Twitter Sentiment Analysis for Celebrities by Data extraction through Web-Scraping on IMDb and Famous Birthdays website

By

SHIVIKA MALIK
Department of Computer Science
University At Albany
Albany, New York
smalik2@albany.edu

A Project Report submitted to

University at Albany

in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

in

Computer Science and Engineering

Prof. Jackson Marques de Carvalho

Date: 15th May 2019

AUTHORIZATION FOR REPRODUCTION OF PROJECT

I grant permission for the reproduction of this project in its entirety, without further authorization from me, on the condition that the person or agency requesting reproduction, absorb the cost and provide proper acknowledgment of authorship.

Date: 15th May 2019

Shivika

Shivika Malik
244 Manning Blvd, Apt #2
Albany, New York, 12206

ABSTRACT

Web Scraping tools and APIs facilitate the significant extraction of information from the Internet. Web scraping that is a common practice to load, prepare and statistically analyze specific structured or unstructured data from the Internet, it is an essential technique in Marketing and Data Science.

In this project, we used Sentiment Analysis to evaluate sentiment of tweets and then find correlation with the rankings of celebrities on different websites These tasks were accomplished by mining tweets using twitter's API and then processed further for analysis.

This Paper covers the available techniques and development in the recent history of Web Scraping. Legal aspects of Web Scraping have been introduced. Software tools used for this technique are listed with a brief summary of their functionalities.

We would emphasize the importance of Open Data and social media data such as Twitter as a scraping target.

Keywords—*Data Science, web scraping, twitter data, open data, sentiment analysis.*

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

1.1 PROBLEM STATEMENT

1.1.1 PURPOSE

- The source information available on the World Wide Web is still in the format of a Hypertext Markup Language (HTML) page.
- Automatic extraction of the information is difficult because the intended reader is always a human.
- With the rapid growth of the World Wide Web, there has been a significant change in the way we share, collect, and publish the data.
- These days, most of the information is being stored online, both in structured and unstructured forms.
- Regarding certain questions or research topics, this has resulted in a new problem—no longer is the concern of data scarcity and inaccessibility but, rather, one of overcoming the tangled masses of online data. These utilizations are often only possible because the existence of automated Web Scraping. Without these techniques, it would be impossible to collect the amount of data repeatedly and in reasonable time. [References 1]

1.1.2 WEBSITE SELECTION

- IMDB and FamousBirthdays websites provide a list of celebrities born on the current date that is March 12th. Below are the links:
  a. IMDb
  b. Famous Birthdays
- Get the list of the celebrities from these webpages using web scraping. We have to extract the below information:
  c. Name of the celebrity
  d. Profession
  e. Rank
  f. Visualization: Using Pyramid graph

1.1.3 TWEET COLLECTION

- Once we have a list of celebrities from both the websites, we run a sentiment analysis on twitter using TextBlob for each celebrity and finally the output is in the below format:
  a. Name of the celebrity
  b. Profession
  c. Rank
  d. Mean Sentiment value on Twitter

1.2 ASSUMPTIONS

- Get a list of top 10 celebrities from IMDb webpage using web scraping. Extract the below information:
    - Name of the celebrity
    - Celebrity Image
    - Profession
    - Best Work
- Once we have this list, run a sentiment analysis on twitter for each celebrity and the final output should be in the below format:
    - Name of the celebrity:
    - Celebrity Image:
    - Profession:
    - Best Work:
- Overall Sentiment on Twitter: Positive, Negative or Neutral should be displayed on the Output console.
- Data Visualization using Pyramid graph, showing the sentiment polarity of the celebrity along with the celebrity rankings on IMDb.
- This application can be used with Static as well as Dynamic data available on the Internet.
- This application would help us predict a celebrity's future growth by understanding the model built through web scraping and sentiment analysis.
- The data model can predict using linear regression model.
- This application can store information in CSV, JSON and YAML file format.
- This implementation can work on any website that gives a celebrity a rank with the help of users or common people comments, tweets or likes.
-

1.3 PROJECT DEFINITION

- As the Internet has continued to grow, the amount of data publicly available has enlarged into an unbelievable size. This data has a lot of power and potential, but the availability doesn't necessarily translate into the accessibility; many questions cannot easily be answered using this data due to the structure and immense size of the web.
- The automated gathering of data from the Internet is nearly as old as the Internet itself.
  Although web scraping is not a new term, in years past the practice has been more commonly known as screen scraping, data mining, web harvesting, or similar variations. General consensus today seems to favor web scraping, so that is the term I'll use throughout the book, although I will occasionally refer to the web scraping programs themselves as bots.
- In theory, web scraping is the practice of gathering data without interacting with the APIs or, without a human interference. This is accomplished by writing an automated program that queries the web server, requests the data (usually the data in the form of the HTML or files that comprises the web pages), and then parses that data to extract the information that is needed.

- Web scraping provides a great solution to facilitate digestion of the web data, providing specific rules on what data to collect and aggregate. Still, web scraping requires a sophisticated understanding of programming and web technologies like HTML and the structure of data on the web (example: Document Object Model (DOM)).
- As a result, programmers who are collecting the data from the web often find themselves writing tedious and complicated scripts; and the people without technical experience are unable to collect the data. Here, we present a browser-integrated visual interface for automated data acquisition two different web pages with the goal of reducing the current barriers of extracting relevant data from the web.
- Our program, *"IMDb"*, is designed to extract data for celebrities born on 12th March. The Internet Movie Database (IMDB) contains celebrity names, profession and rank. Other website "*FamousBirthdays*" too provides the same information for a different set of celebrities.
- These sites are using a templating engine to transform database records into well formatted, user-readable content. Our tool takes the advantage of the structural similarity across both the web pages to allow users to scrape content after creating a template corresponding to just one example page.
- In the current year large amount of data is transmitted using different social media platform and this data contains information about every event or topic, Twitter being one of the most famous social media websites, provides us micro-blogging service. Millions of users sign in daily to tweets about any event they want to express their feelings for, and Twitter provides them that platform to them. Each tweet contains maximum of 140 characters, each tweet may not be useful but still we can examine the sentiment from those tweets.
- These tasks were accomplished by mining tweets using twitter's API and then process further for analysis. For tweets sentiment, we used Vader sentiment analysis.
- In this paper, we describe the implementation of our system, present our observations and results from using IMDb.py, FamousBirthdays.py and Twitter.py and discuss the challenges of designing a simple, flexible tool for data acquisition on the web.
- In practice, web scraping encompasses a wide variety of programming techniques and technologies, such as data analysis and information security.

2 SOLUTION FLOW DIAGRAM

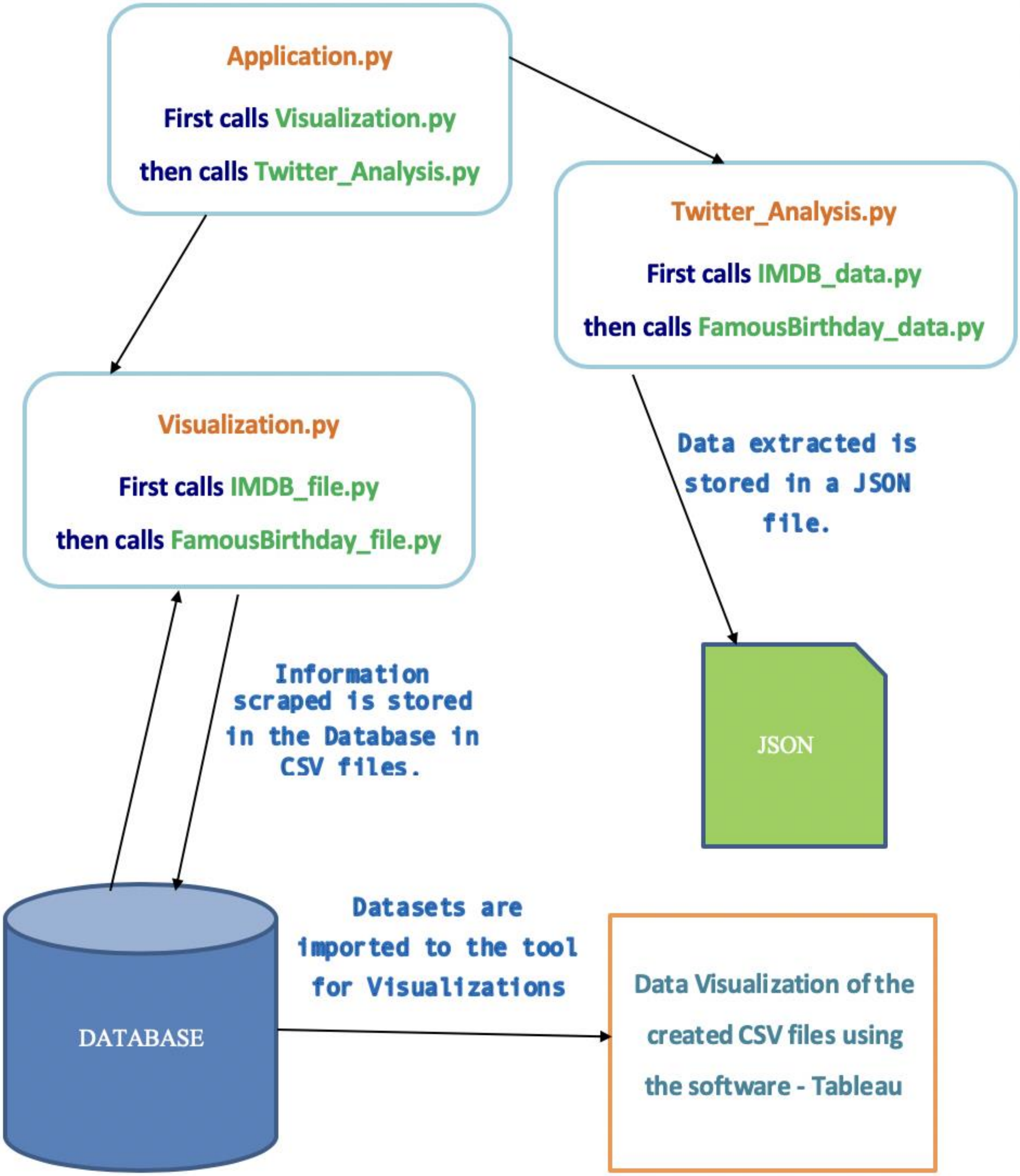

Application.py
First calls Visualization.py
then calls Twitter_Analysis.py
Twitter_Analysis.py
First calls IMDB_data.py
then calls FamousBirthday_data.py
Visualization.py
First calls IMDB_file.py
then calls FamousBirthday_file.py
Data extracted is stored in a JSON file.
Information scraped is stored in the Database in CSV files.
JSON
DATABASE
Datasets are imported to the tool for Visualizations
Data Visualization of the created CSV files using the software - Tableau

Figure 1 - Workflow of the Project

2.1 APPLICATION.PY

Runs the whole application at Once. It will run Visualization.py and Twitter_Analysis.py files which further perform Data Visualization and Sentiment Analysis.

2.2 TWITTER_ANALYSIS.PY

Calls FamousBirthday_data.py and IMDB_data.py to retrieve the celebrity names using Web Scraping and it then fetches tweets for the celebrities through Twitter API and performs Sentiment Analysis using TextBlob. The final output is stored in a json file called Result.json.

2.3 VISUALIZATION.PY

Calls FamousBirthday_file.py and IMDB_file.py to fetch the names, profession and rank of the celebrities using Web Scraping and is stored in two CSV files, called imdb.csv and famousbirthdays.csv. The csv files are used to perform Data Visualization for the celebrities which are common on IMDb and Famous Birthdays websites.

2.4 DATABASE

It stores the output files namely, Result.json, imdb.csv, famousbirthdays.csv.

2.5 TABLEAU SOFTWARE

This software was used for further visualizations. imdb.csv and famousbirthday.csv files are imported to Tableau to get number of professions according to the ranks for both the websites (screenshots of the Tableau output are attached in the Conclusion section).

3 CHALLENGES AND SOLUTIONS

3.1 SCRAPING THE DYNAMIC CONTENT

- The IMDb website contains content which is dynamic in nature and it can only be loaded by JavaScript.
- The web scraping tool BeautifulSoup does not invoke the JavaScript. Therefore, scraping the data was initially problem.
- To solve this problem, the webpage was opened using Selenium which can invoke JavaScript and load the dynamic content.
- The page source from Selenium driver is then passed to BeautifulSoup scraper to solve the problem.

3.2 WHY NOT API?!

- They can provide a convenient stream of well-formatted data from one server to another.
- You can find an API for many different types of data you might want to use such as Twitter posts or Wikipedia pages. In general, it is preferable to use an API (if one exists), rather than build a bot to get the same data.
- However, there are several reasons why an API might not exist: You are gathering data across a collection of sites that do not have a cohesive API.
- The data you want is a fairly small, finite set that the webmaster did not think warranted an API.
- The source does not have the infrastructure or technical ability to create an API.
- Even when an API does exist, request volume and rate limits, the types of data, or the format of data that it provides might be insufficient for your purposes.

3.3 OPINION MINING

Sentiment Analysis basically deals with automatic scanning of the text and establishing the purpose of the scanned text. Basically, it is important to determine if the text scraped and extracted from the website is useful or not.

3.4 CONTEXT AND POLARITY

- Analyzing the sentiments of the users without context can be difficult. However, machines cannot learn about the contexts if they are not mentioned explicitly. Therefore, the names of the celebrities scraped from the 2 websites are passed through the Twitter API.
- The other problem that arise from context is changes in the polarity of the sentiments time to time. Hence, the celebrity data was analyzed only for the month they were born in and for a particular to be more specific with the data that is supposed to analyzed.

3.5 Emojis

- Emojis play an important role in the sentiments of the text. According to Guibon et al., there are two types of emojis. Western emojis (example- :P) are encoded in only one character or in a combination of a couple of them whereas Eastern emojis (e.g. ¯ \ _ (ツ) _ / ¯) are a longer combination of characters of a vertical nature.
- Sentiment analysis performed over tweets on Twitter require a special attention to character-level as well as word-level. Therefore, we require a lot of preprocessing. For example, we might want to preprocess social media content and transform Western as well as Eastern emojis into tokens and whitelist them in order to help improve the performance of sentiment analysis.

3.6 Defining Neutral

- When we train the sentiment analysis model, defining positive, negative and neutral tags to the classifiers is important. Defining what neutral means is a challenge to tackle in order to perform accurate sentiment analysis. Since tagging the data requires that tagging criteria should be consistent.
- Here's some ideas on what a neutral tag might contain:
    - a. **Objective texts:** These texts do not contain explicit sentiments, so you should include those texts into the neutral category. Below are some examples:
        - i. The package is nice.
        - ii. The package is red.
    - b. **Irrelevant information:** If the data isn't preprocessed to filter out irrelevant information, it can be tagged as neutral. Our data did not require preprocessing.
    - c. **Texts containing wishes:** Some wishes like *"I wish I could meet this celebrity"* are generally considered as neutral.

4 SOFTWARE TOOLS USED

4.1 PYTHON 2.7 (64-BIT)

Programming language was used to build the application.

4.2 BEAUTIFULSOUP4

Python library for pulling data out of HTML and XML files.

4.3 TWEEPY

Open Source python library for Twitter API.

4.4 TWITTER API

A Python wrapper for performing API requests such as searching for users and downloading tweets. This library handles all of the OAuth and API queries for you and provides it to you in a simple Python interface.

4.5 SELENIUM

The web driver kit emulates a web-browser and executes JavaScript to load the dynamic content.

4.6 TEXTBLOB

Python library to perform Text Sentiment Analysis.

4.7 XML

A fast html and xml parser for BeautifulSoup4.

4.8 GOOGLE CHROME

Web Browser to perform web scraping.

4.9 CHROME DRIVER

Driver for Selenium to invoke Google Chrome.

4.10 TABLEAU PUBLIC

Data Visualizations for the data extracted using Web Scraping.

5 WEB SCRAPING

5.1 WHAT IS WEB SCRAPING?

- The automated gathering of the data from the Internet is nearly as old as the Internet. In the past years, web scraping was commonly known as screen scraping, data mining, web harvesting, or similar variations.

[References 1]

- Web scraping is a practice of gathering data through any means other than a program interacting with an API or through a human using a web browser. This technique is accomplished by writing an automated program that queries a web server, requests data (in the form of HTML or files comprising of web pages), and then parses that data to extract the information needed. Web scraping encompasses a wide variety of programming techniques and technologies, like data analytics and information security.

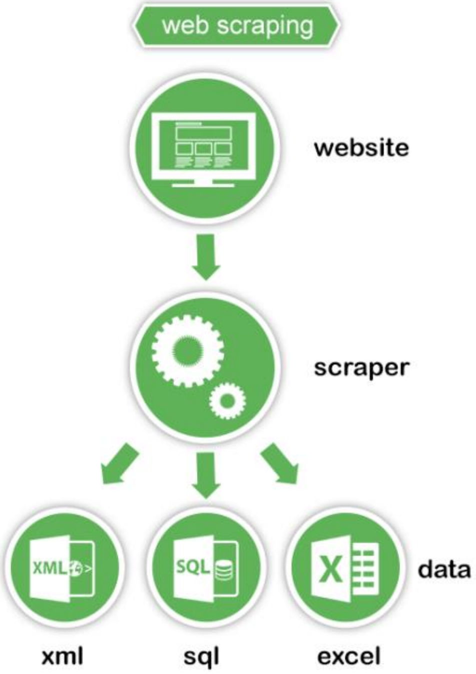


Figure 2 – Web Scraping Method

5.2 PURPOSE OF WEB SCRAPING

- Enormous amounts of source information, available on the World Wide Web, are still in the format of a Hypertext Markup Language (HTML) page. Automated extraction is difficult, because the intended reader was a human.
- Browsers are handy for executing JavaScript, displaying images, and arranging objects in a more human-readable format, web scrapers are excellent at gathering and processing large amounts of data.

- Rather than viewing one page at a time through the narrow window of a monitor, one can view databases spanning thousands or millions of pages at once.
- Web scrapers can go places that traditional search engines cannot. A Google search for “cheapest flights to Boston” will result with advertisements and popular flight search sites. Google only knows what these websites say on their content pages, and not the exact results of various queries entered into a flight search application. However, a well-developed web scraper can chart the cost of a flight to Boston over time, across a variety of websites, and tell you the best time to buy your ticket. [References 2]
- Rapid growth of the World Wide Web has significantly changed the way we share, collect, and publish data. Vast amount of information is being stored online, both in structured and unstructured forms. Regarding certain questions or research topics, this has resulted in a new problem—no longer is the concern of data scarcity and inaccessibility but, rather, one of overcoming the tangled masses of online data. These utilizations are often only possible because the existence of automated Web Scraping. Without these techniques, it would be impossible to collect the amount of data repeatedly and in reasonable time.
- There are obviously many extremely practical applications of having access to nearly unlimited data: market forecasting, machine language translation, and even medical diagnostics have benefited tremendously from the ability to retrieve and analyze data from news sites, translated texts, and health forums, respectively.
- Even in the art world, web scraping has opened up new frontiers for creation. The 2006 project “We Feel Fine” by Jonathan Harris and Sep Kamvar, scraped a variety of English-language blog sites for phrases starting with “I feel” or “I am feeling.” This led to a popular data visualization, describing how the world was feeling day by day and minute by minute. Regardless of your field, there is almost always a way web scraping can guide business practices more effectively, improve productivity, or even branch off into a brand-new field entirely.

6 SENTIMENT ANALYSIS OF TWEETS

6.1 WHAT IS SENTIMENT ANALYSIS?

- Sentiment Analysis (*Opinion Mining)* is a field within Natural Language Processing (NLP) that builds a system that identify and extract opinions within the text. Usually, besides identifying the opinion, these systems extract attributes of the following expressions:
    - *Polarity*: *positive* or *negative* opinion,
    - *Subject*: the thing that is being talked about
    - *Opinion holder*: the person, or entity that expresses the opinion.
- Since public and private information available over the Internet is constantly growing, a large number of texts expressing opinions are available for review on different websites, forums, blogs, and social media.
- With the help of sentiment analysis systems, the unstructured information could be transformed automatically into structured data for public opinions about products, services, brands, politics, or any topic that people can express their opinions about. This data can be very useful for commercial applications like marketing analysis, public relations, product reviews, net promoter scoring, product feedback, and customer service. [References 3]

6.2 FINE-GRAINED SENTIMENT ANALYSIS

- There are many types of sentiment analysis and tools that focus on polarity (positive, negative, neutral) to systems that detect feelings and emotions (neutral, happy, sad etc.) or identify intentions (e.g. interested or not interested).
- If one is interested in being precise about the level of polarity of the opinion, so instead of just talking about positive, neutral, or negative opinions you could consider the following categories:
    - Positive
    - Neutral
    - Negative
- This is usually referred to as fine-grained sentiment analysis. This could be, for example, mapped onto a 5-star rating in a review, e.g.: Very Positive = 5 stars and Very Negative = 1 star.
- Textblob provides different flavors of polarity by identifying if the positive or negative sentiment is associated with a particular feeling, such as, negative feelings: anger, sadness or worries or positive feelings: happiness, love or enthusiasm.

6.3 SENTIMENT ANALYSIS WITH WEB SCRAPING

[References 5]

- While data scraping is quite challenging in itself, we would reflect on how opinion mining/sentiment analysis can help the enterprise clients better.
- Sentiment Analysis deals with automatic scanning of texts and establishing its nature or its purpose.
- It is important to determine if the text scraped and extracted from a website is useful or not; or whether it relates with the subject that is mentioned in the title.
- The function of sentiment analysis is to analyze user entries (customer reviews, product feedbacks, service feedback forms) and indicate the feelings expressed (happiness, dissatisfied etc.).
- On a simple scale, this can be achieved by establishing a scoring system from 1 – 5 with 5 being most positive (or such similar measure) where the words are associated with an emotion. The score of each word, and whole text, is then calculated to see what the opinion or sentiment indicated.

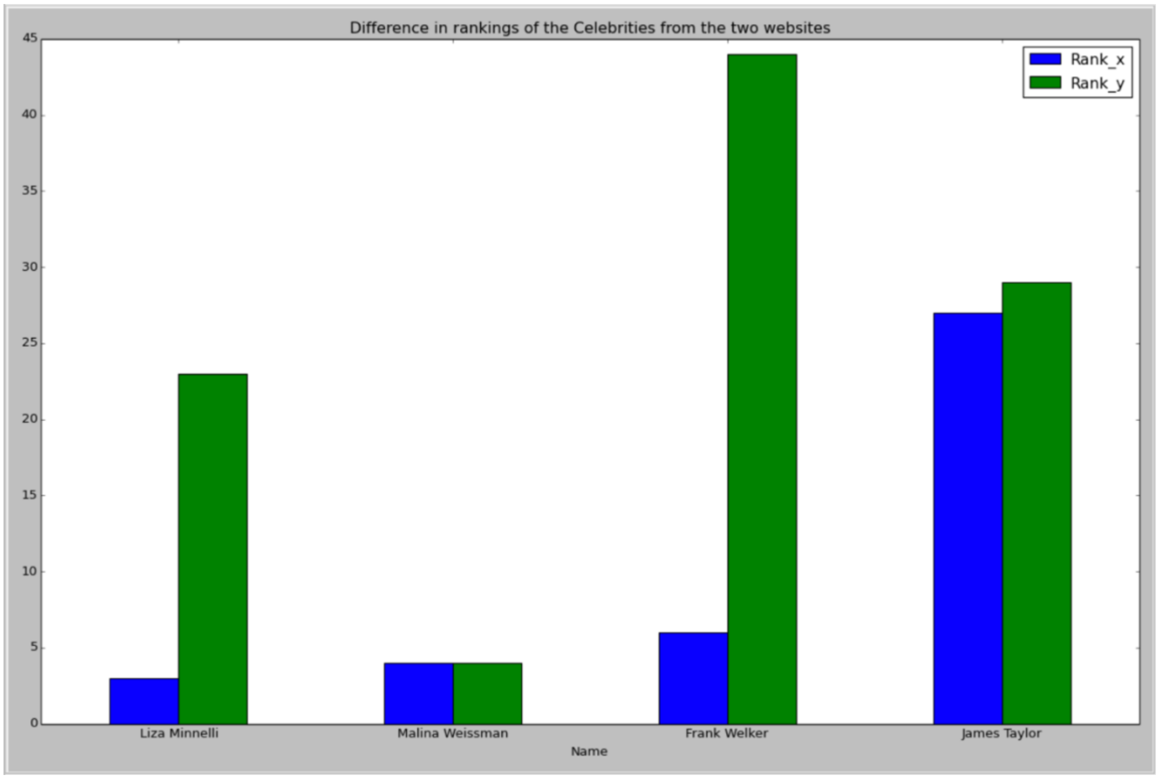
7 Conclusions

7.1 Comparisons of celebrity rankings on the 2 different websites

- After scraping the Celebrity names, Profession and Ranks from the 2 websites and storing them in the Database, the 2 datasets were merged using Inner join to find the celebrity names which were common on both the website.

🟦 Rank of Celebrities on IMDb website

🟩 Rank of Celebrities on Famous Birthdays website

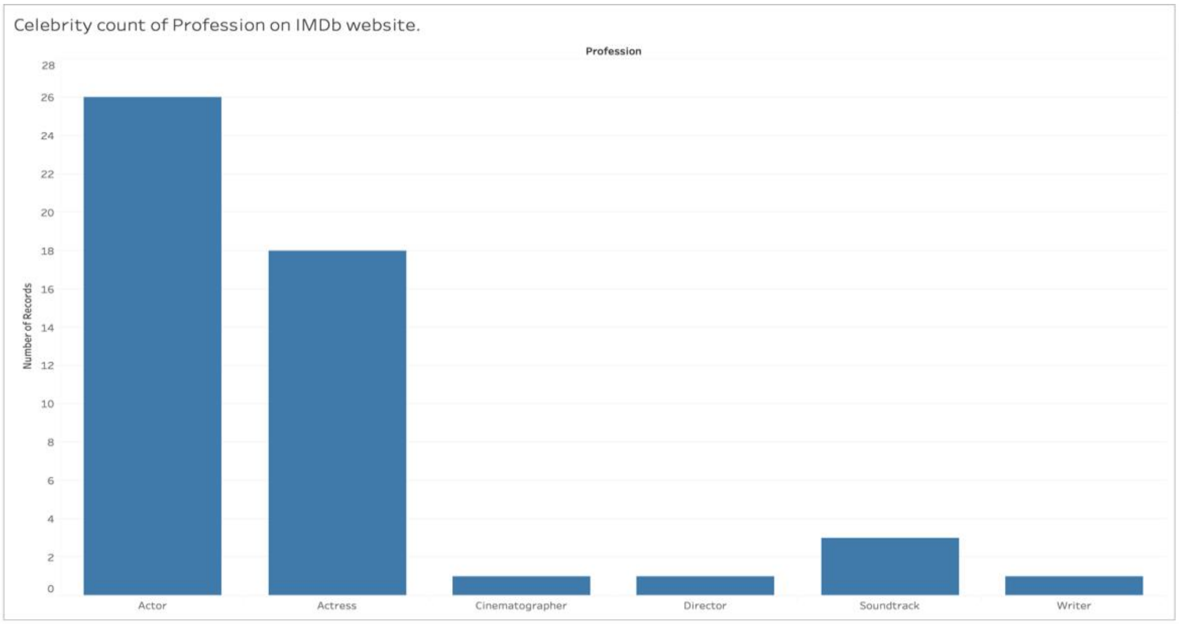


```
               Name Profession_x  Rank_x   Profession_y  Rank_y
0     Liza Minnelli   Soundtrack     2.0  Movie Actress      22
1   Malina Weissman      Actress     4.0     TV Actress       4
2      Frank Welker        Actor     6.0    Voice Actor      43
3      James Taylor   Soundtrack    37.0    Folk Singer      29
```

- We can conclude that there was an extreme difference in the rankings of the celebrities on both the websites even after knowing the fact that the rankings were given to the celebrities according to the viewer's choice.
- Only one celebrity ***Malina Weissman*** got the same ranking on both the websites whereas the ranking of ***Frank Welker*** differed by **37** as his ranking on IMDb is 6 and on Famous Birthdays is 43.

7.2 NUMBER OF RECORDS FOR EACH PROFESSION ON BOTH THE WEBSITES

- We can see in this plotted graph that there are total 50 records of celebrities and there are 6 different kind of Professions on IMDb website.



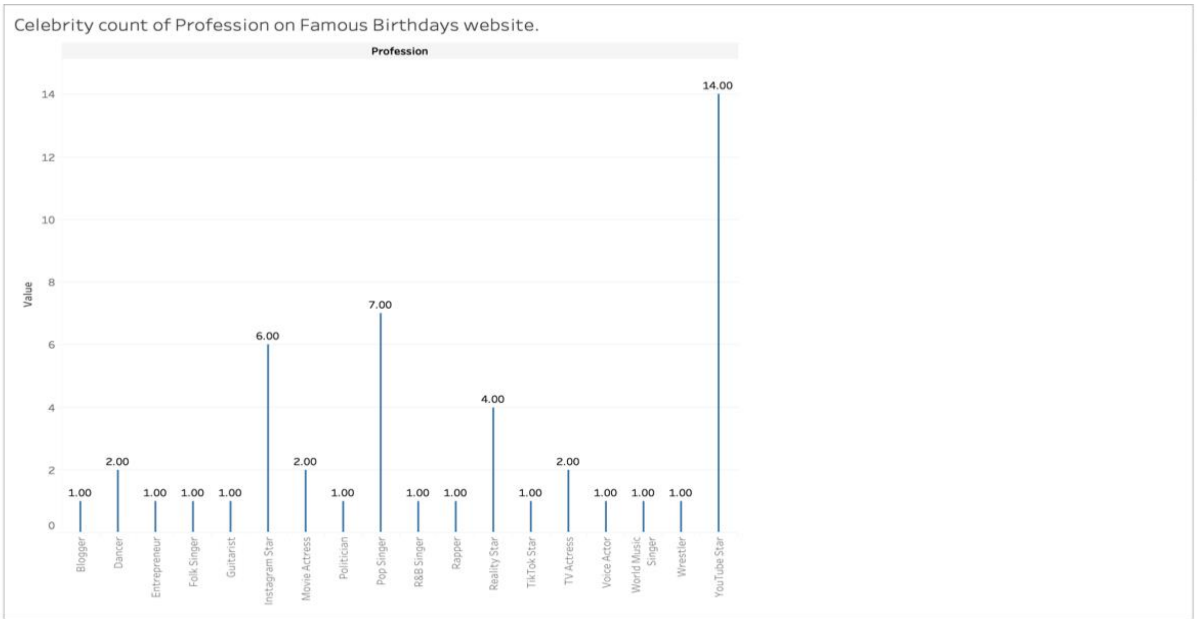
Summary

Count:	6
SUM(Number of Recor...	
Sum:	50
Average:	8.33
Minimum:	1
Maximum:	26
Median:	2.00

Caption

Sum of Number of Records for each Profession.

- We can see in the below plotted graph that there are total 48 records of celebrities and there are 18 different kind of Professions on Famous Birthdays website.



Summary

Count:	18
Measure Values	
Sum:	48.00
Average:	2.67
Minimum:	1.00
Maximum:	14.00
Median:	1.00
SUM(Number of Recor...	
Sum:	48
Average:	2.67
Minimum:	1
Maximum:	14
Median:	1.00

Caption

Number of Records for each Profession.

- From the above 2 graphs we can conclude that IMDb has marked the celebrities with the integral profession whereas Famous Birthdays has marked the profession in segments or sections.

  For example: Malina has been distinguished as an Actress on IMDb whereas specifically a TV actress on Famous Birthdays website.

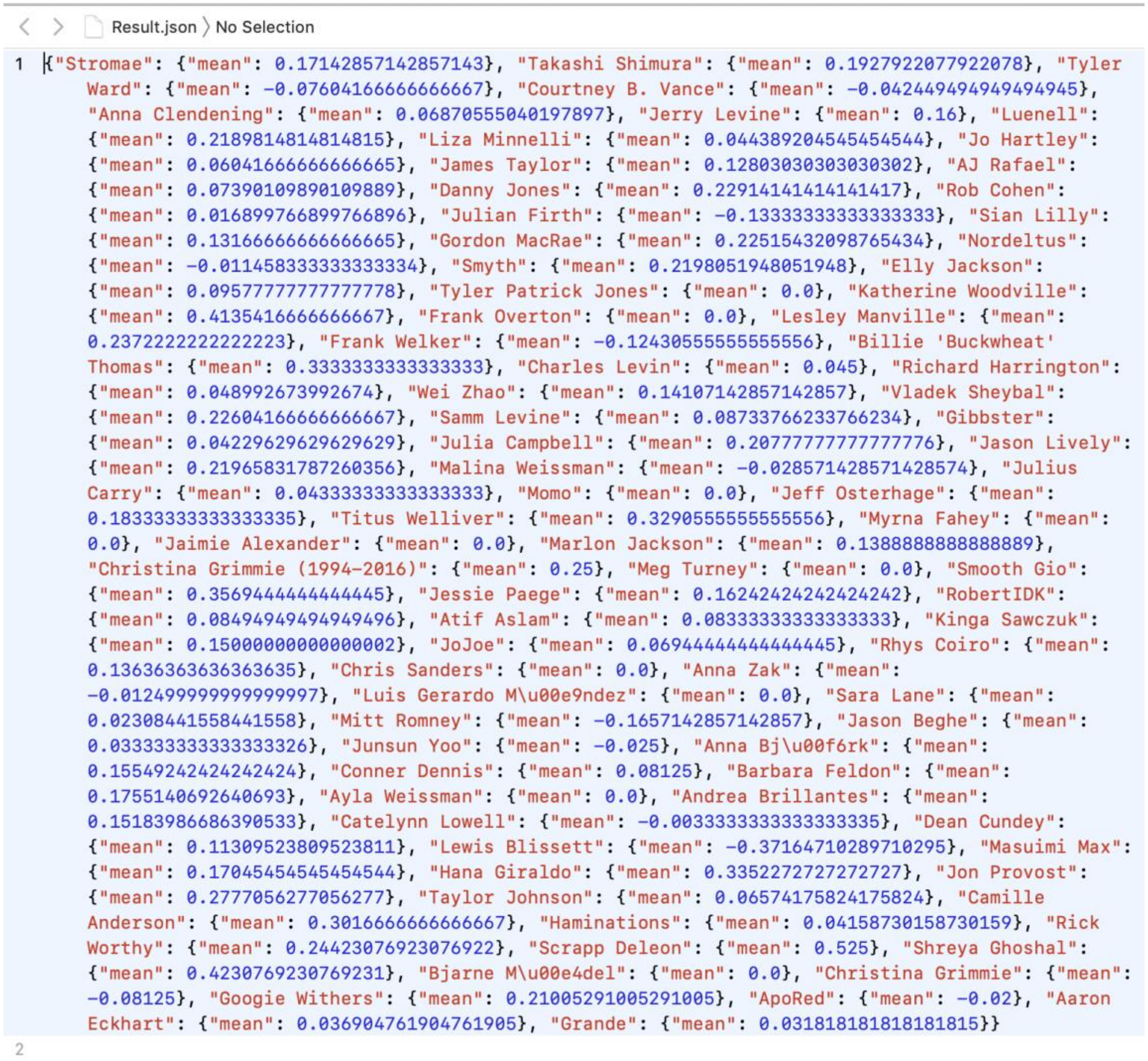
- Also, if a celebrity is in the Music Industry, then that celebrity has been grouped into the Soundtrack category by IMDb whereas Famous birthdays has distinguished them into sub professions such as, world music singer, pop singer, folk singer, guitarist etc.

7.3 TWITTER SENTIMENT ANALYSIS

- The tweets were collected using twitter API, which responses in JSON format, then we required authentication to fetch data from twitter, and need following configuration for it.

```
{
        “consumer_key”: “____”,
        “consumer_secret”: “____”,
        “access_token”: “____”,
        “access_token_secret”: “____”
}
```

- After fetching data from twitter, we ran TextBlob for Sentiment Analysis and calculated the Sentiment polarity.
    - POSITIVE: 0 < polarity <= 1
    - NEGATIVE: polarity = 0
    - NEUTRAL: -1 <= polarity < 0
- The celebrity names along with the sentiment polarities were stored in the json file (further visualization can be done but it is not a part of the project for now).



```
{"Stromae": {"mean": 0.17142857142857143}, "Takashi Shimura": {"mean": 0.1927922077922078}, "Tyler
    Ward": {"mean": -0.07604166666666667}, "Courtney B. Vance": {"mean": -0.042449494949494945},
    "Anna Clendening": {"mean": 0.06870555040197897}, "Jerry Levine": {"mean": 0.16}, "Luenell":
    {"mean": 0.2189814814814815}, "Liza Minnelli": {"mean": 0.044389204545454544}, "Jo Hartley":
    {"mean": 0.06041666666666665}, "James Taylor": {"mean": 0.12803030303030302}, "AJ Rafael":
    {"mean": 0.07390109890109889}, "Danny Jones": {"mean": 0.22914141414141417}, "Rob Cohen":
    {"mean": 0.016899766899766896}, "Julian Firth": {"mean": -0.13333333333333333}, "Sian Lilly":
    {"mean": 0.13166666666666665}, "Gordon MacRae": {"mean": 0.22515432098765434}, "Nordeltus":
    {"mean": -0.011458333333333334}, "Smyth": {"mean": 0.2198051948051948}, "Elly Jackson":
    {"mean": 0.09577777777777778}, "Tyler Patrick Jones": {"mean": 0.0}, "Katherine Woodville":
    {"mean": 0.4135416666666667}, "Frank Overton": {"mean": 0.0}, "Lesley Manville": {"mean":
    0.2372222222222223}, "Frank Welker": {"mean": -0.12430555555555556}, "Billie 'Buckwheat'
    Thomas": {"mean": 0.3333333333333333}, "Charles Levin": {"mean": 0.045}, "Richard Harrington":
    {"mean": 0.048992673992674}, "Wei Zhao": {"mean": 0.14107142857142857}, "Vladek Sheybal":
    {"mean": 0.22604166666666667}, "Samm Levine": {"mean": 0.08733766233766234}, "Gibbster":
    {"mean": 0.04229629629629629}, "Julia Campbell": {"mean": 0.20777777777777776}, "Jason Lively":
    {"mean": 0.21965831787260356}, "Malina Weissman": {"mean": -0.028571428571428574}, "Julius
    Carry": {"mean": 0.04333333333333333}, "Momo": {"mean": 0.0}, "Jeff Osterhage": {"mean":
    0.18333333333333335}, "Titus Welliver": {"mean": 0.3290555555555556}, "Myrna Fahey": {"mean":
    0.0}, "Jaimie Alexander": {"mean": 0.0}, "Marlon Jackson": {"mean": 0.1388888888888889},
    "Christina Grimmie (1994-2016)": {"mean": 0.25}, "Meg Turney": {"mean": 0.0}, "Smooth Gio":
    {"mean": 0.3569444444444445}, "Jessie Paege": {"mean": 0.16242424242424242}, "RobertIDK":
    {"mean": 0.08494949494949496}, "Atif Aslam": {"mean": 0.08333333333333333}, "Kinga Sawczuk":
    {"mean": 0.15000000000000002}, "JoJoe": {"mean": 0.06944444444444445}, "Rhys Coiro": {"mean":
    0.13636363636363635}, "Chris Sanders": {"mean": 0.0}, "Anna Zak": {"mean":
    -0.012499999999999997}, "Luis Gerardo M\u00e9ndez": {"mean": 0.0}, "Sara Lane": {"mean":
    0.02308441558441558}, "Mitt Romney": {"mean": -0.1657142857142857}, "Jason Beghe": {"mean":
    0.033333333333333326}, "Junsun Yoo": {"mean": -0.025}, "Anna Bj\u00f6rk": {"mean":
    0.15549242424242424}, "Conner Dennis": {"mean": 0.08125}, "Barbara Feldon": {"mean":
    0.1755140692640693}, "Ayla Weissman": {"mean": 0.0}, "Andrea Brillantes": {"mean":
    0.15183986686390533}, "Catelynn Lowell": {"mean": -0.0033333333333333335}, "Dean Cundey":
    {"mean": 0.11309523809523811}, "Lewis Blissett": {"mean": -0.37164710289710295}, "Masuimi Max":
    {"mean": 0.17045454545454544}, "Hana Giraldo": {"mean": 0.3352272727272727}, "Jon Provost":
    {"mean": 0.2777056277056277}, "Taylor Johnson": {"mean": 0.06574175824175824}, "Camille
    Anderson": {"mean": 0.3016666666666667}, "Haminations": {"mean": 0.04158730158730159}, "Rick
    Worthy": {"mean": 0.24423076923076922}, "Scrapp Deleon": {"mean": 0.525}, "Shreya Ghoshal":
    {"mean": 0.4230769230769231}, "Bjarne M\u00e4del": {"mean": 0.0}, "Christina Grimmie": {"mean":
    -0.08125}, "Googie Withers": {"mean": 0.21005291005291005}, "ApoRed": {"mean": -0.02}, "Aaron
    Eckhart": {"mean": 0.036904761904761905}, "Grande": {"mean": 0.031818181818181815}}
```

8 REFERENCES

[1] Web Scraping with Python: Collecting Data from the Modern Web by Ryan Mitchell

[2] Web Scraping Data Extraction from websites By: Vojtech Draxl Student Number: 1310256058 Supervisor: DI Dr. Gerd Holweg Wien, 04.02.2018

[3] https://monkeylearn.com/sentiment-analysis/

[4] https://medium.com/the-andela-way/introduction-to-web-scraping-using-selenium-7ec377a8cf72

[5] https://www.promptcloud.com/scraping-webpage-for-sentiment-analysis/

[6] http://fruchter.co/post/53164489086/python-headless-web-browser-scraping-on-amazon

[7] https://textblob.readthedocs.io/en/dev/

[8] https://github.com/maneeshd/Twitter-Sentimental-Analysis

[9] https://github.com/haiderAli62/amazon.com-web-scraping-projects/blob/master/Project%2016%20get%20the%20titles%20of%20the%20100%20best%20seller%20books.py

[10] https://beautiful-soup-4.readthedocs.io/en/latest/

[11] https://blog.hartleybrody.com/web-scraping-cheat-sheet/#using-beautifulsoup

[12] https://medium.com/the-andela-way/introduction-to-web-scraping-using-selenium-7ec377a8cf72