

In [2]:

```
pip install wordcloud
```

```
Collecting wordcloud
  Note: you may need to restart the kernel to use updated packages.
    Downloading wordcloud-1.9.2-cp39-cp39-win_amd64.whl (153 kB)
Requirement already satisfied: pillow in c:\users\hp\anaconda3\lib\site-packages (from wordcloud) (8.4.0)
Requirement already satisfied: numpy>=1.6.1 in c:\users\hp\anaconda3\lib\site-packages (from wordcloud) (1.20.3)
Requirement already satisfied: matplotlib in c:\users\hp\anaconda3\lib\site-packages (from wordcloud) (3.4.3)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.4)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: cycler>=0.10 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.3.1)
Requirement already satisfied: six in c:\users\hp\anaconda3\lib\site-packages (from cycl
er>=0.10->matplotlib->wordcloud) (1.16.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.9.2
```

In [4]:

```
pip install cufflinks
```

```
Collecting cufflinks
  Downloading cufflinks-0.17.3.tar.gz (81 kB)
Requirement already satisfied: numpy>=1.9.2 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (1.20.3)
Requirement already satisfied: pandas>=0.19.2 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (1.3.4)
Collecting plotly>=4.1.1
  Downloading plotly-5.15.0-py2.py3-none-any.whl (15.5 MB)
Requirement already satisfied: six>=1.9.0 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (1.16.0)
Collecting colorlover>=0.2.1
  Downloading colorlover-0.3.0-py3-none-any.whl (8.9 kB)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (58.0.4)
Requirement already satisfied: ipython>=5.3.0 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (7.29.0)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\users\hp\anaconda3\lib\site-packages (from cufflinks) (7.6.5)
Requirement already satisfied: decorator in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.1.0)
Requirement already satisfied: backcall in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: pickleshare in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.7.5)
Requirement already satisfied: pygments in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (2.10.0)
Requirement already satisfied: jedi>=0.16 in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.18.0)
Requirement already satisfied: traitlets>=4.2 in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (5.1.0)
Requirement already satisfied: prompt-toolkit!=3.0.0,!>=3.0.1,<3.1.0,>=2.0.0 in c:\users
```

```
\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (3.0.20)
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.4.4)
Requirement already satisfied: matplotlib-inline in c:\users\hp\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.1.2)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\hp\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (5.1.3)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\hp\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (1.0.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\users\hp\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (3.5.1)
Requirement already satisfied: ipython-genutils~0.2.0 in c:\users\hp\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (0.2.0)
Requirement already satisfied: ipykernal>=4.5.1 in c:\users\hp\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (6.4.1)
Requirement already satisfied: tornado<7.0,>=4.2 in c:\users\hp\anaconda3\lib\site-packages (from ipykernal>=4.5.1->ipywidgets>=7.0.0->cufflinks) (6.1)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\users\hp\anaconda3\lib\site-packages (from ipykernal>=4.5.1->ipywidgets>=7.0.0->cufflinks) (1.4.1)
Requirement already satisfied: jupyter-client<8.0 in c:\users\hp\anaconda3\lib\site-packages (from ipykernal>=4.5.1->ipywidgets>=7.0.0->cufflinks) (6.1.12)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\hp\anaconda3\lib\site-packages (from jedi>=0.16->ipython>=5.3.0->cufflinks) (0.8.2)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\hp\anaconda3\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (2.8.2)
Requirement already satisfied: pyzmq>=13 in c:\users\hp\anaconda3\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (22.2.1)
Requirement already satisfied: jupyter-core>=4.6.0 in c:\users\hp\anaconda3\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (4.8.1)
Requirement already satisfied: pywin32>=1.0 in c:\users\hp\anaconda3\lib\site-packages (from jupyter-core>=4.6.0->jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets>=7.0.0->cufflinks) (228)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\hp\anaconda3\lib\site-packages (from nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (3.2.0)
Requirement already satisfied: attrs>=17.4.0 in c:\users\hp\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (21.2.0)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\users\hp\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets>=7.0.0->cufflinks) (0.18.0)
Requirement already satisfied: pytz>=2017.3 in c:\users\hp\anaconda3\lib\site-packages (from pandas>=0.19.2->cufflinks) (2021.3)
Requirement already satisfied: packaging in c:\users\hp\anaconda3\lib\site-packages (from plotly>=4.1.1->cufflinks) (21.0)
Collecting tenacity>=6.2.0
    Downloading tenacity-8.2.2-py3-none-any.whl (24 kB)
Requirement already satisfied: wcwidth in c:\users\hp\anaconda3\lib\site-packages (from prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0->ipython>=5.3.0->cufflinks) (0.2.5)
Requirement already satisfied: notebook>=4.4.1 in c:\users\hp\anaconda3\lib\site-packages (from widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (6.4.5)
Requirement already satisfied: terminado>=0.8.3 in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.9.4)
Requirement already satisfied: Send2Trash>=1.5.0 in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.8.0)
Requirement already satisfied: prometheus-client in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.11.0)
Requirement already satisfied: nbconvert in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (6.1.0)
```

```
Requirement already satisfied: argon2-cffi in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (20.1.0)
Requirement already satisfied: jinja2 in c:\users\hp\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.11.3)
Requirement already satisfied: pywinpty>=0.5 in c:\users\hp\anaconda3\lib\site-packages (from terminado>=0.8.3->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.7)
Requirement already satisfied: cffi>=1.0.0 in c:\users\hp\anaconda3\lib\site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.14.6)
Requirement already satisfied: pycparser in c:\users\hp\anaconda3\lib\site-packages (from cffi>=1.0.0->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (2.20)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\hp\anaconda3\lib\site-packages (from jinja2->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.1.1)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.4.3)
Requirement already satisfied: testpath in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.0)
Requirement already satisfied: jupyterlab-pygments in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.1.2)
Requirement already satisfied: defusedxml in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.3)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.3)
Requirement already satisfied: bleach in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (4.0.0)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\hp\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.8.4)
Requirement already satisfied: async-generator in c:\users\hp\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.10)
Requirement already satisfied: nest-asyncio in c:\users\hp\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (1.5.1)
Requirement already satisfied: webencodings in c:\users\hp\anaconda3\lib\site-packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.0.0->cufflinks) (0.5.1)
Requirement already satisfied: pyparsing>=2.0.2 in c:\users\hp\anaconda3\lib\site-packages (from packaging->plotly>=4.1.1->cufflinks) (3.0.4)
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=68734 sha256=1ca53d8072fcc8c0bd53601994d8e12e7e93757254fe6db90b3ef06f2f8989a7
  Stored in directory: c:\users\hp\appdata\local\pip\cache\wheels\29\b4\f8\2fd2206eeeba6ccad8167e4e8894b8c4ec27bf1342037fd136
Successfully built cufflinks
Installing collected packages: tenacity, plotly, colorlover, cufflinks
```

Successfully installed colorlover-0.3.0 cufflinks-0.17.3 plotly-5.15.0 tenacity-8.2.2
Note: you may need to restart the kernel to use updated packages.

In [42]:

```
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)
```

In [43]:

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

In [44]:

```
df.head()
```

Out[44]:

	Unnamed: 0	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_v
0	0	NaN	4	No issues.	23-07-2014	138	0	0	
1	1	0mie	5	Purchased this for my device, it worked as adv...	25-10-2013	409	0	0	
2	2	1K3	4	it works as expected. I should have sprung for...	23-12-2012	715	0	0	
3	3	1m2	5	This think has worked out great.Had a diff. br...	21-11-2013	382	0	0	

Unnamed: 0	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_vote
4	4 2&1/2Men	5	Bought it with Retail Packaging, arrived legit...	13-07-2013	513	0	0	

In [45]:

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

Out[45]:

	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_vote
2031	Hyoun Kim "Faluzure"	5	[[UPDATE - 6/19/2014]] So my lovely wife boug...	05-01-2013	702	1952	68	2020
3449	NLee the Engineer	5	I have tested dozens of SDHC and micro-SDHC ca...	26-09-2012	803	1428	77	1505
4212	SkincareCEO	1	NOTE: please read the last update (scroll to ...	08-05-2013	579	1568	126	1694
317	Amazon Customer "Kelly"	1	If your card gets hot enough to be painful, it...	09-02-2012	1033	422	73	495
4672	Twister	5	Sandisk announcement of the first 128GB micro ...	03-07-2014	158	45	4	49



In [46]:

```
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'])
    missing_df = pd.DataFrame(missing_df)
    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,'~'))
print(df.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~~~~~

|              | 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    |

In [47]:

```

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)

```

Out[47]:

|   | Variable     | Classes |
|---|--------------|---------|
| 0 | reviewText   | 4912    |
| 1 | reviewerName | 4594    |
| 2 | reviewTime   | 690     |

|    | Variable             | Classes |
|----|----------------------|---------|
| 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       |

In [ ]:

```

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 =[[{"types": "xy"}, {"types": '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),
                          name = column_name,
                          textposition = 'auto',
                          showlegend = False,
                          marker = dict (color = constraints,
                                         line=dict(color = '#DBE6EC',
                                                   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 = 18),
                          textposition='auto',
                          showlegend= False,
                          name=column_name,
                          marker =dict(color=constraints)),
                  row =1,col= 2)
    fig.update_layout(title={'text':column_name,
                            'y':0.9,
                            'x':0.5,
                            'xanchor':'center',
                            'yanchor':'top'},
                      template = 'plotly_white')
    iplot(fig)

```

In [ ]:

```
categorical_variable_summary(df , 'overall')
```

In [75]:

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

Out[75]:

```

2031    updatesomylovelywifeboughtmeasamsunggalaxytabf...
3449    i havetesteddozensofsdhccardsdhccardsoneidis...
4212    notepleasereadthe lastupdate scrolltothebottomim...

```

```
317      if your card gets hotenoughtobepainfulitisdefectiv...
4672      sandisk announcement of the first gbmicrosd took into...
Name: reviewText, dtype: object
```

In [76]:

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

Out[76]:

'updatesomylovelywifeboughtmeasamsunggalaxytabforfathersdayandivebeenlovingiteversinceju  
stasotherwithsamsungproductsthegalaxytabhastheabilitytoaddamicrosdcardtoexpandthememoryo  
nthedevicessinceitsbeenoverayearidecidedtodosomemoreresearchtoseeifsandiskofferedanything  
newasoftheirproductlineupformicrosdcardsfromworsttobestperformancewisearetheasfollowssan  
disksandiskultrasandiskultraplussandiskextremesandiskextremeplussandiskextremepronowthed  
ifferencebetweenallofthesecardsaresimplythespeedinwhichyoucanreadwritedataothecardyesth  
epublishedratingofmostallthesecardsexceptthesandiskregularareclassuhsibusbutthatsjustaratin  
gactualrealworldperformancedoesgetbetterwitheachmodelbutwithfastercardscomemoreexpensive  
processsinceamazondoesntcarrytheultraplusmodelofmicrosdcardihadtododirectcomparisonsbetwe  
enthesandiskultraextremeandextremeplusasmentionedinmyearlierreviewipurchasedthesandiskul  
traformmygalaxysmyquestionwasdidiwanttopayovermoreforacardthatisfasterthantheoneialreadyo  
wnedoricouldpayalmostdoubletogetsandisksndmostfastestmicrosdcardtheultraworksperfectlyfi  
neformystyleofusagetostringcapturingpictureshdvideoandmovieplaybackonmyphonesintheendien  
dedupjustbuyinganothersandiskultragbcardiusemycellphonemorethanidomitabletandifthecardis  
goodenoughformyphoneitsgoodenoughformytabletidonotownakhdcameraoranythinglikethatsoihones  
tlydidntseeaneedtogetoneofthefastercardsatthistimeiamnowaprudentownerofsandiskultracardsan  
dhavеabsolutelyissueswithitinmysamsungdevicesoriginalreviewihaventhadtobuyamicrosdcardin  
alongtimethelasttimeiboughtonewasformycellphoneoveryearsagobutsincemycellularcontractwas  
upiknewiwouldhavetogetanewercardinadditiontomynewphonethesamsunggalaxysreasonforthisisbe  
causeiknewmysmallgbmicrosdcardwasntgoingtocutitdoingresearchonthegalaxysiwantedtogettheb  
estcardpossiblethathaddecentcapacityborgreaterthistledmetofindthatthegalaxyssupportsthem  
icrosdxcclassuhsicardwhichisthefastestpossiblelegiventhatclasssearchingforthatspecifically  
onamazonagavemerelsofonlyvendorsasofaprilthatmakesthesemicrosdxcclassuhsicardstheyaresa  
ndiskthemajoritysamsungandlexarnobodyelsemakesthesethataresoldonamazonseeinghowsandiskis  
aprettygoodnameoutoftheiveusedthemthemostidecidetuponthesandiskbecauselexarwasoverpriced  
andthesamsungonewasoverpricedaswellasnoteligibleforamazonprimebutthescarythingisthatwhen  
youfilterbythesandiskyouliterallygetdozensofoptionsallofthemhavedifferentmodelnumbersdi  
fferentsizesetcuenttherestthatconfusionofwhatthendifferencebetweendshcsdxsdhcvcssdxcdhc  
standforsecuredigitalhighcapacityandsdxctandsforsecuredigitalextendedcapacityessential  
lythesetwocardsarethesamewiththeexceptionthatsdhonlysupportscapcitiesuptogbandisformat  
edwiththefatfilesystemthesdxcardsarereformattedwiththeexfatfilesystemifyouuseansdxcardsi  
nadeviceitmustssupportthatfilesystemotherwiseitmaynotberecognizableandoryouhavetoreforma  
tthecardtofatfatvsexfatthendifferencesbetweenthetwofilesystemsmeansthatfathasamaximumfil  
esizeofgblimitedbythatfilesystemxfatonthetherhandssupportsfilesizesuptotbterabytestheo  
nlythingyouneedtoknowhererereallyisthatitspossibleyourdevicedoesntsupportxfatifthatsthe  
asejustreformatittofatrememberformattingerasessalldatatoclarifythemodelnumbersiihoppedov  
ertothesandiskofficialwebpagewhatifoundthereisthattheyoffertwohighspeedoptionsforsandis  
kcardstheseariesandiskextremeproandsandiskultrasandiskextremeproisalinethatsupportsreads  
peedsuptombsechowevertheyaresdhonlytomakethingsworsestheyarecurrentlyonlyavailableingbg  
bcapacitiessinceoneofmyrequirementswastohavealotofstorageiruledtheseseouttheremainingdevi  
ceslistedonamazonssearchwerethesandiskultralinebuttherconfusionsetsinbecausesandisksepa  
ratesthesecardstotwodifferentdevicescamerasmobiledevicesistherearealdifferencebetweenth  
etwooristhisjustamarketingstuntunfortunatelyimnotsurebutidoknowthepricedifferencebetwee  
nthewororangefromacouplecentstoafewdollarssinceiwasntsureioptedfortheonespecificallytarg  
etedformobiledevicesjustincasethereissomekindofcompatibilityissuetofindtheexactmodelnum  
beriwouldgotosandiskswebpagesandiskcomandcomparetheirexistingproductlineupfromthereyoug  
etexactmodelnumbersandyoucanthensearchamazonforthesemodelnumbersthatishowigotminesdsdq  
ugasforspeedtestsihaventrunkanyspecifictestingbutcopyinggbworthofdatafrommypctothecardli  
terallytookjustafewminutesonlastnoteisthatamazonattachesadditionalcharacterstotheendfo  
rexamplesdsdqquagaffpavssdsdqaguathedifferencebetweenthetwoisthattheaffpameansamazonfru  
strationfreepackagingotherthanthattheseareexactlythesameifyourewonderingwhatigotandwant  
touseitinyourgalaxysigotthesdsdqquaguanditworkslikecharm'

```
In [62]: review_example = re.sub("[^a-zA-Z]", ' ', review_example)
review_example
#all punctuation and numbers are removed from data
```

```
Out[62]: 'UPDATESomylovelywifeboughtmeaSamsungGalaxyTabforFathersDayandIvebeenlovingiteversinceJu
stasotherwithSamsungproductstheGalaxyTabhastheabilitytoaddamicroSDcardtoexpandthememoryo
nthedeviceSinceitsbeenoverayearIdecidedtodosomemoreresearchtoseeifSanDiskofferedanything
newAsoftheirproductlineupformicroSDcardsfromworsttobestperformancewisearetheasfollowsSan
DiskSanDiskUltraSanDiskUltraPLUSSanDiskExtremeSanDiskExtremePLUSSanDiskExtremePRONowthed
ifferencebetweenallofthesecardsaresimplythespeedinwhichyoucanreadwritedataothecardYesth
epublishedratingofmostallofthesecardsexcepttheSanDiskregularareClassUHSIbutthatsjustaratin
gActualrealworldperformancedoesgetbetterwitheachmodelbutwithfastercardscomemoreexpensive
pricesSinceAmazondoesntcarrytheUltraPLUSmodelofmicroSDcardIhadtododirectcomparisonsbetwe
entheSanDiskUltraExtremeandExtremePLUSAmentionedinmyearlierreviewIpurchasedtheSanDiskUl
traformyGalaxySMyquestionwasdidIwanttopayovermoreforacardthatisfasterthantheoneIalreadyo
wnedOrIcouldpayalmostdoubletogetSanDisksmostfastestmicroSDcardTheUltraworksperfectlyfi
neformystyleofusagestoringcapturingpicturesHDvideoandmovieplaybackonmyphoneSointheendIen
dedupjustbuyinganotherSanDiskUltraGBcardIusemycellphonemorethanIdomytabletandifthecardis
goodenoughformyphoneitsgoodenoughformytabletIdontownaKHDcameraoranythinglikethatsoIphones
tlydidntseeaneedtogetoneofthefastercardsatthistimeIamnowaproudownerofSanDiskUltracardsan
dhaveabsolutelyissueswithitinmySamsungdevicesORIGINALREVIEWIhaventhadtobuyamicroSDcardin
alongtimeThelasttimeIboughtonewasformycellphoneoveryearsagoButsinceImycellularcontractwas
upIknewIwouldhavetogetanewercardinadditiontomymewhonetheSamsungGalaxySReasonforthisisbe
causeIknewmysmall1GBmicroSDcardwasntgoingtocutitDoingresearchontheGalaxySIwantedtogetthel
estcardpossiblethathaddecentcapacityGBorgreaterThisledmetofindthattheGalaxySsupportsthem
icroSDXCClassUHSIcardwhichisthefastestpossiblegiventhatclassSearchingforthatspecifically
onAmazongavemeresultsofonlyvendorsasofofAprilthatmakesthesemicroSDXCClassUHScardsTheyareSa
ndiskthemajoritySamsungandLexarNobodyelsemakesthesethataresoldonAmazonSeeinghowSanDiskis
aprettygoodnameoutoftheIveusedthemthemostIdecidedupontheSanDiskbecauseLexarwasoverpriced
andtheSamsungonewasoverpricedaswellasnoteligibleforAmazonPrimeButthescarythingisthatwhen
youfilterbytheSanDiskyouliterallygetDOZENsofoptionsAllofthemhavedifferentmodelnumbersdi
fferentsizesetcThentheresthatconfusionofwhatsthedifferencebetweenSDHCSDXCSDHCvsSDXCSDHC
standforSecureDigitalHighCapacityandSDXCstandsforSecureDigitaleXtendedCapacityEssential
lythesetwocardsaresameswiththeexceptionthatSDHConlysupportscapcitiesuptoGBandisformat
edwiththeFATfilesystemTheSDXCcardsarereformattedwiththeexFATfilesystemIfyouuseanSDXCcardi
nadeviceitmustssupportthatfilesystemotherwiseitmaynotberecognizableandoryouhavetoreforma
tthecardtoFATFATvsexFATThedifferencesbetweenthetwofilesystemsmeansthatFAThasamaximumfil
esizeofGBlimitedbythatfilesystemexFATontheotherhandsupportsfilesizesuptoTBterabytesTheo
nlythingyouneedtoknowhererellyisthatitspossibleyourdevicedoesntsupportexFATIfthec
asejustreformatittoFATREMEMBERFORMATTINGERASESALLDATATOclarifythemodelnumbersIIhoppedov
ertotheSanDiskofficialwebpageWhatIfoundtheresthattheyoffertwohighspeedoptionsforSanDis
kcardsTheseareSanDiskExtremeProandSanDiskUltraSanDiskExtremeProisanethatsupportsreads
peedsuptoMBsechowevertheyareSDHConlyTomakethingsworsetheyarecurrentlyonlyavailableinGBG
BcapacitiesSinceoneofmyrequirementswastohavealotofstorageIruledtheseoutTheremainingdevi
ceslistedonAmazonssearchweretheSanDiskUltralineButthereconfusionsetsinbecauseSanDisksepa
ratesthesecardstotwodifferentdevicesCamerasmobiledevicesIstherearealldifferencebetweenth
etwooristhisjustamarketingstuntUnfortunatelyImnotsurebutIdoknowthepricedifferencebetwee
nthetworangefromacouplecentstoafewdollarsSinceIwasntsureIoptedfortheonespecificallytarg
etedformobiledevicesjustincasethereissomekindofcompatibilityissueTofindtheexactmodelnum
berIwouldgotoSandiskswebpagesandiskcomandcomparetheirexistingproductlineupFromthereyoug
etexactmodelnumbersandyoucanthensearchAmazonforthesemodelnumbersThatishowIgotmineSDSDQU
AGAsforspeedtestsIhavenrunanyspecifictestingbutcopyingGBworthofdatafrommyPCtothecardli
terallytookjustafewminutesOnelastnoteisthatAmazonattachesadditionalcharacterstotheendfo
rexampleSDSDQUAGAFFPAvsSDSDQUAGUAThedifferencebetweenthetwoisthattheAFFPmeansAmazonFru
strationFreePackagingOtherthanthattheseareexactlythesameIfyourewonderingwhatIgotandwant
touseitinyourGalaxySIgottheSDSDQUAGuAanditworkslikecharm'
```

```
In [ ]:
```

In [77]:

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

Out[77]:

|      | reviewerName                  | overall | reviewText                                           | reviewTime | day_diff | h... |
|------|-------------------------------|---------|------------------------------------------------------|------------|----------|------|
| 2031 | Hyoun Kim<br>"Faluzure"       | 5       | updatesomylovelywifeboughtmeasamsunggalaxytabf...    | 05-01-2013 | 702      |      |
| 3449 | NLee the<br>Engineer          | 5       | i havetesteddozensofsdhcandmicrosdhccardsoneidis...  | 26-09-2012 | 803      |      |
| 4212 | SkincareCEO                   | 1       | noteplease readthelastupdate scrolltothebottomim...  | 08-05-2013 | 579      |      |
| 317  | Amazon<br>Customer<br>"Kelly" | 1       | if your card getshotenoughtobepainfulitisdefectiv... | 09-02-2012 | 1033     |      |
| 4672 | Twister                       | 5       | sandiskannouncemetnofthefirstgbmicrosdtookinte...    | 03-07-2014 | 158      |      |

```
from sentiment.vader import SentimentIntensityAnalyzer
df[['polarity','subjectivity']] = df['reviewText'].apply(lambda Text:pd.Series(TextBlob(Text).sentiment))
for index , row in df['reviewText'].iterrows():

    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"
```

In [84]:

```
df[df["sentiment"] == "Positive"].sort_values("wilson_lower_bound",ascending = False).head(10)
```

Out[84]:

|      | reviewerName         | overall | reviewText | reviewTime | day_diff | helpful_yes | helpful_no | total_vote | scor... |
|------|----------------------|---------|------------|------------|----------|-------------|------------|------------|---------|
| 3234 | minh thong<br>cao    | 5       | good       | 07-07-2014 | 154      | 0           | 0          | 0          |         |
| 3605 | peter Metcalf        | 5       | perfect    | 07-02-2014 | 304      | 0           | 0          | 0          |         |
| 4308 | Stephane<br>Gauthier | 5       | super      | 15-02-2013 | 661      | 0           | 0          | 0          |         |
| 3741 | RASHAWN              | 5       | great      | 07-12-2014 | 1        | 0           | 0          | 0          |         |
| 705  | Brandon<br>Warren    | 5       | yes        | 14-07-2014 | 147      | 0           | 0          | 0          |         |

In [79]:

```
pip install vaderSentiment
```

Requirement already satisfied: vaderSentiment in c:\users\hp\anaconda3\lib\site-packages (3.3.2)  
Note: you may need to restart the kernel to use updated packages.

Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\site-packages (from vaderSentiment) (2.26.0)  
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\hp\anaconda3\lib\site-packages (from requests->vaderSentiment) (1.26.7)  
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\lib\site-packages (from requests->vaderSentiment) (3.2)  
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hp\anaconda3\lib\site-packages (from requests->vaderSentiment) (2021.10.8)  
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\hp\anaconda3\lib\site-packages (from requests->vaderSentiment) (2.0.4)

In [80]:

```
pip install nltk
```

Requirement already satisfied: nltk in c:\users\hp\anaconda3\lib\site-packages (3.6.5)  
Requirement already satisfied: click in c:\users\hp\anaconda3\lib\site-packages (from nltk) (8.0.3)  
Requirement already satisfied: joblib in c:\users\hp\anaconda3\lib\site-packages (from nltk) (1.1.0)  
Requirement already satisfied: regex>=2021.8.3 in c:\users\hp\anaconda3\lib\site-packages (from nltk) (2021.8.3)  
Requirement already satisfied: tqdm in c:\users\hp\anaconda3\lib\site-packages (from nltk) (4.62.3)  
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-packages (from click->nltk) (0.4.4)

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

In [68]:

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
df[['polarity','subjectivity']] = df['reviewText'].apply(lambda Text:pd.Series(TextBlob
for index , row in df['reviewText'].iteritems():

    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"
```

In [69]:

```
df[df["sentiment"] == "Positive"].sort_values("wilson_lower_bound",ascending = False).h
```

Out[69]:

|      | reviewerName   | overall | reviewText | reviewTime | day_diff | helpful_yes | helpful_no | total_vote | score |
|------|----------------|---------|------------|------------|----------|-------------|------------|------------|-------|
| 3234 | minh thong cao | 5       | good       | 07-07-2014 | 154      | 0           | 0          | 0          |       |
| 3605 | peter Metcalf  | 5       | perfect    | 07-02-2014 | 304      | 0           | 0          | 0          |       |

|      | reviewerName      | overall | reviewText | reviewTime | day_diff | helpful_yes | helpful_no | total_vote | scor |
|------|-------------------|---------|------------|------------|----------|-------------|------------|------------|------|
| 4308 | Stephane Gauthier | 5       | super      | 15-02-2013 | 661      | 0           | 0          | 0          | 0    |
| 3741 | RASHAWN           | 5       | great      | 07-12-2014 | 1        | 0           | 0          | 0          | 0    |
| 705  | Brandon Warren    | 5       | yes        | 14-07-2014 | 147      | 0           | 0          | 0          | 0    |

In [ ]: categorical\_variable\_summary(df, 'sentiment')

In [ ]:

In [ ]: