```
!pip install vaderSentiment
!pip install scipy pandas pingouin
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-w
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
     |████████████████████████████████| 125 kB 5.6 MB/s
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/d
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-w
Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-package
Collecting pingouin
  Downloading pingouin-0.5.2.tar.gz (185 kB)
     |████████████████████████████████| 185 kB 3.2 MB/s
Requirement already satisfied: numpy<1.23.0,>=1.16.5 in /usr/local/lib/python3
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: matplotlib>=3.0.2 in /usr/local/lib/python3.7/d
Requirement already satisfied: seaborn>=0.11 in /usr/local/lib/python3.7/dist-
Collecting statsmodels>=0.13
  Downloading statsmodels-0.13.2-cp37-cp37m-manylinux_2_17_x86_64.manylinux201
     |████████████████████████████████| 9.8 MB 7.7 MB/s
Requirement already satisfied: scikit-learn<1.1.0 in /usr/local/lib/python3.7/
Collecting pandas_flavor>=0.2.0
  Downloading pandas_flavor-0.3.0-py3-none-any.whl (6.3 kB)
Collecting outdated
  Downloading outdated-0.2.1-py3-none-any.whl (7.5 kB)
Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /us
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/d
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/d
Requirement already satisfied: xarray in /usr/local/lib/python3.7/dist-package
Collecting pandas_flavor>=0.2.0
  Downloading pandas_flavor-0.2.0-py2.py3-none-any.whl (6.6 kB)
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.7/dis
Collecting littleutils
  Downloading littleutils-0.2.2.tar.gz (6.6 kB)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/d
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7
```

```
  Building wheels for collected packages: pingouin, littleutils
    Building wheel for pingouin (setup.py) ... done
    Created wheel for pingouin: filename=pingouin-0.5.2-py3-none-any.whl size=1
    Stored in directory: /root/.cache/pip/wheels/11/5a/63/a6d32fc26fa462c731f65
    Building wheel for littleutils (setup.py) ... done
    Created wheel for littleutils: filename=littleutils-0.2.2-py3-none-any.whl
```

```
import numpy as np
import pandas as pd
import re
import matplotlib.pyplot as plt
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

import math
from scipy.stats import mannwhitneyu



filepath = '/content/dataset.csv'


dataframe = pd.read_csv(filepath,on_bad_lines='skip')



dataframe.head()
```

	Text	Author_id	Id	Username	Location	Account
0	@boredowntw @snapchatsupport yea i just contac...	1.340000e+18	1.550000e+18	PENIELMARAJ	🇬🇭 ǁ **blk** ǁ **he** ǁ **bi**	snapchat
1	@alexolix05 @lostsoulnw @snapchatsupport T'as ...	1.540000e+18	1.550000e+18	NaN	NaN	snapchat
	@snapchatsupport I					

```
analyser = SentimentIntensityAnalyzer()



#cleaning the tweets
def remove_pattern(input_txt, pattern):
    r = re.findall(pattern, input_txt)
    for i in r:
        input_txt = re.sub(i, '', input_txt)
    return input_txt
def clean_tweets(tweets):
    #remove twitter Return handles (RT @xxx:)
```

```
    tweets = np.vectorize(remove_pattern)(tweets, "RT @[\w]*:")

    #remove twitter handles (@xxx)
    tweets = np.vectorize(remove_pattern)(tweets, "@[\w]*")

    #remove URL links (httpxxx)
    tweets = np.vectorize(remove_pattern)(tweets, "https?://[A-Za-z0-9./]*")

    #remove special characters, numbers, punctuations (except for #)
    tweets = np.core.defchararray.replace(tweets, "[^a-zA-Z]", " ")

    return tweets


dataframe['Text'] = clean_tweets(dataframe['Text'])
dataframe['Text'].head()
```

```
0      yea i just contacted customer support on the...
1       T’as de la chance, moi impossible de ne ser...
2     I stil can’t log in to my account it says bc ...
3        I have not been received an email or anyth...
4     Help i have been having this issue since yest...
Name: Text, dtype: object
```

```
scores = []
# Declare variables for scores
compound_list = []
positive_list = []
negative_list = []
neutral_list = []
for i in range(dataframe['Text'].shape[0]):
#print(analyser.polarity_scores(sentiments_pd['text'][i]))
    compound = analyser.polarity_scores(dataframe['Text'][i])["compound"]
    pos = analyser.polarity_scores(dataframe['Text'][i])["pos"]
    neu = analyser.polarity_scores(dataframe['Text'][i])["neu"]
    neg = analyser.polarity_scores(dataframe['Text'][i])["neg"]

    scores.append({"Compound": compound,
                       "Positive": pos,
                       "Negative": neg,
                       "Neutral": neu
                   })


sentiments_score = pd.DataFrame.from_dict(scores)
dataframe = dataframe.join(sentiments_score)
dataframe.head()
```

Text	Author_id	Id	Username	Location	Account	Compound
yea i just						

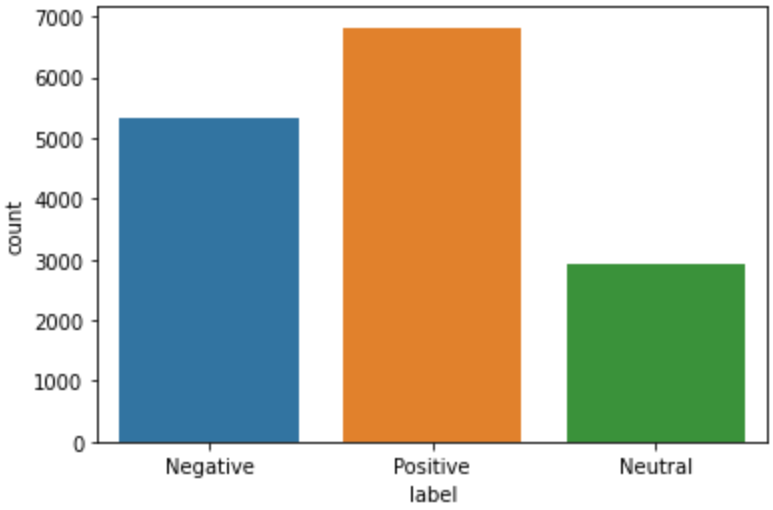
0	contacted customer support on the...	1.340000e+18	1.550000e+18	PENIELMARAJ	🇬🇭 ll **blk** ll **he** ll **bi**	snapchat	-0.6597
1	T’as de la chance, moi impossible	1.540000e+18	1.550000e+18	NaN	NaN	snapchat	0.2500

```
def sentimentPredict(sentiment):
  if sentiment >= 0.05:
    return "Positive"
  elif sentiment <= -0.05:
    return "Negative"
  else:
    return "Neutral"
dataframe['label'] =dataframe['Compound'].apply(lambda x: sentimentPredict(x))
dataframe.head(10)
```

	Text	Author_id	Id	Username	Location	Account	Compound
0	yea i just contacted customer support on the...	1.340000e+18	1.550000e+18	PENIELMARAJ	🇬🇭 ll **blk** ll **he** ll **bi**	snapchat	-0.6597
1	T’as de la chance, moi impossible de ne ser...	1.540000e+18	1.550000e+18	NaN	NaN	snapchat	0.2500
2	I stil can’t log in to my account it says bc ...	1.400000e+18	1.550000e+18	gamermika19	Roermond, Nederland	snapchat	0.1855

```
import seaborn as sns
sns.countplot(dataframe['label'], label='count')
```

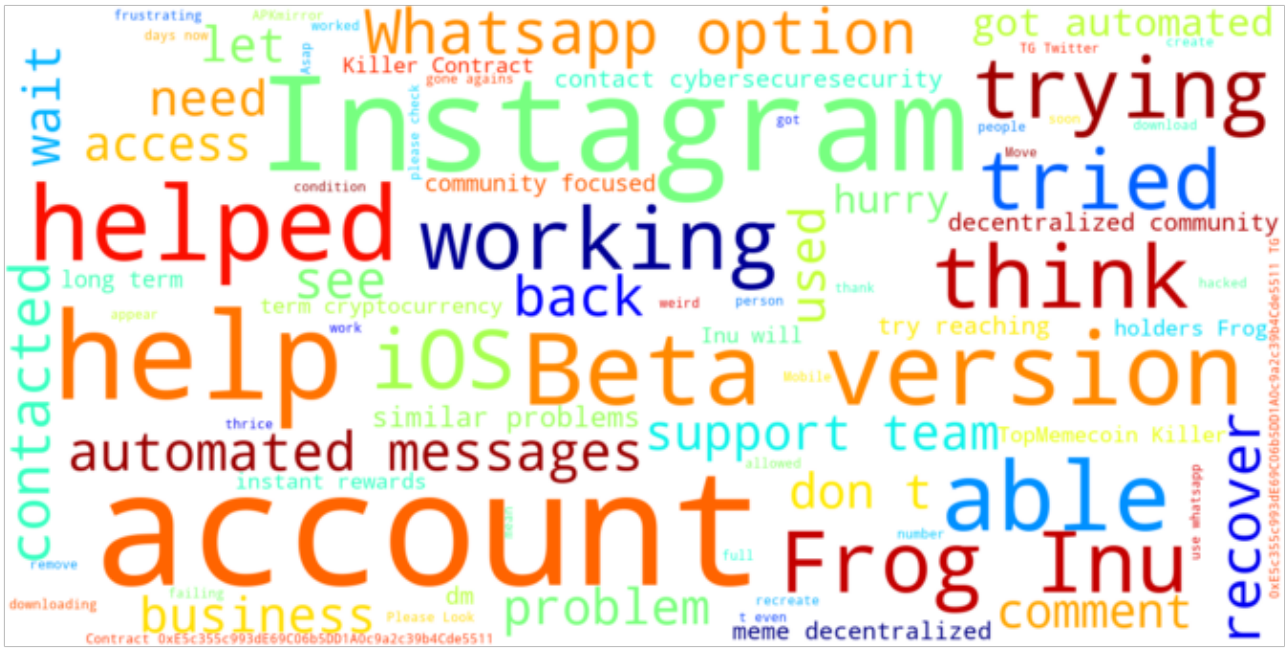
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarni
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b1af70810>
```




```
positive = dataframe[dataframe['label']== 'Positive']
negative = dataframe[dataframe['label']== 'Negative']
neutral  = dataframe[dataframe['label']=='Neutral']
```

```
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
def word_cloud(wd_list):
    stopwords = set(STOPWORDS)
    all_words = ' '.join([text for text in wd_list])
    wordcloud = WordCloud(
```

```
    wordcloud = WordCloud(
        background_color='white',
        stopwords=stopwords,
        width=1600,
        height=800,
        random_state=1,
        colormap='jet',
        max_words=80,
        max_font_size=200).generate(all_words)
    plt.figure(figsize=(12, 10))
    plt.axis('off')
    plt.imshow(wordcloud, interpolation="bilinear");
word_cloud(dataframe['Text'])
```



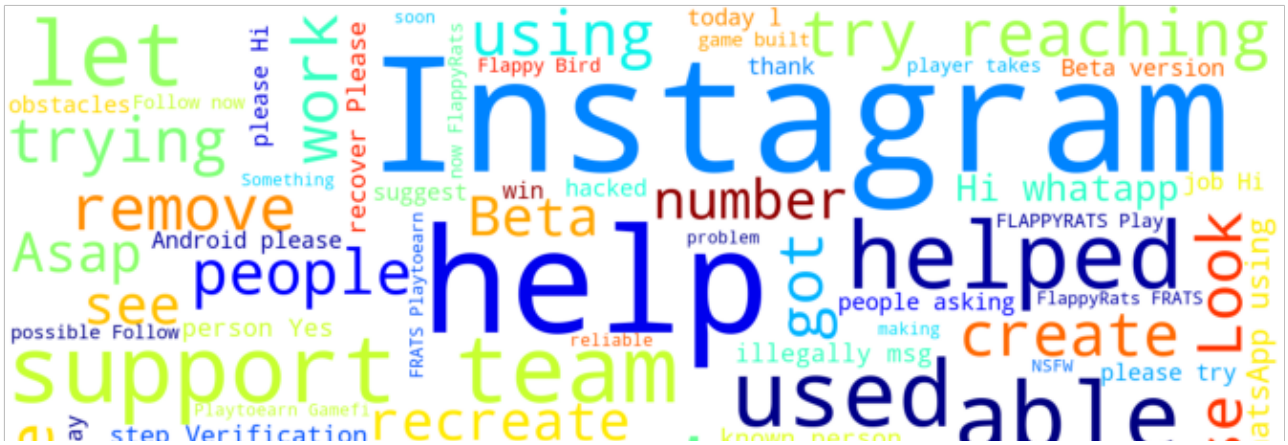

```
neg_sentences = negative['Text'].tolist()
neg_sentences_as_one_string = " ".join(neg_sentences)
plt.figure(figsize = (15,15))
wordcloud = WordCloud(
        background_color='white',
        width=1600,
        height=800,
        random_state=1,
        colormap='jet',
        max_words=80,
        max_font_size=200).generate(neg_sentences_as_one_string)
plt.figure(figsize=(12, 10))
plt.axis('off')
plt.imshow(wordcloud, interpolation="bilinear");
plt.imshow(wordcloud)
```

```
<matplotlib.image.AxesImage at 0x7f6b17686b90>
```

```
<Figure size 1080x1080 with 0 Axes>
```

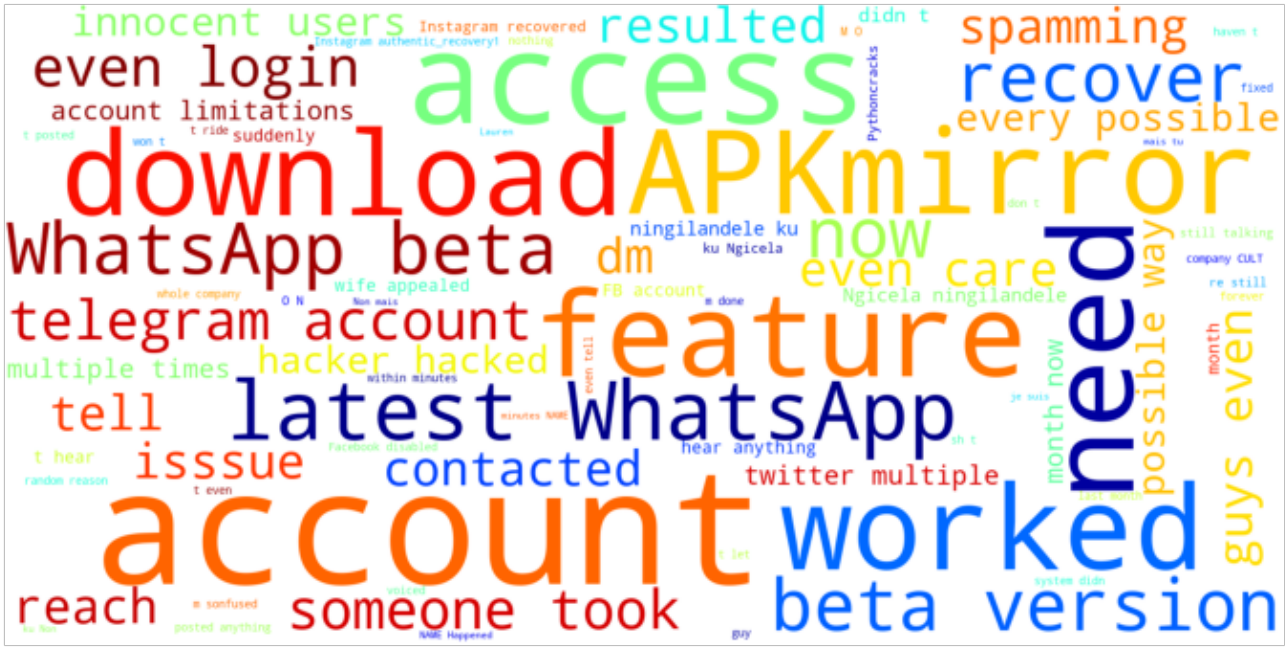

```
pos_sentences = positive['Text'].tolist()
pos_sentences_as_one_string = " ".join(pos_sentences)
plt.figure(figsize = (15,15))
wordcloud = WordCloud(
        background_color='white',
        width=1600,
        height=800,
        random_state=1,
        colormap='jet',
        max_words=80,
        max_font_size=200).generate(pos_sentences_as_one_string)
plt.figure(figsize=(12, 10))
plt.axis('off')
plt.imshow(wordcloud, interpolation="bilinear");
plt.imshow(wordcloud)
```

```
<matplotlib.image.AxesImage at 0x7f6b17b7ef50>
<Figure size 1080x1080 with 0 Axes>
```




```
neu_sentences = neutral['Text'].tolist()
neu_sentences_as_one_string = " ".join(neu_sentences)
plt.figure(figsize = (15,15))
wordcloud = WordCloud(
        background_color='white',
        width=1600,
        height=800,
        random_state=1,
        colormap='jet',
        max_words=80,
        max_font_size=200).generate(neu_sentences_as_one_string)
plt.figure(figsize=(12, 10))
plt.axis('off')
plt.imshow(wordcloud, interpolation="bilinear");
plt.imshow(wordcloud)
```

```
<matplotlib.image.AxesImage at 0x7f6b19910110>
<Figure size 1080x1080 with 0 Axes>
```




```
HT_positive = []
def hashtag_extract(x):
    hashtags = []
    # Loop over the words in the tweet
    for i in x:
```

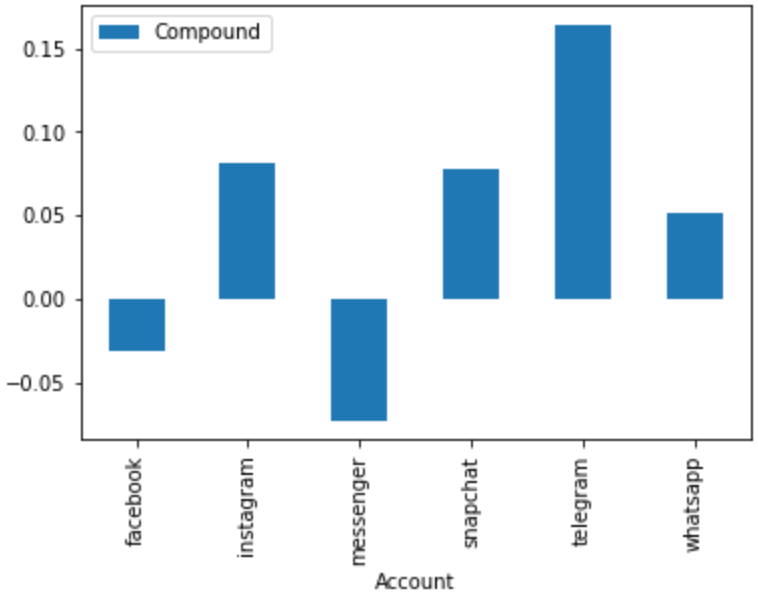
```
    for i in x:
        ht = re.findall(r"#(\w+)", i)
        hashtags.append(ht)
    return hashtags
# extracting hashtags from positive tweetsHT_positive = hashtag_extract(df_tws['te›
# extracting hashtags from  tweets
HT_positive = hashtag_extract(dataframe['Text'][dataframe['Compound'] > 0.5])
# unnesting list
HT_positive = sum(HT_positive,[])
#HT_positive


#Collect the compound values for each news source
score_table = dataframe.pivot_table(index='Account',  values="Compound", aggfunc =
score_table
```

Account	Compound
facebook	-0.031306
instagram	0.081103
messenger	-0.072470
snapchat	0.077882
telegram	0.164028
whatsapp	0.051440

```
score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b18a002d0>
```



```
#Collect the negative values for each news source
neg_score_table = dataframe.pivot_table(index='Account',  values="Negative", aggfur
neg_score_table
```

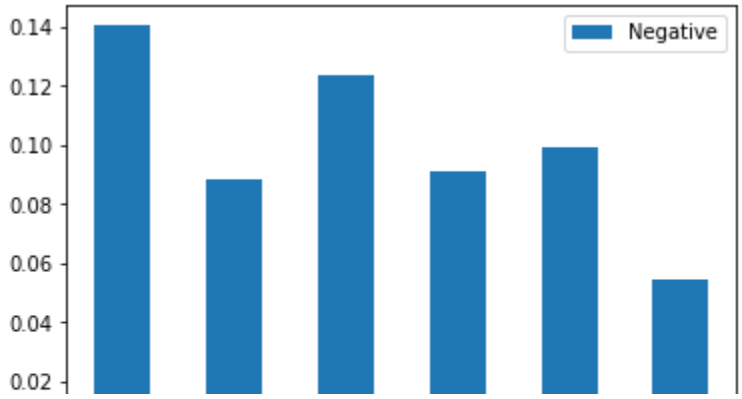
Account	Negative
facebook	0.140315
instagram	0.088443
messenger	0.123800
snapchat	0.091146
telegram	0.099500
whatsapp	0.054700

```
# pos_sentences = dataframe['text'].tolist()
# pos_sentences_as_one_string = " ".join(pos_sentences)

#  wordcloud = WordCloud(
#         background_color='white',
#         stopwords=stopwords,
#         width=1600,
#         height=800,
#         random_state=1,
#         colormap='jet',
#         max_words=80,
#         max_font_size=200).generate(all_words)
#     plt.figure(figsize=(12, 10))
#     plt.axis('off')
#     plt.imshow(wordcloud, interpolation="bilinear");
```

```
neg_score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b175ff890>
```

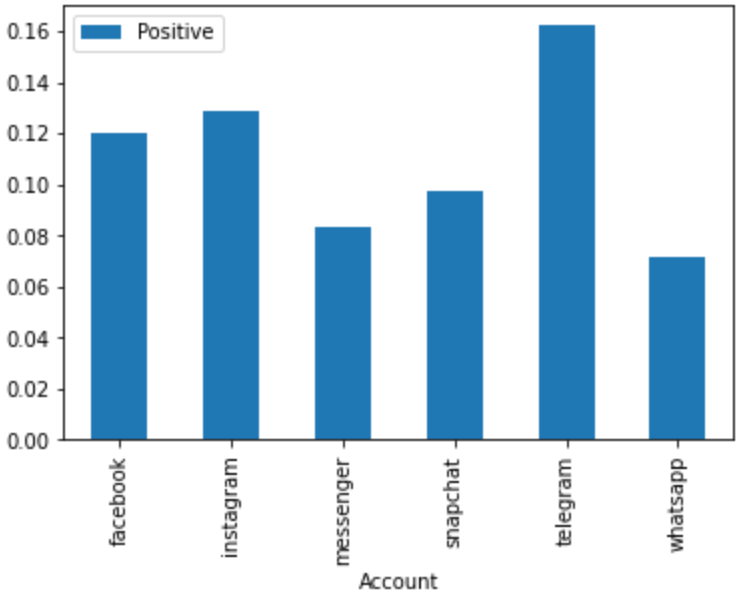



```
 #Collect the negative values for each news source
pos_score_table = dataframe.pivot_table(index='Account',  values="Positive", aggfur
pos_score_table
```

Account	Positive
facebook	0.119925
instagram	0.128725
messenger	0.083400
snapchat	0.097439
telegram	0.162247
whatsapp	0.071500

```
pos_score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b1763c150>
```




```
loca_df = dataframe[dataframe['Location'].notnull()]
len(loca_df)
```

```
9149
```

```
loca_df
```

	Text	Author_id	Id	Username	Location	Accou
0	yea i just contacted customer support on the...	1.340000e+18	1.550000e+18	PENIELMARAJ	🇬🇭 ll **blk** ll **he** ll **bi**	snapch
2	I stil can't log in to my account it says bc ...	1.400000e+18	1.550000e+18	gamermika19	Roermond, Nederland	snapch
9	when switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapch
10	Mine still ain't working and y'all couldn't	1.120000e+18	1.550000e+18	lexiLeighaHart	Florida, USA	snapch

```
temp_df = loca_df['Location'].str.split(',', expand=True)
temp_df.columns= ['city','country']
new_loc_df = pd.concat([loca_df,temp_df], axis=1,ignore_index=False)
new_loc_df
```

	Text	Author_id	Id	Username	Location	Accou
0	yea i just contacted customer support on the...	1.340000e+18	1.550000e+18	PENIELMARAJ	🇬🇭 ll **blk** ll **he** ll **bi**	snapch
2	I stil can't log in to my account it says bc ...	1.400000e+18	1.550000e+18	gamermika19	Roermond, Nederland	snapch
9	when switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapch
10	Mine still ain't working and y'all couldn't h...	1.120000e+18	1.550000e+18	lexiLeighaHart	Florida, USA	snapch
13	fix it omg it's been a fuckimg day	1.290000e+18	1.550000e+18	sxdek	United Kingdom	snapch
...	...	...	...	...	...	
15065	So Facebook disabled my account	1.197863e+08	1.560000e+18	TCapitalG	CincinnatiIL.A.ICorpus Christi	facebo

	for some rando...					
15067	And you will pay the price one day for lyi...	7.732281e+08	1.560000e+18	HAdamsen	Hillerød, Danmark	facebo
15068	is there someway to contact you guys directly... \n\n\I'll	1.023073e+08	1.560000e+18	OdrarEth	Florida, USA	facebo

```
len(new_loc_df)
```

```
9149
```

```
new_loc_df = new_loc_df[new_loc_df['country'].notnull()]
```

```
new_loc_df
```

	Text	Author_id	Id	Username	Location	Account	Compou
2	I stil can’t log in to my account it says bc ...	1.400000e+18	1.550000e+18	gamermika19	Roermond, Nederland	snapchat	0.18
9	when switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapchat	0.54
10	Mine still ain’t working and y’all couldn’t h...	1.120000e+18	1.550000e+18	lexiLeighaHart	Florida, USA	snapchat	0.40
19		1.499320e+09	1.550000e+18	JacobBlackey	Belmont, NH	snapchat	0.00
	when switching						

20	switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapchat	0.54
...	...	...	...	...	...	...	

```
len(new_loc_df)
```

```
4617
```

```
loc_based_neg_score_table = new_loc_df.pivot_table(index=['country','Account'],  va
```

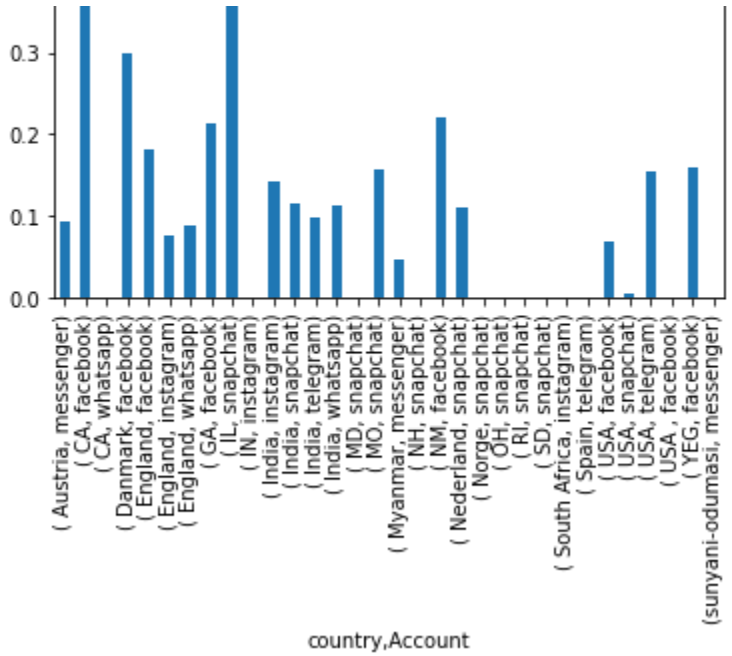
```
loc_based_neg_score_table
```

		Negative
country	**Account**	
Austria	**messenger**	0.093000
CA	**facebook**	0.447000
	whatsapp	0.000000
Danmark	**facebook**	0.298000

England	**facebook**	0.182000
	instagram	0.075000
	whatsapp	0.088000
GA	**facebook**	0.213178
IL	**snapchat**	0.508000
IN	**instagram**	0.000000
India	**instagram**	0.141196
	snapchat	0.116000
	telegram	0.097000
	whatsapp	0.113333
MD	**snapchat**	0.000000
MO	**snapchat**	0.156000
Myanmar	**messenger**	0.045000
NH	**snapchat**	0.000000
NM	**facebook**	0.221000
Nederland	**snapchat**	0.109000
Norge	**snapchat**	0.000000
OH	**snapchat**	0.000000
RI	**snapchat**	0.000000
SD	**snapchat**	0.000000
South Africa	**instagram**	0.000000
Spain	**telegram**	0.000000
USA	**facebook**	0.067355
	snapchat	0.003967
	telegram	0.153000
USA	**facebook**	0.000000

```
loc_based_neg_score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b17edf8d0>
```




```
loc_based_pos_score_table = new_loc_df.pivot_table(index=['country','Account'],  va
```

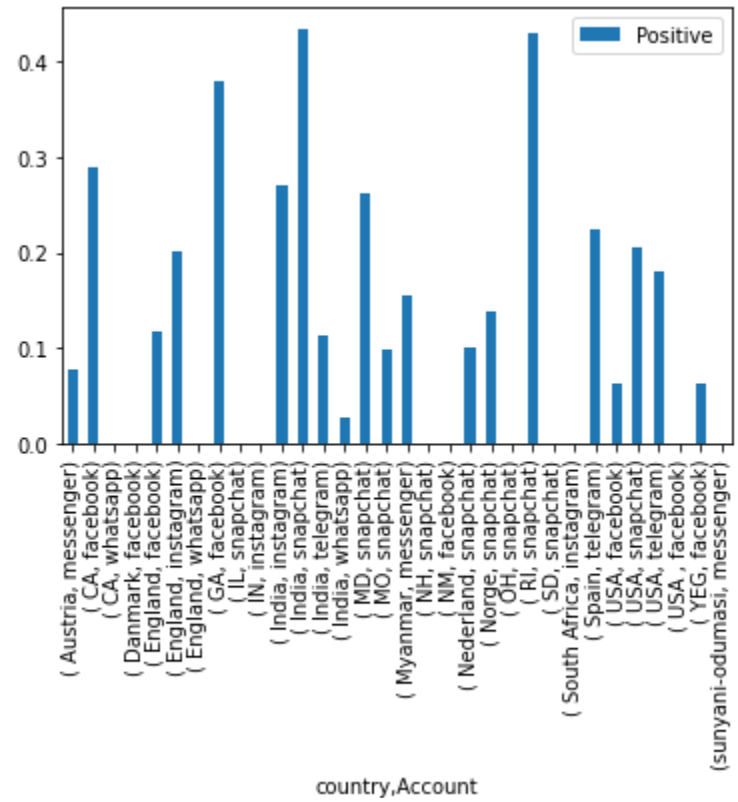
```
loc_based_pos_score_table
```

		Positive
country	**Account**	
Austria	**messenger**	0.078000
CA	**facebook**	0.289000
	whatsapp	0.000000
Danmark	**facebook**	0.000000
England	**facebook**	0.117000
	instagram	0.201000
	whatsapp	0.000000
GA	**facebook**	0.380493
IL	**snapchat**	0.000000
IN	**instagram**	0.000000
India	**instagram**	0.270298
	snapchat	0.435000
	telegram	0.113000
	whatsapp	0.027333

MD	**snapchat**	0.261500
MO	**snapchat**	0.098000
Myanmar	**messenger**	0.154000
NH	**snapchat**	0.000000
NM	**facebook**	0.000000
Nederland	**snapchat**	0.101000
Norge	**snapchat**	0.139000
OH	**snapchat**	0.000000
RI	**snapchat**	0.431000
SD	**snapchat**	0.000000
South Africa	**instagram**	0.000000
Spain	**telegram**	0.224000
USA	**facebook**	0.063424
	snapchat	0.205000
	telegram	0.181000
USA	**facebook**	0.000000

```
loc_based_pos_score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b18859850>
```



```
loc_based_neu_score_table = new_loc_df.pivot_table(index=['country','Account'],  va
```

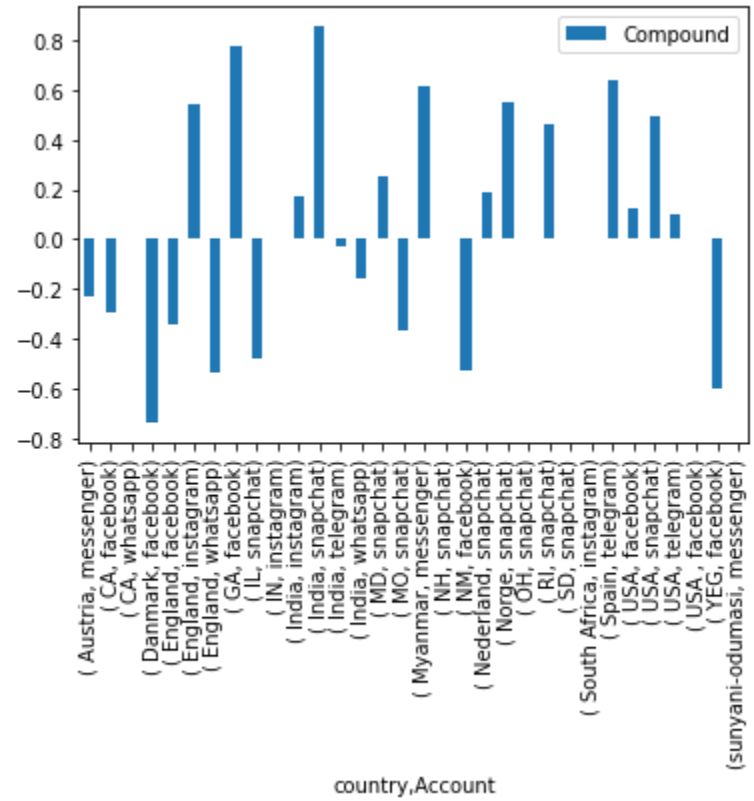
```
loc_based_neu_score_table
```

		Compound
country	**Account**	
Austria	**messenger**	-0.226300
CA	**facebook**	-0.296000
	whatsapp	0.000000
Danmark	**facebook**	-0.735100
England	**facebook**	-0.338200
	instagram	0.542300
	whatsapp	-0.537900
GA	**facebook**	0.773538
IL	**snapchat**	-0.476700
IN	**instagram**	0.000000
India	**instagram**	0.168860
	snapchat	0.853700
	telegram	-0.025800
	whatsapp	-0.157967
MD	**snapchat**	0.254000
MO	**snapchat**	-0.368500
Myanmar	**messenger**	0.612400
NH	**snapchat**	0.000000
NM	**facebook**	-0.526700
Nederland	**snapchat**	0.185500
Norge	**snapchat**	0.549900
OH	**snapchat**	0.000000
RI	**snapchat**	0.464800

SD	**snapchat**	0.000000
South Africa	**instagram**	0.000000
Spain	**telegram**	0.636900
USA	**facebook**	0.122091
	snapchat	0.491700
	telegram	0.102700
USA	**facebook**	0.000000

```
loc_based_neu_score_table.plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6b17706210>
```




```
type(loc_based_neu_score_table)
```

```
pandas.core.frame.DataFrame
```

```
loc_based_neu_score_table.index
```

```
MultiIndex([(       ' Austria', 'messenger'),
            (            ' CA',  'facebook'),
            (            ' CA',  'whatsapp'),
            (       ' Danmark',  'facebook'),
            (       ' England',  'facebook'),
            (       ' England', 'instagram'),
            (       ' England',  'whatsapp'),
```

```
(          ' England',  'whatsapp'),
(             ' GA',  'facebook'),
(             ' IL',  'snapchat'),
(             ' IN', 'instagram'),
(          ' India', 'instagram'),
(          ' India',  'snapchat'),
(          ' India',  'telegram'),
(          ' India',  'whatsapp'),
(             ' MD',  'snapchat'),
(             ' MO',  'snapchat'),
(        ' Myanmar', 'messenger'),
(             ' NH',  'snapchat'),
(             ' NM',  'facebook'),
(      ' Nederland',  'snapchat'),
(          ' Norge',  'snapchat'),
(             ' OH',  'snapchat'),
(             ' RI',  'snapchat'),
(             ' SD',  'snapchat'),
(   ' South Africa', 'instagram'),
(          ' Spain',  'telegram'),
(            ' USA',  'facebook'),
(            ' USA',  'snapchat'),
(            ' USA',  'telegram'),
(           ' USA ',  'facebook'),
(            ' YEG',  'facebook'),
 ('sunyani-odumasi', 'messenger')],
names=['country', 'Account'])
```

```
new_loc_df
```

	Text	Author_id	Id	Username	Location	Account	Compou
2	I stil can’t log in to my account it says bc ...	1.400000e+18	1.550000e+18	gamermika19	Roermond, Nederland	snapchat	0.18
9	when switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapchat	0.54
10	Mine still ain’t working and y’all couldn’t h...	1.120000e+18	1.550000e+18	lexiLeighaHart	Florida, USA	snapchat	0.40
19		1.499320e+09	1.550000e+18	JacobBlackey	Belmont, NH	snapchat	0.00

20	when switching between cameras on my s22, i c...	1.090000e+18	1.550000e+18	DrMagnusW1	Steinkjer, Norge	snapchat	0.54
...	...	...	...	...	...	...	

```
#l1 = new_loc_df[[' England',  'instagram']]

df_tp1 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
final_list1 = l1+l2

df_tp3 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp4 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
l3 = df_tp3['Compound'].tolist()
l4 = df_tp4['Compound'].tolist()
final_list2 = l3 + l4


len(final_list1)
```

```
450
```

```
len(final_list2)
```

```
len(final_list2)
```

```
1576
```

```
data_dict = {'Eng':final_list1,"Ind":final_list2}

df = pd.DataFrame({'Eng': pd.Series(final_list1), 'Ind': pd.Series(final_list2)})
df = df.replace(np.nan, 0)
```

```
results = mannwhitneyu(df['Eng'], df['Ind'])
results
```

```
MannwhitneyuResult(statistic=1429139.0, pvalue=3.1077084834410555e-15)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp2 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
final_list1 = l1+l2

df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
df_tp4 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l3 = df_tp3['Compound'].tolist()
l4 = df_tp4['Compound'].tolist()
final_list2 = l3 + l4
```

```
data_dict = {'Ind':final_list1,"USA":final_list2}

df = pd.DataFrame({'Ind': pd.Series(final_list1), 'USA': pd.Series(final_list2)})
df = df.replace(np.nan, 0)
```

```
results = mannwhitneyu(df['Ind'], df['USA'])
results
```

```
MannwhitneyuResult(statistic=445.0, pvalue=1.0416942005059925e-172)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' Danmark') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
l3 = df_tp3['Compound'].tolist()
```

```
final_list1 = l1+l2+l3

df_tp4 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp5 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp6 = new_loc_df[(new_loc_df["country"] == ' South Africa') & (new_loc_df["Accou
l4 = df_tp4['Compound'].tolist()
l5 = df_tp5['Compound'].tolist()
l6 = df_tp6['Compound'].tolist()
final_list2 = l4 + l5 +l6



data_dict = {'Fb':final_list1,"Inst":final_list2}

df_fb_inst = pd.DataFrame({'Fb': pd.Series(final_list1), 'Inst': pd.Series(final_li
df_fb_inst = df_fb_inst.replace(np.nan, 0)




results = mannwhitneyu(df_fb_inst['Fb'], df_fb_inst['Inst'])
results

    MannwhitneyuResult(statistic=143144.5, pvalue=0.036003201427796325)


df_tp1 = new_loc_df[(new_loc_df["country"] == ' Nederland') & (new_loc_df["Account'
df_tp2 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
l3 = df_tp3['Compound'].tolist()
final_list1 = l1+l2+l3

df_tp4 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp5 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp6 = new_loc_df[(new_loc_df["country"] == ' South Africa') & (new_loc_df["Accou
l4 = df_tp4['Compound'].tolist()
l5 = df_tp5['Compound'].tolist()
l6 = df_tp6['Compound'].tolist()
final_list2 =  l4 +l5+l6



data_dict = {'Sn':final_list1,"Ins":final_list2}

df_fb_inst = pd.DataFrame({'Sn': pd.Series(final_list1), 'Ins': pd.Series(final_lis
df_fb_inst = df_fb_inst.replace(np.nan, 0)


results = mannwhitneyu(df_fb_inst['Sn'], df_fb_inst['Ins'])
```

```
results = mannwhitneyu(df_fb_inst['Sn'], df_fb_inst['Wh'])
results
```

```
MannwhitneyuResult(statistic=88320.0, pvalue=3.933134314972951e-06)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()

final_list1 = l1+l2
df_tp2 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l3 = df_tp3['Compound'].tolist()
l4 = df_tp4['Compound'].tolist()
final_list2 = l3 + l4
df_fb_inst = pd.DataFrame({'Wh': pd.Series(final_list1), 'Sn': pd.Series(final_list
df_fb_inst = df_fb_inst.replace(np.nan, 0)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:12: DeprecationWa
  if sys.path[0] == '':
```

```
results = mannwhitneyu(df_fb_inst['Wh'], df_fb_inst['Sn'])
results
```

```
MannwhitneyuResult(statistic=0.0, pvalue=1.3311138118554098e-14)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' Danmark') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
l3 = df_tp3['Compound'].tolist()
final_list1 = l1+l2+l3

df_tp4 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp5 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
l4 = df_tp4['Compound'].tolist()
l5 = df_tp5['Compound'].tolist()
final_list2 = l4 + l5

df_fb_inst = pd.DataFrame({'Fb': pd.Series(final_list1), 'What': pd.Series(final_li
df_fb_inst = df_fb_inst.replace(np.nan, 0)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:15: DeprecationWa
  from ipykernel import kernelapp as app
```

```
results = mannwhitneyu(df_fb_inst['Fb'], df_fb_inst['What'])
results
```

```
    MannwhitneyuResult(statistic=162892.5, pvalue=0.06173878799528974)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp3 = new_loc_df[(new_loc_df["country"] == ' South Africa') & (new_loc_df["Accou
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
l3 = df_tp3['Compound'].tolist()
final_list1 =  l1 +l2+l3

df_tp4 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp5 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
l4 = df_tp4['Compound'].tolist()
l5 = df_tp5['Compound'].tolist()
final_list2 = l4 + l5


df_fb_inst = pd.DataFrame({'Ins': pd.Series(final_list1), 'What': pd.Series(final_l
df_fb_inst = df_fb_inst.replace(np.nan, 0)
```

```
    /usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:16: DeprecationWa
      app.launch_new_instance()
```

```
results = mannwhitneyu(df_fb_inst['Ins'], df_fb_inst['What'])
results
```

```
    MannwhitneyuResult(statistic=66690.0, pvalue=0.00024544970366060256)
```

```
df_tp1 = new_loc_df[(new_loc_df["country"] == ' Danmark') & (new_loc_df["Account"]
df_tp2 = new_loc_df[(new_loc_df["country"] == ' England') & (new_loc_df["Account"]
df_tp3 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l1 = df_tp1['Compound'].tolist()
l2 = df_tp2['Compound'].tolist()
l3 = df_tp3['Compound'].tolist()
final_list1 = l1+l2+l3

df_tp4 = new_loc_df[(new_loc_df["country"] == ' India') & (new_loc_df["Account"] ==
df_tp5 = new_loc_df[(new_loc_df["country"] == ' USA') & (new_loc_df["Account"] == '
l4 = df_tp4['Compound'].tolist()
l5 = df_tp5['Compound'].tolist()
final_list2 = l4 + l5


df_fb_inst = pd.DataFrame({'Fb': pd.Series(final_list1), 'Sn': pd.Series(final_list
df_fb_inst = df_fb_inst.replace(np.nan, 0)


results = mannwhitneyu(df_fb_inst['Fb'], df_fb_inst['Sn'])
results
```

```
results
```

```
MannwhitneyuResult(statistic=159894.0, pvalue=0.22390255208847598)
```