**Python Project Sentiment Analysis [using WebScraping]**

**Problem Statement**

IMDB provides a list of celebrities born on the current date. Below is the link: <http://m.imdb.com/feature/bornondate>

Get the list of these celebrities from this webpage using web scraping (the ones that are displayed i.e top 10). You have to extract the below information:

1. Name of the celebrity
2. Celebrity Image
3. Profession
4. Best Work

Once you have this list, run a sentiment analysis on twitter for each celebrity and finally the output should be in the below format

1. Name of the celebrity:
2. Celebrity Image:
3. Profession:
4. Best Work:
5. Overall Sentiment on Twitter: Positive, Negative or Neutral

Hint: Use IMDB scrapping sample example as reference for scraping the mentioned web page. For sentiment analysis use the Twitter sentiment code as reference.

**Tools and Packages Used**

* Version: Python 3.6**[VERY IMPORTANT]**
* Tweepy: An easy-to-use Python library for accessing the Twitter API.
* Text Blob: Text Blob is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.
* Beautiful Soup [Symbol] Beautiful Soup provides a few simple methods and Pythonic idioms for navigating, searching, and modifying a parse tree using Python parsers like lxml. It automatically converts incoming documents to Unicode and outgoing documents to UTF-8. Here's the documentation.
* OpenCV --- for showing the reading and showing the images

**Challenges Faced during the project**

Course content was good enough to pull out this project out in almost a week’s time. However, there are some hindrances while coding the assignment

1. Beautiful Soup with Parameters was not working as expected.
2. Installation of Tweepy and compatibility with Python I had used.

But I was new Python programming so coding standard wise I was not very comfortable initially but it took me couple code refinements to bring the coding style to a decent one.

**Reference**:

1. Used text blob library to evaluate the sentiment analysis of tweets:
2. Open CV Documentation: <https://docs.opencv.org/2.4/index.html>
3. Tweepy: <http://docs.tweepy.org/en/v3.5.0/>
4. Sentiment Analysis --- <https://dev.to/rodolfoferro/sentiment-analysis-on-trumpss-tweets-using-python->

Implementation Specific Notes

1. For sentiment analysis I have implemented using 2 approaches

Approach 1

Using the TextBlob Sentiment analysis API I have printed the polarity in the console which is lot more realistic

Approach 2

Is naïve approach which is based on word count(which is discussed as part of curriculum)

1. Images gets changes every five seconds. (which is hardcoded I as well new that I can show all images in one template use Matplot lib but due to want of time I have not implemented it)

Solution Flow Diagram

 There are two major files and corresponding important classes…

1. Web Scraping
2. Analytics --🡪
   1. Main(TweetListener)
   2. SentimentAnalysisWC (based on word count)

Solution Flow Diagram

