Christian Tomford

Dr. Yates

2/8/2019

Databases - Homework #1

SQL SELECT Queries

1. *Run the following queries and turn in the result of running each along with a description of what you think the query is finding:*

a. Result:

```
cooking=> SELECT title FROM Posts WHERE posttypeid = 1 ORDER BY score DESC LIMIT 5;
                            title
-------------------------------------------------------------
 How can I chop onions without crying?
 Translating cooking terms between US / UK / AU / CA / NZ
 Why add salt to the water when cooking pasta?
 Why would boiling milk in an electric kettle break the kettle?
 What is this colored fiber in my chicken?
(5 rows)
```

Description: This query is selecting the title from posts on the cooking database and because posttypeid = 1 the database is looking at answers and not questions. It is then ordered by the score and put into descending order and limiting the output that we can see to 5 rows.

b. Result:

```
cooking=> SELECT reputation,displayname FROM users ORDER BY reputation DESC LIMIT 5;
 reputation | displayname
------------+--------------
      77759 | rumtscho
      67203 | SAJ14SAJ
      59803 | Joe
      52299 | Cascabel
      50322 | Jolenealaska
(5 rows)
```

Description: This query is selecting the reputation and display name from users on the cooking database. It orders the reputation in descending order and outputs the top 5 users display names with the most reputation.

c. Result:

```
cooking=> SELECT displayname,ownerdisplayname FROM users INNER JOIN pos
ts ON users.id = owneruserid WHERE displayname <> ownerdisplayname;
     displayname      | ownerdisplayname
----------------------+------------------
 davidscolgan         | dvcolgan
 lamwaiman1988        | gunbuster363
 Danubian Sailor      | lechlukasz
 Hank                 | Henry Jackson
 Artificial Stupidity | Yasashii Eirian
 Netizen              | IamSJ
 Fahed                | Fahed Fahed
 Roux A               | Hershe A
(8 rows)
```

Description: This query is selecting the display name and the owner of the display name from users in the cooking database and selects keywords with the record that have matching values from both of the tables displayname and ownerdisplayname.

2. *Construct queries to answer the following questions. Turn your query, the result, and a description of how your query works.*

a. Find the question with the most tags. Result:

```
cooking=> SELECT title,((CHAR_LENGTH(tags) - CHAR_LENGTH(REPLACE(tags, '<', ''))) / CHAR_LENGTH('<'))
 AS numtags FROM posts WHERE posttypeid = 1 ORDER BY numtags DESC LIMIT 1;
                  title                   | numtags
------------------------------------------+---------
 What happens if I brine my turkey for 2 days? |       5
(1 row)
```

Description: This query is selecting the title of a question and the tags to that question in the cooking database. It and will output the question with the most tags. It does this by finding the length of the tags characters and then replaces the "<" symbol with " " (nothing). Then subtracts the number of characters of -- old tags without the "<" symbol from the new number of characters of tags with the "<" symbol. Here is an example:

original tags = <food><pasta><pie> | There are 18 characters here

replace all "<" with " "

new tags = food>pasta>pie> | There are 15 characters here

original tags(18) - new tags(15) = 3 tags

It is also important to note that the maximum tags on a question can only be 5. That is why there are not questions with hundreds or even thousands of tags.

b. *How many questions were asked on a Friday.* Result:

```
cooking=> SELECT COUNT(*) FROM posts WHERE posttypeid = 1 AND EXTRACT(DOW FROM creationdate) = 4;
 count
-------
  2960
(1 row)
```

Description: This query counts the number of rows from posts where there is a question and extracts the day of the week from creationdate, in this case it's Fridays only. posttypeid = 1 means to only look at questions and not answers. If posttypeid = 2 then it would look at answers.

c. *Which month of the year sees the most activity.* Result:

```
cooking=> SELECT COUNT(posts),EXTRACT(MONTH FROM creationdate) FROM posts GROUP
 BY date_part ORDER BY count DESC;
 count | date_part
-------+-----------
  7958 |         7
  7084 |         8
  6348 |         1
  6040 |        12
  5971 |        11
  5870 |         9
  5770 |         3
  5732 |        10
  5663 |         2
  4991 |         4
  4689 |         5
  4325 |         6
(12 rows)
```

Description: First, I am going to define activity as posting either a question or an answer. So, this query is counting the posts from the extracted months from creationdate. GROUP BY date_part results in combining a group, in this case, the same month, and then outputting in descending order so we can see which month of the year sees the most activity. In this case, July sees the most activity.

3. *Write three different queries of your own. Turn in the query, the result, and an English description of the information the query was accessing.*

a. Result:

```
cooking=> SELECT displayname,upvotes FROM users ORDER BY upvotes DESC LIMIT 5;
 displayname | upvotes
-------------+---------
 Megha       |   16920
 Joe         |   10801
 rumtscho    |    9842
 Ward        |    8434
 Sue         |    5043
(5 rows)
```

Description: This query is selecting the displayname and upvotes from all the users in the cooking database and ordering them in descending order so we are able to tell which user has the

most upvotes. LIMIT will only allow it output the top 5. As we can see from the result the user "Megha" has the most upvotes.

b. Result:

```
cooking=> SELECT id,reputation FROM users WHERE reputation > 110 AND reputation < 112 LIMIT 5;
 id  | reputation
-----+------------
 237 |        111
 233 |        111
 279 |        111
 564 |        111
 593 |        111
(5 rows)
```

Description: This query is fairly simple. It shows the id of users with exactly 111 reputation. It does this by selecting id and reputation from users and where reputation is greater than 110 and where reputation is less than 112 and then limit the output to only show 5 rows. All of the users within this result have a reputation of 111.

c. Result:

```
cooking=> SELECT COUNT(viewcount),EXTRACT(DOW FROM creationdate) FROM posts
GROUP BY date_part ORDER BY count DESC;
 count | date_part
-------+-----------
  3063 |         1
  3020 |         3
  2960 |         4
  2956 |         2
  2910 |         0
  2900 |         5
  2609 |         6
(7 rows)
```

Description: This query shows that Monday has the most views out of all the days of the week. It does this by counting viewcount and extracting DOW(Day Of Week) from creationdate and then, from posts GROUP BY date_part results, which will combine a group, in this case, the same day of the week, and then output in descending order from the most amount so we can see which day

of the week gets the most views. In this case, Monday gets the most views out of all the other days of the week.