F1 METRICS : CASE STUDY

SQL-BASED ANALYSIS OF F1 RACE DYNAMICS

A. The total number of races conducted at each circuit, categorized by circuit name?

Here we considered Circuits and Races . The races are recorded from 1950 to 2018. With a data of approx 1000 races, I found that the top 6 racing venues were European and their trend of participation over the years has been consistent as compared to other venues.

```
SELECT
    r.name, COUNT(*) AS number_of_races, location
FROM
    races r
        JOIN
    circuits c ON r.circuit_id = c.circuitid
GROUP BY Name
ORDER BY number_of_races DESC;
```

Sample Query Result

name	number_of_races	location
British Grand Prix	69	Silverstone
Italian Grand Prix	69	Monza
Monaco Grand Prix	65	Monte-Carlo

B. How many races have drivers participated in over their careers?

Here , by joining driverstandings table and driver details, we can gauge the number race appearances individually. Only top 3 appeared in 300+ races.

```
SELECT
    dr.full_name, COUNT(ds.driver_id) AS Appearances
FROM
    driverstandings AS ds
        JOIN
    drivers AS dr ON ds.driver_id = dr.driver_id
GROUP BY ds.driver_id
ORDER BY Appearances desc;
```

Sample Query Result

full_name	Appearances
Jenson Button	316
Michael Schumacher	311
Rubens Barrichello	308

Insight :

Two of the top 3 drivers belong to Ferrari racing team, indicating stronger credibilty of the team on drivers to keep participating. In addition, attention to fitness regimen & starting early in the career helped. [1]

C. Which driver has won the most races in the history of the game ?

Here, by knowing the standings of the drivers as per the races, we can get the idea of races won accordingly. The outright dominance is enjoyed by the German driver Michael Schumacher , where the next follow up driver trails by almost 35% , even after Michael suffered a brain injury.

```
SELECT
    full_name, COUNT(full_name) AS total_wins, dr.driver_id
FROM
    driverstandings AS ds
        JOIN
    drivers dr ON ds.driver_id = dr.driver_id
WHERE
    position = 1
GROUP BY full_name
ORDER BY total_wins DESC;
```

Sample Query Result

full_name	Total_wins	driver_id
Michael Schumacher	121	30
Alain Prost	79	117
Lewis Hamilton	70	1

D. What is the distribution of race winners by nationality based on the constructors ?

The constructor results presented a new insight for us to see the constructor wins divided by nationality. The Italian and British lead the pack, whereas Japanese Honda is the first non European constructor to feature in the list.

```
SELECT
 Name, Nationality, COUNT(name) AS Wins
FROM
    constructorstandings AS cs
        JOIN
    constructors AS c ON cs.Constructor_id = c.Constructor_id
WHERE
    position = 1
GROUP BY name
ORDER BY Wins DESC;
```

name	Nationality	Wins
Ferrari	Italian	75
Williams	British	68
McLaren	British	60

Insight :

The Italian Ferrari's dominanace can be credited to its legendary 312T chassis from 70s and 1975 season. They kept modifying it and stayed competitive throughout.

For other European constructors more powerful engine and its advanced Michelin radial tyres could make their presence known.

@anikketsingh

E. Use a window function to calculate the percentage of total points distributed to each driver in each race.

Points scored by each driver in a race is tracked and then divided by the total points scored in that race.

```
SELECT
 Race_id, ds.driver_id, d.Full_name,
 CONCAT (ROUND((points/SUM(points)
  OVER  (PARTITION BY race_id) * 100),2), '%') AS
Percentage_of_points_scored
FROM driverstandings
   AS ds
JOIN
   drivers d
ON
   ds.driver_id = d.driver_id ;
```

Race_id	driver_id	Full_name	Percentage_of_points_scored
355	20	Sebastian Vettel	13.34%
355	4	Fernando Alonso	13.13%
355	1	Lewis Hamilton	12.51%

F. Use a CTE to calculate the number of races each driver participated in and the number of races they finished in the top 5 ? Then, use this CTE to find drivers with the highest percentage of top 5 finishes ?

```
WITH races_participated AS
 (
  SELECT dr.full_name, count(ds.driver_id) AS Participated
  FROM driverstandings
     AS ds
  JOIN drivers AS dr
     ON ds.driver_id = dr.driver_id
  GROUP BY ds.driver_id
),
top_5 AS
 (
  SELECT full_name, count(ds.driver_id) AS Times_in_top_5
  FROM driverstandings AS ds
  JOIN drivers AS dr
     ON ds.driver_id = dr.driver_id
  WHERE position >= 1 AND position <= 5
     GROUP BY ds.driver_id
)
```

```
SELECT
  t.Full_name, rp.Participated, t.Times_in_top_5,
 (ROUND((t.times_in_top_5 / rp.participated * 100),
 2))
AS Win_percentage
FROM
 races_participated AS rp
 JOIN
 top_5 AS t
 ON rp.full_name = t.full_name
 ORDER BY Win_percentage DESC;
```

Query Result

Full_name	Participated	Times_in_top_5	**Win_percentage**
Lewis Hamilton	207	189	91.30
Alain Prost	204	175	85.78
Jackie Stewart	104	86	82.69

G. Which circuits have hosted the most races won by a specific driver?

Here , we are considering the data available for ‘Michael Schumacher’ because he is the most successful driver at the moment with respect to our data.

```
WITH cu AS
(
SELECT year, name, ds.driver_id, d.full_name
FROM races rs
JOIN driverstandings ds
ON rs.Raceid = ds.race_id
jJOIN drivers d
ON d.driver_id = ds.driver_id
WHERE position = 1
)
SELECT
dense_rank() OVER(partition by full_name order by COUNT(name)
desc)
AS Positioned_Rank, name, COUNT(name) as Races_won
FROM cu
WHERE full_name = 'Michael Schumacher'
GROUP  BY name
```

Query Result

Positioned_Rank	**name**	**Races_won**
1	Monaco Grand Prix	8
1	Canadian Grand Prix	8
1	French Grand Prix	8