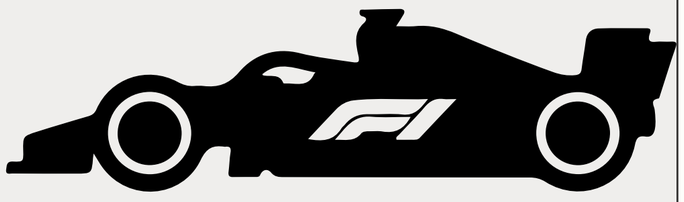
Analyzing Formula 1 Race Performance: Strategy, Speed, and Consistency

YUQI GE



PROJECT INTRODUCTION

My project analyzes Formula 1 race performance from 2022 to 2024, focusing on how driver pace, pit stop strategy, and qualifying position influence final race outcomes.

Multiple data sources (Kaggle + OpenF1 API) are combined to study patterns and correlations

As an Formula 1 Fan, the correlation between the variables in races are always interesting to me

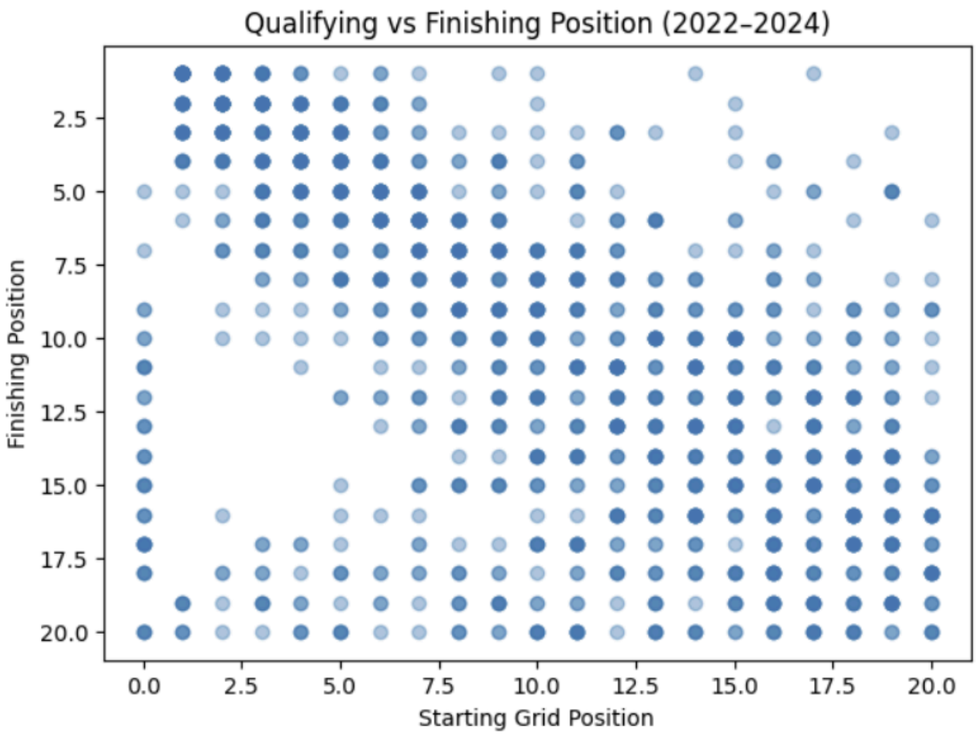
The goals: Does starting position influence final race outcome?

- Which teams consistently perform fastest in pit stops?
- Which drivers show the most consistent lap performance?

DATA SOURCES

Data Source	Data Format	Process Approach	Data Size Processed
Kaggle F1 World Championship Dataset	CSV (drivers, constructors, laps, qualifying)	Imported via pandas, filtered for 2022–2024 seasons; merged race results, lap times, and driver info.	Imported via pandas, filtered for 2022–2024 seasons; merged race results, lap times, and driver info.
F1 Pit Stop Dataset (Kaggle)	CSV (pit stops, tire compounds, stint duration, timing)	Loaded with pandas; merged pit stop timing with race results to analyze strategy impact.	~30,000+ rows (pit stop entries from 2012–2024, filtered to 2022-2024)
OpenF1 API (telemetry, sessions, drivers)	JSON (API responses)	Queried via Python requests; dynamically retrieved session keys, driver data, lap info for selected races.	Dynamic per session (~1,000–10,000 rows per race; multiple sessions queried)

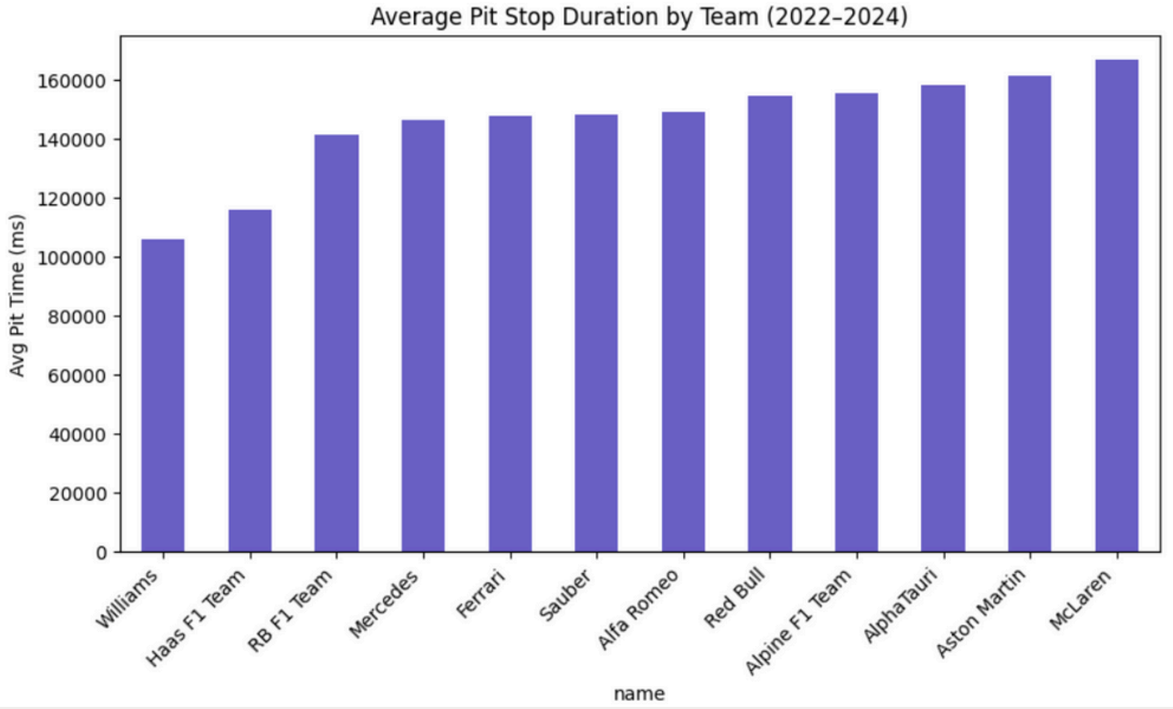
RESULTS



grid = The starting grid position of the driver in the race, Integer (1–20)
Lower is better (e.g. 1 = pole position)

- There is a strong positive correlation between qualifying position and final race result from 2022 to 2024.
- Drivers starting in the top 5 consistently finish within the top 6–7 positions.
- Mid-grid starters (P8–P14) show wide variation, indicating that strategy and race events have a larger influence.
- Backmarkers (P15+) rarely finish in the points unless there is a safety car or high attrition

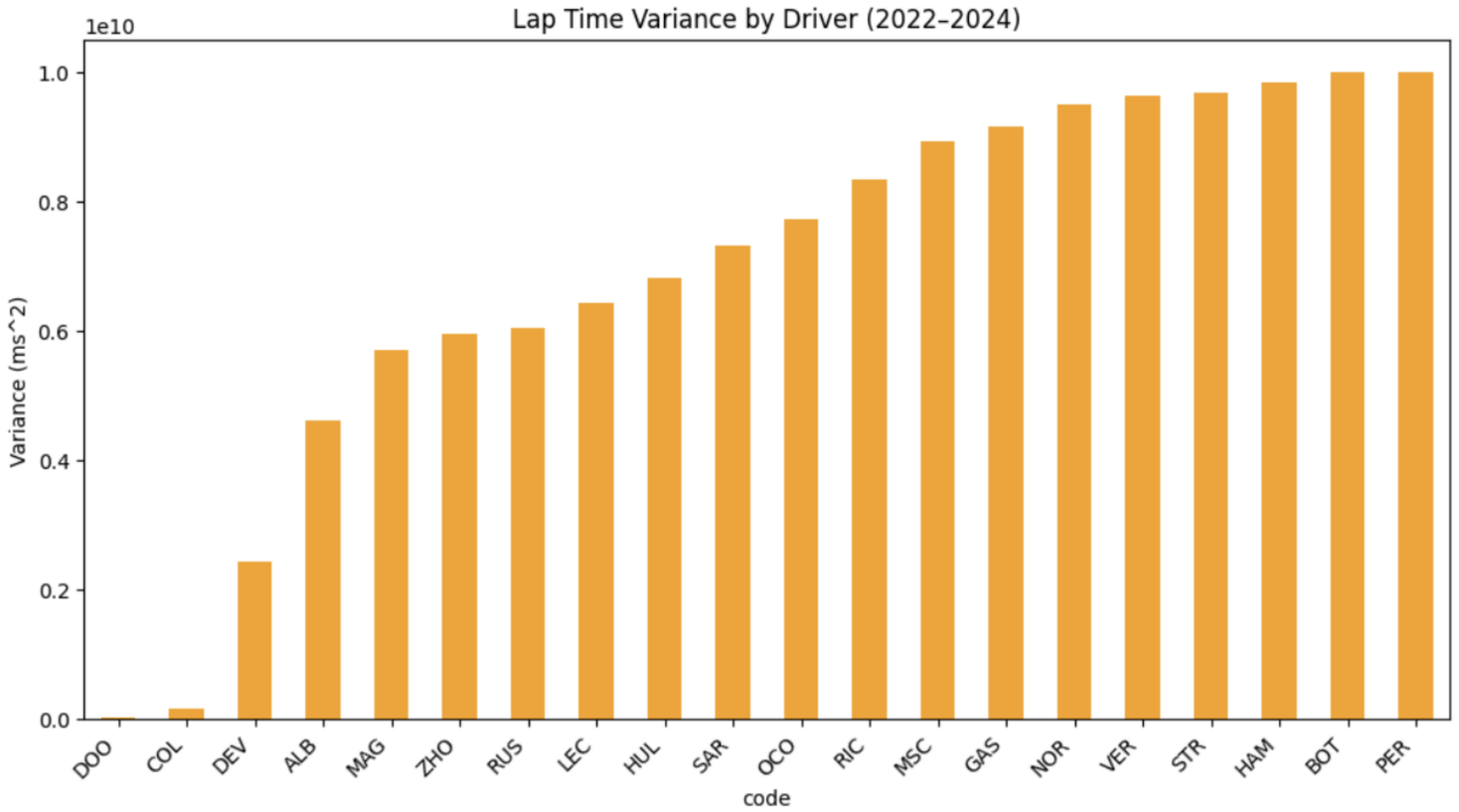
avg_pit_ms = Average pit stop time for a driver/team across races in Milliseconds
Aggregated per team or driver; lower is better



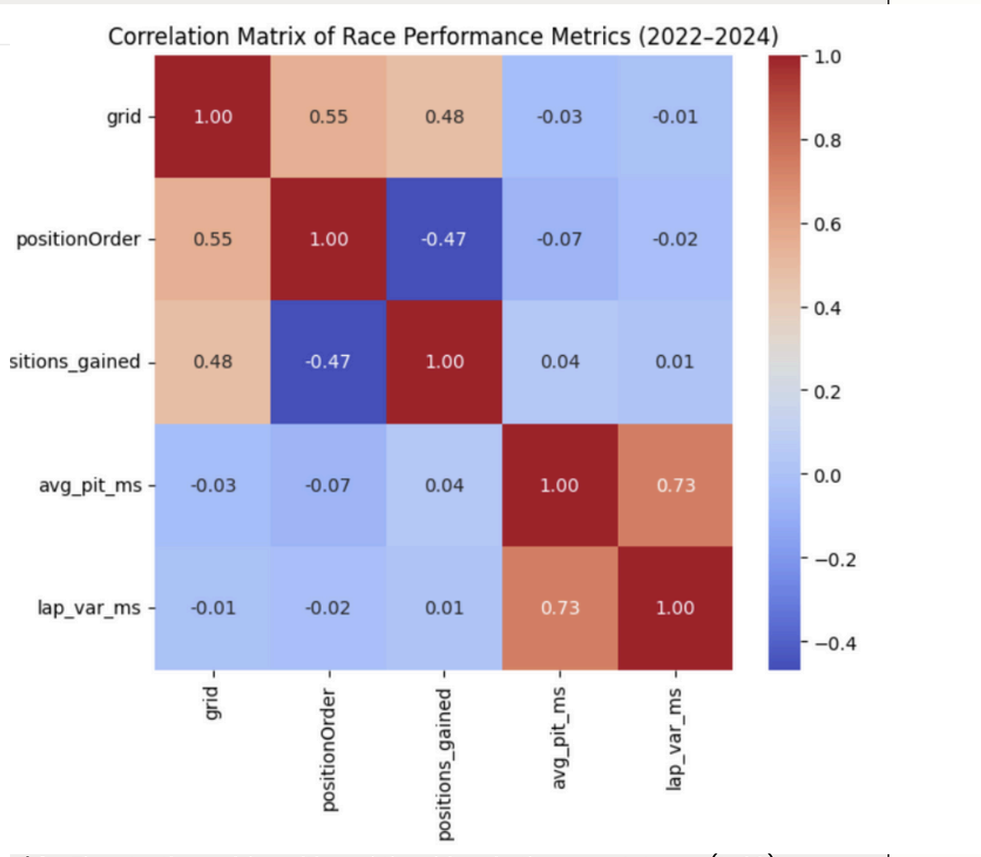
- Teams with the fastest pit stops (Red Bull, McLaren) consistently gain a competitive advantage.
- Teams like Williams / Haas show slower average stops (+0.6–1.2 seconds slower), often leading to position losses.
- Across 2022–2024, pit stop performance remained highly stable for top teams, showing strong operational consistency.
- Variability in pit stop time directly correlates with in-race undercut / overcut effectiveness.
- Some teams improved over time (e.g., Ferrari in 2024 showing fewer errors).

Valtteri Bottas
PER
Sergio Pérez

RESULTS

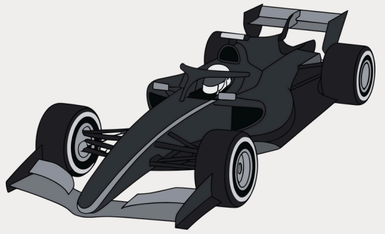


DOO = Jack Doohan, COL = Logan Sargeant, DEV = Nyck de Vries, ALB = Alexander Albon, MAG = Kevin Magnussen, ZHO = Zhou Guanyu, RUS = George Russell, LEC = Charles Leclerc, HUL = Nico Hülkenberg, SAR = Felipe Drugovich, OCO = Esteban Ocon, RIC = Daniel Ricciardo, MSC = Mick Schumacher, GAS = Pierre Gasly, NOR = Lando Norris, VER = Max Verstappen, STR = Lance Stroll, HAM = Lewis Hamilton, BOT = Valtteri Bottas, PER = Sergio Pérez



grid = The starting grid position of the driver in the race, Integer (1–20)
Lower is better (e.g. 1 = pole position)
position Order = Final race position after the race ends, Integer (1–20)
Used to assess the race result outcome
positions_gained = Number of positions gained from start to finish (grid - positionOrder)
Integer (positive or negative)
Positive = gained places; **Negative** = lost positions
avg_pit_ms = Average pit stop time for a driver/team across races in Milliseconds
Aggregated per team or driver; lower is better
lap_var_ms = Lap time variance for a driver over all laps, Milliseconds squared
Lower values indicate higher driving consistency

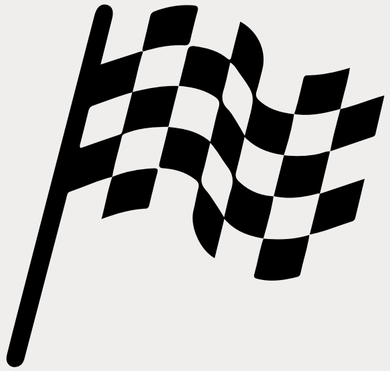
RESULTS ANALYSIS & CHALLENGES



- Drivers with more consistent lap times tend to finish higher.
- High lap-time variance usually means tire degradation or race instability.
- Grid position has the strongest correlation with the final result.
- Slower pit stops and higher lap variance both correlate with poorer finishing positions.
- Pace consistency + operational execution play a key role in race outcomes.

Challenges:

- Merging inconsistent IDs from different data sources
- Handling outliers in pit stop and lap time data
- Managing API rate limits and incomplete records



THANK YOU

QUESTIONS?