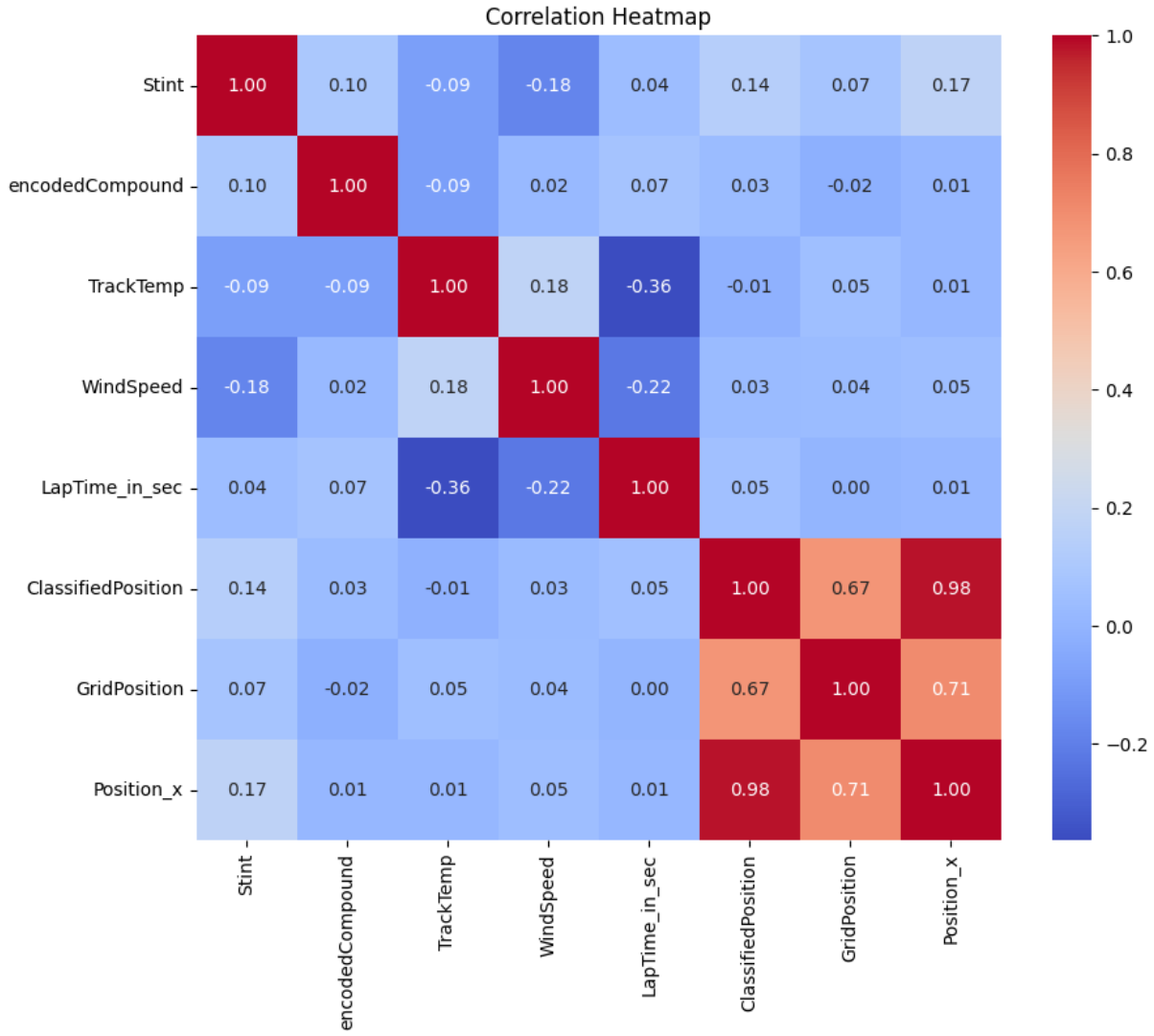
Grid Position and Final Position: There's a strong positive correlation between GridPosition and ClassifiedPosition. This means that where a driver starts on the grid significantly impacts their final position. Drivers starting at the front are more likely to finish in higher positions.
Lap Time: LapTime_in_sec shows a moderate positive correlation with ClassifiedPosition, indicating that faster lap times generally lead to better final positions.
Track Temperature and Wind Speed: These factors seem to have little to no correlation with the final position, suggesting that they might not be as critical in determining race outcomes.
Tire Compounds: EncodedCompound has a slight impact on GridPosition. The choice of tire compounds can influence race strategy and performance.
Stint: The number of stints (periods between pit stops) also shows some correlation, indicating that pit stop strategy plays a role in the final outcome.



Bonus Question Merged Data

The image you provided is a correlation heatmap that illustrates the relationship between various factors and their impact on the final position.

Grid Position and Final Position: There's a strong positive correlation (0.67) between GridPosition and ClassifiedPosition. This indicates that where a driver starts on the grid significantly influences their final position. Starting higher up generally leads to a better finishing position.
Lap Time and Final Position: LapTime_in_sec also shows a slight positive correlation with ClassifiedPosition. Faster lap times are associated with better final positions, which is expected as consistent speed is crucial in racing.
Track Temperature: Interestingly, TrackTemp has a very low correlation (-0.01) with the final position, suggesting that track temperature doesn't significantly impact the race outcome.
Wind Speed: WindSpeed shows a slight negative correlation with ClassifiedPosition. While not very strong, it indicates that higher wind speeds might slightly hinder performance.
Tire Compounds: The EncodedCompound factor, which likely represents different tire compounds, shows some correlation with final positions. The choice of tire can influence race strategy and performance.
Stint: The Stint factor, representing different segments of the race, shows varying correlations with other factors, indicating its role in race strategy and performance management.
Overall, the heatmap provides valuable insights for teams to strategize their races by focusing on factors like grid position, lap times, and tire choices, while not worrying too much about track temperature. This data-driven approach can help in making informed decisions to improve race outcomes.



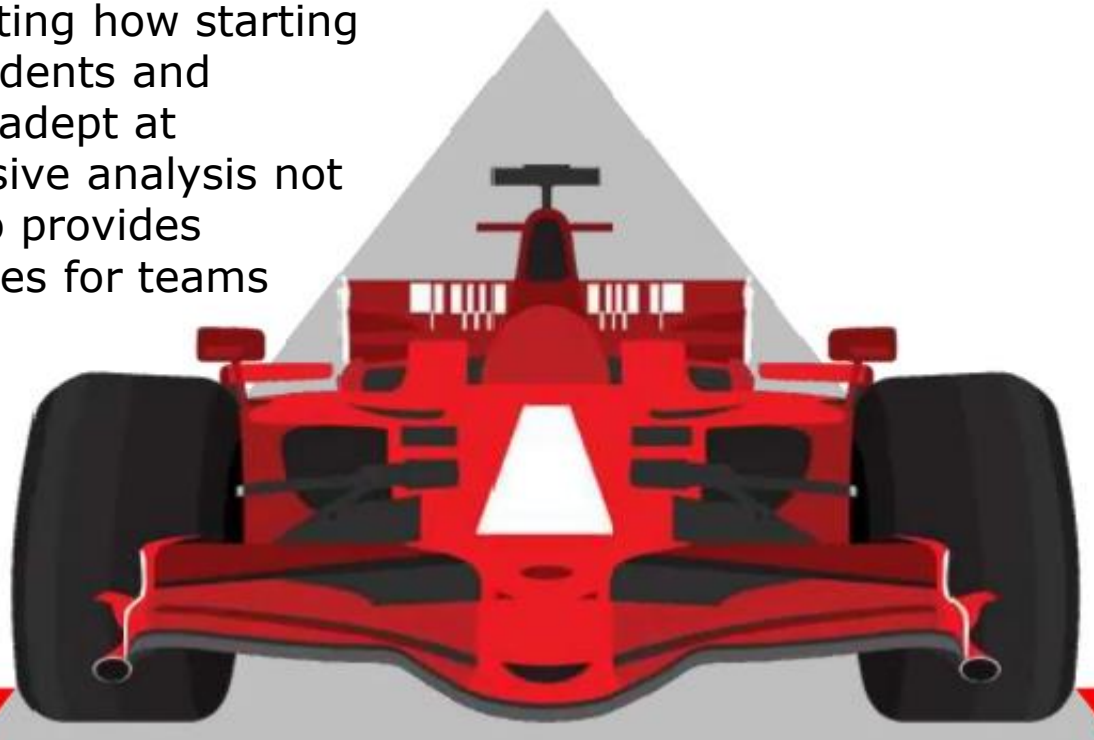
Conclusion


INEOS
PETRONAS
AMG
Einhell
IWC
PETRONAS
Synthium
SOLERA
PETRONAS

Conclusion

The data analysis of the 2024 Formula 1 season reveals critical insights into the dynamics of racing performance influenced by various factors such as car specifications, tire strategies, and environmental conditions. The strong correlations identified between speed, RPM, and gear underscore the mechanical intricacies that teams must navigate to optimize performance on track. Additionally, understanding tire usage patterns highlights strategic decisions made by teams in response to race conditions, which can significantly impact overall race outcomes.

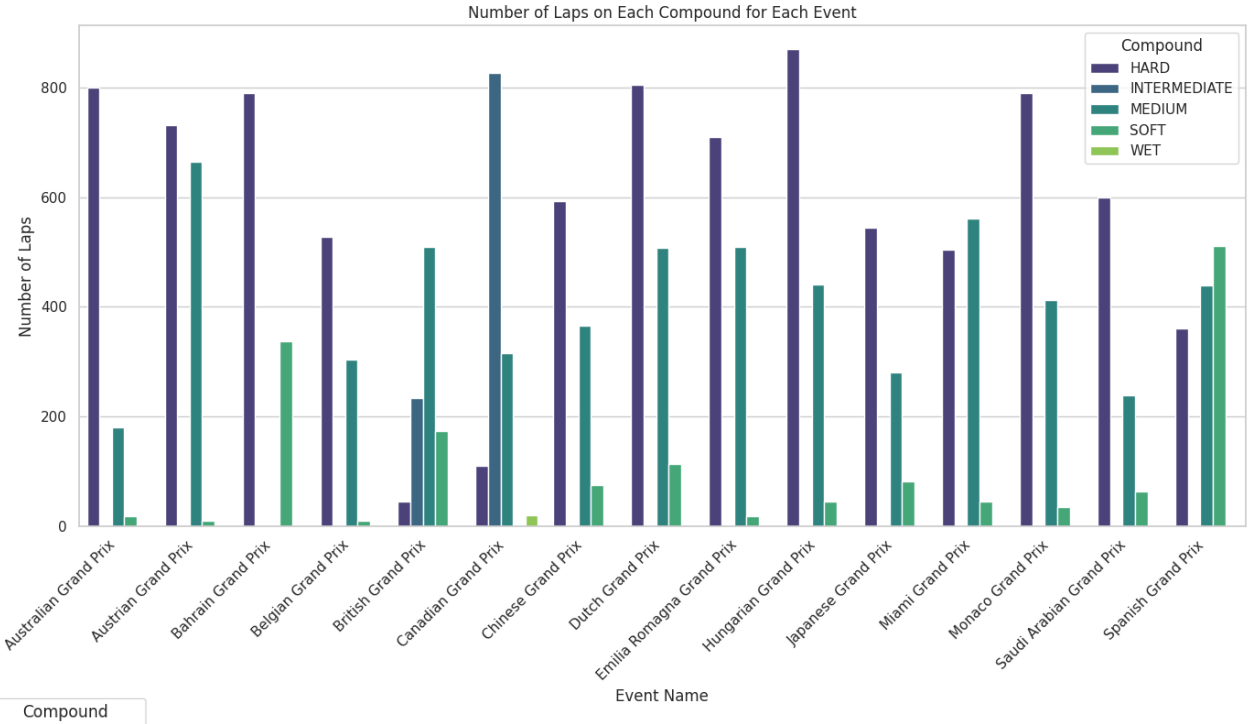
Moreover, the consistent performance of top drivers like Max Verstappen emphasizes the competitive nature of Formula 1 while illustrating how starting positions can dictate race results. With a low incidence of accidents and retirements reported, it suggests that teams are increasingly adept at managing race strategies effectively. Overall, this comprehensive analysis not only enhances our understanding of Formula 1 racing but also provides valuable data-driven insights that could inform future strategies for teams aiming for success on the track.

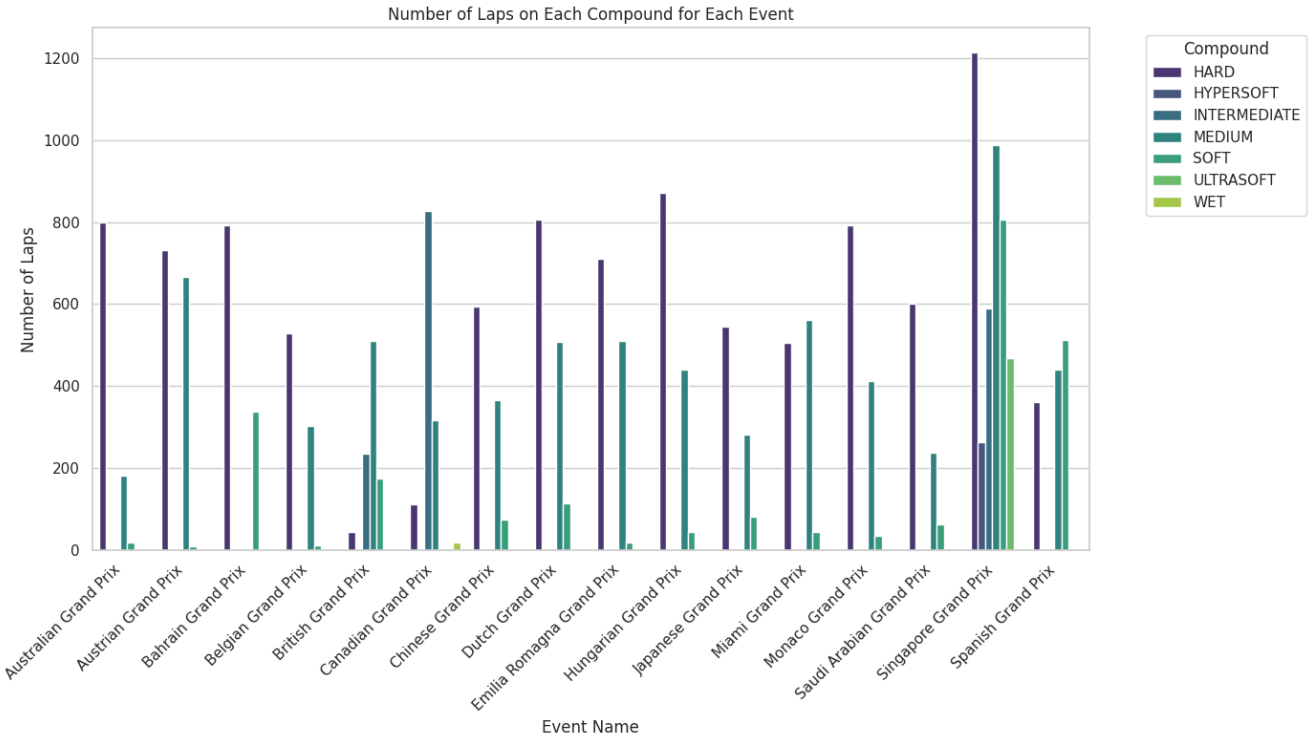


Appendix


EPSON
CROWDSTRIKE
44
PETRONAS
PRIMAX
PETRONAS
Qualcomm
snapdragon
AMG
IWC
125

Appendix


Number of Laps on Each Compound for Each Event
Compound
HARD
INTERMEDIATE
MEDIUM
SOFT
WET
Number of Laps
0
200
400
600
800
Australian Grand Prix
Austrian Grand Prix
Bahrain Grand Prix
Belgian Grand Prix
British Grand Prix
Canadian Grand Prix
Chinese Grand Prix
Dutch Grand Prix
Emilia Romagna Grand Prix
Hungarian Grand Prix
Japanese Grand Prix
Miami Grand Prix
Monaco Grand Prix
Saudi Arabian Grand Prix
Spanish Grand Prix
Event Name


Number of Laps on Each Compound for Each Event
Compound
HARD
HYPERSOFT
INTERMEDIATE
MEDIUM
SOFT
ULTRASOFT
WET
Number of Laps
0
200
400
600
800
1000
1200
Australian Grand Prix
Austrian Grand Prix
Bahrain Grand Prix
Belgian Grand Prix
British Grand Prix
Canadian Grand Prix
Chinese Grand Prix
Dutch Grand Prix
Emilia Romagna Grand Prix
Hungarian Grand Prix
Japanese Grand Prix
Miami Grand Prix
Monaco Grand Prix
Saudi Arabian Grand Prix
Singapore Grand Prix
Spanish Grand Prix
Event Name


Thank You For Reading