CODE FIRST GIRLS

Cognizant Foundation

Formula 1 Analysis

By: Kaede, Suzilah, Judy, Katie

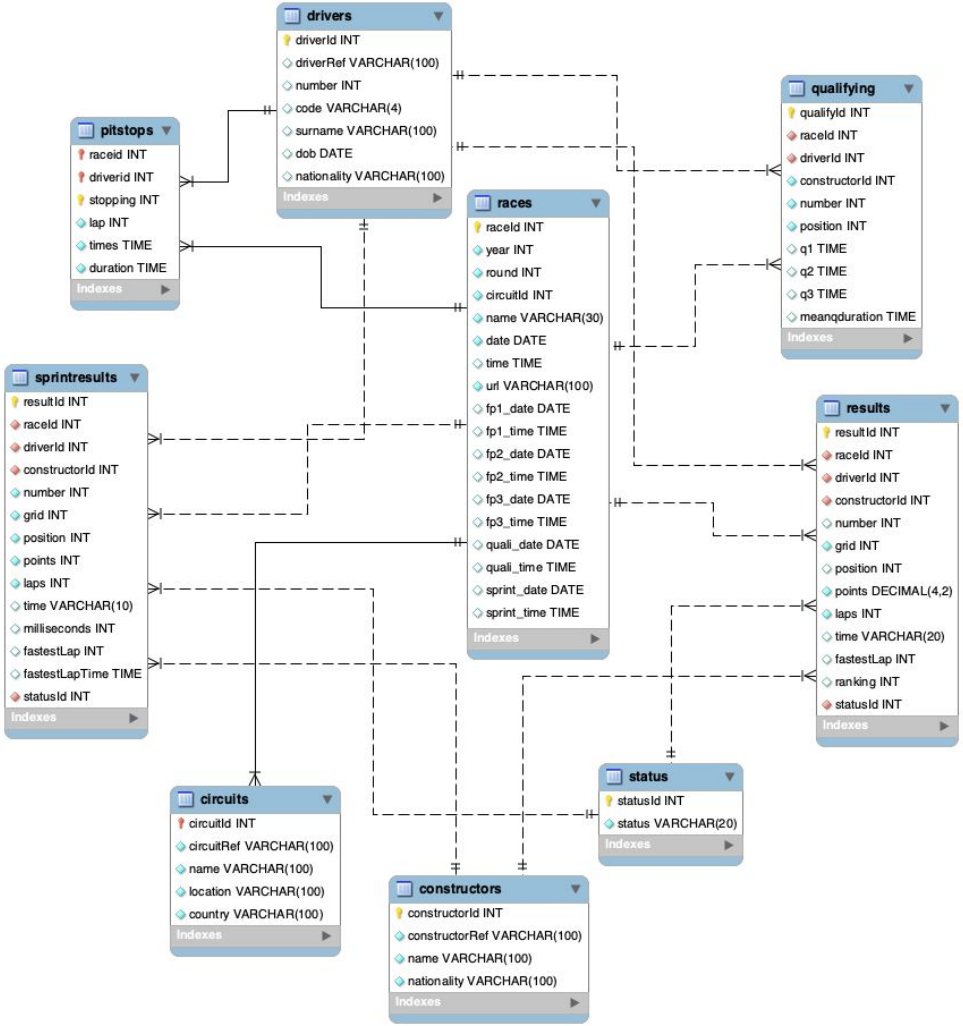
F1 is the world's most prestigious motor racing competition, for determining the **winner**.

How to win?

→ Relevant analysis and table formulation

→ data analyst for consulting?

ER Diagram


drivers
driverId INT
driverRef VARCHAR(100)
number INT
code VARCHAR(4)
surname VARCHAR(100)
dob DATE
nationality VARCHAR(100)
Indexes
pitstops
raceid INT
driverid INT
stopping INT
lap INT
times TIME
duration TIME
Indexes
races
raceId INT
year INT
round INT
circuitId INT
name VARCHAR(30)
date DATE
time TIME
url VARCHAR(100)
fp1_date DATE
fp1_time TIME
fp2_date DATE
fp2_time TIME
fp3_date DATE
fp3_time TIME
quali_date DATE
quali_time TIME
sprint_date DATE
sprint_time TIME
Indexes
qualifying
qualifyId INT
raceId INT
driverId INT
constructorId INT
number INT
position INT
q1 TIME
q2 TIME
q3 TIME
meanqduration TIME
Indexes
sprintresults
resultId INT
raceId INT
driverId INT
constructorId INT
number INT
grid INT
position INT
points INT
laps INT
time VARCHAR(10)
milliseconds INT
fastestLap INT
fastestLapTime TIME
statusId INT
Indexes
results
resultId INT
raceId INT
driverId INT
constructorId INT
number INT
grid INT
position INT
points DECIMAL(4,2)
laps INT
time VARCHAR(20)
fastestLap INT
ranking INT
statusId INT
Indexes
circuits
circuitId INT
circuitRef VARCHAR(100)
name VARCHAR(100)
location VARCHAR(100)
country VARCHAR(100)
Indexes
status
statusId INT
status VARCHAR(20)
Indexes
constructors
constructorId INT
constructorRef VARCHAR(100)
name VARCHAR(100)
nationality VARCHAR(100)
Indexes

Best driver/constructor?

→ For analysing techniques of top scorers

Winning record

e.g.,

```
CREATE TABLE wondrivers AS
SELECT d.driverId, d.surname, d.dob,
d.nationality, COUNT(d.driverId) as racetotal,
COUNT(r.position=1) as wonamount
FROM drivers d
INNER JOIN results r ON r.driverId = d.driverId
GROUP BY d.driverId;

ALTER TABLE wondrivers
ADD COLUMN wonratio DEC(10,2);

UPDATE wondrivers w
SET w.wonratio = (wonamount/racetotal);
```

Average position

e.g.,

```
CREATE TABLE dmeanposition AS
SELECT d.nationality, d.surname,
d.dob, AVG(r.position) AS
meanposition
FROM drivers d
INNER JOIN results r ON d.driverId =
r.driverId
WHERE r.position IS NOT NULL
GROUP BY d.driverId;
```

Total points

e.g.,

```
CREATE TABLE pointsdrivers AS
SELECT d.nationality, d.surname,
d.dob, SUM(r.points) as totalpoints
FROM drivers d
INNER JOIN results r ON d.driverId
= r.driverId
GROUP BY d.driverId;
```

Best drivers

Absolute win

Surname	Wonamount
Alonso	285
Raikkonen	285
Hamilton	276
Vettel	254
Button	242

Ratio win

Surname	Wonratio
Takahashi	1.00
Brown	1.00
Hasemi	1.00
Suzuki	1.00
Kiesa	1.00

Average position

Surname	Meanposition
Serafini	2.00
Amick	2.00
Ascari	2.18
Fangio	2.25
Farina	2.81

Total points

Surname	Totalpoints
Hamilton	4308.50
Vettel	3077.00
Alonso	2021.00
Raikkonen	1873.00
Verstappen	1792.50

→ Depending on calculation, top 5 are all different

Best constructors

Absolute win

Surname	Wonamount
Ferrari	1663
McLaren	1285
Williams	1129
Red Bull	558
Sauber	532

Ratio win

Surname	Wonratio
Meskowskii	1.00
Ewing	1.00
ENB	1.00
Behra-Porsche	1.00
Moore	1.00

Average position

Surname	Meanposition
Matra-Ford	3.66
Brabham-Repco	3.87
Brawn	4.06
Mercedes	4.11
Vanwall	4.19

Total points

Surname	Totalpoints
Ferrari	9924.27
Mercedes	6726.64
McLaren	6085.50
Red Bull	6027.00
Williams	3593.00

→ Again, top 5 are all different

→ Other methods, combining them, prioritising

Not finishing race?

→ To avoid disqualification

```
SELECT s.statusId, s.status,
COUNT(r.statusId) AS totalnumber
FROM status s
INNER JOIN results r ON s.statusId = r.statusId
GROUP BY r.statusId
ORDER BY totalnumber desc;
```

Status	Total number
+1 Lap	3823
Engine	2005
+2 Laps	1591
Accident	1041
Did not qualify	1025

Time per circuit?

→ For knowing target time to meet

```
CREATE TABLE circuitsduration AS
SELECT c.name, c.location, c.country, c.circuitid,
MAX(c.meanqduration) AS maxduration, MIN(c.meanqduration)
AS minduration, c.meanqduration
FROM circuitstest c
GROUP BY c.circuitid;
* ((q.q1+q.q2+q.q3)/3) AS meanqduration
```

Name	Location	Max	Min	Mean
Albert Park Grand Prix	Melbourne	14134	11768	12567
Sepang International	Kuala Lumpur	20200	13034	13467
Bahrain International	Sakhir	15501	5301	13201

Combined data for researcher?

→ For further analysis or information for others

```
CREATE TABLE forresearcher AS
SELECT r.time AS timefinished, a.year, a.name AS racename, a.date, a.time, a.url, d.surname, d.dob, d.nationality AS drivernationality,
c.name AS constructorname, c.nationality AS constructornationality, ci.name AS circuitname
FROM results r
INNER JOIN races a ON r.raceId = a.raceId
INNER JOIN drivers d ON d.driverId = r.driverId
INNER JOIN constructors c ON r.constructorId = c.constructorId
INNER JOIN circuits ci ON ci.circuitId = a.circuitId
WHERE r.position=1;
```

Time finish	Racename	Date	Time	Url	Surname	Dob	Driver national	Construc torname	Construct ornational	Circuitname
34:50.6	Australian Grand Prix	2008-03-16	04:30:00	http://	Hamilton	1985-01-07	British	McLaren	British	Albert Park Grand Prix
00:42.7	Monaco Grand Prix	2008-05-25	12:00:00	http://	Raikkonen	1979-10-17	Finnish	Ferrari	Italian	Sepang International
39:09.4	British Grand Prix	2008-07-06	12:00:00	http://	Massa	1981-04-25	Brazilian	Ferrari	Italian	Bahrain International

Correlational analysis

→ Identify relevant factors for winning

E.g.,

1.) Create table with x and y variables

```
CREATE TABLE pitstopsVSresults AS
SELECT AVG(p.duration) AS meanduration, p.raceid, p.driverid,
p.stopping, r.raceId AS raceid1, r.driverId AS driverid1, r.position
FROM pitstops p
INNER JOIN results r
ON p.raceid=r.raceId AND p.driverid=r.driverId
WHERE r.position IS NOT NULL
GROUP BY p.raceid, p.driverid;
```

2.) Assign relevant values as variables

```
SELECT @ax := avg(meanduration),
       @ay := avg(position),
       @div := (stddev_samp(meanduration) *
stddev_samp(position))
FROM pitstopsVSresults;
```

3.) Apply correlation coefficient equation

```
SELECT SUM( (meanduration - @ax) * (position - @ay) ) /
((count(meanduration) -1) * @div) FROM pitstopsVSresults;
```

Pitstop duration V.S. Position

r = -0.05

Qualifying duration V.S. Position

r = 0.03

Sprint duration V.S. Position

r = 0.17

→ None of them are determining factor!

Summary and Conclusion

- All analysis below lead to advices:
  - Best drivers/constructors
  - Not finishing race
  - Time taken per circuit
  - Big table with all info
  - Correlational analysis
- Many noted potential issue (with extension study) or surprising result

Use those data to make the best decision!

Thanks!

Any questions?