Disney_EDA_Report

June 13, 2023

1 Disney Review Topic Modeling EDA Report

Author: Amelia Tang

1.1 Read Python Packages

[1]:
```
import numpy as np
import pandas as pd
import altair as alt
import string
import nltk
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk.corpus import wordnet
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from wordcloud import WordCloud, STOPWORDS
import tkinter
import matplotlib
import matplotlib.pyplot as plt
from PIL import Image
from textblob import TextBlob
from nltk.sentiment.vader import SentimentIntensityAnalyzer

alt.renderers.enable("html")
alt.data_transformers.disable_max_rows()
matplotlib.use('TkAgg')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('omw-1.4')
nltk.download('punkt')
nltk.download('brown')
nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/amelia/nltk_data…
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /Users/amelia/nltk_data…
[nltk_data]   Package wordnet is already up-to-date!
```

```
[nltk_data] Downloading package omw-1.4 to /Users/amelia/nltk_data…
[nltk_data]   Package omw-1.4 is already up-to-date!
[nltk_data] Downloading package punkt to /Users/amelia/nltk_data…
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package brown to /Users/amelia/nltk_data…
[nltk_data]   Package brown is already up-to-date!
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     /Users/amelia/nltk_data…
[nltk_data]   Package vader_lexicon is already up-to-date!
```

[1]: True

1.2 Read the data

1.2.1 There are 42,656 lines of non-null data. Among the three branches of Disney, the California branch has the most review data.

[2]:
```
disney_df = pd.read_csv("../data/raw/DisneylandReviews.csv")
disney_df
```

[2]:

	Review_ID	Rating	Year_Month	Reviewer_Location	\
0	670772142	4	2019-4	Australia	
1	670682799	4	2019-5	Philippines	
2	670623270	4	2019-4	United Arab Emirates	
3	670607911	4	2019-4	Australia	
4	670607296	4	2019-4	United Kingdom	
…	…	…	…	…	
42651	1765031	5	missing	United Kingdom	
42652	1659553	5	missing	Canada	
42653	1645894	5	missing	South Africa	
42654	1618637	4	missing	United States	
42655	1536786	4	missing	United Kingdom	

	Review_Text	Branch
0	If you've ever been to Disneyland anywhere you…	Disneyland_HongKong
1	Its been a while since d last time we visit HK…	Disneyland_HongKong
2	Thanks God it wasn t too hot or too humid wh…	Disneyland_HongKong
3	HK Disneyland is a great compact park. Unfortu…	Disneyland_HongKong
4	the location is not in the city, took around 1…	Disneyland_HongKong
…	…	…
42651	i went to disneyland paris in july 03 and thou…	Disneyland_Paris
42652	2 adults and 1 child of 11 visited Disneyland …	Disneyland_Paris
42653	My eleven year old daughter and myself went to…	Disneyland_Paris
42654	This hotel, part of the Disneyland Paris compl…	Disneyland_Paris
42655	I went to the Disneyparis resort, in 1996, wit…	Disneyland_Paris

[42656 rows x 6 columns]

[3]:
```
disney_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42656 entries, 0 to 42655
Data columns (total 6 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   Review_ID          42656 non-null  int64 
 1   Rating             42656 non-null  int64 
 2   Year_Month         42656 non-null  object
 3   Reviewer_Location  42656 non-null  object
 4   Review_Text        42656 non-null  object
 5   Branch             42656 non-null  object
dtypes: int64(2), object(4)
memory usage: 2.0+ MB
```

We collected the most data on the reviews for the California branch.

[4]:
```
alt.Chart(disney_df).mark_bar().encode(
    x='count()',
    y=alt.Y('Branch', sort='-x'))
```

[4]:
```
alt.Chart(…)
```

1.2.2 Here are some sample review texts:

[5]:
```
disney_df['Review_Text'][89]
```

[5]:
```
'Hong Kong Disneyland is definitely a must see. It is a smaller park but what it
loses in size it makes up for in quality. Hong Kong is currently home to the
only Marvel themed rides and they do not disappoint. Hong Kong has many of the
standard rides shows you expect from Disney but it also maintains a variety of
exclusives. Hong Kong also has a lot of exclusive merchandise especially pins.
Also most if not all of the rides are Bilingual. '
```

[6]:
```
disney_df['Review_Text'].head(10)
```

[6]:
```
0    If you've ever been to Disneyland anywhere you…
1    Its been a while since d last time we visit HK…
2    Thanks God it wasn   t too hot or too humid wh…
3    HK Disneyland is a great compact park. Unfortu…
4    the location is not in the city, took around 1…
5    Have been to Disney World, Disneyland Anaheim …
6    Great place! Your day will go by and you won't…
7    Think of it as an intro to Disney magic for th…
8    Feel so let down with this place,the Disneylan…
9    I can go on talking about Disneyland. Whatever…
Name: Review_Text, dtype: object
```

1.3 Preprocess corpus

1.3.1 To clean the corpus (Disney review texts) for analysis, I performed the following tasks:

- Changed all the letters to lowercase.
- Ensured consistent spelling of "hongkong".
- Replaced the abbreviation "n't" with "not".
- Removed all digits from the corpus.

[7]:
```
corpus=''
for i in disney_df['Review_Text']:
    corpus+=str(i)
```

[8]:
```
def preprocess(corpus):
    corpus = corpus.lower()
    corpus = corpus.replace('hk', 'hongkong')
    corpus = corpus.replace('hong kong', 'hongkong')
    corpus = corpus.replace("n't", " not")
    corpus = ''.join([i for i in corpus if not i.isdigit()])
    new_corpus =''
    for i in corpus:
        if i not in string.punctuation:
            new_corpus+=i
    return new_corpus
```

1.3.2 Lemmatization and Stemming

- Stemming is a crude method for cataloging related words, such as "boat," "boats," and "boating." One of the most popular stemming tools is Porter's Algorithm, developed by Martin Porter in 1980 in his paper titled "An algorithm for suffix stripping." The algorithm employs five phases of word reduction (Porter, 1980).

- Lemmatization looks beyond word reduction and considers the vocabulary, part of speech, and its use in a sentence.

[9]:
```
def lemmatize_stem(corpus):
    lemmatizer = WordNetLemmatizer()
    stemmer = PorterStemmer()
    new_corpus = lemmatizer.lemmatize(corpus)
    new_corpus = stemmer.stem(new_corpus)
    return new_corpus
```

1.3.3 Remove stopwords

Stopwords refer to commonly used words in a language that are often removed from text during natural language processing (NLP) tasks. These words are considered to have little or no significant meaning and are typically filtered out to focus on the more important and relevant words in a document or corpus. Examples of stopwords in English include "the," "is," "and," "in," "a," and

"an." The removal of stopwords helps reduce noise and improve the efficiency and effectiveness of NLP tasks such as text classification, information retrieval, and sentiment analysis.

I also added customized stopwords related to Disney because they appear so frequently that don't have significant meanings.

[10]:
```
print(stopwords.words("english"))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're",
"you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he',
'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's",
'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what',
'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is',
'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having',
'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or',
'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about',
'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above',
'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under',
'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why',
'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some',
'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very',
's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now',
'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn',
"couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn',
"hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't",
'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn',
"shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn',
"wouldn't"]
```

[11]:
```
stopwords_700 = pd.read_csv('stopwords.txt', header= None, delimiter = "\t")
stopwords_700 = set(stopwords_700[0])
```

[12]:
```
def remove_stopwords(corpus):
    words = nltk.word_tokenize(corpus)
    stop_words = set(stopwords.words("english"))
    stop_words_customize = {'disney', 'disneyland', 'land', 'park', 'parks',␣
 ↪'world', 'disneyworld', 'disney world'}
    stop_words_all = stop_words.union(stopwords_700, stop_words_customize)
    new_corpus = [word for word in words if not word in stop_words_all]
    return new_corpus
```

[13]:
```
new_corpus = preprocess(corpus)
new_corpus = lemmatize_stem(new_corpus)
new_corpus = remove_stopwords(new_corpus)
```

1.4 Most frequent words in the corpus

[14]:
```
from collections import Counter
new_disney_df = pd.DataFrame()
new_disney_df['word'] = list(Counter(new_corpus).keys())
new_disney_df['count']= list(Counter(new_corpus).values())
new_disney_df.sort_values(by=['count'], ascending=False).head(20)
```

[14]:

	word	count
10	rides	33898
24	time	28441
15	day	27421
420	ride	17498
38	great	15659
125	food	13985
161	kids	13824
83	place	13108
55	like	11868
201	see	11566
25	visit	11257
181	people	10878
214	went	10792
945	long	10086
184	fast	10001
409	pass	9690
402	wait	9565
365	back	9432
272	fun	9265
462	first	8777

1.5 WordCloud

1.5.1 Unigram WordCloud

I first generated a unigram WordCloud to visualize frequently appeared words.

[15]:
```
micky_mask = np.array(Image.open('mickey.png'))
```

[16]:
```
text = ''
for i in new_corpus:
    text+=' '
    text+=i
```

[17]:
```
wordcloud = WordCloud(stopwords = STOPWORDS,
                      mask = micky_mask,
                      random_state = 123,
                      collocations=False,
                      max_words = 100).generate(text)
```

[18]:
```
wordcloud.to_file('wordcloud.png')
```

[18]:
```
<wordcloud.wordcloud.WordCloud at 0x7fcaf89dafd0>
```

1.5.2 Bigram WordCloud

A bigram WordCloud is usually more informative than a unigram WordCloud because it provides same contexts to the words.

[19]:
```
bi_gram_cloud = WordCloud(stopwords = STOPWORDS,
                          background_color='white',
                          random_state = 123,
                          collocation_threshold = 3,
                          max_words = 100
                          ).generate(text)
```

[20]:
```
bi_gram_cloud.to_file('bi_gram_wordcloud.png')
```

[20]:
```
<wordcloud.wordcloud.WordCloud at 0x7fcaf89dd1c0>
```

1.6 Sentiment Polarity Analysis

1.6.1 Using TextBlob

Polarity is defined between -1 and 1 with -1 defines a negative sentiment and 1 defines a positive sentiment.

[21]:
```
def clean(Review):
    Review = Review.str.lower()
    Review = Review.replace('hk', 'hongkong', regex = True)
    Review = Review.replace('hong kong', 'hongkong', regex = True)
    Review = Review.replace("n't", " not", regex = True)
    return Review
```

[22]:
```
disney_df['Review_Text'] = clean(disney_df['Review_Text'])
```

[23]:
```
disney_df['polarity'] = disney_df['Review_Text'].map(lambda text:␣
 ↪TextBlob(text).sentiment.polarity)
```

[24]:
```
disney_df.head(10)
```

[24]:

	Review_ID	Rating	Year_Month	Reviewer_Location	\
0	670772142	4	2019-4	Australia	
1	670682799	4	2019-5	Philippines	
2	670623270	4	2019-4	United Arab Emirates	
3	670607911	4	2019-4	Australia	
4	670607296	4	2019-4	United Kingdom	
5	670591897	3	2019-4	Singapore	

```
6  670585330       5     2019-4                India
7  670574142       3     2019-3             Malaysia
8  670571027       2     2019-4            Australia
9  670570869       5     2019-3                India

                                          Review_Text               Branch  \
0  if you've ever been to disneyland anywhere you…  Disneyland_HongKong
1  its been a while since d last time we visit ho…  Disneyland_HongKong
2  thanks god it wasn   t too hot or too humid wh…  Disneyland_HongKong
3  hongkong disneyland is a great compact park. u…  Disneyland_HongKong
4  the location is not in the city, took around 1…  Disneyland_HongKong
5  have been to disney world, disneyland anaheim …  Disneyland_HongKong
6  great place! your day will go by and you wo no…  Disneyland_HongKong
7  think of it as an intro to disney magic for th…  Disneyland_HongKong
8  feel so let down with this place,the disneylan…  Disneyland_HongKong
9  i can go on talking about disneyland. whatever…  Disneyland_HongKong

   polarity
0  0.243981
1  0.236131
2  0.160498
3  0.189286
4  0.266667
5 -0.065476
6  0.185000
7  0.054722
8 -0.067284
9  0.204497
```

[25]:
```
(alt.Chart(disney_df, title ="Polarity Distribution Using TextBlob").mark_bar().
 ↪encode(
    alt.X('polarity', bin=alt.Bin(maxbins=30)),
    y='count()'))
```

[25]:
```
alt.Chart(…)
```

[26]:
```
print('3 random reviews with the positive sentiment polarity: \n')
pos_review = disney_df.loc[disney_df.polarity > 0, ['Review_Text']].sample(3).
 ↪values
for review in pos_review:
    print(review)
```

```
3 random reviews with the positive sentiment polarity:

['nicest, cleanest park ever, but so expensive.  love tower of terror.  love all
the characters and character dining.  nice downtown disney area']
["visited hongkong disneyland while my boyfriend was visiting from the states.
```

```
being born in southern california, disneyland was a staple in my childhood. i've
also been to disney world and disneyland paris and i have to say that hongkong
disneyland probably provides the best value. after reading the reviews, we
decided that going during the week was our best option and also determined that
we probably would not need a full day. the park is significantly smaller, but
you get what you pay for as the admission price is about $50 usd per person. we
took the mtr from causeway bay and the disneyland resort line to get to the park
arriving at 1:30pm. the place was crowded, but no where near as bad as the other
disney parks i've been to. the longest wait we had was about 45 minutes for the
rc racer ride in toy story land. all other attractions had posted wait times of
about an hour, but we honestly never waited that long. maybe 30 minutes tops in
most cases. many people complained about the wait times, but we thoughts they
were short compared to 2  hours for space mountain in california. we were able
to see of the shows and rides we had planned on viewing and left at 8:30pm (yes,
before the fireworks). the food for purchase was cheap compared to the other
parks, but expensive from a hongkong perspective. the fruit cups are a great
value at $25hongkongd. even water is cheap ($15 18hongkongd per bottle) compared
to the other parks. they say no outside food or drink is allowed in the park,
but we learned that this excludes water. so stock up at 7 eleven in central
hongkong station before making your way to the park to save a couple
bucks.overall, we were really happy with our day as we got to experience the
disney magic at about half the price back in the states. the only problem we had
was with the amount of pushing and line cutting attempts we experienced in the
lines for attractions. we had read the reviews from other posters alerting us to
this and were prepared to stand our ground. we were given many unhappy looks
from the kids (and adults) who unsuccessfully tried to jump ahead of us, but
lots of people said they appreciated our efforts. the ride attendants would make
announcements that line jumping was discouraged, but i think they should
implement the same policy they have the states where if you're caught jumping
the line you get kicked out of the park. many of the  culprits  were very mean
and aggressive taking no care as to who they pushed (we witnessed a little girl
about 5 years old be hit to the ground by an adult trying to jump ahead). i
think disney realizes the problem, but more needs to be done. in the end we
thought it was funny and it did not ruin our experience, but we prepared to
fight for your spot."]
["i've been to disney land several times and it's always fun, even though it
gets very crowded. it's fun to experience at night and if you're somebody who is
not particularly interested in parades, that is the best time to find short
lines. choose what side of  main street  you want to be on before the parade
starts and you'll find that lines to the rides are much shorter, with a large
majority of people crowding to watch the parade. utilize fast passes as much as
possible, during the day and go as early as you can because it only gets more
crowded as midday approaches."]
```

[27]:
```
print('3 random reviews with the negative sentiment polarity: \n')
neg_review = disney_df.loc[disney_df.polarity < 0, ['Review_Text']].sample(3).
  ↪values
```

```
for review in neg_review:
    print(review)
```

```
3 random reviews with the negative sentiment polarity:

['we visited disneyland two days after christmas and it was the worst time we
have ever had. it began with the traffic around and in the park. it was
horrible. we expected huge crowds but when you ca not even walk around the park
because they have the lines for the attractions winding everywhere and even
worse pushing a wheel chair (which there was no room to do)   it was not a
magical day. then at the end of the day we tried to catch our ride to the hotel
and it turned into a free for all. there was no organization and when the bus
pulled, everyone rushed for it pushing senior citizens aside. the next day we
went to downtown disney and they wanted to charge us if we stayed more then two
hours   what!   we have never experienced that in orlando. our whole group of
five agreed that this trip made us realize that we never want to come back here.
we will stick with disneyworld where even on holidays we never had such a
horrible experience. we even overheard cast members discussing how certain cast
members should have never been cast as characters in the park because they were
terrible at it.']
['visit to disneyland is a must for anyone who visits hongkong. it is designed
for the entertainment of one and all, irrespective of the age. various rides and
shows keep you engaged throughout the day. pity we could not stay until
fireworks, as we had a child with us.']
["i'm a 20 year old young adult when i went here, and i almost forgot that i am
20. sad thing was that we do not have a lot of time to roam around. hongkong
disneyland is really big and i was like i went back on my younger years. "]
```

1.6.2 Using NLTK VADER

- Valence Aware Dictionary for Sentiment Reasoning (VADER) is a model used to analyze the sentiment of text that are sensitive to polarity and emotional intensity (Hutto, 2015)
- NLTK has a direct implementation of the VADER model.

[28]:
```
sid = SentimentIntensityAnalyzer()
```

[29]:
```
disney_df['NLTK_polarity'] = disney_df['Review_Text'].map(lambda text: sid.
↪polarity_scores(text)['compound'])
```

[30]:
```
disney_df.head(10)
```

[30]:

	Review_ID	Rating	Year_Month	Reviewer_Location	\
0	670772142	4	2019-4	Australia	
1	670682799	4	2019-5	Philippines	
2	670623270	4	2019-4	United Arab Emirates	
3	670607911	4	2019-4	Australia	
4	670607296	4	2019-4	United Kingdom	
5	670591897	3	2019-4	Singapore	

```
6  670585330       5     2019-4                India
7  670574142       3     2019-3             Malaysia
8  670571027       2     2019-4            Australia
9  670570869       5     2019-3                India

                                          Review_Text               Branch  \
0  if you've ever been to disneyland anywhere you…  Disneyland_HongKong
1  its been a while since d last time we visit ho…  Disneyland_HongKong
2  thanks god it wasn   t too hot or too humid wh…  Disneyland_HongKong
3  hongkong disneyland is a great compact park. u…  Disneyland_HongKong
4  the location is not in the city, took around 1…  Disneyland_HongKong
5  have been to disney world, disneyland anaheim …  Disneyland_HongKong
6  great place! your day will go by and you wo no…  Disneyland_HongKong
7  think of it as an intro to disney magic for th…  Disneyland_HongKong
8  feel so let down with this place,the disneylan…  Disneyland_HongKong
9  i can go on talking about disneyland. whatever…  Disneyland_HongKong

   polarity  NLTK_polarity
0  0.243981         0.7069
1  0.236131         0.9892
2  0.160498         0.9920
3  0.189286         0.8489
4  0.266667         0.2846
5 -0.065476         0.9653
6  0.185000         0.7489
7  0.054722         0.8345
8 -0.067284         0.5195
9  0.204497         0.9645
```

[31]:
```
(alt.Chart(disney_df, title ="NLTK Polarity Score Distribution").mark_bar().
 ↪encode(
    alt.X('NLTK_polarity', bin=alt.Bin(maxbins=30)),
    y='count()'))
```

[31]:
```
alt.Chart(…)
```

[32]:
```
print('3 random reviews with the positive sentiment polarity using NLTK: \n')
pos_review_NLTK = disney_df.loc[disney_df.NLTK_polarity > 0, ['Review_Text']].
 ↪sample(3).values
for review in pos_review_NLTK:
    print(review)
```

```
3 random reviews with the positive sentiment polarity using NLTK:

["i'm a regular visitor to disney world but had not been to disneyland in years.
it was nice to see where it all started.the crowds were great  not sure if that
was just the time of year we went or if it's always so much better than the
```

```
florida parks."]
["loved everything about the park! feels like your childhood is coming to life.
even though there were quite a few closures when we visited we still has an
incredible time. the parades and characters meets were amazing and i strongly
recommend sticking around for park closing for disney dreams! it is expensive so
you'd be better off taking your own drinks into the park but as it disney it was
expected. i ca not wait to go back!"]
['the happiest place in the world, indeed!! i try to go at least twice a year as
i live in phoenix and it never disappoints.']
```

[33]:
```
print('3 random reviews with the negative sentiment polarity using NLTK: \n')
neg_review_NLTK = disney_df.loc[disney_df.NLTK_polarity < 0, ['Review_Text']].
 ↪sample(3).values
for review in neg_review_NLTK:
    print(review)
```

```
3 random reviews with the negative sentiment polarity using NLTK:

["disneyland does not need another review, but i'm going to add something: it's
getting really, really tired, guys. you have to throw some infrastructure
spending at it. (but that will only be a payback for the management in a few
years, so i do not expect it to happen.)first visited: 1991. i'm pretty sure
that this will be the last time. caveats: the middle of the holiday season, but
even so the queues were bloody ridiculous. "]
["i was last here 12 years ago. it has not changed! well apart from the unclean
water and the general washed out feeling this park has. needs money spent on it!
the rides are so so old hat. it's crying out for a face lift! even speaking to
some of the staff who could be bothered to speak with you said it's awful! i
must check out the french park next time around…asterix. you do not see any of
the disney characters walking around has you once did. apparently they hang
around the mega expensive disney hotel and await to be paid 60 euros for kids to
have tea with them! the wait to get on rides is completely terrible. young kids
baking in the sun with no cover. also…not many staff on hand.
honestly…if this park did not have the disney name you would not give it
a second look. just save your money and take the kids elsewhere! we will not be
returning unless things dramatically change!!!!!!!!!!!!!!"]
['who dont like mickey, donald, buzz light year, pooh, etc step aside…hahaha
is everyone theme park from children to adult, the show is getting better with
additional show which is started last year, the parade show is on 19.45pm called
light up the night  is a must see show parade !!, one thing that i dont like in
this theme park is the food is quite expensive hehehehe']
```

1.6.3 TextBlob vs NLTK VADER

Similarities

- In total, 38,800 (about 90% of the data set) polarity scores obtained by using TextBlob and NLTK VADER indicate the same polarity (positive / negative).

- Both methods are lexicon-based. In the lexicon approach, words are mapped to sentiment, and the overall sentiment of a sentence is determined by aggregating the sentiment of each individual term.
- Lexicon sentiment analysis produces a polarity score ranging from -1 to 1, where -1 indicates highly negative sentiment and 1 indicates highly positive sentiment. A value close to 0 indicates a neutral sentiment.

Differences

- In total, 3,856 (about 10% of the data set) polarity scores obtained by using TextBlob and NLTK VADER indicate the same polarity (positive / negative).
- VADER is specifically designed for analyzing social media content. As a result, VADER dedicates significant effort to identify sentiments in content commonly found on social media platforms, including emojis, repetitive words, and punctuation marks (such as exclamation marks).

Finding I investigated the reviews on which TextBlob and NLTK VADER disagreed by drawing random samples. These reviews contain both negative and positive sentiments, which could be the reason why the two methods cannot agree on the overall sentiment.

Reference: https://pub.towardsai.net/textblob-vs-vader-for-sentiment-analysis-using-python-76883d40f9ae#:~:text=A%20critical%20difference%20between%20TextBlob,exclamation%20marks%2C%20for%2

[34]:
```
disney_df['compare'] = disney_df['polarity'].mul(disney_df['NLTK_polarity']).
↪ge(0)
```

[35]:
```
disney_df.head(10)
```

[35]:

	Review_ID	Rating	Year_Month	Reviewer_Location	\
0	670772142	4	2019-4	Australia	
1	670682799	4	2019-5	Philippines	
2	670623270	4	2019-4	United Arab Emirates	
3	670607911	4	2019-4	Australia	
4	670607296	4	2019-4	United Kingdom	
5	670591897	3	2019-4	Singapore	
6	670585330	5	2019-4	India	
7	670574142	3	2019-3	Malaysia	
8	670571027	2	2019-4	Australia	
9	670570869	5	2019-3	India	

	Review_Text	Branch	\
0	if you've ever been to disneyland anywhere you…	Disneyland_HongKong	
1	its been a while since d last time we visit ho…	Disneyland_HongKong	
2	thanks god it wasn t too hot or too humid wh…	Disneyland_HongKong	
3	hongkong disneyland is a great compact park. u…	Disneyland_HongKong	
4	the location is not in the city, took around 1…	Disneyland_HongKong	
5	have been to disney world, disneyland anaheim …	Disneyland_HongKong	
6	great place! your day will go by and you wo no…	Disneyland_HongKong	

```
7  think of it as an intro to disney magic for th…  Disneyland_HongKong
8  feel so let down with this place,the disneylan…  Disneyland_HongKong
9  i can go on talking about disneyland. whatever…  Disneyland_HongKong

   polarity  NLTK_polarity  compare
0  0.243981         0.7069     True
1  0.236131         0.9892     True
2  0.160498         0.9920     True
3  0.189286         0.8489     True
4  0.266667         0.2846     True
5 -0.065476         0.9653    False
6  0.185000         0.7489     True
7  0.054722         0.8345     True
8 -0.067284         0.5195    False
9  0.204497         0.9645     True
```

[36]:
```
disney_df['compare'].value_counts()
```

[36]:
```
True     38800
False     3856
Name: compare, dtype: int64
```

[37]:
```
print('3 random reviews with different sentiment by TextBlob and NLTK Vader:␣
 ↪\n')
diff_sent_review = disney_df.loc[disney_df['compare'] == 0, ['Review_Text']].
 ↪sample(3).values
for review in diff_sent_review:
    print(review)
```

```
3 random reviews with different sentiment by TextBlob and NLTK Vader:

['i go every year (sometimes more than once) and never get tired of it! there is
so much to see and do that a few days are never enough. even if you feel like
you need to rush to make it through all the lines, make sure to slow down once
in a while to catch the numerous street performers and musicians that are hidden
throughout the park.']
["wow…went to both parks over the last week and could count the cast
members smiling on one hand, where's the disney magic?so, so disappointed, the
parks are tired and down town is so small it gets over crowded.disney management
need to get hold of all its staff and give them some customer service training,
if you put a like for like staff member against those in florida …not one
french staff member would have a job…sorry…the only bonus was the
impromptu character appearances around the park and late night show…other
than that…go to florida (already booking mine), unless you have unruly kids
then stay in france please…"]
["if you think you're only going for the kids you're wrong, absolutely loved it!
or maybe its my inner kid coming out. very expensive, especially the food but i
```

```
expected it! worth every penny as its a memory you and the family will never
forget"]
```

References Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. Porter, M. F. (1980). An algorithm for suffix stripping. Program, 14(3), 130–137. https://doi.org/10.1108/eb046814