

sentiment analysis

August 11, 2024

[5]: `pip install textblob`

```
Requirement already satisfied: textblob in c:\users\sujal
ojha\anaconda3\lib\site-packages (0.18.0.post0)
Requirement already satisfied: nltk>=3.8 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from textblob) (3.8.1)
Requirement already satisfied: click in c:\users\sujal ojha\anaconda3\lib\site-
packages (from nltk>=3.8->textblob) (8.0.4)
Requirement already satisfied: joblib in c:\users\sujal ojha\anaconda3\lib\site-
packages (from nltk>=3.8->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from nltk>=3.8->textblob) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\sujal ojha\anaconda3\lib\site-
packages (from nltk>=3.8->textblob) (4.65.0)
Requirement already satisfied: colorama in c:\users\sujal
ojha\anaconda3\lib\site-packages (from click->nltk>=3.8->textblob) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

[6]: `pip install wordcloud`

```
Collecting wordcloud
  Obtaining dependency information for wordcloud from https://files.pythonhosted
.org/packages/f5/b0/247159f61c5d5d6647171bef84430b7efad4db504f0229674024f3a4f7f2
/wordcloud-1.9.3-cp311-cp311-win_amd64.whl.metadata
  Downloading wordcloud-1.9.3-cp311-cp311-win_amd64.whl.metadata (3.5 kB)
Requirement already satisfied: numpy>=1.6.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from wordcloud) (1.24.3)
Requirement already satisfied: pillow in c:\users\sujal ojha\anaconda3\lib\site-
packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in c:\users\sujal
ojha\anaconda3\lib\site-packages (from wordcloud) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
```

```

Requirement already satisfied: packaging>=20.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (23.1)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from python-
dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Downloading wordcloud-1.9.3-cp311-cp311-win_amd64.whl (300 kB)
----- 0.0/300.2 kB ? eta -:--:--
- ----- 10.2/300.2 kB ? eta -:--:--
--- ----- 30.7/300.2 kB 660.6 kB/s eta 0:00:01
----- 61.4/300.2 kB 469.7 kB/s eta 0:00:01
----- 112.6/300.2 kB 656.4 kB/s eta 0:00:01
----- 112.6/300.2 kB 656.4 kB/s eta 0:00:01
----- 163.8/300.2 kB 656.4 kB/s eta 0:00:01
----- 204.8/300.2 kB 692.4 kB/s eta 0:00:01
----- 300.2/300.2 kB 845.3 kB/s eta 0:00:00
Installing collected packages: wordcloud
Successfully installed wordcloud-1.9.3
Note: you may need to restart the kernel to use updated packages.

```

[9]: `pip install Seaborn`

```

Requirement already satisfied: Seaborn in c:\users\sujal
ojha\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from Seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from Seaborn) (2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from Seaborn) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn)
(0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn)
(4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (9.4.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\sujal

```

```

ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->Seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from pandas>=0.25->Seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from pandas>=0.25->Seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.1->Seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

```

```
[12]: pip install cufflinks
```

```

Collecting cufflinks
  Downloading cufflinks-0.17.3.tar.gz (81 kB)
    ----- 0.0/81.7 kB ? eta -:-:--
    ----- 20.5/81.7 kB 217.9 kB/s eta 0:00:01
    ----- 30.7/81.7 kB 325.1 kB/s eta 0:00:01
    ----- 81.7/81.7 kB 656.0 kB/s eta 0:00:00
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: numpy>=1.9.2 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (1.24.3)
Requirement already satisfied: pandas>=0.19.2 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (2.0.3)
Requirement already satisfied: plotly>=4.1.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (5.9.0)
Requirement already satisfied: six>=1.9.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (1.16.0)
Collecting colorlover>=0.2.1 (from cufflinks)
  Obtaining dependency information for colorlover>=0.2.1 from https://files.pyth
onhosted.org/packages/9a/53/f696e4480b1d1de3b1523991dea71cf417c8b19fe70c704da164
f3f90972/colorlover-0.3.0-py3-none-any.whl.metadata
  Downloading colorlover-0.3.0-py3-none-any.whl.metadata (421 bytes)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (68.0.0)
Requirement already satisfied: ipython>=5.3.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (8.15.0)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from cufflinks) (8.0.4)
Requirement already satisfied: backcall in c:\users\sujal
ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: decorator in c:\users\sujal
ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.18.1)
Requirement already satisfied: matplotlib-inline in c:\users\sujal

```

ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.1.6)
 Requirement already satisfied: pickleshare in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.7.5)
 Requirement already satisfied: prompt-toolkit!=3.0.37,<3.1.0,>=3.0.30 in
 c:\users\sujal ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks)
 (3.0.36)
 Requirement already satisfied: pygments>=2.4.0 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (2.15.1)
 Requirement already satisfied: stack-data in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.2.0)
 Requirement already satisfied: traitlets>=5 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.7.1)
 Requirement already satisfied: colorama in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.4.6)
 Requirement already satisfied: ipykernel>=4.5.1 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (6.25.0)
 Requirement already satisfied: widgetsnbextension~=4.0 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (4.0.5)
 Requirement already satisfied: jupyterlab-widgets~=3.0 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (3.0.5)
 Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from pandas>=0.19.2->cufflinks) (2.8.2)
 Requirement already satisfied: pytz>=2020.1 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from pandas>=0.19.2->cufflinks) (2023.3.post1)
 Requirement already satisfied: tzdata>=2022.1 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from pandas>=0.19.2->cufflinks) (2023.3)
 Requirement already satisfied: tenacity>=6.2.0 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from plotly>=4.1.1->cufflinks) (8.2.2)
 Requirement already satisfied: comm>=0.1.1 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (0.1.2)
 Requirement already satisfied: debugpy>=1.6.5 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (1.6.7)
 Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (7.4.9)
 Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (5.3.0)
 Requirement already satisfied: nest-asyncio in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (1.5.6)
 Requirement already satisfied: packaging in c:\users\sujal
 ojha\anaconda3\lib\site-packages (from
 ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (23.1)
 Requirement already satisfied: psutil in c:\users\sujal ojha\anaconda3\lib\site-
 packages (from ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (5.9.0)

```

Requirement already satisfied: pyzmq>=20 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (23.2.0)
Requirement already satisfied: tornado>=6.1 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (6.3.2)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from jedi>=0.16->ipython>=5.3.0->cufflinks)
(0.8.3)
Requirement already satisfied: wcwidth in c:\users\sujal
ojha\anaconda3\lib\site-packages (from prompt-
toolkit!=3.0.37,<3.1.0,>=3.0.30->ipython>=5.3.0->cufflinks) (0.2.5)
Requirement already satisfied: executing in c:\users\sujal
ojha\anaconda3\lib\site-packages (from stack-data->ipython>=5.3.0->cufflinks)
(0.8.3)
Requirement already satisfied: asttokens in c:\users\sujal
ojha\anaconda3\lib\site-packages (from stack-data->ipython>=5.3.0->cufflinks)
(2.0.5)
Requirement already satisfied: pure-eval in c:\users\sujal
ojha\anaconda3\lib\site-packages (from stack-data->ipython>=5.3.0->cufflinks)
(0.2.2)
Requirement already satisfied: entrypoints in c:\users\sujal
ojha\anaconda3\lib\site-packages (from jupyter-
client>=6.1.12->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (0.4)
Requirement already satisfied: platformdirs>=2.5 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from jupyter-
core!=5.0.*,>=4.12->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (3.10.0)
Requirement already satisfied: pywin32>=300 in c:\users\sujal
ojha\anaconda3\lib\site-packages (from jupyter-
core!=5.0.*,>=4.12->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (305.1)
Downloading colorlover-0.3.0-py3-none-any.whl (8.9 kB)
Building wheels for collected packages: cufflinks
  Building wheel for cufflinks (setup.py): started
  Building wheel for cufflinks (setup.py): finished with status 'done'
  Created wheel for cufflinks: filename=cufflinks-0.17.3-py3-none-any.whl
size=68725
sha256=203767d79e46ecef9ea9c39d0416d0dff6c0639525802a9289cd92549f08aa95
  Stored in directory: c:\users\sujal\ojha\appdata\local\pip\cache\wheels\c3\bf\
8f\afa0730b29e347e8c1071049f89748c1262e7a14a838820ae6
Successfully built cufflinks
Installing collected packages: colorlover, cufflinks
Successfully installed colorlover-0.3.0 cufflinks-0.17.3
Note: you may need to restart the kernel to use updated packages.

```

```

[14]: 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)

```

```

[15]: df=pd.read_csv("amazon.csv")
df.head()

```

```

[15]:   Unnamed: 0  reviewerName  overall  \
0           0           NaN         4
1           1           Omie         5
2           2           1K3         4
3           3           1m2         5
4           4  2&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

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

```
[18]:
```

	reviewerName	overall	\
2031	Hyouun 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

```
[24]: 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, '~'))
numeric_cols = df.select_dtypes(include=['number'])
print(numeric_cols.quantile([0, .05, 0.50, .95, .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	5.000000	5.000000
day_diff	1.0	98.0	431.0	748.000000	943.000000	1064.000000
helpful_yes	0.0	0.0	0.0	1.000000	3.000000	1952.000000
helpful_no	0.0	0.0	0.0	0.000000	2.000000	183.000000
total_vote	0.0	0.0	0.0	1.000000	4.000000	2020.000000
score_pos_neg_diff	-130.0	0.0	0.0	1.000000	2.000000	1884.000000
score_average_rating	0.0	0.0	0.0	1.000000	1.000000	1.000000

wilson_lower_bound	0.0	0.0	0.0	0.206549	0.34238	0.957544
--------------------	-----	-----	-----	----------	---------	----------

```
[28]: def check_class(dataframe):
        unique_df=pd.DataFrame({'Variable':dataframe.columns,'Classes':
        ↪[dataframe[i].unique()\\

                                                                    for i in_

        ↪dataframe.columns]})
        unique_df=unique_df.sort_values('Classes',ascending=False)
        unique_df=unique_df.reset_index(drop=True)
        return unique_df

check_class(df)
```

```
[28]:
```

	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

```
[49]: constraints=['#B34D22','#EBE00C','#1FEB0C','#0C92EB','#EB0CD5']
def categorical_variable_summary(df,column_name):
    fig=make_subplots(rows=1,cols=2,
        subplot_titles=('Countplot','Percentage'),
        specs=[[{"type":"xy"},{'type':'domain'}]])

    fig.add_trace(go.Bar(y=df[column_name].value_counts().values.tolist(),
        x=[str(i) for i in df[column_name].value_counts().
        ↪index],
        text=df[column_name].value_counts().values.tolist(),
        textfont=dict(size=14),
        textposition='auto',
        showlegend=False,
        marker=dict(color=constraints,
            line=dict(color='#DBE63C',
                width=1))),
        row=1,col=1)
    fig.add_trace(go.Pie(labels=df[column_name].value_counts().keys(),
        values=df[column_name].value_counts().values,
        textfont=dict(size=20),
```

```

        textposition='auto',
        showlegend=False,
        name=column_name,
        marker=dict(colors=constraints)),
        row=1,col=2)
fig.update_layout(title={'text':column_name,
                        'yanchor':'top'},
                  template='plotly_white')

iplot(fig)

```

```
[50]: categorical_variable_summary(df, 'overall')
```

```
[51]: df.reviewText.head()
```

```
[51]: 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
```

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

```
[53]: '[[ 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 capacities up to 32GB and is formatted 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 SDSAQUA-064G-AFFP-A vs SDSAQUA-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 SDSAQUA-064G-u46A and it works like charm.'

```
[54]: review_example=re.sub("[^a-zA-Z]",'',review_example)
review_example
```

```
[54]: 'UPDATESomylovelywifeboughtmeaSamsungGalaxyTabforFathersDayandIvebeenlovingiteve
rsinceJustasotherwithSamsungproductstheGalaxyTabhastheabilitytoaddamicroSDcardto
expandthememoryonthedeviceSinceitsbeenoverayearIdecidedtodosomemoreresearchtosee
ifSanDiskofferedanythingnewAsoftheirproductlineupformicroSDcardsfromworsttobestp
erformancewisearetheasfollowsSanDiskSanDiskUltraSanDiskUltraPLUSSanDiskExtremeSa
nDiskExtremePLUSSanDiskExtremePRONowthedifferencebetweenallofthesecardsaresimply
thespeedinwhichyoucanreadwritedata to the card Yesthe published rating of most all the se ca
rdsexcepttheSanDiskregularareClassUHSIbutthatsjustaratingActualrealworldperforma
ncedoesgetbetterwith each model but with faster cards come more expensive prices Since Amazo
ndoesnt carry the Ultra PLUS model of microSD card I had to do direct comparisons between the San
Disk Ultra Extreme and Extreme PLUS as mentioned in my earlier review I purchased the San Disk Ul
tra 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 San Disk s and most fastest microSD card The Ultra
works perfectly fine for my style of use storing capturing pictures HD video and movie playba
ck on my phone So in the end I ended up just buying another San Disk Ultra GB card I use my cell phone m
ore than I do my tablet and if the card is good enough for my phone it's good enough for my tablet I do
n't own a KHD 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 San Disk Ultra cards and have absolutely no issues with it in my S
amsung devices ORIGINAL REVIEW I haven't had to buy a microSD card in a long time The last time I bou
ght one was for my cell phone over years ago But since my cellular contract was up I knew I would hav
e to get a new card in addition to my new phone the Samsung Galaxy S Reason for this is because I kn
ew my small GB microSD card was not going to cut it Doing research on the Galaxy S I wanted to get the b
est card possible that had decent capacity GB or greater This led me to find that the Galaxy S supp
ort the microSDXC Class UHS I 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 make these microSD
XC Class UHS cards They are San Disk the majority Samsung and Lexar Nobody else makes these that a
resold on Amazon Seeing how San Disk is a pretty good name out of the I've used them the most I decide
d upon the San Disk 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 San Disk you literally
get DOZENS of options All of them have different model numbers different sizes etc Then there's
that confusion of what the difference between SDHC SDXC SDHC vs SDXC SDHC stands for Secure Digi
tal High Capacity and SDXC stands for Secure Digital eXtended Capacity Essentially these two c
ards are the same with the exception that SDHC only supports capacities up to 32GB and is formatted wi
th the FAT file system The SDXC cards are formatted with the exFAT file system If you use a SDXC ca
rd in a device it must support that file system otherwise it may not be recognizable and or you hav
e to reformat the card to FAT FATvsexFAT The differences between the two file systems mean that
FAT has a maximum file size of 4GB limited by that file system exFAT on the other hand supports file
```

sizes upto TB terabytes The only thing you need to know here really is that it's possible your device does not support exFAT. If that's the case, just reformat it to FAT. REMEMBER FORMATTING ERASES ALL DATA. To clarify the model numbers, I hopped over to the SanDisk official webpage. What I found there is that they offer two high-speed options for SanDisk cards. These are SanDisk Extreme Pro and SanDisk Ultra. SanDisk Extreme Pro is a line that supports read speeds up to 150 MB/s. However, they are SDHC only. To make things worse, they are currently only available in 16GB 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 there is confusion set in because SanDisk separates these cards into two different devices: Camera and Mobile. 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 ranges 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 web pages and look at the 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 haven't run any specific testing, but copying GB worth of data from my PC to the card literally took just a few minutes. One last note: I see 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 you're wondering what I got and want to use it in your Galaxy S, I got the SDSDQUAGUA and it works like charm.

```
[62]: review_example=review_example.str.lower().split()
```

```
-----
AttributeError                                Traceback (most recent call last)
Cell In[62], line 1
----> 1 review_example=review_example.str.lower().split()

AttributeError: 'list' object has no attribute 'str'
```

```
[59]: review_example
```

```
[59]: ['updates my lovely wife bought me a Samsung Galaxy Tab for Father's Day and I've 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 it's 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 performance wise are the as follows: SanDisk, SanDisk Ultra, SanDisk Ultra Plus, and SanDisk Extreme. SanDisk Extreme Plus and SanDisk Extreme Pro now the difference between all of these cards is 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-leading, 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 Extreme and Extreme Plus as mentioned in my earlier review. I purchased the SanDisk Ultra for my Galaxy. My question was: did I want to pay over more for a card that is faster than the one I already owned or could I pay almost double to get SanDisk's most fastest microSD card? The Ultra works perfectly fine for my style of usage: storing, capturing pictures, HD video, and movie playback.
```

ackonmyphonesointheendiendedupjustbuyinganothersandiskultragbcardiusemycellphone morethanidomytabletandifthecardisgoodenoughformyphoneitsgoodenoughformytabletido ntownakhdcameraoranythinglikethatsoihonestlydidntseeaneedtogetoneofthefastercard satthistimeiamnowaproudownerofsandiskultracardsandhaveabsolutelyissueswithitinmy samsungdevicesoriginalreviewihaventhadtobuyamicrosdcardinalongtimethelasttimeibo ughtonewasformycellphoneoveryearsagobutsincemycellularcontractwasupiknewiwouldha vetogetanewercardinadditiontomynewphonethesamsunggalaxysreasonforthisisbecauseik newmysmallgbmicrosdcardwasntgoingtocutitdoingresearchonthegalaxysiwantedtogetthe bestcardpossiblethathaddecentcapacitygborgreaterthisledmetofindthatthegalaxyssup portsthemicrosdxccllassuhsicardwhichisthefastestpossiblegiventhatclasssearchingfo rthatspecificallyonamazongavemerestultsofonlyvendorsasofaprilthatmakesthesemicros dxccllassuhsicardstheyaresandiskthemaajoritysamsungandlexarnobodyelsemakesthesethat aresoldonamazonseeinghowsandiskisaprettygoodnameoutoftheiveusedthemthemostidecid eduponthesandiskbecauselexarwasoverpricedandthesamsungonewasoverpricedaswellasno teligibleforamazonprimebutthescarythingisthatwhenyoufilterbythesandiskyouliteral lygetdozensofoptionsallofthemhavedifferentmodelnumbersdifferentsizesetcthen there sthatconfusionofwhatsthe differencebetween sdhcsdxcsdhcvssdxcsdhcstandforsecuredig italhighcapacityandsdxcstandsforsecuredigitalextendedcapacityessentiallythesetwo cardsarethesamewiththeexceptionthatsdhonlysupportscapcitiesuptogbandisformattedw iththefatfilesyste thesdxc cards are formatted with the exfat file system if you use an sdxc ardinadevice it must support that file system otherwise it may not be recognizable and you ha vetoreformatthecardtofatfatvsexfatthedifferencesbetweenthe two file systems meanstha tfathasamaximumfilesizeofgblimitedbythatfile systemexfatontheotherhandssupportsfil esizesuptotbterabyte the only thing you need to know where really is that it is possible your dev icedoesnt support exfat if that is the case just reformat it to fat remember formatting erases al l data to clarify the model numbers i hopped over to the sandisk official web page what i found th ere is that they offer two high speed options for sandisk cards these are sandisk extreme pro and sandisk ultra sandisk extreme pro is a line that supports read speeds up to 100mb/sec however they ar esdh only to make things worse they are currently only available in gb gb capacities since one o fmy requirements was to have a lot of storage i ruled these out the remaining devices listed on am azons search were the sandisk ultra line but there is confusion sets in because sandisk separate st hesecardstotwodifferent devices camera 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 differen cebetween the two range from a couple cents to a few dollars since i wasnt sure i opted for the ones pecifically 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 web pages sandisk command compare the exist ing product line up from there you get exact model numbers and you can then search amazon for the s emodel number that is how i got mines dsd quagas for speed tests i havent run any specific testin gb but copying gb worth of data from my pct to the card literally took just a few minutes one last note isthatamazonattachesadditionalcharacterstotheendforexamplesdsdquagaffpavssdsdqua guathedifferencebetweenthe two isthatthe affpameansamazon frustration free packagingot herthanthattheseareexactlythesameifyourewonderingwhatigotandwanttouseitinyourg al axysigotthesdsdquaguaanditworkslikecharm']

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

```
df.head()
```

```
[63]:
```

	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

```
[67]: pip install vaderSentiment
```

Collecting vaderSentiment

Obtaining dependency information for vaderSentiment from <https://files.pythonhosted.org/packages/76/fc/310e16254683c1ed35eeb97386986d6c00bc29df17ce280aed64d55537e9/vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata>

Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572 bytes)

Requirement already satisfied: requests in c:\users\sujal

ojha\anaconda3\lib\site-packages (from vaderSentiment) (2.31.0)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\sujal

ojha\anaconda3\lib\site-packages (from requests->vaderSentiment) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\users\sujal

ojha\anaconda3\lib\site-packages (from requests->vaderSentiment) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\sujal

ojha\anaconda3\lib\site-packages (from requests->vaderSentiment) (1.26.16)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\sujal

ojha\anaconda3\lib\site-packages (from requests->vaderSentiment) (2024.2.2)

Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)


```

----- 0.0/126.0 kB ? eta -:-:--
--- ----- 10.2/126.0 kB ? eta -:-:--
----- 41.0/126.0 kB 487.6 kB/s eta 0:00:01
----- 126.0/126.0 kB 1.1 MB/s eta 0:00:00

```

Installing collected packages: vaderSentiment

Successfully installed vaderSentiment-3.3.2

Note: you may need to restart the kernel to use updated packages.

```

[70]: from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
df[['polarity', 'subjectivity']] = df['reviewText'].apply(lambda Text: pd.
    ↳Series(TextBlob(Text).sentiment))
for index, row in df['reviewText'].items():
    score = SentimentIntensityAnalyzer().polarity_scores(row)
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']
    if neg > pos:
        df.loc[index, 'sentiment'] = "Negative"
    elif pos > neg:
        df.loc[index, 'sentiment'] = "Positive"
    else:
        df.loc[index, 'sentiment'] = "Neutral"

```

```

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

```

```

[72]:
      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	polarit	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

```
[74]: categorical_variable_summary(df, 'sentiment')
```

```
[ ]:
```