INTERMEDIATE SQL

WHERE are the subqueries?

Mona Khalil
Curriculum Lead, DataCamp

What is a subquery?

- A query nested inside another query

- Can be in any clause -- SELECT, FROM, WHERE, GROUP BY (or even CASE)

- Can contain scalar quantities, or multiple rows/columns

```
SELECT
    season,
    AVG(CASE WHEN hometeam_id = 9789 THEN home_goal END) AS dortmund_home,
    (SELECT AVG(home_goal) FROM match) AS overall_avg
FROM match
GROUP BY season;
```

season	dortmund_home	overall_avg
2011/2012	2.588235	1.554802
2012/2013	2.352941	1.554802
2013/2014	2.411765	1.554802
2014/2015	1.529412	1.554802

Why subqueries?

- Comparing groups to summarized values

- Reshaping data

- Combining data that cannot be joined

Simple Subqueries

- Can be evaluated independently from the outer query

```
SELECT AVG(home_goal) FROM match;
```

- Is only processed once in the entire statement

```
SELECT
    season,
    AVG(CASE WHEN hometeam_id = 9789
            THEN home_goal END) AS dortmund_home_avg,
    (SELECT AVG(home_goal) FROM match) AS overall
FROM match
GROUP BY season;
```

Subqueries in the WHERE clause

```sql
SELECT
    date,
    hometeam_id,
    awayteam_id,
    home_goal,
    away_goal
FROM match
WHERE
    season = '2012/2013'
    AND home_goal > (SELECT AVG(home_goal)
                     FROM match);
```

date	hometeam_id	awayteam_id	home_goal	away_goal
2012-07-28	9998	1773	5	2
2012-07-29	9987	9984	3	3
2012-10-05	9993	9991	2	2

Subqueries in WHERE with IN

```sql
SELECT
    team_long_name,
    team_short_name
FROM team
WHERE
    team_api_id IN
    (SELECT hometeam_id
     FROM match
     WHERE country_id = 15722);
```

team_long_name	team_short_name
Ruch Chorzów	CHO
Jagiellonia Białystok	BIA
Lech Poznań	POZ
P. Warszawa	PWA
Cracovia	CKR
Górnik Łęczna	LEC
Polonia Bytom	GOR
Zagłębie Lubin	ZAG
Pogoń Szczecin	POG
Widzew Łódź	WID
Śląsk Wrocław	SLA

INTERMEDIATE SQL

Practice Time!

INTERMEDIATE SQL

Subqueries in the FROM statement

Mona Khalil

Curriculum Lead, DataCamp

Subqueries in FROM

- Restructure and transform your data

- Calculating aggregates of aggregates

Building subqueries in the FROM clause

```
SELECT
    c.name,
    MAX(home_goal) AS max_home,
    MAX(away_goal) AS max_away
FROM country AS c
LEFT JOIN match AS m
ON c.id = m.country_id
GROUP BY c.name;
```

name	max_home	max_away
Belgium	6	7
England	8	6
France	6	7
Germany	9	7
Italy	7	7
Netherlands	7	6
Poland	6	6
Portugal	6	6
Scotland	6	6
Spain	9	8
Switzerland	6	6

Subqueries in the FROM clause

```sql
SELECT
    ROUND(AVG(max_home),2) AS avg_max_home,
    ROUND(AVG(max_away),2) AS avg_max_away
FROM
    (SELECT
        c.name,
        MAX(home_goal) AS max_home,
        MAX(away_goal) AS max_away
    FROM country AS c
    LEFT JOIN match AS m
    ON c.id = m.country_id
    GROUP BY c.name);
```

avg_max_home	avg_max_away
6.91	6.55

INTERMEDIATE SQL

Let's Practice!

INTERMEDIATE SQL

Subqueries in SELECT

Mona Khalil
Curriculum Lead, DataCamp

Subqueries in SELECT

```sql
SELECT
    season,
    COUNT(id) AS matches,
    (SELECT COUNT(id) FROM match) as total_matches
FROM match
GROUP BY season;
```

season	matches	total_matches
2011/2012	3220	12837
2012/2013	3260	12837
2013/2014	3032	12837
2014/2015	3325	12837

Subqueries in SELECT

- ## Without a subquery

```
SELECT
    date,
    home_goal + away_goal AS goals,
    home_goal + away_goal - 2.73 AS diff
FROM match
WHERE season = '2011/2012';
```

- ## With a subquery

```
SELECT
    date,
    home_goal + away_goal AS goals,
    (home_goal + away_goal) -
        (SELECT AVG(home_goal + away_goal) FROM match) AS diff
FROM match
WHERE season = '2011/2012';
```

INTERMEDIATE SQL

Let's Practice!

INTERMEDIATE SQL

Multiple Subqueries

Mona Khalil

Curriculum Lead, DataCamp

Multiple simple subqueries

- Watch your filters!

```
SELECT
    country_id,
    ROUND(AVG(m.home_goal + m.away_goal),2) AS avg_goals,
    (SELECT ROUND(AVG(home_goal + away_goal),2)
      FROM match WHERE season = '2013/2014') AS overall_avg
FROM match AS m
WHERE m.season = '2013/2014'
    AND (AVG(m.home_goal + m.away_goal) >
        (SELECT AVG(home_goal + away_goal)
         FROM match WHERE season = '2013/2014')
GROUP BY country_id;
```

Best Practices

- Consider the number of subqueries you include

- Clearly format and annotate your query

INTERMEDIATE SQL

Let's Practice!