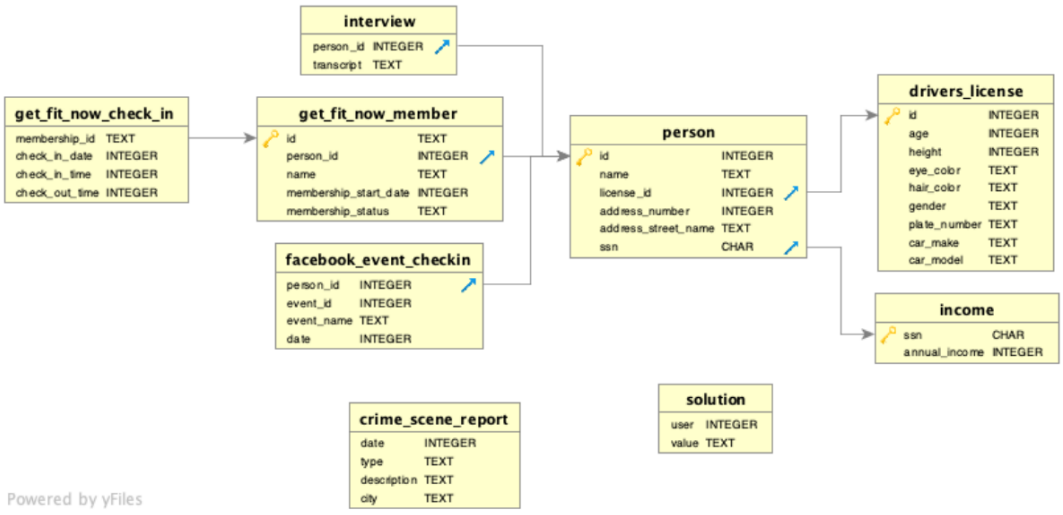
Report on SQL Murder Mystery

PrepInsta Winter Internship on Data Analytics - Week 5 Project

Link to Website: https://mystery.knightlab.com/

Problem Statement:
A crime has taken place and the detective needs our help. The detective has given the crime scene report, but it's now lost. We have information that the crime occurred sometime on January 15, 2018, and took place in SQL City.

Here is the schema diagram to show us the available tables and their relationships within the database:



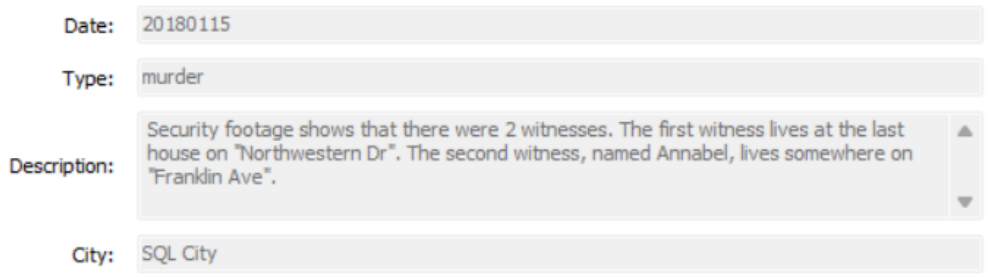
Approach:

- I filtered the database, extracting specifics—crime type, city, and date—to efficiently grasp the crime scene's scope. This focused approach aimed to streamline the investigation's relevant data retrieval.

  Query:

```
# Murder occured on 15th Jan 2018 in SQL City
USE murder_mystery;
SELECT *
FROM crime_scene_report
WHERE date = "20180115"
AND type = "murder"
AND city = "SQL City";
```

Output:

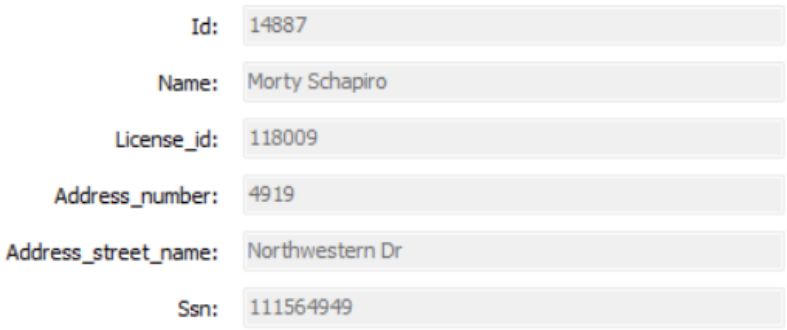


- To locate the first witness, I narrowed down residences on "Northwestern Dr" by querying the street name and using DESC to identify the highest house number among the listings.

  Query:

```
# 2 witnesses
-- 1st witness
SELECT *
FROM person
WHERE address_street_name = "Northwestern Dr"
ORDER BY address_number DESC
LIMIT 1;
```

  Output:



- To retrieve details about the second witness, Annabel, I executed a query on the "person" table specifically targeting Annabel's information.

  Query:

```
-- 2nd witness
SELECT *
FROM person
WHERE address_street_name = "Franklin Ave"
AND name LIKE '%Ann%';
```

Output:

- With narrowed information on the two witnesses, I proceeded to delve deeper by querying the "interview" table using their respective IDs for a more comprehensive understanding.

  Query:

```
# Finding the two witness' interview transcript
SELECT *
FROM interview
WHERE person_id IN (14887, 16371);
```

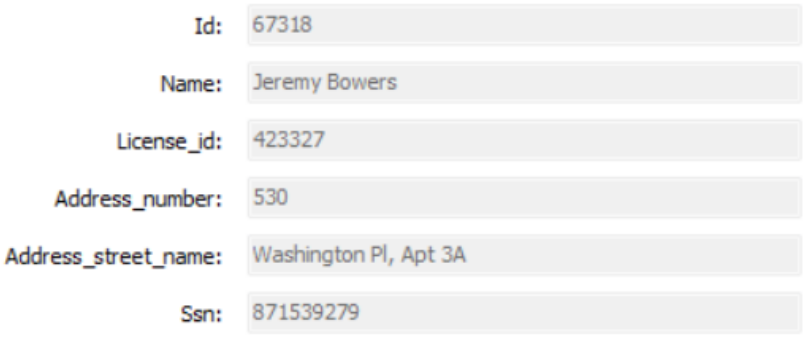
  Output:

- Utilizing an INNER JOIN operation across three tables, I discerned the murderer as a Get Fit Now Gym gold member, their membership initiating with "48Z," and owning a vehicle plate containing "H42W." This meticulous correlation unveiled the perpetrator's identity.

  Query:

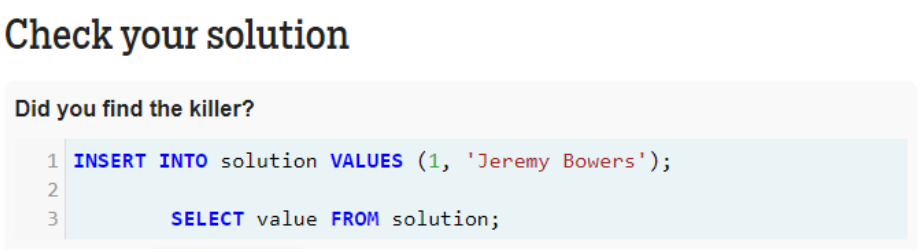
```
# "Get Fit Now Gym" membership started with "48Z" Gold member
# Car Plate includes "H42W"
# Murderer was in gym on 9th Jan
SELECT p.*
FROM drivers_license AS dl
INNER JOIN person AS p ON dl.id = p.license_id
INNER JOIN get_fit_now_member AS member ON p.id = member.person_id
INNER JOIN get_fit_now_check_in AS ci ON member.id = ci.membership_id
WHERE plate_number LIKE '%H42W%'
AND check_in_date = "20180109";
```

Output:

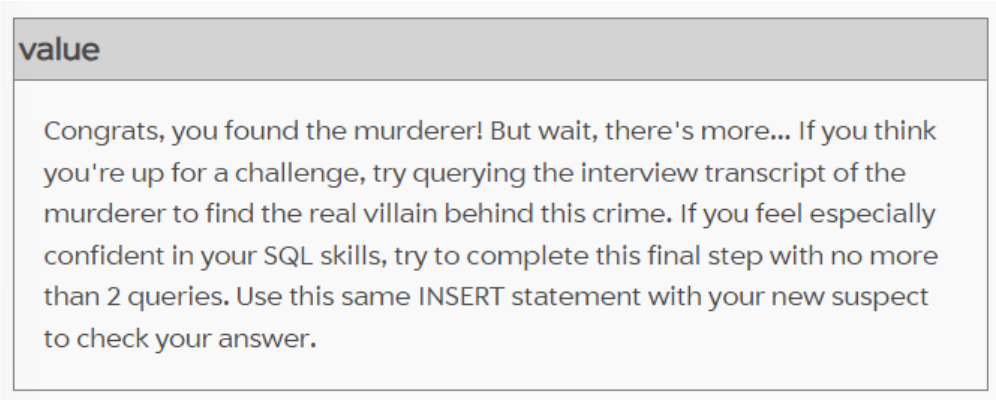


- Verifying the outcome, I cross-checked the suspect's name by inserting "Jeremy Bowers" into the query below, confirming his identification as the perpetrator.

  Query:



  Output:



- To uncover the mastermind, I sought Jeremy's interview by referencing his ID, aiming to extract further details and potentially reveal the orchestrator behind the murderer's actions.

  Query:

```
SELECT *
FROM interview
WHERE person_id = "67318";
```

Output:

Person_id:	67318
Transcript:	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.

- To pinpoint the mastermind, I constructed a complex query utilizing multiple criteria based on Jeremy's information: a woman, height between 65" to 67", red hair, owner of a Tesla Model S, and attended the SQL Symphony Concert thrice in December 2017. This intricate query aims to narrow down the search and reveal the individual behind the murder.

  Query:

```sql
# Rich woman, height 5'5" (65") or 5'7" (67"), red hair, Tesla Model S car, attended SQL Symphony Concert 3 times in Dec 2017
WITH CTE AS (
SELECT person_id,
COUNT(*) AS visits
FROM facebook_event_checkin
WHERE date BETWEEN 20171201 AND 20171231
AND event_name = "SQL Symphony Concert"
GROUP BY person_id
HAVING COUNT(*) = 3
)

SELECT p.*, fb.*
FROM drivers_license AS dl
INNER JOIN person AS p ON dl.id = p.license_id
INNER JOIN CTE AS fb ON p.id = fb.person_id
WHERE height BETWEEN 65 AND 67
AND hair_color = "red"
AND car_make = "Tesla"
AND car_model = "Model S"
AND gender = "female";
```

  Output:

Id:	99716
Name:	Miranda Priestly
License_id:	202298
Address_number:	1883
Address_street_name:	Golden Ave
Ssn:	987756388
Person_id:	99716
Visits:	3

- Now, let's verify the outcome.

Query:

Check your solution

Did you find the killer?

```
1 INSERT INTO solution VALUES (1, 'Miranda Priestly');
2
3         SELECT value FROM solution;
```

Output:

value
Congrats, you found the brains behind the murder! Everyone in SQL City hails you as the greatest SQL detective of all time. Time to break out the champagne!

The culprit behind the crime was Miranda Priestly, who had employed Jeremy for the nefarious act.

-By Sakshi Choudhary.