In [1]:

```
pip install textblob
```

```
Requirement already satisfied: textblob in c:\users\hp\anaconda3\lib\site-
packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in c:\users\hp\anaconda3\lib\site
-packages (from textblob) (3.7)
Requirement already satisfied: click in c:\users\hp\anaconda3\lib\site-pac
kages (from nltk>=3.1->textblob) (8.0.4)
Requirement already satisfied: tqdm in c:\users\hp\anaconda3\lib\site-pack
ages (from nltk>=3.1->textblob) (4.64.1)
Requirement already satisfied: regex>=2021.8.3 in c:\users\hp\anaconda3\li
b\site-packages (from nltk>=3.1->textblob) (2022.7.9)
Requirement already satisfied: joblib in c:\users\hp\anaconda3\lib\site-pa
ckages (from nltk>=3.1->textblob) (1.1.1)
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-
packages (from click->nltk>=3.1->textblob) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:

```
pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\hp\anaconda3\lib\site
-packages (1.9.2)
Requirement already satisfied: matplotlib in c:\users\hp\anaconda3\lib\sit
e-packages (from wordcloud) (3.7.0)
Requirement already satisfied: numpy>=1.6.1 in c:\users\hp\anaconda3\lib\s
ite-packages (from wordcloud) (1.23.5)
Requirement already satisfied: pillow in c:\users\hp\anaconda3\lib\site-pa
ckages (from wordcloud) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\hp\anaconda3\l
ib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\hp\anaconda3\l
ib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\hp\anaconda3\lib\s
ite-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: packaging>=20.0 in c:\users\hp\anaconda3\li
b\site-packages (from matplotlib->wordcloud) (22.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hp\anacond
a3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3
\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: six>=1.5 in c:\users\hp\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:

```
pip install cufflinks
```

```
Requirement already satisfied: cufflinks in c:\users\hp\anaconda3\lib\sit
e-packages (0.17.3)
Requirement already satisfied: pandas>=0.19.2 in c:\users\hp\anaconda3\li
b\site-packages (from cufflinks) (1.5.3)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\users\hp\anaconda3
\lib\site-packages (from cufflinks) (7.6.5)
Requirement already satisfied: ipython>=5.3.0 in c:\users\hp\anaconda3\li
b\site-packages (from cufflinks) (8.10.0)
Requirement already satisfied: six>=1.9.0 in c:\users\hp\anaconda3\lib\si
te-packages (from cufflinks) (1.16.0)
Requirement already satisfied: colorlover>=0.2.1 in c:\users\hp\anaconda3
\lib\site-packages (from cufflinks) (0.3.0)
Requirement already satisfied: plotly>=4.1.1 in c:\users\hp\anaconda3\lib
\site-packages (from cufflinks) (5.9.0)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\hp\anaconda
3\lib\site-packages (from cufflinks) (65.6.3)
Requirement already satisfied: numpy>=1.9.2 in c:\users\hp\anaconda3\lib
\site-packages (from cufflinks) (1.23.5)
Requirement already satisfied: backcall in c:\users\hp\anaconda3\lib\site
```

In [62]:

```
pip install vaderSentiment
```

```
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
     ------------------------------------ 126.0/126.0 kB 7.2 MB/s eta 0:
00:00
Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\site-
packages (from vaderSentiment) (2.28.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\hp\anacon
da3\lib\site-packages (from requests->vaderSentiment) (1.26.14)
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\lib\s
ite-packages (from requests->vaderSentiment) (3.4)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\hp\ana
conda3\lib\site-packages (from requests->vaderSentiment) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hp\anaconda3
\lib\site-packages (from requests->vaderSentiment) (2023.5.7)
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
Note: you may need to restart the kernel to use updated packages.
```

In [4]:

```
pip install plotly
```

```
Requirement already satisfied: plotly in c:\users\hp\anaconda3\lib\site-pa
ckages (5.9.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp\anaconda3\li
b\site-packages (from plotly) (8.0.1)
Note: you may need to restart the kernel to use updated packages.
```

In [5]:

```
pip install plotly-orca
```

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

ERROR: Could not find a version that satisfies the requirement plotly-orca
(from versions: none)
ERROR: No matching distribution found for plotly-orca
```

In [6]:

```
pip install warnings
```

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

ERROR: Could not find a version that satisfies the requirement warnings (f
rom versions: none)
ERROR: No matching distribution found for warnings
```

In [17]:

```
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

pd.set_option('display.max_columns', None)
```

In [18]:

```
#importing the data Set

df = pd.read_csv("C:/Users/Hp/Desktop/python/amazon.csv")
```

In [19]:

```
df.head()
```

Out[19]:

	Unnamed: 0	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_nc
0	0	NaN	4	No issues.	23-07-2014	138	0	(
1	1	0mie	5	Purchased this for my device, it worked as adv...	25-10-2013	409	0	(
2	2	1K3	4	it works as expected. I should have sprung for...	23-12-2012	715	0	(
3	3	1m2	5	This think has worked out great.Had a diff. br...	21-11-2013	382	0	(
4	4	2&1/2Men	5	Bought it with Retail Packaging, arrived legit...	13-07-2013	513	0	(

In [20]:

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

Out[20]:

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

In [21]:

```python
#make a function for missing values

import pandas as pd

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=Tru
    missing_df = pd.concat([n_miss, pd.Series(ratio, name='Ratio')], axis=1)
    missing_df = pd.DataFrame(missing_df)
    return missing_df


# Assuming you have defined the missing_values_analysis function

def check_dataframe(df):
    separator = '~' * 82  # Create a line of '~' characters
    print('SHAPE', separator)
    print('Rows: {}'.format(df.shape[0]))
    print('Columns: {}'.format(df.shape[1]))
    print("TYPES".center(82, '~'))
    print(df.dtypes)
    print("".center(82, '~'))

    # Assuming you have defined the missing_values_analysis function
    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)

# Assuming you have a DataFrame named 'df' that you want to analyze
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
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
~~~~~~~~
              0     Ratio
reviewerName  1  0.020346
reviewText    1  0.020346
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~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     5.0000
00
day_diff                1.0  98.0  431.0  748.000000  943.00000  1064.0000
00
helpful_yes             0.0   0.0    0.0    1.000000    3.00000  1952.0000
00
helpful_no              0.0   0.0    0.0    0.000000    2.00000   183.0000
00
total_vote              0.0   0.0    0.0    1.000000    4.00000  2020.0000
00
score_pos_neg_diff   -130.0   0.0    0.0    1.000000    2.00000  1884.0000
00
score_average_rating    0.0   0.0    0.0    1.000000    1.00000     1.0000
00
wilson_lower_bound      0.0   0.0    0.0    0.206549    0.34238     0.9575
44

C:\Users\Hp\AppData\Local\Temp\ipykernel_20552\393582109.py:30: FutureWarn
ing:

The default value of numeric_only in DataFrame.quantile is deprecated. In
a future version, it will default to False. Select only valid columns or s
pecify the value of numeric_only to silence this warning.
```

In [22]:

```
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[22]:

	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

In [25]:

```
colors = ['red', 'blue', 'green', 'yellow', 'orange']
def catagorical_variable_summary(df, column_name):
    fig = make_subplots(rows=1, cols=2,
                        subplot_titles=('Countplots', 'Percentage'),
                        specs=[[{"type": "xy"}, {"type": "domain"}]])

    # Add the Bar chart
    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=colors,  # Use 'colors' variable here
                    line=dict(color='red', width=1))
    ), row=1, col=1)

    # Add the Pie chart
    fig.add_trace(go.Pie(
        labels=df[column_name].value_counts().keys(),
        values=df[column_name].value_counts().values,
        textfont=dict(size=18),
        name=column_name,
        textinfo='percent+label',
        showlegend=True,
        marker=dict(line=dict(color='white', width=2))
    ), row=1, col=2)

    fig.update_layout(title={'text': column_name,
                             'y': 0.9,
                             'x': 0.5,
                             'xanchor': 'center',
                             'yanchor': 'top'},
                      template='plotly_white')

    # Use the 'show' method to display the figure
    fig.show()

# Now you can call your function
catagorical_variable_summary(df, 'overall')
```

In [26]:

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

Out[26]:

```
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
```

In [52]:

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

Out[52]:

```
'[[ 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 w
ith Samsung products, the Galaxy Tab 4 has the ability to add a microSD c
ard 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 bes
t (performance-wise) are the as follows:SanDiskSanDisk UltraSanDisk Ultra
PLUSSanDisk ExtremeSanDisk Extreme PLUSSanDisk Extreme PRONow, the differ
ence between all of these cards are simply the speed in which you can rea
d/write data to the card.  Yes, the published rating of most all these ca
rds (except the SanDisk regular) are Class 10/UHS-I but that\'s just a ra
ting... Actual real world performance does get better with each model, bu
t with faster cards come more expensive prices.  Since Amazon doesn\'t ca
rry 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
```

In [53]:

```
# Remove non-alphabetical characters
review_example = re.sub("[^a-zA-Z]", '', review_example)

# Print the modified text
review_example
```

Out[53]:

```
'UPDATESomylovelywifeboughtmeaSamsungGalaxyTabforFathersDayandIvebeenlovin
giteversinceJustasotherwithSamsungproductstheGalaxyTabhastheabilitytoaddam
icroSDcardtoexpandthememoryonthedeviceSinceitsbeenoverayearIdecidedtodosom
emoreresearchtoseeifSanDiskofferedanythingnewAsoftheirproductlineupformicr
oSDcardsfromworsttobestperformancewisearetheasfollowsSanDiskSanDiskUltraSa
nDiskUltraPLUSSanDiskExtremeSanDiskExtremePLUSSanDiskExtremePRONowthediffe
rencebetweenallofthesecardsissimplythespeedinwhichyoucanreadwritedatatoth
ecardYesthepublishedratingofmostallthesecardsexcepttheSanDiskregularareCla
ssUHSIbutthatsjustaratingActualrealworldperformancedoesgetbetterwitheachmo
delbutwithfastercardscomemoreexpensivepricesSinceAmazondoesntcarrytheUltra
PLUSmodelofmicroSDcardIhadtododirectcomparisonsbetweentheSanDiskUltraExtre
meandExtremePLUSAsmentionedinmyearlierreviewIpurchasedtheSanDiskUltraformy
GalaxySMyquestionwasdidIwanttopayovermoreforacardthatisfasterthantheoneIal
readyownedOrIcouldpayalmostdoubletogetSanDisksndmostfastestmicroSDcardTheU
ltraworksperfectlyfineformystyleofusagestoringcapturingpicturesHDvideoandm
ovieplaybackonmyphoneSointheendIendedupjustbuyinganotherSanDiskUltraGBcard
IusemycellphonemorethanIdomytabletandifthecardisgoodenoughformyphoneitsgoo
denoughformytabletIdontownaKHDcameraoranythinglikethatsoIhonestlydidntseea
needtogetoneofthefastercardsatthistimeIamnowaproudownerofSanDiskUltracards
andhaveabsolutelyissueswithitinmySamsungdevicesORIGINALREVIEWIhaventhadtob
uyamicroSDcardinalongtimeThelasttimeIboughtonewasformycellphoneoveryearsag
oButsincemycellularcontractwasupIknewIwouldhavetogetanewercardinadditionto
mynewphonetheSamsungGalaxySReasonforthisisbecauseIknewmysmallGBmicroSDcard
wasntgoingtocutitDoingresearchontheGalaxySIwantedtogetthebestcardpossiblet
hathaddecentcapacityGBorgreaterThisledmetofindthattheGalaxySsupportsthemic
roSDXCClassUHSIcardwhichisthefastestpossiblegiventhatclassSearchingforthat
specificallyonAmazongavemeresultsofonlyvendorsasofAprilthatmakesthesemicro
SDXCClassUHScardsTheyareSandiskthemajoritySamsungandLexarNobodyelsemakesth
esethataresoldonAmazonSeeinghowSanDiskisaprettygoodnameoutoftheIveusedthem
themostIdecideduponthesanDiskbecauseLexarwasoverpricedandtheSamsungonewaso
verpricedaswellasnoteligibleforAmazonPrimeButthescarythingisthatwhenyoufil
terbytheSanDiskyouliterallygetDOZENSofoptionsAllofthemhavedifferentmodelnu
mbersdifferentsizesetcThentheresthatconfusionofwhatsthedifferencebetweenSD
HCSDXCSDHCvsSDXCSDHCstandforSecureDigitalHighCapacityandSDXCstandsforSecur
eDigitaleXtendedCapacityEssentiallythesetwocardsarethesamewiththeexception
thatSDHConlysupportscapcitiesuptoGBandisformatedwiththeFATfilesystemTheSDX
CcardsareformattedwiththeexFATfilesystemIfyouuseanSDXCcardinadeviceitmusts
upportthatfilesystemotherwiseitmaynotberecognizableandoryouhavetoreformatt
hecardtoFATFATvsexFATThedifferencesbetweenthetwofilesystemsmeansthatFAThas
amaximumfilesizeofGBlimitedbythatfilesystemexFATontheotherhandsupportsfile
sizesuptoTBterabytesTheonlythingyouneedtoknowherereallyisthatitspossibleyo
urdevicedoesntsupportexFATIfthatsthecasejustreformatittoFATREMEMBERFORMATT
INGERASESALLDATATOclarifythemodelnumbersIIhoppedovertotheSanDiskofficialwe
bpageWhatIfoundthereisthattheyoffertwohighspeedoptionsforSanDiskcardsThese
areSanDiskExtremeProandSanDiskUltraSanDiskExtremeProisalinethatsupportsrea
dspeedsuptoMBsechowevertheyareSDHConlyTomakethingsworsetheyarecurrentlyonl
yavailableinGBGBcapacitiesSinceoneofmyrequirementswastohavealotofstorageIr
uledtheseoutTheremainingdeviceslistedonAmazonssearchweretheSanDiskUltralin
eButhereconfusionsetsinbecauseSanDiskseparatesthesecardstotwodifferentdevi
cesCamerasmobiledevicesIstherearealdifferencebetweenthetwooristhisjustamar
ketingstuntUnfortunatelyImnotsurebutIdoknowthepricedifferencebetweenthetwo
rangefromacouplecentstoafewdollarsSinceIwasntsureIoptedfortheonespecifical
lytargetedformobiledevicesjustincasethereissomekindofcompatibilityissueTof
indtheexactmodelnumberIwouldgotoSandiskswebpagesandiskcomandcomparetheirex
istingproductlineupFromthereyougetexactmodelnumbersandyoucanthensearchAmaz
onforthesemodelnumbersThatishowIgotmineSDSDQUAGAsforspeedtestsIhaventrunan
yspecifictestingbutcopyingGBworthofdatafrommyPCtothecardliterallytookjusta
fewminutesOnelastnoteisthatAmazonattachesadditionalcharacterstotheendforex
ampleSDSDQUAGAFFPAvsSDSDQUAGUAThedifferencebetweenthetwoisthattheAFFPAmean
sAmazonFrustrationFreePackagingOtherthanthattheseareexactlythesameIfyourew
```

In [34]:

```
review_example = review_example.lower().split()
```

onderingwhatIgotandwanttouseitinyourGalaxySIgottheSDSDQUAGuAanditworkslike
charm

In [55]:

```
review_example
```

Out[55]:

```
['updatesomylovelywifeboughtmeasamsunggalaxytabforfathersdayandivebeenlovi
ngiteversincejustasotherwithsamsungproductsthegalaxytabhastheabilitytoadda
microsdcardtoexpandthememoryonthedevicesinceitsbeenoverayearidecidedtodoso
memoreresearchtoseeifsandiskofferedanythingnewasoftheirproductlineupformic
rosdcardsfromworsttobestperformancewisearetheasfollowssandisksandiskultras
andiskultraplussandiskextremesandiskextremeplussandiskextremepronowthediff
erencebetweenallofthesecardsaresimplythespeedinwhichyoucanreadwritedatatot
hecardyesthepublishedratingsmostallthesecardsexceptthesandiskregulararecl
assunsibutthatsjustratingactualrealworldperformancedoesgetbetterwitheachm
odelbutwithfastercardscomemoreexpensivepricessinceamazondoesntcarrytheultr
aplusmodelofmicrosdcardihadtododirectcomparisonsbetweenthesandiskultraextr
emeandextremeplusasmentionedinmyearlierreviewipurchasedthesandiskultraform
ygalaxysmyquestionwasdidiwanttopayovermoreforacardthatisfasterthantheoneia
lreadyownedorwouldiwantalmostdoubletogetsandisksndmostfastestmicrosdcardtheda
ultraworksperfectlyfineformystyleofusagestoringcapturingpictureshdvideoand
movieplaybackonmyphonesointheendiendedupbuyinganothersandiskultraforar
diusemycellphonemorethanidomytabletandifthecardisgoodenoughformyphoneitsgo
odenoughformytabletidontownahdcameraoranythinglikethatsoihonestlydidntsee
aneedtogetoneofthefastercardsatthistimeiamnowaproudownerofsandiskultracard
sandhaveabsolutelyissueswithitinmysamsungdevicesoriginalreviewihaventhadto
buyamicrosdcardinalongtimethelasttimeiboughtonewasformycellphoneoveryearsa
gobutsincemycellularcontractwasupiknewiwouldhavetogetanewercardinadditiont
omynewphonethesamsunggalaxysreasonforthisisbecauseiknewmysmallgbmicrosdcar
dwasntgoingtocutitdoingresearchonthegalaxysiwantedtogetthebestcardpossible
thathaddecentcapacitygborgreaterthisledmetofindthatthegalaxyssupportsthemi
crosdxcclassuhsicardwhichisthefastestpossiblegiventhatclasssearchingfortha
tspecificallyonamazongavemeresultsofonlyvendorsasofaprilthatmakesthesemicr
osdxcclassuhscardstheyaresandiskthemajoritysamsungandlexarnobodyelsemakest
hesethataresoldonamazonseeinghowsandiskisaprettygoodnameoutoftheiveusedthe
mthemostidecideduponthesandiskbecauselexarwasoverpricedandthesamsungonewas
overpricedaswellasnoteligibleforamazonprimebutthescarythingisthatwhenyoufi
lterbythesandiskyouliterallygetdozensofoptionsallofthemhavedifferentmodeln
umbersdifferentsizesetcthentheresthatconfusionofwhatsthedifferencebetweens
dhcsdxcsdhcvssdxcsdhcstandforsecuredigitalhighcapacityandsdxcstandsforsecu
redigitalextendedcapacityessentiallythesetwocardsarethesamewiththeexceptio
nthatsdhconlysupportscapcitiesuptogbandisformatedwiththefatfilesystemthesd
xccardsareformattedwiththeexfatfilesystemifyouuseansdxccardinadeviceitmust
supportthatfilesystemotherwiseitmaynotberecognizableandoryouhavetoreformat
thecardtofatfatvsexfatthedifferencesbetweenthetwofilesystemsmeansthatfatha
samaximumfilesizeofgblimitedbythatfilesystemexfatontheotherhandsupportsfil
esizesuptotbterabytestheonlythingyouneedtoknowherereallyisthatitspossibley
ourdevicedoesntsupportexfatifthatsthecasejustreformatittofatrememberformat
tingerasesalldatatoclarifythemodelnumbersiihoppedovertothesandiskofficialw
ebpagewhatifoundthereisthattheyoffertwohighspeedoptionsforsandiskcardsthes
earesandiskextremeproandsandiskultrasandiskextremeproisalinethatsupportsre
adspeedsuptombsechowevertheyaresdhconlytomakethingsworsetheyarecurrentlyon
lyavailableingbgbcapacitiessinceoneofmyrequirementswastohavealotofstoragei
ruledtheseouttheremainingdeviceslistedonamazonssearchwerethesandiskultrali
nebuthereconfusionsetsinbecausesandiskseparatesthesecardstotwodifferentdev
icescamerasmobiledevicesistherearealdifferencebetweenthetwooristhisjustama
rketingstuntunfortunatelyimnotsurebutidoknowthepricedifferencebetweenthetw
orangefromacouplecentstoafewdollarssinceiwasntsureioptedfortheonespecifica
llytargetedformobiledevicesjustincasethereissomekindofcompatibilityissueto
findtheexactmodelnumberiwouldgotosandiskswebpagesandiskcomandcomparetheire
xistingproductlineupfromthereyougetexactmodelnumbersandyoucanthensearchama
zonforthesemodelnumbersthatishowigotminesdsdquagasforspeedtestsihaventruna
nyspecifictestingbutcopyinggbworthofdatafrommypctothecardliterallytookjust
afewminutesonelastnoteisthatamazonattachesadditionalcharacterstotheendfore
xamplesdsdquagaffpavssdsdquaguathedifferencebetweenthetwoisthattheaffpamea
nsamazonfrustrationfreepackagingotherthanthatthesereexactlythesameifyoure
```

In [81]:

```python
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[81]:

	reviewerName	overall	reviewText	reviewTime	da
2031	Hyoun Kim "Faluzure"	5	updatesomylovelywifeboughtmeasamsunggalaxytab...	05-01-2013	
3449	NLee the Engineer	5	ihavetesteddozensofsdhcandmicrosdhccardsonedis...	26-09-2012	
4212	SkincareCEO	1	notepleasereadthelastupdatescrolltothebottomim...	08-05-2013	
317	Amazon Customer "Kelly"	1	ifyourcardgetshotenoughtobepainfulitisdefectiv...	09-02-2012	
4672	Twister	5	sandiskannouncementoffirst[illegible]...	03-07-2014	

In [82]:

```python
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
df[['polarity', 'subjectivity']] =df["reviewText"].apply(lambda Text:pd.Series(TextBlob(Te

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'] = "Netural"
```

wonderingwhatigotandwanttouseitinyourgalaxysigotthesdsdquaguaanditworkslik
echarm']

In [83]:

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

Out[83]:

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

In [84]:

```
df[df['sentiment'] == 'Negative'].sort_values("wilson_lower_bound", ascending= True).head
```

Out[84]:

	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_vote

In [80]:

```
catagorical_variable_summary(df, 'sentiment')
```

In [ ]:

```

```