Bonus Assignment Option A: Handling Tweets with R and SQL

There are two videos posted on Blackboard. The first shows how to set up a Twitter account and a Twitter Developed App to receive the credentials needed to use the Twitter API through R. The second details how to use the file Script.R *(also on Blackboard) to download Tweets, export them to Access, where to put your answers to the problems, and what you need to deliver for the bonus assignment. Details on how to submit your assignment will be posted on Blackboard later. You can work in teams of two on this assignment.*

Checklist for submission of the assignment

You need to submit a ZIP-file that contains:

- The file *Script.R* with the code you created to solve the problems placed in the marked areas
- All CSV and PNG files you generate through R as part of the assignment (see video)
- An Access Database (*.accdb) that contains the imported dataset as well as a separate query for each of the SQL problems in the assignment (see video)

Grading

The points achieved will be divided by 100, multiplied by 0.5 (the maximum number of grade points achievable), and rounded to the closest decimal. Hence, you need

90 points or more for 0.5

70 points or more for 0.4

50 points or more for 0.3

30 points or more for 0.2

10 points or more for 0.1

Tasks only give full or no points, it is not possible to partially fulfil the requirement of a task.

Tasks

Follow the video description to download 500 Tweets on a topic of your choice, export them as CSV and import the CSV to Access. All of these steps are described in the videos.

R part

1) Create a list that has the number of characters in each Tweet as elements (Hint: use the function nchar). Add the list as a column to the data frame and save the data frame to a new CSV-file. [20 points]
2) Create a data frame that only contains those tweets with an even ID number using a FOR-loop and IF-ELSE constructs. Save the resulting data frame as a new CSV-file. [20 points]
3) Create a word cloud of the contents of the tweets and save the word cloud as a PNG file. You will need to research, install, and use new packages for this task. [15 points]

Access/SQL part

1) Create a query that returns the text and retweet count for all tweets that are not retweets and eliminates any duplicates. [4 points]
2) Create a query that shows the average number of retweets overall all of the tweets returned in the first query. You may not create a new table to achieve this, it has to be done by executing a single query on the original table. [4 points]
3) Create a query that returns the screen name and the number of tweets posted of all users contained in the original table. The column with the number of tweets posted should be called "numberPosts". The output should be sorted first by the number of tweets posted (descending) and then alphabetically be the screen name (ascending). You will need to familiarize yourself with the alias and GROUP BY functions of SQL in Access. [14 points]
4) Create a query that returns the average date/time the tweets in the original table were created (this will be a decimal number since the date/time is stored as an integer timestamp). [2 points]
5) Create a query that creates anew table ("New_Tweets") that only contains those tweets (but with all attributes) that were created after this average date/time. Again, this must be a single query that is executed on the original table. [11 points]
6) Create a query that returns the number of records in this newly created table. [2 points]
7) What does the result from the previous query imply? [3 points]