Formula 1 Qualifying Analysis

Mary Virginia Nye

December 17, 2024

1 Introduction

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 Methods

The project methodology consists of three core components: data collection, data processing and transformation, and an interactive visualization

2.1 Data Collection

The data collection system centers around a custom F1DataFetcher class that uses the FastF1 API. This class implements comprehensive error handling. For example, when requesting data, the system attempts up to three retries with a 5-second delay between attempts, ensuring robust collection even with API instabilities or connection issues. The fetcher maintains a local cache to optimize performance and reduce API load. For each Formula 1 season, the system creates a separate CSV file containing qualifying session data including lap times from Q1, Q2, and Q3, driver information, team affiliations, and track conditions. This separated storage approach allows for targeted updates of specific seasons while maintaining data integrity across the full dataset. The system logs all collection activities, tracking successful retrievals and documenting any failed attempts for later investigation or reprocessing.

2.2 Data Processing

The data processing file transforms raw qualifying data through a series of specialized functions to make it useful in the visualization. The first transformation converts string-formatted lap times into standardized time delta objects, enabling precise mathematical comparisons accurate to milliseconds. For each qualifying session, the system calculates three key performance metrics: The gap to pole position calculation determines the time difference between each driver's fastest lap and the pole position time. This metric uses the best lap time from any qualifying segment (Q1, Q2, or Q3) when comparing against the pole time, accounting for cases where a driver's fastest lap might occur in an earlier session segment. The teammate comparison analysis examines relative performance within teams. The code identifies teammate pairs and calculates the time differential between their best laps when that data is available. If either teammate is missing data, the gap is recorded as NaN. This calculation accounts for cases where teammates might set their best times in different qualifying segments or if one teammate does not participate in qualifying.

The processed data is structured into a JSON format where each entry represents a driver's season. Each season entry contains the driver's basic information (name, team, year), an events array with individual qualifying session results (position, gap to pole, teammate gap), season-level statistics (averages for qualifying position, gap to pole, and teammate gap), and a data completeness metric indicating the proportion of events with complete teammate comparison data."

2.3 Visualization

The visualization framework uses Panel and HoloViews to create an interactive dashboard that presents the processed data in easy to understand metrics and a graph. The system implements a hierarchical display structure that organizes information by season and event. The primary interface provides driver selection functionality that triggers the generation of yearly performance timelines. Each year's visualization has several integrated elements: A position-tracking scatter plot serves as the central visualization component. This plot maps qualifying positions across all races, with connecting lines highlighting performance trends. The Y-axis inverts to match Formula 1 conventions, placing position 1 at the top of the scale.

The statistical summary section presents aggregated performance metrics including average qualifying position, mean gap to pole, and average teammate lap differences. These statistics update dynamically based on

the selected time period. An interactive race selector enables detailed metrics from specific events. When selected, this component displays comprehensive session information including: exact qualifying position, time gap to pole position, time gap between teammates. The visualization code implements sophisticated performance optimizations through selective data loading and efficient filtering mechanisms. The interface automatically adjusts to different screen sizes while maintaining data visibility and interaction capabilities.

3 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
- Event Name: 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.

4 Output

The interface presents a comprehensive view of each driver's qualifying career through a hierarchical organization by year. At the top level, users can select any driver from a dropdown menu, immediately generating their complete performance timeline. Each year's data is displayed in a dedicated card that combines summary statistics with a position visualization.

The central component of each yearly view is a position-tracking plot that maps qualifying positions across all races. This visualization employs a dual-layer approach: a semi-transparent red line connects consecutive results to highlight trends, while distinct points mark exact qualifying positions. Following Formula 1 conventions, the Y-axis is inverted with position 1 at the top, allowing for intuitive interpretation of performance improvements as upward movements in the graph. Like the example below, gaps are shown when the driver did not compete in qualifying.

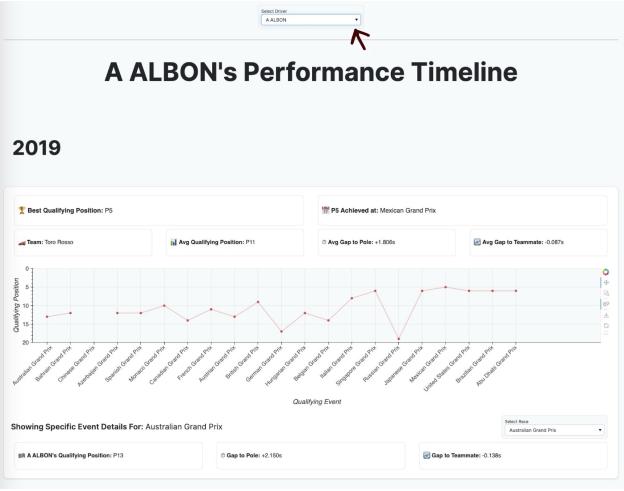

Figure 1: Dashboard example with driver selection

Above the graph, key performance metrics are displayed:

- Best qualifying position and the specific races where it was achieved
- Team affiliation
- Average qualifying position
- Mean gap to pole position in seconds
- Average time difference to teammates

At the most detailed level, users can select specific races to examine particular qualifying sessions in depth. The system displays the exact qualifying position, gap to pole position, and time difference to teammates for the selected event.

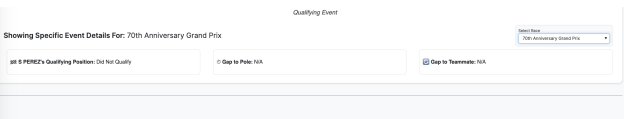

Figure 2: Results for driver that did not compete in event

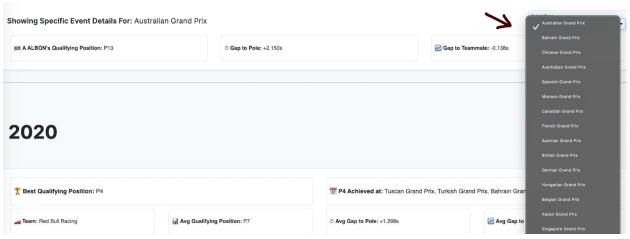

Figure 3: Results for driver that competed in event

The visualization system effectively captures both continuity and change in driver careers. Lando Norris's timeline provides a unique opportunity to analyze long-term progression, as his consistent team affiliation with McLaren from 2019 to 2024 allows observers to track both driver development and team advancement without the variable of team changes. His performance shows remarkable improvement over this period, progressing from an average qualifying position of P9 in 2019, to regularly achieving front-row starts in 2024, including multiple pole positions. This evolution is particularly evident in his qualifying pace relative to pole position, improving from an average gap of +1.573 seconds in 2019 to just +0.332 seconds in 2024. The improvement in qualifying performance is evident through both the position-tracking graph and the summary statistics.

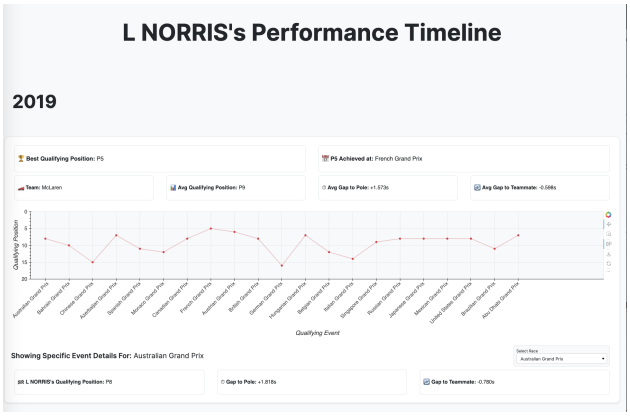

Figure 4: L Norris's 2019 view

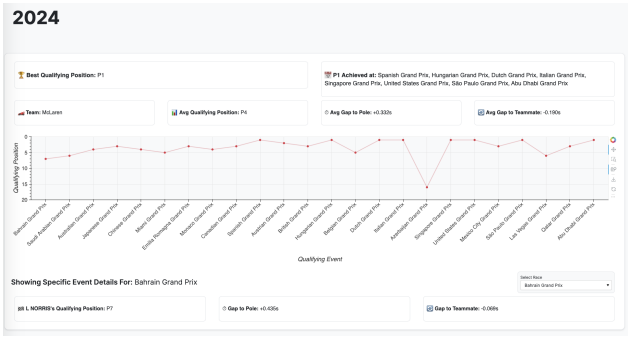

Figure 5: L Norris's 2024 view

5 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.