

Stat 142 MP Light 4

1st Semester, A.Y. 2025-26

First Name: Anne Christine

Last Name: Amores

Student Number: 2021-04650

Section: WFU

```
# Import Twitter dataset
twt_data <- read.csv("Twitter_Data.csv")

# Set up connection
library(RSQLite)
library(DBI)
con <- dbConnect(RSQLite::SQLite(), ":memory:")
dbWriteTable(con, "twt_data", twt_data)
```

Item 1

Query the twitter dataset such that you will be able to extract rows from male users aged 45 and below. Retain only the replies, retweets, likes, and the twitter ID. Save this as a data frame named `twt_sub1` and write this in your database (i.e., in memory). Answer the following questions by querying from `twt_sub1` and by displaying the specific answer only via SQL queries.

```
SELECT twitter_id, replies, retweets, likes
FROM twt_data
WHERE (gender = "Male" AND age <= 45)
```

```
dbWriteTable(con, "twt_sub1", twt_sub1)
```

- How many in total are the observations extracted from item 1?

```
SELECT COUNT () AS n_observations
FROM twt_sub1
```

Table 1: 1 records

n_observations
1868

- b. How many observations are extracted per user? Show the number of observations and the user.

```
SELECT COUNT () AS n_observations, twitter_id
  FROM twt_sub1
 GROUP BY twitter_id
```

Table 2: 3 records

n_observations	twitter_id
1148	GovAndyBeshear
598	GovRonDeSantis
122	GovofCO

- c. What is the average number of retweets, likes, and replies per user? Make sure to get the average of each variable and show each user (this should be in a single SQL query only).

```
SELECT twitter_id, AVG(retweets), AVG(likes), AVG(replies)
  FROM twt_sub1
 GROUP BY twitter_id
```

Table 3: 3 records

twitter_id	AVG(retweets)	AVG(likes)	AVG(replies)
GovAndyBeshear	70.70035	365.9625	67.14286
GovRonDeSantis	384.39298	1331.2609	442.33946
GovofCO	179.27049	837.8525	127.69672

- d. Who is the user with the most likes? Show the number of likes and the user.

```
SELECT twitter_id, SUM(likes) AS total_likes
  FROM twt_sub1
 GROUP BY twitter_id
 ORDER BY total_likes DESC
 LIMIT 1
```

Table 4: 1 records

twitter_id	total_likes
GovRonDeSantis	796094

Item 2

Query the twitter dataset such that you will be able to extract rows from female users aged 50 and above. Retain only the text, party, and the twitter ID. Save this as a data frame named `twt_sub2` and write this in your current database (i.e., in memory). Answer the following questions by querying from `twt_sub2` and by displaying the specific answer only via SQL queries.

```
SELECT twitter_id, text, party
  FROM twt_data
 WHERE gender = "Female" AND age >= 50
```

```
dbWriteTable(con, "twt_sub2", twt_sub2)
```

- a. How many of the tweets (`text` column) in `twt_sub2` has the word “job” in it?

```
SELECT COUNT() AS n_tweets
  FROM twt_sub2
 WHERE text LIKE '%job%'
```

Table 5: 1 records

n_tweets
64

- b. Who is the user who mentioned the word “job” more than the others? Show the number and the user.

```
SELECT twitter_id, COUNT() AS n_job_mentions
  FROM twt_sub2
 WHERE text LIKE '%job%'
 GROUP BY twitter_id
 ORDER BY n_job_mentions DESC
 LIMIT 1
```

Table 6: 1 records

twitter_id	n_job_mentions
GovLauraKelly	23

Item 3

Query the twitter dataset such that you will be able to extract the variables for twitter ID, `from_user`, and `from_content`. Save this as a data frame named `twt_sub3` and write this in your database (i.e., in memory). Answer the following questions by querying from `twt_sub3` and by displaying the specific answer only via SQL queries.

```
SELECT twitter_id, from_user, from_content
  FROM twt_data
```

```
dbWriteTable(con, "twt_sub3", twt_sub3)
```

- a. How many observations have content from the variable from_user (hint: explore the data to understand what it means to ‘have content’ and ‘not have content’ from the variable from_user)?

```
SELECT COUNT() AS n_observations  
FROM twt_sub3  
WHERE from_user != "None"
```

Table 7: 1 records

n_observations
2673

- b. How many observations have content from the variable from_content (same hint as item b)?

```
SELECT COUNT() AS n_observations  
FROM twt_sub3  
WHERE from_content != "None"
```

Table 8: 1 records

n_observations
2662