Progress report

Project scope update

My data respiratory name is "Data-analysis-for-Formula-1-Race-Performance"
The link to my repository:
https://github.com/JeffGe911/Data-analysis-for-Formula-1-Race-Performance-
The project scope has not changed. I will analyze Formula 1 races between 2022 and 2024 to understand how pit stop strategies, driver performance, and starting grid positions affect race outcomes. I plan to visualize these patterns and look for performance trends across teams and drivers.

Data sources
I am using three main datasets:

Kaggle F1 World Championship Dataset – downloaded successfully; includes race results, lap times, driver/team info, and qualifying data.

F1 Pit Stop Dataset (Kaggle) – also downloaded; includes pit stop timing, tire info, stint details, and weather.

OpenF1 API –I switched to this data. It is a public REST API that provides detailed Formula 1 race data, including drivers, telemetry, race sessions, and event information. I tested this API using Python (requests) and successfully retrieved JSON data for a specific race session (2023 F1 Bahrain race). This API will be used to collect telemetry and driver performance metrics for the seasons I analyze.

API used

The Open F1 API provides structured F1 data from 1950 to present. I tested it using Python and successfully pulled data for the 2023 Round 1 race for the test. I plan to use this API to gather race results and standings across multiple years to support my analysis.
 Issues/difficulties

 I first tried to use the Ergast API, but it couldn't pass the test.py, and I kept fixing the code and I found out that this site blocked me, and I was unable to access it. Maybe it's due to my network, but I had to change to another source. To solve this issue, I switched to the OpenF1 API, which works consistently and provides even more detailed Formula 1 data. Possible challenges I expect moving forward include merging data from different sources, and I need to unite the data and set the data granularity. Another one would be cleaning inconsistent fields (such as driver IDs or session formats) because there are some changes in the driver roster, and practice sessions are different from sprint sessions and racing sessions.