ZIMING PING (ZP584)
TINGYU YANG (TY2069)
XINYU ZHAO (XZ2671)

SENTIMENT ANALYSIS

--THREE CHOSEN AUTOMOBILE COMPANIES

INTRODUCTION

BASIC IDEA

- NLP-TEXTBLOB
- Public attitudes towards 3 famous automobile companies in the USA
- Twitter API

INTRODUCTION & NECESSARY STEPS

- Clean up the raw data
- NLP model to calculate the polarity and subjectivity of each meaningful word (TextBlob)
- define a kind of metric to calculate the sentiment scores
- every Positive tweet +2 points; every Neutral tweet +0.5 point; every Negative tweet -2 points
- create a word cloud

TASK PERFORMANCE PROCESS

- Selection of the top 3 automobile companies
- Mazda USA
- Buick
- Ford

TASK PERFORMANCE PROCESS

```
Enter the Automobile Company you want to analyze: >? Buick
Number of tweets you want to analyze(No more than 400 so that you can search for last month): >? 350

Show recent Buick tweets with full text:

                                                Tweets  Subjectivity  \
0    there. team discuss enclave assist can. please...      0.888889
1    thank bringing attention. content contrary gm’...      0.250000
2    hi, maddie. team happy help find right lacross...      0.767857
```

TASK PERFORMANCE PROCESS

```python
def clean_data(text):
    text = re.sub(r'https?:\/\/\S+', '', text)
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = re.sub(r'^\s+|\s+$', '', text)
    text = re.sub(r'#', '',text)
    text = re.sub(r'RT[\s]+', '',text)
    text = re.sub(r'RT[\s]+','',text)
    words = text.split(" ")
    words = [w for w in words if len(w) > 2]  # ignore a, an, be, ...
    words = [w.lower() for w in words]
    words = [w for w in words if w not in STOPWORDS]
    clean_word = ' '.join(words)
    return clean_word
```

TASK PERFORMANCE PROCESS

- importing TextBlob package
- add the corresponding subjectivity and polarity into the previous DataFrame and display the Scatter diagram.
- set the scoring rule that Positive—gain 2 points, Neutral-gain 0.5 point and Negative lose 2 points in order to calculate the total scores of the specific company

TASK PERFORMANCE PROCESS

- generate the relevant graph regarding to the score results
- generate the graph of scores of this company

TASK PERFORMANCE PROCESS-WORD CLOUD

```
#Create the Wordcloud
final_word = ','.join([twts for twts in df['Tweets']])
wcloud_3 = WordCloud(height=500, width=400,random_state=21,background_color="black", max_words=500, max_font_size=120)
wcloud_3.generate(final_word)

#Show image of Wordcloud
f = plt.figure(figsize=(50,50))
plt.imshow(wcloud_3, interpolation='bilinear')
plt.title('Twitter Generated Cloud', size=170)
plt.axis("off")
plt.show()
```

SCORES, GRAPHS WORD CLOUD

MAZDA USA—451.5 POINTS

```
There are  231  positive results.
There are  91  neutral results.
There are  28  negative results.

The Positive Score:  462
The Neutral Score:  45.5
The Negative Score:  -56
The Final Sentiment Score of this automobile company is:  451.5

The Positive Percentage:  66.0 %
The Neutral Percentage:  26.0 %
The Negative Percentage:  8.0 %
```

happy
road
excited
ride
thank
never
always
drive
way
mazda
hope
still
love
come
made
nice
best
make
one
time
soon
better assist
recommend contacting
everything
driver
journey
together
customer experience
experience center
contacting customer
help steer
right direction
help provide
thanks steering
pleasure
mile
see soon
good
back
see
proud
place
allow
ready
right
center better
honored
allow help
first
around
nice work
provide safe
glad see
pick
fun
join journey
working tirelessly
assist you
put
sometimes
healthcare heroes
steer direction
long haul
may ended
weekend
well
glad
start
help
service
got
take
do
thing
turn
know
begin
communities
steer
suv
fast
long
clean fun
good clean
along nonetheless
financial services
contacting customer
everything
learn
allow
place
honored
allow help
center better
driver
proud
journey
made
come
nice
ready
right
amp
give
virtual
feeling
local
social
desire
another
want
grow fast
us
mickey done
move
you reached
right companion
may compact
giving
understand
nice work
slick
forever hope
rejoin road
sorry near
nope rejoin
pick
call
much
word
front lines
direction
steering right

SCORES, GRAPHS WORD CLOUD

BUICK—487.5 POINTS

```
There are  230  positive results.
There are  107  neutral results.
There are  13  negative results.

The Positive Score:  460
The Neutral Score:  53.5
The Negative Score:  -26
The Final Sentiment Score of this automobile company is:  487.5

The Positive Percentage:  66.0 %
The Neutral Percentage:  31.0 %
The Negative Percentage:  4.0 %
```

SCORES, GRAPHS WORD CLOUD

FORD—333.5 POINTS

```
There are  199  positive results.
There are  95  neutral results.
There are  56  negative results.

The Positive Score:  398
The Neutral Score:  47.5
The Negative Score:  -112
The Final Sentiment Score of this automobile company is:  333.5

The Positive Percentage:  56.99999999999999 %
The Neutral Percentage:  27.0 %
The Negative Percentage:  16.0 %
```

ANALYSIS & CONCLUSION

- Sensitivity refers to the ability to designate each tweet in the samples with as positive.
- A higher sensitive result means that this tweet tends to become more positive compared to other samples showing a lower sensitive number.

ANALYSIS & CONCLUSION

- Our scoring rules basically depends on the results of polarity
- If the company comes out more positive results and less negative results and it can gain more scores.

COMPANY'S STRATEGY ANALYSIS:

- Buick and Mazda show a much better sentiment figure than Ford, at least on Twitter

  successfully build up their reputation by offering good services

- Ford usually brings more issues or bad experience to its customers

  improve their customer service and build a much better population to the public.

REFERENCE

- *Twitter API Documentation*. (2010). Docs | Twitter Developer. https://developer.twitter.com/en/docs/twitter-api
- Jain, S. (2020, December 8). *Natural Language Processing for Beginners: Using TextBlob*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/
- Hsu, S. (2018, September 11). *Introduction to Data Science: Custom Twitter Word Clouds*. Medium. https://medium.com/@shsu14/introduction-to-data-science-custom-twitter-word-clouds-704ec5538f46

THANK YOU

ZIMING PING (ZP584)

TINGYU YANG (TY2069)

XINYU ZHAO (XZ2671)