Formula 1 Qualifying Analysis

Mary Virginia Nye

December 17, 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 only have a few chances to deliver their fastest possible lap, making qualifying one of the purest tests of driver and machine performance.

1.1 Why Qualifying Matters

In Formula 1, qualifying sessions are high-pressure events where drivers compete to set the fastest possible lap time. These sessions are critical because:

- Starting position strongly influences race outcomes, especially on tracks where passing is difficult
- Drivers only get a few attempts to set their fastest time
- The sessions reveal pure speed without race-day factors like tire management and fuel saving that influence car and driver performance

Qualifying is split into three sessions for each Grand Prix. All 20 drivers compete in Q1. The five slowest drivers are eliminated, securing the last five grid positions (16-20). The remaining 15 drivers compete in Q2. Again, the five slowest drivers are eliminated, securing grid positions 11-15. The top 10 drivers compete in Q3 to determine the top 10 positions. Pole position refers to position 1.

1.2 What This Project Does

This project creates an interactive dashboard that shows how Formula 1 drivers perform in qualifying sessions: the crucial time trials that determine starting positions for each race. By analyzing data from 2018 to 2024, this dashboard looks at how drivers' qualifying performance develops over their careers at different tracks and through team changes. 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

Each year's FastF1 data is stored in separate CVS files. The dataset captures timing information for each qualifying session of that year. The columns in the CSV files are as follows:

- Driver Number: A unique identifier for each driver
- Broadcast Name: How the driver is officially named during broadcasts
- Team Name: The racing team representing the driver
- Qualifying Positions: Final starting order for the race
- Q1: Fastest time in first qualifying session
- Q2: Fastest time in second qualifying session
- Q3: Fastest time in third qualifying session
- Year: Year of event
- EventName: Grand Prix Name
- Wet Session: True if it was a wet qualifying session (it rained), False if not

The yearly data for the Formula 1 qualifying performance dashboard is reformatted into a consolidated JSON file, `career_timeline.json`. The JSON data structure is a list of dictionaries representing the qualifying performance of a driver at a specific Grand Prix event. The key information captured includes:

- Year of the event
- Driver's name
- Driver's team affiliation
- Qualifying position
- Gap to pole position (time difference between driver's best lap and pole lap)
- Time delta to the driver's teammate (if available)

Additionally, the JSON file includes season-level summary statistics for each driver, providing an overview of their average qualifying position, average gap to pole, and average gap to teammate across the entire season. The datacompletness field in the JSON helps track situations where a driver is missing data for certain performance metrics at particular events within a season.

Note: In some cases, a driver may not have complete data for a given event, such as when they did not participate in qualifying due to technical issues. In these instances, the relevant fields are set to `null` to maintain data consistency.

3 Future Work

Moving forward, the research offers opportunities to enhance the analysis by incorporating weather and track condition details. The original data source included this information but came with significant challenges, including timing inconsistencies with lap times and substantial information gaps that made it difficult to conduct any real analysis.

In finding a new data source, the project turned to a more lap-time-focused analysis. Future work will require identifying alternative data sources to collect comprehensive weather, car behavior, and track data, potentially differentiating between wet and dry track conditions. This approach would enable a more complete analysis of track performance, revealing the specific environments where drivers and teams perform most effectively.

The research could also be expanded through additional analytical approaches. An enhanced teammate comparison would provide valuable insights with side-by-side metric displays and performance visualizations. Additionally, tracking car development throughout the season would be a promising avenue of investigation. By documenting when significant technical modifications were made to the cars throughout the season, researchers could analyze how these changes impact performance. This combined with more track information would offer a dynamic view of team development and strategic adaptations.