

Zigbang 원룸 매물 데이터 수집

```
In [44]: import requests, json
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 [11]: 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]

#위도 경도
lat, lng = data['lat'], data['lng']
```

2. 위도 경도로 geohash 알아내기

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

```
Collecting geohash2
  Downloading geohash2-1.1.tar.gz (15 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Requirement already satisfied: docutils>=0.3 in c:\users\user\anaconda3\lib\site-packages (from
geohash2) (0.18.1)
Building wheels for collected packages: geohash2
  Building wheel for geohash2 (setup.py): started
  Building wheel for geohash2 (setup.py): finished with status 'done'
  Created wheel for geohash2: filename=geohash2-1.1-py3-none-any.whl size=15554 sha256=0bf7c5f
e53735c0849b2910518ffab0c94f455e22138e86a20bfb3633a64872f
  Stored in directory: c:\users\user\appdata\local\pip\cache\wheels\f6\7c\c4\1b3c6fea0ebc53bf7
30dc86bbbee7a713d501455dfb4c1f0623
Successfully built geohash2
Installing collected packages: geohash2
Successfully installed geohash2-1.1
```

```
In [15]: import geohash2 # 위도, 경도의 점으로 부터 영역
```

```
In [16]: lat, lng
```

```
Out[16]: (37.556785583496094, 126.9013442993164)
```

```
In [19]: # precision : 영역범위 : 커질수록 영역이 작아짐
geohash = geohash2.encode(lat, lng, precision=5)
geohash
```

```
Out[19]: 'wydjx'
```

3. geohash로 매물 아이디 가져오기

```
In [31]: 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)
response
```

```
Out[31]: <Response [200]>
```

```
In [43]: items = response.json()['items']
ids = [item['itemId'] for item in items]
len(ids), ids[:5]
```

```
Out[43]: (682, [40042434, 40058332, 40114131, 40184360, 40177503])
```

4. 매물 아이디로 매물 정보 가져오기

```
In [46]: url = 'https://apis.zigbang.com/v2/items/list'
params = {'domain': 'zigbang',
          'item_ids': ids}

response = requests.post(url, params)
response
```

```
Out[46]: <Response [200]>
```

```
In [52]: response.text[:500]
```

```
Out[52]: '{"items":[{"item_id":40042434,"section_type":null,"images_thumbnail":"https://ic.zigbang.com/
ic/items/40042434/1.jpg","sales_type":"월세","sales_title":"월세","deposit":1000,"rent":50,"si
ze_m2":22.48,"공급면적":{"m2":22.48,"p":"6.8"},"전용면적":{"m2":22.48,"p":"6.8"},"계약면적":nu
ll,"room_type_title":null,"floor":"5","floor_string":"5","building_floor":"9","title":"당산역
부근 깔끔함 신축급컨디션 일단클릭","is_first_movein":null,"room_type":"01","status":true,"tag
s":["추천"],"service_type":"원룸","random_location":{"lat":37.52910515'
```

```
In [53]: _data = response.json()['items']
```

```
In [60]: # 최대 보여 줄 수 있는 컬럼
# pd.options.display.max_columns
```

```
In [65]: jib_data = pd.DataFrame(_data)
df = jib_data[jib_data['address1'].str.contains(addr)].reset_index(drop=True)
cols = ['item_id', 'sales_type', 'deposit', 'rent', 'size_m2', 'title', 'address']
df = df[cols]
df
```

Out[65]:

	item_id	sales_type	deposit	rent	size_m2		title	address
0	40232488	전세	18900	0	28.69	반려동물 중기청 LH 버팀목 보증금월세 조정OK 1.5룸		마포구 망원동
1	40137773	월세	22800	30	26.68		보증보험가능 투룸	망원동
2	40137788	월세	22800	30	26.68		보증보험가능 투룸	망원동
3	39612133	월세	29600	14	29.67	한강변 2년된 신축 보증보험 가입가능한 매물		마포구 망원동
4	39914755	월세	29600	13	29.67	신축급 강변쪽 2룸 거실 욕1 풀옵션 용사무		마포구 망원동
...
77	40192841	전세	39600	0	47.10	신축급 컨디션의 3룸		마포구 망원동
78	40232795	월세	2000	72	39.67	반려동물 가능한 레어 투룸		마포구 망원동
79	40178585	전세	33000	0	28.14	보증보험 가능한 화이트톤 인테리어 분리형원룸		마포구 망원동
80	40216392	전세	32800	0	29.61	각종 전세대출, 보증보험 O 화이트톤 투룸		마포구 망원동
81	40165513	전세	32500	0	29.95	보증보험가능, 역세권\r\n실평15, 한강도보, 6호선		마포구 망원동

82 rows × 7 columns

```
In [7]: # 함수
        # 패키지
```

```
In [7]: %%writefile zigbang.py
import pandas as pd
import requests
import geohash2
def oneroom(addr):

    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']

    geohash = geohash2.encode(lat, lng, precision=5)

    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']
    ids = [item['itemId'] for item in items]

    url = 'https://apis.zigbang.com/v2/items/list'
    params = {'domain': 'zigbang', 'item_ids': ids[:900]}
```

```

response = requests.post(url, params)
items = response.json()['items']
df = pd.DataFrame(items)
df = df[df['address1'].str.contains(addr)].reset_index(drop=True)
columns = ['item_id', 'address1', 'sales_type', 'deposit', 'rent', 'size_m2', 'title', 'ma
return df[columns]

```

Overwriting zigbang.py

In [8]: %ls zigbang.py

C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호: FCA9-41A6

C:\Users\User\croling 디렉터리

03/19/2024	02:19 PM	1,088	zigbang.py
	1개 파일	1,088	바이트
	0개 디렉터리	158,726,975,488	바이트 남음

In [1]: import zigbang as zb

In [6]: zb.oneroom('도봉동')

Out[6]:

	item_id	address1	sales_type	deposit	rent	size_m2	title	manage_cost
0	40115186	서울시 도봉구 도봉동	월세	2000	70	55.29	SW방학역 2층쓰리룸 17평 베란다 한층한 세대 애완가능	0
1	40117206	서울시 도봉구 도봉동	전세	18000	0	55.29	SW도봉동쓰리룸전세 대출가능 17평넓음	1
2	40177745	서울시 도봉구 도봉동	월세	1000	30	23.14	H 1.5룸구조 저렴한 지상층 H	0
3	40168712	서울시 도봉구 도봉동	월세	1000	70	41.73	반려동물 가능한 리모델링 쓰리룸 주차ok	0
4	40077791	서울시 도봉구 도봉동	월세	100	32	16.53	SW 도봉역4분 엘베있는 풀옵션원룸 다용도실 주차안됨	6
5	40112780	서울시 도봉구 도봉동	월세	100	40	23.14	가성비최고, 깔끔하고 풀옵션, 빨리나가는원룸, 엘베유	5
6	39956830	서울시 도봉구 도봉동	월세	2000	135	46.64	11.합리적가격 대형3룸 즉시입주 보증금조절가능	5
7	40231715	서울시 도봉구 도봉동	월세	2000	145	46.64	첫입주, 매우 깔끔한 쓰리룸 월세, 보증금조정가능, 주차가	5
8	40177077	서울시 도봉구 도봉동	월세	4000	60	46.28	도봉역 도보 이용가능한 3룸	2
9	39829048	서울시 도봉구 도봉동	월세	2300	80	40.25	도봉역 도보2분 도봉산뷰 아파트구조 완벽 풀옵션	7
10	40181841	서울시 도봉구 도봉동	월세	100	30	19.83	도봉역2분거리 저렴한 월세 풀옵션원룸	5

In [5]: # ipython special command # 검색

In [6]: %%time
data1 = 1CPU times: total: 0 ns
Wall time: 0 ns

In []: