1. Print the 5 most active senders:

After reading the input file as list, extracted the sender data and applied the counter function in Python to calculate the most active senders. The image below shows the 5 most active senders.

```
lst2 = [item[0] for item in tweet]
Sender = Counter(lst2).most_common(5)
Sender
```

```
[('klansmen4trump', 90),
 ('facists4trump', 55),
 ('dawngpsalm63', 26),
 ('jessnatenuff', 24),
 ('skyjones55', 24)]
```

2. Print the 10 most re-tweeted tweets:

The next step was to extract tweets data only and get the tweets that begin with 'RT'. Stored the retweets in a list and applied the counter function once again to calculate the most retweeted tweet. The image below shows the 10 most retweeted tweets:

The 10 most re-tweeted tweets

```
lst3 = [item[2] for item in tweet]
lst3
r = []
for x in lst3:
    if "RT" in x:
        r.append(x)
RT = Counter(r).most_common(10)
RT
```

```
3]: [('RT @zzzeeshaan: Alan Rickman died when he was 69, David Bowie died when he was 69, Donald Trump is currently 69, @ God ht
    tps://t.co/5oCNCyC\x89Û_',
      1642),
     ("RT @twaimz: 2016.   2 stick horses (one unicorn) 0 dates (people not the fruit) 1 bitch (me) 6 something. i don't know 66
    6 the devil donald\x89Û_",
      820),
     ('RT @deray: Donald Trump. 2016. https://t.co/xgteBpQ5KO', 548),
     ('RT @revivaIariana: David Bowie died at 69. Alan Rickman died at 69. Donald Trump is 69. https://t.co/SDdW4PtGSE',
      415),
     ("RT @NathanZed: I cut together Donald Trump's rally and the scene from The Interview when the little girl sings bout Kim J
    ong Un https://t.c\x89Û_",
      379),
     ('RT @leezachariah: Very sad to report that Donald Trump, 69, remains in good health.',
      357),
     ('RT @TheTweetOfGod: SPOT THE MISSING NUMBER  David Bowie (1947 - 2016) Alan Rickman (1946 - 2016) Donald Trump (1946 -
    )',
      182),
     ('RT @mylastdilemma: David Bowie: 69 ans Alan Rickman: 69 ans Donald Trump: https://t.co/OwIwtDAuvI',
      139),
     ('RT @AnneAnneAss: Michel Delpech : 69 ans. David Bowie : 69 ans. Alan Rickman : 69 ans.  Donald Trump a 69 ans, on croise
    les doigts.',
      135),
     ('RT @sahilkapur: Staggering statistic in the NBC/WSJ poll  % of GOP voters who can see themselves supporting Trump  March
    2015: 23% January \x89Û_',
      120)]
```

3. Print the 5 most cited screennames

To get the most cited screennames, the list was split into single words and extracted words that started with '@'. However, there was one screenname that had space between @ and the name. The screenname is @God which can be seen in the first tweet in the above image of retweets. The image below shows the 5 most cited screennames:

```
res = []
res1 = []
sname = []
for x in lst3:
    temp = x.split()

    for ele in temp:
        if (ele[0] == '@') and (len(ele) > 2):
            res.append(ele)
        elif (ele == "God"):
            res1.append(ele)
res
res1
aces = ["@" + r for r in res1]
for x in aces:
    res.append(x)
screenname = Counter(res).most_common(5)
screenname
```

```
[('@God', 1704),
 ('@zzzeeshaan:', 1642),
 ('@twaimz:', 846),
 ('@realDonaldTrump', 725),
 ('@deray:', 653)]
```

4. Print the 10 most popular hashtags words:

To get the most popular hashtag words, the above logic was used. Split the tweets into single words and extracted words that started with #.
The image below shows the 10 most popular hashtag words:

```
res1 = []

for x in lst3:
    temp = x.split()

    for ele in temp:
        if (ele[0] == '#') and (len(ele) > 2):
            res1.append(ele)
Hashtag = Counter(res1).most_common(10)
Hashtag
```

```
[('#Trump', 411),
 ('#Trump2016', 362),
 ('#MakeAmericaGreatAgain', 161),
 ('#GOPDebate', 154),
 ('#RealDonaldTrump', 146),
 ('#tcot', 93),
 ('#TRUMP', 88),
 ('#trump', 82),
 ('#DonaldTrump', 59),
 ('#TCOT', 58)]
```

5. Dividing the dataset into 5 subsets and create a wordcloud for each subset:

The first step was to sort the dataset by timestamp and divide into 5 subsets. After dividing the dataset into subset, eliminated the stop words from all subsets and then applied the wordcloud function from the wordcloud library. Below are the 5 wordclouds from 5 subsets:

I. Subset 1:
The below image show that the possible topics from the first subset would be:
   i. Presidential elections
   ii. Trumps campaign about 'Make America Great Again'

II. Subset 2:

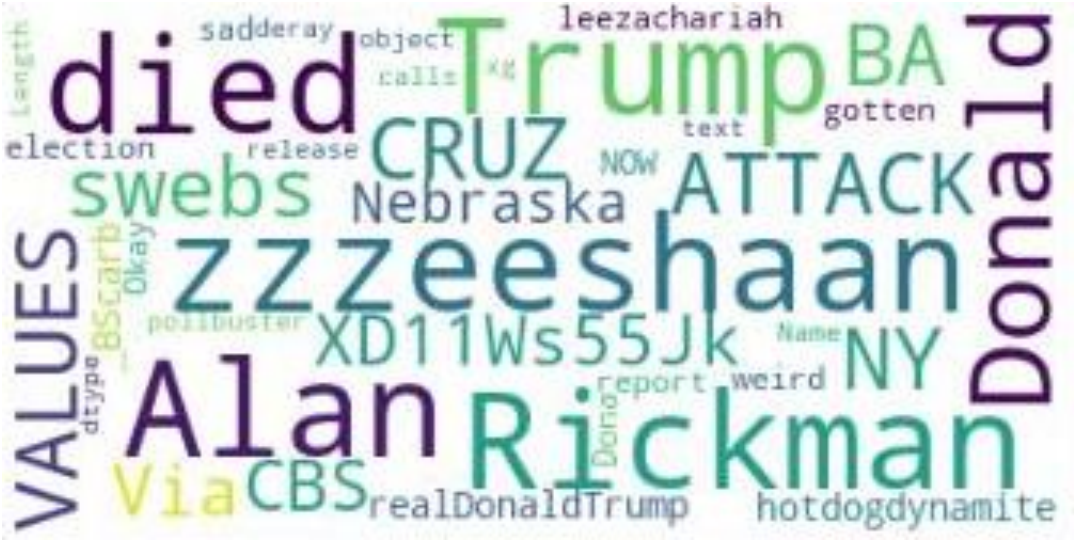
The possible topics in subset 2:

i. Presidential election
ii. McKay Coppins
iii. Kasey Denny

III. Subset 3:

The possible topics in subset 3:

i. Death of Alan Rickman
ii. Via Cabs
iii. Elections



IV. Subset 4:

The possible topics in subset 4:

i. Ted talks

ii. Topics about Israel

iii. Refugee

V. Subset 5:

The possible topics in subset 2:

i. Voting against Donald Trump

ii. Politics