PROJECT-2

POKEMON GO!! ANALYTICS

Group -20

1. Gaurav Kolekar
2. Pooja Lahoti
3. Venkatesh
4. Sai Aishwarya

**INTRODUCTION**

**Background:** Pokémon Go! became a very famous augmented reality (AR) game in 2016 summer. In this project, we want to understand the success of the mobile app game by predicting different models.

**Project Overview:** We have extracted data from HTML files provided and formed a data frame. Using the data frame , created a scatter matrix and found correlation between variables. We then used different regression Algorithms to predict the future values.

**WORKFLOW**

**Data Extraction:** From the HTML files provided , we created a function to read the files from folders using os.path.join and os.path.isdir(). Used Beautiful soup library to extract data by web scraping. We identified the unique html tags to scrape IOS and Android variables . Used try and except to handle the improper files and missing values. We store the data in CSV(read\_csv) ,JSON(to\_json) and Excel(to\_excel) format.

**Data Frame:** We used pandas library to create the data frame using data time as index.

**Handling missing values:** As we had many blank values , we used front fill (ffill) method which fills the previous value to all the blank spaces.

**Data Exploration:** We used describe() function to get the different parameter values for each of the 11 variables . Used scatter matrix() to find the variables with any positive or negative correlation. We used numpy module to calculate the Pearson’s correlation between the correlated pairs found from scatter matrix. Using matplotlib, we created a time series plot.

**Prediction Model:**

The models which we used to predict are :

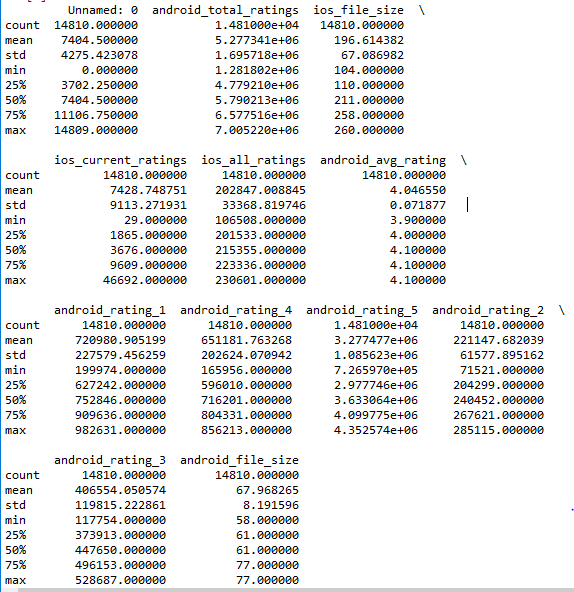
* Linear Regression
* Lasso
* Ridge
* Random Forest

**Tensor Flow:** We again used web scraping technique and used if-else condition to extract the unique screen shots from both Android and HTML files. Then we used tensor floe to get the tags from all the extracted screen shots.

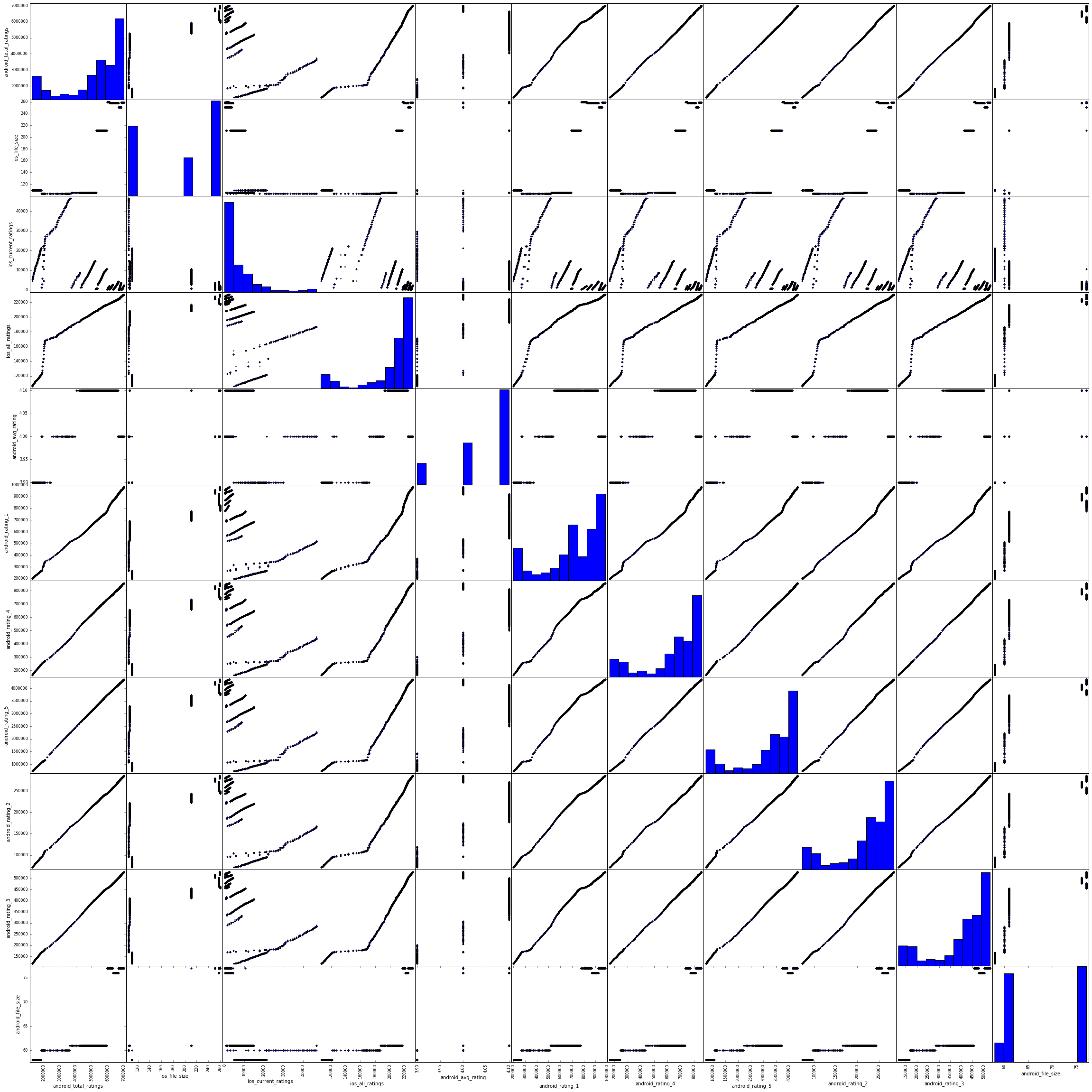
**RESULTS**

**Describe Function:**

s.describe()

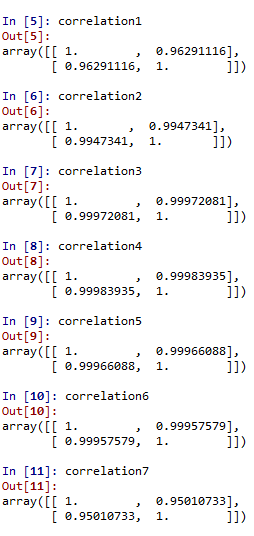


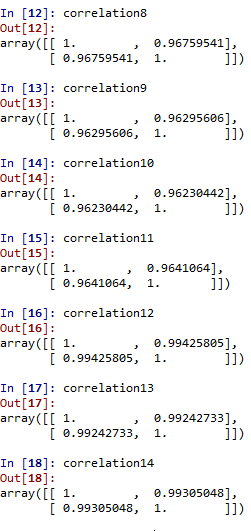
Scatter Matrix:

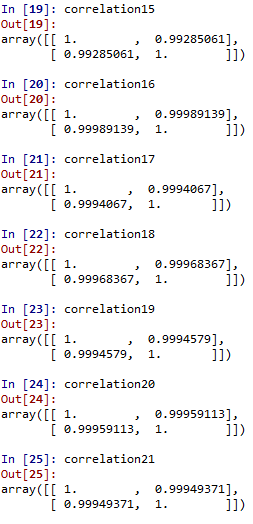


Pearson Coefficient Correlation:



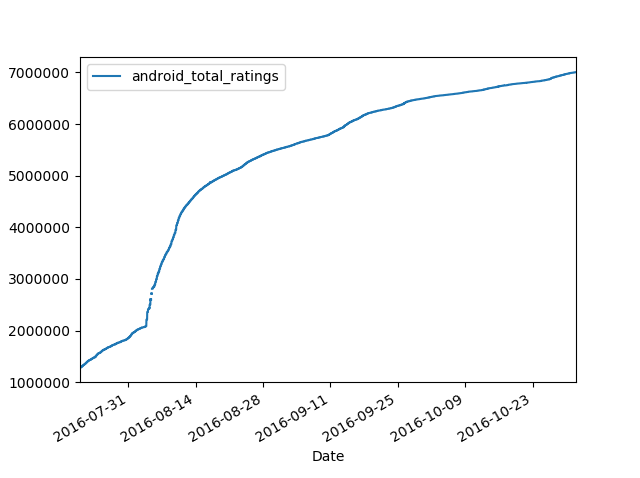


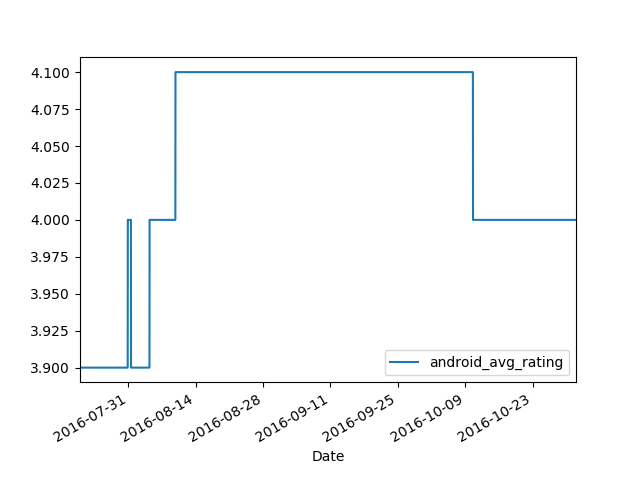


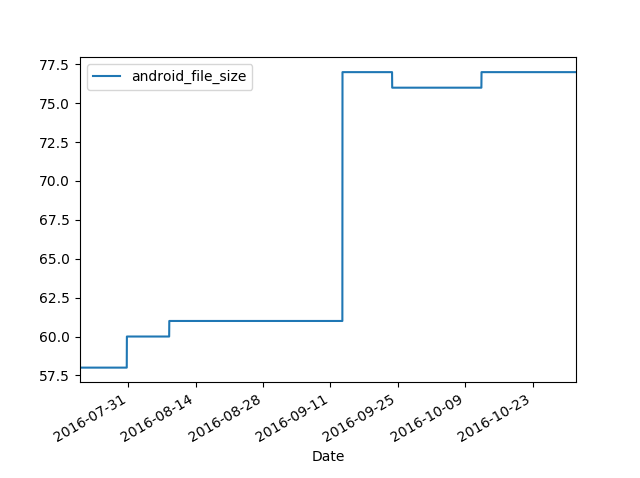


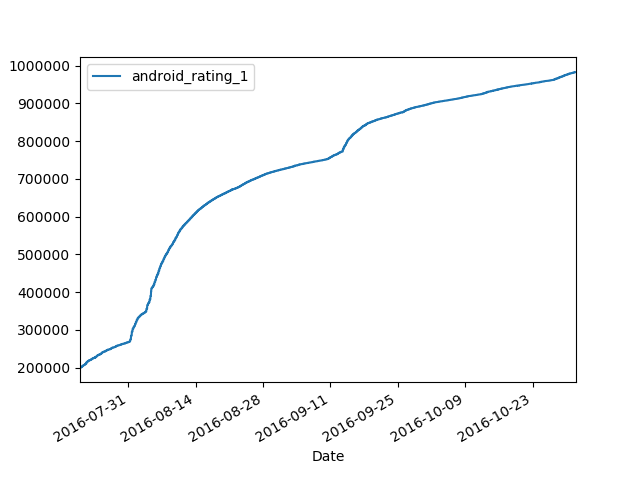
Time Series:

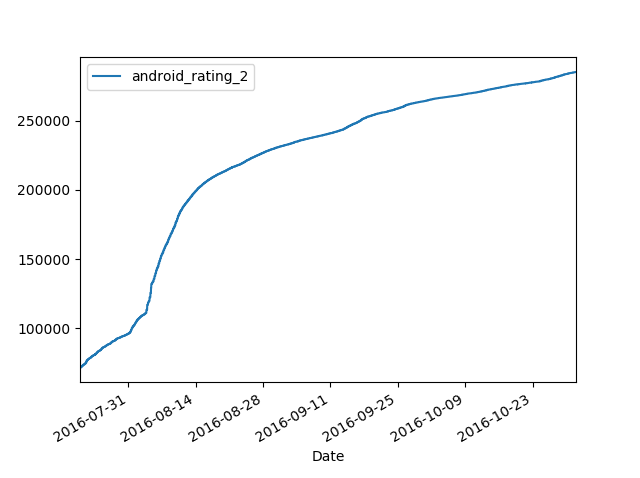
Below are time series graph for each of the 11 variables

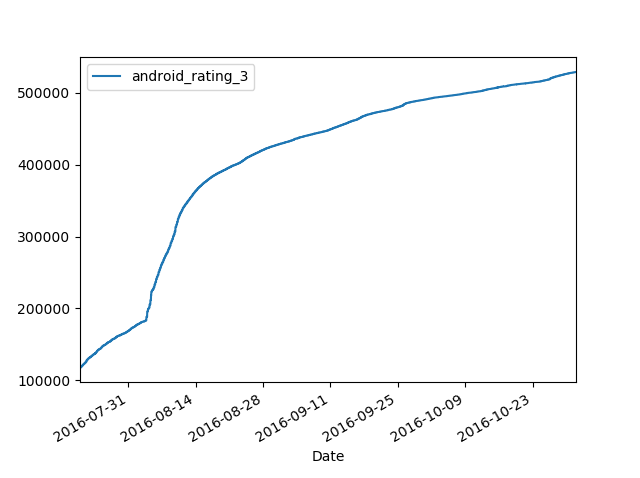


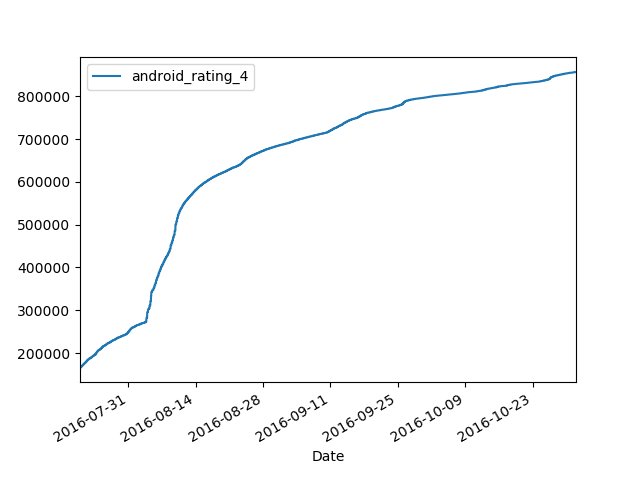


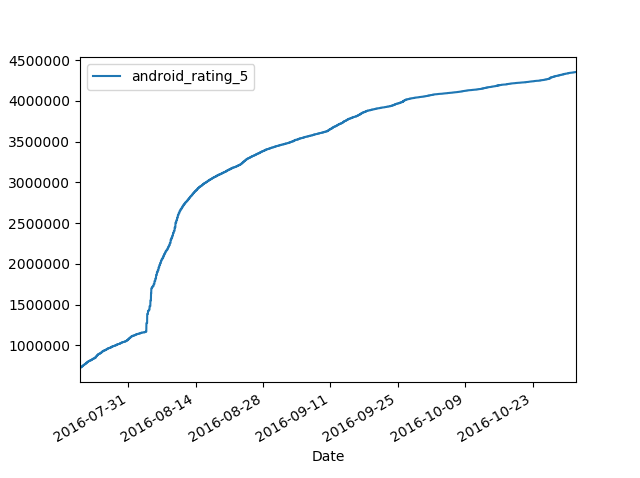


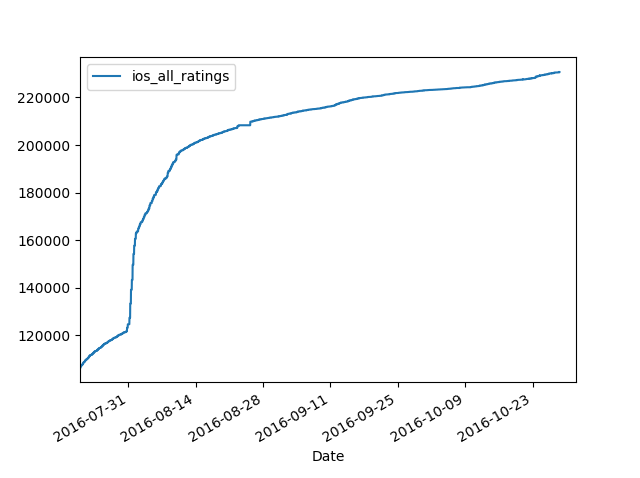


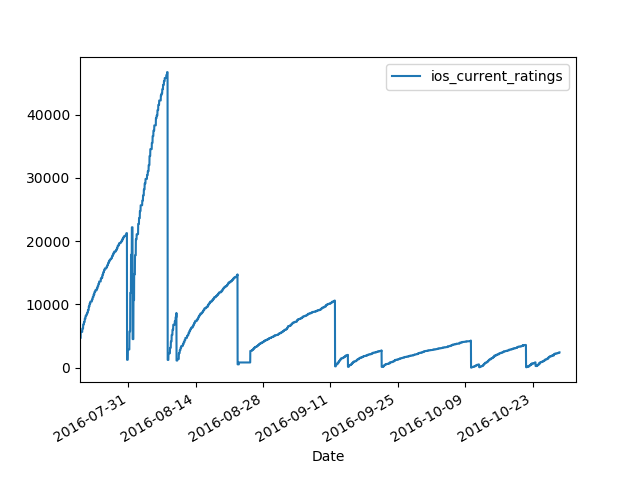


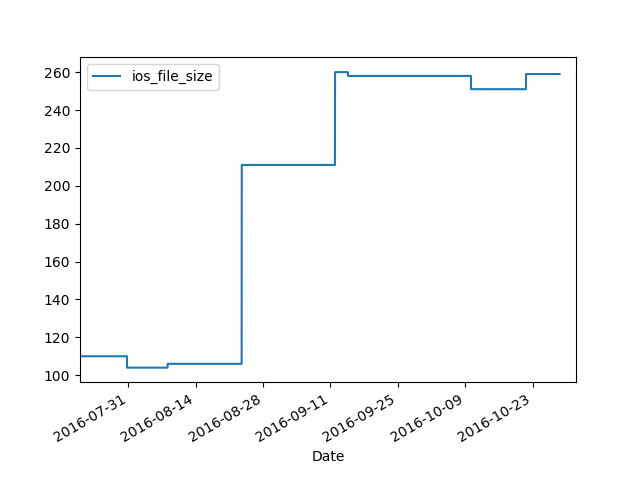












Prediction models:

Linear Regression :

Android\_total\_ratings = 7005220

IOS\_total\_ratings = 225650.92877735

Lasso Regression:

Android\_total\_ratings = 7013087.03581924

IOS\_total\_ratings = 225650.92877818

Ridge :

Android\_total\_ratings = 7005219.99999986

Ios\_total\_ratings = 225650.92840677

Random Forest:

Android\_total\_ratings = 7004878.9

IOS\_total\_ratings = 230592.635

**TENSOR FLOW:**

Here are the outputs for the 22 unique image file after scraping:

* **Android Files:**

file1- and\_screen1.jpg

monitor (score = 0.34349)

screen, CRT screen (score = 0.12488)

desktop computer (score = 0.07624)

web site, website, internet site, site (score = 0.05729)

television, television system (score = 0.02949)

file2- and\_screen1.jpg

lawn mower, mower (score = 0.17193)

golf ball (score = 0.11031)

croquet ball (score = 0.08029)

mountain tent (score = 0.03014)

bow (score = 0.02765)

file3- and\_screen1.jpg

web site, website, internet site, site (score = 0.49719)

monitor (score = 0.07830)

notebook, notebook computer (score = 0.05803)

iPod (score = 0.03292)

desktop computer (score = 0.02498)

file4- and\_screen1.jpg

web site, website, internet site, site (score = 0.59586)

comic book (score = 0.03351)

iPod (score = 0.02989)

screen, CRT screen (score = 0.02483)

television, television system (score = 0.02017)

file5- and\_screen1.jpg

web site, website, internet site, site (score = 0.62170)

television, television system (score = 0.08697)

monitor (score = 0.04946)

screen, CRT screen (score = 0.03223)

hand-held computer, hand-held microcomputer (score = 0.02868)

* **IOS Files:**

file1-ios\_screen1

web site, website, internet site, site (score = 0.42241)

comic book (score = 0.03248)

carousel, carrousel, merry-go-round, roundabout, whirligig (score = 0.02089)

fountain (score = 0.01781)

safety pin (score = 0.01440)

file2-ios\_screen2

web site, website, internet site, site (score = 0.12342)

maze, labyrinth (score = 0.07149)

comic book (score = 0.04789)

joystick (score = 0.04421)

television, television system (score = 0.03758)

file3-ios\_screen3

ashcan, trash can, garbage can, wastebin, ash bin, ash-bin, ashbin, dustbin, tra

sh barrel, trash bin (score = 0.15498)

joystick (score = 0.06405)

cannon (score = 0.03585)

maraca (score = 0.02727)

pedestal, plinth, footstall (score = 0.02715)

file4-ios\_screen4

web site, website, internet site, site (score = 0.58624)

monitor (score = 0.07197)

television, television system (score = 0.05955)

comic book (score = 0.04756)

teapot (score = 0.01425)

file5-ios\_screen5

comic book (score = 0.19361)

maze, labyrinth (score = 0.19330)

web site, website, internet site, site (score = 0.05236)

monitor (score = 0.02957)

book jacket, dust cover, dust jacket, dust wrapper (score = 0.02767)

file6-ios\_screen6

space shuttle (score = 0.23042)

joystick (score = 0.05992)

racer, race car, racing car (score = 0.05626)

scoreboard (score = 0.04957)

airliner (score = 0.04576)

file7-ios\_screen7

fountain (score = 0.20303)

carousel, carrousel, merry-go-round, roundabout, whirligig (score = 0.08314)

comic book (score = 0.05171)

toyshop (score = 0.03343)

monitor (score = 0.03227)

file8-ios\_screen8

web site, website, internet site, site (score = 0.60886)

television, television system (score = 0.05665)

monitor (score = 0.01996)

notebook, notebook computer (score = 0.01607)

iPod (score = 0.01180)

file9-ios\_screen9

web site, website, internet site, site (score = 0.11637)

laptop, laptop computer (score = 0.08080)

notebook, notebook computer (score = 0.05349)

joystick (score = 0.04791)

monitor (score = 0.04169)

file10-ios\_screen10

web site, website, internet site, site (score = 0.36779)

envelope (score = 0.16914)

binder, ring-binder (score = 0.05812)

tray (score = 0.01764)

monitor (score = 0.01721)

file11-ios\_screen11

web site, website, internet site, site (score = 0.88357)

menu (score = 0.00803)

slot, one-armed bandit (score = 0.00404)

washer, automatic washer, washing machine (score = 0.00371)

hand-held computer, hand-held microcomputer (score = 0.00296)

file12-ios\_screen12

web site, website, internet site, site (score = 0.36619)

safety pin (score = 0.02004)

sunglasses, dark glasses, shades (score = 0.01677)

toilet seat (score = 0.01562)

washer, automatic washer, washing machine (score = 0.01438)

file 13-ios\_screen13

aircraft carrier, carrier, flattop, attack aircraft carrier (score = 0.09968)

pole (score = 0.03657)

wing (score = 0.02655)

lakeside, lakeshore (score = 0.02437)

magnetic compass (score = 0.02396)

file14-ios\_screen14

web site, website, internet site, site (score = 0.89077)

menu (score = 0.00364)

monitor (score = 0.00185)

screen, CRT screen (score = 0.00184)

analog clock (score = 0.00177)

file15-ios\_screen15

web site, website, internet site, site (score = 0.94092)

analog clock (score = 0.00367)

envelope (score = 0.00291)

monitor (score = 0.00225)

screen, CRT screen (score = 0.00217)

file16-ios\_screen16

web site, website, internet site, site (score = 0.22753)

envelope (score = 0.09163)

Band Aid (score = 0.03712)

pinwheel (score = 0.02946)

airship, dirigible (score = 0.02486)

file17-ios\_screen17

laptop, laptop computer (score = 0.49859)

web site, website, internet site, site (score = 0.10646)

monitor (score = 0.06384)

screen, CRT screen (score = 0.02985)

notebook, notebook computer (score = 0.02801)

**References:**

* <http://stackoverflow.com/>
* **Class Code**
* **https://preinventedwheel.com/easy-python-time-series-plots-with-matplotlib/**