DATA ANALYTICS AND VISUALIZATION

DATA WRANGLING ASSESSMENT

Brayden O' Neill

K00224576

15/NOV/2021

Game Design and Development - BSc (Hons)

Sentiment Analysis

What is Sentiment Analysis?

The concept of sentiment analysis involves "gathering natural language processing, text analysis, computational linguistics, and biometrics, and using the data collected to identify, extract, quantify, and study affective states and subjective information". This information can then be used to categorize the opinions expressed and determine their polarity (positive, negative, neutral).

Sentiment analysis is a very powerful tool when utilized for business or marketing. It is possible for businesses to understand the sentiment of customers, what is trending and what conversations they are having. This is done to quantify attitudes, opinions, and emotions related to the business. For this reason, sentiment analysis is sometimes referred to as opinion mining. With Twitter, sentiment analysis is achieved by collecting a range of tweets associated to a similar topic, computing the sentiment polarity and strong points of each tweet, and finally, combining the results of such tweets.

How is sentiment analysis performed?

Sentiment analysis can be achieved with natural language processing and machine learning. Messages are broken down into sections of specific topics, where each topic is assigned a sentiment score. Assume the following twitter post was made: *"Yesterday I went to Starbucks for the first time, and I really enjoyed it. The coffee is so amazing! Their cookies were not the best though."* A sentiment analysis tool would gather the text taken from the post and divide it into sections, assigning scores to each topic. For example:

- Starbucks…really enjoyed = +2
- Coffee…amazing = +4
- Cookies…not the best = -2

Public opinion can be monitored in real-time as a new marketing campaign, product, or service is in process. Using the above example, imagine Starbucks were experimenting with a new cookie recipe. Processing tweets such as the one above could give real-time feedback after launching new products and used to track public opinion, allowing for automated/accurate market research to be conducted.

Using Sentiment Analysis Libraries

Sentiment Analysis can be explored in Python using libraries such as Textblob and VADER. These libraries can be combined with other tools, such as pandas, for accurate data analysis. Textblob is an open-source python library used to process textual data. It performs operations such as noun phrase extraction, classification, translation, etc. VADER is a tool that is specifically tailored to sentiments expressed on social media, such as Twitter posts. VADER can be used to give a Positivity/ Negativity score and can be used to portray how positive or negative the sentiment is.

These libraries can be utilized to expand upon the Data Wrangling Assignment (Figures 1,2,3) to sort data based on polarity. The analyzers mentioned above use real numbers between 1 and -1 to represent the polarity of the sentiment using key words found in the tweet's text field ('1' meaning positive feedback and '-1' meaning negative feedback). It could be possible to iterate through the collected tweets and run them through the analyzers, adding a new polarity score to their respective row in the CSV file. From here, collected data could be sorted based on polarity score in Figure 3.

```
Enter word or hashtag to search: navi
Enter total tweets to search: 200
Searching..
[[datetime.datetime(2021, 11, 16, 9, 44, 19, tzinfo=datetime.timezone.utc), b'\xe6\x9c\x80\xe8\xbf\x91i
Pad\xe3\x82\x92\xe4\xbd\xbf\xe3\x81\xa3\xe3\x81\xa6\xe3\x81\x84\xe3\x82\x8b\xe3\x81\xae\xe3\x81\xa7\xe3
\x80\x81\xe6\xb5\x85\xe8\x8d\x89\xe4\xb9\x9d\xe5\x80\xb6\xe6\xa5\xbd\xe9\x83\xa8HOTEL\xe3\x81\xa7\xe9\x
a3\x9f\xe3\x81\xb9\xe3\x81\x9f\xe6\x9c\x9d\xe9\xa3\x9f\xe5\x86\x99\xe7\x9c\x9f\xe3\x81\xab\xe8\x90\xbd\
xe6\x9b\xb8\xe3\x81\x8d\xe3\x81\x97\xe3\x81\xa6\xe3\x81\xbf\xe3\x81\x9f\xe2\x80\xa6\xef\xbc\x81\n\n\xe3
\x81\x8a\xe8\x8c\xb6\xe6\xbc\xac\xe3\x81\x91\xe3\x82\x92\xe9\xa3\x9f\xe3\x81\xb9\xe3\x81\x9f\xe3\x81\x8
4\xe3\x81\xa8\xe6\x80\x9d\xe3\x81\x86\xe3\x81\x93\xe3\x81\xa8\xe3\x81\x8c\xe6\xbb\x85\xe5\xa4\x9a\xe3\x
81\xab\xe3\x81\xaa\xe3\x81\x84\xe3\x82\x93\xe3\x81\xa0\xe3\x81\x91\xe3\x82\x93\xe3\x81\x80\xe3\x81\x80\x9b\
xe3\x83\x86\xe3\x83\xab\xe3\x81\xae\xe9\x9b\xb0\xe5\x9b\xb2\xe6\xb0\x97\xe7\x9a\x84\xe3\x81\xab\xe3\x82
\x82\xe5\x92\x8c\xe3\x81\xa3\xe3\x81\xbd\xe3\x81\x84\xe9\xa3\x9f\xe4\xba\x8b\xe3\x81\x8c\xe3\x81\x97\xe
3\x81\x9f\xe3\x81\x8f\xe3\x81\xa6\xe3\x80\x82\xe3\x81\x84\xe3\x82\x8d\xe3\x82\x93\xe3\x81\xaa\xe3\x81\x
8a\xe3\x81\x8b\xe3\x81\x9a\xe3\x81\x8c\xe7\x9b\x9b\xe3\x82\x89\xe3\x82\x8c\xe3\x81\xa6\xe3\x81\xa6\xe8\
x89\xb2\xe3\x81\xa8\xe3\x82\x8a\xe3\x81\xa9\xe3\x82\x8a\xe3\x81\xa7\xe5\x8f\xaf\xe6\x84\x9b\xe3\x81\x8b
\xe2\x80\xa6 https://t.co/wFHTJwI2xn', 0, 0], [datetime.datetime(2021, 11, 16, 9, 44, 19, tzinfo=dateti
me.timezone.utc), b'\xe3\x81\xb2\xe3\x82\x87\xe3\x81\x88\xef\xbc\x9f\xef\xbc\x81\xe3\x81\xa3\xe3\x81\xa
6\xe5\xa3\xb0\xe3\x81\x8c\xe3\x81\xa3', 0, 0], [datetime.datetime(2021, 11, 16, 9, 44, 7, tzinfo=dateti
me.timezone.utc), b'@EpihinS @natusvincere \xd0\x92\xd0\xbe\xd1\x82 \xd1\x82\xd1\x8b \xd1\x81\xd0\xbc\x
d0\xb5\xd0\xb5\xd1\x88\xd1\x8c\xd1\x81\xd1\x8f, \xd0\xb0 \xd1\x82\xd1\x83\xd1\x82 https://t.co/fYeEQcpY
if', 0, 0], [datetime.datetime(2021, 11, 16, 9, 44, 3, tzinfo=datetime.timezone.utc), b'MacGeneration :
Navi, une app de traduction temp r\xc3\xa9el avec SharePlay #iOS https://t.co/hQptafkvKA', 0, 0], [date
time.datetime(2021, 11, 16, 9, 44, 3, tzinfo=datetime.timezone.utc), b'MacGeneration : Navi, une app de
traduction temp r\xc3\xa9el avec SharePlay #mac https://t.co/i2sUFxSuoJ', 0, 0], [datetime.datetime(202
1, 11, 16, 9, 44, 2, tzinfo=datetime.timezone.utc), b'\xe5\x98\x98\xe5\x81\xbd\xe3\x82\x8a\xe3\x81\xaa\
xe3\x81\x84\xe4\xb8\x80\xe7\x95\xaa\xe5\xae\x89\xe3\x81\x84\xe6\xa0\xbc\xe5\xae\x89SIM(MVNO)\xe6\x96\x9
9\xe9\x87\x91\xe3\x83\xa9\xe3\x83\xb3\xe3\x82\xad\xe3\x83\xb3\xe3\x82\xb0[03112118] https://t.co/6Wn9Mq
gnEU \xe2\x80\xa6 #\xe6\xa0\xbc\xe5\xae\x89SIM #\xe6\xa0\xbc\xe5\xae\x89\xe3\x82\xb9\xe3\x83\x9e\xe3\x8
3\x9b #MVNO #\xe6\x96\x99\xe9\x87\x91\xe3\x83\x97\xe3\x83\xa9\xe3\x83\xb3 #sim_navi', 0, 0], [datetime.
datetime(2021, 11, 16, 9, 43, 55, tzinfo=datetime.timezone.utc), b'\xe3\x80\x90#\xe5\xae\x9a\xe6\x9c\x9
f\xe3\x80\x91#\xe6\x8b\xa1\xe6\x95\xa3\xe5\xb8\x8c\xe6\x9c\x9b\n\xe5\xa5\xb3\xe6\x80\xa7\xe3\x81\xab\xe
4\xba\xba\xe6\xb0\x97\xe3\x81\xae\xe3\x80\x8c\xe9\x85\xb8\xe7\xb4\xa0\xe3\x83\x80\xe3\x82\xa4\xe3\x82\x
a8\xe3\x83\x83\xe3\x83\x88\xe6\xaf\x94\xe8\xbc\x83\xe3\x83\x8a\xe3\x83\x93\xe3\x80\x8d(https://t.co/QFi
vnyWZrP)\xe3\x82\x92\xe3\x81\x94\xe7\xb4\xb9\xe4\xbb\x8b\xe3\x81\x97\xe3\x81\xbe\xe3\x81\x99\xe3\x80\x8
2\n\n#\xe3\x83\x80\xe3\x82\xa4\xe3\x82\xa8\xe3\x83\x83\xe3\x83\x88 #\xe9\x85\xb5\xe7\xb4\xa0\xe3\x80\x8
0#\xe7\x97\xa9\xe3\x81\x9b\xe3\x82\x8b\xe3\x80\x80#\xe6\xa5\xbd\xe3\x81\x97\xe3\x81\xa6\xe7\x97\xa9\xe3
\x81\x9b\xe3\x82\x8b #\xe5\xa5\xb3\xe5\xad\x90\xe3\x80\x80#\xe4\xbe\xbf\xe7\xa7\x98\xe8\xa7\xa3\xe6\xb6
\x88\n#\xe9\x80\x9a\xe8\xb2\xa9 #\xe4\xba\x8c\xe3\x81\xae\xe8\x85\x95 #\xe3\x81\x8a\xe8\x85\xb9 #\xe4\x
bd\x93\xe9\x87\x8d', 0, 0], [datetime.datetime(2021, 11, 16, 9, 43, 43, tzinfo=datetime.timezone.utc),
b'https://t.co/JAi2WCrl8n\n\xe3\x81\x9d\xe3\x81\xae\xe7\x94\x9f\xe6\xb4\xbb\xe8\xbe\x9b\xe3\x81\x8f\xe3
\x81\xaa\xe3\x81\x84\xe3\x81\xa7\xe3\x81\x99\xe3\x81\x8b\xef\xbc\x9f\n\xe5\xa4\xa7\xe5\xb9\x85\xe3\x81\
xab\xe6\xb8\x9b\xe9\xa1\x8d\xe3\x81\x97\xe3\x81\x9f\xe3\x81\x84\xe6\x96\xb9\xe3\x81\xaf\xe3\x81\x93\xe3
\x81\xae\xe3\x82\xa2\xe3\x82\xab\xe3\x82\xa6\xe3\x83\xb3\xe3\x83\x88\xe3\x81\xa7\xe2\x98\x80\n\n\xe2\xa
c\x87\xe3\x80\x80\xe2\xac\x87\xe3\x80\x80\xe2\xac\x87\n@saimu_navi\n\n\xe2\x9c\xb4\xe2\x9c\xb4\n\xe2\x9
c\xb1\n\xe2\x9a\xa0', 0, 0], [datetime.datetime(2021, 11, 16, 9, 43, 37, tzinfo=datetime.timezone.utc),
b'\xe6\x9c\xa0\xe6\x9d\x91\xe6\x9b\x93\xe5\x92\x80 \xe6\x97\xa5\xe7\x94\xa3 CM \xe3\x82\x84\xe3\x81\xa3
```

Figure 1 – Searching for tweets containing the given word or hashtag

A1	▼	⋮	✕ ✓ ƒx	Date

	A	B	C	D	E	F	G	H	I	J
1	Date	Text	Favorites	Retweets						
2										
3	2021-11-1	b'\xe6\x9c	0	0						
4										
5	2021-11-1	b'\xe3\x81	0	0						
6										
7	2021-11-1	b'@Epihin!	0	0						
8										
9	2021-11-1	b'MacGen	0	0						
10										
11	2021-11-1	b'MacGen	0	0						
12										
13	2021-11-1	b'\xe5\x98	0	0						
14										
15	2021-11-1	b'\xe3\x80	0	0						
16										
17	2021-11-1	b'https://t	0	0						
18										
19	2021-11-1	b'\xe6\x9c	0	0						
20										
21	2021-11-1	b'\xe3\x80	0	1						
22										
23	2021-11-1	b'Navi Hea	1	0						
24										
25	2021-11-1	b'\xe7\xa6	0	0						
26										

navi_tweets   ⊕

Figure 2 – The CSV file created after searching for a specific word or hashtag

```
Enter filename to open: navi_tweets.csv

Select an option to sort the data:
    1 - Date (ascending)
    2 - Date (descending)
    3 - Favorites (ascending)
    4 - Favorites (descending)
    5 - Retweets (ascending)
    6 - Retweets (descending)
    7 - Exit
4
                        Date                                                     Text  Favorites  Retweets
129  2021-11-16 09:30:02+00:00  b'\xe8\x80\xb3\xe3\x81\xae\xe5\x8b\x95\xe3\x81...         40        10
117  2021-11-16 09:30:58+00:00  b'Good Morning\n#karlnapwastaken #dreamkarlnap...         18         5
88   2021-11-16 09:34:44+00:00  b'\xe4\xbb\x8a\xe6\x97\xa5\xe3\x82\x82\xe3\x81...         10         0
187  2021-11-16 09:22:38+00:00  b'@NIPCS if you sign es3tag i will go support ...          6         0
131  2021-11-16 09:30:00+00:00  b'11\xe6\x9c\x88\xe3\x82\x82\xe3\x80\x8c\xe3\x...          5         3
41   2021-11-16 09:40:01+00:00  b'\xea\xb8\xb8\xeb\xb2\xa0 \xeb\xb0\x94\xeb\x8...          5         0
141  2021-11-16 09:28:43+00:00  b'\xe7\xab\xb9\xe5\xb3\xb6\xe5\xae\x8f\xe3\x81...          4         1
69   2021-11-16 09:36:36+00:00                        b'https://t.co/T2IBHL2R2k'          4         0
67   2021-11-16 09:36:37+00:00  b'\xec\xb5\x9c\xec\x84\xa0\xec\x9d\x84 \xeb\x8...          3         0
167  2021-11-16 09:25:58+00:00  b'\xec\x97\x98\xeb\xa0\x88\xeb\x85\xb8\xec\x96...          3         0
110  2021-11-16 09:31:30+00:00  b'@YuzukiMuroi \xe8\xbb\xbd\xe8\x87\xaa\xe5\x8...          2         0
165  2021-11-16 09:26:34+00:00  b'@CreacionesNavi A quienes veais este mensaje...          2         1
96   2021-11-16 09:33:37+00:00  b'\xe3\x82\xe3\x81\x81\xe7\xa2\xba\xe3\x81...          2         1
77   2021-11-16 09:35:50+00:00  b'\xe6\x98\xa8\xe6\x97\xa5\xe3\x81\xb0\xe3\x83...          1         0
72   2021-11-16 09:36:26+00:00                        b'https://t.co/T2IBHL2R2k'          1         0
172  2021-11-16 09:25:07+00:00  b'\xe9\x81\x8b\xe5\x96\xb6\xe3\x81\xa3\xe3\x81...          1         0
145  2021-11-16 09:28:11+00:00  b'#tuesdaymotivations D digital loan lending c...          1         2
176  2021-11-16 09:24:41+00:00  b'Navi da guerra d\xe2\x80\x99Israele si eserc...          1         1
136  2021-11-16 09:29:01+00:00  b'@saoodsangle30 @CheemsCricket We want the sk...          1         0
120  2021-11-16 09:30:45+00:00  b'\xe3\x80\x90\xe6\x84\x9b\xe5\xa6\xbb\xe3\x81...          1         0
111  2021-11-16 09:31:30+00:00          b'@realillumi_navi Best good morning ever'          1         0
108  2021-11-16 09:31:36+00:00  b'@smdq_1114 \xe3\x81\x88\xe3\x82\x89\xe3\x81\...          1         0
183  2021-11-16 09:23:05+00:00  b'@AR__Ao3o (\xec\x95\x84\xec\xb0\x8a!) \xeb\x...          1         0
17   2021-11-16 09:42:49+00:00  b'\xe3\x80\x90\xe8\xa6\x8f\xe6\xa0\xbc\xe5\xa4...          1         0
10   2021-11-16 09:43:33+00:00  b'Navi Health Insurance \xe0\xa4\xa8\xe0\xa4\x...          1         0
190  2021-11-16 09:21:35+00:00             b'@NoFlox_ @Riafnbr \xf0\x9f\x94\xa5'          1         0
109  2021-11-16 09:31:32+00:00  b'\xe5\x8f\xa4\xe4\xbb\xa3\xef\xbd\x9e\xe8\xbf...          1         1
196  2021-11-16 09:21:04+00:00  b'\xe7\xa5\x9e\xe3\x82\xad\xad\xe3\x83\xab\xe9\x9b...          1         0
95   2021-11-16 09:33:54+00:00  b'\xe3\x80\x90XM\xe3\x83\x9d\xe3\x82\xa4\xe3\x...          1         0
137  2021-11-16 09:28:59+00:00  b'@freedom_vv_navi \xe9\x81\x8b\xe8\xbb\xa2\xe...          0         0
132  2021-11-16 09:29:57+00:00  b'RT @chariloto_PR: \xe3\x80\x90Perfecta Navi\...          0         2
142  2021-11-16 09:28:37+00:00  b'Tesla Model 3 LONG RANGE | 4% BIJTELLING | N...          0         0
143  2021-11-16 09:28:26+00:00  b'\xe5\xae\x89\xe5\x85\xa8\xe3\x81\xab\xe7\x9f...          0         0
134  2021-11-16 09:29:03+00:00  b'\xe3\x80\x90\xe6\x95\x99\xe8\x82\xb2\xe7\x86...          0         0
133  2021-11-16 09:29:15+00:00  b'\xe3\x83\x81\xe3\x83\x93\xe3\x81\xaa\xe3\x82...          0         0
```

Figure 3 - Reading the data from the CSV and sorting the data

References

1. En.wikipedia.org. 2021. *Sentiment analysis - Wikipedia*. [online] Available at: <https://en.wikipedia.org/wiki/Sentiment_analysis> [Accessed 15 November 2021].

2. Medium. 2021. *Two Sentiment Analysis Libraries and How They Perform*. [online] Available at: <https://towardsdatascience.com/two-sentiment-analysis-libraries-and-how-they-perform-3de4a06342ec> [Accessed 15 November 2021].

3. Youtube.com. 2021. [online] Available at: <https://www.youtube.com/watch?v=seNEPVW7NPA> [Accessed 15 November 2021].

4. Sharma, H., 2021. *Let's Learn TextBlob Quickstart - A Python Library For Processing Textual Data*. [online] Analytics India Magazine. Available at: <https://analyticsindiamag.com/lets-learn-textblob-quickstart-a-python-library-for-processing-textual-data/> [Accessed 15 November 2021].

5. GeeksforGeeks. 2021. *Python | Sentiment Analysis using VADER - GeeksforGeeks*. [online] Available at: <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/> [Accessed 15 November 2021].