```
import numpy as np
import pandas as pd
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import re
from textblob import TextBlob
from wordcloud import WordCloud
import seaborn as sns
import matplotlib.pyplot as plt
import cufflinks as cf
%matplotlib inline
from plotly.offline import init_notebook_mode, iplot
init_notebook_mode (connected = True)
cf.go_offline();
import plotly.graph_objs as go
from plotly.subplots import make_subplots
import warnings
warnings.filterwarnings ("ignore")
warnings.warn("this will not show")
pd.set_option('display.max_columns', None)

df = pd.read_csv("amazon.csv")

df.head()
```

```
   Unnamed: 0  reviewerName  overall  \
0           0           NaN        4
1           1          0mie        5
2           2           1K3        4
3           3           1m2        5
4           4  2&amp;1/2Men        5

                                          reviewText  reviewTime
day_diff  \
0                                        No issues.  23-07-2014
138
1  Purchased this for my device, it worked as adv...  25-10-2013
409
2  it works as expected. I should have sprung for...  23-12-2012
715
3  This think has worked out great.Had a diff. br...  21-11-2013
382
4  Bought it with Retail Packaging, arrived legit...  13-07-2013
513

   helpful_yes  helpful_no  total_vote  score_pos_neg_diff  \
0            0           0           0                   0
1            0           0           0                   0
2            0           0           0                   0
3            0           0           0                   0
```

```
4             0           0           0                    0

   score_average_rating  wilson_lower_bound
0                   0.0                 0.0
1                   0.0                 0.0
2                   0.0                 0.0
3                   0.0                 0.0
4                   0.0                 0.0

df

      Unnamed: 0  reviewerName  overall  \
0              0           NaN        4
1              1          0mie        5
2              2           1K3        4
3              3           1m2        5
4              4  2&amp;1/2Men        5
...          ...           ...      ...
4910        4910        ZM "J"        1
4911        4911            Zo        5
4912        4912     Z S Liske        5
4913        4913      Z Taylor        5
4914        4914           Zza        5

                                                reviewText  reviewTime
day_diff  \
0                                               No issues.  23-07-2014
138
1     Purchased this for my device, it worked as adv...  25-10-2013
409
2     it works as expected. I should have sprung for...  23-12-2012
715
3     This think has worked out great.Had a diff. br...  21-11-2013
382
4     Bought it with Retail Packaging, arrived legit...  13-07-2013
513
...                                                 ...         ...
...
4910  I bought this Sandisk 16GB Class 10 to use wit...  23-07-2013
503
4911  Used this for extending the capabilities of my...  22-08-2013
473
4912  Great card that is very fast and reliable. It ...  31-03-2014
252
4913  Good amount of space for the stuff I want to d...  16-09-2013
448
4914  I've heard bad things about this 64gb Micro SD...  01-02-2014
310

      helpful_yes  helpful_no  total_vote  score_pos_neg_diff  \
```

```
0                0           0           0                    0
1                0           0           0                    0
2                0           0           0                    0
3                0           0           0                    0
4                0           0           0                    0
...            ...         ...         ...                  ...
4910             0           0           0                    0
4911             0           0           0                    0
4912             0           0           0                    0
4913             0           0           0                    0
4914             0           0           0                    0
```

```
      score_average_rating  wilson_lower_bound
0                      0.0                 0.0
1                      0.0                 0.0
2                      0.0                 0.0
3                      0.0                 0.0
4                      0.0                 0.0
...                    ...                 ...
4910                   0.0                 0.0
4911                   0.0                 0.0
4912                   0.0                 0.0
4913                   0.0                 0.0
4914                   0.0                 0.0

[4915 rows x 12 columns]
```

```
df = df.sort_values("wilson_lower_bound", ascending = False)
df.drop ('Unnamed: 0', inplace = True, axis=1)
df.head()
```

```
                 reviewerName  overall  \
2031     Hyoun Kim "Faluzure"        5
3449        NLee the Engineer        5
4212              SkincareCEO        1
317   Amazon Customer "Kelly"        1
4672                  Twister        5

                                              reviewText  reviewTime
day_diff  \
2031  [[ UPDATE - 6/19/2014 ]]So my lovely wife boug...  05-01-2013
702
3449  I have tested dozens of SDHC and micro-SDHC ca...  26-09-2012
803
4212  NOTE:  please read the last update (scroll to ...  08-05-2013
579
317   If your card gets hot enough to be painful, it...  09-02-2012
1033
4672  Sandisk announcement of the first 128GB micro ...  03-07-2014
158
```

```
      helpful_yes  helpful_no  total_vote  score_pos_neg_diff  \
2031         1952          68        2020                1884   
3449         1428          77        1505                1351   
4212         1568         126        1694                1442   
317           422          73         495                 349   
4672           45           4          49                  41   

      score_average_rating  wilson_lower_bound  
2031              0.966337            0.957544  
3449              0.948837            0.936519  
4212              0.925620            0.912139  
317               0.852525            0.818577  
4672              0.918367            0.808109  
```

```python
def missing_values_analysis(df):
    na_columns = [col for col in df.columns if df[col].isnull().sum() > 0]
    n_miss = df[na_columns].isnull().sum().sort_values(ascending=True)
    ratio = (df[na_columns].isnull().sum() / df.shape[0] * 100).sort_values(ascending=True)
    missing_df = pd.concat([n_miss, np.round(ratio, 2)], axis=1, keys=['Missing Values', 'Ratio'])
    return missing_df

def check_dataframe(df, head=5, tail=5):
    print("SHAPE".center(82, '~'))
    print('Rows: {}'.format(df.shape[0]))
    print('Columns: {}'.format(df.shape[1]))
    print("TYPES".center(82, '~'))
    print(df.dtypes)
    print("".center(82, '~'))
    print(missing_values_analysis(df))
    print('DUPLICATED VALUES'.center(83, '~'))
    print(df.duplicated().sum())
    print("QUANTILES".center(82, '~'))

    # Filter numeric columns before calculating quantiles
    numeric_columns = df.select_dtypes(include=[np.number]).columns
    print(df[numeric_columns].quantile([0, 0.05, 0.50, 0.95, 0.99, 1]).T)

check_dataframe(df)
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~SHAPE~~~~~~~~~~~~~~~~~~~~~~~~~~~
~~~~~~~~~~~~
Rows: 4915
Columns: 11
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TYPES~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
~~~~~~~~~~~~
reviewerName              object
overall                    int64
reviewText                object
reviewTime                object
day_diff                   int64
helpful_yes                int64
helpful_no                 int64
total_vote                 int64
score_pos_neg_diff         int64
score_average_rating     float64
wilson_lower_bound       float64
dtype: object
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
~~~~~~~~~~~~
              Missing Values  Ratio
reviewerName               1   0.02
reviewText                 1   0.02
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DUPLICATED
VALUES~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
0
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~QUANTILES~~~~~~~~~~~~~~~~~~~~~~~~~
~~~~~~~~~~~~
                         0.00  0.05   0.50        0.95        0.99
1.00
overall                   1.0   2.0    5.0    5.000000    5.00000
5.000000
day_diff                  1.0  98.0  431.0  748.000000  943.00000
1064.000000
helpful_yes               0.0   0.0    0.0    1.000000    3.00000
1952.000000
helpful_no                0.0   0.0    0.0    0.000000    2.00000
183.000000
total_vote                0.0   0.0    0.0    1.000000    4.00000
2020.000000
score_pos_neg_diff     -130.0   0.0    0.0    1.000000    2.00000
1884.000000
score_average_rating      0.0   0.0    0.0    1.000000    1.00000
1.000000
wilson_lower_bound        0.0   0.0    0.0    0.206549    0.34238
0.957544

def check_class (dataframe):
    nunique_df = pd. DataFrame({ 'Variable': dataframe.columns,
                                 'Classes': [dataframe[i].nunique() \
                                             for i in dataframe.columns]})
    nunique_df = nunique_df.sort_values ('Classes', ascending = False)
    nunique_df = nunique_df.reset_index (drop = True)
    return nunique_df
check_class (df)
```

```
                 Variable  Classes
0              reviewText     4912
1            reviewerName     4594
2              reviewTime      690
3                day_diff      690
4     wilson_lower_bound       40
5   score_average_rating       28
6     score_pos_neg_diff       27
7             total_vote       26
8            helpful_yes       23
9             helpful_no       17
10               overall        5
```

```
import matplotlib.pyplot as plt

def categorical_variable_summary(df, column_name):
    # Create a bar chart
    plt.figure(figsize=(10, 5))
    df[column_name].value_counts().plot(kind='bar', color=constraints)
    plt.title(f'{column_name} Countplot')
    plt.xlabel(column_name)
    plt.ylabel('Count')
    plt.show()

    # Create a pie chart
    plt.figure(figsize=(8, 8))
    df[column_name].value_counts().plot(kind='pie', autopct='%1.1f%%',
colors=constraints)
    plt.title(f'{column_name} Percentage')
    plt.ylabel('')
    plt.show()
```

categorical_variable_summary(df, 'overall')

```
df.reviewText.head()
```

```
2031    [[ UPDATE - 6/19/2014 ]]So my lovely wife boug...
3449    I have tested dozens of SDHC and micro-SDHC ca...
4212    NOTE:  please read the last update (scroll to ...
317     If your card gets hot enough to be painful, it...
4672    Sandisk announcement of the first 128GB micro ...
Name: reviewText, dtype: object
```

```
review_example = df.reviewText [2031]
review_example
```

'[[UPDATE - 6/19/2014]]So my lovely wife bought me a Samsung Galaxy Tab 4 for Father\'s Day and I\'ve been loving it ever since. Just as other with Samsung products, the Galaxy Tab 4 has the ability to add a microSD card to expand the memory on the device. Since it\'s been over a year, I decided to do some more research to see if SanDisk offered anything new. As of 6/19/2014, their product lineup for microSD cards from worst to best (performance-wise) are the as follows:SanDiskSanDisk UltraSanDisk Ultra PLUSSanDisk ExtremeSanDisk Extreme PLUSSanDisk Extreme PRONow, the difference between all of these cards are simply the speed in which you can read/write data to the card. Yes, the published rating of most all these cards (except the SanDisk regular) are Class 10/UHS-I but that\'s just a rating... Actual real world performance does get better with each model, but with faster cards come more expensive prices. Since Amazon doesn\'t carry the Ultra PLUS model of microSD card, I had to do direct comparisons between the SanDisk Ultra ($34.27), Extreme ($57.95), and Extreme PLUS ($67.95).As mentioned in my earlier review, I purchased the SanDisk Ultra for my Galaxy S4. My question was, did I want to pay over $20 more for a card that is faster than the one I already owned? Or I could pay almost double to get SanDisk\'s 2nd-most fastest microSD card.The Ultra works perfectly fine for my style of usage (storing/capturing pictures & HD video and movie playback) on my phone. So in the end, I ended up just buying another SanDisk Ultra 64GB card. I use my cell phone *more* than I do my tablet and if the card is good enough for my phone, it\'s good enough for my tablet. I don\'t own a 4K HD camera or anything like that, so I honestly didn\'t see a need to get one of the faster cards at this time.I am now a proud owner of 2 SanDisk Ultra cards and have absolutely 0 issues with it in my Samsung devices.[[ORIGINAL REVIEW - 5/1/2013]]I haven\'t had to buy a microSD card in a long time. The last time I bought one was for my cell phone over 2 years ago. But since my cellular contract was up, I knew I would have to get a newer card in addition to my new phone, the Samsung Galaxy S4. Reason for this is because I knew my small 16GB microSD card wasn\'t going to cut it.Doing research on the Galaxy S4, I wanted to get the best card possible that had decent capacity (32 GB or greater). This led me to find that the Galaxy S4 supports the microSDXC Class 10 UHS-I card, which is the fastest possible given that class. Searching for that specifically on Amazon gave me results of only 3 vendors (as of April) that makes these microSDXC Class 10 UHS-1 cards. They are Sandisk (the majority), Samsung and Lexar. Nobody else makes these that are sold on Amazon.Seeing how SanDisk is a pretty good name out of the 3 (I\'ve used them the most), I decided upon the SanDisk because Lexar was overpriced and the Samsung one was overpriced (as well as not eligible for Amazon Prime).But the scary thing is that when you filter by the SanDisk, you literally get DOZENS of options. All of them have different model numbers, different sizes, etc. Then there\'s that confusion of what\'s the difference between SDHC & SDXC?SDHC vs SDXC:SDHC stand for "Secure Digital High Capacity" and SDXC stands for "Secure Digital eXtended Capacity". Essentially these two cards are

```
the same with the exception that SDHC only supports capcities up to 32GB and is formated with the FAT32 file system. The SDXC cards are formatted with the exFAT file system. If you use an SDXC card in a device, it must support that file system, otherwise it may not be recognizable and/or you have to reformat the card to FAT32.FAT32 vs exFAT:The differences between the two file systems means that FAT32 has a maximum file size of 4GB, limited by that file system. exFAT on the otherhand, supports file sizes up to 2TB (terabytes). The only thing you need to know here really is that it\'s possible your device doesn\'t support exFAT. If that\'s the case, just reformat it to FAT32. REMEMBER FORMATTING ERASES ALL DATA!To clarify the model numbers, I I hopped over to the SanDisk official webpage. What I found there is that they offer two "highspeed" options for SanDisk cards. These are SanDisk Extreme Pro and SanDisk Ultra. SanDisk Extreme Pro is a line that supports read speeds up to 95MB/sec, however they are SDHC only. To make things worse, they are currently only available in 16GB & 8GB capacities. Since one of my requirements was to have a lot of storage, I ruled these out.The remaining devices listed on Amazon\'s search were the SanDisk Ultra line. But here, confusion sets in because SanDisk separates these cards to two different devices. Cameras & mobile devices. Is there a real difference between the two or is this just a marketing stunt? Unfortunately I\'m not sure but I do know the price difference between the two range from a couple cents to a few dollars. Since I wasn\'t sure, I opted for the one specifically targeted for mobile devices (just in case there is some kind of compatibility issue). To find the exact model number, I would go to Sandisk\'s webpage (sandisk.com) and compare their existing product lineup. From there, you get exact model numbers and you can then search Amazon for these model numbers. That is how I got mine (SDSDQUA-064G).As for speed tests, I haven\'t run any specific testing, but copying 8 GB worth of data from my PC to the card literally took just a few minutes.One last note is that Amazon attaches additional characters to the end (for example SDSDQUA-064G-AFFP-A vs SDSDQUA-064G-U46A). The difference between the two is that the "AFFP-A" means "Amazon Frustration Free Packaging". Other than that, these are exactly the same.  If you\'re wondering what I got (and want to use it in your Galaxy S4), I got the SDSDQUA-064G-u46A and it works like charm.'

review_example = df.reviewText[2031]

# Apply the substitution to remove non-alphabetic characters
review_example = re.sub("[^a-zA-Z ]", '', review_example)



# Split the string into individual words
review_example = review_example.lower().split()
```

```
review_example
```

```
['update',
 'so',
 'my',
 'lovely',
 'wife',
 'bought',
 'me',
 'a',
 'samsung',
 'galaxy',
 'tab',
 'for',
 'fathers',
 'day',
 'and',
 'ive',
 'been',
 'loving',
 'it',
 'ever',
 'since',
 'just',
 'as',
 'other',
 'with',
 'samsung',
 'products',
 'the',
 'galaxy',
 'tab',
 'has',
 'the',
 'ability',
 'to',
 'add',
 'a',
 'microsd',
 'card',
 'to',
 'expand',
 'the',
 'memory',
 'on',
 'the',
 'device',
 'since',
 'its',
 'been',
```

```
 'over',
 'a',
 'year',
 'i',
 'decided',
 'to',
 'do',
 'some',
 'more',
 'research',
 'to',
 'see',
 'if',
 'sandisk',
 'offered',
 'anything',
 'new',
 'as',
 'of',
 'their',
 'product',
 'lineup',
 'for',
 'microsd',
 'cards',
 'from',
 'worst',
 'to',
 'best',
 'performancewise',
 'are',
 'the',
 'as',
 'followssandisksandisk',
 'ultrasandisk',
 'ultra',
 'plussandisk',
 'extremesandisk',
 'extreme',
 'plussandisk',
 'extreme',
 'pronow',
 'the',
 'difference',
 'between',
 'all',
 'of',
 'these',
 'cards',
```

```
 'are',
 'simply',
 'the',
 'speed',
 'in',
 'which',
 'you',
 'can',
 'readwrite',
 'data',
 'to',
 'the',
 'card',
 'yes',
 'the',
 'published',
 'rating',
 'of',
 'most',
 'all',
 'these',
 'cards',
 'except',
 'the',
 'sandisk',
 'regular',
 'are',
 'class',
 'uhsi',
 'but',
 'thats',
 'just',
 'a',
 'rating',
 'actual',
 'real',
 'world',
 'performance',
 'does',
 'get',
 'better',
 'with',
 'each',
 'model',
 'but',
 'with',
 'faster',
 'cards',
 'come',
```

```
 'more',
 'expensive',
 'prices',
 'since',
 'amazon',
 'doesnt',
 'carry',
 'the',
 'ultra',
 'plus',
 'model',
 'of',
 'microsd',
 'card',
 'i',
 'had',
 'to',
 'do',
 'direct',
 'comparisons',
 'between',
 'the',
 'sandisk',
 'ultra',
 'extreme',
 'and',
 'extreme',
 'plus',
 'as',
 'mentioned',
 'in',
 'my',
 'earlier',
 'review',
 'i',
 'purchased',
 'the',
 'sandisk',
 'ultra',
 'for',
 'my',
 'galaxy',
 's',
 'my',
 'question',
 'was',
 'did',
 'i',
 'want',
```

```
 'to',
 'pay',
 'over',
 'more',
 'for',
 'a',
 'card',
 'that',
 'is',
 'faster',
 'than',
 'the',
 'one',
 'i',
 'already',
 'owned',
 'or',
 'i',
 'could',
 'pay',
 'almost',
 'double',
 'to',
 'get',
 'sandisks',
 'ndmost',
 'fastest',
 'microsd',
 'cardthe',
 'ultra',
 'works',
 'perfectly',
 'fine',
 'for',
 'my',
 'style',
 'of',
 'usage',
 'storingcapturing',
 'pictures',
 'hd',
 'video',
 'and',
 'movie',
 'playback',
 'on',
 'my',
 'phone',
 'so',
```

```
 'in',
 'the',
 'end',
 'i',
 'ended',
 'up',
 'just',
 'buying',
 'another',
 'sandisk',
 'ultra',
 'gb',
 'card',
 'i',
 'use',
 'my',
 'cell',
 'phone',
 'more',
 'than',
 'i',
 'do',
 'my',
 'tablet',
 'and',
 'if',
 'the',
 'card',
 'is',
 'good',
 'enough',
 'for',
 'my',
 'phone',
 'its',
 'good',
 'enough',
 'for',
 'my',
 'tablet',
 'i',
 'dont',
 'own',
 'a',
 'k',
 'hd',
 'camera',
 'or',
 'anything',
```

```
 'like',
 'that',
 'so',
 'i',
 'honestly',
 'didnt',
 'see',
 'a',
 'need',
 'to',
 'get',
 'one',
 'of',
 'the',
 'faster',
 'cards',
 'at',
 'this',
 'timei',
 'am',
 'now',
 'a',
 'proud',
 'owner',
 'of',
 'sandisk',
 'ultra',
 'cards',
 'and',
 'have',
 'absolutely',
 'issues',
 'with',
 'it',
 'in',
 'my',
 'samsung',
 'devices',
 'original',
 'review',
 'i',
 'havent',
 'had',
 'to',
 'buy',
 'a',
 'microsd',
 'card',
 'in',
```

```
 'a',
 'long',
 'time',
 'the',
 'last',
 'time',
 'i',
 'bought',
 'one',
 'was',
 'for',
 'my',
 'cell',
 'phone',
 'over',
 'years',
 'ago',
 'but',
 'since',
 'my',
 'cellular',
 'contract',
 'was',
 'up',
 'i',
 'knew',
 'i',
 'would',
 'have',
 'to',
 'get',
 'a',
 'newer',
 'card',
 'in',
 'addition',
 'to',
 'my',
 'new',
 'phone',
 'the',
 'samsung',
 'galaxy',
 's',
 'reason',
 'for',
 'this',
 'is',
 'because',
```

```
 'i',
 'knew',
 'my',
 'small',
 'gb',
 'microsd',
 'card',
 'wasnt',
 'going',
 'to',
 'cut',
 'itdoing',
 'research',
 'on',
 'the',
 'galaxy',
 's',
 'i',
 'wanted',
 'to',
 'get',
 'the',
 'best',
 'card',
 'possible',
 'that',
 'had',
 'decent',
 'capacity',
 'gb',
 'or',
 'greater',
 'this',
 'led',
 'me',
 'to',
 'find',
 'that',
 'the',
 'galaxy',
 's',
 'supports',
 'the',
 'microsdxc',
 'class',
 'uhsi',
 'card',
 'which',
 'is',
```

```
 'the',
 'fastest',
 'possible',
 'given',
 'that',
 'class',
 'searching',
 'for',
 'that',
 'specifically',
 'on',
 'amazon',
 'gave',
 'me',
 'results',
 'of',
 'only',
 'vendors',
 'as',
 'of',
 'april',
 'that',
 'makes',
 'these',
 'microsdxc',
 'class',
 'uhs',
 'cards',
 'they',
 'are',
 'sandisk',
 'the',
 'majority',
 'samsung',
 'and',
 'lexar',
 'nobody',
 'else',
 'makes',
 'these',
 'that',
 'are',
 'sold',
 'on',
 'amazonseeing',
 'how',
 'sandisk',
 'is',
 'a',
```

```
 'pretty',
 'good',
 'name',
 'out',
 'of',
 'the',
 'ive',
 'used',
 'them',
 'the',
 'most',
 'i',
 'decided',
 'upon',
 'the',
 'sandisk',
 'because',
 'lexar',
 'was',
 'overpriced',
 'and',
 'the',
 'samsung',
 'one',
 'was',
 'overpriced',
 'as',
 'well',
 'as',
 'not',
 'eligible',
 'for',
 'amazon',
 'primebut',
 'the',
 'scary',
 'thing',
 'is',
 'that',
 'when',
 'you',
 'filter',
 'by',
 'the',
 'sandisk',
 'you',
 'literally',
 'get',
 'dozens',
 'of',
```

```
'options',
'all',
'of',
'them',
'have',
'different',
'model',
'numbers',
'different',
'sizes',
'etc',
'then',
'theres',
'that',
'confusion',
'of',
'whats',
'the',
'difference',
'between',
'sdhc',
'sdxcsdhc',
'vs',
'sdxcsdhc',
'stand',
'for',
'secure',
'digital',
'high',
'capacity',
'and',
'sdxc',
'stands',
'for',
'secure',
'digital',
'extended',
'capacity',
'essentially',
'these',
'two',
'cards',
'are',
'the',
'same',
'with',
'the',
'exception',
'that',
```

```
 'sdhc',
 'only',
 'supports',
 'capcities',
 'up',
 'to',
 'gb',
 'and',
 'is',
 'formated',
 'with',
 'the',
 'fat',
 'file',
 'system',
 'the',
 'sdxc',
 'cards',
 'are',
 'formatted',
 'with',
 'the',
 'exfat',
 'file',
 'system',
 'if',
 'you',
 'use',
 'an',
 'sdxc',
 'card',
 'in',
 'a',
 'device',
 'it',
 'must',
 'support',
 'that',
 'file',
 'system',
 'otherwise',
 'it',
 'may',
 'not',
 'be',
 'recognizable',
 'andor',
 'you',
 'have',
```

```
 'to',
 'reformat',
 'the',
 'card',
 'to',
 'fatfat',
 'vs',
 'exfatthe',
 'differences',
 'between',
 'the',
 'two',
 'file',
 'systems',
 'means',
 'that',
 'fat',
 'has',
 'a',
 'maximum',
 'file',
 'size',
 'of',
 'gb',
 'limited',
 'by',
 'that',
 'file',
 'system',
 'exfat',
 'on',
 'the',
 'otherhand',
 'supports',
 'file',
 'sizes',
 'up',
 'to',
 'tb',
 'terabytes',
 'the',
 'only',
 'thing',
 'you',
 'need',
 'to',
 'know',
 'here',
 'really',
```

```
 'is',
 'that',
 'its',
 'possible',
 'your',
 'device',
 'doesnt',
 'support',
 'exfat',
 'if',
 'thats',
 'the',
 'case',
 'just',
 'reformat',
 'it',
 'to',
 'fat',
 'remember',
 'formatting',
 'erases',
 'all',
 'datato',
 'clarify',
 'the',
 'model',
 'numbers',
 'i',
 'i',
 'hopped',
 'over',
 'to',
 'the',
 'sandisk',
 'official',
 'webpage',
 'what',
 'i',
 'found',
 'there',
 'is',
 'that',
 'they',
 'offer',
 'two',
 'highspeed',
 'options',
 'for',
 'sandisk',
```

```
 'cards',
 'these',
 'are',
 'sandisk',
 'extreme',
 'pro',
 'and',
 'sandisk',
 'ultra',
 'sandisk',
 'extreme',
 'pro',
 'is',
 'a',
 'line',
 'that',
 'supports',
 'read',
 'speeds',
 'up',
 'to',
 'mbsec',
 'however',
 'they',
 'are',
 'sdhc',
 'only',
 'to',
 'make',
 'things',
 'worse',
 'they',
 'are',
 'currently',
 'only',
 'available',
 'in',
 'gb',
 'gb',
 'capacities',
 'since',
 'one',
 'of',
 'my',
 'requirements',
 'was',
 'to',
 'have',
 'a',
```

```
 'lot',
 'of',
 'storage',
 'i',
 'ruled',
 'these',
 'outthe',
 'remaining',
 'devices',
 'listed',
 'on',
 'amazons',
 'search',
 'were',
 'the',
 'sandisk',
 'ultra',
 'line',
 'but',
 'here',
 'confusion',
 'sets',
 'in',
 'because',
 'sandisk',
 'separates',
 'these',
 'cards',
 'to',
 'two',
 'different',
 'devices',
 'cameras',
 'mobile',
 'devices',
 'is',
 'there',
 'a',
 'real',
 'difference',
 'between',
 'the',
 'two',
 'or',
 'is',
 'this',
 'just',
 'a',
 'marketing',
```

```
 'stunt',
 'unfortunately',
 'im',
 'not',
 'sure',
 'but',
 'i',
 'do',
 'know',
 'the',
 'price',
 'difference',
 'between',
 'the',
 'two',
 'range',
 'from',
 'a',
 'couple',
 'cents',
 'to',
 'a',
 'few',
 'dollars',
 'since',
 'i',
 'wasnt',
 'sure',
 'i',
 'opted',
 'for',
 'the',
 'one',
 'specifically',
 'targeted',
 'for',
 'mobile',
 'devices',
 'just',
 'in',
 'case',
 'there',
 'is',
 'some',
 'kind',
 'of',
 'compatibility',
 'issue',
 'to',
```

```
 'find',
 'the',
 'exact',
 'model',
 'number',
 'i',
 'would',
 'go',
 'to',
 'sandisks',
 'webpage',
 'sandiskcom',
 'and',
 'compare',
 'their',
 'existing',
 'product',
 'lineup',
 'from',
 'there',
 'you',
 'get',
 'exact',
 'model',
 'numbers',
 'and',
 'you',
 'can',
 'then',
 'search',
 'amazon',
 'for',
 'these',
 'model',
 'numbers',
 'that',
 'is',
 'how',
 'i',
 'got',
 'mine',
 'sdsdquagas',
 'for',
 'speed',
 'tests',
 'i',
 'havent',
 'run',
 'any',
```

```
 'specific',
 'testing',
 'but',
 'copying',
 'gb',
 'worth',
 'of',
 'data',
 'from',
 'my',
 'pc',
 'to',
 'the',
 'card',
 'literally',
 'took',
 'just',
 'a',
 'few',
 'minutesone',
 'last',
 'note',
 'is',
 'that',
 'amazon',
 'attaches',
 'additional',
 'characters',
 'to',
 'the',
 'end',
 'for',
 'example',
 'sdsdquagaffpa',
 'vs',
 'sdsdquagua',
 'the',
 'difference',
 'between',
 'the',
 'two',
 'is',
 'that',
 'the',
 'affpa',
 'means',
 'amazon',
 'frustration',
 'free',
```

```
 'packaging',
 'other',
 'than',
 'that',
 'these',
 'are',
 'exactly',
 'the',
 'same',
 'if',
 'youre',
 'wondering',
 'what',
 'i',
 'got',
 'and',
 'want',
 'to',
 'use',
 'it',
 ...]
```

```
rt = lambda x: re.sub("[^a-zA-Z]",' ' ,str(x))
df["reviewText"] = df["reviewText"].map(rt)
df ["reviewText"]= df["reviewText"].str.lower()
df.head()
```

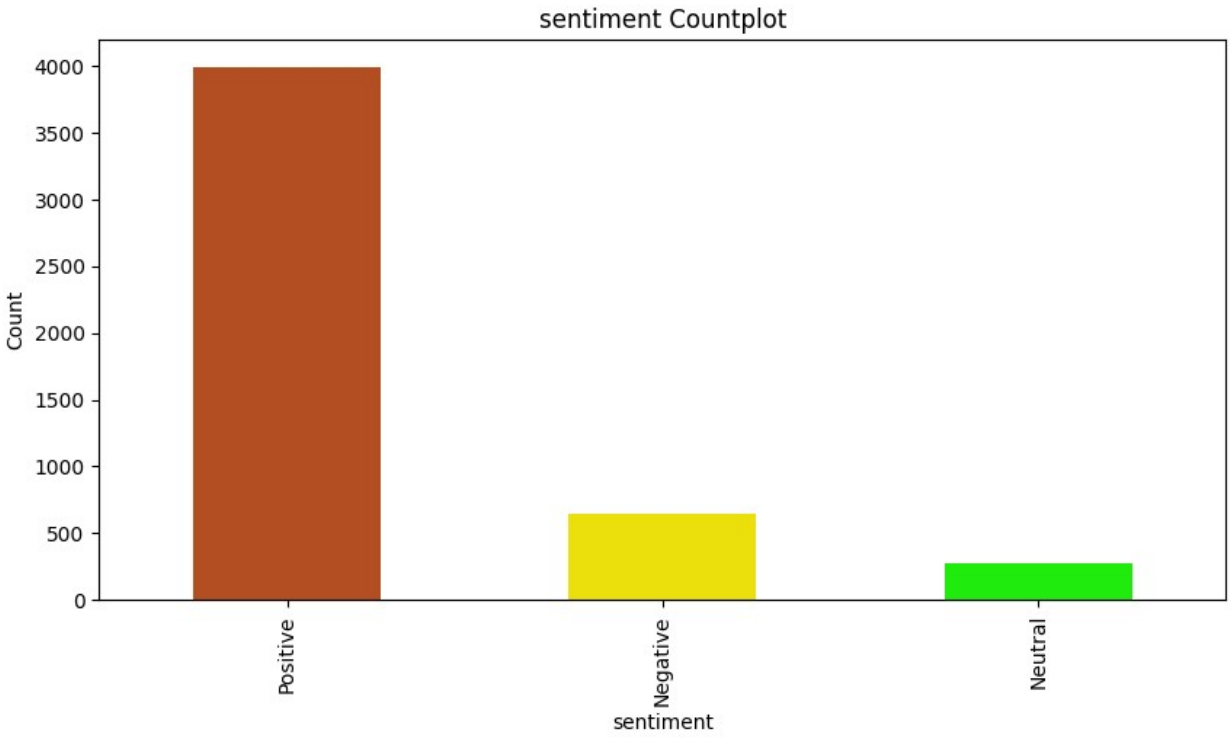
```
                 reviewerName  overall  \
2031     Hyoun Kim "Faluzure"        5
3449        NLee the Engineer        5
4212              SkincareCEO        1
317   Amazon Customer "Kelly"        1
4672                  Twister        5

                                              reviewText  reviewTime
day_diff  \
2031     update                so my lovely wife boug...  05-01-2013
702
3449  i have tested dozens of sdhc and micro sdhc ca...  26-09-2012
803
4212  note   please read the last update  scroll to ...  08-05-2013
579
317   if your card gets hot enough to be painful  it...  09-02-2012
1033
4672  sandisk announcement of the first    gb micro ...  03-07-2014
158

      helpful_yes  helpful_no  total_vote  score_pos_neg_diff  \
2031         1952          68        2020                1884
3449         1428          77        1505                1351
```

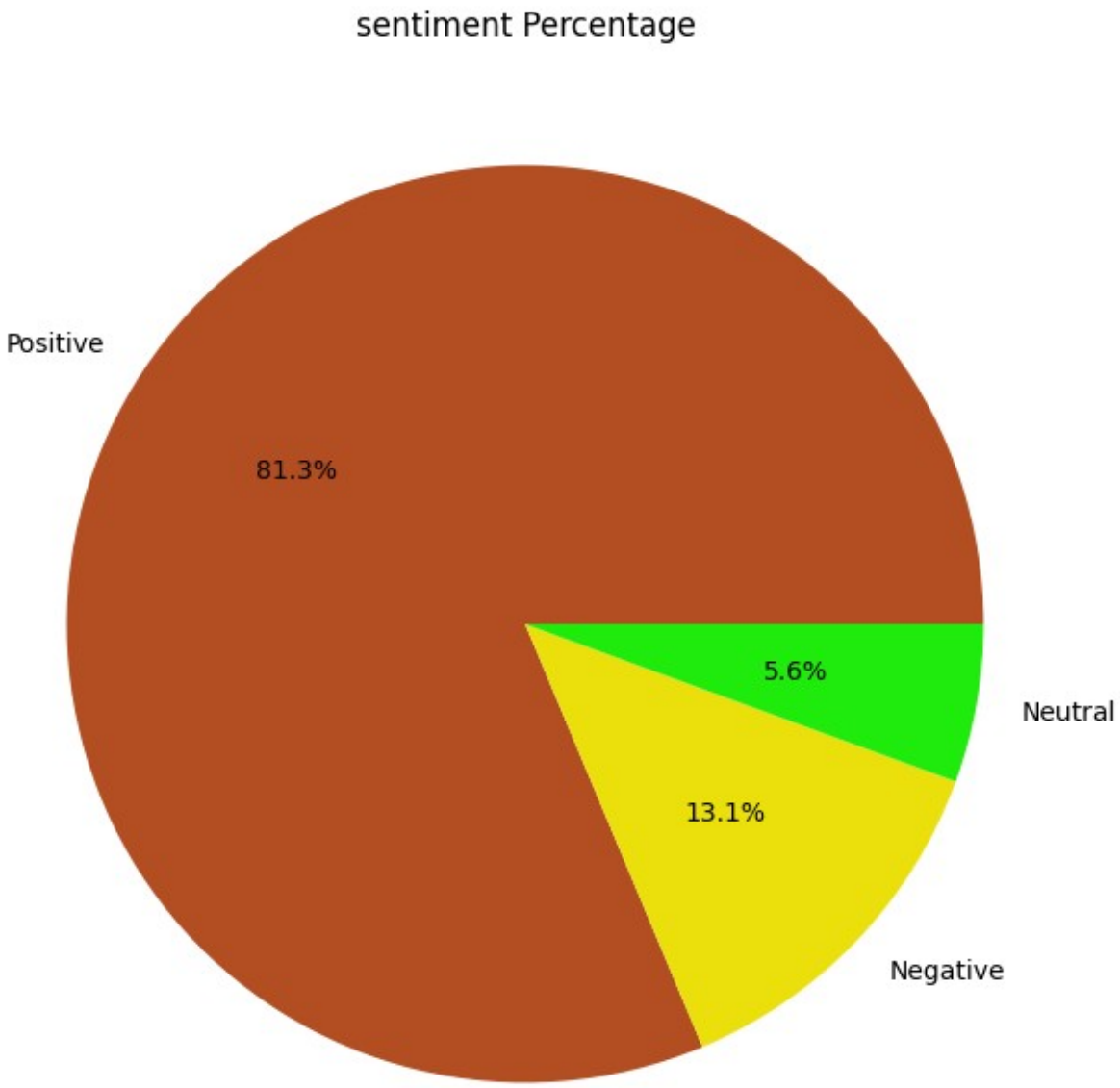
```
4212          1568          126         1694                  1442
317            422           73          495                   349
4672            45            4           49                    41

      score_average_rating  wilson_lower_bound
2031              0.966337            0.957544
3449              0.948837            0.936519
4212              0.925620            0.912139
317               0.852525            0.818577
4672              0.918367            0.808109
```

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer


# Sentiment Analysis using TextBlob
df[['polarity', 'subjectivity']] = df['reviewText'].apply(lambda text:
pd.Series(TextBlob(text).sentiment))

# Sentiment Analysis using VaderSentiment
analyzer = SentimentIntensityAnalyzer()

for index, row in df.iterrows():
    score = analyzer.polarity_scores(row['reviewText'])
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']

    if neg > pos:
        df.at[index, 'sentiment'] = "Negative"
    elif pos > neg:
        df.at[index, 'sentiment'] = "Positive"
    else:
        df.at[index, 'sentiment'] = "Neutral"


df[df['sentiment'] == 'Positive'].sort_values(by="wilson_lower_bound",
                                              ascending =
False).head (5)
```

```
                  reviewerName  overall  \
2031     Hyoun Kim "Faluzure"        5
3449        NLee the Engineer        5
4212              SkincareCEO        1
317   Amazon Customer "Kelly"        1
4672                  Twister        5

                                            reviewText  reviewTime
day_diff  \
2031     update               so my lovely wife boug...  05-01-2013
702
3449  i have tested dozens of sdhc and micro sdhc ca...  26-09-2012
```

```
803
4212  note   please read the last update  scroll to ...  08-05-2013
579
317   if your card gets hot enough to be painful  it...  09-02-2012
1033
4672  sandisk announcement of the first    gb micro ...  03-07-2014
158

      helpful_yes  helpful_no  total_vote  score_pos_neg_diff  \
2031         1952          68        2020                1884
3449         1428          77        1505                1351
4212         1568         126        1694                1442
317           422          73         495                 349
4672           45           4          49                  41

      score_average_rating  wilson_lower_bound  polarity  subjectivity
\
2031              0.966337            0.957544  0.163859      0.562259

3449              0.948837            0.936519  0.103870      0.516435

4212              0.925620            0.912139  0.212251      0.505394

317               0.852525            0.818577  0.143519      0.494207

4672              0.918367            0.808109  0.172332      0.511282


     sentiment
2031  Positive
3449  Positive
4212  Positive
317   Positive
4672  Positive

categorical_variable_summary(df,'sentiment')
```


sentiment Countplot
Count
4000
3500
3000
2500
2000
1500
1000
500
0
Positive
Negative
Neutral
sentiment


sentiment Percentage
Positive
81.3%
5.6%
Neutral
13.1%
Negative