

1.0-jo-extract-prostate-data

November 23, 2020

1 Extract Prostate Cancer Data From Textbook URL

```
[1]: import urllib.request
import os.path
```

```
[2]: # open connection to URL
webUrl = 'http://www.cnachtsheim-text.csom.umn.edu/Kutner/
↳Appendix%20C%20Data%20Sets/APPENC05.txt'
response = urllib.request.urlopen(webUrl)
data = response.read()

# decode bytes to string
data = data.decode('utf-8')
data = data.replace('\r', '')
```

```
[3]: # write to .txt file and store
filename = 'APPENC05'
out_filename = '../data/raw/' + filename + '.txt'
if not os.path.isdir(out_filename):
    output_file = open(out_filename, 'w')
    output_file.write(data)
    output_file.close()
```