

finz.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share M

RAM Disk Editing

Files

sample_data

finviz.csv

finviz_sentiment.csv

list_of_company.csv

finviz data collection

```
[ ] def random_char(y):
    return ''.join(random.choice(string.ascii_letters) for x in range(y))

[ ] import requests
import string
import random
import pandas as pd

company_name = []

proxies = {
    "http": None,
    "https": None,
}

headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/41.0.2228.0 Safari/537.36'
}

session = requests.Session()
session.trust_env = False

try:
    for i in range(3000):
        randomLetter = random_char(1)
        r = session.get("https://finviz.com/api/suggestions.ashx?input=" + randomLetter, proxies=proxies, headers=headers)
        print(i, randomLetter)

        for data in r.json():
            company_name.append([data['company'], data['ticker']])
except e:
    print('error')

try:
    for i in range(3000):
        randomLetter = random_char(2)
        r = session.get("https://finviz.com/api/suggestions.ashx?input=" + randomLetter, proxies=proxies, headers=headers)
        print(i, randomLetter)

        for data in r.json():
            company_name.append([data['company'], data['ticker']])
except e:
    print('error')

try:
    for i in range(3000):
        randomLetter = random_char(3)
        r = session.get("https://finviz.com/api/suggestions.ashx?input=" + randomLetter, proxies=proxies, headers=headers)
        print(i, randomLetter)

        for data in r.json():
            company_name.append([data['company'], data['ticker']])
except e:
    print('error')

try:
    for i in range(3000):
        randomLetter = random_char(4)
        r = session.get("https://finviz.com/api/suggestions.ashx?input=" + randomLetter, proxies=proxies, headers=headers)
        print(i, randomLetter)

        for data in r.json():
            company_name.append([data['company'], data['ticker']])
except e:
    print('error')

try:
    for i in range(3000):
        randomLetter = random_char(5)
        r = session.get("https://finviz.com/api/suggestions.ashx?input=" + randomLetter, proxies=proxies, headers=headers)
        print(i, randomLetter)

        for data in r.json():
            company_name.append([data['company'], data['ticker']])
except e:
    print('error')

company_name = [list(i) for i in set(map(tuple, company_name))]
print(len(company_name))
print(company_name)

df = pd.DataFrame(company_name, columns=['company_name', 'company_ticker'])
df.to_csv('list_of_company.csv')

2945 qUPUH
2946 qBqOu
2947 qhwm
2948 qTVWS
2949 qJdsh
2950 UWf1D
2951 UJK7R
2952 l1IBP
2953 MBxxQ
2954 FjvvB
2955 BTYtx
2956 ebdOg
2957 nijxtZ
2958 QJiro
2959 WxvYv
2960 qWUQG
2961 RBHaA
2962 N2VhK
2963 nVug
2964 dzGpd
2965 GCoaE
2966 aEBo
2967 ocbAa
2968 CQszV
2969 bdXnd
2970 P2lFO
2971 paipg
2972 DSGaG
2973 SjnB
2974 uCimy
2975 BWjAv
2976 BdJMe
2977 xUHaq
2978 YCstj
2979 MHCQg
2980 ctsba
2981 KoCpd
2982 caWfF
2983 FctTz
2984 UBxtl
2985 Bzca...
```

```

2985 DrEqz
2986 wbCuq
2987 IHBii
2988 ojaM
2989 ASpRd
2990 lCuue
2991 hFydi
2992 HFT2Q
2993 kqkPx
2994 KfWxV
2995 wpNds
2996 xiUhs
2997 ElgGN
2998 Icsop
2999 YIAKq
6229
[ 'SolarEdge Technologies, Inc.', 'SEDG' ], [ 'Westlake Corporation', 'WLK' ], [ 'ZIVO Bioscience, Inc.', 'ZIVO' ], [ 'Qorvo, Inc.', 'QRVO' ], [ 'North American Construction Group Ltd.', 'NACG' ]

```

```

from urllib.request import urlopen, Request
from bs4 import BeautifulSoup
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import pandas as pd
import matplotlib.pyplot as plt
from lxml import etree
import time

df = pd.read_csv('/content/list_of_company.csv')
title_list = []

for index, row in df.iterrows():
    company_name = row['company_name']
    company_ticker = row['company_ticker']

    print(index, row['company_ticker'])

    try:
        url = f'https://finviz.com/quote.ashx?t={company_ticker}&ty=c&ta=1&p=d'
        req = Request(url=url, headers={'user-agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/109.0.0.0 Safari/537.36'})
        response = urlopen(req)

        time.sleep(0.5)
        html = BeautifulSoup(response, features='html.parser')

        dom = etree.HTML(str(html))

        for title in dom.xpath('//*[@*[contains(concat( " ", @class, " " ), concat( " ", "tab-link-news", " " ))]]'):
            title_list.append({
                'title': title.text,
                'company_name': company_name,
                'company_ticker': company_ticker,
                'source': 'finviz'})
    except Exception as e:
        print(e)

df = pd.DataFrame(title_list)
df

```

```

6154 APE
6155 GNT
6156 CACO
6157 CQP
6158 TTCF
6159 NABL
6160 ASLN
6161 MAA
6162 CXE
6163 ATY
6164 IEO
6165 USB
6166 RKL
6167 EA
6168 JOET
6169 JOE
6170 CPSI
6171 EFTV
6172 GLRE
6173 XSOE
6174 GXG
6175 CODA
6176 AKAN
6177 CFFE
6178 AG
6179 FNDA
6180 NUW
6181 XMVM
6182 FCF
6183 XNET
6184 OESX
6185 DIVO
6186 EXG
6187 RNER
6188 RVLP
6189 FFEB
6190 KNOP
6191 PUTW
6192 C
6193 DMXF
6194 KNBE
6195 GPP
6196 QMOM
6197 GP
6198 TKLF
6199 BPAC
6200 SRAX
6201 NM
6202 FRGT
6203 HYGH
6204 EOI
6205 LEGR
6206 KCCA
6207 UBT
6208 UPV
6209 KN
...

```

```

[3] df.to_csv('finviz.csv')

```

finviz sentiment analysis

```

[ ] import nltk
nltk.download('vader_lexicon')

[nltk_data]  Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]  Package vader_lexicon is already up-to-date!
True

[19] df = pd.read_csv('/content/finviz.csv')

[20] vader = SentimentIntensityAnalyzer()
    df['temp'] = df['title'].astype(str)
    f = lambda title: vader.polarity_scores(title)['compound']
    df['compound'] = df['temp'].apply(f)
    df['sentiment'] = df['compound'].apply(lambda x: 'positive' if x >= 0 else 'negative')

```

Unnamed: 0	title	company_name	company_ticker	source	temp	compound	sentiment
0	0 The Zacks Analyst Blog Highlights JinkoSolar H...	SolarEdge Technologies, Inc.	SEDG	finviz	The Zacks Analyst Blog Highlights JinkoSolar H...	0.0000	positive
1	1 SolarEdge (SEDG) Commences Shipping of New Bat...	SolarEdge Technologies, Inc.	SEDG	finviz	SolarEdge (SEDG) Commences Shipping of New Bat...	0.0000	positive
2	2 SolarEdge Begins Shipment of New Battery Cell ...	SolarEdge Technologies, Inc.	SEDG	finviz	SolarEdge Begins Shipment of New Battery Cell ...	0.2732	positive
3	3 JinkoSolar (JKS) Introduces Upgraded Tiger Neo...	SolarEdge Technologies, Inc.	SEDG	finviz	JinkoSolar (JKS) Introduces Upgraded Tiger Neo...	0.0000	positive
4	4 12 Best Up-and-Coming Stocks to Buy Now	SolarEdge Technologies, Inc.	SEDG	finviz	12 Best Up-and-Coming Stocks to Buy Now	0.6369	positive
...
460387	460387 The Real Good Food Company Reports Third Quart...	The Real Good Food Company, Inc.	RGF	finviz	The Real Good Food Company Reports Third Quart...	0.4404	positive
460388	460388 The Real Good Food Company to Report Third Qua...	The Real Good Food Company, Inc.	RGF	finviz	The Real Good Food Company to Report Third Qua...	0.4404	positive
460389	460389 The Real Good Food Company to Ring Nasdaq Clos...	The Real Good Food Company, Inc.	RGF	finviz	The Real Good Food Company to Ring Nasdaq Clos...	0.4404	positive
460390	460390 Real Good Food Prices IPO At \$12/Share, Below ...	The Real Good Food Company, Inc.	RGF	finviz	Real Good Food Prices IPO At \$12/Share, Below ...	0.4404	positive
460391	460391 The Real Good Food Company Announces Pricing o...	The Real Good Food Company, Inc.	RGF	finviz	The Real Good Food Company Announces Pricing o...	0.4404	positive

460392 rows x 8 columns



```
[21] vader.polarity_scores('you are a bad boy')

{'neg': 0.538, 'neu': 0.462, 'pos': 0.0, 'compound': -0.5423}
```

Preprocessing

```
[24] df.drop(['Unnamed: 0', 'temp'], axis = 1, inplace=True)
```

```
[25] df['company_name'].apply(lambda x: x.lower())
df
```

	title	company_name	company_ticker	source	compound	sentiment	edit
0	0 The Zacks Analyst Blog Highlights JinkoSolar H...	SolarEdge Technologies, Inc.	SEDG	finviz	0.0000	positive	
1	1 SolarEdge (SEDG) Commences Shipping of New Bat...	SolarEdge Technologies, Inc.	SEDG	finviz	0.0000	positive	
2	2 SolarEdge Begins Shipment of New Battery Cell ...	SolarEdge Technologies, Inc.	SEDG	finviz	0.2732	positive	
3	3 JinkoSolar (JKS) Introduces Upgraded Tiger Neo...	SolarEdge Technologies, Inc.	SEDG	finviz	0.0000	positive	
4	4 12 Best Up-and-Coming Stocks to Buy Now	SolarEdge Technologies, Inc.	SEDG	finviz	0.6369	positive	
...	
460387	460387 The Real Good Food Company Reports Third Quart...	The Real Good Food Company, Inc.	RGF	finviz	0.4404	positive	
460388	460388 The Real Good Food Company to Report Third Qua...	The Real Good Food Company, Inc.	RGF	finviz	0.4404	positive	
460389	460389 The Real Good Food Company to Ring Nasdaq Clos...	The Real Good Food Company, Inc.	RGF	finviz	0.4404	positive	
460390	460390 Real Good Food Prices IPO At \$12/Share, Below ...	The Real Good Food Company, Inc.	RGF	finviz	0.4404	positive	
460391	460391 The Real Good Food Company Announces Pricing o...	The Real Good Food Company, Inc.	RGF	finviz	0.4404	positive	

460392 rows x 6 columns

```
[26] df.to_csv('finviz_sentiment.csv', index=False)
```



Colab paid products - Cancel contracts here

✓ 2s completed at 04:02

