

Zigbang 원룸 매물 데이터 수집

```
In [1]: import requests
import pandas as pd
```

Process

- 동이름으로 위도 경도 구하기
- 위도 경도로 geohash 알아내기
- geohash로 매물 아이디 가져오기
- 매물 아이디로 매물 정보 가져오기

1. 동이름으로 위도 경도 구하기

```
In [2]: addr = "망원동"
url = f"https://apis.zigbang.com/v2/search?leaseYn=N&q={addr}\
&serviceType=원룸"
response = requests.get(url)
data = response.json()["items"][0]
lat, lng = data["lat"], data["lng"]
lat, lng
```

```
Out [2]: (37.556785583496094, 126.9013442993164)
```

2. 위도 경도로 geohash 알아내기

```
In [3]: # install geohash2
# !pip install geohash2
```

```
In [4]: import geohash2
```

```
In [5]: # precision이 커질수록 영역이 작아짐
geohash = geohash2.encode(lat, lng, precision=5)
geohash
```

```
Out [5]: 'wydix'
```

3. geohash로 매물 아이디 가져오기

```
In [6]: url = f'https://apis.zigbang.com/v2/items/oneroom?geohash={geohash}\
&depositMin=0&rentMin=0&salesTypes[0]=전세&salesTypes[1]=월세\
&domain=zigbang&checkAnyItemWithoutFilter=true'
response = requests.get(url)
items = response.json()["items"]
# len(items), items[0]
ids = [item["itemId"] for item in items]
len(ids), ids[:5]
```

Out [6]: (638, [39817718, 39817840, 39858920, 39791961, 39792041])

4. 매물 아이디로 매물 정보 가져오기

```
In [7]: # 1000개 넘어가면 나눠서 수집해야 함
url = "https://apis.zigbang.com/v2/items/list"
params = {
    "domain": "zigbang",
    "item_ids": ids
}
response = requests.post(url, params)
response
```

Out [7]: <Response [200]>

```
In [8]: data = response.json()["items"]
df = pd.DataFrame(data)
df.tail(2)
```

Out [8]:

	item_id	section_type	images_thumbnail	sales_type	sales_tit
--	---------	--------------	------------------	------------	-----------

636	39949470	None	https://ic.zigbang.com/ic/items/39949470/1.jpg	월세	올
-----	----------	------	--	----	---

637	39865827	None	https://ic.zigbang.com/ic/items/39865827/1.jpg	월세	올
-----	----------	------	--	----	---

2 rows × 35 columns

```
In [9]: # 필요한 컬럼만 필터링
columns = ["item_id", "sales_type", "deposit", "rent", "size_m2",
           "floor", "building_floor", "address1", "manage_cost"]
filtered_column_df = df[columns]
filtered_column_df.tail(2)
```

```
Out[9]:
```

	item_id	sales_type	deposit	rent	size_m2	floor	building_floor	address1	manage_cost
636	39949470	월세	1000	45	14.04	1	4	서울시 마포구 중동	
637	39865827	월세	2000	90	48.81	5	6	서울시 마포구 성산동	

```
In [10]: # 주소에 망원동이 있는 데이터만 필터링
result_df = filtered_column_df[
    filtered_column_df["address1"].str.contains("망원동")
]
result_df = result_df.reset_index(drop=True)
result_df.tail(2)
```

```
Out[10]:
```

	item_id	sales_type	deposit	rent	size_m2	floor	building_floor	address1	manage_cost
63	39920882	월세	5000	30	18.18	2	4	서울시 마포구 망원동	
64	39800258	월세	500	55	19.83	1	2	서울시 마포구 망원동	