Comprehensive Notes on Formula 1 Data Analysis Project

Overview

This project involved analyzing the Formula 1 dataset spanning from 1950 to 2024. The goal was to answer 10 business questions using SQL queries while ensuring the data was clean and structured for efficient analysis. A variety of insights regarding drivers, constructors, circuits, and race trends were uncovered.

Data Cleaning and Preparation

1. Null values were removed to maintain data integrity.
2. Column names were standardized for consistency across tables.
3. The dataset was imported into MySQL for structured querying.
4. Relationships between tables (e.g., drivers and results, constructors and standings) were validated to ensure data accuracy.
5. Assumed 'position = 1' in the results table indicates a race win.

SQL Queries

The following SQL queries were executed to answer the business questions:

1. Identified the top 10 drivers with the most race wins.
2. Determined the constructor with the most race wins.
3. Listed the top 5 circuits with the highest number of races hosted.
4. Analyzed the number of races per year to understand Formula 1's growth over time.
5. Found the youngest drivers to win a race, highlighting early achievers.

6. Analyzed fastest lap times in sprint races to identify exceptional performance.

7. Identified the driver with the most race participations, reflecting longevity.

8. Listed constructors with the most championship wins to showcase consistent dominance.

9. Explored the average number of laps completed across different circuits.

10. Highlighted drivers with the most pole positions in qualifying.

****Python Code Integration****

To automate the process:

1. Python was used to upload CSV files into MySQL tables.

2. Queries were executed within Python, and results were fetched for analysis.

3. A well-documented Python script was created to support repeatable analysis.

****Presentation****

A PowerPoint presentation was prepared with slides covering:

1. An introduction to the dataset and objectives.

2. Data cleaning and assumptions made during preparation.

3. One slide per business question, featuring SQL queries, screenshots, and insights.

****Deliverables****

1. A Python script for SQL upload and analysis.

2. SQL queries for answering the 10 business questions.

3. A PowerPoint presentation with all findings.

4. This PDF document summarizing the entire workflow and results.

****Conclusion****

This project provides a comprehensive look at Formula 1's rich history through data. By combining SQL and Python,
we gained insights into the sport's drivers, teams, circuits, and evolution over time. The results demonstrate the power
of data analysis in uncovering trends and answering meaningful business questions.