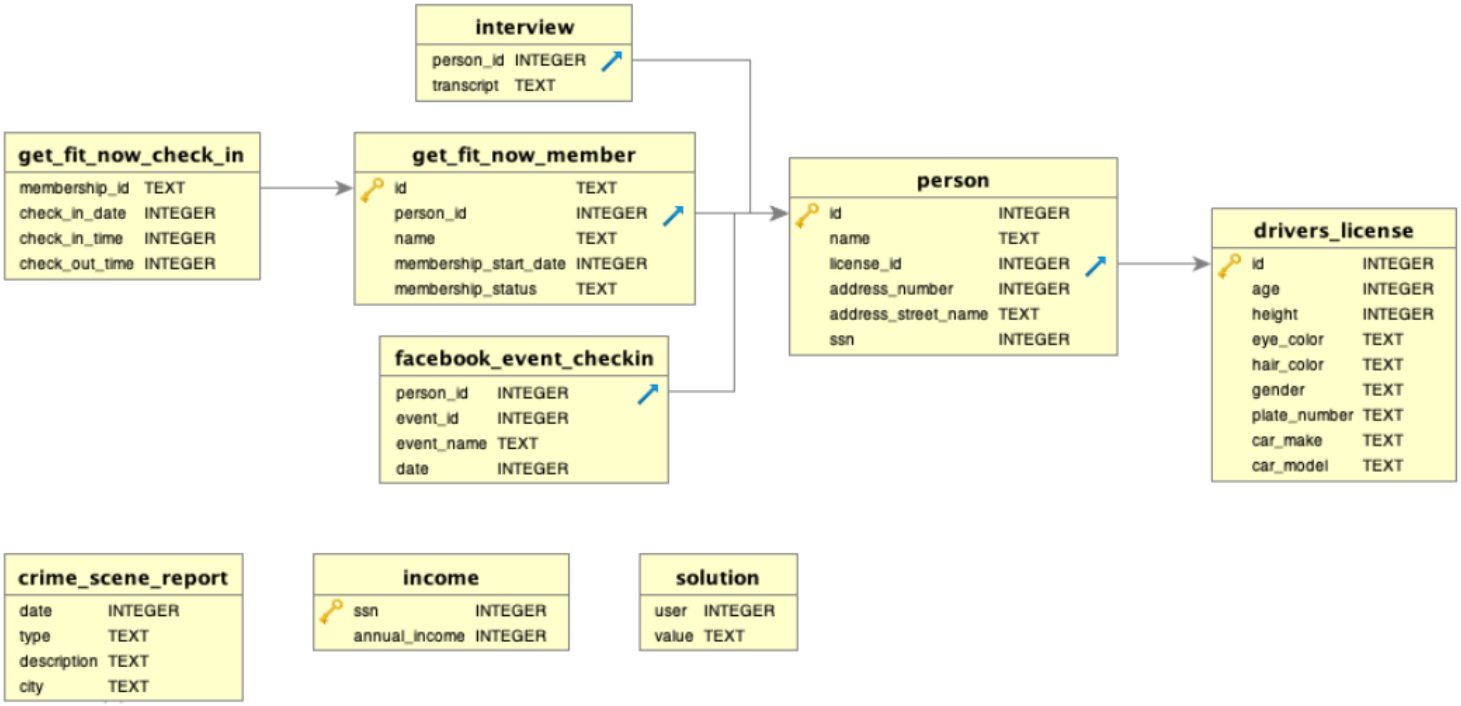
SQL Murder Mystery notes

https://mystery.knightlab.com/

Starting the challenge, here's the first important tidbit of info I'm given:

"You vaguely remember that the crime was a **murder** that occurred sometime on **Jan.15, 2018** and that it took place in **SQL City**."

I'm also given the schema diagram for the relational databases:



Knowing that the murder took place on Jan 15 2018 in SQL City, I look for the crime scene report

```
select * from crime_scene_report
where date = 20180115
        and city = 'SQL City';
```

"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"."

I decide to look into the first witness. So I'll need to find the last house on Northwestern Dr:

```
select MAX(address_number)
from person
where address_street_name = "Northwestern Dr";
```

4919 is the highest address number on the street.

So going into the person table, let's see who lives at 4919 Northwestern Dr.

```
select *
from person
where address_street_name = "Northwestern Dr"
and address_number = 4919;
```

This must be our first witness:

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

Let's find their testimony in their interview

```
select *
from interview
where person_id = 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"."

There's a table of Get Fit Now's members that has their membership status and member id, so let's see who meets that description.

```
select *
from get_fit_now_member
where membership_status = 'gold'
and id like '48Z%';
```

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

Now running for info based on "*a plate that included "H42W"*"

```
select *
from drivers_license
where plate_number like '%H42W%';
```

eliminate female based on witness description.

id	age	height	eye_color	hair_color	gender	plate_number	car_make	car_model
~~183779~~	~~21~~	~~65~~	~~blue~~	~~blonde~~	~~female~~	~~H42W0X~~	~~Toyota~~	~~Prius~~
423327	30	70	brown	brown	male	0H42W2	Chevrolet	Spark LS

664760	21	71	black	black	male	4H42WR	Nissan	Altima

Looking up the two names that matched the gym membership info

```
select *
from person
where name = 'Joe Germuska'
or name = "Jeremy Bowers";
```

id	name	license_id	address_number	address_street_name	ssn
28819	Joe Germuska	173289	111	Fisk Rd	138909730
67318	Jeremy Bowers	423327	530	Washington Pl, Apt 3A	871539279

Jeremy Bowers is a license plate match and a membership bag match. Suspect #1. Let's see what he was up to that day.

```
select *
from facebook_event_checkin
where person_id = 67318
and date = 20180115;
```

person_id	event_id	event_name	date
67318	4719	The Funky Grooves Tour	20180115

(That wasn't very helpful) Hmm... did he do an interview?

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

He did!

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.

All the info I've highlighted in green for Suspect #2 can be found in drivers_license table.

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

If I can take the *5'5" (65") or 5'7" (67")* literally, I could narrow it down to just the last entry in the table, but I think it should be taken more colloquially. So I will keep all three of these possible suspects in mind.

id	age	height	eye_color	hair_color	gender	plate_number	car_make	car_model
202298	68	66	green	red	female	500123	Tesla	Model S
291182	65	66	blue	red	female	08CM64	Tesla	Model S
918773	48	65	black	red	female	917UU3	Tesla	Model S

Ok, so let's see if any of these ladies "*attended the SQL Symphony Concert 3 times in December 2017*" and checked in on FB.
First, what's the event_id of SQL Symphony Concert? (*This step was probably unnecessary, and I could have just searched by event_name in upcoming queries.)*

```
select event_id
from facebook_event_checkin
where event_name = "SQL Symphony Concert";
```

1143

From here I tried to look up whether any of those three ladies attended the concert in December 2017

```
select *
from facebook_event_checkin
where (event_id = 1143)
and date like "201712__"
```

```
and (person_id = 202298 or person_id = 291182 or person_id = 918773);
```

But I was having trouble figuring out why I kept getting no data returned. Until I realized:
Those 3 ids are their driver_license id, not their *person_id*s to look for in the fb event database. That's why none of them were showing up as having been there. So let's go back to the person table and figure out the person_id for each of these ladies.

```
select *
from person
where (license_id = 202298 or license_id = 291182 or license_id = 918773);
```

id	name	license_id	address_number	address_street_name	ssn
78881	Red Korb	918773	107	Camerata Dr	961388910
90700	Regina George	291182	332	Maple Ave	337169072
99716	Miranda Priestly	202298	1883	Golden Ave	987756388

Alright now we have person_id and names etc. for each candidate for suspect #2. Time to try again at the SQL concert

```
select *
from facebook_event_checkin
where (event_id = 1143)
and date like "201712__"
and (person_id = 78881 or person_id = 90700 or person_id = 99716);
```

person_id	event_id	event_name	date
99716	1143	SQL Symphony Concert	20171206
99716	1143	SQL Symphony Concert	20171212
99716	1143	SQL Symphony Concert	20171229

The only person from the three that was at the SQL Symphony Concert was person 99716 --> Miranda Priestly! This is our Suspect #2.
I wonder if the police interviewed her already.

```
select *
from interview
```

```
where person_id = 99716;
```

No data returned. So it appears not. Suspect #1 also said that she was wealthy. Let's double check that in the income table, since we know her ssn from our person table query earlier.

```
select *
from income
where ssn = 987756388;
```

annual_income: 310000. Yup, making 310k a year she's fairly wealthy.
Time to check if my solution is right!

Check your solution

Did you find the killer?

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

RUN ⇩ RESET

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!

Success! Yay!

I was about to sign off, when I realized I never followed one of the two early leads: I never checked out the other witness and was curious. So to find out who it was:

```
select *
from person
where name like "Annabel%"
and address_street_name = "Franklin Ave";
```

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

Now I can search for her interview

```
select *
from interview
where person_id = 16371;
```

transcript: "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 if I had followed this lead instead (or in addition to, and maybe it would have narrowed down an upcoming step), I would have gone into the get_fit_now_check_in table to see who checked in on January 9th, and gone from there.

ToDo: More advanced SQL challenges to try out?
https://discuss.codecademy.com/t/data-science-independent-project-4-home-value-trends/419948
https://discuss.codecademy.com/t/data-science-independent-project-5-analyze-airfare-data/419949