CS660-PA3 Extra Write Up

Gur Asees Singh Chandok -U95489771

Part 1) A. No of tweets that have data somewhere in the text:

Code for this problem can be found in part1_a.py .

Output: Count of Tweets that contain data somewhere in their text = 72

Snippet:

```
for tweets in all_tweets:
    if re.search('(?<=(?<=(?<=d)a)t)a', tweets['text'], re.IGNORECASE) is not None:
        count_data =count_data+1
```

B. Data Related objects with geo_enabled

Code for this problem can be found in part1_b.py .

Output: Count of data related objects that are geo_enabled = 9

Snippet:

```
for tweets in all_tweets:
    if re.search('(?<=(?<=(?<=d)a)t)a', tweets['text'], re.IGNORECASE) is not None:
        user = tweets['user']
        if user['geo_enabled']:
            count_geo_enabled += 1
```

C. Sentiment analysis on tweet

Code for this problem can be found in part1_C.py .

Portion of Output:

Positive | https://t.co/r2oRkmuYca Thanks to @gritgrindhustle @haldaume3 @orangerose #bigdata

Neutral | RT @SachinLulla: 10 Things You Need to Know About #AI, #BigData, and Analytics in 2018 https://t.co/95fmQl8PTZ #DataScience #IoT…

Positive |The latest The Project management Daily!

Positive | https://t.co/Cscbx1FFZw Thanks to @CorrectDEV @TruckPlantSales @rvvargas #bigdata #blockchain

Negative | RT @Ronald_vanLoon: Machine Learning Is Not Magic: It’s All About Math, Stats, Data, and Programming | #MachineLearning #Python #RT…

Snippet:

```
for tweets in all_tweets:
    if re.search('(?<=(?<=(?<=d)a)t)a', tweets['text'], re.IGNORECASE) is not None:
        blob = TextBlob(tweets['text'])
        for sentence in blob.sentences:
            if sentence.sentiment.polarity < 0:
                print("Negative  | "+str(sentence))
            elif sentence.sentiment.polarity > 0:
                print("Positive | "+str(sentence))
```

```
            else:
                print("Neutral   | "+str(sentence))
```

Part 2.A.

Code for this problem is located in part2_a.py

```
streamer.filter(locations=[-125,25,-65,48]) #was used to fetch tweets from usa

#used to keep only fields that have coordinates
coordinatesField=str(datajson['coordinates'])

#used to limit the number of tweets to 10000
if coordinatesField!="None":
    db.usa_tweets_collection.insert(datajson)
    no_of_tweets=db.usa_tweets_collection.find().count()
    print("No of Records in DB="+str(no_of_tweets))
    if no_of_tweets>9999:
        return False
```

B.

1. **What are the top 15 emojis used in the entire tweets?**
2. **What are the top 5 states for the emoji 🎄?**
3. **What are the top 5 emojis for MA?**
4. **What are the top 5 states that use emojis?**

Code for these problems is in part2_b.py

Output:

Top 15 Emoji's Used are as below:

[('▓', 218), ('👇', 164), ('🔥', 101), ('❤', 91), ('🎄', 76), ('😂', 71), ('😭', 64), ('💯', 55), ('▓', 53), ('▒', 52), ('▒', 51), ('😍', 45), ('😱', 44), ('🌸', 38), ('😁', 32)]

Top 5 states for emoji '🎄' are:

[('CA', 21), ('FL', 9), ('MD', 8), ('IL', 6), ('NJ', 5)]

Top 5 emojis for MA are:

[('▒', 3), ('♀', 3), ('💝', 3), ('💸', 2), ('👉', 2)]

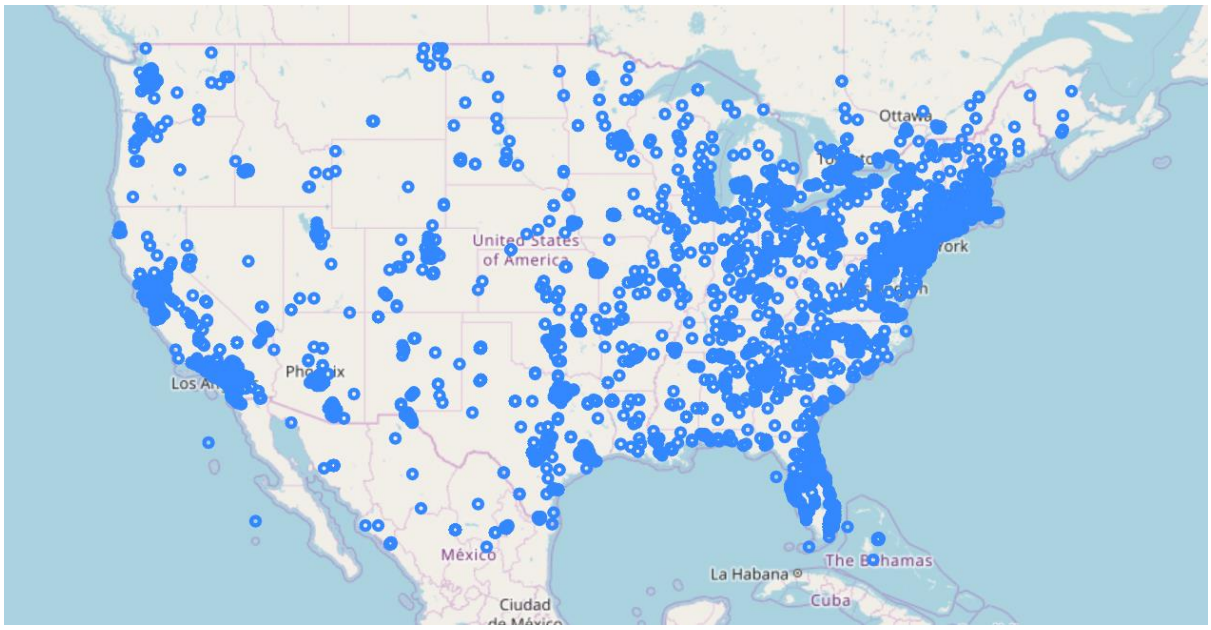
Top 5 states that use emojis are

[('CA', 850), ('NY', 541), ('FL', 291), ('TX', 226), ('GA', 105)]

Top 2 emoji's for each state are listed below, these can be plotted on a map using show_map_extra_credit.py

{'CA': [('▓', 168), ('👇', 162)], 'FL': [('▒', 11), ('😱', 11)], 'NY': [('🔥', 42), ('🌸', 38)], 'GA': [('🔥', 12), ('⚡', 10)], 'NV': [('🎰', 3), ('▓', 2)], 'VA': [('☝', 3), ('❌', 2)], 'OR': [('🎄', 4), ('💔', 2)], 'MS': [('😂', 3), ('💯', 1)], 'IL': [('😍', 8), ('🎄', 6)], 'NJ': [('🔥', 6), ('🎄', 5)], 'WA': [('🍁', 3), ('🔥', 2)], 'PA': [('🀄', 13), ('😱', 9)], 'TN': [('🍁', 3), ('🎸', 3)], 'TX': [('😭', 61), ('🍺', 9)], 'LA': [('❤', 2), ('💪', 1)], 'CO': [('🔥', 3), ('😊', 3)], 'MI': [('💯', 12), ('🛍', 2)], 'AZ': [('😍', 3), ('😊', 3)], 'MA': [('▒', 3), ('♀', 3)], 'UT': [('❤', 5), ('🏰', 2)], 'MN': [('⚙', 3), ('❤', 2)], 'MO': [('✈', 4), ('🀄', 2)], 'DC': [('📈', 5), ('😍', 3)], 'OH': [('💯', 15), ('👀', 3)], 'AL': [('▓', 3), ('💗', 3)], 'IN': [('😂', 5), ('💯', 3)], 'MD': [('🎄', 8),

('❤', 3)], 'NC': [('▓', 8), ('▓', 6)], 'WI': [('🔥', 4), ('🎶', 4)], 'RI': [('❤', 3), ('🎄', 2)], 'SD': [('☺', 1)], 'KS': [('👂', 2), ('😎', 1)], 'CT': [('☺', 3), ('❤', 1)], 'KY': [('🎄', 2), ('🎁', 1)], 'SC': [('☺', 1), ('🀄', 1)], 'AR': [('😂', 4), ('💞', 1)], 'NH': [('❤', 2), ('⚡', 1)], 'OK': [('😷', 1), ('♂', 1)], 'IA': [('🌪', 1), ('🍷', 1)]}



d. Map of tweets for code. It is saved in part2_d_map.html

C. Use MongoDB queries within PyMongo API to answer the following:

Code can be found in part2_c.py

1. What are the top 5 states that have tweets?

Top 5 states that have tweets are

[('CA', 1424), ('NY', 931), ('TX', 546), ('FL', 544), ('IL', 498)]

2. In the state of California, what are the top 5 cities that tweet?

In the state of California, the top 5 cities that tweet are

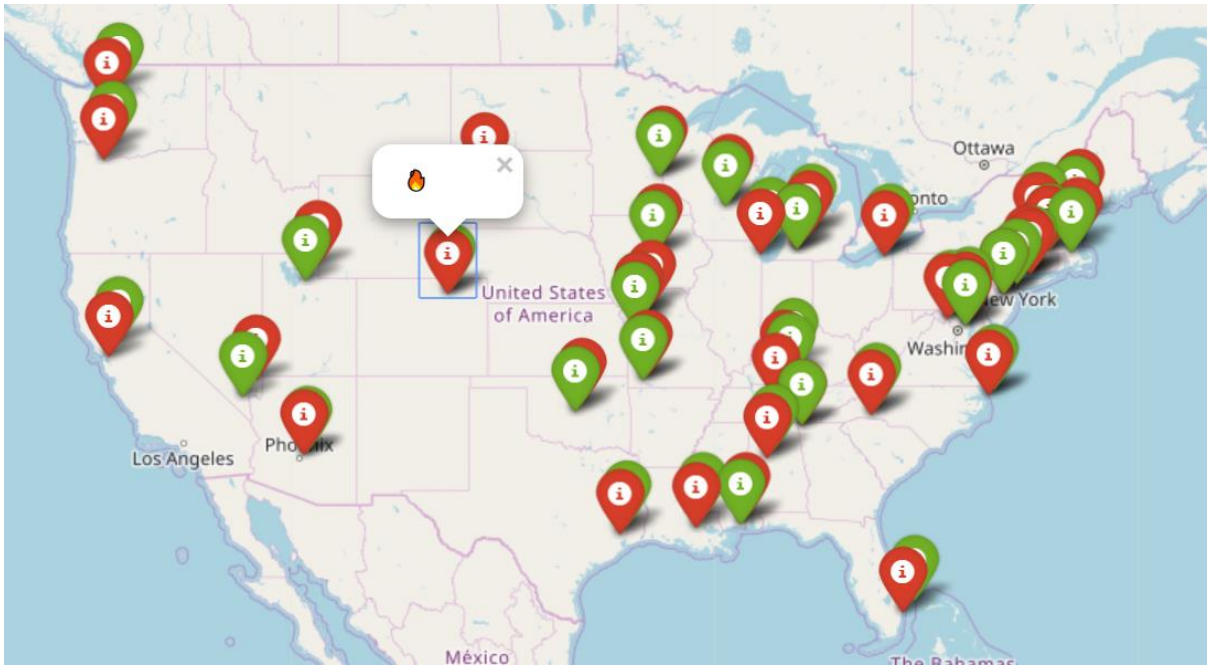
[('Los Angeles', 678), ('San Francisco', 145), ('San Diego', 78), ('Anaheim', 42), ('Oakland', 21)]

D.Extra credit

Map was plotted using folium and csv named usa_top_emoji_tweets.csv was created

Code can be found in show_map_extra_credit.py

Map on next page


Ottawa
United States
of America
Los Angeles
México