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.

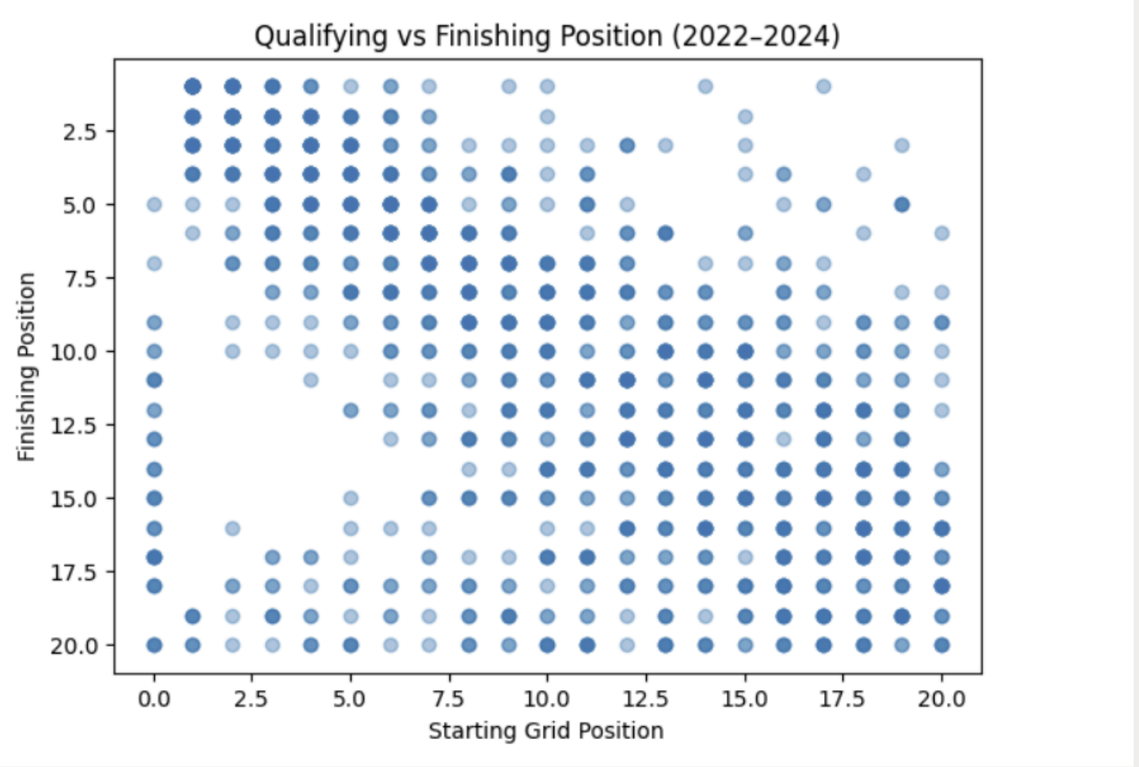
Multiple data sources (Kaggle + OpenF1 API) are combined to study patterns in lap times, pit stop efficiency, and driver consistency.

The goal is to identify the key factors that contribute to race results and compare performance trends across teams and drivers with visualization and data analysis.

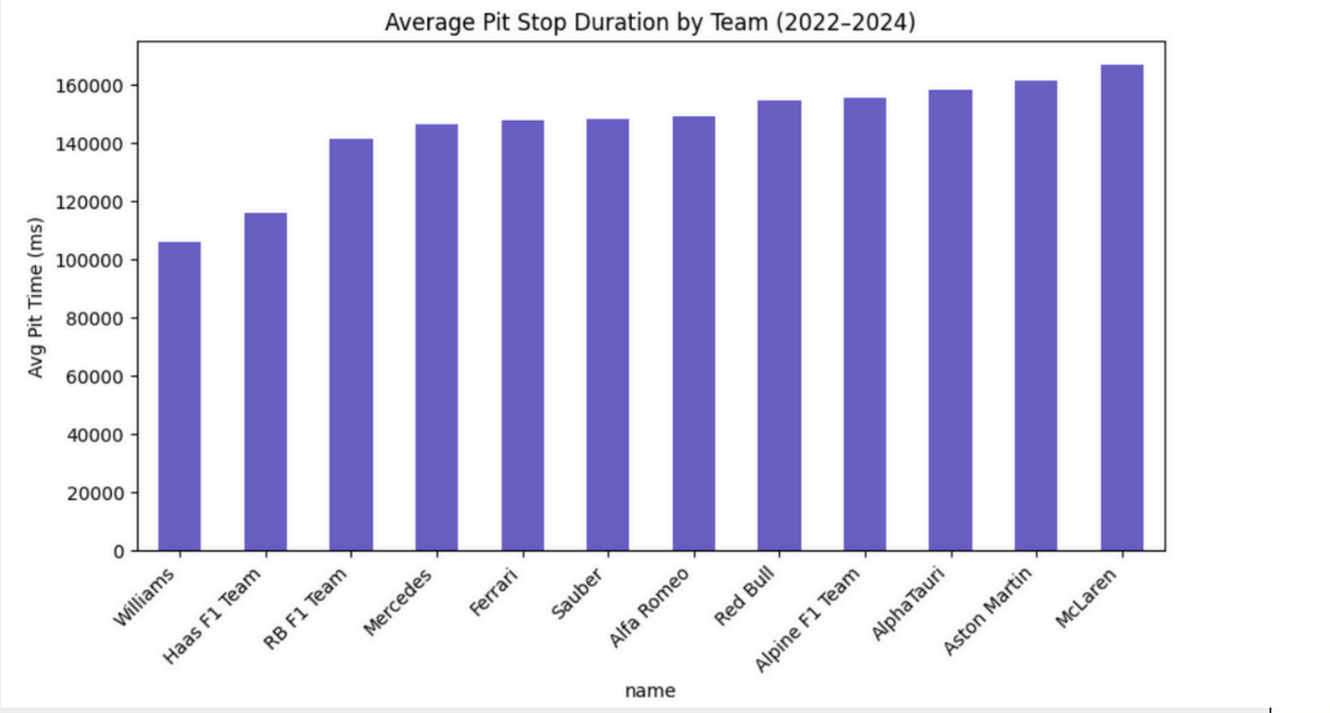
DATA SOURCES

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

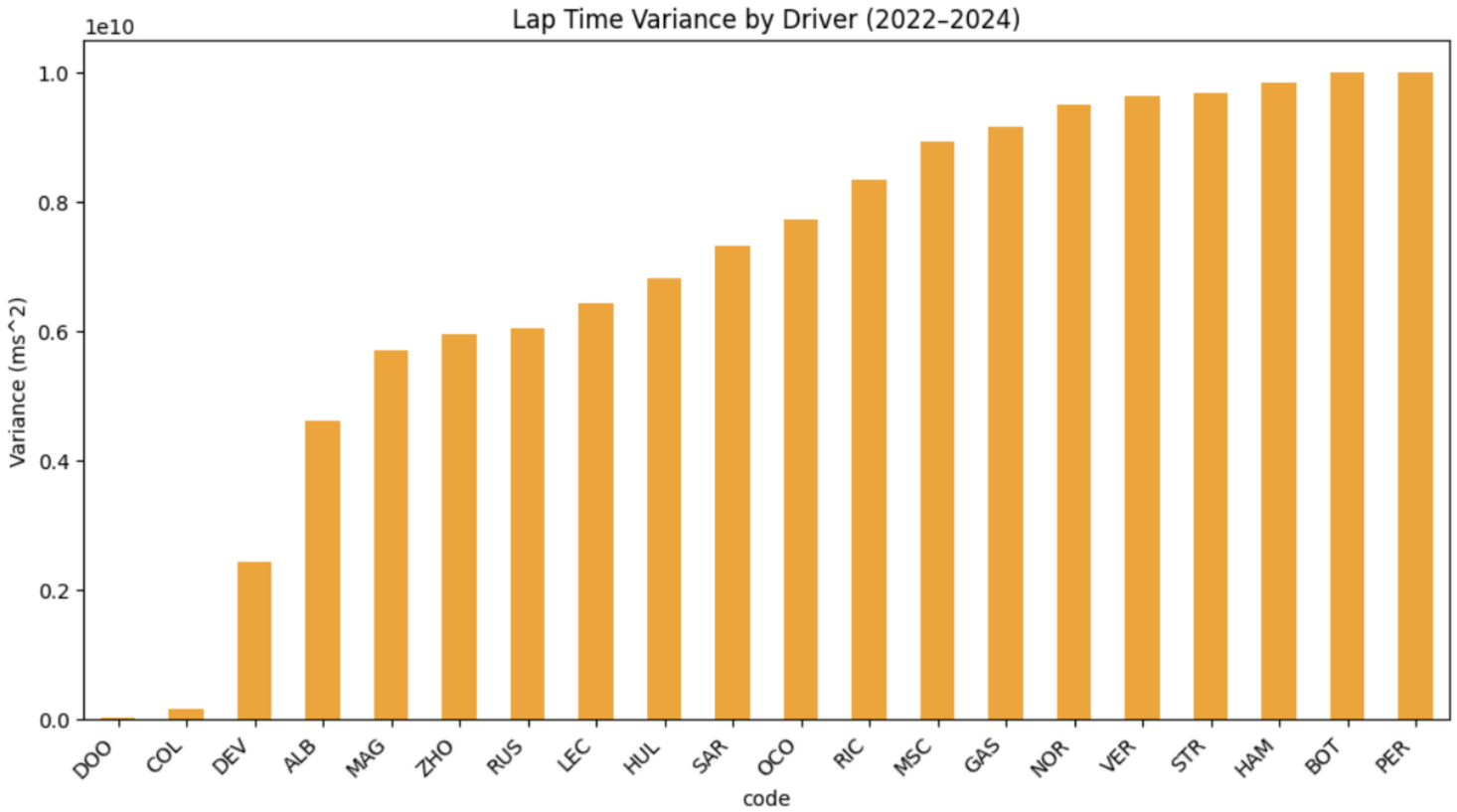


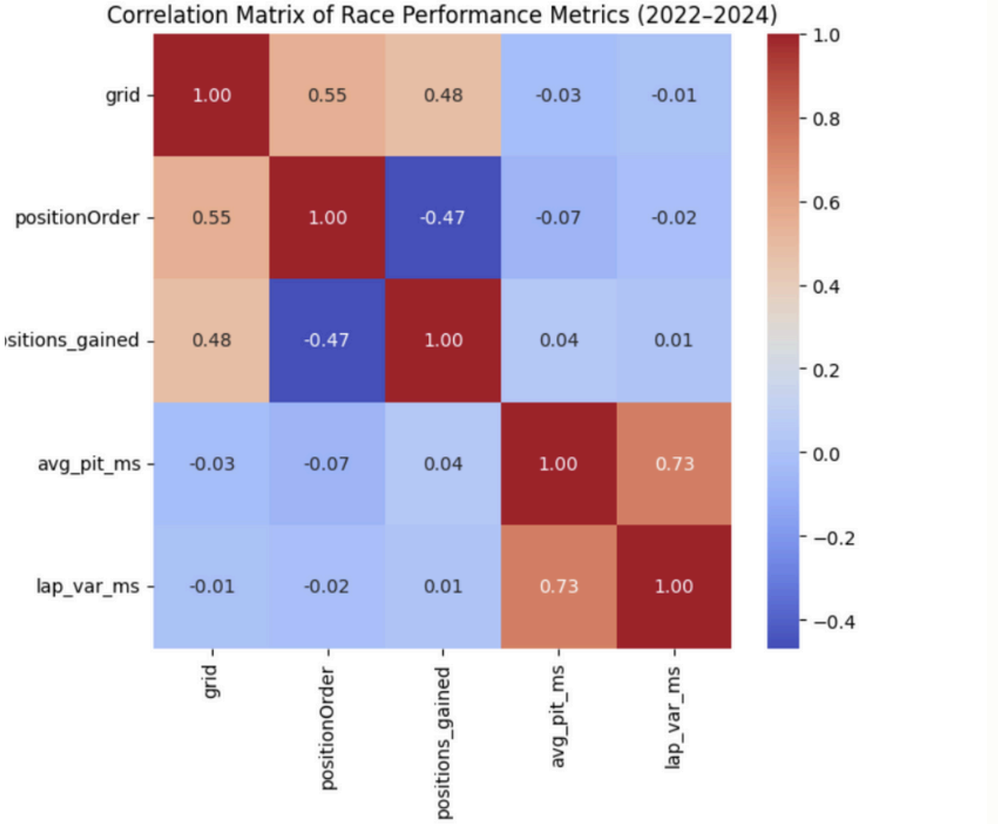
- There is a strong positive correlation between qualifying position and final race result from 2022 to 2024.
- Drivers starting in the top 5 consistently finish within the top 6–7 positions.
- Mid-grid starters (P8–P14) show wide variation, indicating that strategy and race events have a larger influence.
- Backmarkers (P15+) rarely finish in the points unless there is a safety car or high attrition.
- 2023 shows the strongest grid-to-finish correlation due to Red Bull dominance.



- 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).

RESULTS


Lap Time Variance by Driver (2022–2024)
1e10
Variance (ms^2)
1.0
0.8
0.6
0.4
0.2
0.0
DOO
COL
DEV
ALB
MAG
ZHO
RUS
LEC
HUL
SAR
OCO
RIC
MSC
GAS
NOR
VER
STR
HAM
BOT
PER
code


Correlation Matrix of Race Performance Metrics (2022–2024)
grid
positionOrder
sitions_gained
avg_pit_ms
lap_var_ms
1.00 0.55 0.48 -0.03 -0.01
0.55 1.00 -0.47 -0.07 -0.02
0.48 -0.47 1.00 0.04 0.01
-0.03 -0.07 0.04 1.00 0.73
-0.01 -0.02 0.01 0.73 1.00
grid
positionOrder
positions_gained
avg_pit_ms
lap_var_ms
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

RESULTS ANALYSIS

- 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.

Used class methods: data cleaning, merging, groupby analysis, and visualization (Matplotlib/Seaborn).

Performed lap time variance, groupby operations, feature engineering, heatmaps, and correlation analysis across 2022–2024.

Integrated OpenF1 API using the techniques taught in class (requests + JSON parsing).

THANK YOU

QUESTIONS?