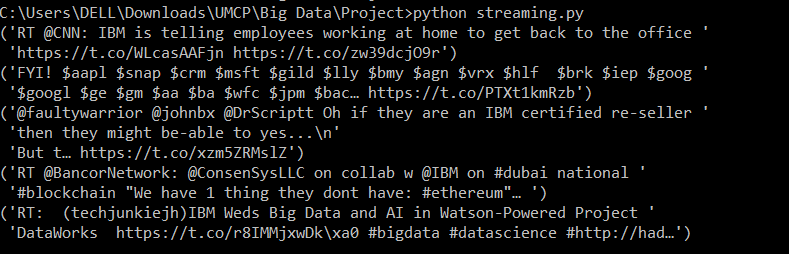
**Read-me:**

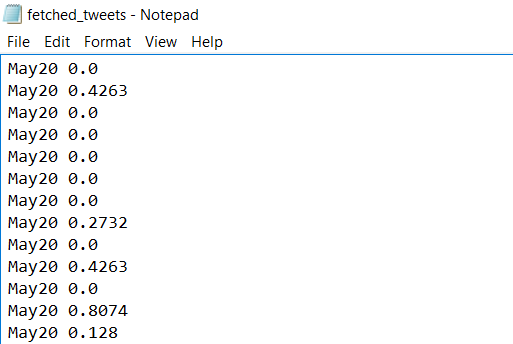
**1.Collect tweets using streaming API-**

Run python script streaming.py on your terminal. The script outputs tweet text on the terminal and also writes an output file in the directory where you saved the script. The output file has records in the form of <Date Sentiment compound score> (eg. May05 0.54). The file has filter set to get tweets for ‘IBM’. Alternatively, you can change this line in the code (stream.filter(track=['ibm'])) to change to any other filter you want. The script keeps running until you stop it using CTRL + C / D.

Tweet text output on terminal-



Output file fetched\_tweets.txt (written in the same directory where you save the python script)-



**2. Put the output file fetched\_tweets.txt to HDFS-**

Start Hadoop on your terminal and put the file onto your HDFS.

$cd $HADOOP\_HOME

$./bin/start-all.sh

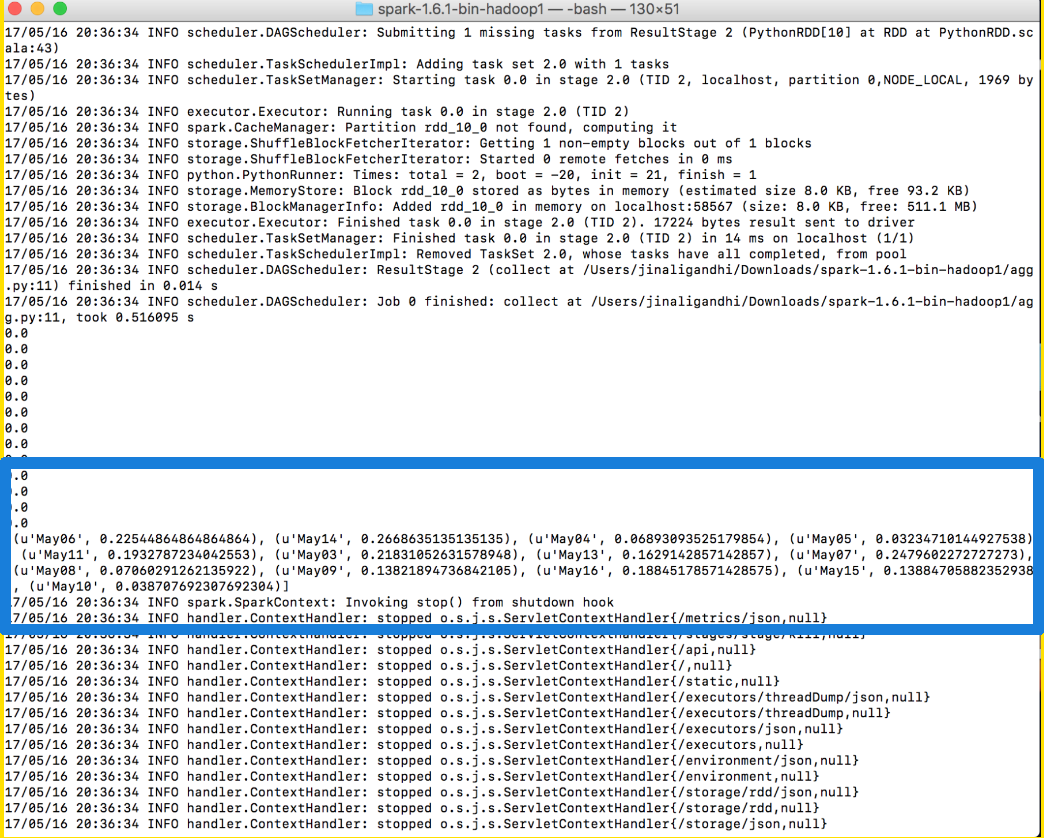
$./bin/hadoop fs -put <path where you saved fetched\_tweets.txt> <absolute or relative path in HDFS where you want to save this file (eg. /user/<user\_id>/.)>

**3. Run script to aggregate sentiment for a day for a particular company**

$spark-submit <path where you saved agg.py script>

This script outputs list of key-value pair of **day-average** sentiment for the company using the fetched\_tweets.txt file saved on HDFS.

Run the **pyspark** script agg.py:

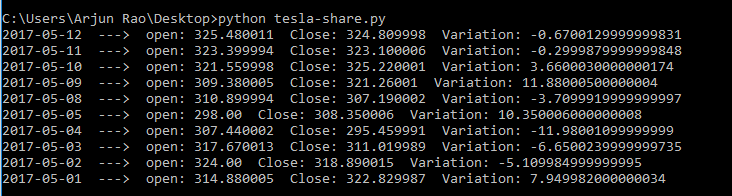


The highlighted blue box shows the aggregated sentiment score of each day. We then stored each of these aggregated outputs into a text file and stored them all in HDFS as well as on our local system.

4. **Get financial information from yahoo finance**

Navigate to location where the python script files are stored. In this case, we are running the tesla-share.py file

python tesla-share.py

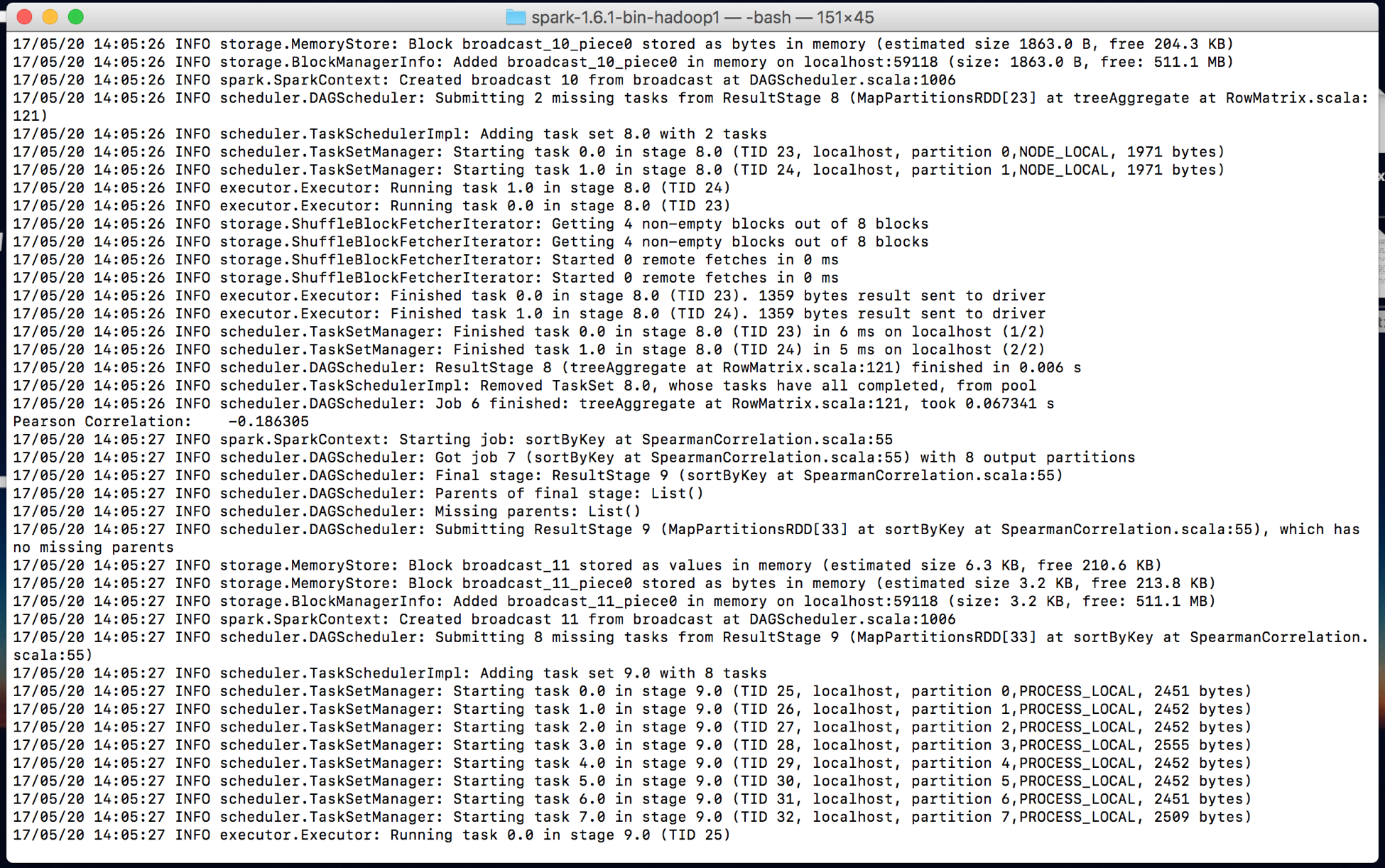


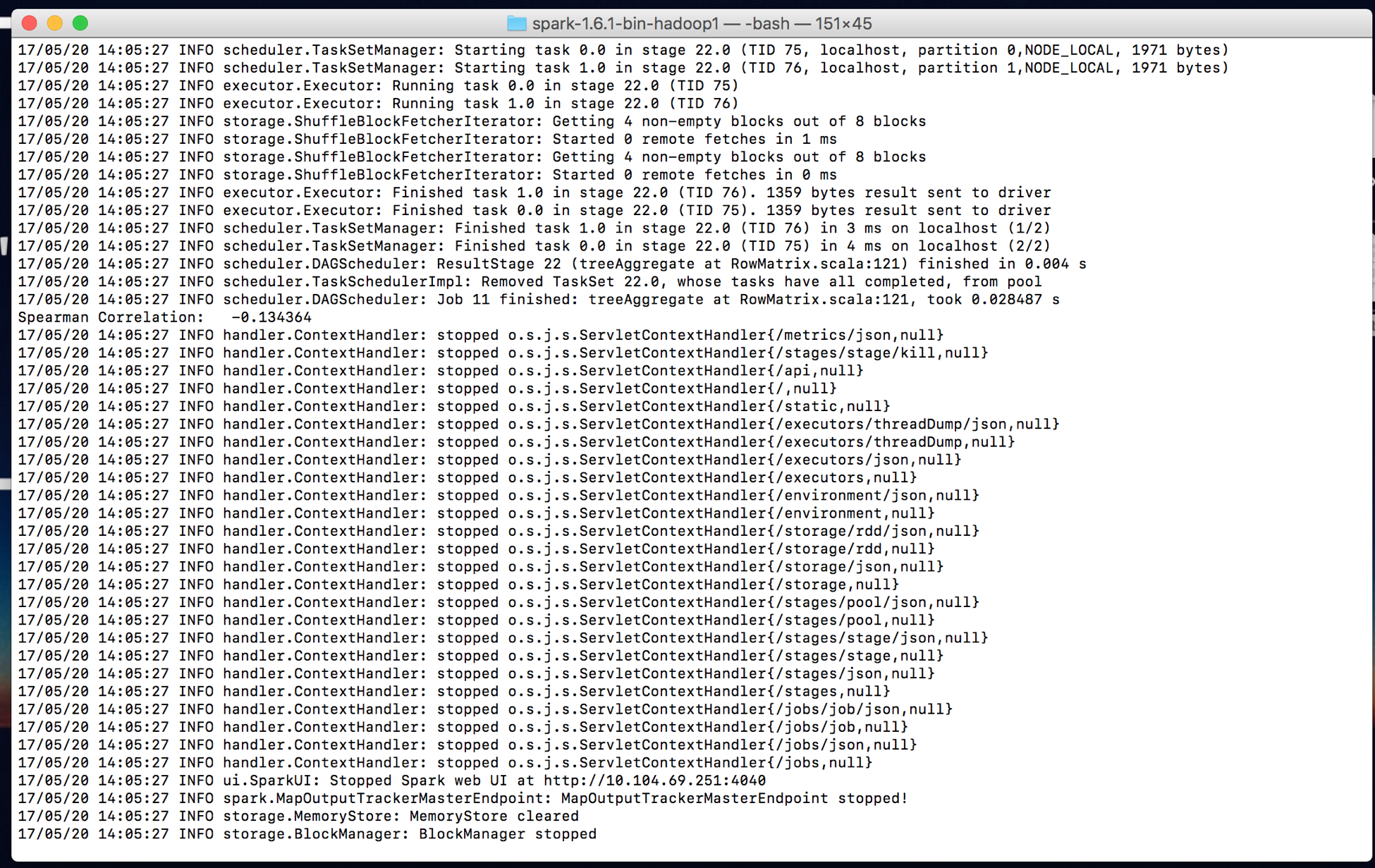
5. **Using the output of key-value pair of date and average sentiment, find pearson and spearman correlation**

The .txt files of average sentiment and stock price by date are then accessed by the correlation pyspark program.

$spark-submit <path where you save corr.py script>

This script gives the correlation output (Pearson and Spearman) on terminal:





6. The key-value pair data of date-average sentiment in step 3, finance data from step 4 and correlation output from step 5 are saved in AWS MySQL database and front end is developed in Django

In order to run the final application, first we need to install django

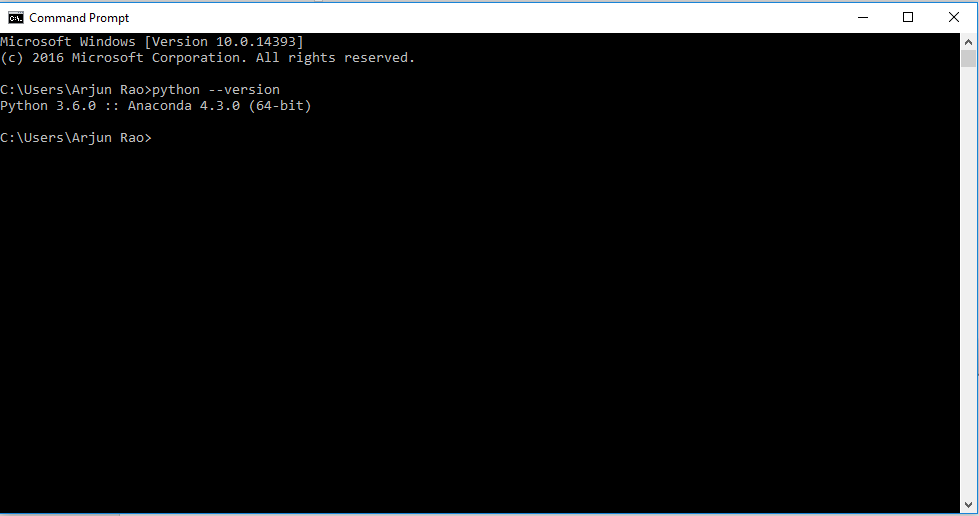
Open Terminal (Mac) or Windows Powershell (run as administration)

Check for python and pip or easy install version (Python 3.4 or above required)

python --version

pip --version

easy\_install --version



Install Django

pip install django

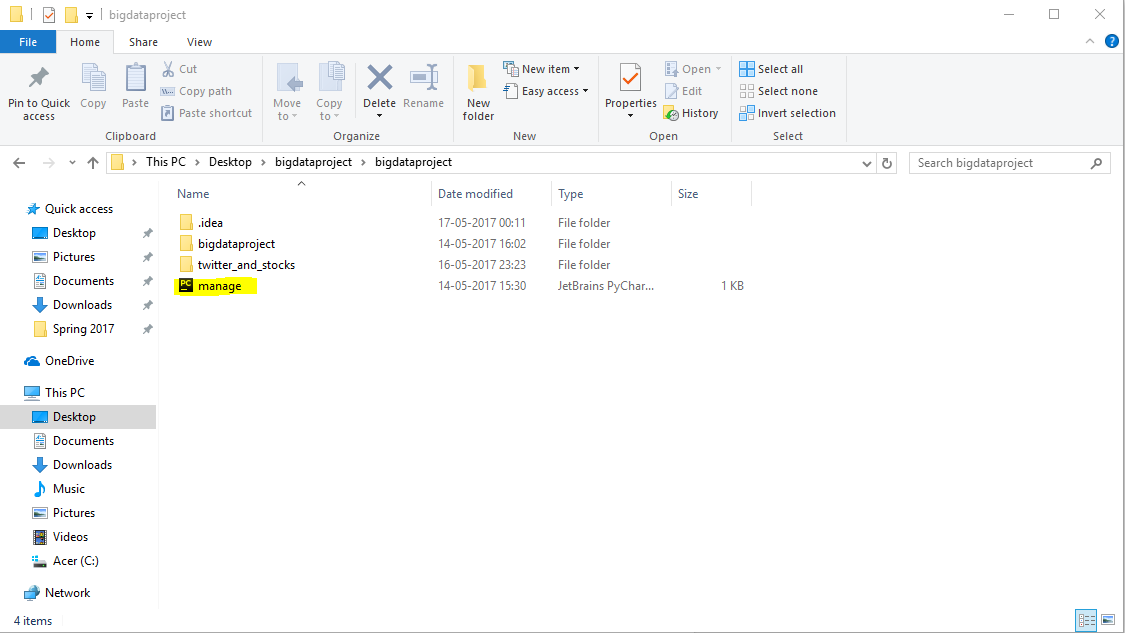
Check Django version

django-admin -version

django version.PNG

To run the application, first unzip to any local directory

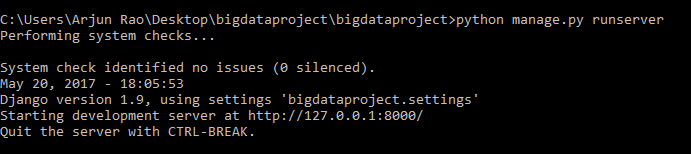
In terminal, navigate to the folder that contains the file ‘manage.py’



navigate terminal.PNG

Once here, us the following command to run the application on local server

python manage.py runserver



\*\* if above command causes python to stop responding, use

python manage.py runserver --noreload

Go to <http://127.0.0.1:8000/twitter_and_stocks/correlation/>