

Lab 9 - APIs

```
In [21]: import pandas as pd
import numpy as np
import requests
from pandas.io.json import json_normalize #special package in pandas
import json
```

Part A: APIs

Wiki Dog Info

```
In [58]: # download the response from Wikipedia API endpoint https://en.wikipedia.org/w/api.php
# to get general information about the Dog Wikipedia article
# ask for the output as json
# HINT: prop = info
url = "https://en.wikipedia.org/w/api.php?action=query&prop=info&titles=Dog&format=json"
dog = requests.get(url)
# save output and print
print(dog.text)
```

```
{
  "batchcomplete": "",
  "query": {
    "pages": {
      "4269567": {
        "pageid": 4269567,
        "ns": 0,
        "title": "Dog",
        "contentmodel": "wikitext",
        "pagelanguage": "en",
        "pagelanguagehtmlcode": "en",
        "pagelanguagedir": "ltr",
        "touched": "2022-11-04T12:07:15Z",
        "lastrevid": 1119844781,
        "length": 114308
      }
    }
  }
}
```

```
In [59]: # convert json string into json object using json.loads (HINT: you need the text from
dog2revs_json = json.loads(dog.text)
```

Wiki Dog Revisions

```
In [60]: # download the response from Wikipedia API endpoint https://en.wikipedia.org/w/api.php
# to get the latest revision information (last revision) for the Dog Wikipedia article
# ask for the output as json
# HINT: prop = revisions
url2 = "https://en.wikipedia.org/w/api.php?action=query&prop=revisions&titles=Dog&format=json"
# save output and print
dog2 = requests.get(url2)
print(dog2.text)
```

```
{
  "batchcomplete": "",
  "query": {
    "pages": {
      "4269567": {
        "pageid": 4269567,
        "ns": 0,
        "title": "Dog",
        "revisions": [
          {
            "revid": 1119844781,
            "parentid": 1119844243,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:28:43Z",
            "comment": "/* Neutering */"
          }
        ]
      }
    }
  }
}
```

```
In [61]: # convert json string into json object using json.loads (HINT: you need the text from
dog2revs_json = json.loads(dog2.text)
```

```
In [62]: # flatten json data into a data frame where each row is a revision (HINT: you should use
# observation)
dog2revs_df = pd.json_normalize(dog2revs_json['query']['pages']['4269567']['revisions'])
```

```
# print dataframe
dog2revs_df
```

```
Out[62]:
```

	revid	parentid	user	timestamp	comment
0	1119844781	1119844243	Surtsicna	2022-11-03T18:28:43Z	/* Neutering */

```
In [63]: # download the last 15 revisions to the dog page
# ask for the output as json
# HINT: use rvlimit
url3 = "https://en.wikipedia.org/w/api.php?action=query&prop=revisions&rvlimit=15&titl
# save output and print
dog3 = requests.get(url3)
print(dog3.text)
```

```
{
  "continue": {
    "rvcontinue": "20220922143322|1111718614",
    "continue": ""
  },
  "query": {
    "page": {
      "4269567": {
        "pageid": 4269567,
        "ns": 0,
        "title": "Dog",
        "revisions": [
          {
            "revid": 1119844781,
            "parentid": 1119844243,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:28:43Z",
            "comment": "/* Neutering */",
            "revid": 1119844243,
            "parentid": 1119843617,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:25:30Z",
            "comment": "Link what you say and say what you link.",
            "revid": 1119843617,
            "parentid": 1119843506,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:21:24Z",
            "comment": "/* Breeds */",
            "revid": 1119843506,
            "parentid": 1119842745,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:20:41Z",
            "comment": "/* Domestication */",
            "revid": 1119842745,
            "parentid": 1119842514,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:15:45Z",
            "comment": "avoiding [[WP:Easter egg|Easter egg links]]",
            "revid": 1119842514,
            "parentid": 1118538971,
            "user": "Surtsicna",
            "timestamp": "2022-11-03T18:13:50Z",
            "comment": "simplifying links per [[WP:NOPIPE]] and [[MOS:NOPIPE]]",
            "revid": 1118538971,
            "parentid": 1118513977,
            "minor": "",
            "user": "Ira Leviton",
            "timestamp": "2022-10-27T15:21:41Z",
            "comment": "Fixed a pmc parameter in a citation. Please see [[Category:CS1 maint: PMC format]].",
            "revid": 1118513977,
            "parentid": 1114919057,
            "minor": "",
            "user": "Rheaxx666",
            "timestamp": "2022-10-27T12:04:31Z",
            "comment": "Added a citation on \"11 fixed genes show variation between wolves and dogs\".",
            "revid": 1114919057,
            "parentid": 1114596197,
            "minor": "",
            "user": "Kind Tennis Fan",
            "timestamp": "2022-10-08T23:23:20Z",
            "comment": "[[Wikipedia:MOSNUMscript|script]]-assisted date audit and style fixes per [[MOS:NUM]]",
            "revid": 1114596197,
            "parentid": 1114501414,
            "user": "Citation bot",
            "timestamp": "2022-10-07T08:52:00Z",
            "comment": "Add: website, title. Changed bare reference to CS1/2. | [[WP:UCB|Use this bot]]. [[WP:DEBUG|Report bugs]]. | Suggested by BrownHairedGirl | #UCB_webform 310/3680",
            "revid": 1114501414,
            "parentid": 1114482291,
            "minor": "",
            "user": "AnomieBOT",
            "timestamp": "2022-10-06T20:39:36Z",
            "comment": "Dating maintenance tags: {{Cn}}",
            "revid": 1114482291,
            "parentid": 1114363930,
            "user": "CandyStalnak",
            "timestamp": "2022-10-06T18:38:38Z",
            "comment": "added 29 years",
            "revid": 1114363930,
            "parentid": 1114347942,
            "user": "Jtrevor99",
            "timestamp": "2022-10-06T03:50:43Z",
            "comment": "Undid revision 1114347942 by [[Special:Contributions/CandyStalnak|CandyStalnak]] ([[User talk:CandyStalnak|talk]]) Not correct; the longest-lived dog on modern record was at least 29 years old",
            "revid": 1114347942,
            "parentid": 1113990620,
            "user": "CandyStalnak",
            "timestamp": "2022-10-06T01:47:47Z",
            "comment": "/* Lifespan */ dog life expectancy up to 22 years",
            "revid": 1113990620,
            "parentid": 1111718614,
            "user": "Rreagan007",
            "timestamp": "2022-10-04T06:59:24Z",
            "comment": ""
          ]
        ]
      }
    }
  }
}
```

```
In [64]: # convert json string into json object using json.loads (HINT: you need the text from
dog3revs_json = json.loads(dog3.text)
```

```
In [65]: # flatten json data into a data frame where each row is a revision
dog3revs_df = pd.json_normalize(dog3revs_json['query']['pages']['4269567']['revisions']
# count the number of rows
dog3revs_df.shape[0]
```

```
Out[65]: 15
```

```
In [66]: # print the first few rows
dog3revs_df.head()
```

Out[66]:

	revid	parentid	user	timestamp	comment	minor
0	1119844781	1119844243	Surtsicna	2022-11-03T18:28:43Z	/* Neutering */	NaN
1	1119844243	1119843617	Surtsicna	2022-11-03T18:25:30Z	Link what you say and say what you link.	NaN
2	1119843617	1119843506	Surtsicna	2022-11-03T18:21:24Z	/* Breeds */	NaN
3	1119843506	1119842745	Surtsicna	2022-11-03T18:20:41Z	/* Domestication */	NaN
4	1119842745	1119842514	Surtsicna	2022-11-03T18:15:45Z	avoiding [[WP:Easter egg Easter egg links]]	NaN

```
In [ ]:
```