

Zigbang 원룸 매물 데이터 수집

In [1]:

```
import requests
import pandas as pd
```

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

In [2]:

```
# 1. 동이름으로 위도 경도 구하기
```

In [3]:

```
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[3]:

```
(37.556785583496094, 126.9013442993164)
```

In [4]:

```
# 2. 위도 경도로 geohash 알아내기
```

In [5]:

```
# install geohash2
# !pip install geohash2
```

In [6]:

```
import geohash2
```

In [7]:

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

Out[7]:

```
'wydjx'
```

In [8]:

```
# 3. geohash로 매물 아이디 가져오기
```

In [9]:

```
url = f"https://apis.zigbang.com/v2/items?deposit_gteq=0&domain=zigbang\
&geohash={geohash}&needHasNoFiltered=true&rent_gteq=0&sales_type_in=전세|월세\
&service_type_eq=원룸"
response = requests.get(url)
data = response.json()["items"]
ids = [item["item_id"] for item in data]
len(ids), ids[:5]
```

Out[9]:

```
(569, [37623131, 37699264, 37849856, 37864041, 37864054])
```

In [10]:

```
# 4. 매물 아이디로 매물 정보 가져오기
```

In [11]:

```
# 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[11]:

```
<Response [200]>
```

In [12]:

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

Out[12]:

	item_id	section_type	images_thumbnail	sales_type	sales_title	d
567	37703323	None	https://ic.zigbang.com/ic/items/37703323/1.jpg	월세	월세	
568	37836529	None	https://ic.zigbang.com/ic/items/37836529/1.jpg	월세	월세	

2 rows × 34 columns

In [13]:

```
# 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[13]:

	item_id	sales_type	deposit	rent	size_m2	floor	building_floor	address1	manage_cost
567	37703323	월세	5000	160	63.97	3	6	서울시 마포구 중동	8
568	37836529	월세	5000	160	61.09	고	6	서울시 마포구 중동	8

In [14]:

```
# 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[14]:

	item_id	sales_type	deposit	rent	size_m2	floor	building_floor	address1	manage_cost
59	37804362	월세	300	72	19.83	2	3	서울시 마포구 망원동	7
60	37811722	월세	2000	110	48.00	1	3	서울시 마포구 망원동	1