

# Pa3-extra\_report

Tianyou Huang

## Part 1

Run pymongo\_tweepy.py to get tweets.

Run pa3\_part1.py

The first part of output is for question C, sentiments for tweets.

The second part is for question A, tweets that have "data" somewhere in text.

The last part is for question B, the number of tweets that have their geo\_enabled field true.

```
neutral sentiment for the tweet: u_t=18x^2u_xx+(18x+5t)u_x+2u https://t.co/rrFFcdZr5Y #bigdata #analytics http
neutral sentiment for the tweet: Conheça o BigID: https://t.co/gE2bUbU6DZ #BigData #ValidacaodeIdentidade #Frau
neutral sentiment for the tweet: Selecteam GmbH: Gesucht: Solution Architect m/w BIG DATA / DATA WAREHOUSE an v
positive sentiment for the tweet: RT @odsc: CatBoost: Yandex's machine learning algorithm is available free of
neutral sentiment for the tweet: What Is EdTech Doing And What's Left? https://t.co/b4X2205zv0 #DataScience #Da
neutral sentiment for the tweet: The internet of things Daily! https://t.co/1B8XWj3dbc #IoT #customerexperience
```

The number of tweets that have 'data' somewhere in the tweet's text(case insensitive): 783 for question A

192 of them are geo\_enabled. for question B

Process finished with exit code 0

## Part 2

A. Run pa3\_part2\_A.py to get tweets.

B.

Run pa3\_part2\_B.py

```
/Users/richard/anaconda/bin/python /Users/richard/Desktop/cs_fall_2017/cs660/pa3_extra/mongo_tweets/pa3_part2_B.py
Top 15 emojis used in the entire tweets:
[{'': 104}, {'': 92}, {'': 71}, {'': 71}, {'': 64}, {'': 55}, {'': 55}, {'': 55}, {'': 53}, {'': 53}, {'': 45}, {'': 43}, {'': 34}, {'': 32}, {'': 30}]
Top 5 states for the emoji :
[{'CA': 20}, {'TX': 9}, {'FL': 4}, {'GA': 4}, {'NC': 3}]
Top 5 emojis for MA:
[{'': 3}, {'': 2}, {'': 2}, {'': 2}, {'': 2}]
Top 5 states that use emojis:
[{'CA': 477}, {'NY': 311}, {'FL': 257}, {'TX': 194}, {'NV': 83}]
Process finished with exit code 0
```

C.

Run pa3\_part2\_C.py

```
/Users/richard/anaconda/bin/python /Users/richard/Desktop/cs_fall_2017/cs660/pa3_extra/mongo_tweets/pa3_part2_C.py
Top 5 states that have tweets :
[{'CA': 1614}, {'NY': 918}, {'TX': 882}, {'FL': 603}, {'IL': 368}]
Top 5 cities that tweet in the state of California:
[{'_id': 'Los Angeles', 'count': 304},
 {'_id': 'San Francisco', 'count': 122},
 {'_id': 'San Diego', 'count': 88},
 {'_id': 'Oakland', 'count': 35},
 {'_id': 'San Jose', 'count': 34}]
Process finished with exit code 0
```

D.

First, run json\_to\_csv.py (I made some changes from the original json\_to\_csv.py, so please run my version here.)

Then, run pa3\_part2\_D.py

It takes quite a time to draw the map, I print the current number of points in terminal, so you can feel the process.

Get the result:

map.html:

