Pokemon Go Analytics

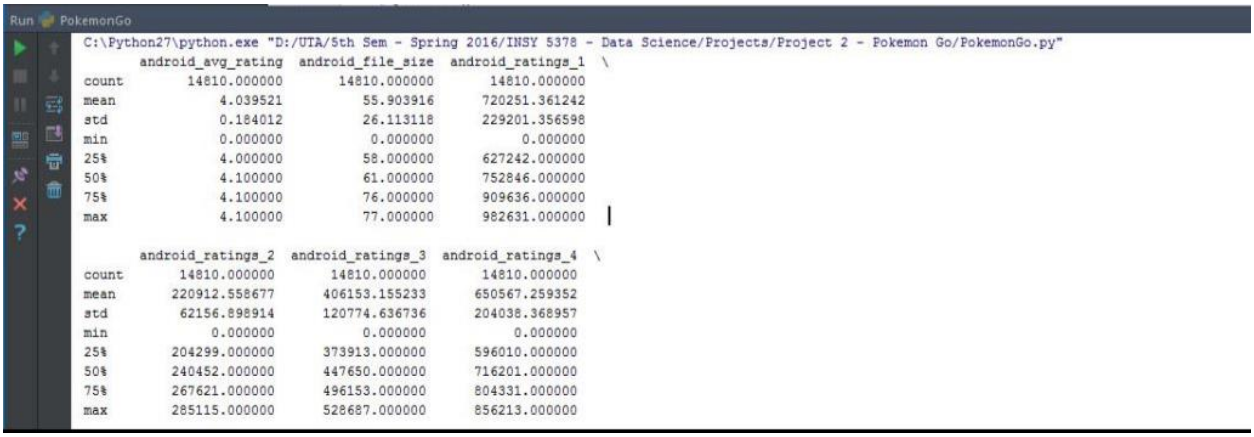
Team members

1. Sagar Chhadia (Computer Science) sagarmahesh.chhadia@mavs.uta.edu
2. Nikita Dhawle (Information Systems) nikita.dhawle@mavs.uta.edu
3. Jasleen Kaur Sandhu (Computer Science) jasleenkaur.sandhu@mavs.uta.edu

High Level Description:

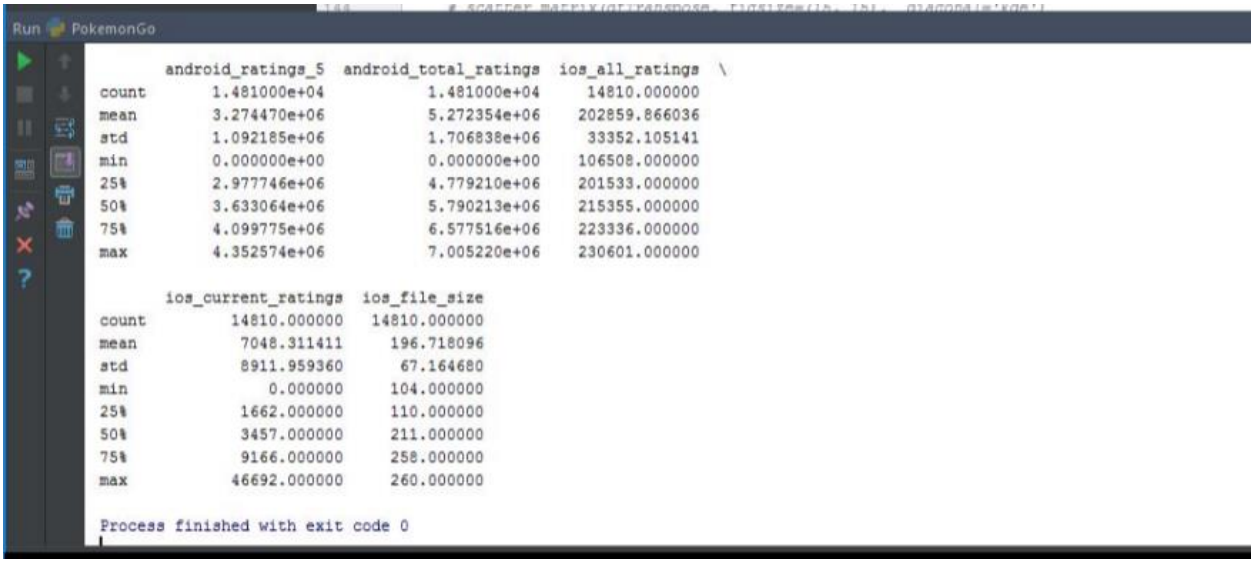
Step 1: We used beautifulSoup to scrape the data from the HTML files. For the android HTML files we extracted android_ratings_1, android_ratings_4, android_ratings_3, android_ratings_5, android_file_size, android_ratings_2, android_avg_rating, android_total_ratings. For the IOS HTML files we scraped ios_all_ratings, ios_current_ratings, ios_file_size. After extracting the data separately for the android file and IOS file we merged the data in a single json files.

Step 2: We created a pandas dataframe using the json file that we created in the step above and calculated the count, mean, standard deviation, minimum, maximum, 25th percentile, 50th percentile and 75th percentile.



```
Run PokemonGo
C:\Python27\python.exe "D:/UTA/5th Sem - Spring 2016/INSY 5378 - Data Science/Projects/Project 2 - Pokemon Go/PokemonGo.py"
       android_avg_rating  android_file_size  android_ratings_1  \
count        14810.000000       14810.000000       14810.000000
mean             4.039521          55.903916      720251.361242
std              0.184012          26.113118      229201.356598
min              0.000000           0.000000           0.000000
25%              4.000000          58.000000      627242.000000
50%              4.100000          61.000000      752846.000000
75%              4.100000          76.000000      909636.000000
max              4.100000          77.000000      982631.000000

       android_ratings_2  android_ratings_3  android_ratings_4  \
count       14810.000000       14810.000000       14810.000000
mean       220912.558677      406153.155233      650567.259352
std         62156.898914      120774.636736      204038.368957
min             0.000000           0.000000           0.000000
25%        204299.000000      373913.000000      596010.000000
50%        240452.000000      447650.000000      716201.000000
75%        267621.000000      496153.000000      804331.000000
max        285115.000000      528687.000000      856213.000000
```



Step 3: We have used pairplot function from seaborn module to create a scatter matrix.

Step 4: For **Correlation Coefficient** we have taken two for loops and taken the unique combinations of all the 11 values present in the rows and columns. For instance if a and b are the only values present then we have taken a to b only once and not considered b to a, a to a and b to b. After taking these unique pairs we have taken only those values whose correlation coefficient is greater than 0.5 or less than -0.5.

Following is our output for the correlation coefficients:

```
android_ratings_1 android_ratings_2
0.994207916269
android_ratings_1 android_ratings_3
```

0.99252151211
android_ratings_1 android_ratings_4
0.993169370519
android_ratings_1 android_ratings_5
0.992937900987
android_ratings_1 android_total_ratings
0.994814032134
android_ratings_1 ios_all_ratings
0.945816955364
android_ratings_1 ios_current_ratings
-0.628205109585
android_ratings_1 ios_file_size
0.869415658223
android_ratings_2 android_ratings_3
0.999474344331
android_ratings_2 android_ratings_4
0.999300877288
android_ratings_2 android_ratings_5
0.999189724188
android_ratings_2 android_total_ratings
0.999472588747
android_ratings_2 ios_all_ratings
0.961465804393
android_ratings_2 ios_current_ratings
-0.628570225768
android_ratings_2 ios_file_size
0.841085175428
android_ratings_3 android_ratings_4
0.999865730963
android_ratings_3 android_ratings_5
0.999479923195
android_ratings_3 android_total_ratings
0.999517402299
android_ratings_3 ios_all_ratings
0.957792549528
android_ratings_3 ios_current_ratings
-0.641592531078
android_ratings_3 ios_file_size
0.84522313658
android_ratings_4 android_ratings_5
0.999657420105
android_ratings_4 android_total_ratings
0.999718210515
android_ratings_4 ios_all_ratings
0.957871435825
android_ratings_4 ios_current_ratings
-0.646088682503
android_ratings_4 ios_file_size
0.848002730235
android_ratings_5 android_total_ratings
0.999834040699
android_ratings_5 ios_all_ratings
0.960330002947
android_ratings_5 ios_current_ratings

```
-0.637181024295
android_ratings_5 ios_file_size
0.846713599239
android_total_ratings ios_all_ratings
0.958803379599
android_total_ratings ios_current_ratings
-0.637605973351
android_total_ratings ios_file_size
0.850359062122
ios_all_ratings ios_current_ratings
-0.511092785895
ios_all_ratings ios_file_size
0.7383534015
ios_current_ratings ios_file_size
-0.638065714728
```

Step 5: For the prediction we have taken time on x- axis , and the value to be predicted on y-axis i.e on y-axis we have android-total-rating for one function and for other function we have ios-all-rating. Then for cross validation we divided our data set into train and test sets and used 80/20 model for prediction of values.

Following are our predicted values for android and ios on 2016/11/01 11:50 PM:

```
android_total_ratings model score: 0.547827308225
Predicted value of android_total_ratings for 2016/11/01 11:50 PM is : [
6083266.42555237]


ios_all_ratings model score: 0.349139557241
Predicted value of ios_all_ratings for 2016/11/01 11:50 PM is : [
215525.90612221]
```

Step 6: Deep Learning

In this step, we scraped all the files and got unique image links from the android and IOS files. Using the links, we downloaded the images and then using tensor flow on each image we extracted the tags with the corresponding probabilities. The images are attached as zip file under the folder name PokemonGoScreenshots

Unique Screenshots for android: 5

Unique Screenshots for IOS: 17

Following are the outputs for the probabilities:

D:\PGScreenShots\android_screenshot_0.jpg
web site, website, internet site, site 0.5976784229278564
sunglasses, dark glasses, shades 0.0469086691737175
electric fan, blower 0.03147434443235397

sunglass 0.021552791818976402
comic book 0.02090362459421158

D:\PGScreenShots\android_screenshot_1.jpg
web site, website, internet site, site 0.6958844661712646
television, television system 0.02656516060233116
monitor 0.01862839236855507
pool table, billiard table, snooker table 0.01564372144639492
screen, CRT screen 0.015389925800263882

D:\PGScreenShots\android_screenshot_2.jpg
lawn mower, mower 0.3052093982696533
golf ball 0.06548774987459183
bow 0.03529775142669678
croquet ball 0.02763860858976841
barrow, garden cart, lawn cart, wheelbarrow 0.0266158115118742

D:\PGScreenShots\android_screenshot_4.jpg
web site, website, internet site, site 0.5401659607887268
monitor 0.0543154813349247
notebook, notebook computer 0.0384797565639019
television, television system 0.03006863221526146
maillot 0.011101976968348026

D:\PGScreenShots\android_screenshot_5.jpg
ant, emmet, pismire 0.1309080868959427
monitor 0.08762867003679276
aircraft carrier, carrier, flattop, attack aircraft carrier 0.07103414088487625
wing 0.05027107894420624
web site, website, internet site, site 0.042926445603370667

D:\PGScreenShots\ios_screenshot_0.jpg
web site, website, internet site, site 0.883571207523346
menu 0.008027322590351105
slot, one-armed bandit 0.004043694585561752
washer, automatic washer, washing machine 0.003706002375110984

hand-held computer, hand-held microcomputer 0.002964386250823736

D:\PGScreenShots\ios_screenshot_1.jpg
comic book 0.19361448287963867
maze, labyrinth 0.1932980716228485
web site, website, internet site, site 0.05235723778605461
monitor 0.029567185789346695
book jacket, dust cover, dust jacket, dust wrapper 0.027674004435539246

D:\PGScreenShots\ios_screenshot_2.jpg
ashcan, trash can, garbage can, wastebin, ash bin, ash-bin, ashbin, dustbin, trash barrel, trash bin 0.15497566759586334
joystick 0.06404933333396912
cannon 0.03585103154182434
maraca 0.02726832963526249
pedestal, plinth, footstall 0.027154820039868355

D:\PGScreenShots\ios_screenshot_3.jpg
web site, website, internet site, site 0.2275332808494568
envelope 0.09162567555904388
Band Aid 0.03712098300457001
pinwheel 0.029456494376063347
airship, dirigible 0.024857865646481514

D:\PGScreenShots\ios_screenshot_4.jpg
laptop, laptop computer 0.49859192967414856
web site, website, internet site, site 0.10645917803049088
monitor 0.06384018808603287
screen, CRT screen 0.029848331585526466
notebook, notebook computer 0.02801426686346531

D:\PGScreenShots\ios_screenshot_5.jpg
web site, website, internet site, site 0.3661857843399048
safety pin 0.020037969574332237
sunglasses, dark glasses, shades 0.0167746189981699
toilet seat 0.015619936399161816
washer, automatic washer, washing machine 0.014380029402673244

D:\PGScreenShots\ios_screenshot_6.jpg
web site, website, internet site, site 0.8907701373100281
menu 0.0036376425996422768
monitor 0.0018526039784774184
screen, CRT screen 0.001841831486672163
analog clock 0.0017735683359205723

D:\PGScreenShots\ios_screenshot_7.jpg
web site, website, internet site, site 0.1163666844367981
laptop, laptop computer 0.0807962566614151
notebook, notebook computer 0.05348580703139305
joystick 0.04790791869163513
monitor 0.04169079661369324

D:\PGScreenShots\ios_screenshot_8.jpg
web site, website, internet site, site 0.42240995168685913
comic book 0.03247756138443947
carousel, carrousel, merry-go-round, roundabout, whirligig 0.02088976837694645
fountain 0.01781134307384491
safety pin 0.014400497078895569

D:\PGScreenShots\ios_screenshot_9.jpg
web site, website, internet site, site 0.6088576316833496
television, television system 0.056650057435035706
monitor 0.019958283752202988
notebook, notebook computer 0.016072208061814308
iPod 0.01179817970842123

D:\PGScreenShots\ios_screenshot_10.jpg
space shuttle 0.2304239720106125
joystick 0.059921521693468094
racer, race car, racing car 0.05625808611512184
scoreboard 0.04957202821969986
airliner 0.04575682431459427

D:\PGScreenShots\ios_screenshot_11.jpg
web site, website, internet site, site 0.12342077493667603
maze, labyrinth 0.07148678600788116
comic book 0.04789261892437935
joystick 0.04420957341790199
television, television system 0.037576720118522644

D:\PGScreenShots\ios_screenshot_12.jpg
fountain 0.20302647352218628
carousel, carrousel, merry-go-round, roundabout, whirligig 0.08313611894845963
comic book 0.05170505866408348
toyshop 0.0334254615008831
monitor 0.03227037563920021

D:\PGScreenShots\ios_screenshot_13.jpg
web site, website, internet site, site 0.9409151673316956
analog clock 0.0036712565924972296
envelope 0.002909436821937561
monitor 0.0022511144634336233
screen, CRT screen 0.0021692963782697916

D:\PGScreenShots\ios_screenshot_14.jpg
web site, website, internet site, site 0.36779189109802246
envelope 0.16913513839244843
binder, ring-binder 0.05812036246061325
tray 0.017636902630329132
monitor 0.017210371792316437

D:\PGScreenShots\ios_screenshot_15.jpg
web site, website, internet site, site 0.5862426161766052
monitor 0.07197427749633789
television, television system 0.05955268442630768
comic book 0.04756322130560875
teapot 0.014249833300709724

D:\PGScreenShots\ios_screenshot_16.jpg
aircraft carrier, carrier, flattop, attack aircraft carrier 0.09968294948339462
pole 0.03657454252243042
wing 0.02655319683253765
lakeside, lakeshore 0.024369418621063232
magnetic compass 0.023960646241903305