



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('Reviews.csv', nrows=500)
df.head(3)
```



	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I hav bought several c th Vitalit canne d.
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Produc arrive labeled a Jumb Salte Peanut.
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia"	1	1	4	1219017600	"Delight" says it all	This is confectio that ha bee




Next steps:

[Generate code with df](#)

[View recommended plots](#)

[New interactive sheet](#)


```
from nltk.corpus import stopwords
from textblob import TextBlob
from textblob import Word
df['Text'] = df['Text'].apply(lambda x: " ".join(x.lower() for x in x.split()))
df['Text'] = df['Text'].str.replace('[^\w\s]', '')
stop = stopwords.words('english')
df['Text'] = df['Text'].apply(lambda x: " ".join(x for x in x.split() if x not in stop))
df['Text'] = df['Text'].apply(lambda x: str(TextBlob(x).correct()))
df['Text'] = df['Text'].apply(lambda x: " ".join([Word(word).lemmatize() for word in x.split()]))
df.Text.head()
```



	Text
0	bought several vitality canned dog food produc...
1	product arrived labelled lumbo halted peanuts....
2	connection around centuries. light, pillow cit...
3	looking secret ingredient robittussin believe f...
4	great staff great price. wide assortment mummy...

dtype: object

```
import nltk
nltk.download('stopwords')
```




[nltk_data] Downloading package stopwords to /root/nltk_data...

[nltk_data] Unzipping corpora/stopwords.zip.

True

```
import nltk
nltk.download('wordnet')
```



[nltk_data] Downloading package wordnet to /root/nltk_data...

True

```
import pandas as pd
import matplotlib.pyplot as plt
reviews = df
reviews.dropna(inplace=True)
reviews.Score.hist(bins=5, grid=False)
plt.show()
print(reviews.groupby('Score').count().Id)
```



```
Score
1    18
2    18
3    18
4    18
5    18
Name: Id, dtype: int64
```

```
wordcloud_negative = WordCloud(background_color='white') \
    .generate(negative_reviews_str)
wordcloud_positive = WordCloud(background_color='white') \
    .generate(positive_reviews_str)

fig = plt.figure(figsize=(10,10))
ax1 = fig.add_subplot(211)
ax1.imshow(wordcloud_negative, interpolation='bilinear')
ax1.axis("off")
ax1.set_title('Reviews with Negative Scores', fontsize=20)
ax2 = fig.add_subplot(212)
ax2.imshow(wordcloud_positive, interpolation='bilinear')
ax2.axis("off")
ax2.set_title('Reviews with Positive Scores', fontsize=20)
plt.show()
```



```
Requirement already satisfied: vaderSentiment in /usr/local/lib/python3.11/dist-packages (3.3.2)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from vaderSentiment) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.4.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2.3.0)
```

```
import seaborn as sns
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
plt.style.use('fivethirtyeight')
cp = sns.color_palette()
analyzer = SentimentIntensityAnalyzer()
emptyline=[]
for row in df['Text']:
    vs=analyzer.polarity_scores(row)
    emptyline.append(vs)
```

```
df_sentiment=pd.DataFrame(emptyline)
df_sentiment.head()
```

1 to 2 of 2 entries (filtered from 5 total entries)

Filter

index	neg	neu	pos	compound
1	0.258	0.644	0.099	-0.5719
2	0.126	0.623	0.25	0.788

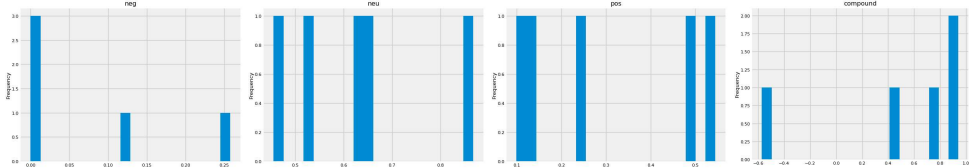
Show

100

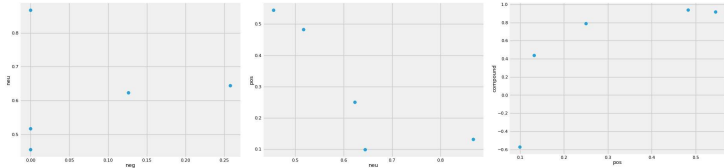
 per page

Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

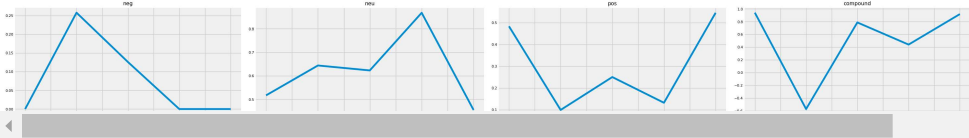
Distributions



2-d distributions




Values




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

```
df_combined=pd.concat([df.reset_index(drop=True),df_sentiment],axis=1)
df_combined.head(3)
```



	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Te
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	boug sever vitali canne dog for produc
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	produ arrive labell lunt halte peanuts.
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	connect arour centurie ligh pillow cit



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

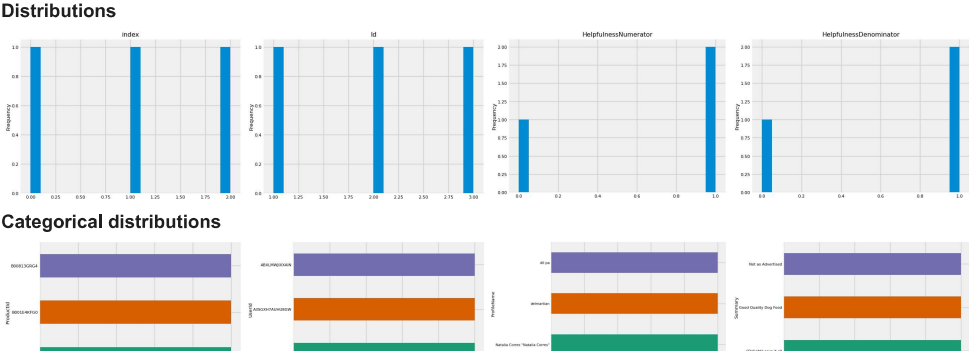
```
df_combined['Sentiment']=np.where(df_combined['compound']>=0,'Positive','Negative')
df_combined.head(3)
```

1 to 3 of 3 entries Filter ?

index	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text	neg
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	bought several vitality canned dog food product found good quality. product look like stew process meat smell better. labrador finicky appreciated product better most.	0.0
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	product arrived labelled lumbo halted peanuts...the peanut actually small sized insulted. sure error vendor intended represent product "lumbo".	0.258
2	3	B000LQOCH0	ABXLMWJIXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	connection around centuries. light, pillow city gelatin nut - case filbert. cut tiny square liberally coated powdered sugar. tiny youthful heaven. chew, flavorful. highly recommend mummy treat. familiar story c.s. lewis' "the lion, witch, wardrobe" - treat reduces edmund selling brother sister witch.	0.126

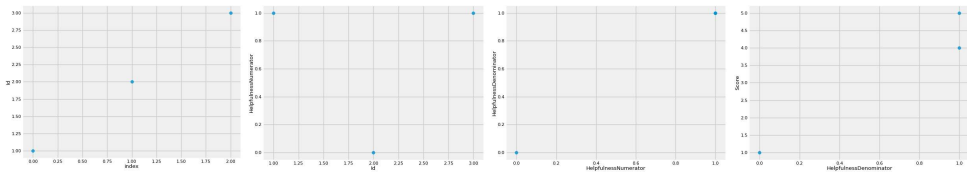
Show 25 per page

Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

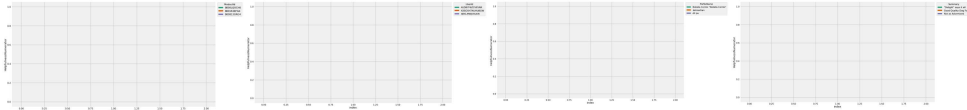




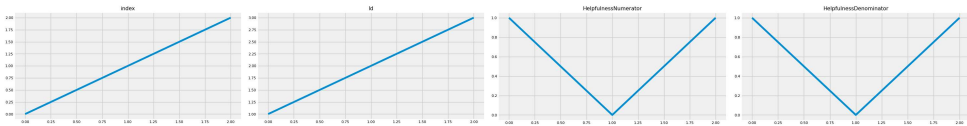
2-d distributions



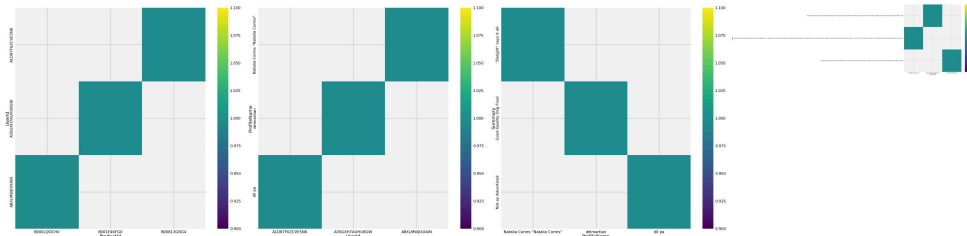
Time series



Values



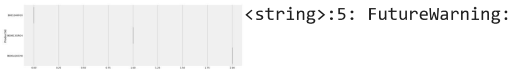
2-d categorical distributions



Faceted distributions

<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legenc



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legenc

