**Project Name:** Twitter reactions on iPhone7 using Sentiment Analysis and Social Media Mining

**Group Number:** SMM Group3

## Team Member:
- Sunil Agrawal (suniagra@umail.iu.edu)
- Pawan Pinjarkar (ppinjark@umail.iu.edu)

## Deliverables:
1) Code for text processing and sentiment analysis - SMMGroup3_TextProcessing_Iphone7.ipynb
2) Lexicon dictionary files
   a) positive-words.txt
   b) negative-words.txt
3) Processed data file for iPhone7 Sentiment analysis
   a) LaunchWeek_Sep7-Sept14_Iphone7.pkl
   b) PostLaunchWeek_Sep15-Oct31_Iphone7.pkl
4) Result files
   a) Result_SMMGroup3_TextProcessing_Iphone_LaunchWeek_Sep07-Sep14.html
   b) Result_SMMGroup3_TextProcessing_Iphone_PostLaunchWeek_Sep15-Oct31.html

## Pre-requisites:
1) Install Python 3.5.2
2) Install Anaconda 4.1.1
3) Import stopwords from nltk. From python cosnole ,execute below commands –
   a) import nltk
   b) nltk.download("stopwords")

## Steps to execute code from jupyter notebook:

1) Make sure all the prerequisites are completed
2) Download and extract zip file SMMGroup3_TextProcessing_Iphone7.zip
3) Go to the folder where the file is extracted on your system
4) Execute *jupyter notebook* command
5) Go to the address http://localhost:8888/tree/ on your browser
6) Execute notebook file SMMGroup3_TextProcessing_Iphone7.ipynb
7) The program will ask for the pickle file with below message. Select the pickle file for launch week or post launch week data
     *Please enter pickle file name with processed dataset for sentiment analysis*

# Results:

**You should get below result for launch week data file.**

```
Please enter pickle file name with processed dataset for sentiment analysis =LaunchWeek_Sep7-Sept14_Iphone7.pkl

The entered dataset filename  is   =   LaunchWeek_Sep7-Sept14_Iphone7.pkl
Reading and processing the dataset .............................................

Total number of Sentimentts =  10825
Total number of Positive Sentiments = 3424
Total number of Negative Sentiments =  1916
Total number of Neutral Sentiments =  4647
Total number of ToDo Sentiments(Noise) which is ignored =  838
................................................................................

Saving the pickle file  :  finalLaunchWeek_Sep7-Sept14_Iphone7.pkl
Saving the csv file :  finalLaunchWeek_Sep7-Sept14_Iphone7.pkl.csv
................................................................................

Final Number of Tweets Considered for Sentiment Analsyis(Exclused ToDo(Noise) classified Tweets) = 9987

Gender and location based sentiments(Positive, Negative, and Neutral) for input file  LaunchWeek_Sep7-Sept14_Iphone7.pkl
................................................................................
```

| | 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 |

**And, for post launch week data the result should be as below –**

```
Please enter pickle file name with processed dataset for sentiment analysis =PostLaunchWeek_Sep15-Oct31_Iphone7.pkl

The entered dataset filename  is   =   PostLaunchWeek_Sep15-Oct31_Iphone7.pkl
Reading and processing the dataset ...........................................

Total number of Sentimentts =  15408
Total number of Positive Sentiments = 5496
Total number of Negative Sentiments =  2151
Total number of Neutral Sentiments =  7065
Total number of ToDo Sentiments(Noise) which is ignored =  696
..........................................................................

Saving the pickle file  :  finalPostLaunchWeek_Sep15-Oct31_Iphone7.pkl
Saving the csv file :  finalPostLaunchWeek_Sep15-Oct31_Iphone7.pkl.csv
..........................................................................

Final Number of Tweets Considered for Sentiment Analsyis(Exclused ToDo(Noise) classified Tweets) = 14712

Gender and location based sentiments(Positive, Negative, and Neutral) for input file  PostLaunchWeek_Sep15-Oct31_Iphone7.pkl
..........................................................................
```

| | 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 | 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 |