

```
import pandas as pd
from textblob import TextBlob
import matplotlib.pyplot as plt

from google.colab import files
import pandas as pd

uploaded = files.upload()
df = pd.read_csv("7817_1.csv")
print(df.head())

Choose Files 7817_1.csv
• 7817_1.csv(text/csv) - 18386219 bytes, last modified: 1/23/2022 - 100% done
Saving 7817_1.csv to 7817_1.csv
   id    asins   brand      categories
0 AVPe7AsMilAPnD_xQ78G B00QJDU3KY Amazon  Amazon Devices,mazon.co.uk
1 AVPe7AsMilAPnD_xQ78G B00QJDU3KY Amazon  Amazon Devices,mazon.co.uk
2 AVPe7AsMilAPnD_xQ78G B00QJDU3KY Amazon  Amazon Devices,mazon.co.uk
3 AVPe7AsMilAPnD_xQ78G B00QJDU3KY Amazon  Amazon Devices,mazon.co.uk
4 AVPe7AsMilAPnD_xQ78G B00QJDU3KY Amazon  Amazon Devices,mazon.co.uk

  colors      dateAdded      dateUpdated \
0  NaN  2016-03-08T20:21:53Z  2017-07-18T23:52:58Z
1  NaN  2016-03-08T20:21:53Z  2017-07-18T23:52:58Z
2  NaN  2016-03-08T20:21:53Z  2017-07-18T23:52:58Z
3  NaN  2016-03-08T20:21:53Z  2017-07-18T23:52:58Z
4  NaN  2016-03-08T20:21:53Z  2017-07-18T23:52:58Z

  dimension      ean      keys ... \
0 169 mm x 117 mm x 9.1 mm NaN kindlepaperwhite/b00qjdusky ...
1 169 mm x 117 mm x 9.1 mm NaN kindlepaperwhite/b00qjdusky ...
2 169 mm x 117 mm x 9.1 mm NaN kindlepaperwhite/b00qjdusky ...
3 169 mm x 117 mm x 9.1 mm NaN kindlepaperwhite/b00qjdusky ...
4 169 mm x 117 mm x 9.1 mm NaN kindlepaperwhite/b00qjdusky ...

  reviews.rating      reviews.sourceURLS \
0      5.0 https://www.amazon.com/Kindle-Paperwhite-High...
1      5.0 https://www.amazon.com/Kindle-Paperwhite-High...
2      4.0 https://www.amazon.com/Kindle-Paperwhite-High...
3      5.0 https://www.amazon.com/Kindle-Paperwhite-High...
4      5.0 https://www.amazon.com/Kindle-Paperwhite-High...

  reviews.text \
0 I initially had trouble deciding between the p...
1 Allow me to preface this with a little history...
2 I am enjoying it so far. Great for reading. Ha...
3 I bought one of the first Paperwhites and have...
4 I have to say upfront - I don't like corporat...

  reviews.title reviews.userCity \
0 Paperwhite voyage, no regrets!      NaN
1 One Simply Could Not Ask For More      NaN
2 Great for those that just want an e-reader      NaN
3 Love / Hate relationship      NaN
4 I LOVE IT      NaN

  reviews.userProvince      reviews.username      sizes      upc      weight
0      NaN      Cristina M      NaN      NaN  205 grams
1      NaN          Ricky      NaN      NaN  205 grams
2      NaN      Tedd Gardner      NaN      NaN  205 grams
3      NaN        Dougal      NaN      NaN  205 grams
4      NaN  Miljan David Tanic      NaN      NaN  205 grams

[5 rows x 27 columns]
```

```
columns_needed = [
    'id', 'asins', 'brand', 'categories', 'dateAdded',
    'reviews.rating', 'reviews.text', 'reviews.title',
    'reviews.username'
]
df_clean = df[columns_needed]
df_clean
```

	id		asins		brand	categories	dateAdded	reviews.rating	reviews.text	reviews.title	reviews.username
0	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I initially had trouble deciding between the p...	Paperwhite voyage, no regrets!	Cristina M		
1	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	Allow me to preface this with a little history...	One Simply Could Not Ask For More	Ricky		
2	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	4.0	I am enjoying it so far. Great for reading. Ha...	Great for those that just want an e-reader	Tedd Gardiner		
3	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I bought one of the first Paperwhites and have...	Love / Hate relationship	Dougal		
4	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I have to say upfront - I don't like corporat...	I LOVE IT	Miljan David Tanic		
...
1592	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	3.0	This is not the same remote that I got for my ...	I would be disappointed with myself if i produ...	GregAmandawith4		
1593	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	I have had to change the batteries in this rem...	Battery draining remote!!!!	Amazon Customer		
1594	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	Remote did not activate, nor did it connect to...	replacing an even worse remote. Waste of time	Amazon Customer		

1595	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	3.0	It does the job but is super over priced. I fe...	Overpriced	Meg Ashley
1596	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	I ordered this item to replace the one that no...	I am sending all of this crap back to amazon a...	DIANE K

1597 rows × 9 columns

Next steps: [Generate code with df_clean](#) [View recommended plots](#) [New interactive sheet](#)

```
[7] df_clean.columns = [
    'review_id', 'asin', 'brand', 'category', 'date_added',
    'rating', 'review_text', 'review_title', 'username'
]
df_clean
```

	review_id	asin	brand	category	date_added	rating	review_text	review_title	username
0	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I initially had trouble deciding between the p...	Paperwhite voyage, no regrets!	Cristina M
1	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	Allow me to preface this with a little history...	One Simply Could Not Ask For More	Ricky
2	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	4.0	I am enjoying it so far. Great for reading. Ha...	Great for those that just want an e-reader	Tedd Gardiner
3	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I bought one of the first Paperwhites and have...	Love / Hate relationship	Dougal
4	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08T20:21:53Z	5.0	I have to say upfront - I don't like corporat...	I LOVE IT	Miljan David Tanic
...
1592	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	3.0	This is not the same remote that I got for my ...	I would be disappointed with myself if i produ...	GregAmandawith4
1593	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	I have had to change the batteries in this item...	Battery draining remote!!!!	Amazon Customer
1594	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	Remote did not activate, nor did it connect to...	replacing an even worse remote. Waste of time	Amazon Customer
1595	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	3.0	It does the job but is super over priced. I fe...	Overpriced	Meg Ashley
1596	AVpfo9ukilAPnD_xfhuj	B00NO8JJZW	Amazon	Amazon Devices & Accessories,Amazon Device Acc...	2016-04-02T14:40:43Z	1.0	I ordered this item to replace the one that no...	I am sending all of this crap back to amazon a...	DIANE K

1597 rows × 9 columns

Next steps: [Generate code with df_clean](#) [View recommended plots](#) [New interactive sheet](#)

```
[8] df_clean['date_added'] = pd.to_datetime(df_clean['date_added'], errors='coerce')
→ <ipython-input-8-03027eaf91f2>:1: SettingWithCopyWarning:
  A value is trying to be set on a copy of a slice from a DataFrame.
  Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
  df_clean['date_added'] = pd.to_datetime(df_clean['date_added'], errors='coerce')
```

```
[9] df_clean.dropna(subset=['review_text'], inplace=True)
→ <ipython-input-9-b05af3e678d6>:1: SettingWithCopyWarning:
  A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
  df_clean.dropna(subset=['review_text'], inplace=True)
```

```
[10] def get_sentiment(text):
    return TextBlob(str(text)).sentiment.polarity
[11] df_clean['sentiment'] = df_clean['review_text'].apply(get_sentiment)
→ <ipython-input-11-58454115324c>:1: SettingWithCopyWarning:
  A value is trying to be set on a copy of a slice from a DataFrame.
  Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
  df_clean['sentiment'] = df_clean['review_text'].apply(get_sentiment)
```

```
[12] df_clean.to_csv("amazon_reviews_cleaned.csv", index=False)
print(df_clean.head())
```

	review_id	asin	brand	category	date_added	rating	review_text
0	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08 20:21:53+00:00	5.0	I initially had trouble deciding between the p...
1	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08 20:21:53+00:00	5.0	Allow me to preface this with a little history...
2	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08 20:21:53+00:00	4.0	I am enjoying it so far. Great for reading. Ha...
3	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08 20:21:53+00:00	5.0	I bought one of the first Paperwhites and have...
4	AVpe7AsMilAPnD_xQ78G	B00QJDU3KY	Amazon	Amazon Devices,mazon.co.uk	2016-03-08 20:21:53+00:00	5.0	I have to say upfront - I don't like corporat...

```

review_title      username  sentiment
0      Paperwhite voyage, no regrets!    Cristina M  0.200259
1 One Simply Could Not Ask For More        Ricky  0.159591
2 Great for those that just want an e-reader   Tedd Gardiner  0.420833
3 Love / Hate relationship       Dougal  0.148821
4           I LOVE IT    Miljan David Tanic  0.267086

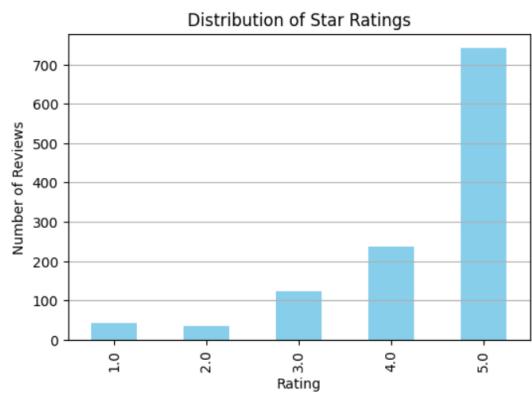
```

```

13] import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
df_clean['rating'].value_counts().sort_index().plot(kind='bar', color='skyblue')
plt.title('Distribution of Star Ratings')
plt.xlabel('Rating')
plt.ylabel('Number of Reviews')
plt.grid(True, axis='y')
plt.show()

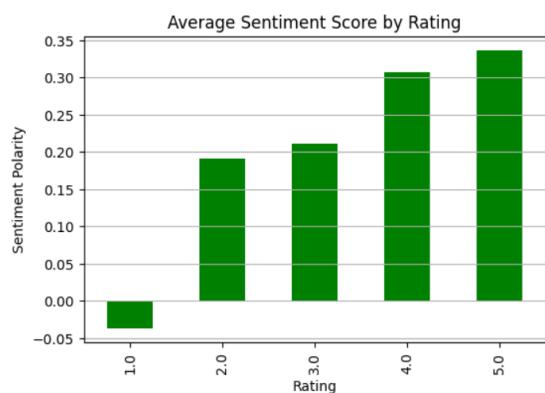
```



```

14] plt.figure(figsize=(6,4))
df_clean.groupby('rating')['sentiment'].mean().plot(kind='bar', color='green')
plt.title('Average Sentiment Score by Rating')
plt.xlabel('Rating')
plt.ylabel('Sentiment Polarity')
plt.grid(True, axis='y')
plt.show()

```

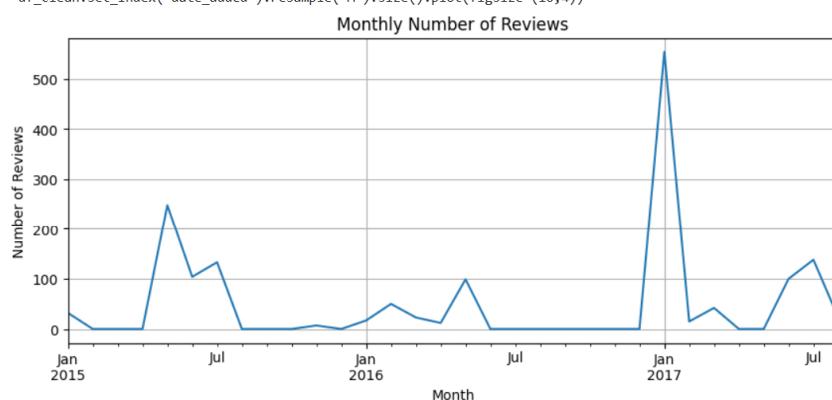


```

15] df_clean.set_index('date_added').resample('M').size().plot(figsize=(10,4))
plt.title('Monthly Number of Reviews')
plt.ylabel('Number of Reviews')
plt.xlabel('Month')
plt.grid(True)
plt.show()

```

<ipython-input-15-44ade211db62>:1: FutureWarning: 'M' is deprecated and will be removed in a future version, please use 'ME' instead.
df_clean.set_index('date_added').resample('M').size().plot(figsize=(10,4))



[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 12:44 PM

