```
In [1]: import requests
import bs4
import pickle

In [2]: def url_to_transcript(url):
    res = requests.get(url)
    data = bs4.BeautifulSoup(res.text,'lxml')

    text = [p.text for p in data.find(class_="elementor-widget-theme-post-content").find_all('p') ]
    print(url)
    return text

In [3]: url = r'https://scrapsfromtheloft.com/2018/12/15/vir-das-losing-it-transcript/'
```

```
In [3]: url = r'https://scrapsfromtheloft.com/2018/12/15/vir-das-losing-it-transcript/'
text = url_to_transcript(url)
```

https://scrapsfromtheloft.com/2018/12/15/vir-das-losing-it-transcript/ (https://scrapsfromtheloft.com/2018/12/15/vir-das-losing-it-transcript/)

In [4]: text

'My mom had to get a job. She had never worked before, so she did the one thing that women who spoke good English in Ind ia did. She read the news for Doordarshan. Americans, Doordarshan is like our CNN, but with credibility. With no training or formal experience, my mother was Doordarshan's top newsreader in one week. That's all it took. Yeah. That's how bad Do ordarshan's English was at the time. I don't know what happened at her first job interview. They're like, "Mrs. Das, can you pronounce 'epsinocage'?" "Do you mean 'espionage'?" "You're hired, madam! We've been wondering for ten years."', 'I changed schools. I went from India's top private boarding school, The Lawrence School, Sanawar… Yeah, yeah, damaged g oods. ...to... From Lawrence School, Sanawar, to Delhi Public School. Yeah, that was a great transition. I went from an Engli sh medium school to a school where they spoke English medium. I went from nouns, pronouns and verbs to chest, shoulder, t riceps. I went from Shakespeare, Byron and Keats to, "Eh, madam, eh!" I couldn't do the examination thing. Here's how bad my results were in Delhi Public School. My first PTA meeting, the teacher sat my parents down and she was like, "Mr... and Mrs. Das... Vir has issues... that need to be... addressed." And then I realized she was talking to my parents really slowly. B ecause I think she just assumed the stupidity ran through my family. And the thing is, I'm not stupid. I'm just dyslexic. You guys know what dyslexia is? It's a reading-writing disorder. You jumble up letters. When I was in school, that shit d idn't exist. Like, if I went to my parents, "Mom, Dad, I'm dyslexic," they'd be like, "Shut up, boys are not meant to be

issue in my life. I can't write joint letters. I remember the first time I told my mom, typical Indian mom, I was like, "Mom, I can't write joint letters." And she was like, "Why don't you write all the letters first and join them later?" "B ecause I'm writing an essay. Mom, not designing a fucking freeway." So I had to write all of my exams in block letters, s

with boys." But... By the way, if you're dyslexic, being gay is a great choice. Plus, LGBTO just sounds like a dyslexic kid trying to spell a big word. Something like "lozenges" or... "logarithm." Like, till date, I can't write cursive. It's a big

```
In [6]: comedians = ['Vir', 'dave', 'ricky', 'bo', 'bill', 'jim', 'john', 'hasan', 'ali', 'anthony', 'mike', 'joe']
```

8/2/2021 31july - Jupyter Notebook

In [7]: transcripts = [url_to_transcript(u) for u in urls]

```
https://scrapsfromtheloft.com/2018/12/15/vir-das-losing-it-transcript/ (https://scrapsfromtheloft.com/2018/12/15/vir-das-losi
ng-it-transcript/)
https://scrapsfromtheloft.com/2020/11/08/dave-chappelle-snl-monologue-2020-transcript/ (https://scrapsfromtheloft.com/2020/1
1/08/dave-chappelle-snl-monologue-2020-transcript/)
https://scrapsfromtheloft.com/2018/03/15/ricky-gervais-humanity-transcript/ (https://scrapsfromtheloft.com/2018/03/15/ricky-gervais-humanity-transcript/)
ervais-humanity-transcript/)
https://scrapsfromtheloft.com/2017/08/06/bo-burnham-make-happy-2016-full-transcript/ (https://scrapsfromtheloft.com/2017/08/0
6/bo-burnham-make-happy-2016-full-transcript/)
https://scrapsfromtheloft.com/2017/05/24/bill-burr-im-sorry-feel-way-2014-full-transcript/ (https://scrapsfromtheloft.com/201
7/05/24/bill-burr-im-sorry-feel-way-2014-full-transcript/)
https://scrapsfromtheloft.com/2017/04/21/jim-jefferies-bare-2014-full-transcript/ (https://scrapsfromtheloft.com/2017/04/21/j
im-jefferies-bare-2014-full-transcript/)
https://scrapsfromtheloft.com/2017/08/02/john-mulaney-comeback-kid-2015-full-transcript/ (https://scrapsfromtheloft.com/2017/
08/02/john-mulanev-comeback-kid-2015-full-transcript/)
https://scrapsfromtheloft.com/2020/05/21/hasan-minhaj-homecoming-king-transcript/ (https://scrapsfromtheloft.com/2020/05/21/h
asan-minhaj-homecoming-king-transcript/)
https://scrapsfromtheloft.com/2017/09/19/ali-wong-baby-cobra-2016-full-transcript/ (https://scrapsfromtheloft.com/2017/09/19/
ali-wong-baby-cobra-2016-full-transcript/)
https://scrapsfromtheloft.com/2017/08/03/anthony-jeselnik-thoughts-prayers-2015-full-transcript/ (https://scrapsfromtheloft.c
om/2017/08/03/anthony-jeselnik-thoughts-prayers-2015-full-transcript/)
https://scrapsfromtheloft.com/2018/03/03/mike-birbiglia-my-girlfriends-boyfriend-2013-full-transcript/ (https://scrapsfromthe
loft.com/2018/03/03/mike-birbiglia-my-girlfriends-boyfriend-2013-full-transcript/)
https://scrapsfromtheloft.com/2017/08/19/joe-rogan-triggered-2016-full-transcript/ (https://scrapsfromtheloft.com/2017/08/19/
joe-rogan-triggered-2016-full-transcript/)
```

```
In [8]: transcripts[3]
 Out[8]: ['[woman on TV] That has been, really, a difficult thing for me. My mother has always been a very difficult person all he
         r life. Very unhappy. I can never remember my mother being happy. [inhales deeply] And as she gets older, however she was
         as a young person, I find it's just now exaggerated and, uh... so, she's, uh, 83 and now very extremely negative, very extr
         emely unhappy, and, uh... that has been hard for me... -[classical music playing]',
          '[audience cheering and applauding] [woman]',
          'Hello, audience. Thank you for coming. You are here because you want to laugh... and you want to forget about your proble
         ms. But I cannot allow it. You should not laugh. You should not forget about your problems. The world is not funny. -We a
         re all dving.',
          '[audience laughing]',
          'The world is not funny. Twelve percent of the world's population does not have access to clean drinking water. The worl
         d is not funny. Guy Fieri owns two functioning restaurants. -[audience laughing and applauding] -The world is not funny.
         So then... now that we understand the context, now that we realize how terribly unfunny the world is... let's do this. -[orga
         n music playing] -[audience cheering and applauding] [man] Entertainers are not here to help you. They are here to contro
         l you. -[audience continues cheering] -Never listen to them. Never give them what they want. And most importantly, never
         make some fucking noise! [audience cheers loudly]',
          '[hip-hop music playing] [audience continues cheering] -[bell dings] -[rapping] ﴾ Ladies! ﴾ ﴾ Ladies, if you feel me Sa
         y, "Hell, yeah!" ♪ [all women] Hell, yeah! ♪ Fellas! Come on, fellas If you feel me, say, "Hell, yeah!" ♪ [all men] Hell,
         yeah! ♪ Virgins, if you haven't felt a person Say, "Hell, yeah!" ♪ [audience] Hell, yeah! ♪ If you can divide by zero Let
         me hear you say, "Hell, yeah!" ♪ [audience] Hell, yeah! -No, you can't. -[audience laughing] Mathematically impossible. L
 In [9]: | for i,c in enumerate(comedians):
             with open(r'C:\Users\raj20\OneDrive\Desktop\AD\transcript/'+c+'.txt','wb') as file:
                 pickle.dump(transcripts[i],file)
In [10]: | data = {}
         for i,c in enumerate(comedians):
             data[c]=transcripts[i]
```

```
In [11]: data
Out[11]: {'Vir': ['I lost 80% of my mind. It's very freeing. You should see the look on your faces right now, by the way. Oh! Good
         evening, San Francisco. Are you guys excited, yeah? All right. Well, my name is Vir Das. We're gonna have such a good tim
         e tonight. I'm so excited. It's gonna be delightful. Oh, this is how I talk now. I just thought it was time to really emb
         race my roots, you know, and to make my comedy more authentically Indian. And really, what could be more Indian than a fa
         ke American accent? I don't think you understand. I have an opportunity to make history tonight, guys. I can. I can be th
         e first ever Indian who comes to California... and then leaves. That's never happened before. Because you guys are sticking
         around. Until they kick you out. Which, going by the news, is about three weeks from now.',
           'Now, I used to work in America in the year 2002. It didn't go very well, so I left for browner pastures. And then, hon
         estly, I didn't think about you guys for 15 years. Because, honestly, there is nothing that you can get in America that I
         can't get in India. And then my government banned beef and I was like, "You know, an international career might not be su
         ch a bad thing!" Make no mistake. I'm just here for the beef. It's been a good couple of years for me. I went on my first
         world tour. I saw the entire world. I went... Yeah. Would you like to know what 33 countries in the world have in common? -
         Would you like to know, yeah? - Yeah! Two things. Number one,\xa0I have now masturbated in all 33 of those countries. Tha
         nk you. I'm like the Genghis Khan of the Holiday Inn chain. My DNA is everywhere. If your hotel has a memory foam mattres
         s, I'm the memory.',
           'And two, no matter where I went in the entire world, people said the same thing. They said, "Oh, my God, you're India
         n? I love Indian people. They're, like, so smart. Indian people, you're, like, so smart." Which leads me to believe the r
         est of the world not that smart. Now, there's no answer to that question. Whenever somebody says, "Indian people, so smar
         t," all we can do is just, "Namaste," and, "Yes." "Okay." Because we all know the reality, right? In reality, 20% of Indi
In [12]: | data.keys()
Out[12]: dict keys(['Vir', 'dave', 'ricky', 'bo', 'bill', 'jim', 'john', 'hasan', 'ali', 'anthony', 'mike', 'joe'])
In [13]: def combine text(list of text):
             combined text = ' '.join(list of text)
             return combined text
```

localhost:8888/notebooks/31julv.ipynb

In [14]: data combined = {key: [combine text(value)] for key,value in data.items()}

In [15]: data_combined

Out[15]: {'Vir': ['I lost 80% of my mind. It's very freeing. You should see the look on your faces right now, by the way. Oh! Good evening, San Francisco. Are you guys excited, yeah? All right. Well, my name is Vir Das. We're gonna have such a good tim e tonight. I'm so excited. It's gonna be delightful. Oh, this is how I talk now. I just thought it was time to really emb race my roots, you know, and to make my comedy more authentically Indian. And really, what could be more Indian than a fa ke American accent? I don't think you understand. I have an opportunity to make history tonight, guys. I can. I can be th e first ever Indian who comes to California... and then leaves. That's never happened before. Because you guys are sticking around. Until they kick you out. Which, going by the news, is about three weeks from now. Now, I used to work in America in the year 2002. It didn't go very well, so I left for browner pastures. And then, honestly, I didn't think about you gu ys for 15 years. Because, honestly, there is nothing that you can get in America that I can't get in India. And then my g overnment banned beef and I was like, "You know, an international career might not be such a bad thing!" Make no mistake. I'm just here for the beef. It's been a good couple of years for me. I went on my first world tour. I saw the entire worl d. I went... Yeah. Would you like to know what 33 countries in the world have in common? - Would you like to know, yeah? -Yeah! Two things. Number one,\xa0I have now masturbated in all 33 of those countries. Thank you. I'm like the Genghis Kha n of the Holiday Inn chain. My DNA is everywhere. If your hotel has a memory foam mattress, I'm the memory. And two, no m atter where I went in the entire world, people said the same thing. They said, "Oh, my God, you're Indian? I love Indian people. They're, like, so smart. Indian people, you're, like, so smart." Which leads me to believe the rest of the world not that smart. Now, there's no answer to that question. Whenever somebody says, "Indian people, so smart," all we can do is just, "Namaste," and, "Yes." "Okay." Because we all know the reality, right? In reality, 20% of India is smart. Eighty percent of India is so stupid, we don't even give them a passport. Eighty percent of India is just Being Human t-shirts a

In [16]: for key,value in data.items(): print(key)

Vir
dave
ricky
bo
bill
jim
john
hasan
ali
anthony
mike
joe

```
In [17]: data_combined['dave']
```

Out[17]: ['Original air date: November 07, 2020 \xa0 \xa0 [Announcer] Ladies and gentlemen — Dave Chappelle! [Cheers and applause] Thank you. Thank you. [Cheers and applause] Thank you all for being here. [Cheers and applause] Pretty incredible day. [C heers and applause] You know, you know what I was thinking about all day today? This is really weird. First of all, I'm n ervous. I should tell you that. You can't tell, but, all inside. I was thinking about a person I never actually met, but I heard about all my life. I was thinking about my great grandfather, who apparently, by all accounts, was a very great m an. He was born a slave in South Carolina. Was a slave for ten years of his life. And when the northerners came down, the y started educating some of the newly freed black children, and he learned how to read. Got enamored with education and d edicated his life to three things: Education, freedom of black people, Jesus Christ. Became a juggernaut in the AME Churc h. It was a pretty amazing story my great grandfather had, and I thought about him all day today because I wish I could s ee him now. And I wish he could see me, because I wonder what he would say. This week I flew to New York on a private jet to host the Saturday Night Live. Netflix started streaming a show that bears his name, The Chappelle Show. And HBO Max is streaming it. And I didn't get paid for any of it. [Laughter] Yeah, if he could see me now he'd probably be like, "this n iggagot bought and sold more than I have." [Laughter] [applause] This morning after the results came in, got a text f rom a friend of mine in London. And she said, "The world feels like a safer place now that America has a new president." And I said, "That's great, but America doesn't. [Laughter] Do you guys remember what life was like before Covid? I do. Th ere was a mass shooting every week. Anyone remember that? Thank god for Covid. [Laughter] Someone had to lock these murde rous whites up, keep them in the house. [Laughter and applause] All summer long, you know what I've been doing? I've been doing shows in Ohio. I live in a small town in Ohio. And a lot of these small towns in America was dying. My town was dyi ng. So what I did is I did shows in my neighbor's cornfield, and these shows were very successful and may have even helpe APPEAR AND STATE

```
In [18]: import pandas as pd
    pd.set_option('max_colwidth',150)
    data_df = pd.DataFrame.from_dict(data_combined).transpose()
    data_df.columns=['transcripts']
```

transcripts

Out[19]:

```
In [19]: data_df
```

Vir I lost 80% of my mind. It's very freeing. You should see the look on your faces right now, by the way. Oh! Good evening, San Francisco. Are you gu... Original air date: November 07, 2020 [Announcer] Ladies and gentlemen — Dave Chappelle! [Cheers and applause] Thank you. Thank you. [Cheers an... dave ricky Hello. Hello! How you doing? Great. Thank you. Wow. Calm down. Shut the fuck up. Thank you. What a lovely welcome. I'm gonna try my hardest tonigh... bo [woman on TV] That has been, really, a difficult thing for me. My mother has always been a very difficult person all her life. Very unhappy. I can... bill [cheers and applause] All right, thank you! Thank you very much! Thank you. Thank you. Thank you. How are you? What's going on? Thank you. It's a ... jim [Car horn honks] [Audience cheering] [Announcer] Ladies and gentlemen, please welcome to the stage Mr. Jim Jefferies! [Upbeat music playing] Hello... john Armed with boyish charm and a sharp wit, the former "SNL" writer John Mulaney offers sly takes on marriage, his beef with babies and the time he m... hasan [theme music: orchestral hip-hop] [crowd roars] What's up? Davis, what's up? I'm home. I had to bring it back here. Netflix said, "Where do you wa... ali Ladies and gentlemen, please welcome to the stage: Ali Wong! Hi. Hello! Welcome! Thank you! Thank you for coming. Hello! Hello. We are gonna have ... anthony Thank you. Thank you, Thank you, San Francisco. Thank you so much. So good to be here. People were surprised when I told 'em I was gonna tape my s... mike Wow. Hey, thank you. Thanks. Thank you, guys. Hey, Seattle. Nice to see you. Look at this. Look at us. We're here. This is crazy. It's insane. So ... joe [rock music playing] [audience cheering] [announcer] Ladies and gentlemen, welcome Joe Rogan. [audience cheering and applauding] What the fuck is ...

Out[20]: transcripts Original air date: November 07, 2020 [Announcer] Ladies and gentlemen — Dave Chappelle! [Cheers and applau se] Thank you. Thank you. [Cheers an...

Name: dave, dtype: object

```
In [21]: data_df.loc['dave']['transcripts']
```

Out[21]: 'Original air date: November 07, 2020 \xa0 \xa0 [Announcer] Ladies and gentlemen — Dave Chappelle! [Cheers and applause] Thank you. Thank you. [Cheers and applause] Thank you all for being here. [Cheers and applause] Pretty incredible day. [C heers and applause] You know, you know what I was thinking about all day today? This is really weird. First of all, I'm n ervous. I should tell you that. You can't tell, but, all inside. I was thinking about a person I never actually met, but I heard about all my life. I was thinking about my great grandfather, who apparently, by all accounts, was a very great m an. He was born a slave in South Carolina. Was a slave for ten years of his life. And when the northerners came down, the y started educating some of the newly freed black children, and he learned how to read. Got enamored with education and d edicated his life to three things: Education, freedom of black people, Jesus Christ. Became a juggernaut in the AME Churc h. It was a pretty amazing story my great grandfather had, and I thought about him all day today because I wish I could s ee him now. And I wish he could see me, because I wonder what he would say. This week I flew to New York on a private jet to host the Saturday Night Live. Netflix started streaming a show that bears his name, The Chappelle Show. And HBO Max is streaming it. And I didn't get paid for any of it. [Laughter] Yeah, if he could see me now he'd probably be like, "this n iggagot bought and sold more than I have." [Laughter] [applause] This morning after the results came in, got a text f rom a friend of mine in London. And she said, "The world feels like a safer place now that America has a new president." And I said, "That's great, but America doesn't. [Laughter] Do you guys remember what life was like before Covid? I do. Th ere was a mass shooting every week. Anyone remember that? Thank god for Covid. [Laughter] Someone had to lock these murde rous whites up, keep them in the house. [Laughter and applause] All summer long, you know what I've been doing? I've been doing shows in Ohio. I live in a small town in Ohio. And a lot of these small towns in America was dying. My town was dyi ng. So what I did is I did shows in my neighbor's cornfield, and these shows were very successful and may have even helpe At the state of the State

```
import re
import string
def clean_text_round1(text):
    text = text.lower()
    text = re.sub('\[.*?\]','',text)
    text = re.sub('[%s]'%re.escape(string.punctuation),'',text)
    text = re.sub('\w*\d\w*','',text)
    return text
```

```
In [23]: round1 = lambda x: clean_text_round1(x)
In [24]: import re import string re.escape(string.punctuation)
```

Out[24]: '!"\\#\\\$%\\&\'\\(\\)*\\+,\\-\\./:;<=>\\?@\\[\\\\\]\\^_`\\{\\|\\}\\~'

In [25]: data_df

Out[25]:

transcripts

Vir I lost 80% of my mind. It's very freeing. You should see the look on your faces right now, by the way. Oh! Good evening, San Francisco. Are you gu... Original air date: November 07, 2020 [Announcer] Ladies and gentlemen — Dave Chappelle! [Cheers and applause] Thank you. Thank you. [Cheers an... dave ricky Hello. Hello! How you doing? Great. Thank you. Wow. Calm down. Shut the fuck up. Thank you. What a lovely welcome. I'm gonna try my hardest tonigh... bo [woman on TV] That has been, really, a difficult thing for me. My mother has always been a very difficult person all her life. Very unhappy. I can... bill [cheers and applause] All right, thank you! Thank you very much! Thank you. Thank you. Thank you. How are you? What's going on? Thank you. It's a ... jim [Car horn honks] [Audience cheering] [Announcer] Ladies and gentlemen, please welcome to the stage Mr. Jim Jefferies! [Upbeat music playing] Hello... john Armed with boyish charm and a sharp wit, the former "SNL" writer John Mulaney offers sly takes on marriage, his beef with babies and the time he m... hasan [theme music: orchestral hip-hop] [crowd roars] What's up? Davis, what's up? I'm home. I had to bring it back here. Netflix said, "Where do you wa... ali Ladies and gentlemen, please welcome to the stage: Ali Wong! Hi. Hello! Welcome! Thank you! Thank you for coming. Hello! Hello. We are gonna have ... anthony Thank you. Thank you, Thank you, San Francisco. Thank you so much. So good to be here. People were surprised when I told 'em I was gonna tape my s... mike Wow. Hey, thank you. Thanks. Thank you, guys. Hey, Seattle. Nice to see you. Look at this. Look at us. We're here. This is crazy. It's insane. So ... joe [rock music playing] [audience cheering] [announcer] Ladies and gentlemen, welcome Joe Rogan. [audience cheering and applauding] What the fuck is ...

```
In [26]: data_clean = pd.DataFrame(data_df.transcripts.apply(round1))
    data_clean
```

Out[26]:

Vir

dave

ricky

bo bill

jim

john

hasan

anthony

mike

joe

ali

transcripts

i lost of my mind it's very freeing you should see the look on your faces right now by the way oh good evening san francisco are you guys excited...

original air date november ladies and gentlemen — dave chappelle thank you thank you all for being here pretty incredible day ...

hello hello how you doing great thank you wow calm down shut the fuck up thank you what a lovely welcome i'm gonna try my hardest tonight you're t...

that has been really a difficult thing for me my mother has always been a very difficult person all her life very unhappy i can never remember my...

all right thank you thank you very much thank you thank you how are you what's going on thank you it's a pleasure to be here in the gre...

ladies and gentlemen please welcome to the stage mr jim jefferies hello sit down sit down sit down sit down thank you boston i appre...

armed with boyish charm and a sharp wit the former "snl" writer john mulaney offers sly takes on marriage his beef with babies and the time he met...

what's up davis what's up i'm home i had to bring it back here netflix said "where do you want to do the special la chicago new york" i was like...

ladies and gentlemen please welcome to the stage ali wong hi hello welcome thank you thank you for coming hello hello we are gonna have to get thi...

thank you thank you thank you san francisco thank you so much so good to be here people were surprised when i told 'em i was gonna tape my special...

wow hey thank you thanks thank you guys hey seattle nice to see you look at this look at us we're here this is crazy it's insane so about five yea...
ladies and gentlemen welcome joe rogan what the fuck is going on san francisco thanks for coming i appreciate it god damn put your phone down ...

```
In [29]: def clean_text_round2(text):
    text=re.sub('\n','',text)
    text = re.sub(' +', ' ',text)
    return text
round2 = lambda x: clean_text_round2(x)
```

```
In [30]: data_clean = pd.DataFrame(data_clean.transcripts.apply(round2))
    data_clean
```

Out[30]:

transcripts

Vir i lost of my mind it's very freeing you should see the look on your faces right now by the way oh good evening san francisco are you guys excited ... dave original air date november ladies and gentlemen — dave chappelle thank you thank you thank you all for being here pretty incredible day you kn... ricky hello hello how you doing great thank you wow calm down shut the fuck up thank you what a lovely welcome i'm gonna try my hardest tonight you're t... bo that has been really a difficult thing for me my mother has always been a very difficult person all her life very unhappy i can never remember my... bill all right thank you thank you very much thank you thank you thank you how are you what's going on thank you it's a pleasure to be here in the gre... jim ladies and gentlemen please welcome to the stage mr jim jefferies hello sit down sit down sit down sit down sit down thank you boston i appreciat... john armed with boyish charm and a sharp wit the former "snl" writer john mulaney offers sly takes on marriage his beef with babies and the time he met... hasan what's up davis what's up i'm home i had to bring it back here netflix said "where do you want to do the special la chicago new york" i was like ... ali ladies and gentlemen please welcome to the stage ali wong hi hello welcome thank you thank you for coming hello hello we are gonna have to get thi... anthony thank you thank you thank you san francisco thank you so much so good to be here people were surprised when i told 'em i was gonna tape my special... mike wow hey thank you thanks thank you guys hey seattle nice to see you look at this look at us we're here this is crazy it's insane so about five yea... ladies and gentlemen welcome joe rogan what the fuck is going on san francisco thanks for coming i appreciate it god damn put your phone down fuc... joe