INTERMEDIATE SQL

Welcome to Intermediate SQL!

Mona Khalil
Curriculum Lead, DataCamp

Topics covered:

- CASE statements

- Simple subqueries

- Correlated subqueries

- Window functions

Prerequisites

- Intro to SQL for Data Science

```sql
SELECT user_id, SUM(sales)
FROM sales_data
WHERE user_id BETWEEN 300 AND 400
GROUP BY user_id;
```

- Joining Tables in PostgreSQL

```sql
SELECT c.country, c.team, SUM(m.goals)
FROM countries AS c
LEFT JOIN matches AS m
ON c.team_id = m.home_team_id
WHERE m.year > 1990
GROUP BY c.country, c.team;
```

Selecting from the European Soccer Database

```
SELECT
    l.name AS league,
    COUNT(m.country_id) as matches
FROM
    league AS l
LEFT JOIN match AS m
ON l.country_id = m.country_id
GROUP BY l.name;
```

league	total_matches
Belgium Jupiler League	732
England Premier League	1520
France Ligue 1	1520
Germany 1. Bundesliga	1224

Selecting from the European Soccer Database

```
SELECT
    date,
    id,
    home_goal,
    away_goal
FROM match
WHERE
    season = "2013/2014";
```

date	id	home_goal	away_goal
2014-03-29 00:00:00	1237	2	0
2014-03-29 00:00:00	1238	0	1
2014-04-05 00:00:00	1239	1	0
2014-04-05 00:00:00	1240	0	0

Selecting from the European Soccer Database

```sql
SELECT
    date,
    id,
    home_goal,
    away_goal
FROM match
WHERE
    season = "2013/2014"
    AND home_team_goal > away_team_goal;
```

date	id	home_goal	away_goal
2014-03-29 00:00:00	1237	2	0
2014-04-05 00:00:00	1239	1	0
2014-04-12 00:00:00	1241	2	1
2014-04-12 00:00:00	1242	2	0

CASE statements

- Contains a WHEN, THEN, and ELSE statement, finished with END

```
CASE WHEN x = 1 THEN "a"
     WHEN x = 2 THEN "b"
     ELSE "c" END AS new_column
```

CASE WHEN

```sql
SELECT
    id,
    home_goal,
    away_goal,
    CASE WHEN home_goal > away_goal THEN "Home Team Win"
         WHEN home_goal < away_goal THEN "Away Team Win"
         ELSE "Tie" END AS outcome
FROM match
WHERE
    season = "2013/2014";
```

id	home_goal	away_goal	outcome
1237	2	0	Home Team Win
1238	0	1	Away Team Win
1239	1	0	Home Team Win
1240	0	0	Tie

INTERMEDIATE SQL

Let's Practice!

INTERMEDIATE SQL

CASE WHEN with multiple conditions

Mona Khalil

Curriculum Lead, DataCamp

CASE WHEN ... and then some!

```sql
SELECT
    date,
    season,
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
        THEN 'Chelsea home win!'
        WHEN  awayteam_id = 8455 AND home_goal < away_goal
        THEN 'Chelsea away win!'
        ELSE 'Loss or tie :(' END AS outcome
FROM match
WHERE
    hometeam_id = 8455 OR awayteam_id = 8455;
```

date	season	outcome
2011-08-14	2011/2012	Loss or tie :(
2011-11-05	2011/2012	Chelsea away win!
2011-11-20	2011/2012	Loss or tie :(
2011-11-26	2011/2012	Chelsea home win!

CASE WHEN things to watch out for

- What's in your `ELSE` clause?

```
SELECT
date,
season,
CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
        THEN 'Chelsea home win!'
    WHEN awayteam_id = 8455 AND home_goal < away_goal
        THEN 'Chelsea away win!'
    ELSE 'Loss or tie :(' END AS outcome
FROM match;
```

- Non-Chelsea matches aren't filtered out

- All losses, ties, and non-Chelsea matches count as "Loss or tie"

CASE WHEN best practices

```
SELECT
    date,
    season,
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
        THEN 'Chelsea home win!'
        WHEN awayteam_id = 8455 AND home_goal < away_goal
        THEN 'Chelsea away win!'
        ELSE 'Loss :(' END AS outcome
FROM match
WHERE
    hometeam_id = 8455 OR awayteam_id = 8455;
```

CASE WHEN best practices

```
SELECT
    date,
    season,
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
            THEN 'Chelsea home win!'
        WHEN awayteam_id = 8455 AND home_goal < away_goal
            THEN 'Chelsea away win!'
        WHEN awayteam_id = 8455 AND home_goal = away_goal
            THEN 'Tie'
        ELSE "Not Chelsea" END AS outcome
FROM match
WHERE
    hometeam_id = 8455
    OR awayteam_id = 8455;
```

CASE WHEN and NULLs

- What are your NULL values doing?

```
SELECT
    date,
    season,
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
        THEN 'Chelsea home win!'
        WHEN awayteam_id = 8455 AND home_goal < away_goal
        THEN 'Chelsea away win!'
        END AS outcome
FROM match;
```

date	season	outcome
2012-01-02	2011/2012	Chelsea away win!
2012-01-14	2011/2012	NULL
2012-01-14	2011/2012	NULL
2012-01-14	2011/2012	Chelsea home win!

CASE WHEN and NULLs

```sql
SELECT
    date,
    season,
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
            THEN 'Chelsea home win!'
         WHEN awayteam_id = 8455 AND home_goal < away_goal
            THEN 'Chelsea away win!' END AS outcome
FROM match
WHERE
    CASE WHEN hometeam_id = 8455 AND home_goal > away_goal
            THEN 'Chelsea home win!'
         WHEN awayteam_id = 8455 AND home_goal < away_goal
            THEN 'Chelsea away win!' END
        IS NOT NULL;
```

date	season	outcome
2011-11-05	2011/2012	Chelsea away win!
2011-11-26	2011/2012	Chelsea home win!
2011-12-03	2011/2012	Chelsea away win!

INTERMEDIATE SQL

Let's Practice!

INTERMEDIATE SQL

CASE WHEN with aggregate functions

Mona Khalil

Curriculum Lead, DataCamp

CASE WHEN with SUM

```
SELECT
    season,
    SUM(CASE WHEN hometeam_id = 8634
        THEN home_goal END) AS barca_homegoals,
    SUM(CASE WHEN awayteam_id = 8634
        THEN away_goal END) AS barca_awaygoals
FROM match
GROUP BY season;
```

season	barca_homegoals	barca_awaygoals
2011/2012	73	41
2012/2013	63	52
2013/2014	64	36
2014/2015	64	46

CASE WHEN with COUNT

```
SELECT
    season,
    COUNT(CASE WHEN hometeam_id = 8634 AND home_goal > away_goal
        THEN id END) AS home_wins,
    COUNT(CASE WHEN awayteam_id = 8634 AND away_goal > home_goal
        THEN id END) AS away_wins
FROM match
GROUP BY season;
```

season	home_wins	away_wins
2011/2012	17	11
2012/2013	18	14
2013/2014	16	11
2014/2015	16	14

COUNTing with CASE WHEN and SUM

```sql
SELECT
    season,
    SUM(CASE WHEN hometeam_id = 8634 AND home_goal > away_goal
        THEN 1 ELSE 0 END) AS home_wins,
    SUM(CASE WHEN awayteam_id = 8634 AND away_goal > home_goal
        THEN 1 ELSE 0 END) AS away_wins
FROM match
GROUP BY season;
```

season	home_wins	away_wins
2011/2012	17	11
2012/2013	18	14
2013/2014	16	11
2014/2015	16	14

CASE WHEN with AVG

```
SELECT
    season,
    AVG(CASE WHEN hometeam_id = 8634
        THEN home_goal END) AS barca_home_goals,
    AVG(CASE WHEN awayteam_id = 8634
        THEN away_goal END) AS barca_away_goals
FROM match
GROUP BY season;
```

season	barca_homegoals	barca_awaygoals
2011/2012	3.84	2.16
2012/2013	3.32	2.74
2013/2014	3.37	1.89
2014/2015	3.37	2.42

Percentages with CASE and AVG

```sql
SELECT
    season,
    AVG(CASE WHEN hometeam_id = 8634 AND home_goal > away_goal THEN 1
             WHEN hometeam_id = 8634 AND home_goal < away_goal THEN 0
             END) AS pct_homewins,
    AVG(CASE WHEN awayteam_id = 8634 AND away_goal > home_goal THEN 1
             WHEN awayteam_id = 8634 AND away_goal < home_goal THEN 0
             END) AS pct_awaywins
FROM match
GROUP BY season;
```

season	pct_homewins	pct_awaywins
2011/2012	0.944444	0.846153
2012/2013	1	0.875
2013/2014	0.941176	0.733333
2014/2015	0.888888	0.875

Percentages with CASE and AVG!

```sql
SELECT
    season,
    ROUND(AVG(CASE WHEN hometeam_id = 8634 AND home_goal > away_goal THEN 1
              WHEN hometeam_id = 8634 AND home_goal < away_goal THEN 0
              END)*100,2) AS pct_homewins,
    ROUND(AVG(CASE WHEN awayteam_id = 8634 AND away_goal > home_goal THEN 1
              WHEN awayteam_id = 8634 AND away_goal < home_goal THEN 0
              END)*100,2) AS pct_awaywins
FROM match
GROUP BY season;
```

season	pct_homewins	pct_awaywins
2011/2012	94.44	84.62
2012/2013	100	87.5
2013/2014	94.12	73.33
2014/2015	88.89	87.5

INTERMEDIATE SQL

Let's Practice!