EECS 495 — Fall 2016 **Peter Scheuermann**

Project No. 1 - SQL

The Assignment

Consider the following schema:

```
APPEARED_IN (STAR, MOVIE)
MADE_MONEY (MOVIE, HOW_MUCH, DAY_OPENED)
MARRIED (COUPLE_NUM, DAY)
DIVORCED (COUPLE_NUM, DAY)
IN_COUPLE (STAR, COUPLE_NUM)
```

Write SQL queries for the following equally weighted questions (10 points per question, +10 for extra credit). For any credit, you must run your query in MySQL, and copy and paste the query and the output into a text file using an editor like Notepad and turn that in Canvas. We will grade you by re-running the queries.

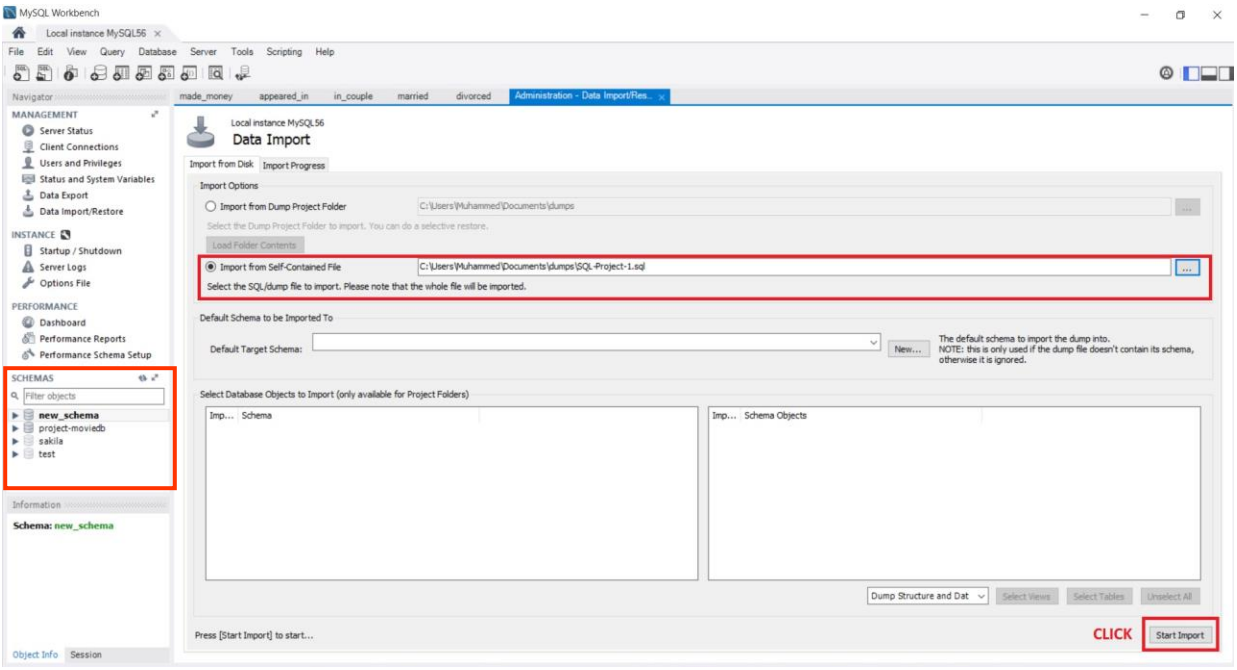
For full credit, your query should work over any database with the above schema, and not just for the one that you test it on.

Assumptions

- Couple is identified by COUPLE_NUM.
- There is only one couple in any movie.
- If a couple remarried each other, it is a different (new) couple from the old one.

Loading Schema & Data

Download the attached file with the assignment (*SQL-Project-1.sql*). You should be able to import the file within your MySQL environment. To do so, you can launch the MySQL Workbench. Go to server -> data import. Select the options "Import from Self-Contained File" and select the *SQL-Project-1.sql* file from your local directory. Then click "start import" button. The screenshot is given below:



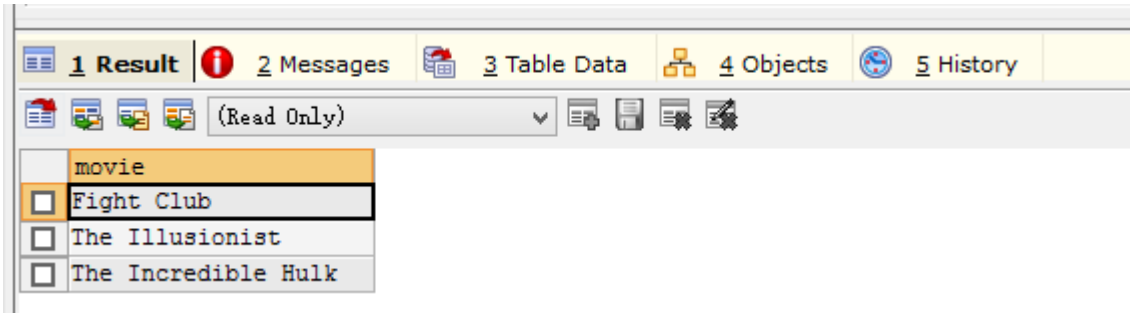
When the import is successfully done, restart the MySQL Workbench. You will find a new schema (*project-moviedb*) in the schema section of your MySQL Workbench (highlighted in the above figure). Use *project-moviedb* schema for all the SQL query testing for this assignment.

Questions

Easy Questions:

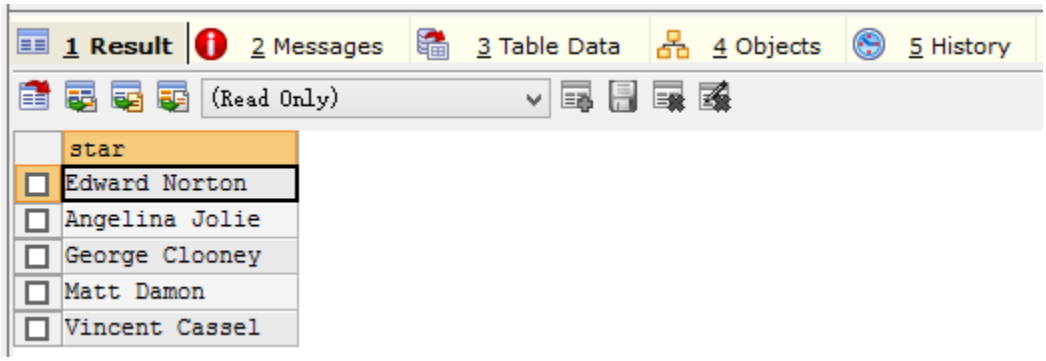
1. What movies has ‘Edward Norton' appeared in?

```
select movie
from appeared_in
where star='Edward Norton';
```



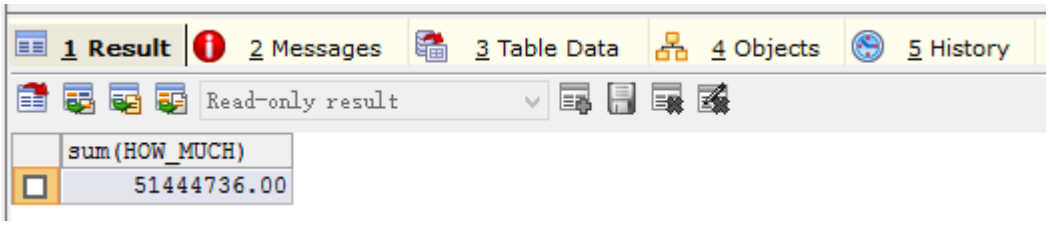
2. Who has starred along with 'Brad Pitt' in the same movie?

```
select distinct a1.star
from appeared_in a1
where a1.movie in (
select movie
from appeared_in
where star='Brad Pitt'
) and a1.star != 'Brad Pitt';
```



3. How much money, in total, have movies earned in which 'Tom Hanks' and 'Rita Wilson' starred together?

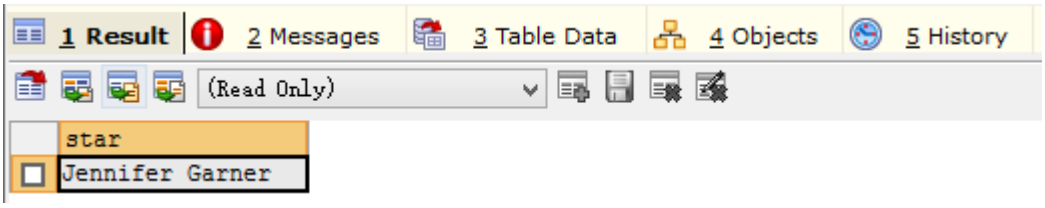
```
select sum(HOW_MUCH)
from made_money
where movie in (
select a1.movie
from appeared_in a1, appeared_in a2
where a1.star='Tom Hanks' and a2.star='Rita Wilson' and a1.movie=a2.movie
);
```



4. Who has (have) divorced 'Ben Affleck'?

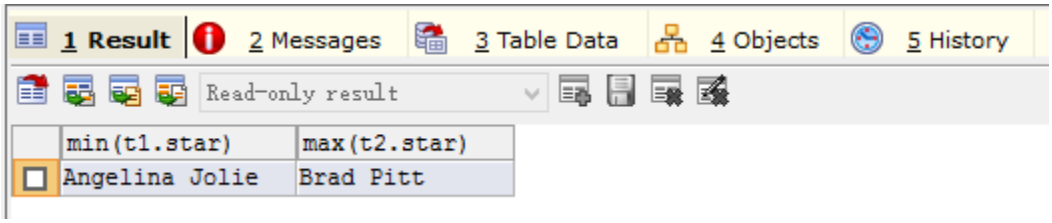
```
select t1.star
from in_couple t1, in_couple t2
```

```
where t1.couple_num in (
select couple_num
from divorced )
and t1.couple_num=t2.couple_num
and t1.star != t2.STAR
and t2.star = 'Ben Affleck';
```



5. Which stars were married and then divorced on the same day?

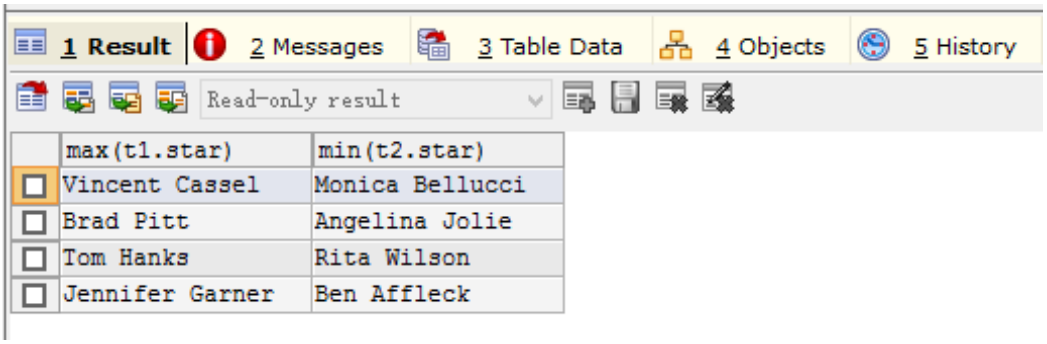
```
select min(t1.star),max(t2.star)
from in_couple t1, in_couple t2, married t3, divorced t4
where t1.couple_num=t3.couple_num
and t1.couple_num=t4.couple_num
and t1.couple_num=t2.couple_num
and t3.day=t4.day
and t1.star !=t2.star;
group by couple_num;
```



Moderate Questions:

6. What stars who married one another could possibly have met while working on the same movie?

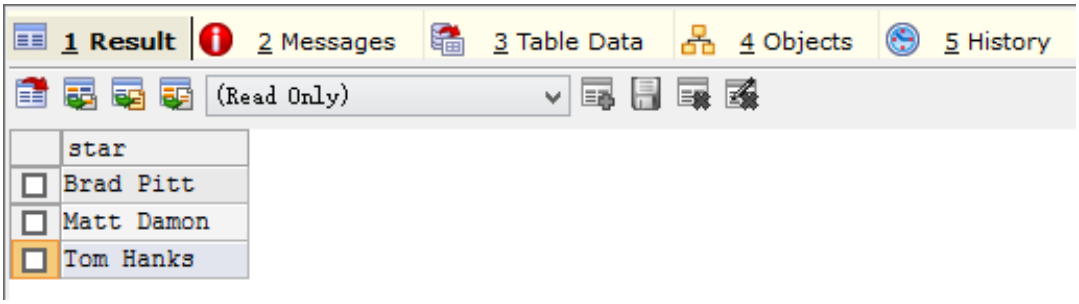
```
select distinct max(t1.star),min(t2.star)
from in_couple t1, in_couple t2, married t3, appeared_in t4, appeared_in t5
where t1.couple_num=t2.couple_num
and t1.couple_num=t3.COUPLE_NUM
and t1.star=t4.star
and t2.star=t5.star
and t4.movie=t5.movie
and t1.star != t2.star
group by t1.couple_num
```



max(t1.star)	min(t2.star)
Vincent Cassel	Monica Bellucci
Brad Pitt	Angelina Jolie
Tom Hanks	Rita Wilson
Jennifer Garner	Ben Affleck

7. What star has appeared in the most movies?

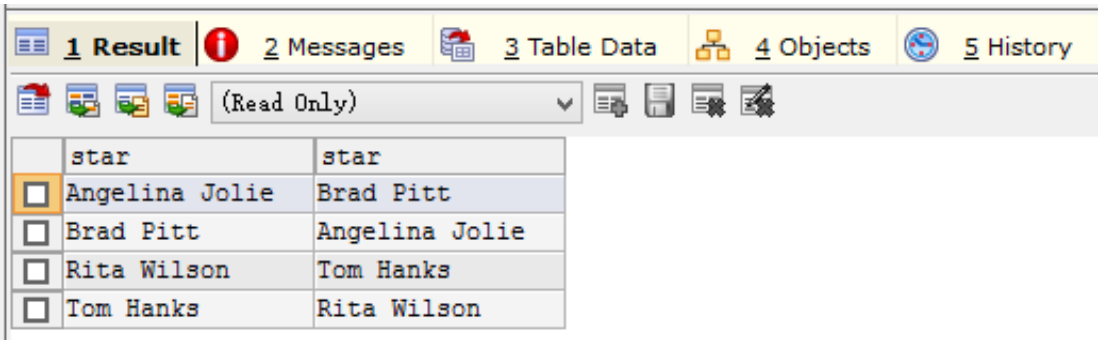
```
select star
from appeared_in
group by star
having count(*)=(
select count(star) as num
from appeared_in
group by star
order by count(star) desc
limit 1 );
```



star
Brad Pitt
Matt Damon
Tom Hanks

8. Which pairs of stars were married to each other more than once?

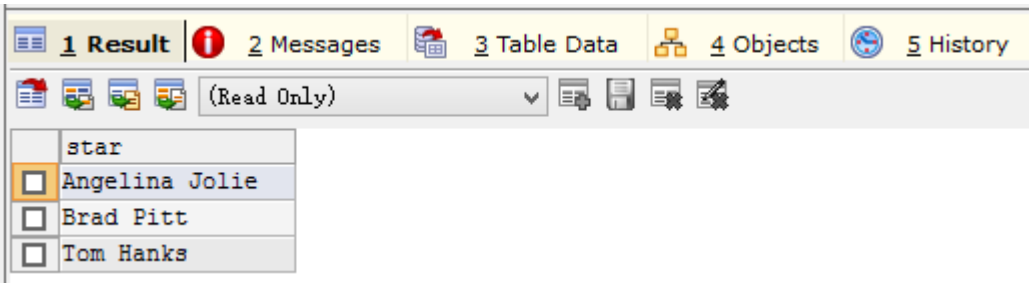
```
select t1.star,t2.star
from in_couple t1,in_couple t2,married t3
where t1.couple_num=t2.couple_num
and t1.couple_num=t3.couple_num
and t1.star != t2.star
group by t1.star,t2.star
having count(*)>1;
```



star	star
Angelina Jolie	Brad Pitt
Brad Pitt	Angelina Jolie
Rita Wilson	Tom Hanks
Tom Hanks	Rita Wilson

9. Which star or stars have been divorced at least two times?

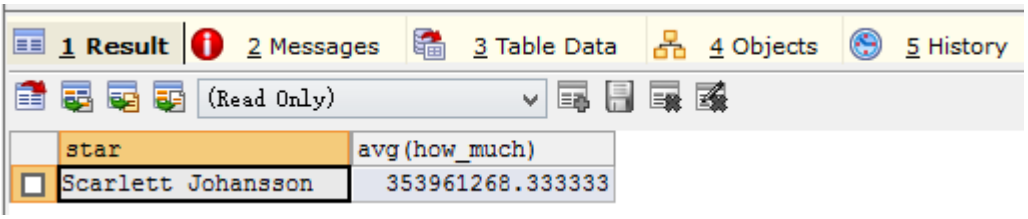
```
select t1.star
from in_couple t1,divorced t2
where t1.couple_num=t2.couple_num
group by t1.star
having count(*)>1;
```



star
Angelina Jolie
Brad Pitt
Tom Hanks

10. Which star has averaged the highest box office for all films appeared in?

```
select t1.star,avg(how_much)
from appeared_in t1,made_money t2
where t1.movie=t2.movie
group by t1.star
order by avg(how_much) desc
limit 1;
```

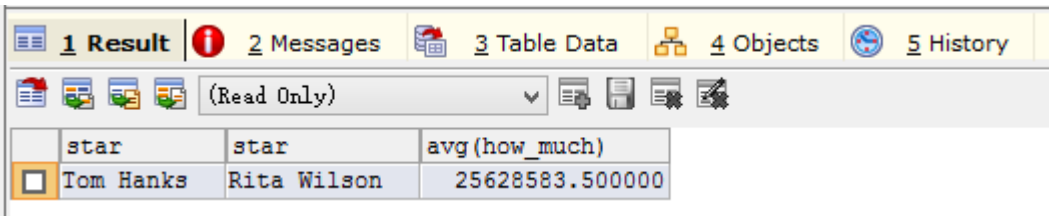


star	avg(how_much)
Scarlett Johansson	353961268.333333

Hardest Question: (for extra credit +10)

11. What couple averaged the most money per film *while they were married*?

```
select t1.star,t2.star,avg(how_much)
from (appeared_in t1,appeared_in t2,in_couple t3, in_couple t4, married t5,
made_money t6) left outer join divorced t7
on t5.couple_num=t7.couple_num
where t1.movie=t2.movie
and t1.star=t3.star
and t2.star=t4.star
and t3.couple_num=t4.couple_num
and t1.movie=t6.movie
and t3.couple_num=t5.couple_num
and t1.star != t2.star
and t5.day<t6.day_opened
and t7.day>t6.day_opened
group by t1.star,t2.star
order by avg(how_much) desc
limit 1;
```



star	star	avg(how_much)
Tom Hanks	Rita Wilson	25628583.500000

Notes

Here are a few more, final notes:

SQL is not truly set-oriented. What this means is that you can ask queries and get back multiples of the same value. To avoid this, add the `DISTINCT` keyword. For example:

```
SELECT DISTINCT(STAR) FROM APPEARED_IN;
```

Often times, it will be useful to create a temporary table that will be used to complete a given query. You can do it by using `CREATE VIEW` command:

```
CREATE VIEW ALL_STARS AS SELECT DISTINCT(STAR) FROM APPEARED_IN;
```

After you do this, you can query the imaginary table ALL_STARS just like it was a regular table. To kill it when you are done, type:

```
DROP VIEW ALL_STARS;
```

You must drop any newly created view/table in your answer if you use create view, or create table, otherwise points will be deducted.

Collaboration

Lastly, and most important, a note on what is allowed and what is not allowed in terms of collaboration. It is OK to speak in generalities about how one might approach these problems with others. In other words, you can bounce ideas off of one another. However, it is NOT OK TO DISCUSS SQL ANSWERS AT ALL, as it relates to these problems, with anyone else in class. It is not OK to look at anyone else's answers, or to specifically discuss answers to the problems with anyone else. Basically, two sets of answers to problems that are turned in with the same SQL queries will be viewed with extreme suspicion.

Submission Guidelines

The assignment is due by the midnight on the day mentioned on the front page. Please submit the assignment electronically via Canvas, but be sure to send the files in a format that we can read (.txt, .doc, .pdf, etc.). You also can (and may be should) optionally upload another text file (apart from the answers) where you can specify any additional assumptions you have made, or describe your method briefly in case it is not producing the desired result.

s