

## OUTPUT:

Q1>

- 1) How many tweets are there in total?

count

-----

110574

(1 row)

- 2) How are these tweets distributed across languages? Write a query that shows, for every language ( `user_lang` ) the number of tweets in that language.

total\_tweets\_per\_language | user\_lang

-----+-----

1405 | ar

7 | ca

2 | cs

75 | de

2 | el

74077 | en

17910 | es

1 | eu

1 | fi

9 | fil

231 | fr

1 | hu

1423 | id

26 | it

9861 | ja

691 | ko

14 | msa

61 | nl

1 | no

```

4 | pl
3977 | pt
396 | ru
2 | sv
233 | th
123 | tr
1 | ur
26 | zh-cn
14 | zh-tw
(28 rows)

```

- 3) Compute, for each language, the *fraction* of total tweets that have that language setting, as well as the fraction of the number of users that have that language setting.

```
user_lang | fraction_of_total_tweets
```

```

-----+-----
ar      | 0.01270642284804746
ca      | 6.330602130699803e-05
cs      | 1.8087434659142295e-05
de      | 0.000678278799717836
el      | 1.8087434659142295e-05
en      | 0.6699314486226419
es      | 0.16197297737261923
eu      | 9.043717329571148e-06
fi      | 9.043717329571148e-06
fil     | 8.139345596614032e-05
fr      | 0.0020890987031309347
hu      | 9.043717329571148e-06
id      | 0.012869209759979742
it      | 0.00023513665056884982
ja      | 0.08918009658690108

```

ko		0.006249208674733662
msa		0.00012661204261399606
nl		0.0005516667571038399
no		9.043717329571148e-06
pl		3.617486931828459e-05
pt		0.03596686381970445
ru		0.0035813120625101742
sv		1.8087434659142295e-05
th		0.002107186137790077
tr		0.001112377231537251
ur		9.043717329571148e-06
zh-cn		0.00023513665056884982
zh-tw		0.00012661204261399606

(28 rows)

Q2>

1) What fraction of the tweets are retweets?

fraction

-----

0.19387921211134626

(1 row)

2) Compute the average number of retweets per tweet.

average

-----

121.27

(1 row)

3) What fraction of the tweets are never retweeted?

fractions\_never\_tweeted

-----

0.6692079512362762

(1 row)

- 4) What fraction of the tweets are retweeted fewer times than the average number of retweets (and what does this say about the distribution)?

retweeted\_fewer\_times

-----

0.9361965742398756

(1 row)

Q3>

- 1) What is the number of distinct hashtags found in these tweets?

count

-----

10158

(1 row)

- 2) What are the top ten most popular hashtags, by number of usages?

distinct\_hashtag | total\_count

-----+-----

ReasonsIFailAtBeingAGirl | 467

RED | 240

oomf | 190

HonestyHour | 172

TeamFollowBack | 139

EresGuapaSi | 130

10PeopleYouTrulyLove | 126

TweetLikeAGirl | 98

ImSingleBecause | 97

WeAllGotThatOneFriend | 96

(10 rows)

Q4>

1) How many tweets are neither replies, nor replied to?

no\_reply

-----

84798

(1 row)