SQL Murder Mystery

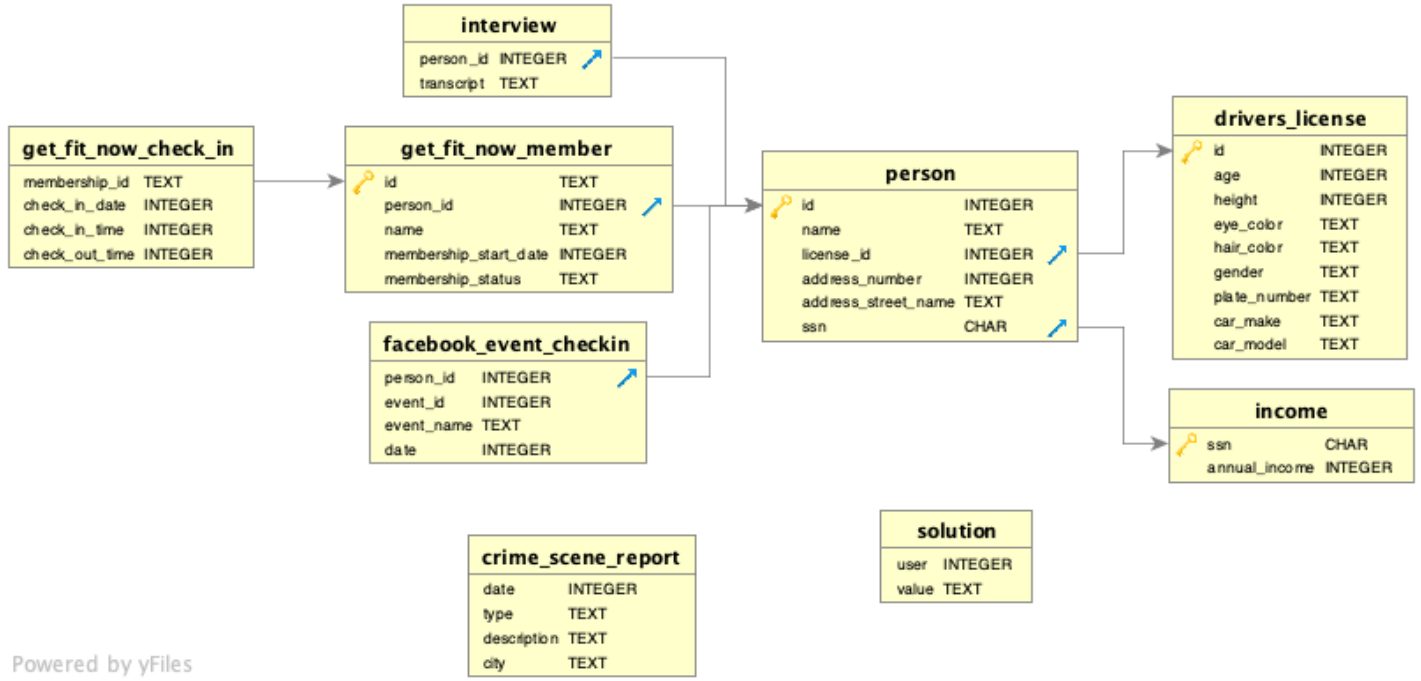
A project completed by Anna Zinkovich, Data Analyst
Contacts: https://www.linkedin.com/in/anna-zinkovich/
Other projects: https://github.com/Anna-Zinkovich?tab=repositories

The SQL Murder Mystery is a game that is aimed to practice SQL concepts and commands while experienced SQL users can work on solving a crime.
More information: https://mystery.knightlab.com/

INTRO

A crime has taken place and the detective needs your help. The detective gave you the crime scene report, but you somehow lost it. You vaguely remember that the crime was a murder that occurred sometime on Jan.15, 2018 and that it took place in SQL City. Start by retrieving the corresponding crime scene report from the police department's database.

EXPLORING THE DATABASE STRUCTURE



SOLUTION

We know that the murder occurred on January 15 2018 in SQL City.
So, let's start by pulling up the crime report from the crime_scene_report dataset:

QUERY:

```
SELECT *
FROM crime_scene_report
WHERE city = 'SQL City' AND date = 20180115;
```

RESULT:

date	type	description
20180115	assault	Hamilton: Lee, do you yield? Burr: You shot him in the side! Yes he yields!
20180115	assault	Report Not Found
20180115	murder	Security footage shows that there were 2 witnesses. The first witness lives at the last house on "Northwestern Dr". The second witness, named Annabel, lives somewhere on "Franklin Ave".

Now we have some information on the witnesses and can retrieve data from other tables!
Let's start with Annabel who lives on Franklin Ave. It turned out that there's only one Annabel there:

QUERY:

```
SELECT *
FROM person
WHERE address_street_name LIKE 'Franklin Ave' AND name LIKE "%Annabel%";
```

RESULT:

id	name	license_id	address_number	address_street_name	ssn
16371	Annabel Miller	490173	103	Franklin Ave	318771143

Now let's see if she was investigated by the police. Let's find her in the interview database using her person_id:

QUERY:

```
SELECT *
FROM interview
WHERE person_id = 16371;
```

RESULT:

person_id	transcript
16371	I saw the murder happen, and I recognized the killer from my gym when I was working out last week on January the 9th.

So, Annabel knows the murderer because she saw this person on January 9.
Now we need a list of people who went to the gym on Jan 9 2018:

QUERY:

```
SELECT *
FROM get_fit_now_check_in
WHERE check_in_date = 20180109;
```

RESULT:

membership_id	check_in_date	check_in_time	check_out_time
X0643	20180109	957	1164
UK1F2	20180109	344	518

XTE42	20180109	486	1124
1AE2H	20180109	461	944
6LSTG	20180109	399	515
7MWHJ	20180109	273	885
GE5Q8	20180109	367	959
48Z7A	20180109	1600	1730
48Z55	20180109	1530	1700
90081	20180109	1600	1700

For now, let's move on to the second witness.
We know that this person lives at the last house on "Northwestern Dr". So, let's pull up everyone who lives on that street from the person database but order results by address_number column in descending order (so that we have the last house first).

QUERY:

```
SELECT *
FROM person
WHERE address_street_name LIKE 'Northwestern Dr'
ORDER BY address_number DESC;
```

RESULT:

id	name	license_id	address_number	address_street_name	ssn
14887	Morty Schapiro	118009	4919	Northwestern Dr	111564949

Let's see if Morty was interviewed by the police and if we can find his testimony in the interview database:

QUERY:

```
SELECT *
FROM interview
WHERE person_id = 14887;
```

RESULT:

person_id	transcript
14887	I heard a gunshot and then saw a man run out. He had a "Get Fit Now Gym" bag. The membership number on the bag started with "48Z". Only gold members have those bags. The man got into a car with a plate that included "H42W".

Thank you, Morty. This is very helpful.

So, we know that the killer is a male with the gold membership and id that starts with "48Z". Referring to the list of gym visitors (2 tables above) from Jan. 9 2018 (when first witness, Annabel, saw the killer), we see that there were only 2 people whose id started with "48Z": 48Z7A and 48Z55.

Let's see who they are in get_fit_now_member database.

QUERY:

```
SELECT *
FROM get_fit_now_member
WHERE id = "48Z7A" OR id = "48Z55";
```

RESULT:

id	person_id	name	membership_start_date	membership_status
48Z55	67318	Jeremy Bowers	20160101	gold
48Z7A	28819	Joe Germuska	20160305	gold

Let's find if any of them have been investigated by the police:

QUERY:

```
SELECT *
FROM interview
WHERE person_id = 67318 OR person_id = 28819;
```

RESULT:

person_id	transcript
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.

So, the first one, Jeremy Bowers, is the actual shooter. However, we still have to find the instigator of killing.

Lets' find all attendees of the SQL Symphony Concert in December 2017 from facebook_event_checkin database and order them by person_id so that it's easier to see how many times a unique user visited this concert during this timeframe:

QUERY:

```
SELECT *
FROM facebook_event_checkin
WHERE date BETWEEN 20171201 AND 20171231
AND event_name = "SQL Symphony Concert"
ORDER BY person_id;
```

RESULT:

person_id	event_id	event_name	date
11173	1143	SQL Symphony Concert	20171223
19260	1143	SQL Symphony Concert	20171214
19292	1143	SQL Symphony Concert	20171213
24397	1143	SQL Symphony Concert	20171208

24556	**1143**	**SQL Symphony Concert**	**20171207**
24556	**1143**	**SQL Symphony Concert**	**20171221**
24556	**1143**	**SQL Symphony Concert**	**20171224**
28582	1143	SQL Symphony Concert	20171220
28582	1143	SQL Symphony Concert	20171215
43366	1143	SQL Symphony Concert	20171207
58898	1143	SQL Symphony Concert	20171220
62596	1143	SQL Symphony Concert	20171225
67318	1143	SQL Symphony Concert	20171206
69325	1143	SQL Symphony Concert	20171206
69699	1143	SQL Symphony Concert	20171214
79312	1143	SQL Symphony Concert	20171203
81526	1143	SQL Symphony Concert	20171202
92343	1143	SQL Symphony Concert	20171212
99716	**1143**	**SQL Symphony Concert**	**20171206**
99716	**1143**	**SQL Symphony Concert**	**20171212**
99716	**1143**	**SQL Symphony Concert**	**20171229**

We found that people with ids 24556 and 99716 went to the SQL Symphony Concert 3 times in December 2017. Let's see who they are:

QUERY:

```
SELECT *
FROM person
WHERE id = 24556 OR id = 99716;
```

RESULT:

id	name	license_id	address_number	address_street_name	ssn
24556	Bryan Pardo	101191	703	Machine Ln	816663882
99716	Miranda Priestly	202298	1883	Golden Ave	987756388

We know that the crime was ordered by a woman. So, let's pull up all the information we can find about her. There's no data on her in interview and get_fit_now_member databases. Let's check other databases:

QUERY:

```
SELECT *
FROM person
JOIN drivers_license ON person.license_id = drivers_license.id
JOIN income ON person.ssn = income.ssn
WHERE person.id = 99716;
```

RESULT:

Since the result table was rather wide, I split it in 3 parts:

Part 1

id	name	license_id	address_number	address_street_name	ssn
99716	Miranda Priestly	202298	1883	Golden Ave	987756388

Part 2

id	age	height	eye_color	hair_color	gender
202298	68	66	green	red	female

Part 3

plate_number	car_make	car_model	ssn	annual_income
500123	Tesla	Model S	987756388	310000

This woman fits perfectly to the description of Jeremy Bowers (the shooter):
- her height is between 65”-67”
- hair color is red
- she drives Tesla Model S
- her annual income of $310K seem like a high income for me.

Let’s check if it’s correct in the solution table:

QUERY:

```
INSERT INTO solution VALUES (1, 'Miranda Priestly');

    SELECT value FROM solution;
```

RESULT:

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!

And just like that using the wonderful SQL we helped to solve the crime :)
Thank you for taking the time to check out my project.

Anna Zinkovich,
Data Analyst