



Spatial and Temporal Analysis on Twitter data using ArcGIS

Kamalika Poddar

Pratheek Chindodi Rajashekhar

Contents

- Introduction
- Related Work
- Data Collection
- Data Processing
- Data Visualization
- Results
- Challenges
- Conclusion



Introduction

- Sentiments can be leveraged to understand variation in people's emotion.
- This project analyzes the sentiment variation within a given time frame of the week.
- Explores different nearby-locations with respect to a tweet.





Related Work

Spatial and Temporal Sentiment Analysis of Twitter data

- Investigated the spatial and temporal distribution of georeferenced Twitter sentiment within the area of 1km buffer.
 - The spatial and Temporal distribution of Twitter Sentiment was mapped using Geographic Information Systems.



Spatial and Temporal Analysis: a tale of two countries

- Retrieval of relevant information from Social media is a difficult challenge as the bar for posting is very low.
- Filtering the flow of messages is critical in times of crisis.
- A spatial-temporal model that collects data from Twitter based on the density of tweets surrounding the area is proposed.
- The possibility of shared user accounts is evaluated by determining the physical distance and velocity between the messages

Tweets of one Twitter account with 8 users



Spatial, Temporal and Content Analysis of Twitter for Wildfire hazards

- Analysis of Wildfire-related Twitter activities in terms of their attributes pertinent to space, time and content.
- Social media data can characterize the wildfire across space and over time, and thus provide useful information on disaster situations.
- Local Authorities play a dominant role in the wild fire retweet network.

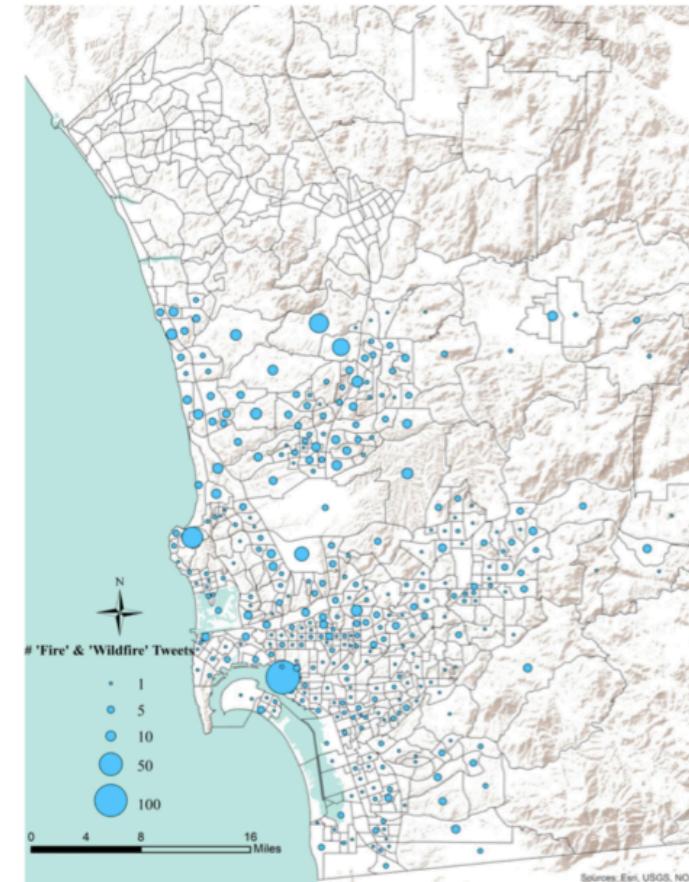


Fig. 3 Spatial distribution of geotagged "fire" and "wildfire" tweets

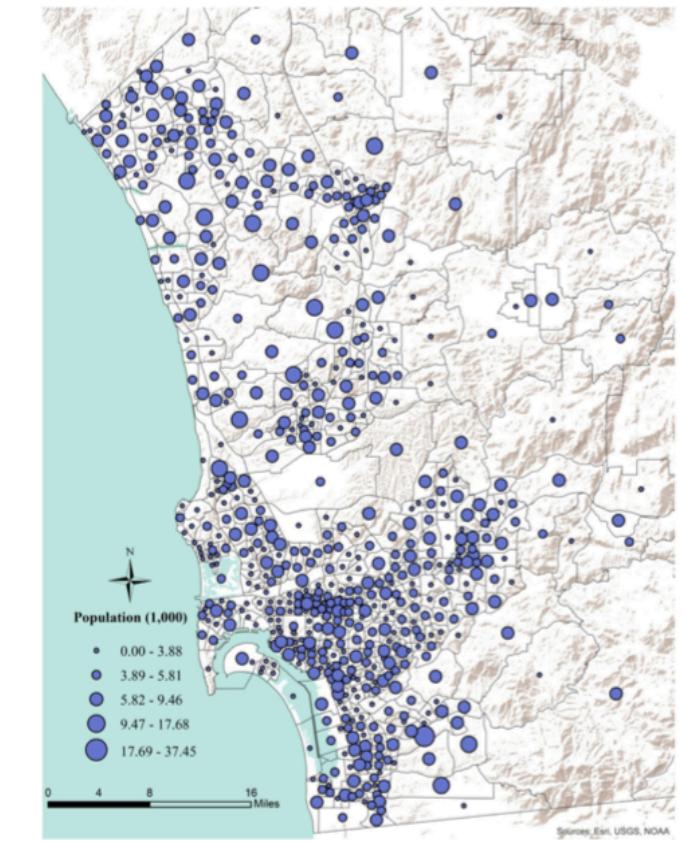


Fig. 6 Spatial distribution of population in San Diego County

Hypothesis

Explores relation between the negative tweets and the temperature of the day.

Data Collection

Used “Tweepy” library to stream current data from twitter.

Collected using positive and negative keywords.

Collected each day evening hours for 1 hour from Monday to Friday.

Keywords

Positive	Negative	Neutral
happy	shame	baffled
enjoy	doubt	authoritative
cheerful	envy	clinical
great	grief	detached
love	fear	nostalgic
enjoying	sadness	objective
challenge	frustration	restrained
learning	guilt	sentimental
curious	disgusted	candid
prefer	failure	frank
demand	afraid	preoccupied
advice	hate	unequivocal
trusting	pain	probing
unique	sick	nonchalant
like	overwhelmed	callous
easy	problem	consoling
nice	stressed	didactic
good	boring	direct
helpful	bothered	impartial
pretty	weird	unambiguous
fun	greedy	understated

Data Processing

The JSON file is converted to a GeoJSON file. From all the features in the data, only the geotagged data with the user information and geo coordinates are obtained.

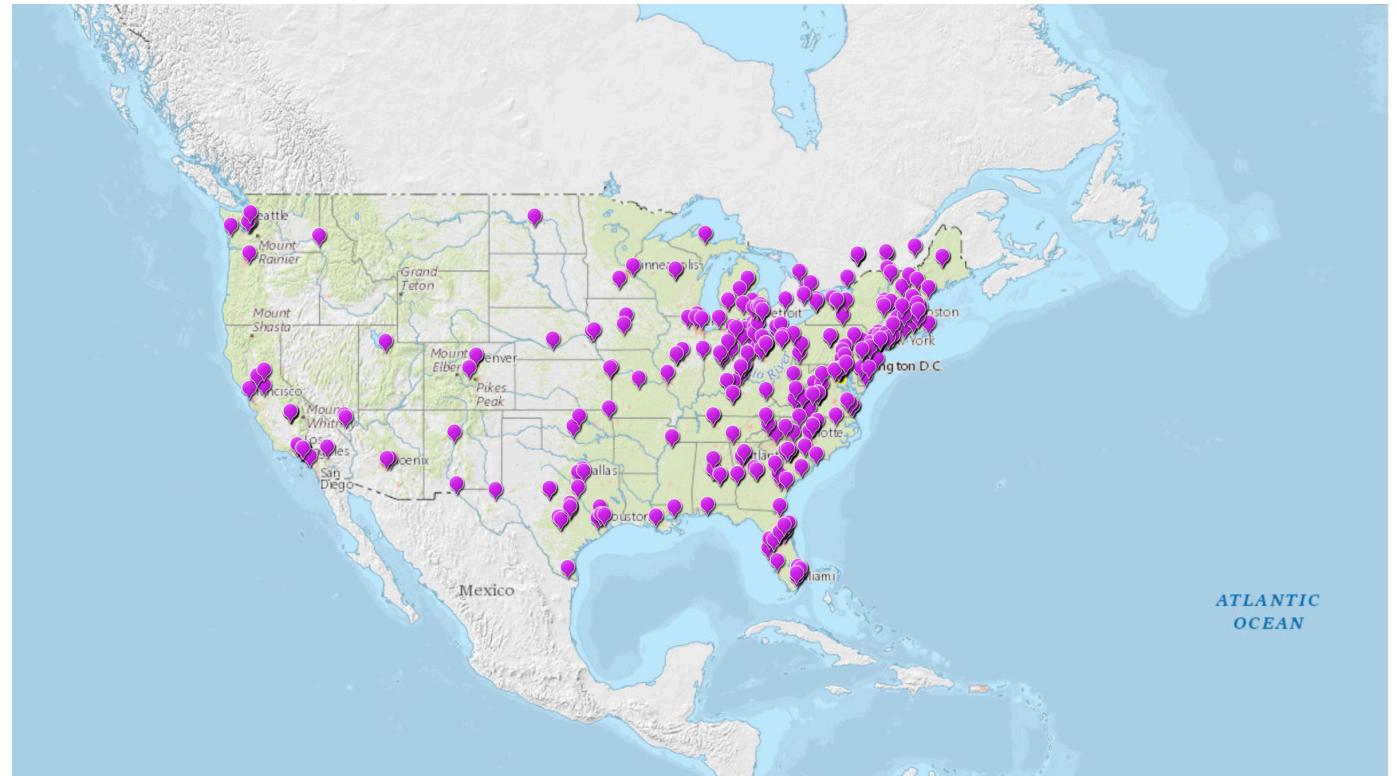


The obtained data that contained the location information was mapped into ArcGIS by adding different layers.

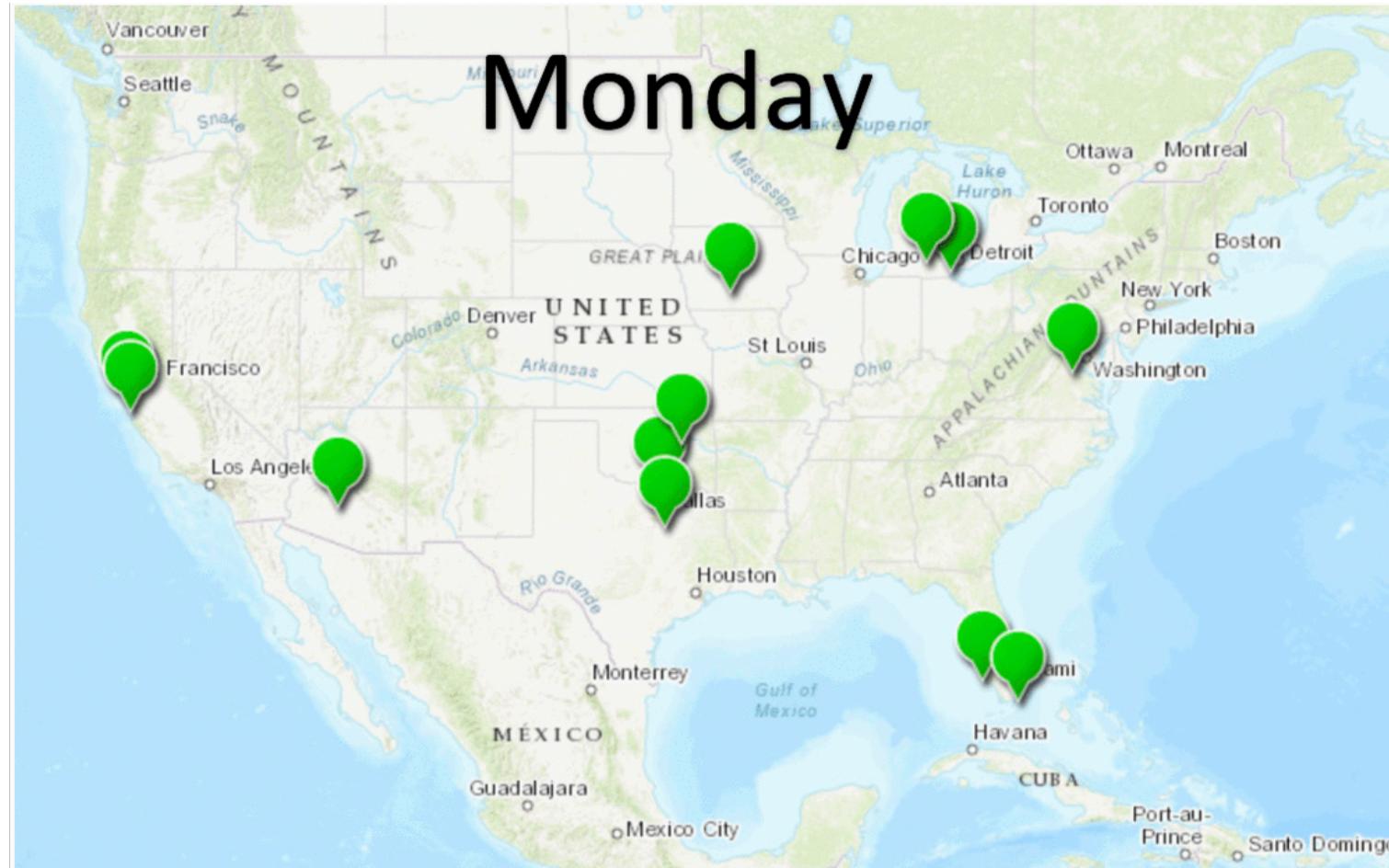


Data Visualization

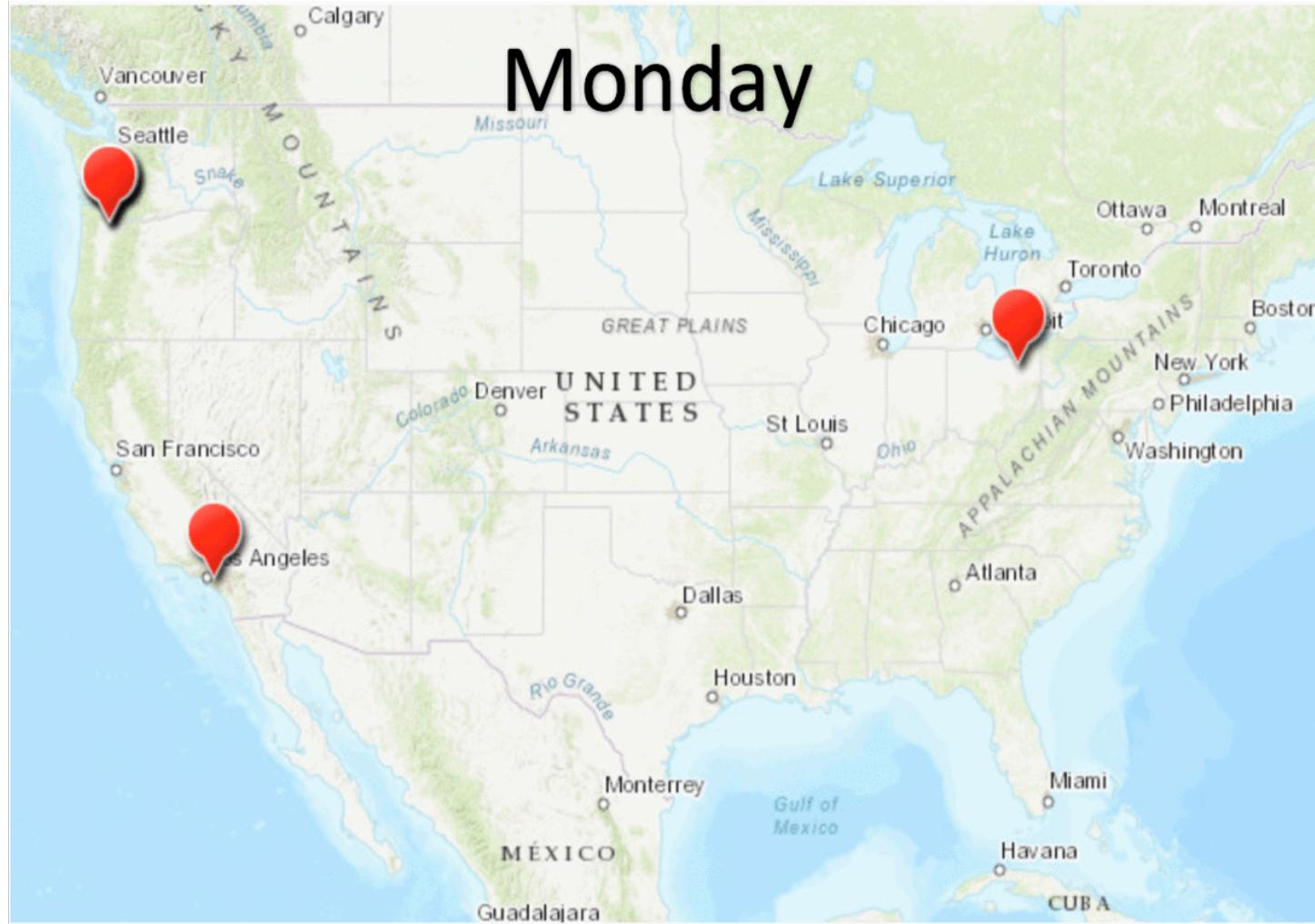
One Minute USA



Positive



Negative



Temporal Analysis

Enables us to examine and model the behavior of a variable in a data set over time.

‘Find Aggregate’ function in “Summarize” was used to count all the points in the boundary.

‘US Country Boundary 2016’ feature layer was being used for this.

Spatial Analysis

Study of entities using their topological, geometric, or geographic properties.

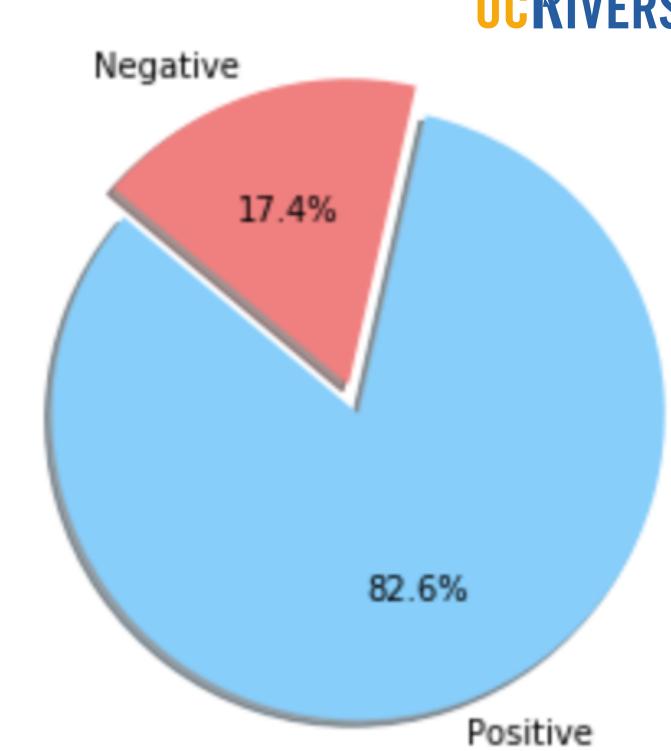
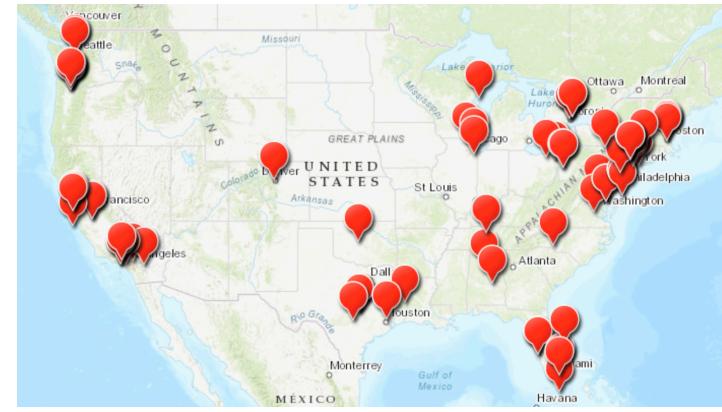
‘Find Nearest’ function within the use ‘Use proximity’ function in ArcGIS’

Find points in the layer which are nearer to Park, University and Hospital.

Analysis of negative tweets to retrieve the pattern of negative tweets origination from different location during each day of the week.

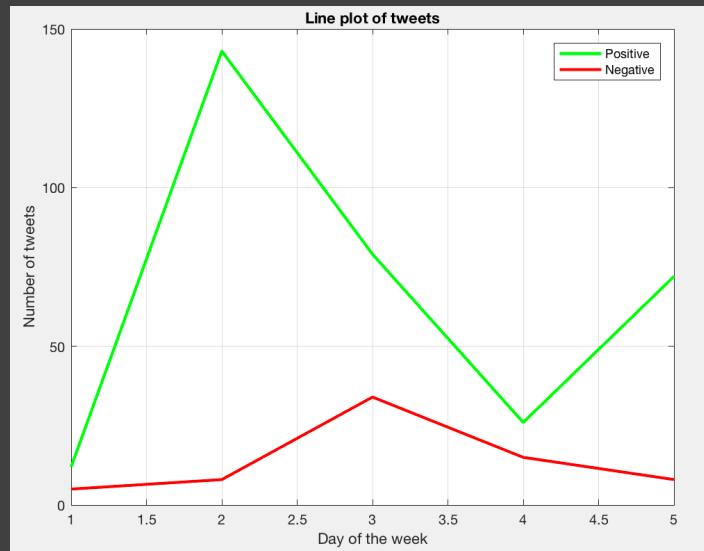


Results



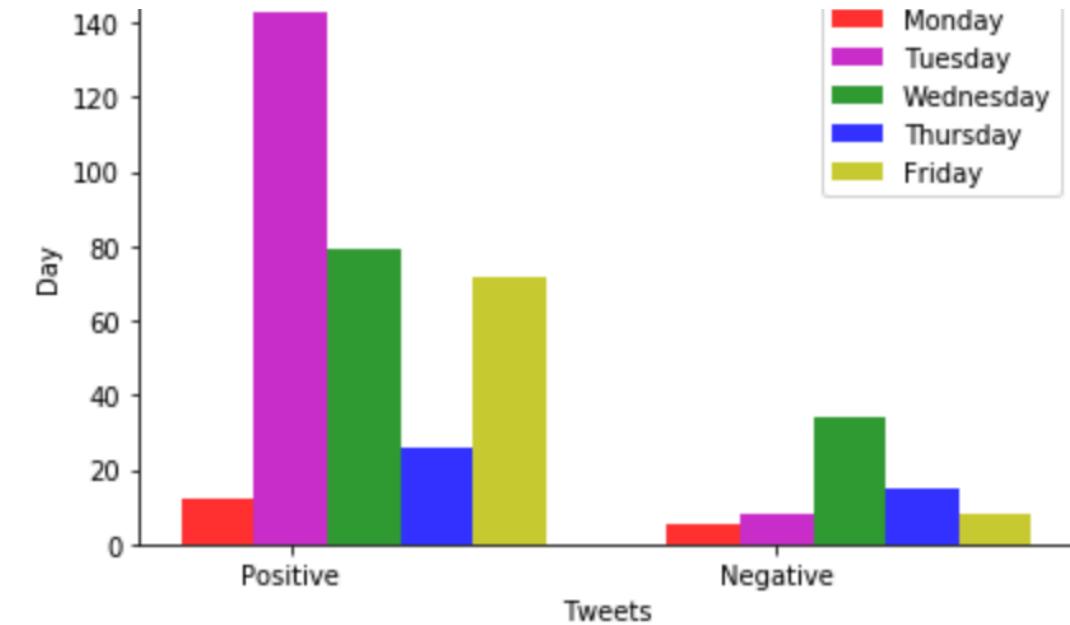
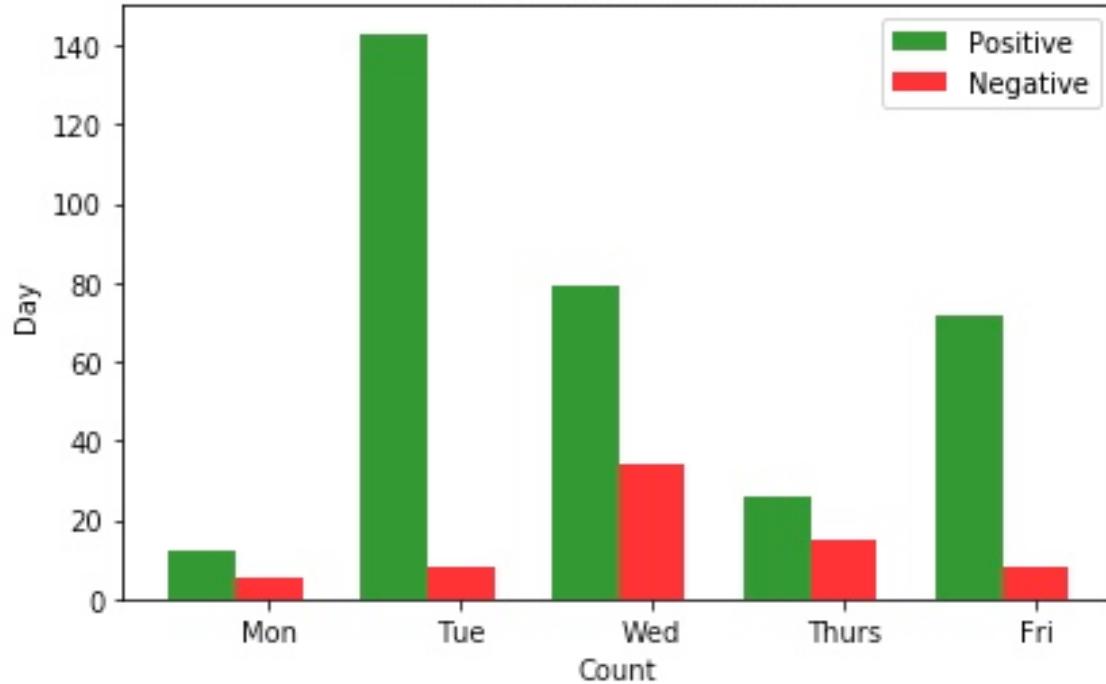
Positive and Negative Tweets

Temporal Analysis



	Positive	Negative
Monday	12	5
Tuesday	143	8
Wednesday	79	34
Thursday	26	15
Friday	72	8
Total	332	70

Bar Graph of tweets



Representation for the Week

Percentage Representation

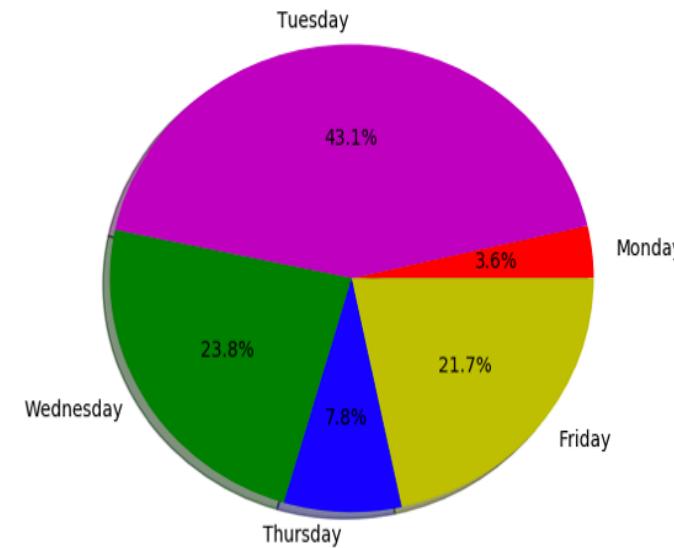


Fig a: Positive

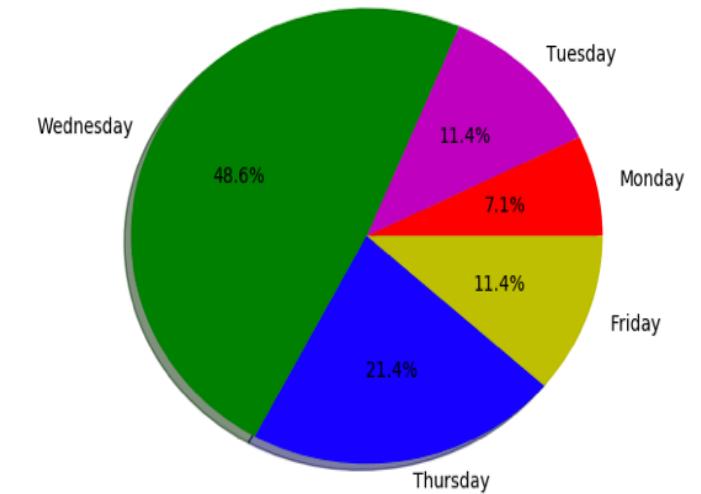
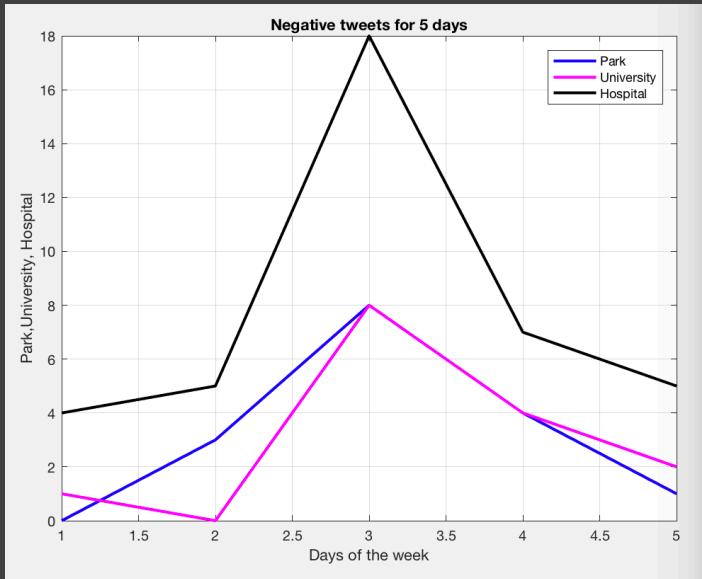
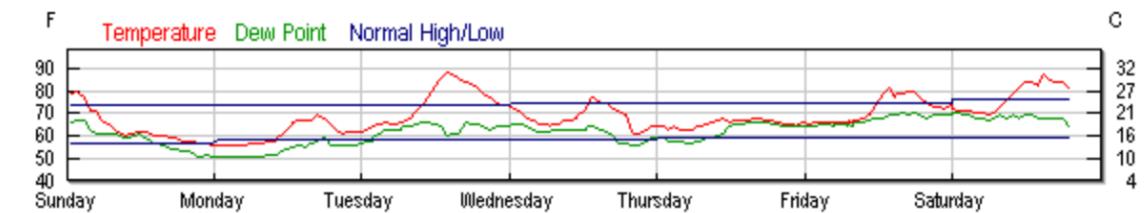
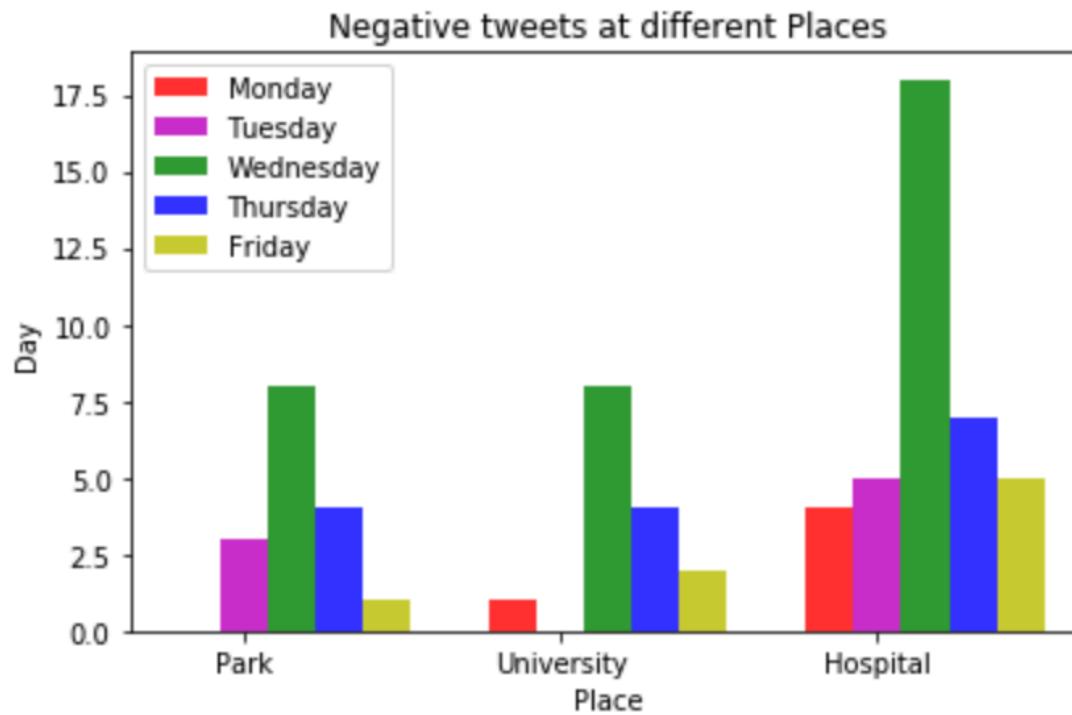


Fig b: Negative

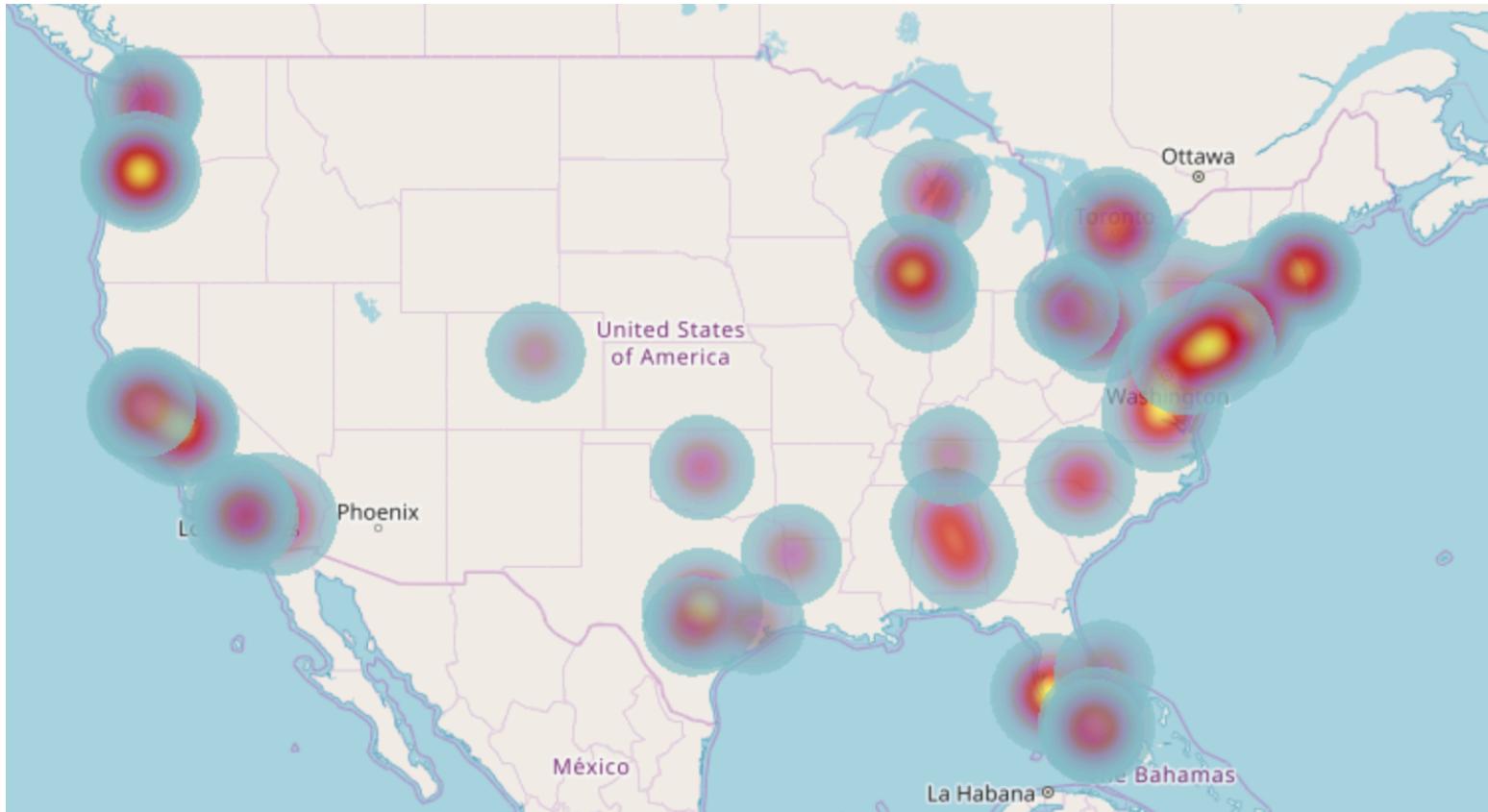
Spatial Analysis



	Park	University	Hospital	Total
Monday	-	1	4	5
Tuesday	3	-	5	8
Wednesday	8	8	18	34
Thursday	4	4	7	15
Friday	1	2	5	8
Total	16	15	39	



Spatial Analysis with Temperature



Negative Tweets

Challenges Faced

Geo co-ordinates are not enabled so not much Tweets could be generated.

Storage Issues

Online Mapping was difficult

Issues with the change of internet conditions.

Conclusion

Negative tweets generated were mostly near hospital region

Highest amount of Negative tweets generated was on **Wednesday**.

Temperature can be one of the cause, as depicted from Weather report.

Percentage of **Positive** tweets were more than the Negative tweets.

Events play an important role in generating positive and negative tweets.

Popularity of a region is a big factor in the generation of tweets.

Q/A

Thank you

