

# Python class 6

## 공공 데이터

이번 수업부터는 공공데이터 포털에서 얻은 데이터를 통해 빅데이터 수업을 진행하겠습니다.

## 서울특별시 관광지 입장정보

이번에는 공공데이터 중 '서울특별시 관광지 입장정보'를 url 요청을 통해 json으로 데이터를 만들어 보겠습니다.

우선 이전 수업과 마찬가지로 url요청을 위해 `urllib.request`를 `import` 하고, `json` 과 `math`를 `import` 하겠습니다.

```
import urllib.request
import json
import math
```

### get\_request\_url

이전과 마찬가지로 url 요청을 위한 `get_request_url`를 선언합니다.

```
def get_request_url(url):
    req=urllib.request.Request(url)
    try:
        response=urllib.request.urlopen(req)
        return response.read().decode('utf8')
    except Exception as e:
        print(e)
        return None
```

### getTourPointVisitor

`getTourPointVisitor`는 관광지에 입장한 방문객들의 정보를 년도, 도시, 구 정보에 따라 parameters에 추가하여 url요청을 하고 얻은 url로 부터 json형식으로 데이터를 로드합니다.

```
def getTourPointVisitor(yyyyymm, sido, gungu, nPagenum, nItems):

    access_key="lUb6ajXWCK0qhkX2v7TMP2TtcGNlJXgtfyoipeGaet9JEpCSgg6CNfqzWN%2BABnLF9gTFVgA
    tjNQe5Kh2ohxzyQ%3D%3D"

    end_point="http://openapi.tour.go.kr/openapi/service/TourismResourceStatsService/getP
    chrgTrsrstVisitorList"
    parameters = "?_type=json&serviceKey=" + access_key
    parameters += "&YM=" + yyyyymm
    parameters += "&SIDO=" + urllib.parse.quote(sido)
    parameters += "&GUNGU=" + urllib.parse.quote(gungu)
    parameters += "&RES_NM=&pageNo=" + str(nPagenum)
```

```

parameters += "&numOfRows=" + str(nItems)
url = end_point + parameters
retData = get_request_url(url)
if (retData == None):
    return None
else:
    return json.loads(retData)

```

### getTourPointData

getTourPointData 는 json 데이터로 부터 관광지 정보를 정리하여 리턴 해 줍니다.

```

def getTourPointData(item, yyyyymm, jsonResult):
    addrCd = 0 if 'addrCd' not in item.keys() else item['addrCd']
    gungu = '' if 'gungu' not in item.keys() else item['gungu']
    sido = '' if 'sido' not in item.keys() else item['sido']
    resNm = '' if 'resNm' not in item.keys() else item['resNm']
    rnum = 0 if 'rnum' not in item.keys() else item['rnum']
    ForNum = 0 if 'csForCnt' not in item.keys() else item['csForCnt']
    NatNum = 0 if 'csNatCnt' not in item.keys() else item['csNatCnt']
    jsonResult.append({'yyyyymm':yyyyymm, 'addrCd':addrCd, 'gungu':gungu, 'sido':sido,
'resNm':resNm, 'rnum':rnum, 'ForNum':ForNum, 'NatNum':NatNum})
    return

```

### main

main 에서는 위 메소드들을 호출하여 서울특별시\_관광지입장정보\_2011\_2017.json 라는 json파일을 생성합니다.

```

def main():
    jsonResult=[]
    sido='서울특별시'
    gungu=''
    nPagenum=1
    nTotal=0
    nItems=100
    nStartYear=2011
    nEndYear=2017

    for year in range(nStartYear, nEndYear):
        for month in range(1, 13):
            yyyyymm = str(year)+str(month)
            nPagenum = 1
            while True:
                jsonData = getTourPointVisitor(yyyyymm, sido, gungu, nPagenum, nItems)
                if (jsonData['response']['header']['resultMsg'] == 'OK'):
                    nTotal = jsonData['response']['body']['totalCount']
                    if nTotal == 0:
                        break
                    for item in jsonData['response']['body']['items']['item']:
                        getTourPointData(item, yyyyymm, jsonResult)
                    nPage = math.ceil(nTotal / 100)

```

```
        if (nPagenum == nPage):
            break
        nPagenum += 1
    else:
        break

    with open("%s_관광지입장정보_%d_%d.json" %(sido, nStartYear, nEndYear-1), 'w',
encoding='utf8') as outFile:
        retJson = json.dumps(jsonResult, indent=4, sort_keys=True,
ensure_ascii=False)
        outFile.write(retJson)

if __name__ == '__main__':
    main()
```