Formula 1 Racing Data

Dataset Sources

https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020

https://www.formula1.com/en/results.html/1950/races/94/great-britain/race-result.html

-Provide car manufacturers

-Provide location site or circuit

-Provide driver

- Race Results

Technology

-MySQL workbench

-Kaggle

-https://tableconvert.com/csv-to-sql

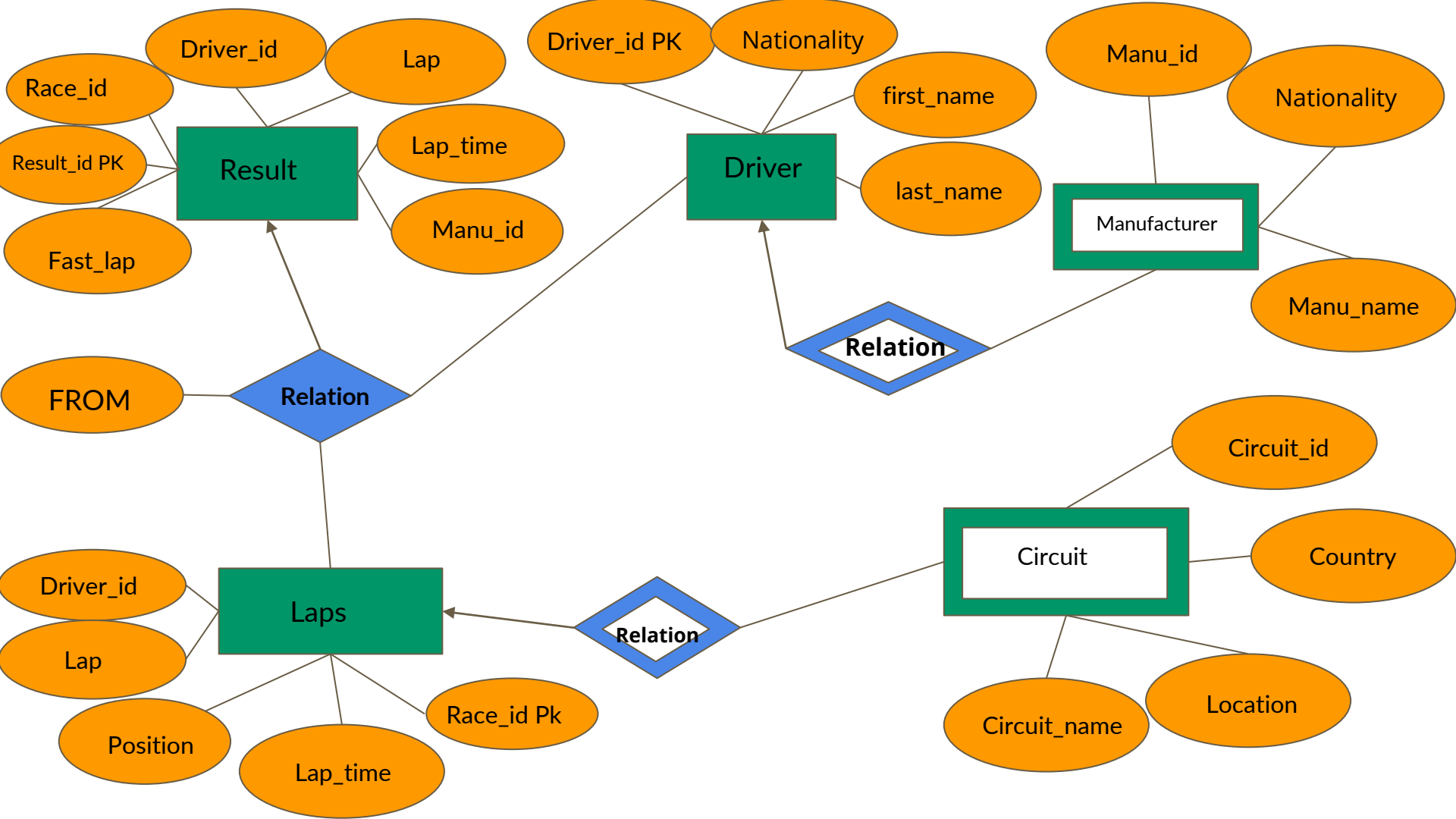
-Microsoft Excell

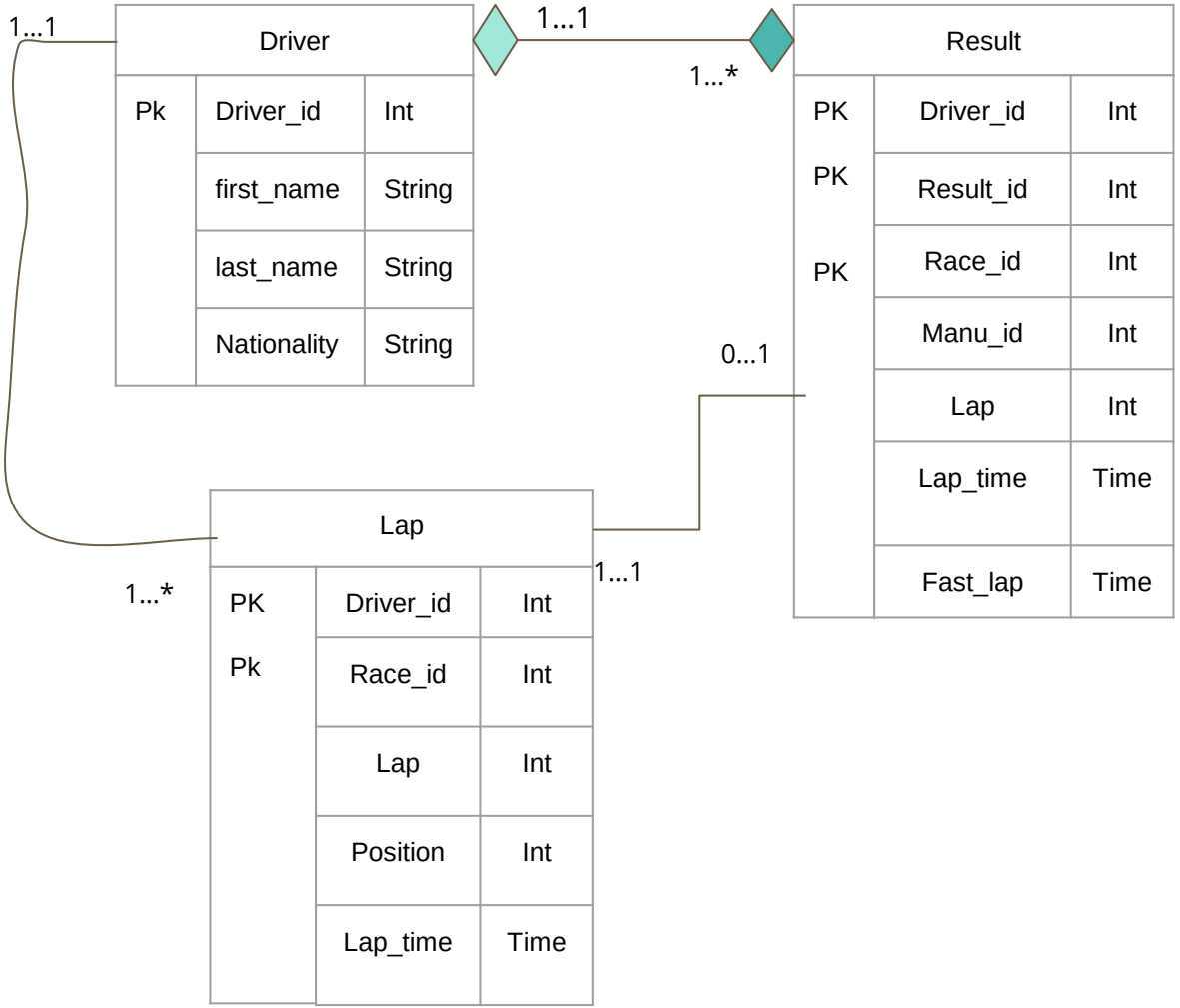
Question We hope to Answer

Which car manufacturer produce the Fastest lap?

Which circuit has produce the fastest lap time?

Who has won the most races?


Driver_id
Lap
Race_id
Result_id PK
Result
Lap_time
Fast_lap
Manu_id
Driver_id PK
Nationality
first_name
Driver
last_name
Manu_id
Nationality
Manufacturer
Manu_name
Relation
FROM
Relation
Circuit_id
Circuit
Country
Driver_id
Laps
Lap
Relation
Position
Lap_time
Race_id Pk
Circuit_name
Location


1...1
Driver
1...1
1...*
Result
Pk
Driver_id
Int
first_name
String
last_name
String
Nationality
String
PK
Driver_id
Int
PK
Result_id
Int
PK
Race_id
Int
Manu_id
Int
Lap
Int
Lap_time
Time
Fast_lap
Time
0...1
1...1
1...*
Lap
PK
Driver_id
Int
Pk
Race_id
Int
Lap
Int
Position
Int
Lap_time
Time

Phase 3 Database Design

-Key Definitions:

- Result: Result_id(INTEGER)

  A unique identifier for the Results of all thing

- Driver: Driver_id(INTEGER),

  A unique record number within the result id

- Race: Race_id(INTEGER),

  A unique record number within the result id

Phase 4: Data Normalization

-Functional Dependency and MDV:

-FD:

*Result_id → All others

-MVD:

*Result_id, Driver_id → All others

*Race_id, Driver_id → Laps

* Driver_id, Nationality → Manufacturer

Normal Forms Violation

-3NF: No Transitive Dependency

-BCNF: No non-prime key

-4NF: Contain MVD

-(Driver_id, Race_id, Result_id) → Lap and Lap_time

Result		
PK	Driver_id	Int
PK	Result_id	Int
PK	Race_id	Int
	Manu_id	Int
	Lap	Int
	Lap_time	Time
	Fast_lap	Time

Normal Forms Violation

-3NF: No Transitive Dependency

-BCNF: No non-prime key

-4NF: Contain no MVD

Driver		
PK	Driver_id	Int
	first_name	String
	last_name	String
	Nationality	String

Normal Forms Violation

-3NF: Transitive Dependency

-Race_id, Lap, Position and Lap_time depend Driver_id

-BCNF: No non-prime key

-4NF: Contain no MVD

Lap		
PK	Driver_id	Int
PK	Race_id	Int
	Lap	Int
	Position	Int
	Lap_time	Time

Phase 5: Querying The Database

Subqueries:

Main Query:

```
SELECT Manu.Manu_name,
AVG(Result.Fastest_lapSpeed) AS
avg_speed
FROM Result
JOIN Manu ON Result.Manu_id =
Manu.Manu_id
GROUP BY Manu.Manu_name
HAVING COUNT(DISTINCT
Result.Race_id) >= 5
ORDER BY avg_speed DESC
LIMIT 10;
```

Subquery 1:

```
SELECT Manu.Manu_name, Driver.first_name,
Driver.last_name, AVG(Result.Fastest_lapSpeed)
AS avg_speed

FROM Result

JOIN Manu ON Result.Manu_id = Manu.Manu_id

JOIN Driver ON Result.Driver_id = Driver.Driver_id

WHERE Driver.Nationality = 'British'

GROUP BY Manu.Manu_name, Driver.first_name,
Driver.last_name

HAVING COUNT(DISTINCT Result.Race_id) >= 5

ORDER BY avg_speed DESC

LIMIT 10;
```

Subquery 2:

```
SELECT Manu.Manu_name, Driver.first_name,
Driver.last_name, AVG(Result.Fastest_lapSpeed) AS
avg_speed

FROM Result

JOIN Manu ON Result.Manu_id = Manu.Manu_id

JOIN Driver ON Result.Driver_id = Driver.Driver_id

WHERE Driver.Nationality = 'Japanese'

GROUP BY Manu.Manu_name, Driver.first_name,
Driver.last_name

HAVING COUNT(DISTINCT Result.Race_id) >= 5

ORDER BY avg_speed DESC

LIMIT 10;
```

The result from left to right are:
-The top 10 fastest lap in races in atleast 10 races
-The second show the top 10 manufactuer with their Nationality in this case it British
-The last query show the top Nationality in this case Japan, driver and their fastest speed.

Phase 5: Querying The Database

Aggregation:

Aggregation 1:

```
SELECT
    Circuit.Country,
    Circuit.Location,
    COUNT(*) AS Number_of_Races,
    AVG(Result.Fastest_lapSpeed) AS
Average_Fastest_Lap_Speed
FROM Circuit
JOIN Result ON Circuit.Circuit_id = Result.Race_id
GROUP BY Circuit.Country, Circuit.Location
ORDER BY Number_of_Races DESC;
```

Aggregation 2:

```
SELECT Driver.first_name, Driver.last_name,
COUNT(Result.Result_id) AS num_wins
FROM Result
JOIN Driver ON Result.Driver_id = Driver.Driver_id
WHERE Result.Lap_time = 1
GROUP BY Driver.first_name, Driver.last_name
ORDER BY num_wins DESC;
```

Aggregation 3:

```
SELECT Manu.Manu_name, COUNT(DISTINCT
Result.Race_id) AS num_races,
AVG(Result.Fastest_lapSpeed) AS avg_speed
FROM Result
JOIN Manu ON Result.Manu_id = Manu.Manu_id
GROUP BY Manu.Manu_name
HAVING COUNT(DISTINCT Result.Race_id) >= 5
ORDER BY num_races DESC;
```

These aggregation show the result from left to right are the number of races and there fastest speed, Count the number of first and last name and the races they won, and the last show the races won by manufacturer

Phase 5: Querying The Database

Insert and Update Query:

```
INSERT INTO Circuit (Circuit_id, Circuit_name, Location, Country)

VALUES (6, 'Monza', 'Monza, Italy', 'Italy');
```

```
INSERT INTO Driver (Driver_id, first_name, last_name,
Nationality)

VALUES (29, 'Chad', 'Owen', ‘American');
```

```
UPDATE Lap

SET Lap_time = '01:30.253'

WHERE Lap_id = 100;
```

```
UPDATE Circuit

SET Location = 'Los Angeles', Country = 'United State'

WHERE Circuit_id = 300;
```

Phase 6/7: Analysis

Here are the result to our Question we mention:

Manu_name	avg_speed
BMW Sauber	202.08397857142856
Racing Point	201.0525263157895
Alpine F1 Team	200.5545909090909
AlphaTauri	198.02559836065564
Brawn	197.2609411764706
Red Bull	195.2266824712644
Haas F1 Team	192.9470902777777
Spyker	190.90673529411762
Toro Rosso	189.44319776119414
Force India	186.52382311320738

Manufacturer that produce the fastest speed

first_name	last_name	num_wins
Sergio	Pérez	47
Felipe	Massa	45
Jenson	Button	41

Who has won the most races?

Country	Location	Number_of_Races	Average_Fastest_Lap_Speed
Canada	Ontario	22	235.62845454545456
Spain	Barcelona	44	233.22029545454544
Belgium	Spa	20	231.40064999999998
UK	Castle Doni...	20	229.76065
South A...	Midrand	20	229.66275000000002
France	Magny Cours	20	226.02519999999998
USA	Nevada	22	223.40436363636366
USA	Austin	22	221.82245454545455
Russia	Sochi	20	218.43024999999997
Brazil	Rio de Janeiro	22	217.7501818181818

Which circuit has produce the fastest lap time?