Last Name: SUH First Name: JOOWON Student ID: 44414081

1. [10pts] Return the handles of Tweeters and their number of Covid-tagged tweets if they’ve used the hashtag “**covid19**” more than 3 times. Your query should normalize the hashtags to lowercase (e.g., **Covid19** should be converted to **covid19** in order to properly consider all Covid-tagged tweets.

a) [7pts] SQL Query:

from Tweeter

where Tweeter.tweeter\_id IN ( select tweeter\_id

from( select count(\*) as cnt, tweeter\_id

from ( select lowercase.hashtag, lowercase.tweet\_id, count(\*)

from ( select LOWER(hashtag) as hashtag, tweet\_id

from Hashtags ) as lowercase

where lowercase.hashtag = 'covid19'

group by tweet\_id ) as covidtweets, Tweet

where covidtweets.tweet\_id = Tweet.tweet\_id

group by tweeter\_id ) as cnttb

where cnttb.cnt >3 )

b) [3pts] Result (2 rows):

Table

Description automatically generated

2. [10pts] Return the handles of Tweeters who have a followers count greater than 500,000 and who have posted a tweet that contains one or more of the top ten most popular hashtags. (Note: You can break popularity ties arbitrarily.)

a) [7pts] SQL Query:

select \*

from Tweeter

where followers\_count > 500000 and Tweeter.tweeter\_id IN (select tweeter\_id

from

(select Hashtags.tweet\_id

from ( select count(\*), lowercase.hashtag

from ( select hashtag as hashtag, tweet\_id from Hashtags ) as lowercase

group by lowercase.hashtag

order by count(\*) desc

limit 10) as topten, Hashtags

where topten.hashtag = Hashtags.hashtag) as tweetwithtopten, Tweet

where tweetwithtopten.tweet\_id = Tweet.tweet\_id)

b) [3pts] Result (3 rows):

Table

Description automatically generated

3. [10pts] Find the tweet ids for tweets that have been **verified** using at least two **different** pieces of evidence and that have a popularity greater than 2.4. Remember from HW1 (😷 ) that the popularity of a tweet can be computed using the formula:

a) [7pts] SQL Query:

b) [3pts] Result (4 rows):

4. **Views** [20 pts]

Congratulations! For obvious reasons, the CTO of ***CheckedTweets.org*** is setting up a data science team to analyze election tweets that contain one or more of the following hashtags: **"election2020", "trump", "biden", "bidenharris2020", "trumppence2020", "pennsylvania", "northcarolina", "wisconsin", "michigan".**  (*You will need to normalize the hashtags to lowercase*.) The CTO has made you the head of that team. As the team leader, you have been asked to create a SQL view so that the rest of the team can simply look at the data and draw meaningful conclusions without having to deal with all of its underlying complexity.

The view should provide simple tabular access to a combination of the following pieces of information:

* Tweeter info (tweeter\_id, handle, followers\_count, verified)
* Tweet info (tweet\_id, tweet\_text, *popularity*, *quality*)

**Remember** that tweet *popularity* and *quality* are derived attributes and can be computed as follows:

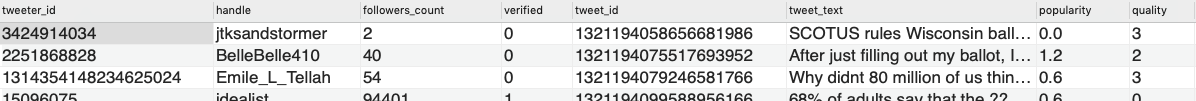
a) [15 pts] Create the desired view (ElectionTweets) by writing an appropriate CREATE VIEW statement.

CREATE VIEW ElectionTweets…;

A picture containing application

Description automatically generated

*(Hint: your view should have 699 rows)*



b) [5 pts] Show the usefulness of your view by writing a SELECT query against the view that prints the Tweet id, the Tweeter’s handle, and the popularity and quality of tweets that have the maximum popularity.

Result (1 row):

5. **Stored Procedures** [20 pts]

a) [15 pts] Create and exercise a SQL stored procedure called RegisterChecker(…) that the application can use to add a brand new checker with an office phone to the database. You **may not** change the signature of this procedure. Hint: To get the current time, use the [NOW()](https://dev.mysql.com/doc/refman/8.0/en/date-and-time-functions.html#function_now) function.

|  |
| --- |
| DELIMITER // **CREATE** **PROCEDURE** RegisterChecker(  user\_id integer,  name\_first varchar(50),  name\_last varchar(50),   email varchar(100),  password varchar(30),  profile\_pic varchar(500),  address\_country varchar(30),  address\_state varchar(30),  address\_city varchar(30),  office\_number varchar(20) ) **BEGIN**   ... **END**; // DELIMITER ; |

A picture containing graphical user interface

Description automatically generated  
b) [5pts] Verify that your new stored procedure works properly by calling it as follows to add a new checker and then running a SELECT query to show the stored procedure’s after-effects:

|  |
| --- |
| **CALL** RegisterChecker (3000, "Peter", "Anteater",  "peter-anteater2020@gmail.com", "pretend-this-is-hashed",  null, "USA", "California", "Irvine", "(949) 824-5011");  **SELECT** U.user\_id, U.email, U.profile\_pic, C.checker\_since, P.number, P.kind **FROM** **User** U, Checker C, Phone P **WHERE** U.user\_id = C.user\_id **AND**   P.user\_id = C.user\_id **AND**  U.user\_id = 3000; |

Result (1 row): Table

Description automatically generated

6. **Alter Table** [10 pts]

As your schema currently stands, evidence can only be submitted in the form of URLs to websites. Your boss would like to enrich the Evidence entity by also allowing books (specifically, 13-character ISBNs) to be used as evidence. This changes your ER model in two ways: 1) URL now becomes an optional field in Evidence, and 2) ISBN is now an additional optional field in Evidence.

Note: *The current datatype for URL is VARCHAR(500).*

a) [5 pts] Write and execute the ALTER TABLE statement(s) needed to modify the Evidence table to reflect the new requirements above. (Hint: Refer to the MySQL documentation online if you need more information about how to use the ALTER TABLE statement.)

Graphical user interface, application

Description automatically generated

b) [5 pts] Execute the following INSERT and SELECT statements to show the effect of your change. Report the results (just the counts) for each SELECT statement.

|  |
| --- |
| **INSERT** **INTO** Evidence (ev\_id, url, isbn) **VALUES** (2000, NULL, "0-1306-3278-3");  **SELECT** **COUNT**(\*) **AS** url\_evidence  **FROM** Evidence **WHERE** url **IS NOT NULL**;  **SELECT COUNT**(\*) **AS** book\_evidence  **FROM** Evidence  **WHERE** isbn **IS NOT NULL**; |

Result:

Graphical user interface, text, application

Description automatically generated

A picture containing graphical user interface

Description automatically generated

7. **Triggers** [20 pts]

a) [15 pts] To help tie your newfound SQL knowledge back to the seemingly mysterious initial ER model, you are tasked with defining a trigger called update\_tweet\_info(…). When raw tweets are deposited into the database by your application, this trigger will insert tuples into the Tweet, Tweeter, and Hashtags tables using the information found in the newly deposited raw tweets. *If a Tweeter already exists at the time of a deposit, you should only update their follower count, display name, and handle.* To specify an update action for an INSERT statement when you have a duplicate primary key, i.e., when an object with that key already exists, see [here](https://dev.mysql.com/doc/refman/8.0/en/insert-on-duplicate.html). A picture containing text

Description automatically generated

Hint 1: To get all tweeter-associated information for an arbitrary single raw tweet, we can perform the query:

|  |
| --- |
| **SELECT** JSON\_UNQUOTE(JSON\_EXTRACT(**content**, '$.user.screen\_name')) **AS** display\_name,  JSON\_UNQUOTE(JSON\_EXTRACT(**content**, '$.user.followers\_count')) **AS** followers\_count,  JSON\_UNQUOTE(JSON\_EXTRACT(**content**, '$.user.name')) **AS** handle,  JSON\_UNQUOTE(JSON\_EXTRACT(**content**, '$.user.id\_str')) **AS** tweeter\_id,  **CASE** **WHEN** JSON\_EXTRACT(**content**, '$.user.verified') **THEN** 1 **ELSE** 0 **END** **AS** verified **FROM** RawTweet T **LIMIT** 1; |

Hint 2: To get all tweet-associated information for an arbitrary single raw tweet, we can perform the query:

|  |
| --- |
| **SELECT** JSON\_UNQUOTE(JSON\_EXTRACT(T.content, '$.created\_at')) **AS** posting\_datetime,  JSON\_EXTRACT(T.content, '$.geo.coordinates[0]') **AS** posting\_location\_latitude,  JSON\_EXTRACT(T.content, '$.geo.coordinates[1]') **AS** posting\_location\_longitude,  JSON\_EXTRACT(T.content, '$.quoted\_status\_id') **AS** quoted\_tweet,  JSON\_EXTRACT(T.content, '$.in\_reply\_to\_status\_id') **AS** replied\_to\_tweet,  JSON\_UNQUOTE(JSON\_EXTRACT(T.content, '$.id')) **AS** tweet\_id,  JSON\_UNQUOTE(JSON\_EXTRACT(T.content, '$.text')) **AS** tweet\_text,  JSON\_UNQUOTE(JSON\_EXTRACT(T.content, '$.user.id\_str')) **AS** tweeter\_id **FROM** RawTweet T **LIMIT** 1; |

Hint 3: To **update** the Hashtag table for a particular raw tweet, we can call the following stored procedure that we have provided for you in the updated load script:

|  |
| --- |
| **CALL** UpdateHashtags(tweet\_id); |

|  |
| --- |
| DELIMITER // **CREATE** **TRIGGER** update\_tweet\_info ... **FOR** **EACH** **ROW** **BEGIN** ... **END**; // DELIMITER ; |

b) [5 pts] Execute the following INSERT and SELECT statements to show the effect of your trigger. Report the follower count, the number of tweets posted, and the number of distinct hashtags associated with the specified tweeter before and after each INSERT.

|  |
| --- |
| **SELECT** TW.followers\_count, **COUNT**(**DISTINCT** T.tweet\_id) **AS** tweets\_posted,  **COUNT**(**DISTINCT** H.hashtag) **AS** total\_hashtags  **FROM** Tweeter TW **LEFT** **OUTER** **JOIN** Tweet T **ON** TW.tweeter\_id = T.tweeter\_id **LEFT** **OUTER** **JOIN** Hashtags H **ON** T.tweet\_id = H.tweet\_id **WHERE** TW.tweeter\_id = T.tweeter\_id **AND** T.tweeter\_id = '1121141389696413697'; |

Result (1 row): Table

Description automatically generated

|  |
| --- |
| **INSERT** **INTO** RawTweet **VALUES** (  "1321203503310762222",  '{"id": "1321203503310762222", "created\_at": "2020-06-12 14:56:40", "extended\_tweet": {"entities": {"hashtags": [{"indices": [136, 145], "text": "criminals"}, {"indices": [147, 158], "text": "Obidengate2020"}, {"indices": [159, 169], "text": "obamagate"}, {"indices": [170, 178], "text": "treason"}, {"indices": [179, 198], "text": "politicalespionage"}, {"indices": [199, 219], "text": "seditiousconspiracy"}, {"indices": [220, 235], "text": "4yrsrussialies"}, {"indices": [236, 262], "text": "fbidocsproveHRCrussiahoax"}], "user\_mentions": [{"indices": [0, 9], "screen\_name": "LLinda\_W", "id\_str": "3303621560", "name": "LindaW- #VoteBlueDownBallot", "id": 3303621560}, {"indices": [10, 23], "screen\_name": "WardDPatrick", "id\_str": "3353220134", "name": "Pat Ward", "id": 3353220134}]}, "full\_text": "@LLinda\_W @WardDPatrick Truly nauseating to hear the excuses and defensive epic fails coming from antiamerican actually voting for this #criminal #Obidengate #obamagate #treason #politicalespionage #seditiousconspiracy #4yrsrussialies #fbidocsproveHRCrussiahoax https://t.co/Q98eyjmpBC"}, "retweet\_count": 0, "retweeted": false, "filter\_level": "low", "in\_reply\_to\_screen\_name": "LLinda\_W", "is\_quote\_status": false, "id\_str": "1321203503310762222", "favorite\_count": 0, "text": "@LLinda\_W @WardDPatrick Truly nauseating to hear the excuses and defensive epic fails coming from antiamerican actu\u2026 https://t.co/tDmveFKMDf", "lang": "en", "quote\_count": 0, "favorited": false, "possibly\_sensitive": false, "truncated": true, "timestamp\_ms": "1603834143603", "reply\_count": 0, "user": {"friends\_count": 911, "profile\_image\_url\_https": "https://pbs.twimg.com/profile\_images/1280133701599404032/9HYrUqHg\_normal.jpg", "listed\_count": 1, "profile\_background\_image\_url": "", "default\_profile\_image": false, "favourites\_count": 9141, "description": "\u26a1 #Seditiousconspiracy", "is\_translator": false, "profile\_background\_image\_url\_https": "", "protected": false, "screen\_name": "WrathchiksMama", "id\_str": "1121141389696413697", "profile\_link\_color": "1DA1F2", "translator\_type": "none", "id": 1121141389696413697, "geo\_enabled": false, "profile\_background\_color": "F5F8FA", "profile\_sidebar\_border\_color": "C0DEED", "profile\_text\_color": "333333", "verified": false, "profile\_image\_url": "http://pbs.twimg.com/profile\_images/1280133701599404032/9HYrUqHg\_normal.jpg", "url": "https://www.altcensored.com/watch?v=9HFxVvrXjCg", "contributors\_enabled": false, "profile\_background\_tile": false, "profile\_banner\_url": "https://pbs.twimg.com/profile\_banners/1121141389696413697/1594042636", "statuses\_count": 7648, "followers\_count": 219, "profile\_use\_background\_image": true, "default\_profile": true, "name": "ThePlotAgainstPresident", "location": "\ud83d\udc47\ud83d\udc47WATCH&SHARE!#SaveOur1A\ud83d\udc47\ud83d\udc47", "profile\_sidebar\_fill\_color": "DDEEF6"}}' );  **SELECT** TW.followers\_count, **COUNT**(**DISTINCT** T.tweet\_id) **AS** tweets\_posted,  **COUNT**(**DISTINCT** H.hashtag) **AS** total\_hashtags  **FROM** Tweeter TW **LEFT** **OUTER** **JOIN** Tweet T **ON** TW.tweeter\_id = T.tweeter\_id **LEFT** **OUTER** **JOIN** Hashtags H **ON** T.tweet\_id = H.tweet\_id **WHERE** TW.tweeter\_id = T.tweeter\_id **AND** T.tweeter\_id = '1121141389696413697'; |

Result (1 row): Table

Description automatically generated

|  |
| --- |
| **SELECT** TW.followers\_count, **COUNT**(**DISTINCT** T.tweet\_id) **AS** tweets\_posted,   **COUNT**(**DISTINCT** H.hashtag) **AS** total\_hashtags  **FROM** Tweeter TW **LEFT** **OUTER** **JOIN** Tweet T **ON** TW.tweeter\_id = T.tweeter\_id **LEFT** **OUTER** **JOIN** Hashtags H **ON** T.tweet\_id = H.tweet\_id **WHERE** TW.tweeter\_id = T.tweeter\_id **AND** T.tweeter\_id = '992109555483103122'; |

Result (1 row): Table

Description automatically generated

|  |
| --- |
| **INSERT** **INTO** RawTweet **VALUES** (  "1321203503310762322",  '{"id": "1321203503310762322", "created\_at": "2020-04-29 05:29:03", "retweet\_count": 0, "retweeted": false, "filter\_level": "low", "is\_quote\_status": false, "id\_str": "1321203503310762322", "favorite\_count": 0, "text": "Harris County early voting hours extended to 10 p.m. until Thursday https://t.co/VF9uvQlJwK) \*\*LETS GO, HARRIS COU\u2026 https://t.co/k7pt7SRIN5", "lang": "en", "quote\_count": 0, "favorited": false, "possibly\_sensitive": false, "truncated": true, "timestamp\_ms": "1603832187724", "reply\_count": 0, "entities": {"urls": [{"display\_url": "chron.com/news/election2\u2026", "indices": [68, 91], "expanded\_url": "https://www.chron.com/news/election2020/article/Harris-County-early-voting-hours-extended-to-10-15677986.php?utm\_campaign=CMS%20Sharing%20Tools%20(Premium", "url": "https://t.co/VF9uvQlJwK"}, {"display\_url": "twitter.com/i/web/status/1\u2026", "indices": [117, 140], "expanded\_url": "https://twitter.com/i/web/status/1321194070757330945", "url": "https://t.co/k7pt7SRIN5"}]}, "user": {"friends\_count": 22468, "profile\_image\_url\_https": "https://pbs.twimg.com/profile\_images/1057463467286773760/wKFC-Ixa\_normal.jpg", "listed\_count": 6, "profile\_background\_image\_url": "http://abs.twimg.com/images/themes/theme1/bg.png", "default\_profile\_image": false, "favourites\_count": 20682, "description": "JUJUS BU PRU VETTED NICE LADY, BUT, PLEASE DONT HIT ON ME. MARRIED CLOSE TO 40 YRS. CRITTER LOVER. PRO CLIMATE 4 ALL KIDS. MAIN ACCT@jjsmokkieBOY57", "is\_translator": false, "profile\_background\_image\_url\_https": "https://abs.twimg.com/images/themes/theme1/bg.png", "protected": false, "screen\_name": "smokesdad289", "id\_str": "992109555483103122", "profile\_link\_color": "FF691F", "translator\_type": "none", "id": 992109555483103122, "geo\_enabled": false, "profile\_background\_color": "000000", "profile\_sidebar\_border\_color": "000000", "profile\_text\_color": "000000", "verified": false, "profile\_image\_url": "http://pbs.twimg.com/profile\_images/1057463467286773760/wKFC-Ixa\_normal.jpg", "contributors\_enabled": false, "profile\_background\_tile": false, "profile\_banner\_url": "https://pbs.twimg.com/profile\_banners/992109555483103232/1541111709", "statuses\_count": 62229, "followers\_count": 20476, "profile\_use\_background\_image": false, "default\_profile": false, "name": "jujus other", "location": "texas", "profile\_sidebar\_fill\_color": "000000"}}' );  **SELECT** TW.followers\_count, **COUNT**(**DISTINCT** T.tweet\_id) **AS** tweets\_posted,  **COUNT**(**DISTINCT** H.hashtag) **AS** total\_hashtags  **FROM** Tweeter TW **LEFT** **OUTER** **JOIN** Tweet T **ON** TW.tweeter\_id = T.tweeter\_id **LEFT** **OUTER** **JOIN** Hashtags H **ON** T.tweet\_id = H.tweet\_id **WHERE** TW.tweeter\_id = T.tweeter\_id **AND** T.tweeter\_id = '992109555483103122'; |

Result (1 row):

Table

Description automatically generated