**Introduction**

Sentiment analysis is a kind of natural language processing (NLT). For this project I wanted to perform sentiment analysis on Michelle Obama tweets. Michelle is a lawyer and writer who served as a first lady of the United States of America from 2009 to 2017. Particularly, the former first lady worked as a campaigner for poverty reduction, education, health, nutrition, and exercise, and many more (1) and served as role model for women out there. After her husband's presidency, she has continued to influence and educate the community through writing books, hosting podcasts, using social medias and more. Personally, she is one of my inspirations. As her influence remains high, looking at the sentiment of her tweets might help us to explore and understand her tweets current reflection and thought. Thus, the purpose of this analysis was to explore whether Michelle’s tweet was mostly positive, neutral, or negative.

**Questions**

Are Michelle’s tweet being mostly positive, neutral or negative?

Which words stand out from the sentiment analysis?

**Methods**

The data source and how it was extracted were described below.

First off, I needed to create a tweeter account as I did not have any. Next, I signed up for developer account which was used to get started with Twitter Application Programming Interface (API). Then, I created a new project and connected to an app while signed into the developer account. Next, using the app, API key and token were retrieved (2). That means, API set up was made to get the necessary credentials (**the consumer key**, **consumer secret**, **access key** and **access secret**). Then, I saved those credential in a csv format to keep the credential private.

Jupiter notebook was used to analyze the data. Necessary libraries were installed and imported as mentioned below. tweepy, textblob, pandas, numpy, matplotlib and re (regular expression) were imported.

To work with word cloud, I have installed wordcloud and imported the necessary libraries like WordCloud, STOPWORDS, ImageColorGenerator, and get stopwords from wordcloud library using stopwords = set(STOPWORDS)

Next, the keys and access tokens were imported using **read\_csv()**. Then, all individual keys were retrieved and assigned with different variable name.

I used **Tweepy** which is one of python library for getting tweets mentioning using the Tweeter API (3,4). This help to authenticate and set the token.

Next, a new variable was assigned to store the tweets extracted from tweeter in a list. To proceed, screen\_name was set to be '**@MichelleObama'** and only 1000 tweets were extracted using Tweepy Curser. Then, selected columns were extracted and appended to the empty list created earlier. Column names were assigned and columns were separated using **split()** method. Next, the data were stored in a data frame to work with python pandas. The first five observations were viewed using **head()** method and **tail()** method was used to view the last five observations.

Data were cleaned using **regular expression (re)**. all the mentions, hashtags, retweets, URLs were removed using regular expressions. Column having null tweets were dropped using **drop** method. Presence of missing data were checked using **isna()** method.

**textBlob** is a python library used to work with text data type. It was used to get the polarity and subjectivity of a text (5).

Text analysis was done to get the sentiment. The polarity score is a float within the range [-1.0, 1.0] whereas the subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective (5).

Polarity score was then categorized into three categories. value less than zero was named as negative sentiment, value equal to zero as neutral and value greater than zero as positive sentiment.

Then, the percentage of each sentiment were calculated.

In order to get insight, I have also visualized the data using matplotlib Pyplot. Pie chart and Bar graphs were used to visualize the sentiment analysis whereas word cloud was used to visualize the most used words from the sentiment analysis. The Jupiter notebook of this report is available in my GitHub repo (<https://github.com/Bitiyaa/DataAnalysisProjects>).

**Results**

Sentiment analysis result of Michelle’s tweets revealed that among the selected 1000 tweets majority of them were positive. The count of the number of tweets by sentiments is shown in figure below (**Figure 1**).

Chart, bar chart

Description automatically generated

Fig 1: Count of Michelle’s tweet for each sentiment

As expected, most of Michelle’s tweets were positive (76.5 %) followed by neutral (17.1 %) and negative (6.4%). This finding is shown in pie chart below (**Figure 2**).

Chart

Description automatically generated

Fig 2: Percentage of Michelle’s tweet Sentiment Analysis

On the other hand, I have used word cloud to identify words that stand out from the sentiment analysis. Based on the word cloud analysis, some of the most mentioned words were vote, today, thank, make, hope, year, one, love, world, know, time, share, work, proud and people. This finding is shown in **Figure 3** below.

Text

Description automatically generated

Fig 3: Total reviews of word count of Michelle’s tweet

**Conclusion**

Most of Michelle’s tweets were positive. Vote, love, today, thank, make, world and hope were some of the words stand out in her tweets.

**References**

1. About-the-white-house/first-families/michelle-obama .<https://www.whitehouse.gov/about-the-white-house/first-families/michelle-obama/>
2. Step-by-step-guide-to-making-your-first-request-to-the-twitter-api-v2. <https://developer.twitter.com/en/docs/tutorials/step-by-step-guide-to-making-your-first-request-to-the-twitter-api-v2>
3. Extraction-of-tweets-using-tweepy.<https://www.geeksforgeeks.org/extraction-of-tweets-using-tweepy/>
4. beginners-guide-to-tweepy.<https://www.loginradius.com/blog/engineering/beginners-guide-to-tweepy/>
5. textblob documentation <https://buildmedia.readthedocs.org/media/pdf/textblob/dev/textblob.pdf>.