

## Reading the web page into python

```
In [1]: import requests # To load the webpage

In [2]: r = requests.get('https://www.nytimes.com/interactive/2017/06/23/opinion/trumps-lies.html')

In [3]: r

Out[3]: <Response [200]>

In [4]: r.text[:500]

Out[4]: <!DOCTYPE html><html><!--[if gt IE 9]>[[[!IE]]> <!--><html lang=en" class=no-js page-interactive section-opinion page-theme-standard tone-opinion page-interactive-default limit-small
out-charge app-interactive" itemid="http://www.nytimes.com/interactive/2017/06/23/opinion/trumps-lies.html" itemtype="http://schema.org/NewsArticle" itemscope xmlns:og="http://
opengraphprotocol.org/schema"><!--[[endif]]--><!--[[if IE 9]]> <html lang=en" class=no-js ie9 1-1e10 page-interactive section-opinion page>

parsing the html using BeautifulSoup

In [5]: from bs4 import BeautifulSoup as bs # It will display the content which are inside the tags

In [6]: soup = bs(r.text, 'html.parser') # to Parsing the content using html tags

In [ ]: print(soup.prettify())

Collecting all the records

In [8]: result = soup.find_all('span', attrs={'class':'short-desc'})

In [9]: len(result)

Out[9]: 189

In [ ]: print(result)

In [11]: result[0]

Out[11]: <span class="short-desc"><strong>Jan. 21 </strong>"I wasn't a fan of Iraq, I didn't want to go into Iraq." <span class="short-truth"><a href="https://www.buzzfeed.com/andrewkaczynski/
k1/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a></span></span>

In [12]: result[1]

Out[12]: <span class="short-desc"><strong>Jan. 21 </strong>"A reporter for Time magazine -- and I have been on their cover 14 or 15 times. I think we have the all-time record in the history
of Time magazine." <span class="short-truth"><a href="https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion be
fore he was against it.)</a></span>

In [13]: result[1:]

Out[13]: <span class="short-desc"><strong>Nov. 11 </strong>"I'd rather have him -- you know, work with him on the Ukraine than standing and arguing about whether or not -- because that's whol
e thing was set up by the Democrats." <span class="short-truth"><a href="https://www.nytimes.com/interactive/2017/12/18/us/politics/trump-and-russia.html" target="_blank">(There is
no evidence that Democrats "set up" Russian interference in the election.)</a></span></span>

Extracting the Dates
```

```
In [14]: result0 = result[0]

In [15]: result0

Out[15]: <strong>Jan. 21 </strong>"I wasn't a fan of Iraq, I didn't want to go into Iraq." <span class="short-truth"><a href="https://www.buzzfeed.com/andrewkaczynski/
k1/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a></span></span>

In [16]: result0.find('strong')

Out[16]: <strong>Jan. 21 </strong>

In [17]: result0.find('strong').text # get_text

Out[17]: 'Jan. 21'

In [18]: result0.find('strong').text[0:7]

Out[18]: 'Jan. 21'

In [19]: result0.find('strong').text[0:7] + ', 2017'

Out[19]: 'Jan. 21, 2017'
```

## Extracting the lie

```
In [20]: result0

Out[20]: <span class="short-desc"><strong>Jan. 21 </strong>"I wasn't a fan of Iraq, I didn't want to go into Iraq." <span class="short-truth"><a href="https://www.buzzfeed.com/andrewkaczynski/
k1/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a></span></span>

In [21]: result0.contents

Out[21]: <strong>Jan. 21 </strong>
" I wasn't a fan of Iraq, I didn't want to go into Iraq." "
" https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion be
fore he was against it.)</a></span>

In [22]: result0.contents[1]

Out[22]: "'I wasn't a fan of Iraq, I didn't want to go into Iraq.'"

In [23]: result0.contents[1][1:-2]

Out[23]: "I wasn't a fan of Iraq, I didn't want to go into Iraq."
```

## Extracting the explanation

```
In [24]: result0

Out[24]: <span class="short-desc"><strong>Jan. 21 </strong>"I wasn't a fan of Iraq, I didn't want to go into Iraq." <span class="short-truth"><a href="https://www.buzzfeed.com/andrewkaczynski/
k1/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a></span></span>

In [25]: result0.find('a')

Out[25]: <a href="https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a>

In [26]: result0.find('a').text[1:-1]

Out[26]: 'He was for an invasion before he was against it.'
```

## Extracting the Links

```
In [27]: result0.find('a')

Out[27]: <a href="https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the" target="_blank">(He was for an invasion before he was against it.)</a>

In [28]: result0.find('a')['href']

Out[28]: 'https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the'
```

## Notes for Beautiful Soup Methods & Attributes

### You can apply these two methods to either initialise soup object or a tag object

- find() :- searches for the first matching tag and returns a tag object
- find\_all() :- searches for all the matching tags and returns a resultant object as a list

### You can extract information from a tag object using these two attributes

- text :- extracts the text of a tag object and returns a string
- contents :- extracts the children of a tag and returns a list of tags and strings

## Iterating

```
In [29]: record = []

for i in result:
    date = i.find('strong').text[0:7] + ', 2017'
    lie = i.contents[1][1:-2]
    exp = i.find('a').text[1:-1]
    link = i.find('a')['href']
    record.append((date,lie,exp,link))

In [30]: record[:3]
```

```
Out[30]: (('Jan. 21, 2017',
'I wasn't a fan of Iraq, I didn't want to go into Iraq.',
'He was for an invasion before he was against it.',
'https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the'),
('Jan. 21, 2017',
'A reporter for Time magazine -- and I have been on their cover 14 or 15 times. I think we have the all-time record in the history of Time magazine.',
'https://www.buzzfeed.com/andrewkaczynski/in-2002-donald-trump-said-he-supported-invading-iraq-on-the',
'https://www.nytimes.com/2017/01/23/us/politics...'),
('Jan. 23, 2017',
'A reporter for Time magazine -- and I have been on their cover 14 or 15 times. I think we have the all-time record in the history of Time magazine.',
'https://www.nytimes.com/2017/01/23/us/politics...'),
('Jan. 23, 2017',
'I'd rather have him -- you know, work with him on the Ukraine than standing and arguing about whether or not -- because that's whole thing was set up by the Democrats.',
'https://www.nytimes.com/2017/12/18/us/politics/donald-trump-congress-democrats.html'))
```

## Applying Tabular Data Structure

```
In [31]: import pandas as pd

In [32]: ny = pd.DataFrame(record, columns=['date','Trump lie','Explanation','Links'])

In [33]: ny.head()
```

```
Out[33]:
```

	date	Trump lie	Explanation	Links
0	Jan. 21, 2017	I wasn't a fan of Iraq, I didn't want to go in...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
1	Jan. 21, 2017	A reporter for Time magazine -- and I have been...	Trump was on the cover 11 times and Nixon appe...	http://nation.time.com/2013/11/06/10-things-ye...
2	Jan. 23, 2017	Between 3 million and 5 million illegal votes ...	There's no evidence of illegal voting.	https://www.nytimes.com/2017/01/23/us/politics...
3	Jan. 25, 2017	Now, the audience was the biggest ever. But th...	Official aerial photos show Obama's 2009 inaug...	https://www.nytimes.com/2017/01/21/us/politics...
4	Jan. 25, 2017	Take a look at the Pew reports (which show vot...	The report never mentioned voter fraud.	https://www.nytimes.com/2017/01/24/us/politics...

```
In [47]: ny.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 189 entries, 0 to 179
Data columns (total 4 columns):
#   column      Non-Null Count  Dtype
---  ---
0   date        189 non-null      datetime64[ns]
1   Trump lie   189 non-null      object
2   Explanation 189 non-null      object
3   Links       189 non-null      object
dtypes: datetime64[ns](1), object(3)
memory usage: 5.8+ MB
```

```
In [48]: ny['date'] = pd.to_datetime(ny['date'])

In [49]: ny.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 189 entries, 0 to 179
Data columns (total 4 columns):
#   column      Non-Null Count  Dtype
---  ---
0   date        189 non-null      datetime64[ns]
1   Trump lie   189 non-null      object
2   Explanation 189 non-null      object
3   Links       189 non-null      object
dtypes: datetime64[ns](1), object(3)
memory usage: 5.8+ MB
```

## Exporting the Dataset into csv file

```
In [50]: ny.to_csv('Trump Lie.csv')

In [188]: df = pd.read_csv('Trump Lie.csv')
df.head()
```

```
Out[188]:
```

Unnamed: 0	date	Trump lie	Explanation	Links	
0	0	2017-01-21	I wasn't a fan of Iraq, I didn't want to go in...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
1	1	2017-01-21	A reporter for Time magazine -- and I have been...	Trump was on the cover 11 times and Nixon appe...	http://nation.time.com/2013/11/06/10-things-ye...
2	2	2017-01-23	Between 3 million and 5 million illegal votes ...	There's no evidence of illegal voting.	https://www.nytimes.com/2017/01/23/us/politics...
3	3	2017-01-25	Now, the audience was the biggest ever. But th...	Official aerial photos show Obama's 2009 inaug...	https://www.nytimes.com/2017/01/21/us/politics...
4	4	2017-01-25	Take a look at the Pew reports (which show vot...	The report never mentioned voter fraud.	https://www.nytimes.com/2017/01/24/us/politics...

```
In [189]: x = df.date.value_counts().sum()
x

Out[189]: 189
```

```
In [190]: # Start index from 1 in DataFrame
df.index = 1

In [191]: df.head()
```

```
Out[191]:
```

Unnamed: 0	date	Trump lie	Explanation	Links	
1	0	2017-01-21	I wasn't a fan of Iraq, I didn't want to go in...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
2	1	2017-01-21	A reporter for Time magazine -- and I have been...	Trump was on the cover 11 times and Nixon appe...	http://nation.time.com/2013/11/06/10-things-ye...
3	2	2017-01-23	Between 3 million and 5 million illegal votes ...	There's no evidence of illegal voting.	https://www.nytimes.com/2017/01/23/us/politics...
4	3	2017-01-25	Now, the audience was the biggest ever. But th...	Official aerial photos show Obama's 2009 inaug...	https://www.nytimes.com/2017/01/21/us/politics...
5	4	2017-01-25	Take a look at the Pew reports (which show vot...	The report never mentioned voter fraud.	https://www.nytimes.com/2017/01/24/us/politics...

```
In [192]: # Deleting unwanted columns
df.drop(columns=['Unnamed: 0'], inplace=True)

In [193]: df.head()
```

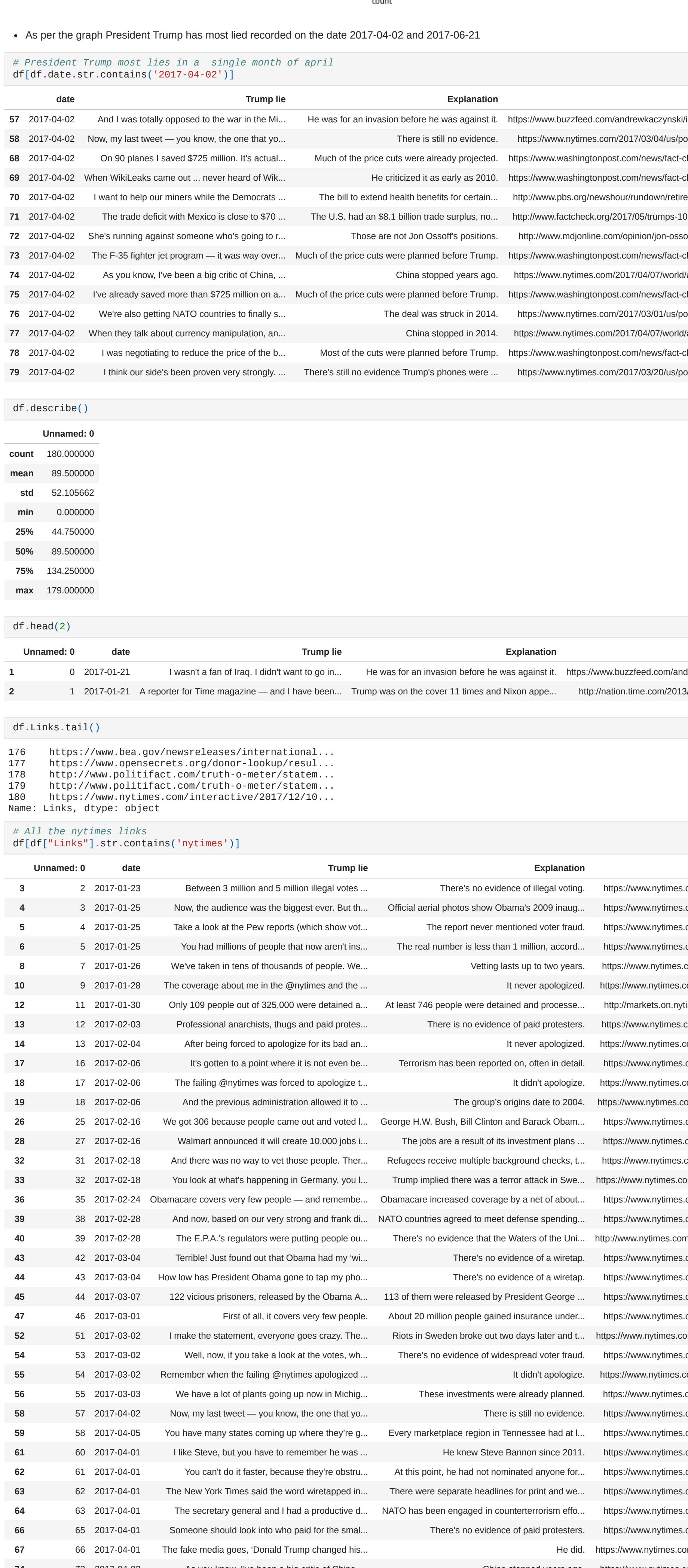
```
Out[193]:
```

	date	Trump lie	Explanation	Links
1	2017-01-21	I wasn't a fan of Iraq, I didn't want to go in...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
2	2017-01-21	A reporter for Time magazine -- and I have been...	Trump was on the cover 11 times and Nixon appe...	http://nation.time.com/2013/11/06/10-things-ye...
3	2017-01-23	Between 3 million and 5 million illegal votes ...	There's no evidence of illegal voting.	https://www.nytimes.com/2017/01/23/us/politics...
4	2017-01-25	Now, the audience was the biggest ever. But th...	Official aerial photos show Obama's 2009 inaug...	https://www.nytimes.com/2017/01/21/us/politics...
5	2017-01-25	Take a look at the Pew reports (which show vot...	The report never mentioned voter fraud.	https://www.nytimes.com/2017/01/24/us/politics...

```
In [56]: import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(15,15))
sns.countplot(date=df, y = 'date')
plt.plot()
```

```
Out[56]:
```



### As per the graph President Trump has most lied recorded on the date 2017-04-02 and 2017-06-21

```
In [173]: # President Trump most lies in a single month of april
df[df.date.str.contains('2017-04-02')]

In [173]:
```

```
Out[173]:
```

	date	Trump lie	Explanation	Links
57	2017-04-02	And I was totally opposed to the war in the M...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
58	2017-04-02	Now, my last tweet -- you know, the one that yo...	There is still no evidence.	https://www.nytimes.com/2017/03/04/us/politics...
68	2017-04-02	On 90 planes I saved \$725 million. It's actual...	Much of the price cuts were already projected.	https://www.washingtonpost.com/news/fact-check...
69	2017-04-02	When WikiLeaks came out ... never heard of Wik...	He criticized it as early as 2010.	https://www.washingtonpost.com/news/fact-check...
70	2017-04-02	I want to help our miners while the Democrats ...	The bill to extend health benefits for certain ...	https://www.pbs.org/newshour/undone/reviewed-co...
71	2017-04-02	The trade deficit with Mexico is close to \$70 ...	The U.S. had an \$8.1 billion trade surplus, no...	http://www.factcheck.org/2017/05/trumps-100-...
72	2017-04-02	She's running against someone who's going to r...	Those are not Jon Ossoff's positions.	http://www.msnonline.com/opinion/ny-on-ossoff-w...
73	2017-04-02	The F-35 fighter jet program -- it was way over...	Much of the price cuts were planned before Trump.	https://www.washingtonpost.com/news/fact-check...
74	2017-04-02	As you know, I've been a big critic of China ...	China stopped years ago.	https://www.nytimes.com/2017/04/01/world/asia...
75	2017-04-02	I've already saved more than \$725 million on a...	Much of the price cuts were planned before Trump.	https://www.washingtonpost.com/news/fact-check...
76	2017-04-02	We're also getting NATO countries to finally s...	The deal was struck in 2014.	https://www.nytimes.com/2017/03/31/us/politics...
77	2017-04-02	When they talk about currency manipulation, an...	China stopped in 2014.	https://www.nytimes.com/2017/04/01/world/asia...
78	2017-04-02	I was negotiating to reduce the price of the b...	Most of the cuts were planned before Trump.	https://www.washingtonpost.com/news/fact-check...
79	2017-04-02	I think our side's been proven very strongly ...	There's still no evidence Trump's phones were ...	https://www.nytimes.com/2017/03/20/us/politics...

```
In [57]: df.describe()

Out[57]:
```

Unnamed: 0	count	mean	std	min	25%	50%	75%	max
0	189	89.500000	52.105662	0	44.750000	89.500000	134.250000	179.000000

```
In [58]: df.head(2)

Out[58]:
```

Unnamed: 0	date	Trump lie	Explanation	Links	
1	0	2017-01-21	I wasn't a fan of Iraq, I didn't want to go in...	He was for an invasion before he was against it.	https://www.buzzfeed.com/andrewkaczynski/in-20...
2	1	2017-01-21	A reporter for Time magazine -- and I have been...	Trump was on the cover 11 times and Nixon appe...	http://nation.time.com/2013/11/06/10-things-ye...

```
In [77]: df.Links.tail()

Out[77]:
```

```
176 https://www.bna.gov/newsreleases/international...
177 https://www.opensecrets.org/donor-lookup/resul...
178 http://www.politifact.com/truth-o-meter/statem...
179 http://www.politifact.com/truth-o-meter/statem...
180 https://www.nytimes.com/interactive/2017/12/10...
Name: Links, dtype: object
```

```
In [73]: # All the nytimes links
df[df["Links"].str.contains('nytimes[]')]

Out[73]:
```

```
Out[73]:
```

Unnamed: 0	date	Trump lie	Explanation	Links	
3	2	2017-01-23	Between 3 million and 5 million illegal votes ...	There's no evidence of illegal voting.	https://www.nytimes.com/2017/01/23/us/politics...
4	3	2017-01-25	Now, the audience was the biggest ever. But th...	Official aerial photos show Obama's 2009 inaug...	https://www.nytimes.com/2017/01/21/us/politics...
5	4	2017-01-25	Take a look at the Pew reports (which show vot...	The report never mentioned voter fraud.	https://www.nytimes.com/2017/01/24/us/politics...
6	5	2017-01-25	You had millions of people that now aren't ins...	The real number is less than 1 million, accord...	https://www.nytimes.com/2017/03/01/us/politics...
8	7	2017-01-26	We've taken in tens of thousands of people. We...	Vetting lists up to two years.	https://www.nytimes.com/interactive/2017/01/2...
10	9	2017-01-28	The coverage about me in the @nytimes and the ...	It never apologized.	https://www.nytimes.com/2016/11/13/us/politics...
12	11	2017-01-30	Only 109 people out of 325,000 were detained ...	At least 746 people were detained and processe...	http://markets.on.nytimes.com/research/blackbo...
13	12	2017-02-03	Professional anarchists, thugs and paid protes...	There is no evidence of paid protesters.	https://www.nytimes.com/2017/02/19/world/euro...
14	13	2017-02-04	After being forced to apologize for its bad an...	It never apologized.	https://www.nytimes.com/2016/11/23/us/politics...
17	16	2017-02-06	It's gotten to a point where it is not even be...	Terrorism has been reported on, often in detai...	https://www.nytimes.com/2017/02/07/us/politics...
18	17	2017-02-06	The failing @nytimes was forced to apologize t...	It didn't apologize.	https://www.nytimes.com/2016/11/13/us/politics...
19	18	2017-02-06	And the previous administration allowed it to...	The group's origins date to 2004.	https://www.nytimes.com/2015/11/19/world/midd...
26	25	2017-02-16	We got 306 because people came out and voted I...	George H.W. Bush, Bill Clinton and Barack Obam...	https://www.nytimes.com/2017/02/01/us/politics...
28	27	2017-02-16	Wal-Mart announced it will create 10,000 jobs I...	The jobs are a result of its investment plans...	https://www.nytimes.com/interactive/2017/busin...
32	31	2017-02-18	And there was no way to vet those people. Ther...	Refugees receive multiple background checks, t...	https://www.nytimes.com/interactive/2017/01/2...
33	32	2017-02-18	You look at what's happening in Germany, you l...	Trump implied there was a terror attack in Swe...	https://www.nytimes.com/2017/02/19/world/euro...
36	35	2017-02-24	ObamaCare covers very very few people -- and r...	ObamaCare increased coverage by a net of about...	https://www.nytimes.com/2017/02/19/us/politics...
39	38	2017-02-28	The E.P.A.'s regulators were putting people o...	There's no evidence that the Waters of the Uni...	https://www.nytimes.com/2017/03/05/business/e...
40	42	2017-03-04	TenBrel just found out that Obama had my pho...	There's no evidence of a wiretap.	https://www.nytimes.com/2017/03/01/us/politics...
44	43	2017-03-04	How low has President Obama gone to tap my pho...	There's no evidence of a wiretap.	https://www.nytimes.com/2017/03/01/us/politics...
45	44	2017-03-07	122 vicious prisoners, released by the Obama A...	113 of them were released by President George...	https://www.nytimes.com/2017/03/05/us/politics...
47	46	2017-03-01	First of all, it covers very few people.	About 20 million people gained insurance unde...	https://www.nytimes.com/2017/03/01/us/politics...
52	51	2017-03-02	I make the statement, everyone goes crazy. The...	Riots in Sweden broke out two days later and ...	https://www.nytimes.com/2017/02/19/world/euro...
54	53	2017-03-02	Well, now, if you take a look at the votes, wh...	There's no evidence of widespread voter frau...	https://www.nytimes.com/2017/01/24/us/politics...
55	54	2017-03-02	Remember when the failing @nytimes apologized...	It didn't apologize.	https://www.nytimes.com/2016/11/23/us/politics...
56	55	2017-03-02	We have a lot of plants going on now in Michi...	These investments were already planned.	https://www.nytimes.com/interactive/2017/busin...
58	57	2017-04-02	Now, my last tweet -- you know, the one that yo...	There is still no evidence.	https://www.nytimes.com/2017/04/01/world/asia...
59	58	2017-04-05	You have many states coming up where they're g...	Every marketplace region in Tennessee had at l...	https://www.nytimes.com/2017/04/01/us/politics...
61	60	2017-04-01	I like Steve, but you have to remember he was ...	He knew Steve Bannon since 2011.	https://www.nytimes.com/2017/04/21/us/politics...
62	61	2017-04-01	You can't do it faster, because they're obstac...	At this point, he had not nominated anyone for...	https://www.nytimes.com/2017/04/21/us/politics...
63	62	2017-04-01	The New York Times said the word wreathed in...	There were separate headlines for print and we...	https://www.nytimes.com/2017/03/23/us/politics...
64	63	2017-04-01	The secretary general and I had a productive d...	NATO has been engaged in no endorsement effo...	https://www.nytimes.com/2017/04/01/us/politics...
66	65	2017-04-01	Someone should look into who paid for the smel...	There's no evidence of counter paid protesters.	https://www.nytimes.com/2017/04/01/world/asia...
67	66	2017-04-01	The fake media goes, "Donald Trump changed his...	He did.	https://www.nytimes.com/2017/04/01/business/med...
74	73	2017-04-02	As you know, I've been a big critic of China ...	China stopped years ago.	https://www.nytimes.com/2017/04/01/world/asia...
76	75	2017-04-02	When they talk about currency manipulation, fin...	The deal was struck in 2014.	https://www.nytimes.com/2017/03/31/us/politics...
77	76	2017-04-02	We're also getting NATO countries to finally s...	China stopped in 2014.	https://www.nytimes.com/2017/04/01/world/asia...
79	78	2017-04-02	I think our side's been proven very strongly ...	There's still no evidence Trump's phones were ...	https://www.nytimes.com/2017/04/25/us/politics...
80	79	2017-05-01	I, well, we are protecting pre-existing conditi...	The bill weakens protections for people with p...	https://www.nytimes.com/2017/04/25/us/politics...
82	81	2017-05-01	First of all, since I started running, they ha...	China stopped years ago.	https://www.nytimes.com/2017/04/25/us/politics...
84	83	2017-05-04	Number two, they're actually not a currency [m...	China stopped years ago.	https://www.nytimes.com/2017/04/25/us/politics...
87	86	2017-05-08	You know we've gotten billions of dollars more...	The deal was struck in 2014.	https://www.nytimes.com/2017/03/31/us/politics...
89	88	2017-05-08	Director Clapper reiterated what everybody's in...	Clapper only said he wasn't aware of an invest...	https://www.nytimes.com/2017/05/01/us/politics...
90	89	2017-05-12	Again, the story that there was collusion betw...	The F.B.I. was investigating before the electi...	https://www.nytimes.com/2017/05/01/us/politics...
91	90	2017-05-12	When James Clapper himself, and virtually ever...	Clapper said he wouldn't have been told of an ...	https://www.nytimes.com/2017/05/01/us/politics...
99	98	2017-06-21	They say it's "bordering" like hell it is. ...	The Paris climate agreement is non-binding -- an...	https://www.nytimes.com/2017/06/21/us/politics...
100	100	2017-06-21	If [Benji] have a problem in the middle of th...	The Obama environmental rule is limit polluti...	https://www.nytimes.com/2017/06/21/us/politics...
104	103	2017-06-21	Gary Cohen just said \$200 million in tax in ord...	Cohn sold Goldman Sachs stock worth \$220 mill...	https://www.nytimes.com/2017/06/23/us/politics...
113	112	2017-07-17	We've spread more lies -- and I'm talking abou...	Clinton, Clinton, Clinton, and F.D.R. had signed...	https://www.nytimes.com/2017