Department of Computing UWI
COMP4621 Programming for Data Science
Lab 03

Task 0:

a) Load the data from the file COMP4621_NKE.csv into a pandas DataFrame and create 2 time series plots for data covering the first and last quarters of 2018.

Task 1:

Load tweet data from the files provided and determine the sentiment of each tweet using the python `textblob` library. Remember to use the utility function provided (and import the `re` library) to `cleanup_tweets()`. *The function is defined below.*

Holness.csv has tweets from Andrew Holness @AndrewHolnessJM MarkGolding.csv has tweets from Mark Golding @MarkJGolding

a) Who has more positive tweets?

Task 3:

a) Retrieve ANY 25 tweets from the file *Jamaica.csv* which has tweet data retrieved using keyword "Jamaica"
b) Extract the following from the tweets and use them to create a DataFrame based on the table details below:

Description	**Tweet Result Property**	**DataFrame Column Name**
The tweet text	`text`	TweetText
The person who sent the tweet	`tweet.user.screen_name`	SentBy
The location of the person who sent the tweet	`user.location`	SenderLocation
The date that the tweet was sent	`created_at`	SendDate

c) Using the DataFrame column `TweetText`, and the `textblob` library, generate the `subjectivity and polarity` of each `TweetText` and add a new columns with `subjectivity` and `polarity` values.
d) Based on this snapshot of data, what is the dominant sentiment currently being expressed about Jamaica (justify your answer).

```
import re
def cleanup_tweet(tweet):
    return''.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z  \t])|(\w+:\/\/\S+)",
  " ", tweet).split())
```