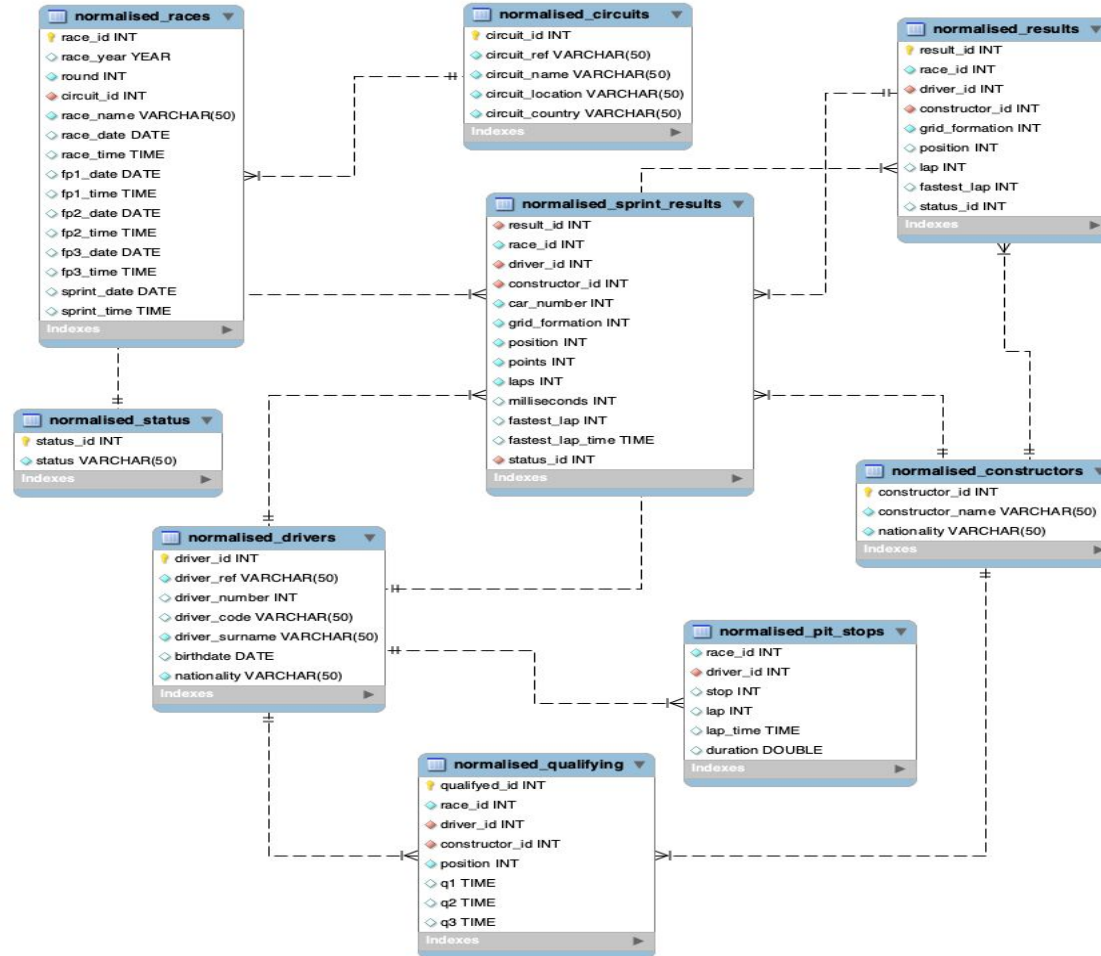


F1 SQL MOOC

TEAM WHITE



Normalisation of datasets, PK and FK





01

F1 BACKGROUND

- Country that hosted F1 races
- Circuit that hosted races and which countries joined them

02

ABOUT THE RACER

- Avg/max/min time difference between the 1st, 2nd and 3rd drivers
- Most constructor wins
- Driver that has won most races and the respective constructor
- Fastest stop in history

.... And the curse of number 13

03

INVESTIGATE RELATIONSHIPS

- Compare nationality of the champion and host country
- F1 top speed record, its correlated circuit and constructor
- Avg/sum/count of pole position and winners
- Correlation between duration of pit stops and time to reach and line
- Avg. pit stop duration by constructor and circuit

01

F1 BACKGROUND

Country that hosted F1 races

```
SELECT circuit_country  
FROM normalised_circuits;
```

```
circuit_country  
Australia  
Malaysia  
Bahrain  
Spain  
Turkey  
Monaco  
Canada  
France  
UK  
Germany  
Hungary  
Spain  
Belgium  
Italy  
Singapore  
Japan
```

Circuits that hosted races and countries that joined

```
SELECT circuit_ref, circuit_country  
FROM normalised_circuits;
```

```
circuit_ref,circuit_country  
albert_park,Australia  
sepang,Malaysia  
bahrain,Bahrain  
catalunya,Spain  
istanbul,Turkey  
monaco,Monaco  
villeneuve,Canada  
magny_cours,France  
silverstone,UK  
hockenheimring,Germany  
hungaroring,Hungary  
valencia,Spain  
spa,Belgium  
monza,Italy  
marina_bay,Singapore
```

Fastest driver amongst first 3 positions

```
SELECT driver_id, position,  
MIN(milliseconds) AS difference  
FROM normalised_sprint_results  
WHERE position BETWEEN 1 and 3  
GROUP BY driver_id, position  
ORDER BY MIN(milliseconds) ASC;
```

driver_id	position	difference
830	1	1538426
1	2	1539856

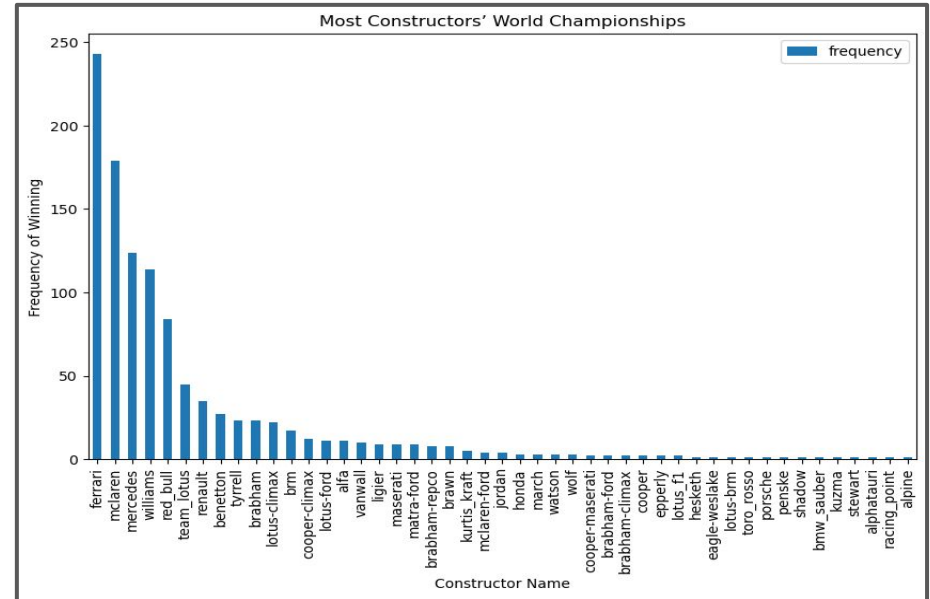
Slowest driver amongst first 3 positions

```
SELECT driver_id, position,  
MAX(milliseconds) AS difference  
FROM normalised_sprint_results  
WHERE position BETWEEN 1 and 3  
GROUP BY driver_id, position  
ORDER BY MAX(milliseconds) DESC;
```

driver_id	position	difference
815	3	1844288
844	2	1842542

Constructor that has won the most races

```
SELECT c.constructor_id, c.constructor_name  
FROM normalised_constructors AS c  
INNER JOIN(SELECT constructor_id,  
COUNT(constructor_id) FROM normalised_results  
WHERE position = 1 GROUP BY constructor_id ORDER  
BY COUNT(constructor_id) DESC LIMIT 1) AS v  
ON c.constructor_id = v.constructor_id;
```



02

ABOUT THE RACER

Driver that has won the most races and respective constructor

```
SELECT d.driver_surname, c.constructor_name,  
r.race_year  
FROM normalised_sprint_results AS sr  
LEFT JOIN normalised_drivers AS d on sr.driver_id  
= d.driver_id  
LEFT JOIN normalised_constructors AS c on  
sr.constructor_id = c.constructor_id  
LEFT JOIN normalised_races AS r on sr.race_id =  
r.race_id  
WHERE sr.position = 1  
ORDER BY r.race_year;
```

driver_surname	constructor_name	race_year
Verstappen	red_bull	2021
Bottas	mercedes	2021
Bottas	mercedes	2021
Verstappen	red_bull	2022
Verstappen	red_bull	2022

Fastest pit stop ever

```
SELECT d.driver_surname, ps.duration, r.race_id,  
r.race_date, r.round, r.race_name  
FROM normalised_pit_stops ps  
LEFT JOIN normalised_drivers d ON ps.driver_id =  
d.driver_id  
LEFT JOIN normalised_races r ON ps.race_id =  
r.race_id  
WHERE ps.duration = (SELECT MIN(duration) FROM  
normalised_pit_stops);
```

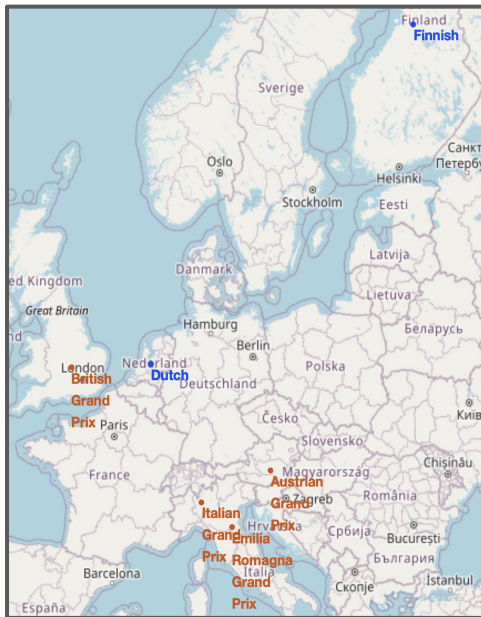
driver_surn ame	duration	race_ id	race_ date	round	race_nam e
Maldonado	12.897	858	2011- 11-13	18	Abu Dhabi Grand Prix

03

INVESTIGATE RELATIONSHIPS

Compare the nationality of the 1st racers and the hosting country

```
SELECT d.nationality,
r.race_name, r.race_id
FROM
normalised_sprint_results AS sr
LEFT JOIN
normalised_drivers AS d
ON sr.driver_id =
d.driver_id
LEFT JOIN
normalised_races AS r
ON sr.race_id =
r.race_id
WHERE sr.position =1
ORDER BY
d.driver_surname;
```



F1 top fastest lap, its correlated circuit and its constructor

```
SELECT ns.fastest_lap_time, ns.milliseconds,
c.circuit_name, co.constructor_name, ns.driver_id
FROM normalised_sprint_results AS ns
INNER JOIN normalised_circuits AS c
ON ns.driver_id = c.circuit_id
LEFT JOIN normalised_constructors AS co
USING (constructor_id)
WHERE ns.fastest_lap_time IS NOT NULL
AND ns.milliseconds IS NOT NULL
ORDER BY ns.fastest_lap_time LIMIT 1;
```

fastest_lap_time	milliseconds	circuit_name	constructor_name	driver_id
01:09:01	1625506	Albert Park Grand Prix Circuit	mercedes	1

03

INVESTIGATE RELATIONSHIPS

The avg of the pole position (q1, q2, q3) and the winners

```
SELECT sec_to_time(AVG(TIME_TO_SEC(q1))) AS average_q1,
sec_to_time(STD(TIME_TO_SEC(q1))) AS std_q1,
sec_to_time(AVG(TIME_TO_SEC(q2))) AS average_q2,
sec_to_time(STD(TIME_TO_SEC(q2))) AS std_q2,
sec_to_time(AVG(TIME_TO_SEC(q3))) AS average_q3,
sec_to_time(STD(TIME_TO_SEC(q3))) AS std_q3
FROM normalised_qualifying;
```

average_q1	std_q1	average_q2	std_q2	average_q3	std_q3
01:28:14.7588	00:15:55.999012	01:25:08.3090	00:18:59.345005	01:25:08.3090	00:18:59.345005

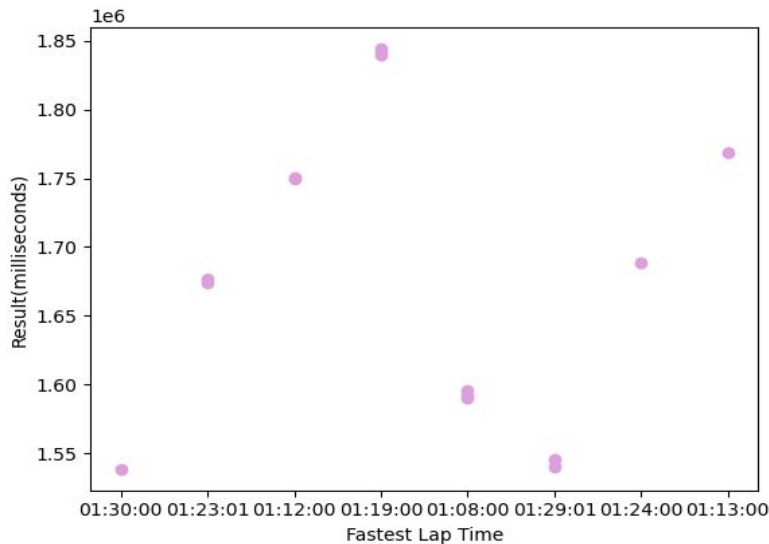
--Find the average and the std of the winners' pole positions --

```
SELECT sec_to_time(AVG(TIME_TO_SEC(q1))) AS average_q1,
sec_to_time(STD(TIME_TO_SEC(q1))) AS std_q1,
sec_to_time(AVG(TIME_TO_SEC(q2))) AS average_q2,
sec_to_time(STD(TIME_TO_SEC(q2))) AS std_q2,
sec_to_time(AVG(TIME_TO_SEC(q3))) AS average_q3,
sec_to_time(STD(TIME_TO_SEC(q3))) AS std_q3
FROM normalised_qualifying
WHERE position = 1;
```

average_q1	std_q1	average_q2	std_q2	average_q3	std_q3
01:28:14.7588	00:15:55.999012	01:25:08.3090	00:18:59.345005	01:25:08.3090	00:18:59.345005

Correlation between duration of pit stops and tie to win race/reach the end line

```
SELECT position,laps AS no_laps,
fastest_lap_time,
milliseconds AS result_milliseconds,
driver_id, constructor_id, race_id
FROM normalised_sprint_results
WHERE position BETWEEN 1 AND 3
ORDER BY position ASC;
```



03

INVESTIGATE RELATIONSHIPS

The average pit stop duration by each constructor in each year

```
SELECT ROUND(AVG(duration), 2) AS
avg_pit_stop_duration, c.constructor_name,
races.race_year
FROM normalised_pit_stops AS ps
INNER JOIN normalised_constructors AS c
    ON ps.driver_id = c.constructor_id
LEFT JOIN normalised_races AS races
    ON ps.race_id = races.race_id
GROUP BY c.constructor_name, races.race_year
ORDER BY c.constructor_name ASC LIMIT 6;
```

avg_pit_stop_duration	constructor_name	race_year
26.86	ags	2011
25.08	ags	2012
23.02	bar	2011
23.91	bar	2013
25.79	bar	2014
24.62	benetton	2011

The average pit stop by circuit

```
SELECT ROUND(AVG(ps.duration), 2) AS avg_pit_stop,
c.circuit_id AS circuit_id, c.circuit_name,
c.circuit_location, c.circuit_country, races.race_name
FROM normalised_pit_stops AS ps
INNER JOIN normalised_circuits AS c
    ON ps.driver_id = c.circuit_id
LEFT JOIN normalised_races AS races
    ON ps.race_id = races.race_id
GROUP BY c.circuit_id, races.circuit_id, c.circuit_name,
c.circuit_location, c.circuit_country, races.race_name
ORDER BY c.circuit_country ASC
LIMIT 3;
```

avg_pit_stop	circuit_id	circuit_name	circuit_location	circuit_country	race_name
23.48	1	Albert Park Grand Prix Circuit	Melbourne	Australia	Australian Grand Prix
23.27	3	Bahrain International Circuit	Sakhir	Bahrain	Chinese Grand Prix
23.72	13	Circuit de Spa-Francorchamps	Spa	Belgium	Italian Grand Prix

03

The curse of number 13*

-- >Does number 13 lost the most races?

```
SELECT *
FROM normalised_sprint_results
WHERE car_number = 13; -- Nobody use car number 13 in
this dataset
```

```
SELECT car_number
FROM normalised_sprint_results
WHERE position = 20;
```

car_number
11
10
9
24
14

-- >Does number 13 have more accidents/disqualify than other racers?

```
SELECT sr.car_number, sr.constructor_id,
sr.driver_id, s.status
FROM normalised_sprint_results sr
RIGHT JOIN normalised_status s
ON sr.status_id = s.status_id
WHERE s.status_id = 2
OR s.status_id = 3
OR s.status_id = 4
OR s.status_id = 73
OR s.status_id = 82;
```

***Formula 1** has banned the number 13. There is also no garage with the number 13 and, previously, when numbers were assigned based on constructors' position in the standings, teams avoided that number.

Why? Because historically, whoever has used 13 in F1 has finished badly. This number records two deaths in a row in 1925 (Paul Torchy) and 1926 (Giulio Masetti).

In recent history, four riders have tried this number and it hasn't gone well: fires, dropouts, failures to qualify for races...

The closest case is that of Pastor Maldonado. The Venezuelan used the 13 in his Lotus during the 2014 and 2015 seasons, and retired from F1 with almost more DNFs/dropouts than points in those two seasons.

car_number	constructor_id	driver_id	status
NULL	NULL	NULL	Disqualified
10	213	842	Accident
NULL	NULL	NULL	Collision
NULL	NULL	NULL	Injured
NULL	NULL	NULL	Injury