

F1 Performance Analysis Report

1. Data Collection and Preparation

The foundation of this analysis began with retrieving historical Formula 1 datasets from a publicly available database through Kaggle. This data encompassed tables like Race Results, Lap Timings, Qualifying Results, Race Schedule, Constructor Standings, and others; each offering a unique perspective on race performance, timing, and team rankings.

Before importing into SQLite, the raw data underwent a necessary cleansing and normalization process. This included handling missing values, standardizing team and circuit names to prevent redundancy, and ensuring consistency across foreign keys. Time fields, often stored in text formats (like "MM:SS.mmm"), were particularly important to convert into numeric types for accurate analysis. These transformations ensured the data was reliable and compatible with SQL-based computations. After the data was clean, queries were designed and run using SQLite to uncover performance trends, comparative metrics, and patterns within the sport.

2. Overview of Key Components

This analysis centered around a structured examination of Formula 1 data, focusing particularly on drivers Lewis Hamilton and Max Verstappen. Utilizing data from several tables, including Race Results, Lap Timings, Qualifying Results, Race Schedule, and more, to draw insights across multiple dimensions: individual driver performance, circuit characteristics, qualifying impact, and team outcomes.

The core analysis involved determining how many races each driver competed in, their best lap times per track, race victories, pole conversion rate (qualifying position 1 vs race result position 1), and cumulative career points. Average lap times were examined by year to uncover long-term trends. At the circuit level, we evaluated average, fastest, and slowest lap times, as well as how often pole position translated into victory. Finally, on the team side, constructor rankings by season and qualifying consistencies (e.g., Q3 appearances) were analyzed to assess team performance over time.

3. Benefits of This Analysis

There are numerous advantages to conducting this kind of multi-layered performance analysis. First, it provides a comprehensive view of driver and team efficiency over time,

helping stakeholders understand not just who is winning, but how and why. For team strategists and engineers, the data can inform setup decisions and driver pairing strategies. For broadcasters and fans, the context and comparisons add narrative depth and excitement to race-day commentary and for someone like me, new to F1 wanting to know more about the sport.

Additionally, long-term trend analysis helps identify shifts in competitive balance, track evolution, and rule impact, allowing F1 teams and the FIA to make more informed decisions about regulations, calendar design, and vehicle development.

4. Challenges Faced

Despite the richness of the dataset, several challenges had to be addressed. A major hurdle was dealing with time data stored in string formats (MM:SS), which required parsing and converting into numerical values for any meaningful arithmetic comparison. This increased the complexity of the SQL queries and the likelihood of calculation errors.

Moreover, the dataset had missing/NULL entries for some circuits, drivers, or seasons—requiring careful data cleansing and the application of conditional filters to ensure the accuracy of computed metrics. Variations in naming conventions (such as team aliases or track names) also demanded standardization before meaningful aggregations could be performed.

5. Conclusions and Insights

This comprehensive SQL-based analysis revealed several layers of insight into Formula 1 dynamics. Hamilton and Verstappen, while both dominant in their own eras, show different strengths—Hamilton's legacy is built on consistency and longevity, while Verstappen's recent seasons demonstrate aggressive domination in both qualifying and race finishes.

Track analysis unearthed how some circuits consistently produce closer finishes, while others reward pole position more heavily. On the team level, it became evident that a few constructors have maintained dominance across decades, while others spike in performance during specific eras.

Overall, this analysis exemplifies how well-structured data combined with thoughtful querying can uncover hidden patterns and deepen our understanding of high-performance sport. As data continues to play an increasingly important role in motorsports, such insights will only become more valuable for teams, viewers, and analysts alike.