

# Formula 1 Qualifying Analysis

Mary Virginia Nye

November 9, 2024

## 1 Project Description, Background, and Motivation

Formula 1 racing represents the pinnacle of motorsport, where teams and drivers push the boundaries of technology and human performance. Qualifying sessions, held the day before the race, determine the start positions and often play a decisive role in race outcomes, particularly on circuits where overtaking is difficult. During these high-pressure sessions, drivers get only a few chances to deliver their fastest possible lap, making qualifying one of the purest tests of driver and machine performance.

This project aims to uncover how drivers perform under the pressure of qualifying by analyzing performance data across multiple seasons (2018-2023), focusing on both driver and track-specific patterns. This longitudinal view could reveal interesting patterns in driver development and adaptation to new cars or teams throughout their career that might not be apparent from race results alone.

These patterns have practical implications for teams' qualifying strategies and driver development programs. For example, identifying that certain drivers consistently perform better at specific types of tracks could influence team decisions about race weekend preparation or even future driver selection. For fans and analysts, this work provides a data-driven approach to evaluating driver performance beyond the typical race results and championship standings.

## 2 Data Description

The data is sourced through FastF1, a Python package that provides access to Formula 1 racing data. FastF1 aggregates data from the Official F1 API, Ergast F1 API, and the Official F1 Live Timing service. The dataset spans qualifying sessions from 2018 onwards.

### 2.1 Data Access Methods

#### General FastF1 Access

The FastF1 package can be installed using `pip install fastf1` and the official documentation is available at <https://docs.fastf1.dev/>. Basic data access can be achieved with just a few lines of code:

```
import fastf1
session = fastf1.get_session(2023, 'Bahrain', 'Q') # 'Q' for qualifying
session.load()
```

## Project-Specific Data Collection

There are two Python scripts for the data collection process. The `data_collection.py` script creates organized CSV files with qualifying session data for specified years. The `data_cleaning.py` script removes unnecessary columns and rows, organizes the data hierarchically by year, race, driver, and lap number, and combines all seasons into a single comprehensive CSV file.

```
python data_collection.py --years 2018 2019 2020 2021
python data_cleaning.py
```

## 2.2 Data Structure and Variables

The dataset captures detailed timing information for each qualifying lap. A lap time represents the duration from crossing the start line to crossing the finish line, with each lap divided into three sectors. These sector times provide insights into performance across different portions of the track. Speed measurements are recorded at four key points: the first intermediate point (SpeedI1), the second intermediate point (SpeedI2), the finish line (SpeedFL), and the track's fastest point, called the speed trap (SpeedST).

Drivers are identified by their name and number (e.g., L. Hamilton, 44), as well as their team affiliations. Performance tracking includes indicators such as personal best laps and stint numbers, while tire data captures compound type and wear information.

Table 1: Summary Statistics for Qualifying Sessions (2018-2024)

Metric	Value
Total Unique Drivers	39
Mean Lap Time	1:30.527
Fastest Lap Time	53.377
Fastest Lap By	V. Bottas
Average Speed (km/h)	267.35

Table 2: Sample of Qualifying Session Data - Abu Dhabi GP 2018

Driver	Lap Time	S1	S2	S3	Speed I1	Speed ST	PB
B. Hartley	1:38.713	17.567	41.797	39.349	319.0	327.0	Yes
B. Hartley	1:38.127	17.513	41.709	38.905	318.0	323.0	Yes
B. Hartley	1:37.994	17.450	41.589	38.955	320.0	328.0	Yes
C. Leclerc	1:38.968	17.271	41.969	39.728	321.0	326.0	Yes
C. Leclerc	2:22.825	29.037	61.227	52.561	190.0	190.0	No

Note: Times in mm:ss.sss format. S1/S2/S3: Sector times. PB: Personal Best lap. All speeds in km/h.

## 3 Progress and Next Steps

So far, I have used my initial exploratory Jupyter Notebook code for data gathering and cleaning to create two Python scripts that handle data collection and preprocessing for user-specified years.

Moving forward, my analysis will focus on two main areas. First, I plan to conduct track-specific analysis of qualifying times. This will require obtaining track metadata through FastF1 to understand how track

characteristics influence performance, analyze historical best laps at each circuit, and identify which drivers consistently excel overall and at specific tracks. Second, I will investigate driver performance evolution over time, examining how drivers' qualifying performances have changed throughout their careers and looking at any performance shifts that coincide with team changes.

Additional analytical directions could include examining the impact of weather conditions on qualifying performance or investigating the effect of car development throughout a season on qualifying pace.