

Short Report: Formula One Race Prediction

Motivation

I started off with one question for my Machine Learning project: Can I predict the outcome of any Formula One race given several factors? I have been a big fan of Formula One for about six years and have loved motorsport in general for much longer than that. The problem of predicting race winners and race placement has been attempted before, but I aim to make it more accurate than previously done by other students. Several GitHub repositories exist, and even university-research department-wide studies have been published on the subject. Also, Formula One race engineers and strategists do this exact project nearly every day. Given that I do not have the vast array of technology and information that race engineers have, I will be on the back foot, but several packages in python can make this incredibly accurate for the level of detail that I aim to achieve.

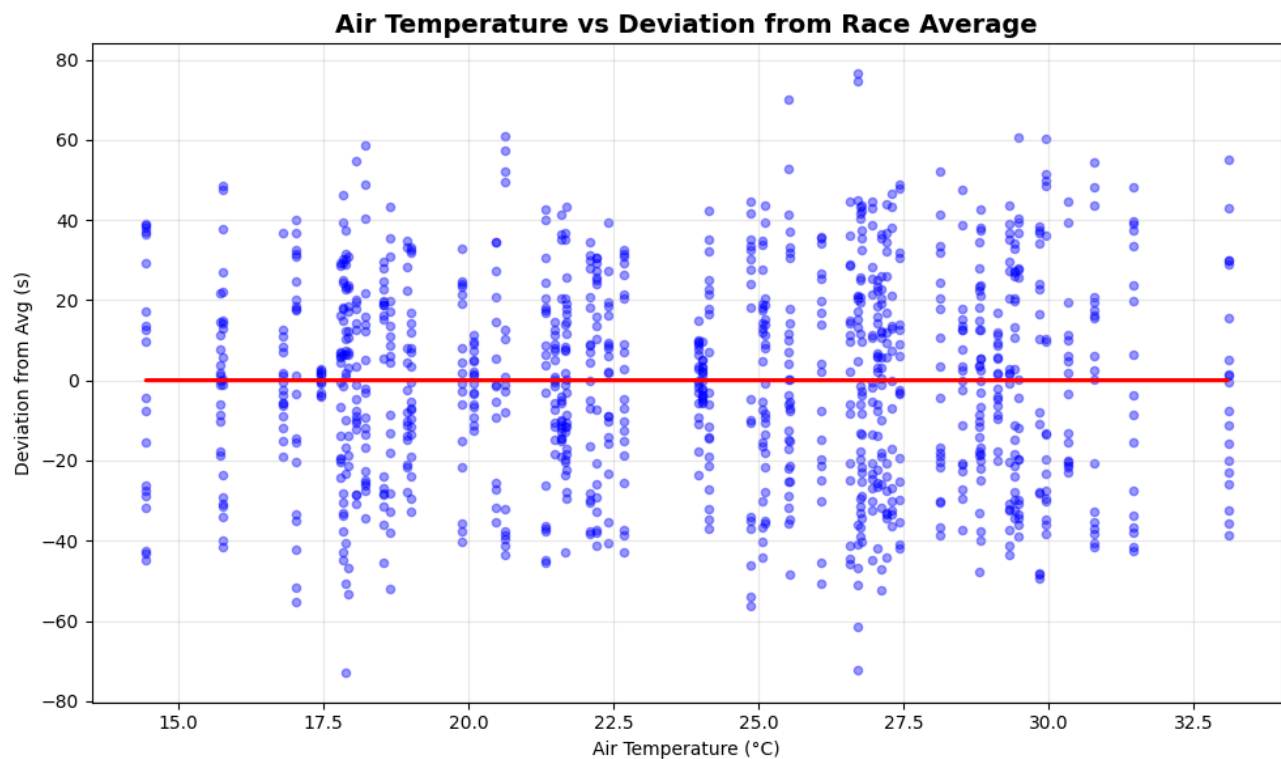
Problem Framing

Several conditions factor into performance over a Formula One race weekend. Weather, tire performance, championship standings, qualifying results, and many more conditions can be attributed to where a driver finishes in a race. By using data from previous races from the last three years, I aim to predict the winner of any Formula One race via the time delta from the average race finish time. The differential from the average race time will create the easiest method of sorting out the finishing order for all 20 drivers as well as all 22 once Cadillac joins in 2026. I will be able to input any condition that is set to be present on race weekend and have a

fair possibility of a successful prediction. These conditions can be taken day of for the most accurate assessment of race prediction.

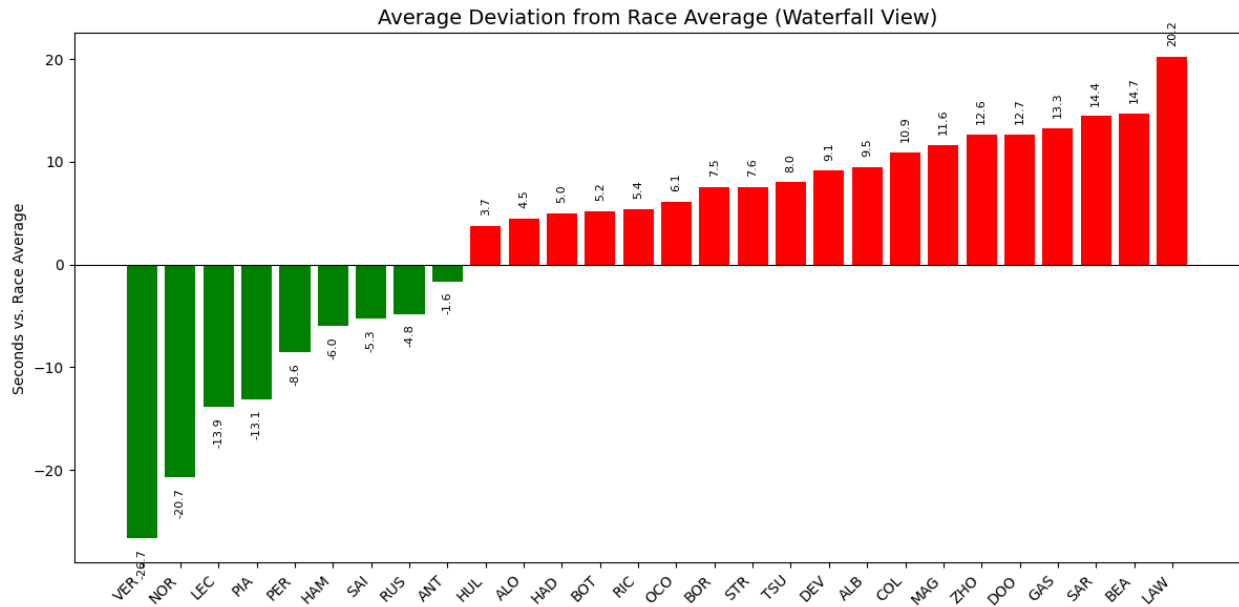
Data Overview

The dataset that I have planned to use for this project is the FastF1 Python package. This package is updated incredibly frequently and holds a vast amount of data. Tire readings, weather information (down to atmospheric pressure), sector specific times for each driver in each lap of a given race, and many more metrics are all available within the package. We can use these conditions to make several insights like the ones below to advance our model.



There are plenty of commands that allow one to collect information from any session over a race weekend such as FP1, FP2, FP3, Sprint Qualifying, Sprint Races, Qualifying, and Race Sessions. The benefit of using the FastF1 package instead of a downloadable dataset is its data refresh rate

as well as its ability to be easily aggregated and appended to other outputs that it generates. The final product and prediction metric will use the time differential of a driver taken from the average race time. The final visualization will be similar to the one below:



Time differential is the best way to show a prediction of each driver is the interval list that is shown above. This way we can accurately predict each driver's finishing result for every race instead of just the winner or podium places.

Bibliography

- frankndungu. (2025). *GitHub - frankndungu/fl-shanghai-prediction-2025: Machine learning model that predicts Formula 1 race results for the 2025 Shanghai Grand Prix using historical performance data, team strengths, and driver characteristics. Features data visualization, team change handling, and position progression forecasting.* GitHub.
<https://github.com/frankndungu/fl-shanghai-prediction-2025>
- Staub, S. (2022). Formula One Race Prediction Model. In *George Washington University*.
George Washington University.
<https://datasci.columbian.gwu.edu/sites/g/files/zaxdzs4746/files/2023-02/fl-final-presentation.pdf>