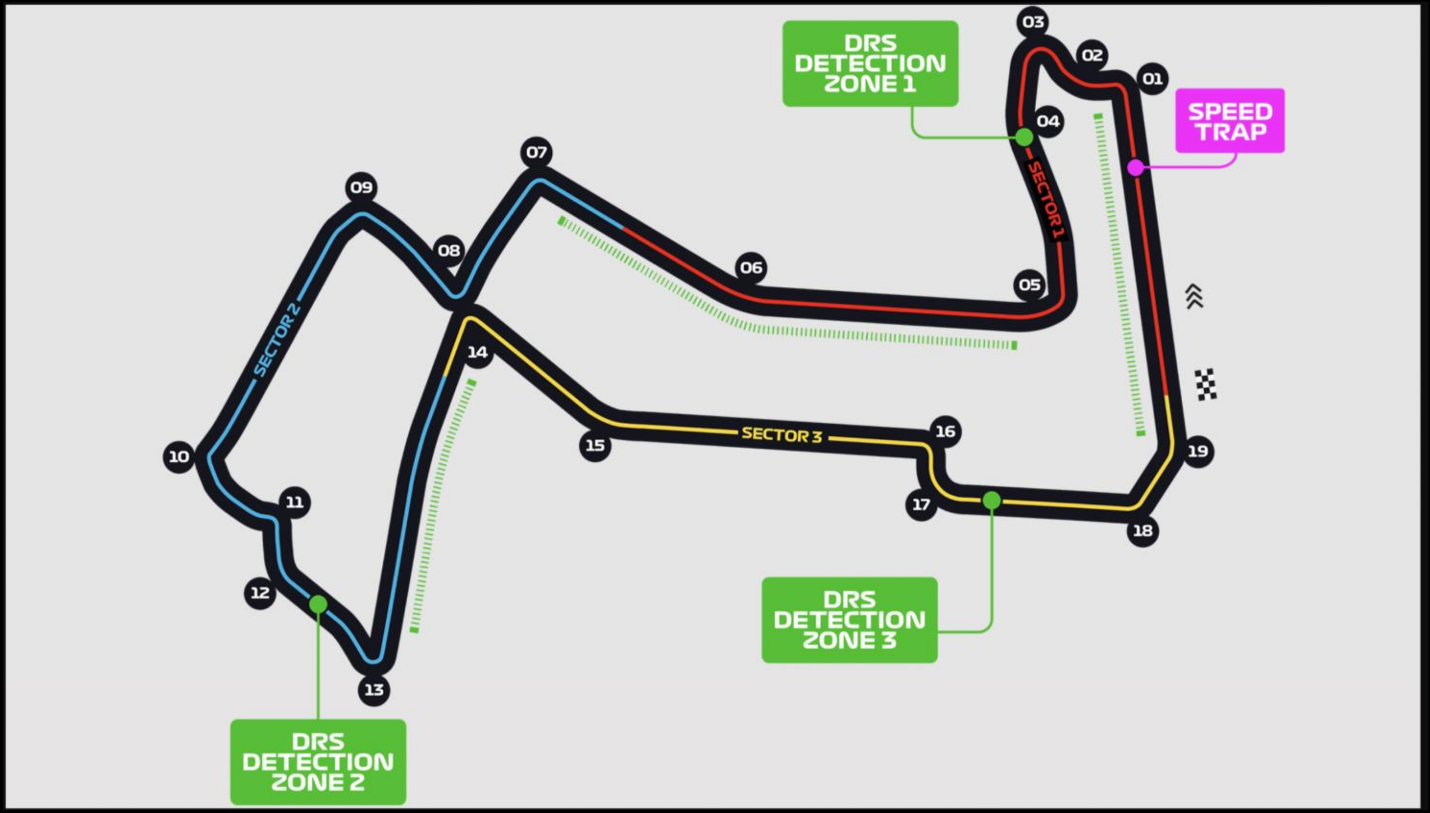
AN ANALYSIS OF THE 2023 FORMULA 1 SINGAPORE GRAND PRIX

BY DANNY SATTERTHWAITE, BERN ADU, AND MASOOD ALI KHAN

INTRODUCTION: ABOUT THE EVENT

What is an F1 race and how is it structured?

- The race is on Sunday. First driver to cross the line after 50+ laps wins.
- Qualifying: Sets the start order. The faster the single lap in the session, the higher up a driver qualifies.
- 3 practice sessions: Used to adjust the car setup to make the car as quick as possible for a single lap (qualifying) and over the course of 50+ laps (for the race).



THE DATA AND OUR GOALS

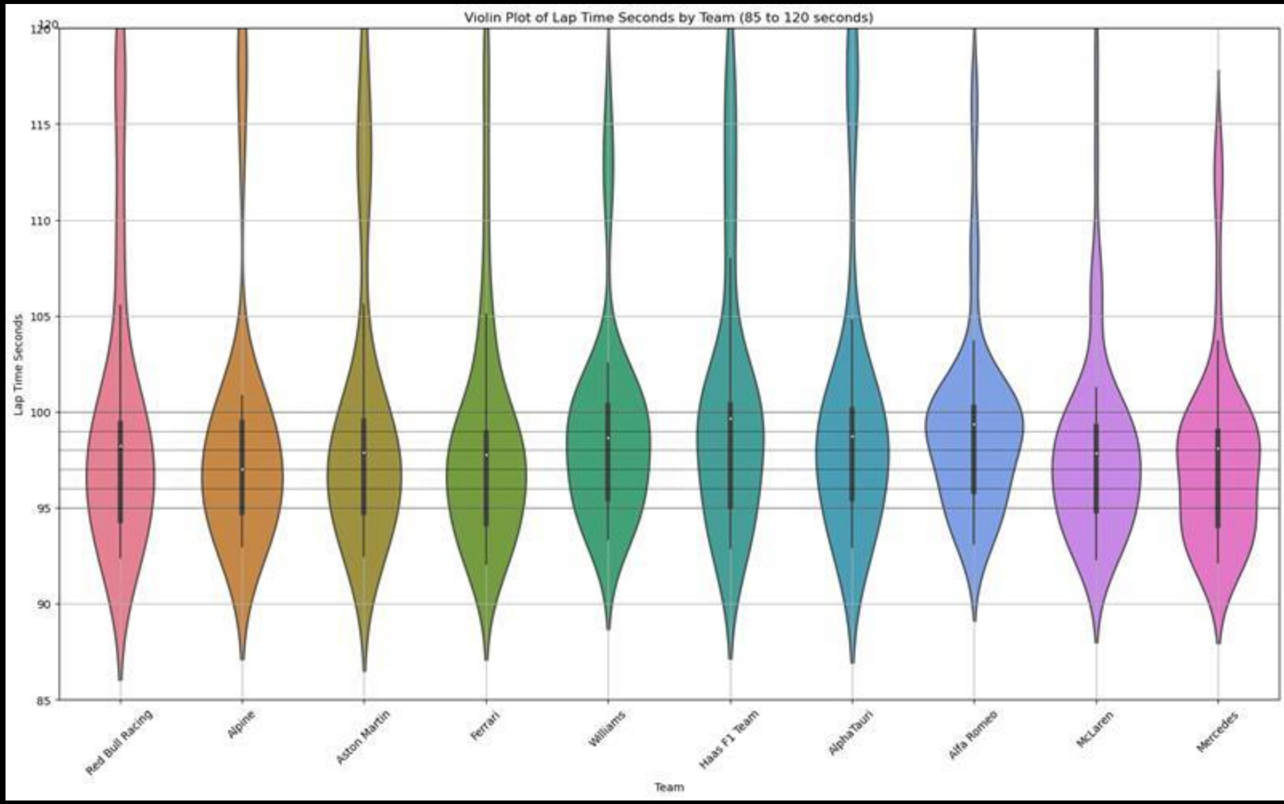
- Goals: What can we learn from these practice sessions to inform our strategies for qualifying and the race? And with real-time lap data during the race, how can we adjust strategies and influence the race’s outcome.
- Variables of interest: Driver, Team, Lap Time, Lap Number, Sector times, Tire Life, Tire Compound

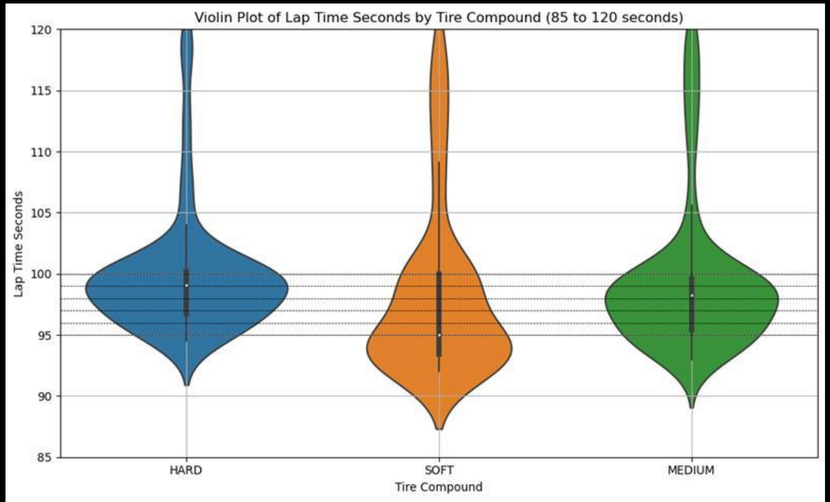
OUR APPROACHES

- EDA to explore potential relationships
- Brief analysis of qualifying results
- Multiple regression to examine the relationship between Lap Time, Tire Life, and Laps Completed (as a substitute for fuel load).
- Logistic regression to model the probability of whether a driver would pit the following lap or not.
- Poisson regression to model the count of the fastest sector times achieved by a driver as a function of Track Status, Tire Life, and other variables.

EDA PRACTICE

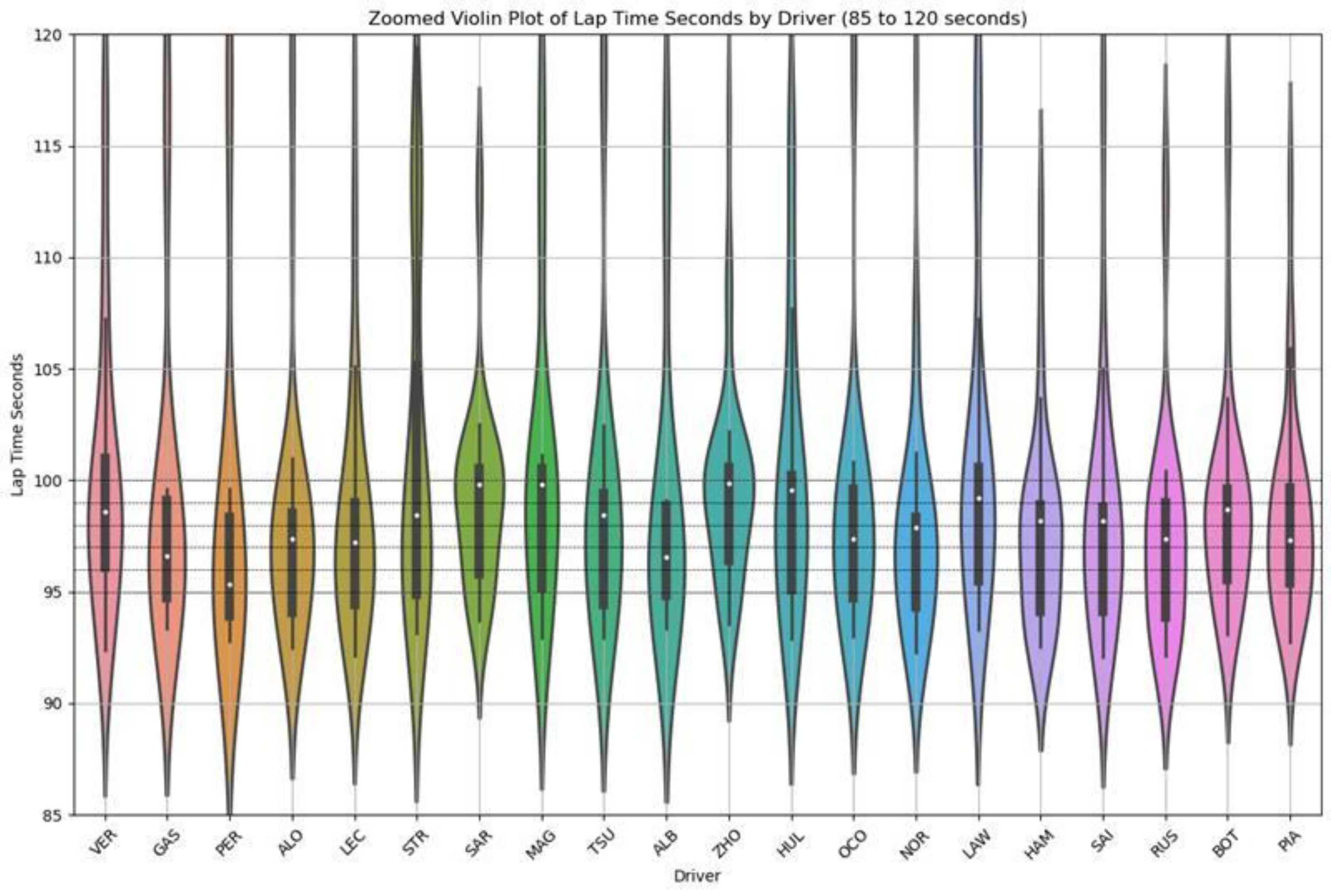
- The soft compound tires are clearly the fastest of the compounds.
- The distribution of the medium and hard tires appears similar with the median lap time for the medium tires being slightly quicker.
- No clear pattern between quickness in practice (by median lap time) and race results when looking by team.





EDA PRACTICE CONTINUED

- Looking at the distribution of lap time across all sessions for each driver we can see that these distributions all appear quite similar.
- Not particularly surprising as each driver and team are experimenting with different car setups and run plans.



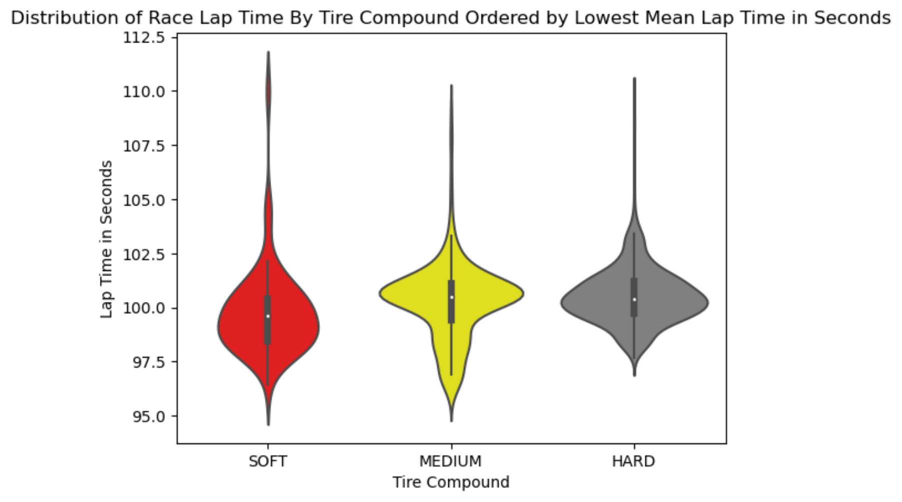
EDA QUALIFYING

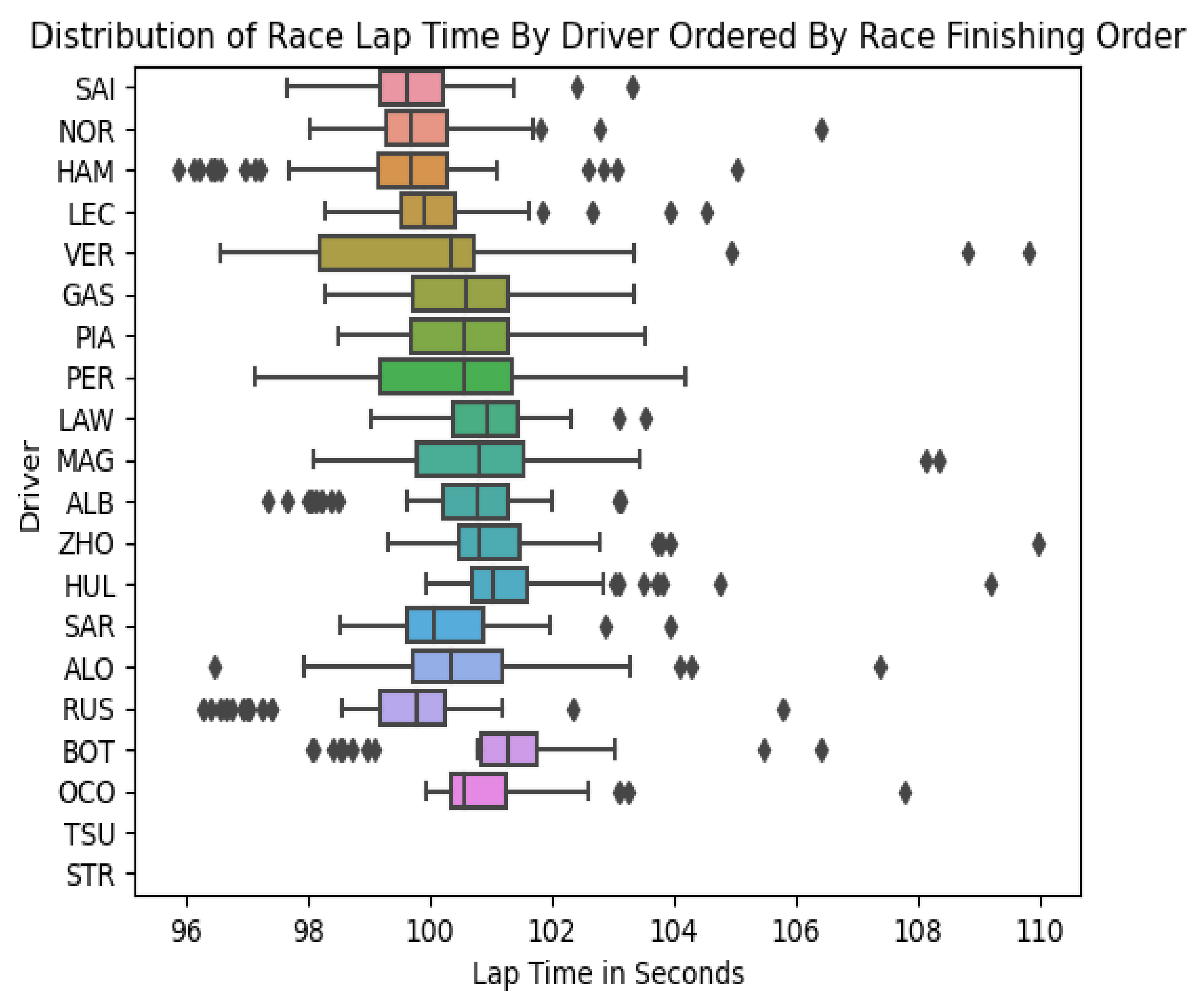
- Qualifying results
- Incredibly close times

Driver	Time in Seconds (Q3)	Difference to Next Driver
Carlos Sainz	90.984	0
George Russell	91.056	0.072
Charles Leclerc	91.063	0.007
Lando Norris	91.27	0.207
Lewis Hamilton	91.485	0.215
Kevin Magnussen	91.575	0.09
Fernando Alonso	91.615	0.04
Esteban Ocon	91.673	0.058
Nico Hulkenberg	91.808	0.135
Liam Lawson	92.268	0.46

EDA RACE

- The softer the tire compound, the quicker on average the lap is.
- The top 3 finishers also had the quickest median lap times
- The two Ferrari drivers (SAI and LEC) had very similar lap time distributions, as we would expect as they are driving the same car.





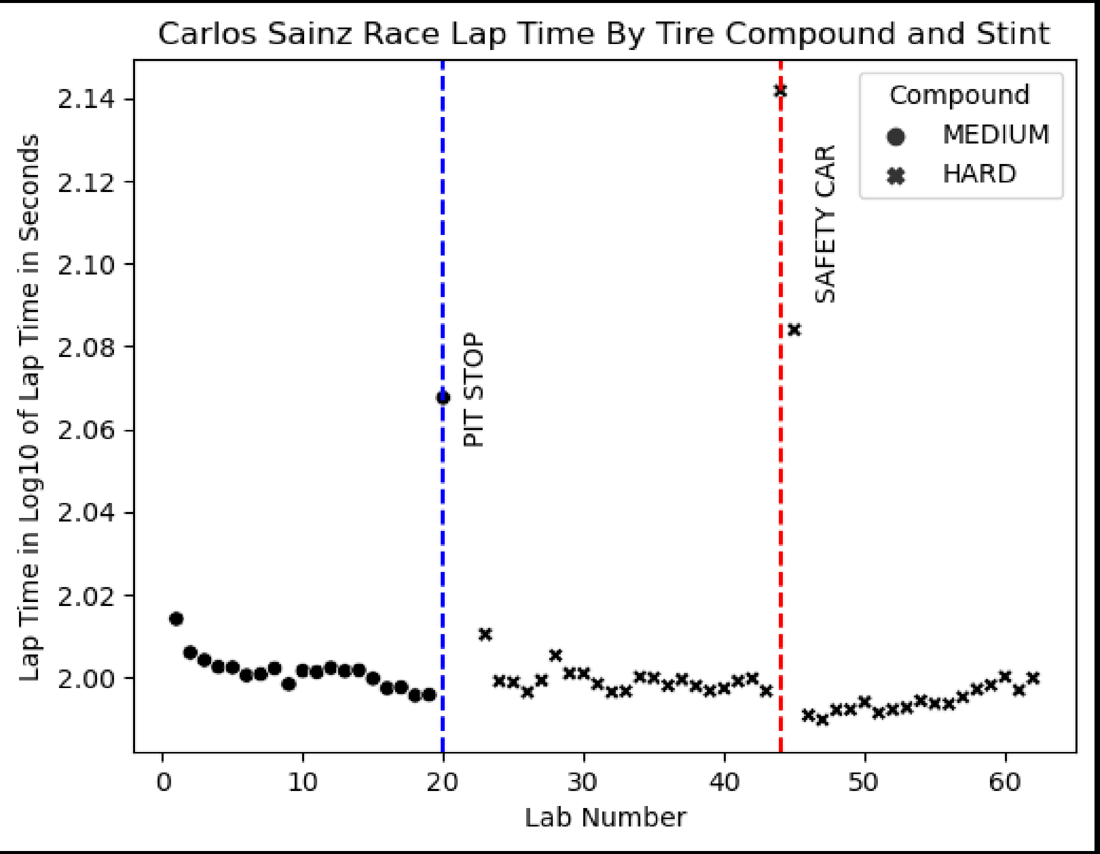
ANALYSIS QUALIFYING

- Teams fastest qualifying time vs their quickest practice time and their theoretical fastest practice time.
- The theoretical fastest lap per team is within +0.2 to +1.6 seconds of the teams' fastest qualifying driver but given the tight margins of the final qualifying laps, this theoretical fastest time is not a particularly useful predictor.
- Ex: The difference between Sainz's qualifying lap and theoretical fastest lap was +0.984 seconds which would have made him qualify 9'th.

Team	Team Fastest Qualifying Time (Seconds)	Team Best Practice Time (Seconds)	Team Theoretical Fastest Time (Seconds)	Difference Between Qualification Time and Fastest Practice Time (Seconds)	Difference Between Qualification Time and Theoretical Fastest Time (Seconds)
Ferrari	90.984	92.12	91.965	1.136	0.981
Mercedes	91.056	92.355	92.349	1.299	1.293
McLaren	91.27	92.711	92.594	1.441	1.324
Haas	91.575	93.017	92.965	1.442	1.39
Aston Martin	91.615	92.478	92.965	0.863	1.35
Alpine	91.808	93.361	93.361	1.553	1.553
AlphaTauri	92.166	93.285	93.133	1.119	0.967
Red Bull	92.17	92.812	92.659	0.642	0.489
Williams	92.668	94.327	93.768	1.659	1.1
Alpha Romeo	92.809	93.105	93.066	0.296	0.257

ANALYSIS RACE PART 1

- Used multiple linear regression to model Lap Time by Stint, Compound, Tire Life, Fresh Tire, and Lap Number
- Iterated over all combinations of these variables. Best model had a very poor R-Squared of 1.9% but Lap Number and Hard Compound Tires were statistically significant predictors.
- Likely a non-linear relationship and as Lap Number was significant it is likely that fuel load had more impact than tire life.



ANALYSIS RACE PART 2

- Utilized the multiple logistic regression to assess the association between Pit Stop and the other predictors.
- We obtained the R-squared of 18.77% which means that about 19% of the variance in the model is explained by the predictors.
- Lap number, Tire Life, and Track Status appear to be statistically significant, meaning that those predictors significantly influence whether a driver would pit or not
- The other predictors (various speed variables and Lap Time) were not significant.

Independent variable	Coefficient (s.e.)	Odds ratio	95% CI	P-value
Intercept	-15.0284 (18.44)			0.42
Lap Number	-0.1244 (0.03)	0.88	0.83 – 0.93	<0.001
Tire Life	0.2025 (0.04)	1.22	1.12 – 1.33	<0.001
Track Status	0.0838 (0.04)	1.09	1.01 – 1.17	0.03
SpeedL1	0.0299 (0.03)	1.03	0.97 – 1.09	0.32
SpeedL2	0.0183 (0.03)	1.02	0.97 – 1.07	0.48
SpeedFL	-0.0382 (0.03)	0.96	0.91 – 1.01	0.16
SpeedST	0.0310 (0.03)	1.03	0.97 – 1.10	0.35
Lap Time (seconds)	-0.0228 (0.10)	0.98	0.24 – 0.51	0.83

Abbreviations: CI-confidence interval; s.e.- standard error. A 2-sided p-value <0.05 indicate statistically significant*Overall p-value<0.001 was based on a likelihood ratio test.R-squared of 18.77% was obtained from the model.

ANALYSIS RACE PART 3

- Conducted further analysis to model the frequency of fastest sector times for each session. Here just to make it simple we aggregated the data to count how many times each driver achieved the fastest time.
- Used Poisson regression to model the frequency of fastest times across all sectors.
- The model had a very good R-squared of about 99.95% but only Tire Life and Track Status were found to be significant. This suggests that Tire Life and Track Status significantly affect the ability of a driver to achieve the fastest sector time, with improvement in those predictors leading to higher counts of fastest times.

Independent variable	Coefficient (s.e.)	Exp(coef)	95% CI	P-value
Intercept	3.5491 (1.38)			0.01
Tire Life	0.0341 (0.01)	1.03	1.02 – 1.05	<0.001
Track Status	0.3381 (0.07)	1.40	1.22 – 1.61	<0.001
SpeedL1	0.0002 (0.01)	1.00	0.99 – 1.00	0.92
SpeedL2	0.0002 (0.01)	1.00	0.99 – 1.00	0.92
Position	0.0022 (0.01)	1.00	0.99 – 1.01	0.65

Abbreviations: CI-confidence interval; s.e.- standard error. A 2-sided p-value <0.05 indicate statistically significant. R-squared of 99.95% was obtained from the model.

DISCUSSION AND CONCLUSION

- Tire Life, Track Status and Tire Compound significantly influence the predictive ability of a race whether looking at pit stops, fastest sector times, or lap times.
- We suspect that using multiples races, would improve our models' predictive abilities by accounting for average pace and consistency of the teams and drivers.