

# Live Coding - Module 3

## Rachel Holman

In this live coding session we will access three different datasets which we can access on the internet without having to supply any API keys or other credentials (we'll cover APIs with credentials next week).

The goal is to examine the structure of the data, decide what is metadata and what constitutes the content of the dataframe we are trying to build, and to use the various tools available to us to convert the data to a pandas dataframe in our Python environment.

We'll work on the following three examples together:

1. [Cocktail recipes](#) from [The Cocktail DB](#)
2. Data on the pages on [Wikipedia](#) that pop up when searching for the term "[Virginia](#)".  
(This search gets the first 500 hits, but there are 210391 results. If time, we will use the `sroffset` parameter described in the [API documentation](#) to get the full list)
3. Articles from [NewsAPI.org](#). We will need to register for a [free API key](#) to be able to use this API. We'll talk about the best practices for storing API keys and keeping them secret during Module 4.

```
In [1]: import numpy as np
import pandas as pd
import json
import requests
```

```
In [2]: url = "https://www.thecocktaildb.com/api/json/v1/1/filter.php?c=Cocktail"
r = requests.get(url)
r #response 200 means it worked without error!
```

```
Out[2]: <Response [200]>
```

```
In [3]: # output what was returned by the requests.get(): r.text
# output in json format: json.loads(r.text)
myjson = json.loads(r.text)
```

```
In [4]: # get drink name in first index of drinks
myjson['drinks'][0]['strDrink']
```

```
Out[4]: '155 Belmont'
```

```
In [5]: numbers = [1,2,3,4,5]
numbers
```

Out[5]: [1, 2, 3, 4, 5]

```
In [6]: # list comprehension
# to square every number, subtract 1, and take sqrt:
[np.sqrt(x**2 - 1) for x in numbers]
```

Out[6]: [0.0,  
1.7320508075688772,  
2.8284271247461903,  
3.872983346207417,  
4.898979485566356]

```
In [7]: # print out name of drink for every drink in the json
[d['strDrink'] for d in myjson['drinks']]
```

```
Out[7]: ['155 Belmont',
'57 Chevy with a White License Plate',
'747 Drink',
'9 1/2 Weeks',
"A Gilligan's Island",
'A True Amaretto Sour',
'A.D.M. (After Dinner Mint)',
'Al',
'Abbey Martini',
'Absolut Summertime',
'Absolutely Fabulous',
'Absolutly Screwed Up',
'Ace',
'Adam & Eve',
'Addington',
'Addison',
'Addison Special',
'Adios Amigos Cocktail',
'Afterglow',
'Alice Cocktail',
'Amaretto fizz',
'Aperol Spritz',
'Apple Highball',
'Apple Karate',
'Applejack',
'Aquamarine',
'Arizona Stingers',
'Arizona Twister',
'Army special',
'Autumn Garibaldi',
'Aviation',
'Bahama Mama',
'Banana Cream Pi',
"Bee's Knees",
'Bijou',
'Blue Hurricane',
'Blueberry Mojito',
'Bombay Cassis',
'Bora Bora',
'Boulevardier',
'Bounty Hunter',
'Brigadier',
'Broadside',
'Brooklyn',
'Butterfly Effect',
"Captain Kidd's Punch",
'Cherry Electric Lemonade',
'Cocktail Horse's Neck',
'Corn n Oil',
'Corpse Reviver',
'Cosmopolitan',
'Cosmopolitan Martini',
'Cream Soda',
'Dark Caipirinha',
'Death in the Afternoon',
'Dirty Martini',
'Dry Martini',
"Duchamp's Punch",
'Elderflower Caipirinha',
"Empellón Cocina's Fat-Washed Mezcal",
```

```
'Espresso Martini',  
'Espresso Rumtini',  
'Figgy Thyme',  
'Flaming Lamborghini',  
'French Martini',  
'French Negroni',  
'Frosé',  
'Funk and Soul',  
'Gagliardo',  
'Garibaldi Negroni',  
'Gimlet',  
'Gin and Soda',  
'Gin Basil Smash',  
'Gin Lemon',  
'Gin Rickey',  
'Gin Tonic',  
'Greyhound',  
'Honey Bee',  
'Hot Toddy',  
'Hunter's Moon',  
'Imperial Cocktail',  
'Irish Curdling Cow',  
'Jitterbug',  
'Kiwi Martini',  
'Lazy Coconut Paloma',  
'Lemon Drop',  
'Lemon Elderflower Spritzer',  
'Malibu Twister',  
'Mango Mojito',  
'Manhattan',  
'Martinez 2',  
'Martinez Cocktail',  
'Martini',  
'Mary Pickford',  
'Mauresque',  
'Miami Vice',  
'Michelada',  
'Midnight Mint',  
'Mojito',  
'Mojito Extra']
```

```
In [8]: # for making a json dataset  
pd.json_normalize(myjson, record_path=['drinks'])
```

Out [8]:

	strDrink	strDrinkThumb	idDrink
0	155 Belmont	https://www.thecocktaildb.com/images/media/dri...	15346
1	57 Chevy with a White License Plate	https://www.thecocktaildb.com/images/media/dri...	14029
2	747 Drink	https://www.thecocktaildb.com/images/media/dri...	178318
3	9 1/2 Weeks	https://www.thecocktaildb.com/images/media/dri...	16108
4	A Gilligan's Island	https://www.thecocktaildb.com/images/media/dri...	16943
...	...	...	...
95	Miami Vice	https://www.thecocktaildb.com/images/media/dri...	13936
96	Michelada	https://www.thecocktaildb.com/images/media/dri...	178343
97	Midnight Mint	https://www.thecocktaildb.com/images/media/dri...	14842
98	Mojito	https://www.thecocktaildb.com/images/media/dri...	11000
99	Mojito Extra	https://www.thecocktaildb.com/images/media/dri...	15841

100 rows x 3 columns

```
In [9]: url = 'https://en.wikipedia.org/w/api.php?action=query&list=search&srsearch=Vir
r = requests.get(url)
r
```

Out[9]: &lt;Response [200]&gt;

```
In [10]: myjson = json.loads(r.text)
#myjson
```

```
In [11]: wikidf = pd.json_normalize(myjson, record_path = ['query', 'search'])
wikidf
```

Out[11]:

	ns	title	pageid	size	wordcount	snippet	t
0	0	Virginia	32432	299870	26084	<span>&lt;span class="searchmatch"&gt;Virginia&lt;/span&gt;, off...</span>	27T
1	0	West Virginia	32905	183464	17649	<span>&lt;span class="searchmatch"&gt;Virginia&lt;/span&gt; is a...</span>	25T
2	0	Virginia Beach, Virginia	91239	141238	12558	<span>&lt;span class="searchmatch"&gt;Virginia&lt;/span&gt; Beac...</span>	27T
3	0	Virginia Woolf	32742	331505	31948	<span>Adeline <span>&lt;span class="searchmatch"&gt;Virginia&lt;/sp...</span></span>	26T
4	0	Virgínia	1392216	2035	26	<span>&lt;span class="searchmatch"&gt;Virgínia&lt;/span&gt; is a...</span>	27T
...	...	...	...	...	...	...	...
495	0	Ballston, Arlington, Virginia	603147	22831	1764	Ballston is a neighborhood in Arlington County...	24T
496	0	Christiansburg, Virginia	137687	28407	3282	Christiansburg (formerly Hans Meadows) is a to...	12T
497	0	Petersburg, Virginia	91268	78607	8248	Petersburg is an independent city in the Commo...	16T
498	0	1964 United States presidential election in Vi...	43161140	77986	1368	The 1964 United States presidential election i...	27T
499	0	Virginia Declaration of Rights	225550	21313	2681	<span>The <span>&lt;span class="searchmatch"&gt;Virginia&lt;/span&gt;</span> ...</span>	11T

500 rows x 7 columns

In [12]:

```
# harder example
# don't do this! in the future, don't paste your API key in your code like this
apikey = "38d9de88dc9d4dc09c5764977a193672"
url = 'https://newsapi.org/v2/everything'
parameters = {'apiKey': apikey,
              'q': 'hasbulla'} #search parameter
r = requests.get(url, params=parameters)
r
```

Out[12]: <Response [200]>

In [13]: myjson = json.loads(r.text)

In [14]: myjson['articles']

```

Out[14]: [{'source': {'id': None, 'name': 'Huffingtonpost.es'},
  'author': 'Sergio Coto',
  'title': 'Le manda esta imagen a su madre, le pregunta si sabe quién es y la
  continuación es ORO',
  'description': '<![CDATA[<p>Las conversaciones de WhatsApp han aparecido par
  a cambiarlo la forma de comunicarse, pero, por si fuera poco, también se han c
  onvertido en un arma de doble filo por las bromas y los memes que se acaban co
  mpartiendo y que terminan en redes sociales...'],
  'url': 'https://www.huffingtonpost.es/virales/le-manda-imagen-madre-le-pregu
  nta-continuacion-oro.html',
  'urlToImage': 'https://img.huffingtonpost.es/files/og_thumbnail/uploads/202
  3/06/24/imagen-de-la-conversacion.jpeg',
  'publishedAt': '2023-06-24T10:53:11Z',
  'content': 'Las conversaciones de WhatsApp han aparecido para cambiarlo la f
  orma de comunicarse, pero, por si fuera poco, también se han convertido en un
  arma de doble filo por las bromas y los memes que se acab... [+1240 chars]'},
  {'source': {'id': None, 'name': 'Clarion Ledger'},
  'author': 'Mississippi Clarion Ledger, David Eckert, Mississippi Clarion Led
  ger',
  'title': 'Ole Miss says goodbye to rivalry with Vanderbilt football with wil
  d video',
  'description': "See Ole Miss football's tribute to its rivalry with Vanderbi
  lt that won't be on the 2024 SEC schedule.",
  'url': 'https://www.clarionledger.com/story/sports/college/ole-miss/2023/06/
  15/ole-miss-football-sec-schedule-release-vanderbilt-video-memes-lane-kiffin/7
  0326945007/',
  'urlToImage': 'https://www.gannett-cdn.com/presto/2021/11/21/PJAM/928a4165-b
  271-4b42-97d1-db8f373b1111-AP21325156477323.jpg?auto=webp&crop=4629,2604,x0,y2
  35&format=pjpg&width=1200',
  'publishedAt': '2023-06-15T20:05:13Z',
  'content': "OXFORD The social media teams around the SEC got creative with t
  heir content in response to the conference's 2024 schedule release on Wednesda
  y. Mississippi State gifted us an AI-generated gator dres... [+1056 chars]"},
  {'source': {'id': 'marca', 'name': 'Marca'},
  'author': 'marca.com',
  'title': 'Prime Tyson Fury or prime Mike Tyson? Evander Holyfield makes his
  pick',
  'description': 'Tyson Fury and Mike Tyson are two names that will forever by
  synonymous with the boxing world, and sport in general. \nTheir names carry si
  gnificant weight and command a certain lev',
  'url': 'https://www.marca.com/en/boxing/2023/06/19/64904840ca474128368b45c3.
  html',
  'urlToImage': 'https://phantom-marca.unidadeditorial.es/10648f392870659578f1
  64b49c9e9019/resize/1200/f/jpg/assets/multimedia/imagenes/2023/06/19/168717743
  98550.jpg',
  'publishedAt': '2023-06-19T12:25:04Z',
  'content': 'Tyson Fury and Mike Tyson are two names that will forever by syn
  onymous with the boxing world, and sport in general. \r\nTheir names carry sig
  nificant weight and command a certain level of respect in a... [+1457 chars]'},
  {'source': {'id': None, 'name': 'Www.abc.es'},
  'author': '(abc)',
  'title': 'El Barça se redime de sus pecados',
  'description': 'PESTAÑA unicaja-barcelona-semis-tercero-acb22/23 Crónica 4 E
  l Barcelona al fin tuvo fe ante la dificultad. Tras un partido de intensidad m
  áxima en Málaga, los azulgranas sacaron a relucir todo su coraje y calidad en
  el último cuarto para acabar cerrando el du...',
  'url': 'https://www.abc.es/deportes/baloncesto/barca-redime-pecados-20230611
  204558-nt.html',
  'urlToImage': 'https://s2.abcstatics.com/abc/www/multimedia/deportes/2023/0
  6/11/niko_20230611203815-U62070467254sYd-1024x512@abc.jpg',

```

```
'publishedAt': '2023-06-11T18:45:58Z',
'content': 'El Barcelona al fin tuvo fe ante la dificultad. Tras un partido de intensidad máxima en Málaga, los azulgranas sacaron a relucir todo su coraje y calidad en el último cuarto para acabar cerrando el d... [+2723 chars]']}]
```

```
In [15]: pd.json_normalize(myjson, record_path=['articles'])
```

Out[15]:	author	title	description	url
0	Sergio Coto	Le manda esta imagen a su madre, le pregunta s...	<![CDATA[<p>Las conversaciones de WhatsApp han...	https://www.huffingtonpost.es/virales/le-manda... https:
1	Mississippi Clarion Ledger, David Eckert, Miss...	Ole Miss says goodbye to rivalry with Vanderbi...	See Ole Miss football's tribute to its rivalry...	https://www.clarionledger.com/story/sports/col... https:
2	marca.com	Prime Tyson Fury or prime Mike Tyson? Evander ...	Tyson Fury and Mike Tyson are two names that w...	https://www.marca.com/en/boxing/2023/06/19/649... https:
3	(abc)	El Barça se redime de sus pecados	PESTAÑA unicaja-barcelona-semis-tercero-acb22/...	https://www.abc.es/deportes/baloncesto/barca-r... https:

```
In [ ]:
```