

```
[In [1]]: import import_ipynb
import numpy as np
import pandas as pd
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import re
import textblob
from textblob import TextBlob
import wordcloud
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 [2]]: import os

[In [3]]: os.getcwd()

[Out[3]]: 'C:\Users\BEHARA AMULYA'

[In [4]]: os.chdir("C:\Users\BEHARA AMULYA")

[In [5]]: os.getcwd()

[Out[5]]: 'C:\Users\BEHARA AMULYA'

[In [6]]: df = pd.read_csv('amazon.csv')

[In [7]]: df.head()
```

Out[7]:

Unnamed: 0	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_vote	score_pos_neg_diff	score_average_rating	wilson_lower_bound
0	0	NaN	4	No issues.	23-07-2014	138	0	0	0	0.0	0.0
1	1	Omie	5	Purchased this for my device, it worked as adv...	25-10-2013	409	0	0	0	0.0	0.0
2	2	1K3	4	it works as expected. I should have sprung for...	23-12-2012	715	0	0	0	0.0	0.0
3	3	1m2	5	This think has worked out great.Had a diff. br...	21-11-2013	382	0	0	0	0.0	0.0
4	4	2&amp;np;12Men	5	Bought it with Retail Packaging, arrived legiti...	13-07-2013	513	0	0	0	0.0	0.0

```
[In [8]]: df

[Out[8]]: Unnamed: 0 reviewerName overall reviewText reviewTime day_diff helpful_yes helpful_no total_vote score_pos_neg_diff score_average_rating wilson_lower_bound
```

0	0	NaN	4	No issues.	23-07-2014	138	0	0	0	0.0	0.0
1	1	Omie	5	Purchased this for my device, it worked as adv...	25-10-2013	409	0	0	0	0.0	0.0
2	2	1K3	4	it works as expected. I should have sprung for...	23-12-2012	715	0	0	0	0.0	0.0
3	3	1m2	5	This think has worked out great.Had a diff. br...	21-11-2013	382	0	0	0	0.0	0.0
4	4	2&amp;np;12Men	5	Bought it with Retail Packaging, arrived legiti...	13-07-2013	513	0	0	0	0.0	0.0

...	...	...	...	...	...	...	...	...	...	...	...
4910	4910	ZM "J"	1	I bought this Sandisk 16GB Class 10 io wear wr...	23-07-2013	503	0	0	0	0.0	0.0
4911	4911	Zo	5	Used this for extending the capabilities of my...	02-02-2014	473	0	0	0	0.0	0.0
4912	4912	Z S Liske	5	Great card that is very fast and reliable. It ...	31-03-2014	252	0	0	0	0.0	0.0
4913	4913	Z Taylor	5	Good amount of space for the stuff I want to d...	16-09-2013	448	0	0	0	0.0	0.0
4914	4914	Zzaa	5	I've heard bad things about this 64gb Micro SD...	01-02-2014	310	0	0	0	0.0	0.0

4915 rows x 12 columns

```
[In [9]]: df = df.sort_values("wilson_lower_bound",ascending = False)
df.dropna(inplace = True, axis=1)
df.head()
```

reviewerName	1	0.02																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
--------------	---	------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```
[In [10]]: 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()
    ratio = (df[na_columns_].isnull().sum())/df.shape[0]
    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('QUANTILES' center(82, '-'))
    print(df.quantile([0,0.5,0.50,0.95,0.95,0.99,1],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: 0.02

reviewerName: 1 0.02

reviewText: 1 0.02

-----DUPLICATED VALUES-----

0

-----QUANTILES-----

	0.00	0.00	0.50	0.50	0.95	0.95
overall	1.0	1.0	5.0	5.0	5.0000000	5.0000000
day_diff	1.0	1.0	431.0	431.0	748.0000000	748.0000000
helpful_yes	0.0	0.0	0.0	0.0	1.0000000	1.0000000
helpful_no	0.0	0.0	0.0	0.0	0.0000000	0.0000000
total_vote	0.0	0.0	0.0	0.0	1.0000000	1.0000000
score_pos_neg_diff	-130.0	-130.0	0.0	0.0	1.0000000	1.0000000
score_average_rating	0.0	0.0	0.0	0.0	1.0000000	1.0000000
wilson_lower_bound	0.0	0.0	0.0	0.0	0.206549	0.206549

0.99 1.00

overall: 5.00000 5.000000

day\_diff: 943.00000 1064.000000

helpful\_yes: 3.00000 1952.000000

helpful\_no: 2.00000 183.000000

total\_vote: 4.00000 2020.000000

score\_pos\_neg\_diff: 2.00000 1884.000000

score\_average\_rating: 1.00000 1.000000

wilson\_lower\_bound: 0.34258 0.957544

```
[In [11]]: def check_class(dataFrame):
    nunique_df = pd.DataFrame({'Variable': dataFrame.columns,
                              'Classes': [df[df[col] == v].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

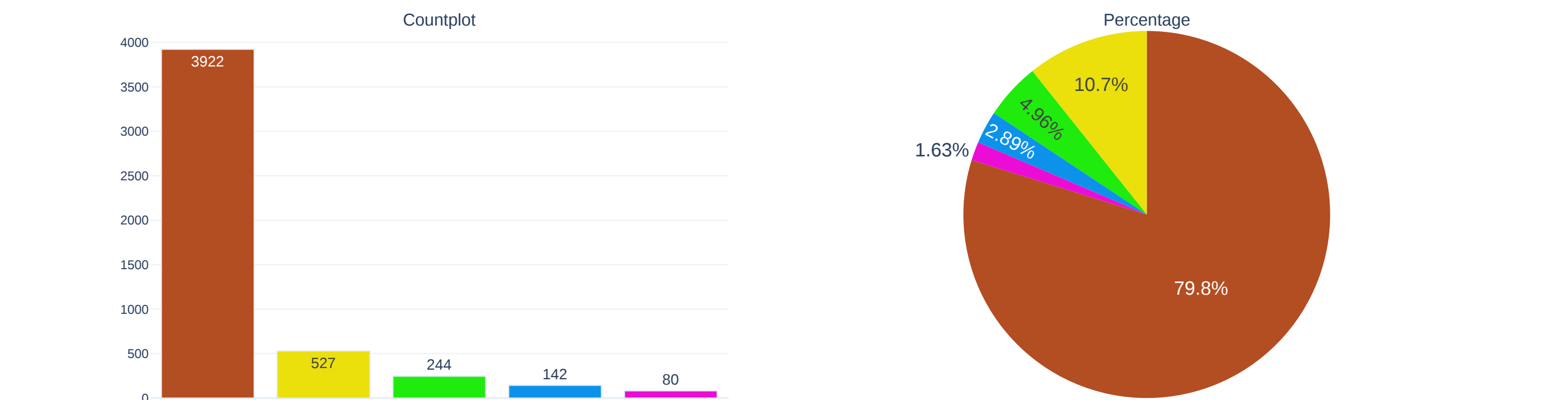
check_class(df)
```

Out[11]:

Variable	Classes
reviewText	4912
reviewerName	4594
reviewTime	690
day_diff	690
wilson_lower_bound	40
score_average_rating	28
score_pos_neg_diff	27
total_vote	26
helpful_yes	23
helpful_no	17
overall	5

```
[In [12]]: import plotly.offline as py
import plotly.graph_objs as go
constraints = [{"#B3D22", "#E8E0C", "#FEB0C", "#BC92EB", "#EBC0C5"}]
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(x = df[column_name].value_counts().tolist(),
                        x = [str(i) for i in df[column_name].value_counts().index],
                        text = df[column_name].value_counts().tolist(),
                        textfont = dict(size=14),
                        name= column_name,
                        textposition = 'auto',
                        showlegend = False,
                        marker=dict(color=constraints,
                                lines=dict(color="#DBE6C",
                                            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,
                        markers=dict(colors=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 [13]]: categorical_variable_summary(df, 'overall')
```



```
[In [14]]: df.reviewText.head()

[Out[14]]: 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 [15]]: review_example = df.reviewText[2031]
review_example

[Out[15]]: '[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 much with Samsung products the Galaxy Tab has the ability to add a microSD card to expand the memory on the device. As of 6/19/2014, their product lineup for microSD cards from worst to best (performance-wise) are the as follows: Sandisk Sandisk Ultra Sandisk Ultra PLUS Sandisk Extreme Sandisk Extreme PLUS Sandisk Extreme PRO Now, 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 fast er 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 movies and moving photos into the end of the day just by using 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-I 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 of the Samsung one 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, y ou 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 difference 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 is 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 hoped 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 90MB/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 n ot 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 de vices (just in case there is some kind of compatibility issue). To find the exact model number, I would go to Sandisk's website (sandisk.com) and compare their existing product li ne. From there, you get exact model numbers and you can then search Amazon for these model numbers. That is how I got mine (SDSQUA-0646). As for speed tests, I haven't run any s peed testing, but you can get 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 SDSQUA-0646-AFPF-A vs SDSQUA-0646-U46A). The difference between the two is that the "AFPFA" 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 SDSQUA-0646-U46A and it works like charm.'
```

```
[In [16]]: review_example = re.sub("([a-zA-Z])+", ' ', review_example)
review_example

[Out[16]]: 'UPDATE So my lovely wife bought the Samsung Galaxy Tab for Father's day and I've been loving it ever since. Just as thoer with the Samsung products the Galaxy Tab has the ability to add a microSD card to expand the memory on the device. Since I've been loving it ever since I decided to do some more research so I found that the new way of the ir product line up for microSD cards from worst to best performance wise are the follows: Sandisk Sandisk Ultra Sandisk Ultra PLUS Sandisk Extreme Sandisk Extreme PLUS Sandisk Extreme PRO Now 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 the 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 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 the Sandisk's 2nd most fastest microSD card. The Ultra works perfectly fine for my style of usage (storing/capturing pictures & HD movies and moving photos into the end of the day just by using 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-I 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 of the Samsung one 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 difference 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 is 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 hoped 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 90MB/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 website (sandisk.com) and compare their existing product line. From there, you get exact model numbers and you can then search Amazon for these model numbers. That is how I got mine (SDSQUA-0646). As for speed tests, I haven't run any speed testing, but you can get 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 SDSQUA-0646-AFPF-A vs SDSQUA-0646-U46A). The difference between the two is that the "AFPFA" 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 SDSQUA-0646-U46A and it works like charm.'
```

```
[In [17]]: review_example = re.sub("([a-zA-Z])+", ' ', review_example)
review_example

[Out[17]]: 'UPDATE So my lovely wife bought the Samsung Galaxy Tab for Father's day and I've been loving it ever since. Just as thoer with the Samsung products the Galaxy Tab has the ability to add a microSD card to expand the memory on the device. Since I've been loving it ever since I decided to do some more research so I found that the new way of the ir product line up for microSD cards from worst to best performance wise are the follows: Sandisk Sandisk Ultra Sandisk Ultra PLUS Sandisk Extreme Sandisk Extreme PLUS Sandisk Extreme PRO Now 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 the 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 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 the Sandisk's 2nd most fastest microSD card. The Ultra works perfectly fine for my style of usage (storing/capturing pictures & HD movies and moving photos into the end of the day just by using 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-I 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 of the Samsung one 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 difference 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 is 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 hoped 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 90MB/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 website (sandisk.com) and compare their existing product line. From there, you get exact model numbers and you can then search Amazon for these model numbers. That is how I got mine (SDSQUA-0646). As for speed tests, I haven't run any speed testing, but you can get 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 SDSQUA-0646-AFPF-A vs SDSQUA-0646-U46A). The difference between the two is that the "AFPFA" 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 SDSQUA-0646-U46A and it works like charm.'
```

```
[In [18]]: review_example = re.sub("([a-zA-Z])+", ' ', review_example)
review_example

[Out[18]]: '[UPDATE So my lovely wife bought the Samsung Galaxy Tab for Father's day and I've been loving it ever since. Just as thoer with the Samsung products the Galaxy Tab has the ability to add a microSD card to expand the memory on the device. Since I've been loving it ever since I decided to do some more research so I found that the new way of the ir product line up for microSD cards from worst to best performance wise are the follows: Sandisk Sandisk Ultra Sandisk Ultra PLUS Sandisk Extreme Sandisk Extreme PLUS Sandisk Extreme PRO Now 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 the 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 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 the Sandisk's 2nd most fastest microSD card. The Ultra works perfectly fine for my style of usage (storing/capturing pictures & HD movies and moving photos into the end of the day just by using 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-I 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 of the Samsung one 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 difference 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 is 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 hoped 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 90MB/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 website (sandisk.com) and compare their existing product line. From there, you get exact model numbers and you can then search Amazon for these model numbers. That is how I got mine (SDSQUA-0646). As for speed tests, I haven't run any speed testing, but you can get 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 SDSQUA-0646-AFPF-A vs SDSQUA-0646-U46A). The difference between the two is that the "AFPFA" 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 SDSQUA-0646-U46A and it works like charm.'
```

```
[In [19]]: rt = lambda x: re.sub("([a-zA-Z])+", ' ', str(x))
df['reviewText'] = df['reviewText'].map(rt)
df['reviewText'] = df['reviewText'].str.lower()
df.head()
```

support capacities up to 4GB, and that the latter is not recognizable and/or you have to reformat the card to FAT32. FAT32 vs exFAT. The difference between the two file systems seems to be that FAT32 has a maximum file size of 4GB, limited by that file system. exFAT on the otherhand, supports file sizes up to 2^48 (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 hopped over to the SanDisk official website. 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 of cards up to 64GB, however they are SDHC only. To make things worse, they are currently only available in 16GB & 32GB 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, but in case there is some kind of compatibility issue, to find the exact model number, I would go 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 (SDSQDQA-64G46). 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 SDDQQA-64G46-AFPF-A) which I believe stands for "Amazon Frustration Free Packaging". Other than that, these are exactly the same. If you're wondering what I got (and what use it in your Galaxy S4), I got the SDDQQA-64G46-u46A and it works like charm.

review\_example = re.sub('([A-Z]{2}-[A-Z]{2})', '\nreview\_example\n', review\_example)

UPDATE: So my lovely wife boughe me a Samsung Galaxy Tab for Father's Day and I've been loving it ever since. Just as a throw in I'll Samsung products the Galaxy Tab has the ability to add micro SD cards to expand the memory on the