DS325 HW Week 5 SQL Name: Quan Nguyen

We will be using SQLite, which provides a standards-compliant SQL implementation. In reality, there are slight variations between the SQL dialects of different vendors (PostgreSQL, MySQL, SQLite, Oracle, Microsoft, etc.) – especially with respect to built-in functions. The SQL tutorial at http://sqlzoo.net/, provides a good introduction to the basic features of SQL. After following this tutorial, you should be able to answer most of the problems in this problem set. A more detailed tutorial is available at https://www.sqlitetutorial.net/ .

The SQLite SELECT documentation at https://sqlite.org/lang_select.html will be helpful to you, and you can access all the other SQLite documentation on that site as well.

Download and install SQLiteStudio https://sqlitestudio.pl/

Feel free to review the in-class notes as a refresher. https://www.w3schools.com/sql/

Download the data https://public.gettysburg.edu/~jpuckett/ds325/data/imdb.db

Dataset : IMDb

Schema:

movies	(tconst, title, year, runtimeminutes,genres)
people	(nconst, name, birthyear, deathyear, primaryprofession)
principals	(tconst, ordering, nconst, category, characters)
ratings	(tconst, averagerating, numberofvotes)

In the above, nconst is the unique person id and tconst is the unique movie id.

Notice that the **movies** table has an id column that uniquely identifies each movie, as well as columns for the title of a movie and the year in which the movie was released. The **people** table also has an id column, and also has columns for each person's name and birth year.

Movie ratings, meanwhile, are stored in the **ratings** table. The first column in the table is tconst: a foreign key that references the id of the movies table. The rest of the row contains data about the rating for each movie and the number of votes the movie has received on IMDb.

Finally, the **principals** tables match people to the movies in which they acted or directed.

The challenge ahead of you is to write SQL queries to answer a variety of different questions by selecting data from one or more of these tables.

For each of the following problems, you should write a single SQL query that outputs the results specified by each problem. Your response must take the form of a single SQL query, though you may nest other queries inside of your query. You **should not** assume anything about the ids of any particular movies or people: your queries should be accurate even if the id of any particular movie or person were different. Finally, each query should return only the data necessary to answer the question: if the problem only asks you to output the names of movies, for example, then your query should not also output the movie's release year.

You're welcome to check your queries' results against IMDb itself, but realize that ratings on the website might differ from those in **imdb.db**, as it has been since we downloaded the data!

What to turn in?

Submit to Moodle a single pdf which clearly contains:

the query question,

your query,

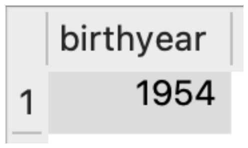
and your answer (either copy the table from SQLite or take a screenshot).

Use word or google doc to compile the pdf. For instance:

Q1. Determine the birth year of Denzel Washington.

MY SQL script is here:
SELECT *
FROM people
More SQL commands

Screen shot of my results or copy the text here



```
  birthyear
1      1954
```

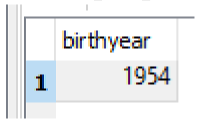
Q2. etc

Exercises: Queries

Q1. Determine the birth year of Denzel Washington.

- Your query should output a table with a single column and a single row (plus optional header) containing Denzel Washington's birth year.
- You may assume that there is only one person in the database with the name Denzel Washington.

select people.birthyear
from people
where people.name like 'Denzel Washington'

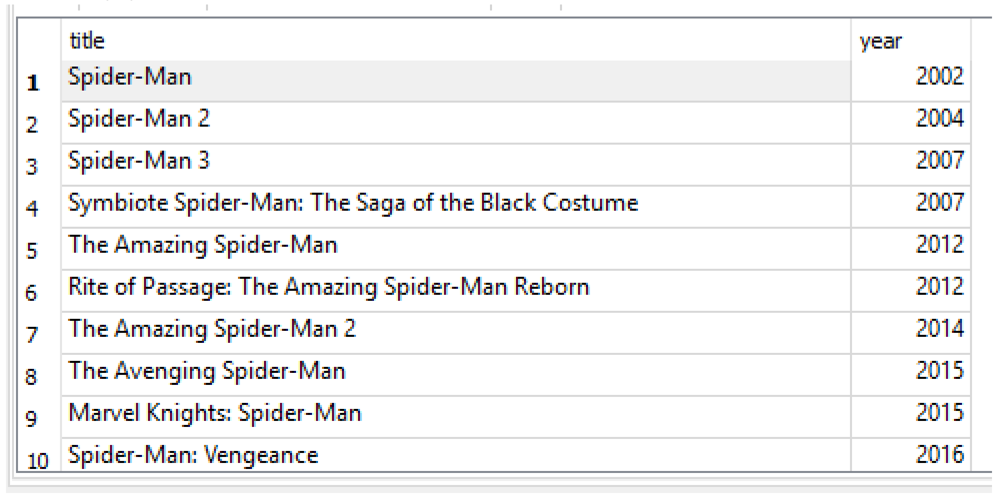


	birthyear
1	1954

Q2. List the titles and release years of all Spider-man movies, <u>in chronological order</u>.

- Your query should output a table with two columns, one for the title of each movie and one for the release year of each movie.
- You may assume that the title of all Spider-man movies will include with the word "Spider-man".
  [19 titles or 16titles depending if you check if movies have ratings]
  [you may screenshot just the top 10]

select title, year
from movies
where title like '%spider-man%'
order by year asc

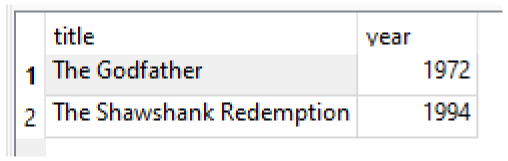


	title	year
1	Spider-Man	2002
2	Spider-Man 2	2004
3	Spider-Man 3	2007
4	Symbiote Spider-Man: The Saga of the Black Costume	2007
5	The Amazing Spider-Man	2012
6	Rite of Passage: The Amazing Spider-Man Reborn	2012
7	The Amazing Spider-Man 2	2014
8	The Avenging Spider-Man	2015
9	Marvel Knights: Spider-Man	2015
10	Spider-Man: Vengeance	2016

Q3. Determine the number of movies with an IMDb rating of greater than 9.0 and more than 50000 number of votes. Sort alphabetically by title. Output a table with the title and year. [2 titles]

select movies.title, movies.year
from ratings
inner join movies

on ratings.tconst = movies.tconst
where averagerating > 9.0 and numberofvotes > 50000

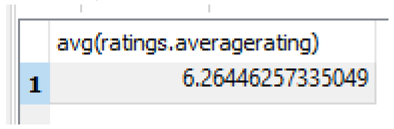


	title	year
1	The Godfather	1972
2	The Shawshank Redemption	1994

Q4. Determine the average rating of all movies released in 2012.

- Your query should output a table with a single column and a single row (plus optional header) containing the average rating.

select avg(ratings.averagerating)
from movies
inner join ratings
on movies.tconst = ratings.tconst
where year=2012

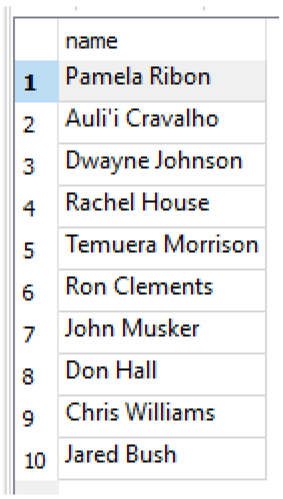


	avg(ratings.averagerating)
1	6.26446257335049

Q5. Write a SQL query to list the names of all people who involved in Moana (2016).

- Your query should output a table with a single column for the name of each person.
- You may assume that there is only one movie in the database with the title Moana. [10 people]

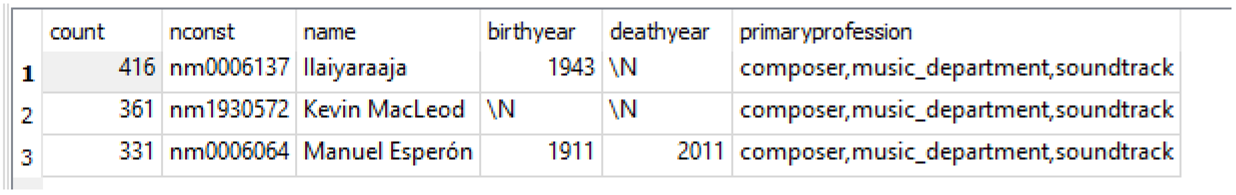
select people.name
from people
inner join principals
on people.nconst = principals.nconst
inner join movies
on principals.tconst = movies.tconst
where movies.title like 'Moana' and movies.year = 2016



	name
1	Pamela Ribon
2	Auli'i Cravalho
3	Dwayne Johnson
4	Rachel House
5	Temuera Morrison
6	Ron Clements
7	John Musker
8	Don Hall
9	Chris Williams
10	Jared Bush

Q6. Which 3 composers contributed to the most movies?

select count(principals.tconst) as count, people.*
from people
inner join principals
on people.nconst = principals.nconst
where people.primaryprofession like '%composer%' and principals.category like 'composer'
group by principals.nconst
order by count desc
limit 3



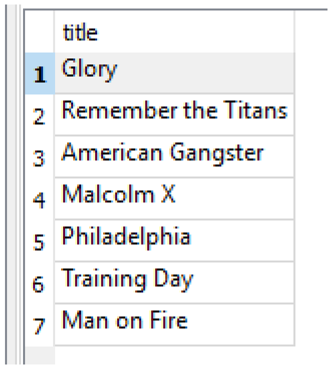
	count	nconst	name	birthyear	deathyear	primaryprofession
1	416	nm0006137	Ilaiyaraaja	1943	\N	composer,music_department,soundtrack
2	361	nm1930572	Kevin MacLeod	\N	\N	composer,music_department,soundtrack
3	331	nm0006064	Manuel Esperón	1911	2011	composer,music_department,soundtrack

Q7. Write a SQL query to list the titles of the seven highest rated movies (in order) that Denzel Washington starred in, starting with the highest rated.

- Your query should output a table with a single column for the title of each movie.
- You may assume that there is only one person in the database with the name Denzel Washington.

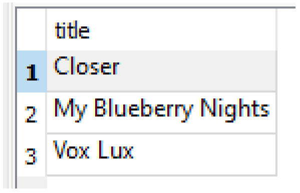
select movies.title
from movies
inner join ratings
on movies.tconst = ratings.tconst
inner join principals
on movies.tconst = principals.tconst
inner join people
on people.nconst = principals.nconst

where people.name like 'Denzel Washington'
group by movies.tconst
order by ratings.averagerating desc
limit 7



	title
1	Glory
2	Remember the Titans
3	American Gangster
4	Malcolm X
5	Philadelphia
6	Training Day
7	Man on Fire

Q8. Print the names of all movies in which “Natalie Portman” and “Jude Law” co-starred (i.e., the movie starred both actors).
[3 films]
select title
from (
select count(principals.nconst) as count,*
from principals
inner join people
on principals.nconst = people.nconst
inner join movies
on principals.tconst = movies.tconst
where (people.name like 'Natalie Portman' or people.name like 'Jude Law')
group by principals.tconst
)
WHERE count = 2

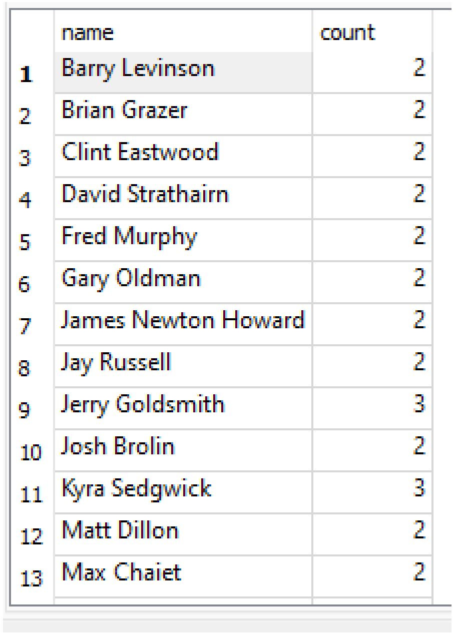


	title
1	Closer
2	My Blueberry Nights
3	Vox Lux

Challenge Question [+5pts]
Q9. Find Kevin Bacon's favorite co-stars. Print all actors as well as the number of movies that actor has co-starred with Kevin Bacon (but only if they've acted together in 2 movies or more). Be sure that Kevin Bacon isn't in your results!
[18 people, you may screenshot just the top 10]

select *
from (
select people.name, count(principals.tconst) as count

```
from people
inner join principals
on people.nconst = principals.nconst
where principals.tconst in
(
select principals.tconst
from principals
inner join people
on principals.nconst = people.nconst
where people.name like 'Kevin Bacon'
)
group by people.name
)
where count >= 2 and name not like 'Kevin Bacon'
```



	name	count
1	Barry Levinson	2
2	Brian Grazer	2
3	Clint Eastwood	2
4	David Strathairn	2
5	Fred Murphy	2
6	Gary Oldman	2
7	James Newton Howard	2
8	Jay Russell	2
9	Jerry Goldsmith	3
10	Josh Brolin	2
11	Kyra Sedgwick	3
12	Matt Dillon	2
13	Max Chaiet	2