

Exercise #2

Twitter Dataset Analysis

This report presents the results of data analysis conducted on a twitter dataset. The analysis focused on identifying the 10 most active senders, the 10 most re-tweeted tweets and the 10 most cited screen names along with commonalities between the top cited screen names and the top senders.

Key findings from the analysis:

Most active senders:

These are the 10 Most Active Senders:

```
klansmen4trump      90
facists4trump       55
dawngpsalm63        26
skyjones55          24
jessnatenuff        24
FedoraPayan         23
Genie115            21
nathanTbernard      19
PatrickAnna         18
teed_chris          18
Name: Sender, dtype: int64
```

Most retweeted tweets:

Top 10 Most Retweeted Texts:

	Text	RetweetCount
0	RT @zzzeeshaan: Alan Rickman died when he was ...	1642
1	RT @twaimz: 2016. 2 stick horses (one unicor...	820
2	RT @deray: Donald Trump. 2016. https://t.co/xg...	548
3	RT @revivaIariana: David Bowie died at 69. Ala...	415
4	RT @NathanZed: I cut together Donald Trump's r...	379
5	RT @leezachariah: Very sad to report that Dona...	357
6	RT @TheTweetOfGod: SPOT THE MISSING NUMBER Da...	182
7	RT @mylastdilemma: David Bowie: 69 ans Alan Ri...	139
8	RT @AnneAnneAss: Michel Delpech : 69 ans. Davi...	135
9	RT @sahilkapur: Staggering statistic in the NB...	120

Most cited screen names:

Top 10 Most Cited Screen-Names:

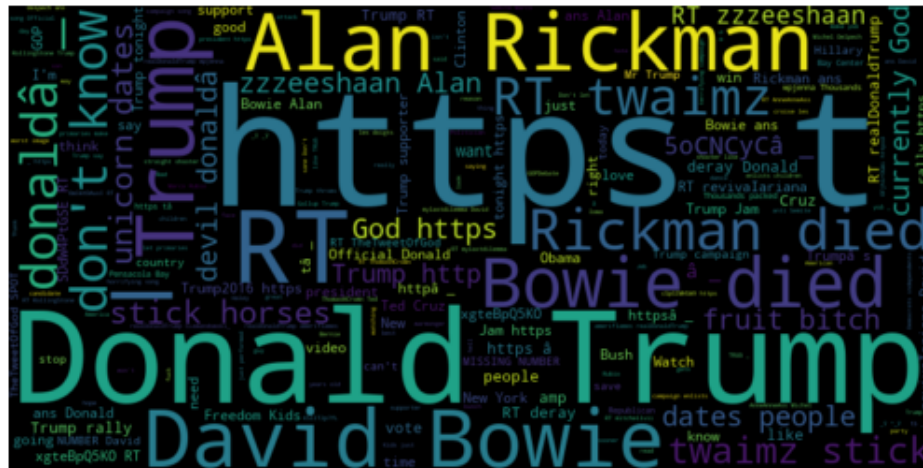
```
zzzeeshaan      1647
realDonaldTrump 1171
twaimz          855
deray           682
revivaIariana   415
NathanZed       384
leezachariah    357
benshapiro      205
TheTweetOfGod   193
CLewandowski_   186
dtype: int64
```

Commonalities between cited screen names and active senders:

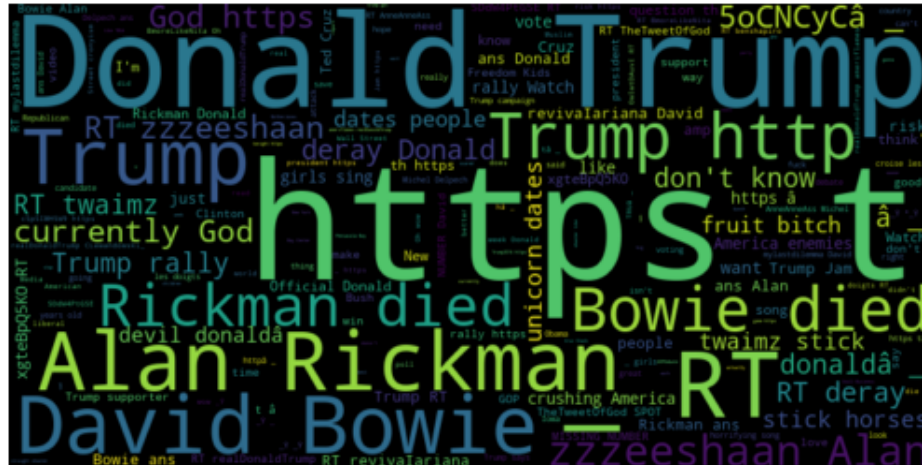
Common Screen-Names in Top Cited and Top Senders:
Index([], dtype='object')

WordClouds:

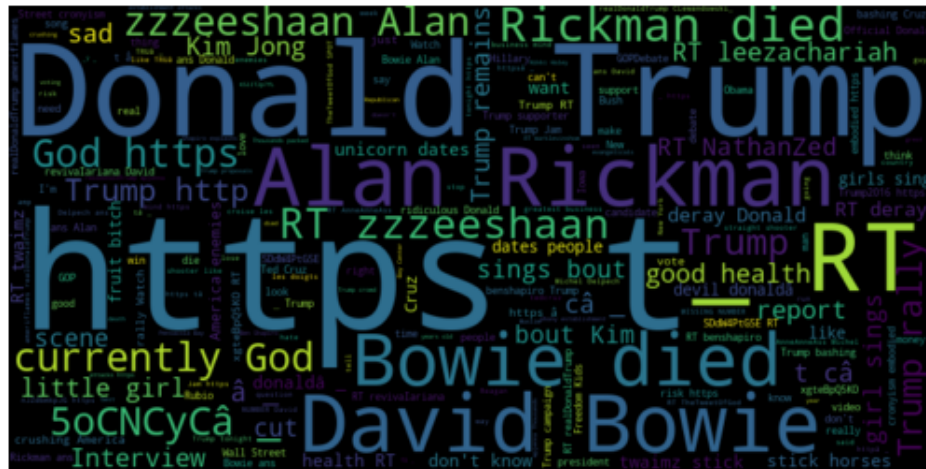
Subset 1 - Start Time: 2016-01-15 02:23:29, End Time: 2016-01-15 02:34:53



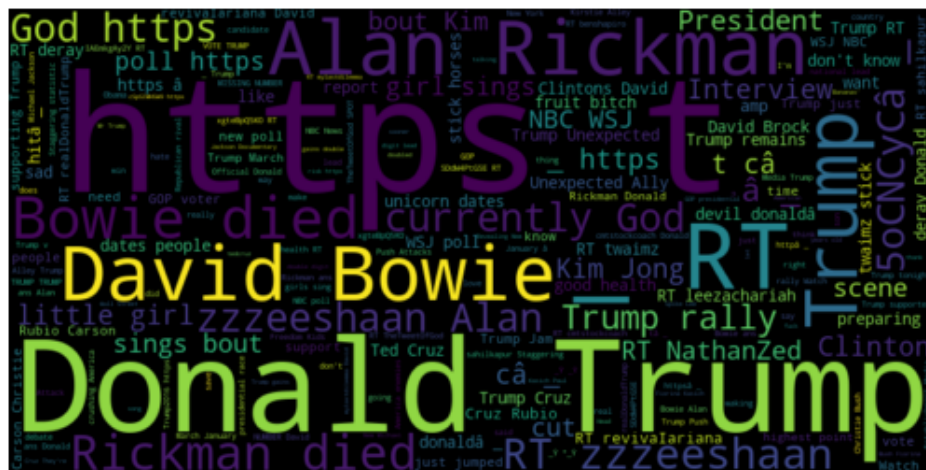
Subset 2 - Start Time: 2016-01-15 02:34:54, End Time: 2016-01-15 02:46:00



Subset 3 - Start Time: 2016-01-15 02:46:00, End Time: 2016-01-15 02:55:43



Subset 4 - Start Time: 2016-01-15 02:55:43, End Time: 2016-01-15 03:16:15



Subset 5 - Start Time: 2016-01-15 03:16:15, End Time: 2016-01-15 03:25:35

