

Public Data Portal Tutorial

Secret Key, Base Url

```
In [ ]: SERVICE_KEY = r'y%2BYI%2F3eYALrp6Z4%2Ff%2FoPhqKdk6hRwGfJqfxHVj98JuneHV6EeE5NqPEw
BASE_URL = 'http://apis.data.go.kr/1300000/MJBGJWJeopSuHH3/list'
```

Import libraries

```
In [ ]: import requests
import numpy as np
import pandas as pd
from bs4 import BeautifulSoup as bs
import xmltodict
import json
```

Get Data From API

```
In [ ]: def getData(pageNo):

    # init setting
    baseUrl = 'http://apis.data.go.kr/1300000/MJBGJWJeopSuHH3/list'
    serviceKey = 'y%2BYI%2F3eYALrp6Z4%2Ff%2FoPhqKdk6hRwGfJqfxHVj98JuneHV6EeE5NqF
    numOfRows = 1000

    # get request
    endPoint = f'{baseUrl}?pageNo={pageNo}&numOfRows={numOfRows}&ServiceKey={ser
    response = requests.get(endPoint)
    if (response.status_code != 200):
        raise Exception('Invalid Error')

    # xml to json
    data_dict = xmltodict.parse(response.text)

    return data_dict
```

Get Sample Data

There's 866 rows

So that to get all data, it's sufficient to get page 1 with
parameter numOfRows 1000

```
In [ ]: sampleData = getData(pageNo=1)
print(f"Total Elements : {sampleData['response']['body']['totalCount']}")
```

Total Elements : 866

Parsing Data

- mojipYy : 모집년도
- iyyjsijakYm : 입영시작월
- mojipYy : 모집년도
- gsteukgiCd : 군사특기명

- seonbalPcnt : 선발인원수

```
In [ ]: df = pd.DataFrame(sampleData['response']['body']['items']['item'])
print(f'Dataframe has shape with {df.shape[0]}')
df.head()
```

Dataframe has shape with 866

Out []:

| | extremes | gstekgiCd | gstekgiNm | gunGbnm | iybudaecdm | iyjjongryoYm | iyjsijal |
|---|----------|-----------|-----------|---------|------------|--------------|----------|
| 0 | 88 | A001 | 동반입대병 | 육군 | 7사단 | 202305 | 202 |
| 1 | 84 | A001 | 동반입대병 | 육군 | 12사단 | 202305 | 202 |
| 2 | 86 | A001 | 동반입대병 | 육군 | 15사단 | 202305 | 202 |
| 3 | 90 | A001 | 동반입대병 | 육군 | 21사단 | 202305 | 202 |
| 4 | 170 | A001 | 동반입대병 | 육군 | 1사단 | 202305 | 202 |

Save Raw Data

```
In [ ]: df.to_csv("raw_data.csv", index=False)
```