Twitter reactions on iPhone7 using Sentiment Analysis and Social Media Mining

Sunil Agrawal
Indiana University
107 S Indiana Ave
Bloomington
IN 47405
suniagra@umail.iu.edu

Pawan Pinjarkar
Indiana University
107 S Indiana Ave
Bloomington
IN 47405
ppinjark@umail.iu.edu

ABSTRACT

Apple Inc., launched its latest flagship smartphone -iPhone 7 in 2016. Apple's iPhone is immensely popular in United States. In 2016, Apple Inc. has introduced significant changes in the device model such as removal of 3.5 mm headphone jack, introduction of wireless headphones, new color options etc. With this paper we introduce an approach to discover user's sentiments from tweets about iPhone7. The sentiments discovered are majorly classified in 3 broad categories - positive, negative and neutral. We further classified these sentiments based on user's gender and location. We focused on US only location. Our approach can be useful for product companies who want to gauge public opinions about their product in the market and determine the target audience based on consumer's gender and location. Our work can be helpful to get more insights on user's sentiments based on i) gender and ii) location. Our method presents the results of supervised machine learning algorithms for classifying the sentiment of tweets. This paper also describes the preprocessing steps required to perform sentiment analysis using social media mining.

KEYWORDS

Social media mining, sentiment analysis, social media, twitter, lexicon dictionary

1. INTRODUCTION

Social media is becoming an immensely popular platform to express views, opinions and share the information among the public/private groups of people on internet. Twitter is one of the popular social media where users express their views in maximum 140 characters in text and/or images. Social media mining is a process to extract the information shared by users on social media platforms and to determine/predict the trends or behaviors from the data gathered. Sentiment analysis is a process to identify, analyze and extract the opinions from the data on social media platforms into various categories such as positive, negative, neutral.

2. TWITTER

Twitter is a popular microblogging site. Twitter was created in March 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams and launched in July 2006. As of March 2016, Twitter had more than 310 million monthly active users. Twitter users can express their thoughts in a short length message called *tweet*. Twitter is becoming popular platform to express reviews about products, movies, services etc. and hence many organizations has started monitoring and analyzing the twitter trends. The data collected from twitter is a great source of information for data scientists to perform data mining and sentiment analysis.

3. SOCIAL MEDIA MINING

"Social Media Mining, introduces basic concepts and principal algorithms suitable for investigating massive social media data; it discusses theories and methodologies from different disciplines such as computer science, data mining, machine learning, social network analysis, network science, sociology, ethnography, statistics, optimization, and mathematics. It encompasses the tools to formally represent, measure, model, and mine meaningful patterns from large-scale social media data."[1]

4. OBJECTIVES

With this paper we tried to achieve below objectives

1. Gender based positive and negative tweets analysis for the launch week of iPhone7
2. US location based positive and negative tweets analysis for the launch week of iPhone7
3. Gender based positive and negative tweets analysis after the launch week of iPhone7
4. US location based positive and negative tweets analysis after the launch week of iPhone7
5. Top 5 US states based on gender for iPhone7 – during and after the launch week
6. Twitter trend after the launch of the iPhone 7

5. RELATED WORK

Apple's iPhone has been quite popular among people across the globe, especially in US. This topic has attracted many data scientists to perform sentiment analysis on iPhone. There has been work done on previous versions of iPhone such as comparing the sentiments about one version over the other. Performing the sentiment analysis using social media mining on product domain has been quite popular among data scientists. The paper by Nithish et al. (2013) [2] mainly talks about market reactions on smart phone domain using sentiment analysis. They focused on determining various factors which impacts popularity and rating of a smart phone in the market. They used the tweets along with NLP to classify positive or negative or neutral sentiments. In the paper by Shruti W. at The University of Akron,[3] they used Wekal data mining tool with positive and negative set of words and checked against the word set available from Twitter. They were more interested on the impact of emoticons.

Classifying sentiment using lexicon dictionary is also a popular way to doing data mining on the data gathered from social platforms such as twitter. Esuli and Sebastiani 2005) used supervised learning method to classify words into two distinct classifiers - Positive and Negative. They used an online dictionary (e.g. WordNet) to include the synonyms and antonyms of the words from training dataset into the seed sets of words. [4]

6. APPROACH

We used Python – a powerful programming language to get the historical data from Twitter. Python has many modules readily available which helps data scientists to perform social media mining and sentiment analysis by allowing to extract the historic data, perform data cleanup, process the data using natural language processing etc. We used supervised machine learning approach to get the sentiment related to iPhone7 between genders across US locations

6.1 List of Datasets and Features

We used below features to get the data for our analysis -

1. Twitter feed collected using Python API
2. Hashtag #iPhone7
3. Username
4. User's first name
5. Gender
6. User's US location
7. Tweet text

6.2 Feature Reduction

A tweet contains a lot of not so useful characters as a part of a tweet such as punctuation marks, URLs, mentions etc. Such characters do not help in sentiment analysis but only causes noise or garbage. We removed punctuation marks, URLs, mentions as a part of preprocessing the tweet text.

6.3 Python Modules

We used Python – a powerful programming language, to collect and analyze the twitter data. We used various python modules like got3, tweepy, geopy, gender_detector, pandas etc.

6.3.1 Got3

Python's GOT3 module is a perfect example of *web scraping* where someone can emulate visiting the website, downloading the HTML, and extracting specific results. The browser version of Twitter Search eliminates the limitation of Twitter Search API and allows searching for older tweets. The got3 module emulates visiting the browser search bar, "executes" a search per the parameters given, and scrapes the results. The results achieved with got3 can be even saved to a file with simple python modules.

Some of the useful criteria,

setQuerySearch – specify the search string(e.g. #iPhone7)
setSince – takes a string in the format ``YYYY-MM-DD`` and sets the start date for the search window
setUntil – takes a string in the format ``YYYY-MM-DD`` and sets the end date for the search window
setLang – takes the two-letter code of a language to restrict the language of the returned tweets

6.3.2 Tweepy

Tweepy is a powerful python module and supports accessing Twitter via OAuth Authentication. With tweepy, it's possible to get any object and use any method that the official Twitter API offers. For example, a User object has its documentation at https://dev.twitter.com/docs/platform-objects/users and following those guidelines, tweepy can get the appropriate information.

6.3.3 Geopy

geopy is a Python 2 and 3 client for several popular geocoding web services. geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

6.3.4 Gender detector

Gender detector is a Python library for guessing a person's gender by his/her first name. This library is based on [beauvoir](https://github.com/jeremybmerrill/beauvoir) with support for United States, United Kingdom, Argentina and Uruguay.

6.3.5 Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

6.4 Visualization with Tableau

Tableau is a visualization software that allows anyone to easily connect to data, visualize and create interactive, sharable dashboards. It's easy enough that any Excel user can learn it, but powerful enough to satisfy even the most complex analytical problems.

7. DATASETS AND METHOD

The historical dataset that we used to analyze the sentiments about iPhone7 in US locations was collected from Twitter.com. To get the data related to iPhone 7, we used hashtag '*#iPhone7'*. Apple's iPhone was launched on Sept 7, 2016. We used this date as a starting point for collecting data for our sentiment analysis for iPhone7. We collected two datasets with '#iPhone7'- one for the launch week of iPhone7 (Sept7, 2016 – Sept14, 2016), and second for the duration of 45 days after the launch week (Sept15, 2016-Oct 31, 2016). We observed that twitter users were very much active on *Twitter* talking about 'iPhone7' for the initial days of iPhone7 launch.

While Twitter Search API is extremely powerful to extract tweet information, it has a major limitation that it returns results from last 7 days when someone wants to get tweets based on hashtags. Our project completely depends on the historical data extracted from hashtags such as #iPhone7 and hence using Twitter Search API alone was not a solution for our problem. Hence we have decided to use python module called 'got3' that allows to execute tweets searches at any date range. The tweets retrieved from got3 gave us information such as id, username, text, date etc. However, for our project work we used only certain parameters as highlighted in below table.

Table 1. Features available with got3 data

Parameter name	Description	Used
id (str)	The integer representation of the unique identifier for this Tweet.	No
permalink (str)	URL of individual tweets	No
username (str)	Login name of the twitter	Yes

text (str)	The actual UTF-8 text of the status update	Yes
date (date)	Date of the tweet tweeted	Yes
retweets (int)	Number of times this Tweet has been retweeted.	No
favorites (int)	Indicates approximately how many times this Tweet has been "liked"	No
mentions (str)	User mentioned in the tweet	No
hashtags (str)	Hashtag mentioned in the particular tweets	No
Geo(str)	Location of the tweet tweeted	No (As most of the results are blanks. Hence we used a different approach described further in the report)

The data retrieved from got3 was alone again not sufficient for our sentiment analysis because we wanted to know the gender and the location of the user. To achieve this we combined *got3* with *tweepy* python module along with *gender_detector* and geopy.

7.1 Datasets

As mentioned above, we analyzed two datasets –

1. Launch week (7 Days)
2. Post Launch week (45 Days)

- **Launch week:** We collected the historical tweet data for the launch week of iPhone7 (Sept7, 2016- Sept14, 2016). We used below parameters to collect the data with got3

  tweetCriteria.setQuerySearch ("#iPhone7")

  tweetCriteria.setSince ("2016-09-07")

  tweetCriteria.setUntil ("2016-09-14")

  tweetCriteria.setLang ("en")

With the above mentioned criteria for the launch week, we got **1,20,885** tweets as a raw data for our sentiment analysis. However, even after using *setLang* criteria as 'en' , we did not get all the tweets in English language only. There were Spanish language tweets as well. For example, *"I love my iPhone7"* in English is equivalent of *"Me encanta mi iPhone7"* in Spanish. Here, both the messages uses same English alphabets. This is because setLang='en' (we believe) only considers English alphabets (a,b,c....z) but not the real language(English or Spanish). As we had to focus on English only data, we used "langdetect" python module. We then got **1,04,788** tweets of English language for the launch week. Our next goal was to determine gender and location. For gender, we used user's first name. There were some users for whom the "gender_guessor" python module could not predict the gender based on user's first name as the firstname had some weird values. We ignored such users and hence the records were again filtered to a smaller amount. Also, we focused only on tweets which can give us user's US location in state, country format (for example: North Carolina, United States of America) because gender and location are the main aspects of our analysis. We used "geopy" python module to get the user's location in 'State, Country' format.

With all these steps we finalized **10,825** tweets with

i) English language only
ii) Gender information available and
iii) US location information available.

Out of **10,825** records, we ignored **838** as noise: we termed these records with sentiments as "To Do" - these were the tweets which we could not confirm either as more positive or more negative (the number of positive and negative words were same in this case. We need to do more analysis of the exact emotion of the word to confirm a tweet is really positive or negative. This was beyond the scope of current project, hence we have ignored that parts. We have considered this task as a future work). This resulted in our final data of **9,987** tweets.

- **45 days post launch week :** We used below parameters to collect the data with got3 for 45 data post the launch week of iPhone7

  tweetCriteria.setQuerySearch ("#iPhone7")
  tweetCriteria.setSince ("2016-09-15")
  tweetCriteria.setUntil ("2016-10-31")
  tweetCriteria.setLang ("en")

With the above a got3 criteria and all the similar steps for launch week applied together, we finalized **15,408** tweets with

i) English language only

ii) Gender information available and

iii)US location information available.

Out of **15,408** records, we ignored **696** as noise: we termed these records with sentiments as "To Do" - these were the tweets which we could not confirm either as more positive or more negative. This resulted in our final data of **14,712** tweets.

Table 2. Final processed data count

Date range	Processed number of final tweets	Noise (Tweets classified as 'ToDo')	Number of tweets considered for sentiment analysis
Launch Week	10,825	838	9,987
45 days post launch week	15,408	696	14,712

7.2 Methods

With reference to the steps from section 7.1 above, we applied python's got3 module to get the English only data as described in Fig1. Our main focus was to get the tweets in English language as it is more convenient to perform data mining and sentiment analysis using lexicon dictionary in English language compared to other languages.

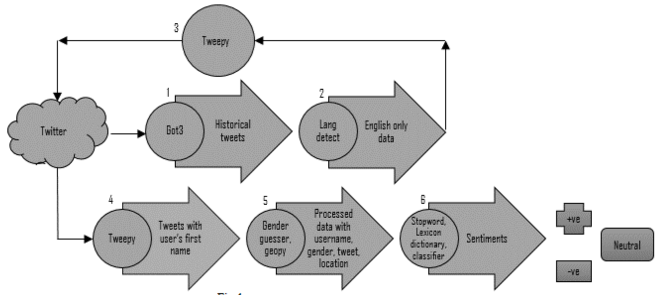

Fig 1

Got3's *setLang* criteria method is not guaranteed to return English only tweets as described in section 7.1 above. Hence, we applied *langdetect* python module (Step 2) on each tweet retrieved from got3 to ensure we only really get English only tweets before starting our sentiment analysis. The *langdetect* python module has over 99% precision for 53 languages such as English, Spanish etc. We also randomly sampled 500 tweets manually to ensure we actually get the English language tweets. For every English only tweet, we used userhandle retrieved from got3 along with *tweepy* python module (Step3 and 4) to get the name (first name, last name) of the user and the location for the user. We extracted only the first name of the user as the first name was the vital information needed for *gender_detector* python module (Step 5). The *gender_detector* module predicts the gender of the user based on the first name into categories such as 'male', 'mostly male', 'female', 'mostly female', 'unknown'. We considered the users under 'male' category for the users which gender_detector classified as 'mostly male'. Similarly, we considered the users under 'female' category for the users which gender_detector classified as 'mostly female'. We ignored results with 'unknown' gender. With the gender data identified for each user, we then used *geopy* python module (Step 5) to get users who belonged to US locations. We filtered and collected the data only for the users such that -

i) The tweet tweeted is in English
ii) The gender of the user is predicted
iii) The US location of the user is known

As a part of preprocessing, on the data collected with above steps, we performed operations such as tokenization/unigram, stopword removal, lower case conversion, removing URLs and hashtags from the tweet's text, remove mentions from tweet's text etc. (Step 6)

With the processed data in hand, we referred the lexicon dictionary by Bing Liu, Minqing Hu and Junsheng Cheng [6] and updated the dictionary files for our sentiment analysis, one representing the positive words such as happy, thrilled, awesome, excited etc. which conveys the positive sentiments of a particular tweet and the other dictionary representing negative words such as not happy, inconvenience, waste of money, bad, disappoint, miss etc. which conveys the negative sentiments of a particular tweet. The tweets which neither suggest positive nor negative sentiment, we tagged them as neutral tweets. We also created a separate classifier called 'ToDo' for the tweets which has equal number of positive and negative words as per our lexicon dictionary. A future work can be done on such tweets to classify the tweets into either positive or negative based on the intensity of each word used in a tweet such that a word suggests more positive or more negative sentiment. We created classifier to count number of positive and negative words in processed tweet text.

- If 'number of +ve words' > 'number of –ve words' , classified as positive sentiment
- If 'number of -ve words' > 'number of +ve words' , classified as negative sentiment
- If 'number of +ve and -ve words' = 0 , classified as neutral sentiment
- if 'number of +ve and -ve words are equal', classified as 'ToDo' (We considered such tweets as noise and ignored for our sentiment analysis)

If a positive word is preceded by a negative word such that the overall sentiments is negative, then our model classifies the sentiment as negative. For example, in the tweet 'I am not happy with removal of 3.5mm headphone jack', the word '*happy'* implying positive sentiment is preceded by the word '*not'* such that the emotion appears to be negative. Hence, our model treats such tweets as negative sentiments.

8. RESULTS

As mentioned in Table 2 above, we finalized a dataset of 9,987 tweets for launch week of iPhone7 and 14,712 tweets for the next 45 days after the launch week of iPhone7. We used supervised learning to train our model on a test data of 200 tweets from the launch week. "In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way."[5]

We read and analyzed each and every tweet from the training and test dataset and manually classified each tweet into 3 categories – Positive, Negative and Neutral sentiments. To achieve higher percentage of accuracy, we first divided the test dataset of 200 tweets into 2 parts – each containing 100 tweets. We later individually analyzed each tweet from the subset of dataset and then also cross validated the classified sentiments. For some of the records from each individual dataset, we found deviation in sentiments classified. We gathered such tweets and discussed together to confirm a sentiment into one final sentiment based on human intelligence. We used the manually classified dataset as an input to develop and train our model. We achieved **84.43%** accuracy with our model.

As per the results achieved with our model, we observed that most of the tweets had neutral sentiment. This was expected in the initial days of launch week because people were talking more about general features, questions and posting images etc. rather than positive or negative reviews about iPhone7. The next major sentiment found was positive. Before starting our sentiment analysis project, we thought that people might not like changes in the device model which Apple has introduced with iPhone7 such as removal of 3.5 mm headphone jack, antenna bar, introduction of wireless headphones etc. However, with our data mining and sentiment analysis we found that people accepted these new features in a positive way (Fig 2).

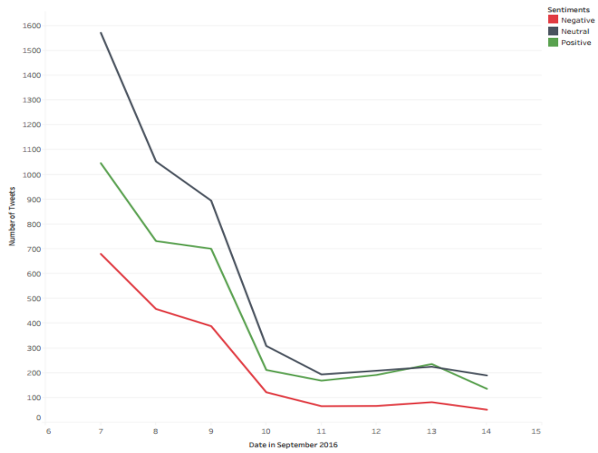

Fig 2: Sentiment trend for launch week

Similar results were seen for the next 45 days of launch week (Fig 3). Here also, sentiments were mostly neutral and then followed by positive. The number negative sentiments was less.

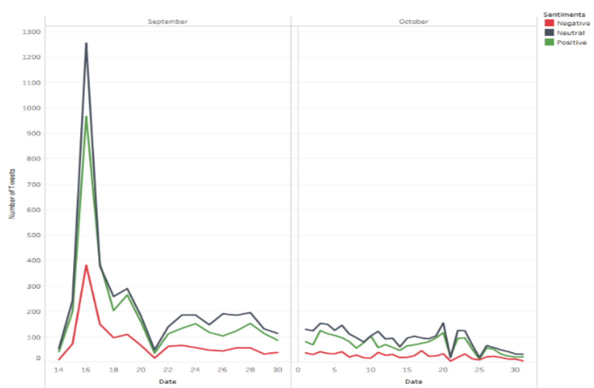

Fig 3: Sentiment trend for 45 days post launch week

Out of **9,987** tweets (after excluding 838 ToDo tweets, Table 2) analyzed for launch week of iPhone7, **34.38%** tweets (**3,424**) were Positive, **19.18%** tweets (**1,916**) were Negative and remaining **46.53%** (**4,647**) were of Neutral sentiments. Out of **9,987** tweets for the launch week, **74.15%** tweets (**7,406**) were posted by men and remaining **25.84%** (**2,581**) were posted by women. We concluded that men were more interested in talking about iPhone7 on twitter. For men, we got **2,579** (**34.82%**) positive sentiments and **1,392** (**18.79%**) negative sentiments. **3,435** tweets represented neutral sentiments. While for females, **845** (**32.73%**) were positive sentiments, **524** (**20.30%**) were negative and **1212** were neutral sentiments (Fig. 4). It appears irrespective of gender, people were more positive towards iPhone7 during the launch week (Fig. 4).

For the duration of 45 days after the launch week of iPhone7, out of **14,712** tweets (after excluding 696 ToDo tweets, Table 2) analyzed for launch week of iPhone7, **37.35%** tweets (**5,496**) were Positive, **14.62%** tweets (**2,151**) were Negative and remaining **48.02%** (**7,065**) were of Neutral sentiments. Out of **14,712** tweets for the launch week, **77.63%** tweets (**11,422**) were posted by men and remaining **22.37%** (**3,290**) were posted by women. We concluded that men were more interested in talking about iPhone7 on twitter. For men, we got **4,108** (**35.96%**) positive sentiments and **1,667** (**14.60%**) negative sentiments. **5,647** tweets represented neutral sentiments. While for females, **1388** (**42.18%**) were positive sentiments, **484** (**14.71%**) were negative and **1418** were neutral sentiments (Fig.

5).

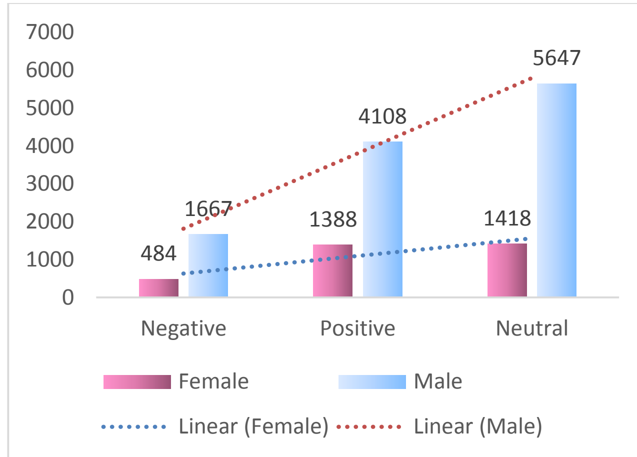

Fig 4: Gender based sentiments for launch week

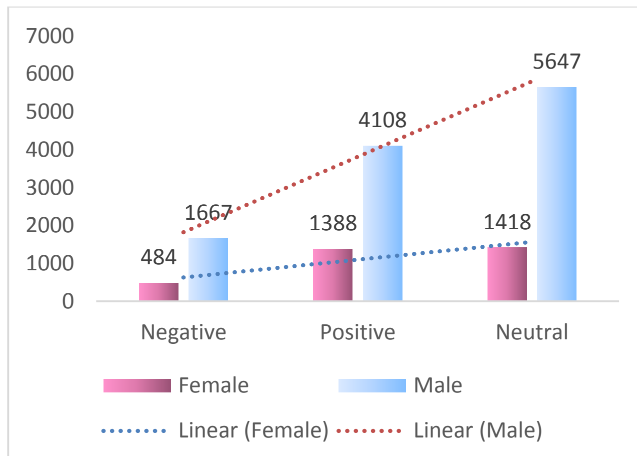

Fig 5: Gender based sentiments for post launch week

We also focused to determining the top 5 US states where users were most proactive talking about iPhone7. For the launch week of iPhone 7, the top 5 US states are New York, California, Texas, Florida and Ohio (Fig.6, Table 3)

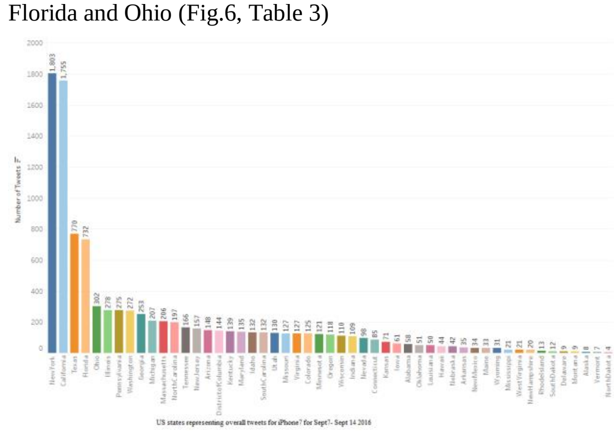

Fig 6: US states representing overall tweets for iPhone7 for Sept7-Sept14 2016

Table 3: Classified sentiments for launch week across locations between genders

	State	Total Sentiments	Total number of positive sentiments	Number of males with positive sentiments	Number of females with positive sentiments	Total number of negative sentiments	Number of males with negative sentiments	Number of females with negative sentiments	Total number of neutral sentiments	Number of males with neutral sentiments	Number of females with neutral sentiments
0	NewYork	1803	536	462	74	241	171	70	1026	816	210
1	California	1755	599	459	140	335	266	69	821	606	215
2	Texas	770	279	221	58	151	119	32	340	251	89
3	Florida	732	244	164	80	172	84	88	316	182	134
4	Ohio	302	109	76	33	57	44	13	136	106	30
5	Illinois	278	96	79	17	63	47	16	119	95	24
6	Pennsylvania	275	103	80	23	56	38	18	116	89	27
7	Washington	272	99	76	23	58	42	16	115	83	32
8	Georgia	253	90	62	28	43	37	6	120	88	32
9	Michigan	207	78	50	28	36	27	9	93	81	12

For the duration of post launch week, the top 5 US states were New York, California, Texas, Florida and Illinois. (Fig.7, Table 4)

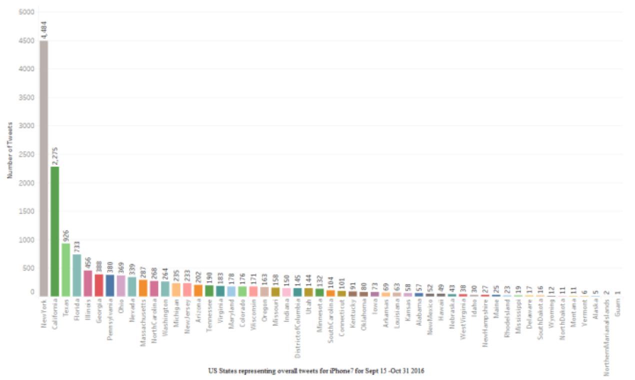

Fig 7: US states representing overall tweets for iPhone7 for Sept7-Sept14 2016

Table 4: Classified sentiments post launch week across locations between genders

	State	Total Sentiments	Total number of positive sentiments	Number of males with positive sentiments	Number of females with positive sentiments	Total number of negative sentiments	Number of males with negative sentiments	Number of females with negative sentiments	Total number of neutral sentiments	Number of males with neutral sentiments	Numbe females with neutral sentime
0	NewYork	4484	1604	1445	159	488	429	59	2392	2197	195
1	California	2275	920	593	327	362	274	88	993	748	245
2	Texas	926	348	246	102	141	96	45	437	320	117
3	Florida	733	255	173	82	110	84	26	368	272	96
4	Illinois	456	173	137	36	87	57	30	196	161	35
5	Georgia	388	103	57	46	57	41	16	228	106	122
6	Pennsylvania	380	137	107	30	54	42	12	189	152	37
7	Ohio	369	148	90	58	56	39	17	165	121	44
8	Nevada	339	94	85	9	48	48	0	197	182	15
9	Massachusetts	287	109	77	32	41	35	6	137	98	39
10	NorthCarolina	268	107	64	43	41	35	6	120	90	30
11	Washington	264	115	85	30	44	30	14	105	71	34
12	Michigan	235	89	46	43	46	35	11	100	59	41

We also determined the distribution of sentiments between genders across US locations.

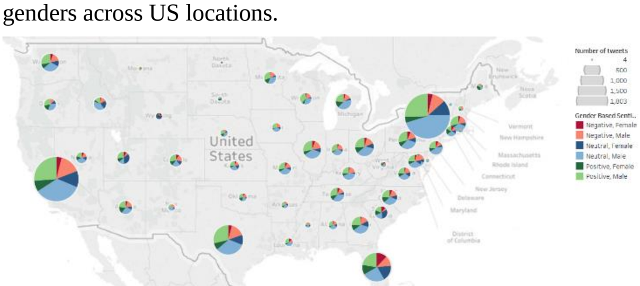

Fig 8: Proportional symbol map demonstrating sentiments across US states for launch week

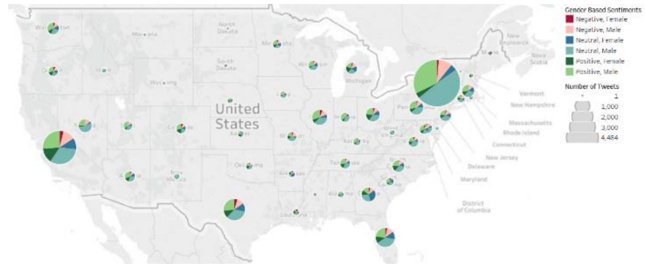

Fig 9: Proportional symbol map demonstrating sentiments across US states after the launch week

9. CONCLUSION AND FUTURE WORK

With all the data gathered, processed and analyzed with various python modules, we could see a pattern that during and after the launch week of iPhone7, twitter users were inclined towards neutral and positive reactions. At the start of our project proposal we had thought that users might not accept new features introduced by Apple, Inc. like removal of 3.5 mm headphone jack. However, we were surprised with the data and the results we got after sentiment analysis which clearly indicated positive reactions. We also observed that men were more active in expressing their views in the form of tweets as compared to women. We had also thought that major metropolitan states like New York, California would be among the top 5 US states where twitter users would be most active. We got our results in line with our assumption for these states.

With the results achieved so far we believe this project has met our objectives however, we only used #iPhone7 to do the analysis. We believe #iPhone7Plus, #iPhoneJetBlack etc. should also be worked on in future. This project has more potential to explore the granularity of the sentiments. For example, in our analysis we encountered a certain number of tweets (although small in number) for which our existing model could not clearly differentiate into one of the three classifiers – Positive, Negative and Neutral. Our model uses the count of words for positive and negative reactions in a tweet to classify the emotion but currently it does not handle a situation where in the number of positive words and negative words are equal in number. We consider determining the real intensity of emotions behind a word to classify a sentiment as more positive or more negative can be implemented as a future work. For example, a tweet that says *"The new IPhone looks awesome. , going to miss the old headphone jack tho #iPhone7"* expresses more positive sentiment when looked with human intelligence. However, as per our model this tweet contains one positive word *'awesome'* and one negative word *'miss'* and hence our model classifies the tweet as 'ToDo'. We have ignored the tweets which resulted in 'ToDo' sentiment classifier as noise. We believe more research can be done to achieve a higher accuracy of sentiment analysis on our model such that the model determines if a word is expressing more positive sentiment or more negative sentiment. A tweet usually contains emoticons such as ☺, ☹ signifying positive and negative sentiments. Our model does not consider emotions for sentiment analysis. This can be another enhancement to model in future. Because of 140 characters limit, a twitter user uses a lot of slangs in the tweets and also the words are often misspelled and contains grammatical errors. The model can be made more robust to handle acronym usage like asap, lol etc. These areas can also be improved with future iterations of our model. Overall, we achieved the objectives we set at the start of our work for this paper.

10. ACKNOWLEDGMENT

We thank our Professor Mr. Vincent Malik, Indiana University for guiding us to complete this project and helping us to learn Python and various methods and aspects for social media mining

11. REFERENCES

[1] Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, Social Media Mining (April 20, 2014), P2, Page 16

[2] Nithish, R., Sabarish, S., Abirami, A.M., Askarunisa, A., Navaneeth Kishen, M. (2013). An Ontology based Sentiment Analysis for mobile products using tweets. In Fifth International Conference on Advanced Computing (p 242–247).

[3] Text Mining for Sentiment Analysis of Twitter Data - Wakade, Shruti; Shekar, Chandra; Liszka, Kathy J; Chan, Chien-Chung.(2012) Proceedings of the International Conference on Information and Knowledge Engineering (IKE): 1-6. Athens: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). (2012)

[4] Esuli, Andrea and Fabrizio Sebastiani. Determining term subjectivity and term orientation for opinion mining in Proceedings of Conf. of the European Chapter of the Association for Computational Linguistics (EACL-2006)2006.

[5] https://en.wikipedia.org/wiki/Supervised_learning

[6] Bing Liu, Minqing Hu and Junsheng Cheng. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan.

[7] Mastering Social Media Mining with Python (July 2016) - Marco Bonzanini - https://marcobonzanini.com/2015/05/17/mining-twitter-data-with-python-part-6-sentiment-analysis-basics/

[8] Twitter API documents : https://dev.twitter.com/docs