21378681

Ibrahim Khan

Final Assignment | ISYS2014

MARCH 30 2025



**DATABASE SYSTEM**

**REPORT**

Contents

[INTRODUCTION 3](#_Toc17726)

[DATABASE DESIGN 4](#_Toc222)

[Relationship Diagrams 4](#_Toc16018)

[Rational Schema 5](#_Toc3627)

[ER Model 6](#_Toc13811)

[IMPLEMENTATION OF THE DATABASE AND ADDING SAMPLE DATA 9](#_Toc10449)

[USE OF THE DATABASE 13](#_Toc18369)

[Advanced Queries 13](#_Toc8648)

[Simple Queries 14](#_Toc22239)

[Advanced Features (Stored Procedure And View) 15](#_Toc30466)

[Database Connectivity and Python Implementation 18](#_Toc32146)

[DISCUSSION 23](#_Toc1384)

# INTRODUCTION

I began by outlining a **relational schema**, a blueprint for creating an **Entity-Relationship (ER) model** for the **Formula 1 database**. As a result of this step, I was able to organize elements like **championships, Grand Prix races, circuits, teams, drivers, and race results** into entities and defined suitable relationships to encapsulate their interactions. In order to ensure smooth connectivity between the entities, I defined **cardinality and participation constraints**, which were applied to both entities and relationships. This structured approach helped establish an **efficient data management system**, ensuring the database is well-optimized for handling Formula 1 race data.

# DATABASE DESIGN

I selected **six main entities** to accurately model the **Formula 1 racing structure**:

| ****Entity**** | ****Purpose**** |
| --- | --- |
| **Championship** | Represents each **Formula 1 season** (e.g., 2025, 2026). |
| **GrandPrix** | Stores details about each **race event** (e.g., Monaco GP, British GP). |
| **Circuit** | Defines the **track locations** where Grand Prix races take place. |
| **Team** | Represents **Formula 1 racing teams** (e.g., Red Bull, Ferrari, Mercedes). |
| **Driver** | Stores information about **Formula 1 drivers**, including their teams. |
| **RaceResult** | Stores **race results**, such as **driver position, points, and fastest lap time**. |

These entities **cover all essential aspects** of Formula 1, allowing users to store and retrieve **meaningful data** about races, drivers, and team performance.

## Relationship Diagrams

Relationships ensure that the **entities are connected logically**, reflecting how **Formula 1 operates**.

| ****Relationship**** | ****Between Which Entities**** | ****Purpose**** |
| --- | --- | --- |
| **Includes** | Championship, GrandPrix | Each **championship** consists of **multiple races**. |
| **HeldAt** | GrandPrix, Circuit | Each **Grand Prix is hosted at a single circuit**. |
| **ParticipatesIn** | Team, Championship | Each **team competes in a championship**. |
| **BelongsTo** | Driver, Team | Each **driver belongs to one team**. |
| **CompetesIn** | Driver, GrandPrix | Drivers **participate in multiple Grand Prix races**. |
| **HasResult** | RaceResult, Driver | Links **race results** to specific **drivers and races**. |

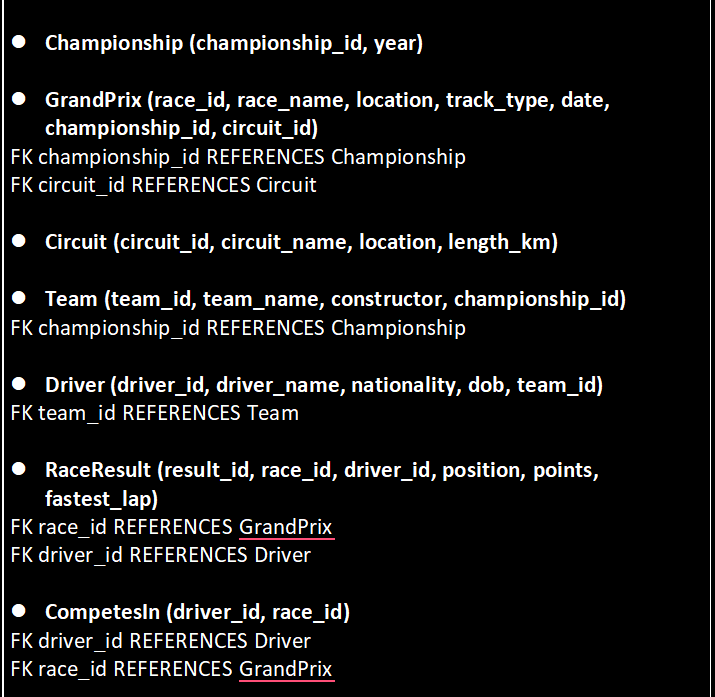
****Entities, Keys, and Attributes****

| ****Entity**** | ****Key**** | ****Attributes**** |
| --- | --- | --- |
| **Championship** | championship\_id | year |
| **GrandPrix** | race\_id | race\_name, location, track\_type, date, championship\_id |
| **Circuit** | circuit\_id | circuit\_name, location, length\_km |
| **Team** | team\_id | team\_name, constructor, championship\_id |
| **Driver** | driver\_id | driver\_name, nationality, dob, team\_id |
| **RaceResult** | result\_id | race\_id, driver\_id, position, points, fastest\_lap |

**Relationships, Cardinality, and Participation Constraints**

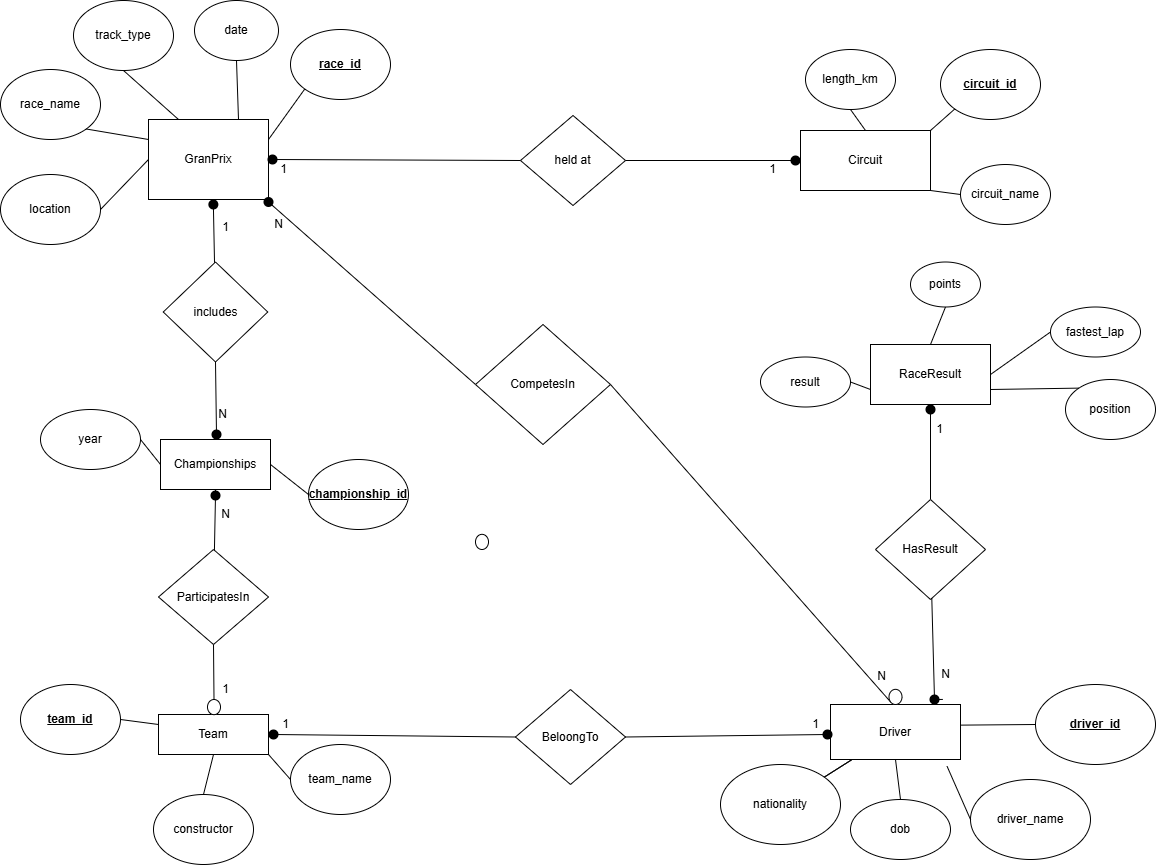
| ****Relationship**** | ****Cardinality Constraint**** | ****Participation Constraint**** |
| --- | --- | --- |
| **Includes** | One to Many (One Championship includes many Grand Prix, each Grand Prix belongs to one Championship) | **Total** = GrandPrix, **Total** = Championship |
| **HeldAt** | One to One (Each Grand Prix is held at one Circuit, and each Circuit can host only one Grand Prix in a season) | **Total** = GrandPrix, **Total** = Circuit |
| **ParticipatesIn** | One to Many (Many Teams participate in one Championship, each Championship has multiple Teams) | **Total** = Team, **Total** = Championship |
| **BelongsTo** | One to Many (One Team has multiple Drivers, but each Driver belongs to only one Team) | **Total** = Driver, **Total** = Team |
| **CompetesIn** | Many to Many (Many Drivers can compete in many Grand Prix races, and each Grand Prix has multiple participating Drivers) | **Total** = GrandPrix, **Total** = Driver |
| **HasResult** | One to Many (One Driver has multiple race results, but each RaceResult belongs to one Driver) | **Total** = RaceResult, **Partial** = Driver |

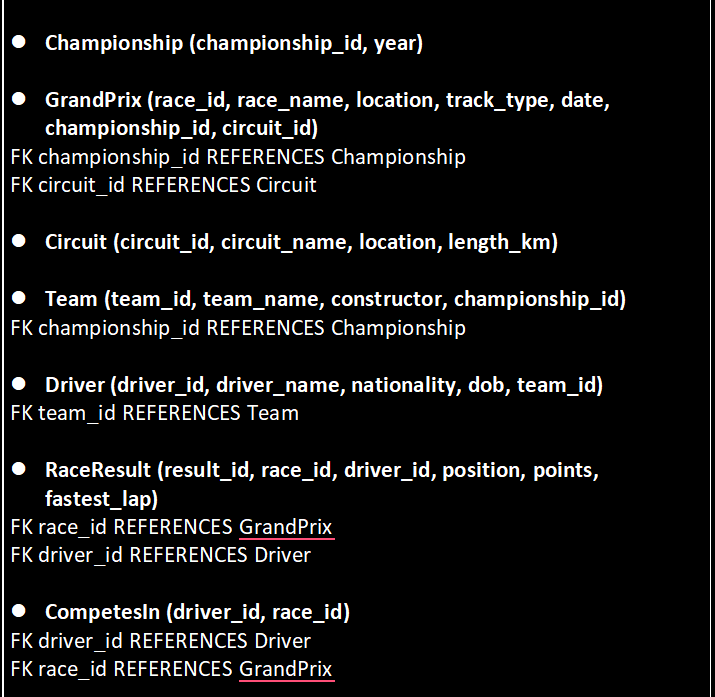
## **Rational Schema**

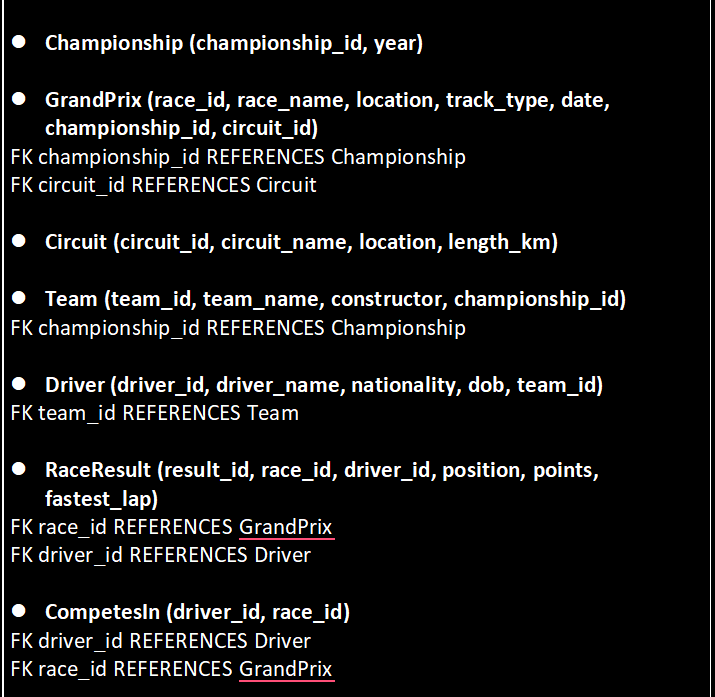
****

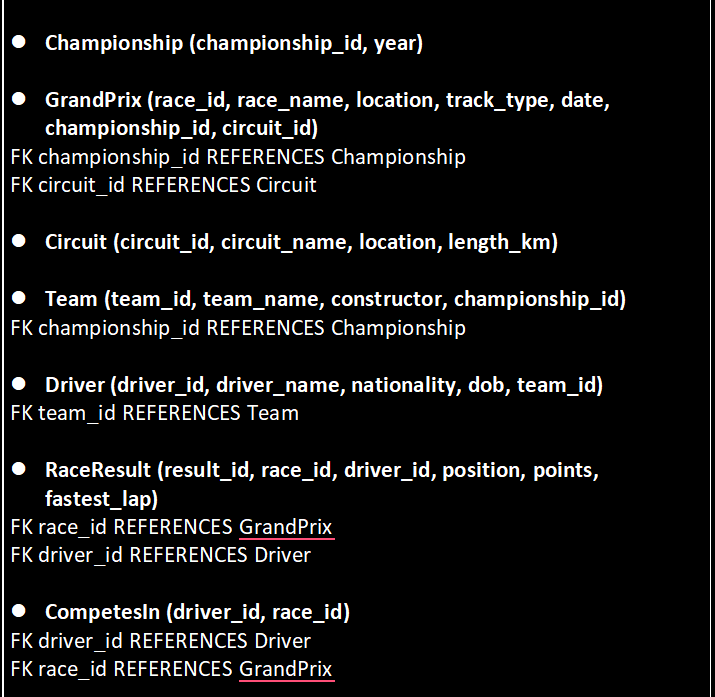
**CompetesIn created due to Many-Many**

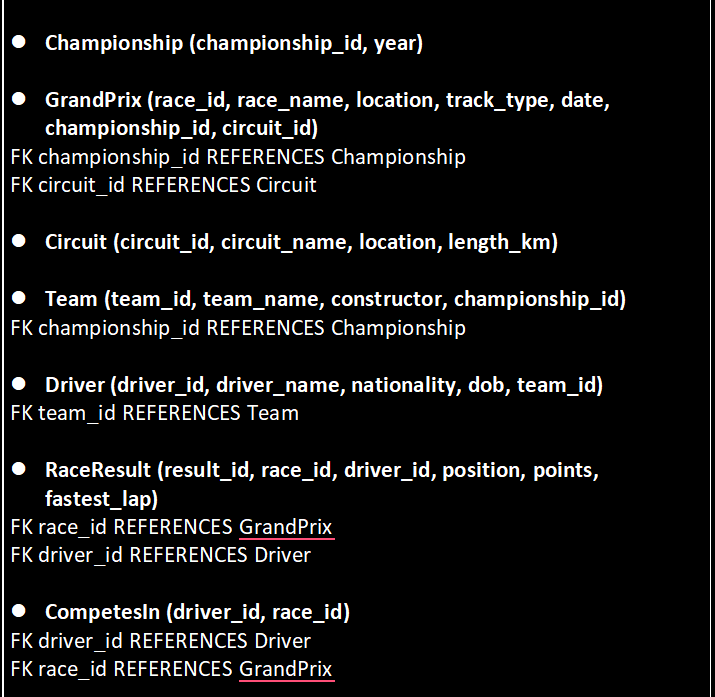
## ER Model





 TOTAL PARTICIPATION





PARTIAL PARTICIPATION

### ****DATA DESCRIPTION****

I selected **data types** that ensure **data accuracy, storage efficiency, and query optimization**.

| ****Attribute**** | ****Data Type**** | ****Reason**** |
| --- | --- | --- |
| INT | Used for **IDs** and **foreign keys** | Ensures **unique identification** of records. |
| VARCHAR(x) | Used for **names, locations, and track types** | Allows storing **variable-length text**. |
| DATE | Used for **race dates and driver birthdates** | Ensures **accurate date handling**. |
| DECIMAL(5,2) | Used for **race points and circuit lengths** | Allows **precision in numerical values**. |
| TIME | Used for **fastest lap records** | Ensures **proper time calculations**. |

### ****GrandPrix****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| race\_id | INT | Unique ID for each Grand Prix | **Primary Key** |
| race\_name | VARCHAR(40) | Name of the Grand Prix |  |
| location | VARCHAR(60) | Location of the race |  |
| track\_type | VARCHAR(30) | Type of circuit (Street, Road, etc.) |  |
| date | DATE | Date of the race |  |
| championship\_id | INT | References Championship table | **Foreign Key** |
| circuit\_id | INT | References Circuit table | **Foreign Key** |

### ****Circuit****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| circuit\_id | INT | Unique ID for each Circuit | **Primary Key** |
| circuit\_name | VARCHAR(50) | Name of the Circuit |  |
| location | VARCHAR(60) | Location of the Circuit |  |
| length\_km | DECIMAL(5,2) | Circuit length in km | CHECK (length\_km > 0) |

### ****Championship****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| championship\_id | INT | Unique ID for each Championship | **Primary Key** |
| year | YEAR | Year of the Championship |  |

### ****Team****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| team\_id | INT | Unique ID for each Team | **Primary Key** |
| team\_name | VARCHAR(50) | Name of the Team | **UNIQUE** |
| constructor | VARCHAR(50) | Name of the Constructor |  |
| championship\_id | INT | References Championship table | **Foreign Key** |

### ****Driver****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| driver\_id | INT | Unique ID for each Driver | **Primary Key** |
| driver\_name | VARCHAR(50) | Name of the Driver |  |
| nationality | VARCHAR(50) | Nationality of the Driver |  |
| dob | DATE | Date of Birth of Driver | CHECK (dob < CURDATE()) |
| team\_id | INT | References Team table | **Foreign Key** |

### ****RaceResult****

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| result\_id | INT | Unique ID for each race result | **Primary Key** |
| race\_id | INT | References GrandPrix table | **Foreign Key** |
| driver\_id | INT | References Driver table | **Foreign Key** |
| position | INT | Finishing position | CHECK (position > 0) |
| points | DECIMAL(5,2) | Points earned in the race | CHECK (points >= 0) |
| fastest\_lap | TIME | Fastest lap time in the race |  |

### ****CompetesIn**** (From GrandPrix and Driver)

| **Attribute** | **Data Type** | **Description** | **Constraints** |
| --- | --- | --- | --- |
| driver\_id | INT | References Driver table | **Foreign Key** |
| race\_id | INT | References GrandPrix table | **Foreign Key** |

IMPLEMENTATION OF THE DATABASE AND ADDING SAMPLE DATA

**Database Creation:**

A MySQL database named khan\_21378681 was created.

**Table Creation:**

Entities such as Championship, GrandPrix, Circuit, Team, Driver, RaceResult, and CompetesIn were designed with primary and foreign keys to ensure referential integrity.

Appropriate data types (INT, VARCHAR, DATE, DECIMAL, TIME) were used for optimal storage and query performance.

**Constraints Applied:**

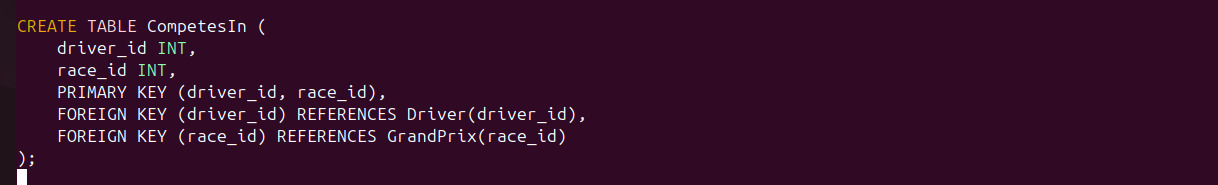
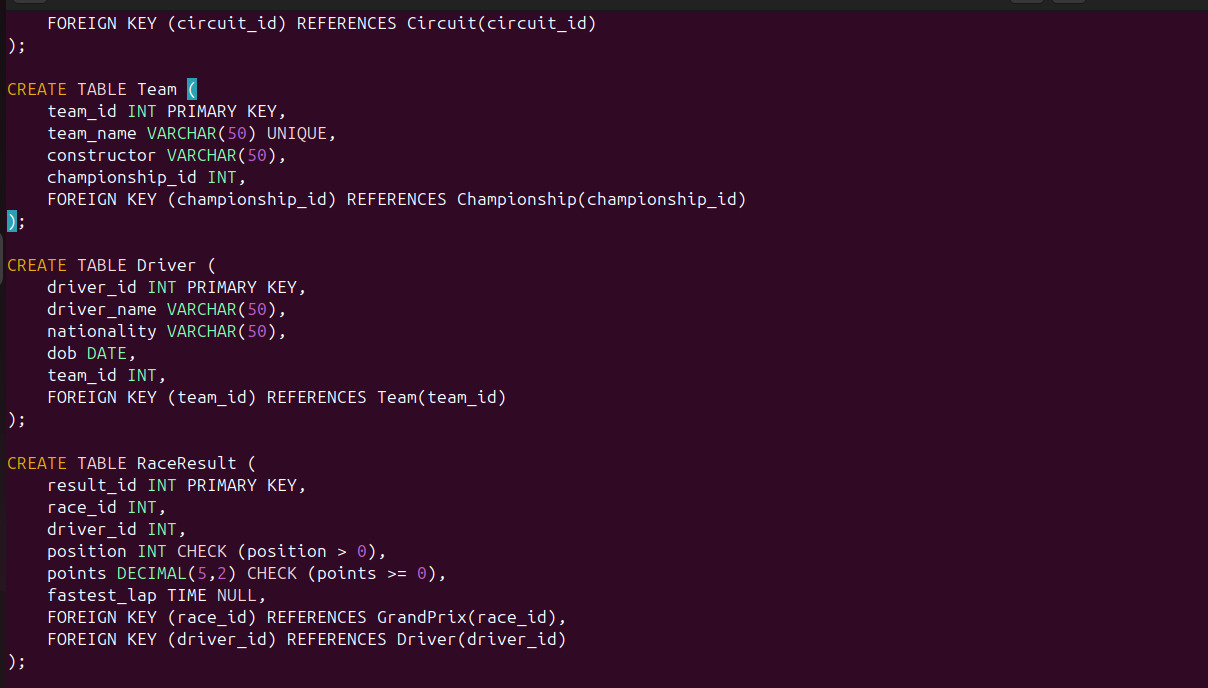
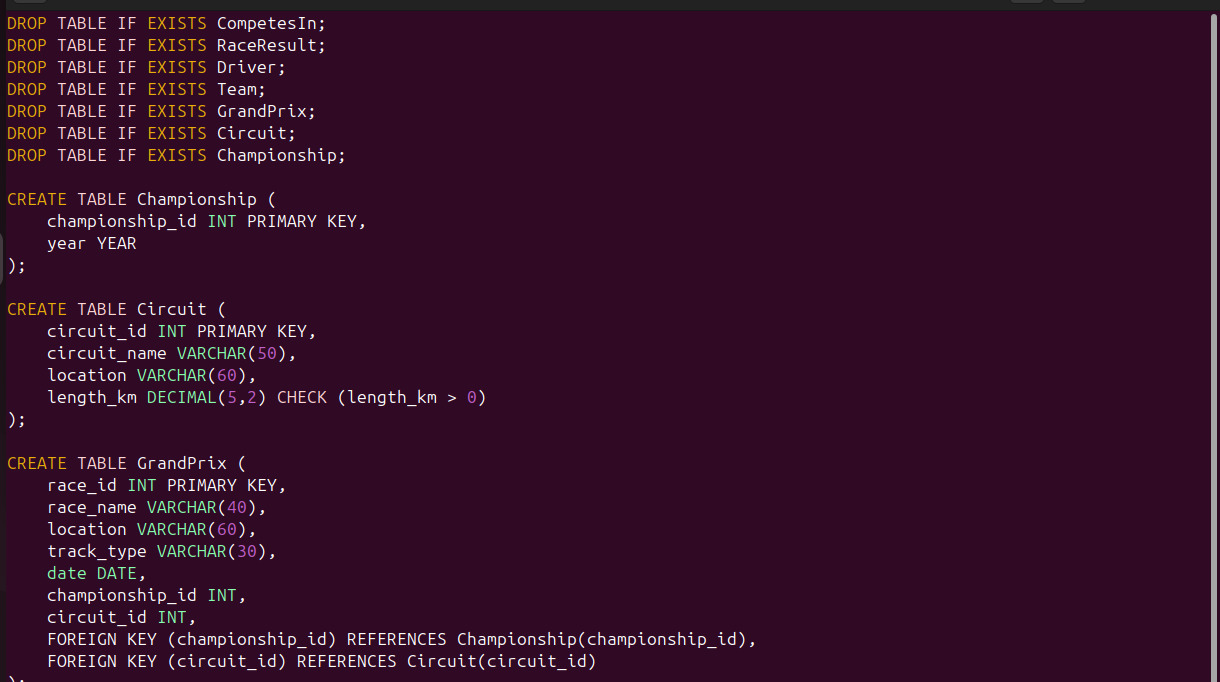
FOREIGN KEY constraints to maintain relationships.

CHECK constraints to enforce data validity (e.g., ensuring circuit length is positive).

UNIQUE constraints for attributes like team\_name to prevent duplicates.

This ensures data consistency, accuracy, and optimized retrieval.

**DELETE CASCADE** was not applied to maintain data integrity and prevent accidental deletions of important historical records. Removing a parent record (such as a Championship, Circuit, or Team) should not automatically delete associated data, as these records hold valuable historical information. Instead, data should be manually reviewed and deleted if necessary, ensuring that essential information is not lost unintentionally.



The sample data was realistically designed to represent actual Formula 1 championships, teams, races, and drivers.

**Championships**:

Includes multiple seasons (2025, 2026, 2027, etc.) to simulate real-world Formula 1 history.

**Circuits**:

Well-known circuits such as Monaco, Silverstone, Suzuka, and Yas Marina were added with their respective locations and lengths.

**Grand Prix Races:**

Each race is assigned to a circuit and linked to a championship year.

Example: "Monaco Grand Prix 2025" held at "Monaco Circuit".

**Teams and Drivers:**

Teams like Red Bull Racing, Mercedes, Ferrari, McLaren were included.

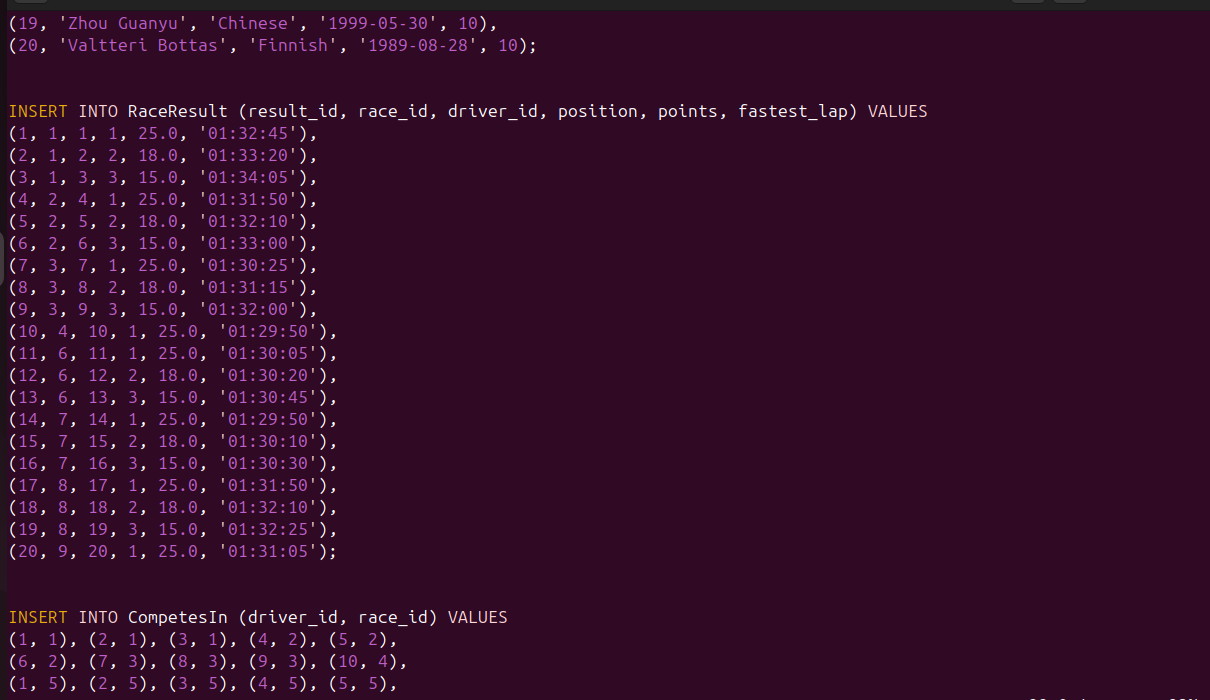
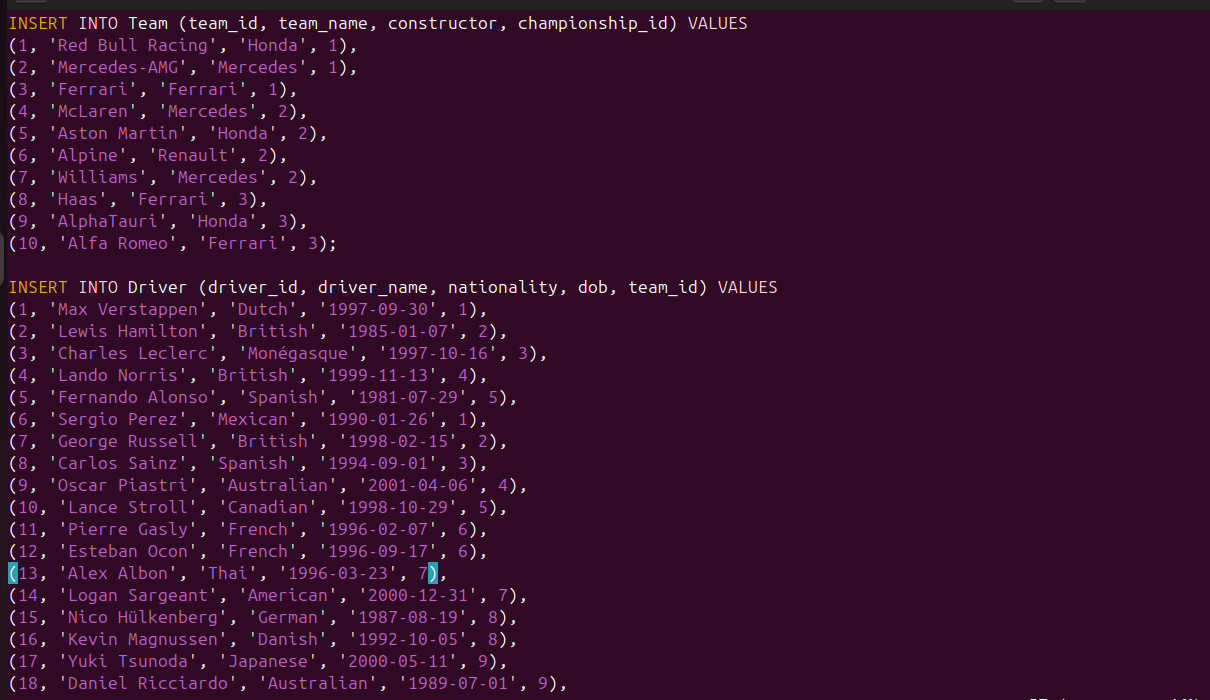
Drivers were assigned to teams with realistic details, including nationality and birthdate.

**Race Results & Driver Participation:**

Driver performances were inserted, including finishing positions, points earned, and fastest lap times.

The CompetesIn table ensures drivers can participate in multiple races across seasons.

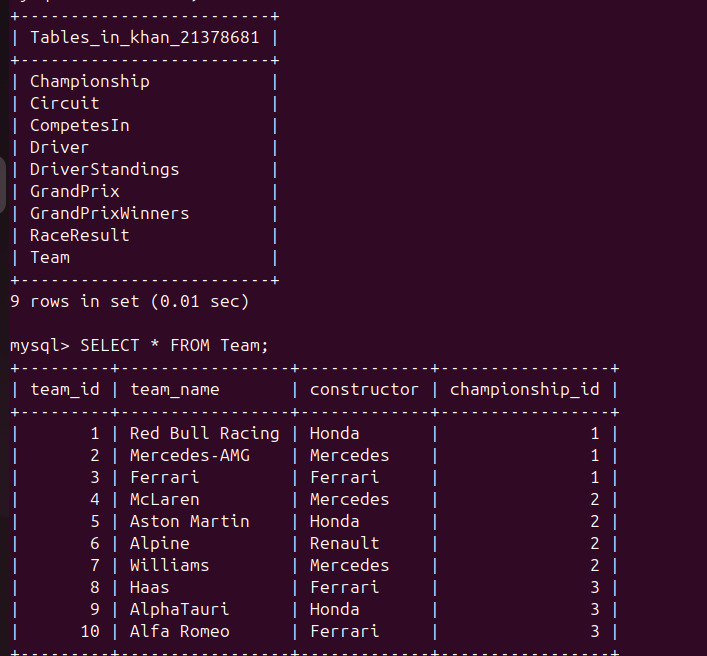
**What Data Was Inserted into the Database:**



**The data was selected from the following websites:**

**<https://en.wikipedia.org/wiki/Formula_One>  
<https://www.formula1.com/>**

**Implemented examples:**

****

**All the tables**

**Team and Driver tables values**

****

# USE OF THE DATABASE

## Advanced Queries

1. **Purpose**: Finds top 3 drivers based on total points for the 2025 season.

Why Its Important: Determines championship leaders and performance ranking.

1. **Purpose**: Retrieves how many races were held in each championship season.

Why Its Important: Helps in comparing season formats and identifying changes in race count over the years.

**3. Purpose**: Identifies which team has won the most races.

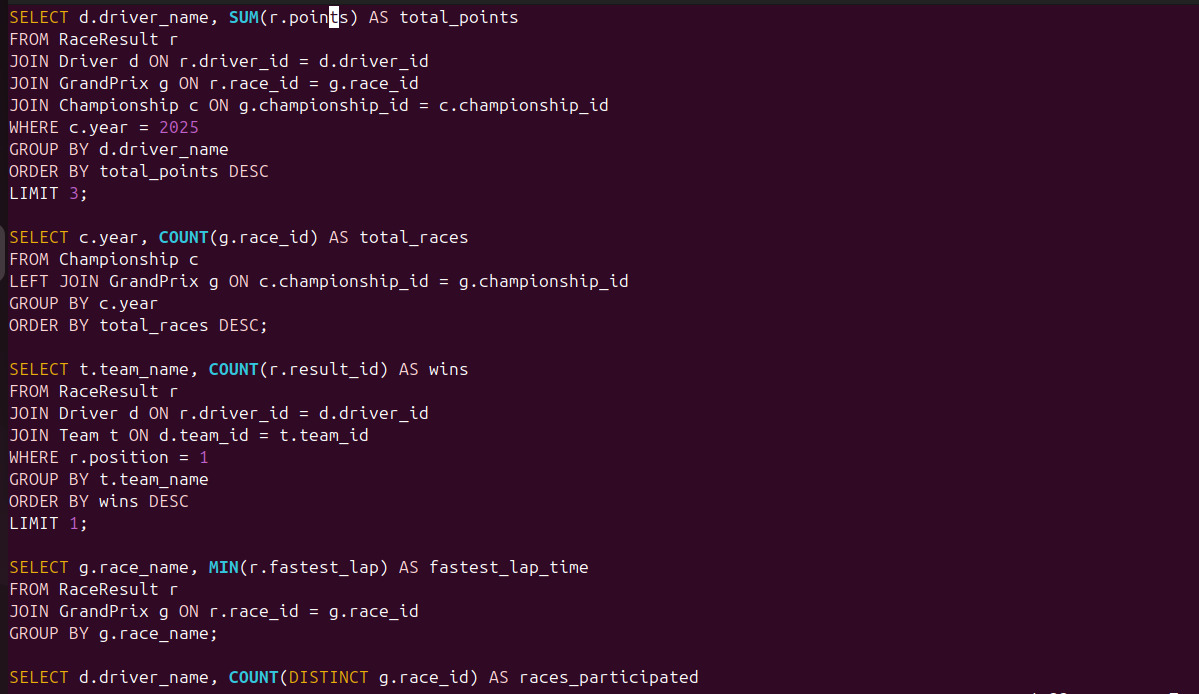
Why Its Important: Determines which constructor is dominating the season.

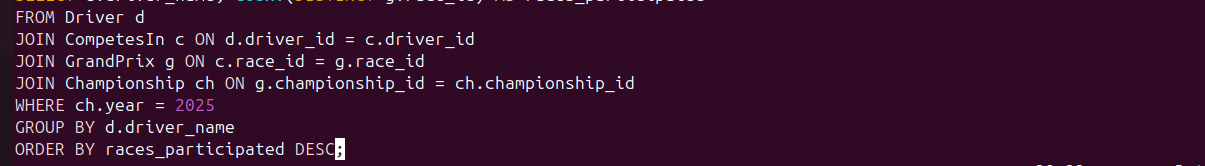
**4.** **Purpose**: Finds the fastest lap recorded per race.

Why Its Important: Useful for analyzing driver speed and track performance.

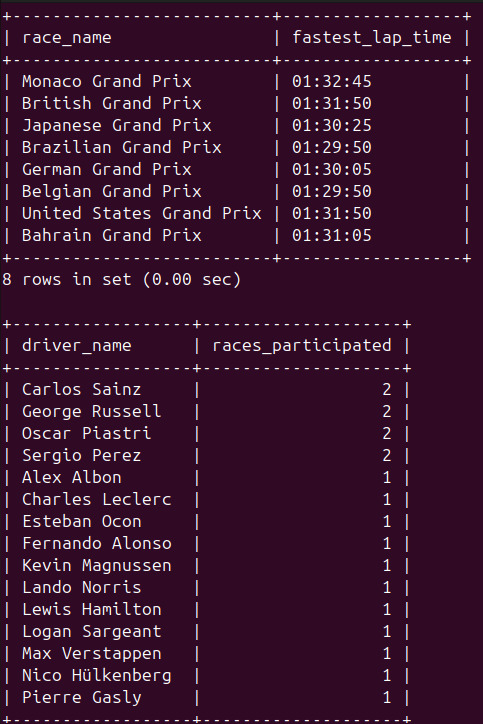
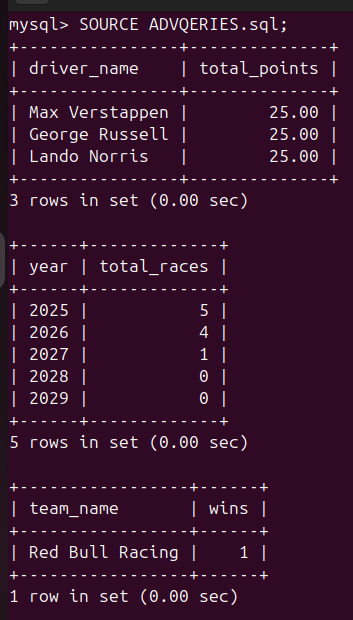
1. **Purpose**: Determines which drivers participated in the most races in 2025.

Why Its Important: Helps in analyzing driver consistency and engagement in the championship.





**SAMPLE OUTPUTS**



## Simple Queries

1. **Purpose:** This query lists all Grand Prix races scheduled in 2025, including their location and date.

Why It’s Important: Helps in analyzing upcoming events for a given season.

1. **Purpose:** Retrieves all drivers of British nationality.

Why It’s Important: Useful for country-based performance analysis or for fans tracking national drivers.

1. **Purpose:** Finds the longest Formula 1 circuit by sorting circuits in descending order of length.

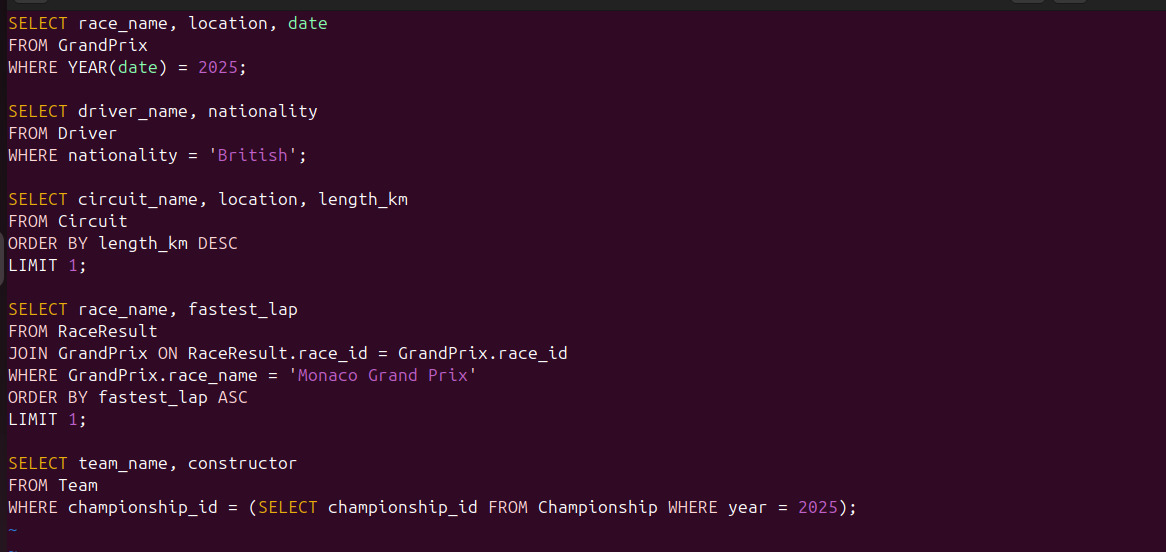
Why It’s Important: Helps in comparing track lengths and determining which circuits are the most challenging.

1. **Purpose:** Retrieves the fastest recorded lap time in the Monaco Grand Prix.

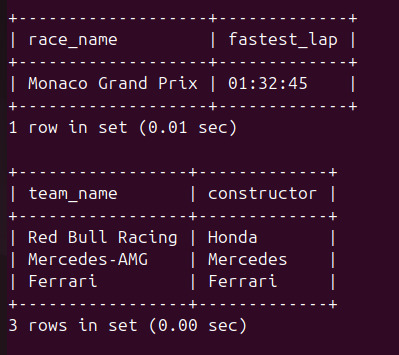
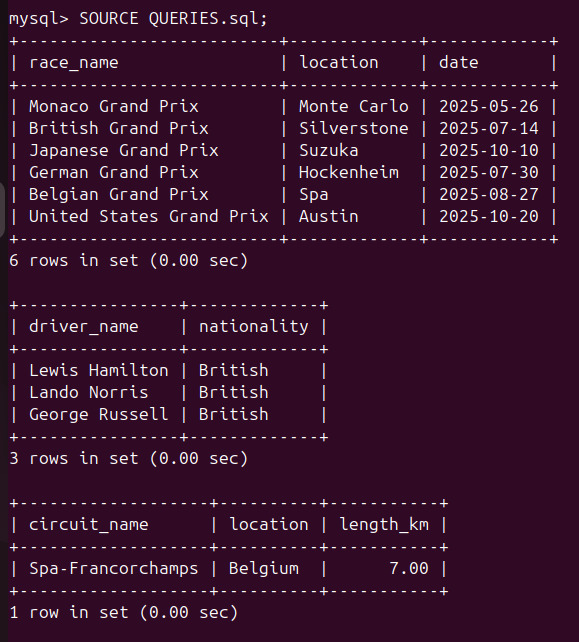
Why It’s Important: Used for tracking driver performance and analyzing the speed of laps on a specific track

1. **Purpose:** Retrieves all teams and their constructors competing in the 2025 Formula 1 Championship.

Why It’s Important: Helps in analyzing team participation per season and understanding constructor involvement.



**SAMPLE OUTPUTS**

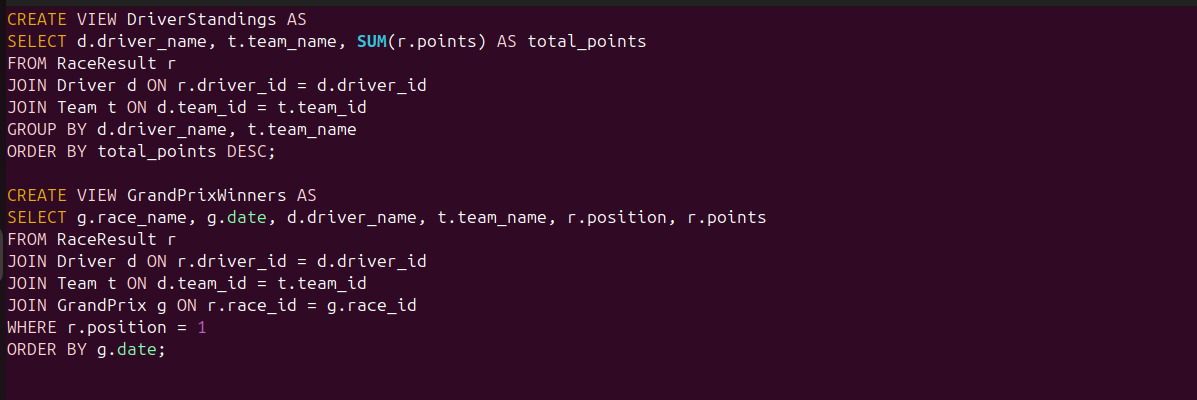
****

## Advanced Features (Stored Procedure And View)

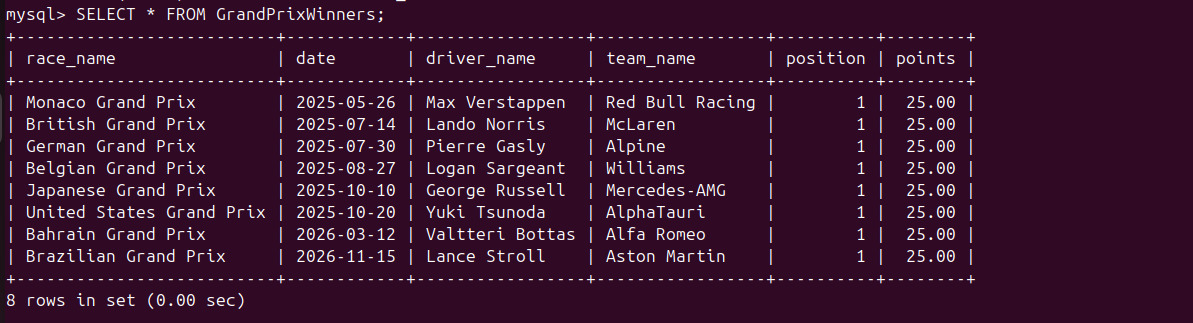
**Views:**

DriverStandings: A precomputed view that dynamically ranks drivers based on total points.

GrandPrixWinners: Retrieves only race winners, making it easy to analyze championship results.

****

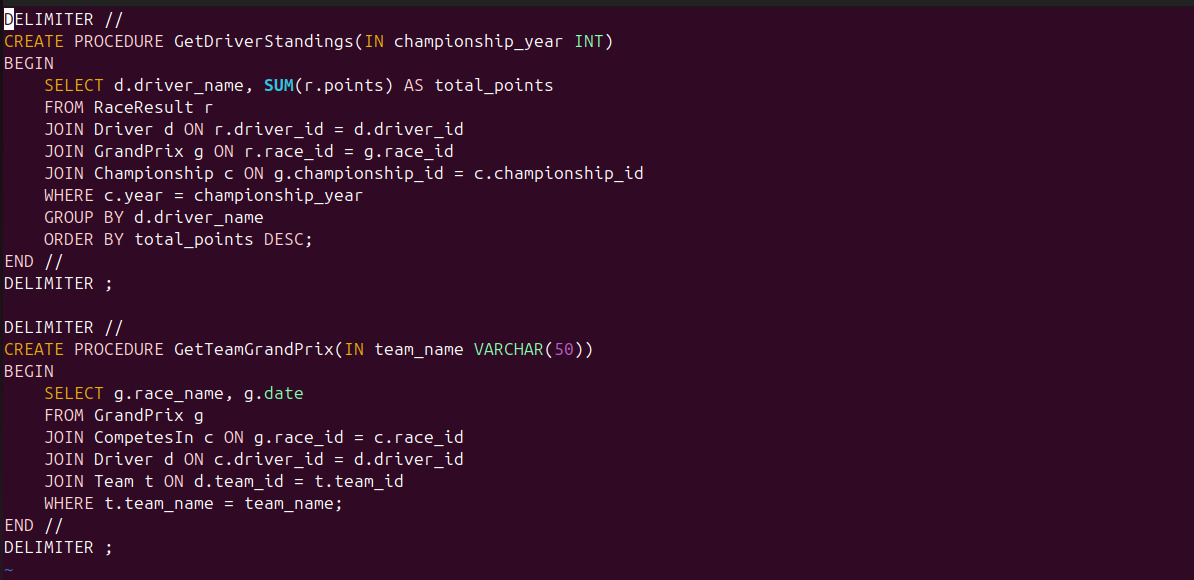
**SAMPLE OUTPUT**

****

**Stored Procedures:**

GetDriverStandings(year): Returns rankings for a given season, simplifying dynamic queries.

GetTeamGrandPrix(team\_name): Lists all races a team has competed in, making team analysis more efficient.



**SAMPLE OUTPUT**





## Database Connectivity and Python Implementation

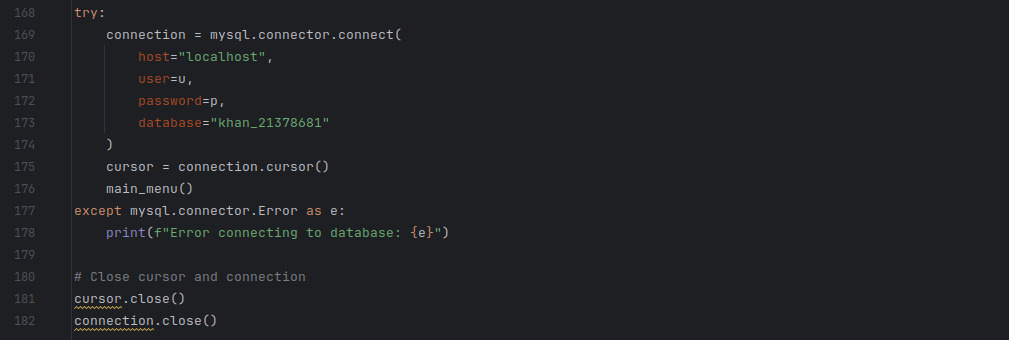
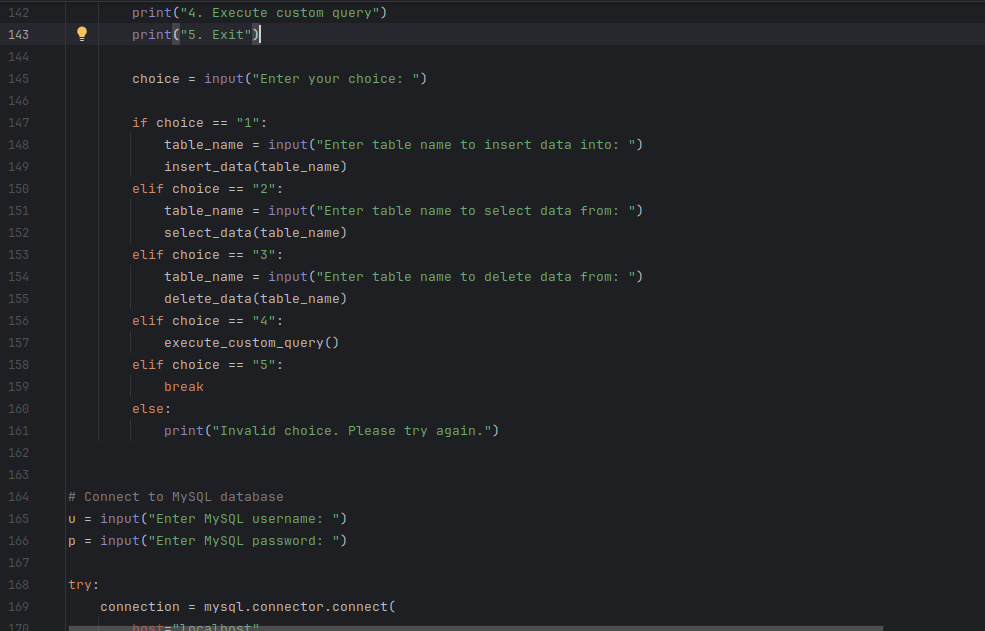
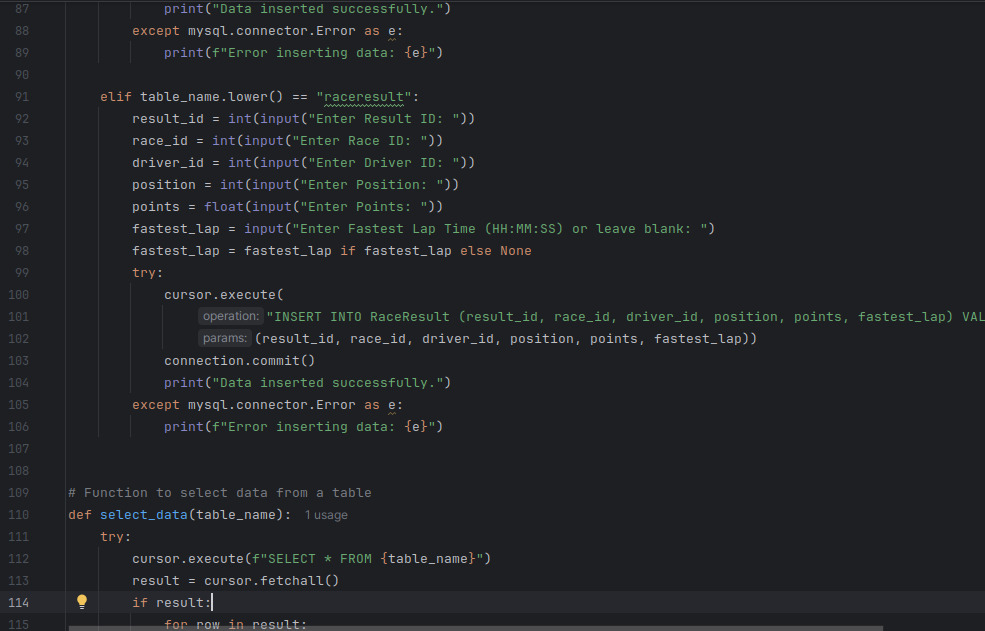
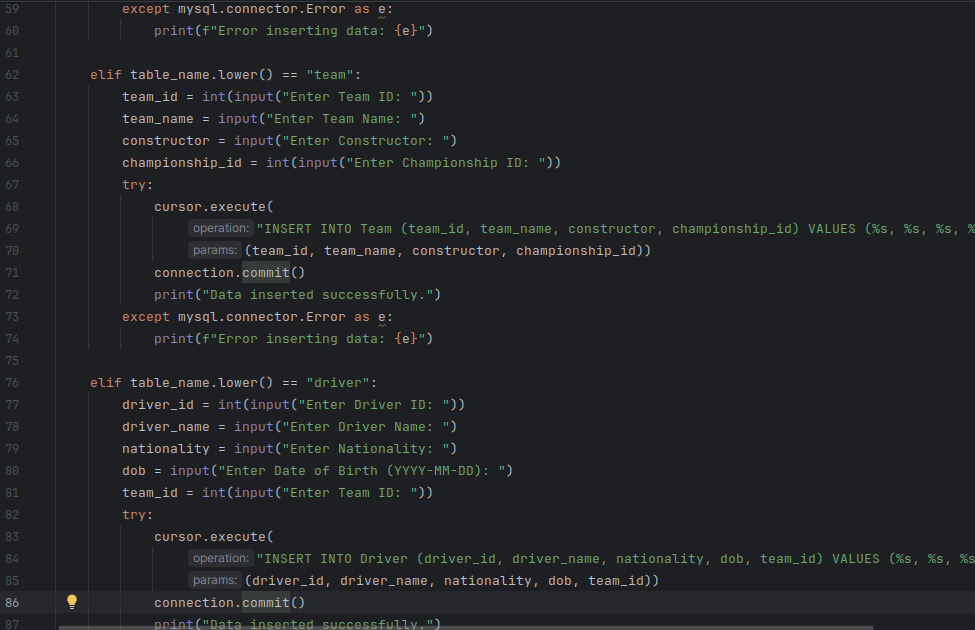
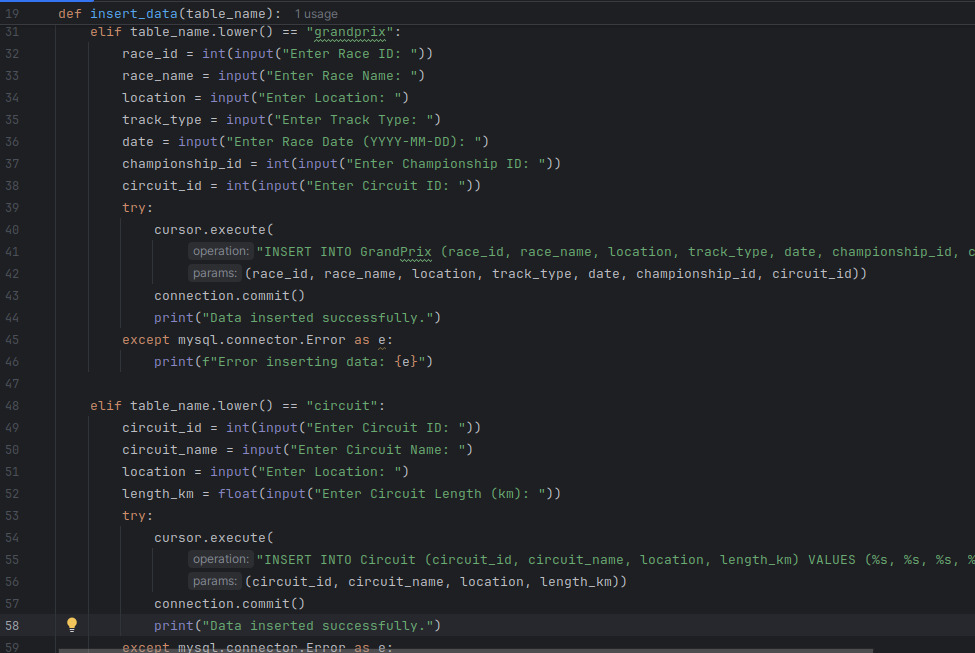
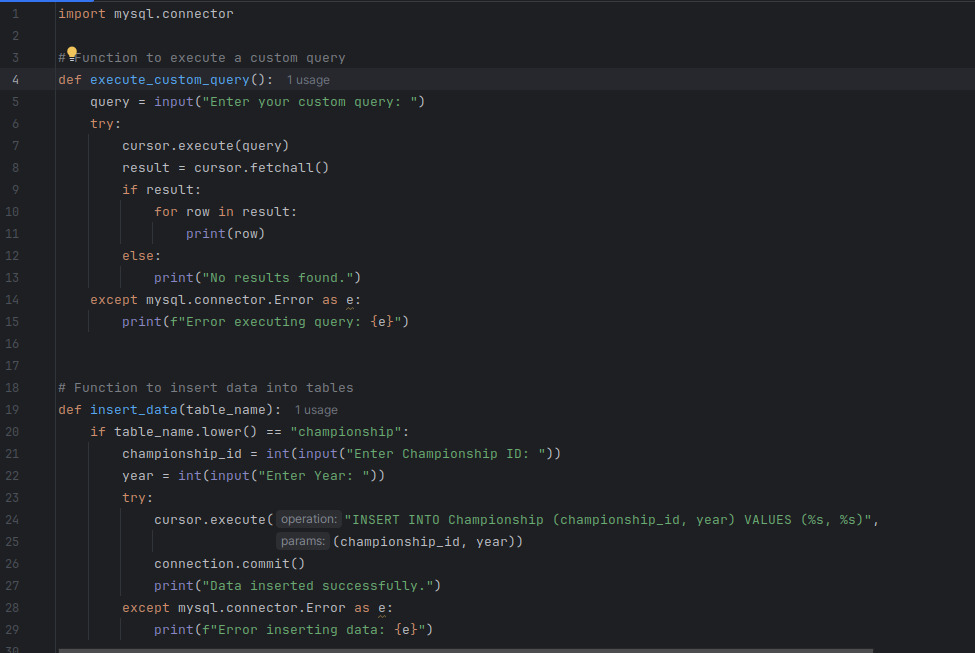
A Python application was developed to interact with the Formula 1 database, allowing users to insert, delete, update, and query data dynamically.

* Used mysql.connector for MySQL connectivity.
* Implemented CRUD operations (Create, Read, Update, Delete) with user input handling.
* Designed a menu-driven interface for ease of use, supporting:

Dynamic data entry (e.g., inserting new drivers, race results, and teams).

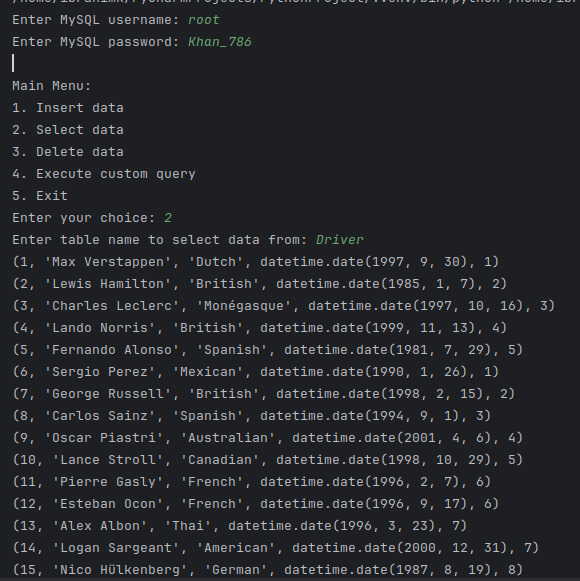
Query execution (e.g., retrieving standings, race results, and participation stats).

Custom query execution for flexibility.

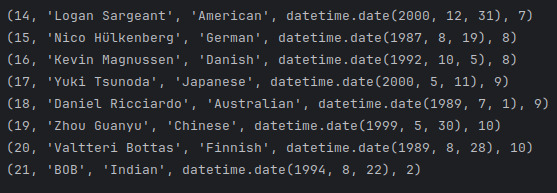
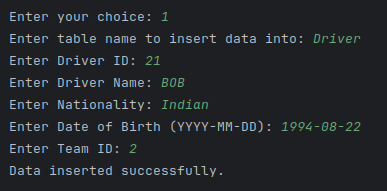


**SAMPLE OUTPUT**

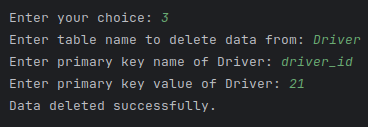
**SELECT DATA**



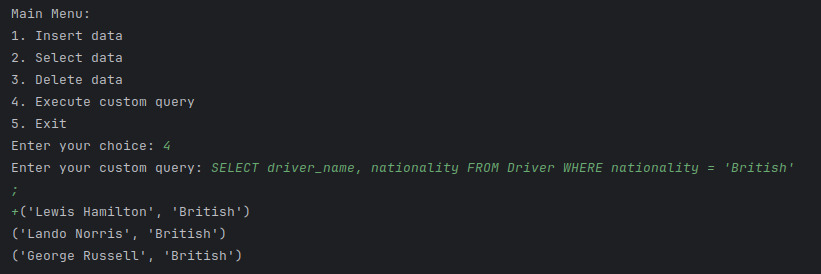
**INSERTING DATA**



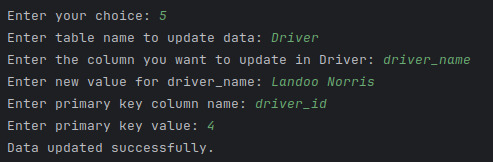
**DELETE DATA**



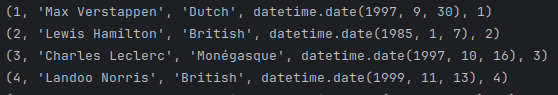
**CUSTOM QUERY**



**UPDATE DATA**

****

**SAMPLE OUTPUT**



# DISCUSSION

Looking back at this project, I was able to successfully design and implement a Formula 1 database that allows storing, retrieving, and analyzing race-related data. The database structure ensures data integrity through well-defined relationships, foreign keys, and constraints. By implementing advanced queries and views, I made it easier to extract meaningful insights, such as driver standings, fastest laps, and team performance.Initially, I faced issues when inserting data due to missing **parent records**. I had to carefully structure the insertion order to ensure referential integrity. Some queries, especially those with **multiple JOINs and aggregations**, were slow. I had to optimize them by using **indexes** and refining **WHERE conditions**.Overall, this project helped me understand database design, SQL querying, and optimization techniques. While there are still areas for improvement, I believe this database provides a solid foundation for managing Formula 1 race data efficiently.