Advanced SQL

⏪ Recap

- `SELECT` columns `FROM` a table `WHERE` conditions are true
  - `LIKE` pattern matching operator (with `%` and `_`)
  - `IN` membership operator
  - `BETWEEN` range operator
- `LIMIT` the number of records returned
- `ORDER BY` columns

⏪ **Recap**

```sql
select * from person
-- suspect person_id from first witness
where id = 67318
-- suspect person_id from second witness
and id in (67318, 28819)
```

```
select * from interview where person_id=67318
```

I was hired by a woman with a lot of money. I don't know her name but I know she's around 5'5" (65") or 5'7" (67"). She has red hair and she drives a Tesla Model S. I know that she attended the SQL Symphony Concert 3 times in December 2017.

Clue #1: Find the person who matches the description

I don't know her name but I know she's around 5'5" (65") or 5'7" (67"). She has red hair and she drives a Tesla Model S.

Clue #1: Find the person who matches the description

```sql
select *
from drivers_license
where hair_color = 'red'
and car_make = 'Tesla'
and car_model = 'Model S'
and height between 65 and 67
```

WHERE column **IN** (list of values)

```
select *
from person
where license_id in (918773, 291182, 202298)
```

Subquery: query inside another query

WHERE IN (subquery to return a list of values)

```sql
select *
from person
where license_id in (918773, 291182, 202298)
```

```sql
select *
from person
where license_id in (
    -- subquery that returns the same list of values
    select id
    from drivers_license
    where hair_color = "red"
    and car_make = "Tesla"
    and car_model = "Model S"
    and height between 65 and 67
)
```

9

The number of columns in subquery must match the number of columns in WHERE clause

```sql
select *
from person
where (license_id, name) in (
    -- subquery that returns the same list of values
    select id, name -- must match the number of columns in WHERE clause
    from drivers_license
    where hair_color = "red"
    and car_make = "Tesla"
    and car_model = "Model S"
    and height between 65 and 67
)
```

1. Pull up the interview of the first witness, who lives at the last house on "Northwestern Dr."**

1. Pull up the interview of the first witness, who lives at the last house on "Northwestern Dr."

```sql
select * from interview
where person_id = (
    select id from person
    where address_street_name = "Northwestern Dr"
    order by address_number DESC limit 1
    )
```

🖥️ 2. Find the first suspect from person table: membership id starts with "48Z" and license plate includes "H42W"

🖥️ 3. Pull up the interview of the second witness. Her name is Annabel and she lives somewhere on "Franklin Ave".

Find the second suspect

4. Membership id of the second witness (Annabel, Franklin Ave)

5. Check in time and check out time of the second witness

6. Find members who were at the gym during the same time as the second witness

```sql
select * from get_fit_now_member
where id in (
    select membership_id from get_fit_now_check_in
    where check_in_time <= (
        select check_out_time from get_fit_now_check_in
        where membership_id = (
            select id from get_fit_now_member
            where person_id = (select id from person where address_street_name = "Franklin Ave" and name like "%Annabel%")
            )
        )
    and check_out_time >= (
        select check_in_time from get_fit_now_check_in
        where membership_id = (
            select id from get_fit_now_member
            where person_id = (select id from person where address_street_name = "Franklin Ave" and name like "%Annabel%")
            )
        )
    )
```

Find the killer

7. From first witness: gym membership id starts with "48Z" and license plate includes "H42W"

8. From second witness

9. Cross-reference the person IDs of the suspects in the `person` table

```sql
select * from person
-- from first witness
where id in (select person_id from get_fit_now_member where id like "%48Z%")
and license_id in (select id from drivers_license where plate_number like "%H42W%")
-- from second witness
and id in (
    select person_id from get_fit_now_member
    where id in (
        select membership_id from get_fit_now_check_in
        where check_in_time <= 1700
        and check_out_time >= 1600
        )
    )
```

Clue #2

I know that she attended the SQL Symphony Concert 3 times in December 2017.

Table: `facebook_event_checkin`

Columns: all

Rows:

- SQL Symphony Concert

- December 2017

```
select *
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
order by person_id
```

Aggregation with functions

- `count()`
- `sum()`
- `avg()`
- `max()`
- `min()`
- `stddev()` : standard deviation
- `variance()` : variance

  ...

```sql
select count(*)
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
order by person_id
```

`select` column `from` table `group by` column

GROUP BY to aggregate data at a group level

```sql
select person_id, count(*)
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
group by person_id
```

HAVING to filter groups

```
select person_id, count(*)
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
group by person_id
having count(*)=3
```

25

Alias `AS`

```sql
select person_id, count(*) as num_attendance, count(*) num_attendance2
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
group by person_id
having num_attendance = 3
```

ORDER BY aggregated value

Who attended SQL Symphony Concert the most in 2017?

```sql
select person_id, count(*)
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
group by person_id
order by count(*) desc
```

DISTINCT to remove duplicates

How many times SQL Symphony Concert was held in 2017?

```sql
select count(distinct date)
from facebook_event_checkin
where event_name like "%SQL Symphony Concert%"
and date between 20170101 and 20171231
```

Facebook Event Checkin

1. When was the first event recorded?

2. how many people attended each event?

3. Who attended the most events in 2017?

🖥️ **List all the events that were held more than three times in 2017. What was the most popular event in 2017?**

1. **Clue #1: People that match the description (** `license_id` **from** `drivers_license` **)**

2. **Clue #2: People that attended SQL Symphony Concert 3 times in 2017 (** `person_id` **from** `facebook_event_checkin` **)**
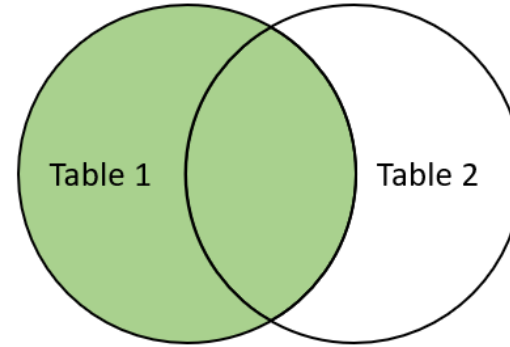
Using Subquery

```sql
select *
from person
where license_id in (
    select id
    from drivers_license
    where hair_color = "red"
    and car_make = "Tesla"
    and car_model = "Model S"
    and height between 65 and 67
)
and id in (
    select person_id
    from facebook_event_checkin
    where event_name like "%SQL Symphony Concert%"
    and date between 20170101 and 20171231
    group by person_id
    having count(*)=3
)
```
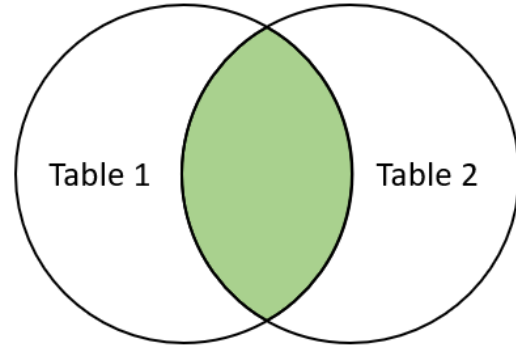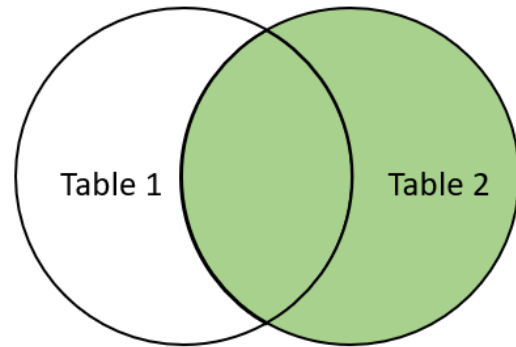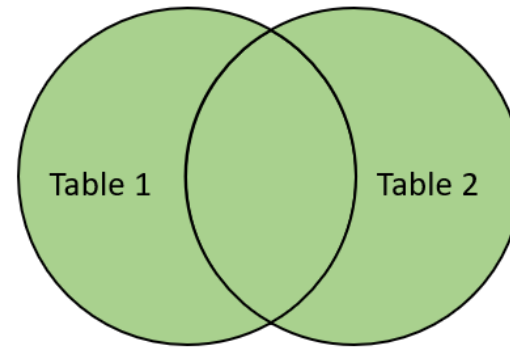
Using `JOIN`

`JOIN ON` (inner join)

- What tables do you want to join?: `person` and `drivers_license`

- What column(s) do you want to join on?: `license_id` from `person` and `id` from `drivers_license`

- Notice `.` to specify which table a column belongs to

```
select person.id, drivers_license.id
from (person join drivers_license on person.license_id = drivers_license.id)
```

What column(s) do you want to join on?

1. Join `person` and `drivers_license` on?

2. Join `person` and `facebook_event_checkin` on?

Join person and drivers_license

```sql
select *
from person
join drivers_license on person.license_id = drivers_license.id
where hair_color = "red"
and car_make = "Tesla"
and car_model = "Model S"
and height between 65 and 67
```

Join **person**, **drivers_license**, and subquery **facebook_event_checkin**

```sql
select *
from person
join drivers_license on person.license_id = drivers_license.id
where hair_color = "red"
and car_make = "Tesla"
and car_model = "Model S"
and height between 65 and 67
and person.id in (
    select person_id
    from facebook_event_checkin
    where event_name like "%SQL Symphony Concert%"
    and date between 20170101 and 20171231
    group by person_id
    having count(*)=3
)
```

Other `JOIN` syntaxes

JOIN USING :

```sql
select * from person join drivers_license using (license_id)
```

NATURAL JOIN :

```sql
select * from person natural join drivers_license
```

WHERE :

```sql
select * from person, drivers_license
where person.license_id = drivers_license.id
```

LEFT JOIN and RIGHT JOIN

left join:

```
select * from person left join get_fit_now_member on person.id = get_fit_now_member.person_id
```

right join:

```
select * from person right join get_fit_now_member on person.id = get_fit_now_member.person_id
```