Disney_EDA_Report

June 16, 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
     from IPython.display import Image, display

     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(…)
```

```
[5]: display(Image(filename='count_branch.png'))
```
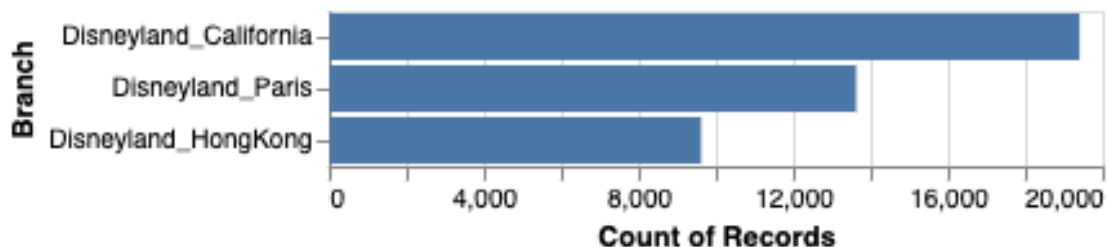

1.2.2 Here are some sample review texts:

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

```
[6]: '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. '
```

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

```
[7]: 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.

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

```
[9]: 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.

```python
[10]: 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.

```python
[11]: 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"]
```

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

```
[13]: 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
```

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

1.4 Most frequent words in the corpus

```
[15]: 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)
```

```
[15]:        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.

```python
[16]: from PIL import Image
      micky_mask = np.array(Image.open('mickey.png'))
```

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

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

```python
[19]: wordcloud.to_file('wordcloud.png')
```

```
[19]: <wordcloud.wordcloud.WordCloud at 0x7f8e88431760>
```

```python
[20]: from IPython.display import Image, display
      display(Image(filename='wordcloud.png'))
```

1.5.2 Bigram WordCloud

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

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

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

```
[22]: <wordcloud.wordcloud.WordCloud at 0x7f8e88431550>
```

```
[23]: from IPython.display import Image, display
      display(Image(filename='bi_gram_wordcloud.png'))
```

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.

```
[24]: 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
```

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

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

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

```
[27]:    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
```

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

```
[28]: alt.Chart(…)
```

```python
[29]: display(Image(filename='textblob_polarity.png'))
```

Polarity Distribution Using TextBlob
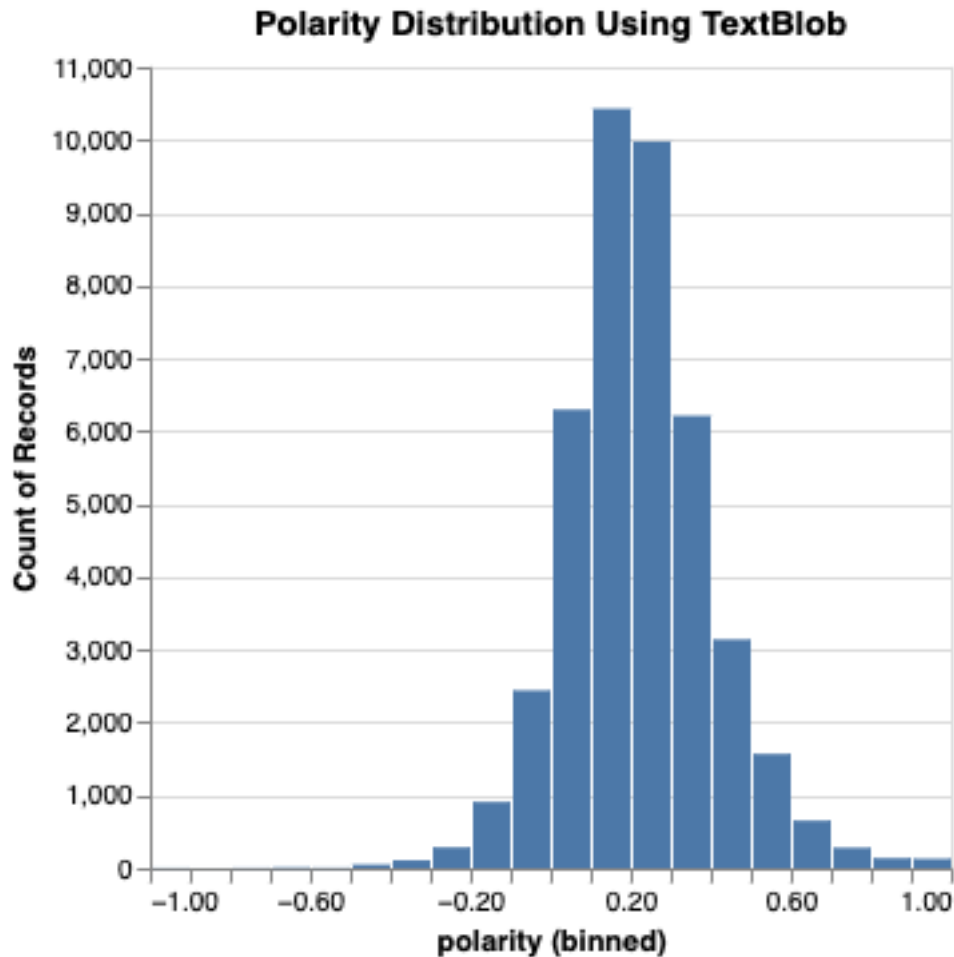

```
[30]: print('3 random reviews with the positive sentiment polarity: \n')
      pos_review = disney_df.loc[disney_df.polarity > 0, ['Review_Text']].sample(3).
        ↪values
      for review in pos_review:
          print(review)
```

3 random reviews with the positive sentiment polarity:

['disneyland was very nice but differant from florida. still great experiance,
did not stay here but wished we did']
["i am in my early 60's. the time when television was introduced to the world as
i was a toddler. i grew up with disneyland and th ecastle that seems to open the
imagination to the real world. i have on my many visits to california visited
disneyland and always seem to be taken back to my childhood, but at the same
time be escorted through the generations with the always evolving focus of the
disney management and development team, right up to my last visit a few years
ago, which was also fantastic as also my many visits to disneyworld in florida.

unfortunately on our trip to california in less then 2 weeks we have not time to stay over to visit disneyland and see the developments since our last visit, but hope to visit once again disneyworld, etc when in florida for the almost 2 weks we are there. simply if you want to be taken on a ride of the mind, body and spirit, allow a little time for the disney universe to share what they have. it will be worth it not just for now, but for a future time of reflection of something very special in your life.. greg keyes"]
['although not as big as the other disney parks (not even paris), we had a lovely day here. all the usual disney refinements were here and if you used the metro you could be on a train carriage with mickey mouse windows. my son loved it.']

```
[31]: 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:

['if you can be there when the park opens, do it. ride the most popular rides before the crowds get too bad. the park will get crowded and your options will narrow as the lines get longer. you might consider going back to the hotel to play in the pool or nap during the middle of the day and return refreshed for the late afternoon and evening.']
["i'm an annual passholder and every time i go to any other theme park i always end up wanting to go back to dl. nobody does it like them."]
['we went on a road trip for a few days over the half term break which included 2 1 2 days in disney. a family group of 8, ranging from 8 months to 56, there really was something for everyone. only thing i would say is that starting at 8am with a character breakfast meant it was a long time until fireworks at 11pm; especially for children. we all took a break in the afternoon but even so only 3 made it to 11pm.']

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.

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

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

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

```
[34]:      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
```

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

```
[35]: alt.Chart(…)
```

```python
[36]: display(Image(filename='NLTK_Polarity.png'))
```

NLTK Polarity Score Distribution
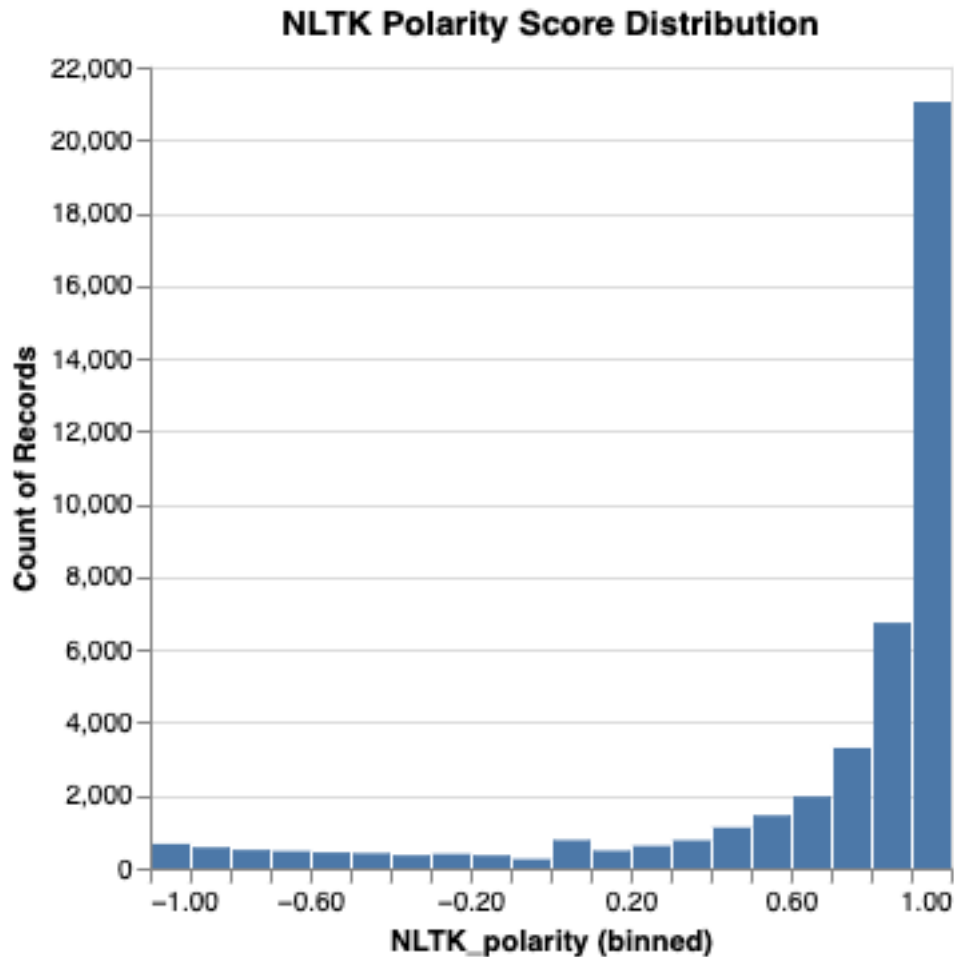

```
[37]: print('3 random reviews with the positive sentiment polarity using NLTK: \n')
      pos_review_NLTK = disney_df.loc[disney_df.NLTK_polarity > 0, ['Review_Text']].
        ↪sample(3).values
      for review in pos_review_NLTK:
          print(review)
```

3 random reviews with the positive sentiment polarity using NLTK:

['disneyland offers an experience that takes you back in time to some of your
most beloved classics while also mixing in some of the newer hits. disneyland,
california adventure park, and downtown disney are all side by side so two days
is plenty to take it all in!']
['i absolutely loved hongkong disneyland. i have been to disneyland in la twice
and while it is small in comparison, i still had an awesome time here. people
had told me not to expect much from this disneyland, but they were wrong.the
lines were extremely reasonable & the rides were a lot of fun. i would highly
recommend those travelling with children come here, it is a great, more relaxed

14

version of la.please visit!']
["there is nothing to dislike. i had a great time at both parks. loved the
fireworks display and the 'world of colour' water feature."]

[38]:
```python
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   happiest place on earth? well hongkong disney is a bit of a
disappointment.to start with getting into the place takes 30 minutes as they
have to check every single bag. when you have thousands waiting to get in, then
the queues they provide is certainly not enough. imagine having kids excitedly
wanting to get in, only to have to queue at the front entrance for 30 45
minutes.also if you did not buy your tickets online, you will also have to queue
to purchase the tickets first. lucky we bought ours online, but the ticket
printing machine printed only 5 of our tickets when we have bought 6.the
hongkong disney is smaller than the us disney, so there are fewer rides and less
things to do once you are in the park.most of the rides require at least 30
minutes of waiting time. on the day we were there, the weather bureau issued a
typhoon 3 warning, and as a result they shut down all the outdoor rides. however
during the entire day we were there, there was hardly a breeze. the management
should be a bit more flexible and reopen some of the rides as there was no risk
or danger. the only plus is that the lion king show is worth a visit. also you
can buy disney merchandise at the hongkong international airport so if you must
get the disney toys, there is an alternative.also food and drinks are very
expensive in there. another minus is the some of the visitors from the mainland
have no clues on what a queue is and their children just cut into the queues.
save your money and go to other disneyland if you must.']
["myself and my partner have dreamt for years of coming to disneyland paris
together for halloween. we are a young hard working couple and saved for over a
year to be able to afford our dream trip. we booked the full disney package in
august 2018 (flights, magical shuttle, hotel, park tickets and full board dining
plan) and could not wait to arrive for halloween week.the problems started the
second we landed at paris cdg airport when our not so magical shuttle bus was
no where to be found. after waiting an hour outside the airport at the shuttle
stop, we decided to ring the hotel (disney santa fe) to find out where it was.
it was only at this point we were informed that the shuttle stopped at 8pm
(which makes me think why did they allow us to book and pay for this, knowing
our flight arrived after this time). the hotel assured us they would fully
refund our taxi when we arrived   which cost 64euros. turns out we had to file
for a refund request which can take 15 working days to get a response from. as
we were on the full board dining plan, we only brought 80euros with us as extra
spends for shops etc. this therefore left us with 16euros for the next 4 days.it
was 10pm when we arrived at our hotel. luckily the canteen was open until

10.30pm, so we were able to pop in for a bite to eat with our dining plan. when
checking in, we were told all blocks had coffee and vending machines, however
once we got to our room, they were all 'out of order', and remained that way for
the entire stay.on our first morning we decided to visit the disneyland parc,
all in all we had a fairly enjoyable day. however the hyperinflation of guests
was unbareable to say the least. queue times were ridiculous, with some being up
to 90 minutes long for a 2 minute ride. throughout the day me and my partner ate
exactly the same meal until the evening when we ate at the hukuna matata
restaurant. i had the meatballs, and he had chicken strips. all seemed fine at
the time until the next morning when he became incredibly ill, from what seems
to be minor food poisoning. this set us back for our entire 2nd day
(halloween).in the evening of halloween, my parter decided some fresh air might
do him some good, so we headed down to the disneyland parc to try and enjoy the
halloween party   the one thing we came to experience. once we got to the gates
we showed our magic pass, only to be denied access! not once were we informed
that this was an additional expense and we could not enter the park without a
prepaid wristband. we were then forced to spend the entire evening sat in our
hotel room.we ca not comment on the price of food as we were full board, however
we have seen many families have to pay out over 80euros for a few sandwiches and
drinks.we have not had any issues with the staff, and were actually susprised at
how accomodating they are to english guests. we are writing this whilst sat in
the hotel room on our 3rd day. we did go into the disneyland parc this morning
however we just had to leave due to the amount of people they had managed to
cram in through the gates. there was no chance of getting on any ride without
having to queue for over an hour in the rain. we even had to queue to leave the
park!all in all it feels that the individual is not valued in disney anymore. it
just seems to be about getting as many people in and making as much money as
possible. i think old walt disney himself would be looking down in
dissapointment.looking forward to going home tomorrow!"]
["well i have to say very disappointing after going to disney in florida and
paris many times.the staff were not nearly as nice and helpful .we had a meal
which was awful something i have never had to say about disney .it was so packed
for a small park and not nearly the same choice in places to eat and rides to go
on the disney castle was so disappointing .well it's back to florida and paris
where they have it spot on ."]

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

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

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

```
[40]:    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
```

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

[41]:
```
True      38800
False      3856
Name: compare, dtype: int64
```

[42]:
```python
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:

['queues were not too busy and kids got to go on everything they wanted to
except thunder mountain which broke down 2 in the one day, once when we were
nearly at the front of the queue and a 2nd time later that day. staff did not
keep anyone informed on what was going on and a lot of people where left queuing
not knowing there was a problem.']
["our like   the entire experience. we are disneyholics. lost count how many
times we've been. dislike   rude, impatient people. it's disney, ca not we all
just have fun :)"]
["my wife and i only have one day at each park. so i booked us this tour for the
morning of our disneyland park day.we did not have a magical time. we had a
pretty slow group and we only got to go on 3 rides during our tour. the tour
went from 9:45 (people showed up late and we we told to be there at 9:15) until
after 2pm. over four hours and only three rides.i knew most of the history
already so the tour itself was not that great for me. when this would be worth
it:*if you have at least 3 days in the parks or come here regularly throughout
the year.*if you come regularly and want a se disney pin*if you want a relaxing
day and do not care about rides or waiting for things.*you really want to see
inside the dream suitewhen this is not worth it:*if you only have one day per
park. we missed out on a bunch of rides because the length of the tour.*if you
love rides and do not really care that much about history*if you have kids. kids
we were with hated is and were so bored.we did not even go near club 33 and did
not see walt's apt or ride in the train.long story short…i liked the tour in
general (because i love all things disney) but it wasted a lot of our day and we

18
```

missed a bunch of rides because of it…oh, and we went in early nov when the
crowds are much lighter…"]

```
[43]: disney_df.to_csv('../data/sentiment.csv') # save the polarity / sentiment␣
       ↪analysis results for further analysis using ChatGPT
```

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