

Twitter Sentiment Analysis

Supriya Ramarao Prasanna | Chaitanya Muppala | Vanisa Achakulvisut

Objective



- The goal of our project is to understand sentiment on work from home(wfh) situation in Covid-19 in the United States and rest of the world on twitter over time.
- **Keywords searched:** “#workfromhome”, “#wfh”, “#remotework”
- Understanding of API, AWS services, Algorithms and Tableau

Workflow



Twitter Data Retrieval



Two ways of streaming

1. StreamListener
2. Tweepy.cursor

StreamListener vs tweepy.cursor

- StreamListener:
 - Features: Upto 25 filters can be used, In-built function which can be customized, Uses “stream.filter” to apply the filter.
- tweepy.cursor:
 - Features: Upto 25 filters can be used, Limit number of tweets we can sweep, Easy way of adding more filters.

Observations:

- Retrieval through tweepy.cursor was quick and easy but Stream Listener did take time until it received new real time tweets.

EC2

<https://aws.amazon.com/>



Services ▼

Search for service

Amazon S3

Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens dashboard](#)

Buckets (2)

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

	Name	AWS Region	Access	Creation date
<input type="radio"/>	ids561	US East (Ohio) us-east-2	Bucket and objects not public	May 1, 2021, 14:10:39 (UTC-05:00)
<input type="radio"/>	sagemaker-studio-068254172638-c4ydupfl8	US East (Ohio) us-east-2	Public	April 30, 2021, 20:04:00 (UTC-05:00)



Amazon EC2

```
{
  "Version": "2008-10-17",
  "Statement": [
    {
      "Sid": "AllowPublicRead",
      "Effect": "Allow",
      "Principal": {
        "AWS": "*"
      },
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::sagemaker-studio-068254172638-c4ydupfl8/*"
    }
  ]
}
```

Step 1: Create EC2 Instance Jupyter Notebook on SageMaker S3 for data storage

Step 2: Connect the S3 to Sagemaker

Option 1: Read Data Files on S3 from Amazon SageMaker

Step 1: Know where you keep your files

Step 2: Get permission to read from S3 buckets

Step 3: Use **boto3** to create a connection

Step 4: Load pickled data directly from the S3 bucket

Option 2: Make S3 Public and access by urllib

Step 1: Set S3 Public

Step 2: Give permission to read from S3 buckets

Step 3: Set up the environment and run codes

#NLTK #Textblob #Vader #Wordcloud

#Pyspark,Pandas,Numpy,Matplotlib

>> %pip , conda

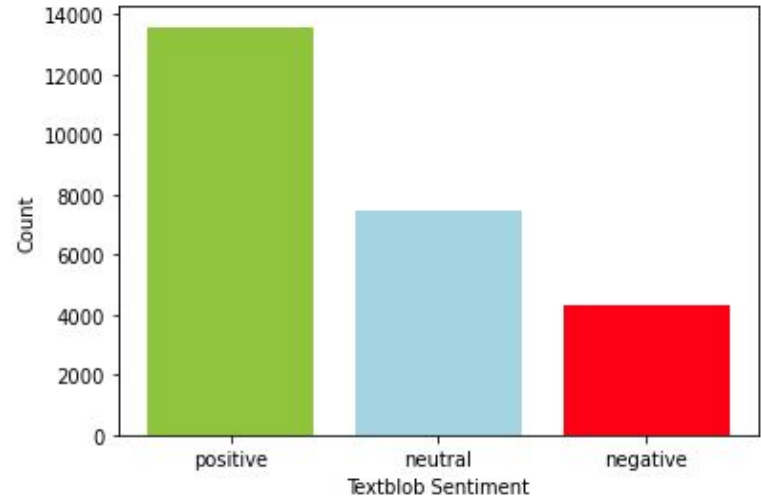
TextBlob

Aims to analyse the polarity of the tweets to understand the sentiment of people towards work from home.

Textblob is used for Sentiment analysis, Classification, Translation, and etc.

Textblob sentiment analyzer returns two output values:

1. **Polarity:** sentiment of sentences.
Range: -1 (negative) and 1 (positive).
2. **Subjectivity:** personal opinion, emotion, or judgment.
Range : 0 (objective) and 1 (subjective).



TextBlob Sentiment

+-----+-----+	
sentiment	count
+-----+-----+	
positive	13586
neutral	7471
negative	4333
+-----+-----+	

Vader

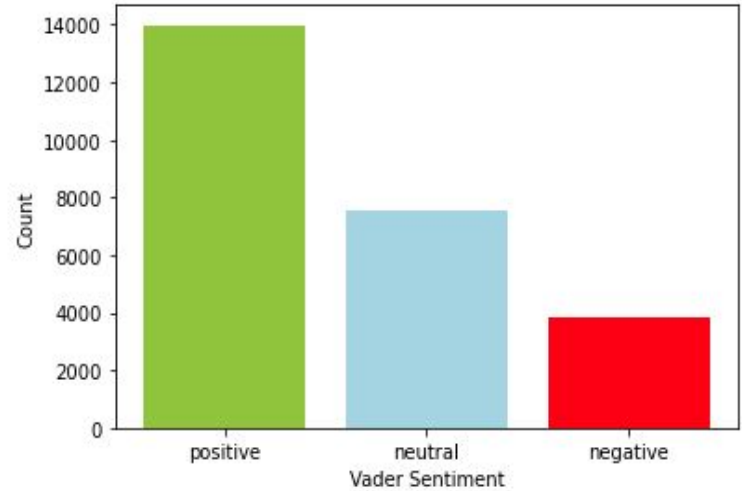
Vader sentiment analyzer returns four output values: positive score, negative score, neutral score and compound score.

Beside the Sentiment Lexicon, in Vader, it is incorporated the heuristic rules; “**Degree modifiers**” which boosted the positivity in the **compound score of a sentence**.

Compound score metric evaluation of positive and negative sentiment classification.

1. positive sentiment: `compound score >= 0.05`
2. neutral sentiment: `(compound score > -0.05) and (compound score < 0.05)`
3. negative sentiment: `compound score <= -0.05`

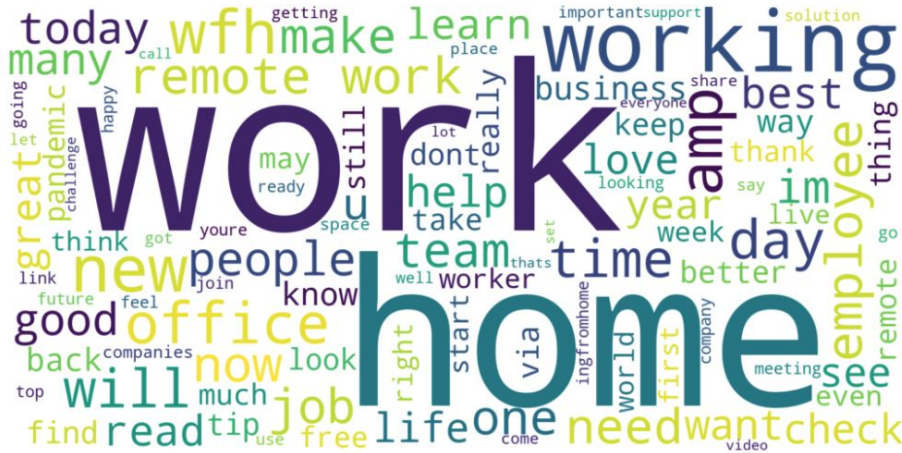
compound score metric



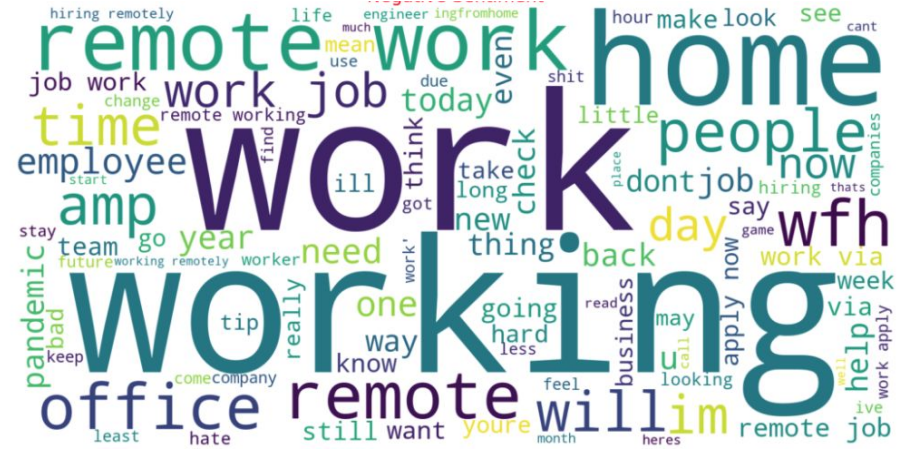
Vader Tweet Sentiment

positive	13980
neutral	7574
negative	3836

Results and Discussion



Positive Sentiment



Negative Sentiment

Conclusion

- We observed that there are differences with the number of tweets and polarity trend w.r.t date, but both of the algorithms show **overall positive sentiment trend** in the analysis.
- This indicates that people are having **positive outlook towards working from home**.
- From our analysis, we recommended that companies can consider hybrid work culture or wfh in the future.

Future Scope

- As part of our future scope, we want to explore AWS Kinesis data stream to stream our tweets with tweepy.
- We want to identify a method to remove advertisements from the tweets collected.

Thank you

Supriya Ramarao Prasanna | Chaitanya Muppala | Vanisa Achakulvisut