

서울특별시 상권분석 (서울특별시 카페 중심)

- 소상공인시장진흥공단에서 제공하는 서울특별시 상가(상권) 정보 데이터를 활용하여 분석
- 필요한 feature들을 추출하여 활용함
- 전처리 이전 데이터 현황을 간단히 파악하기 위한 목적으로 pandas_profiling을 통해 profile_report를 확인

1. Profile report

파악한 내역:

- 일관된 value : 시도명 feature가 하나의 value를 갖고 있음을 파악하고 제거
- High cardinality를 가진 항목 중엔 분석 대상이 있었으므로 유지
- High correlation을 갖고 있는 항목 중 필요한 항목만 추출하여 사용
- 중복된 항목이 있음을 확인 : 추후 drop_duplicate를 통해 제거

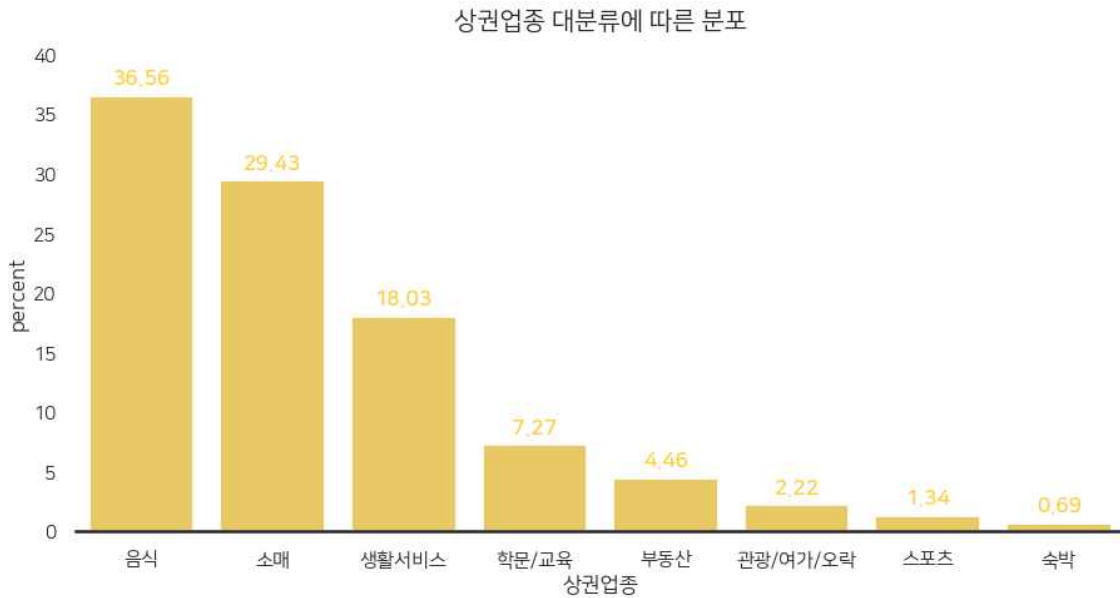
Decision:

- 편의점이 가장 많은 ## 상호명 (CU> GS25> 세븐일레븐 > 카페> 이마트 순)
- 상권업종대분류명 (8종)
- 상권업종소분류명 (89개 항목)
- 상권업종소분류명 (679개 항목)
- 시군명 (25개)
- 시군명에 따른 행정동명, 법정동명
- 대지구분명 (2개 항목) : 대부분이 대지이므로 제거
- 사용한 feature: 상호명, 상권업종대분류명, 상권업종중분류명, 상권업종소분류명, 시군구명, 행정동명, 법정동명, 경도, 위도

2. EDA

- shape: (325880,9)
 - 9개 feature와 325880개 항목에 대한 데이터가 있음을 확인
 - drop_duplicate를 통해 중복행 제거 후 325873개 항목 데이터가 남음
- 결측치 확인
 - 상호명에 1개의 결측치
 - 행정동명에 645개 결측치
 - 시군구명, 행정동명, 법정동명의 high correlation
 - 위도, 경도 데이터가 있으므로 행정동 데이터는 제외하고 사용하기로 결정함
 - 법정동 역시 배제함 (전체적 추이를 파악하기 위함)
- 최종 shape: (325873, 7)

3. 상권업종대분류 분포 확인



서울특별시 상권이 음식, 소매, 생활서비스, 학문/교육, 부동산 순으로 나타나고 있다.

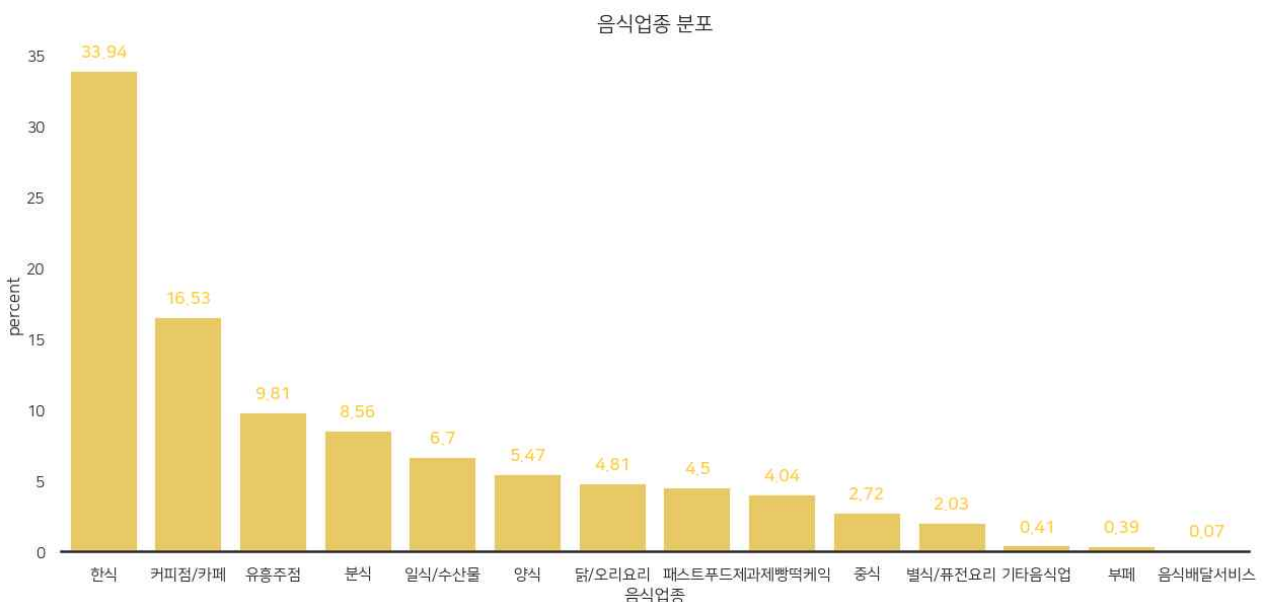
상권 업종 대분류에 따라 데이터프레임을 다시 형성해주었고, 카페 상권분석에 주안점을 두었기 때문에 음식 카테고리의 데이터를 살펴보도록 하겠다. 그 과정에서 상권업종 대분류가 통일되므로 해당 feature를 drop(axis=1)함수를 통해 제거해주었다.

4. [음식 업종]

음식데이터의 shape: (119155,6)

- 119155개 항목의 데이터가 존재하는 것을 확인

음식업종에 대해 분포를 확인한 결과 한식이 33.94%, 그 다음으로 커피점/카페가 16.53%로 많이 차지하는 것을 확인할 수 있다.



5. [서울특별시 카페]

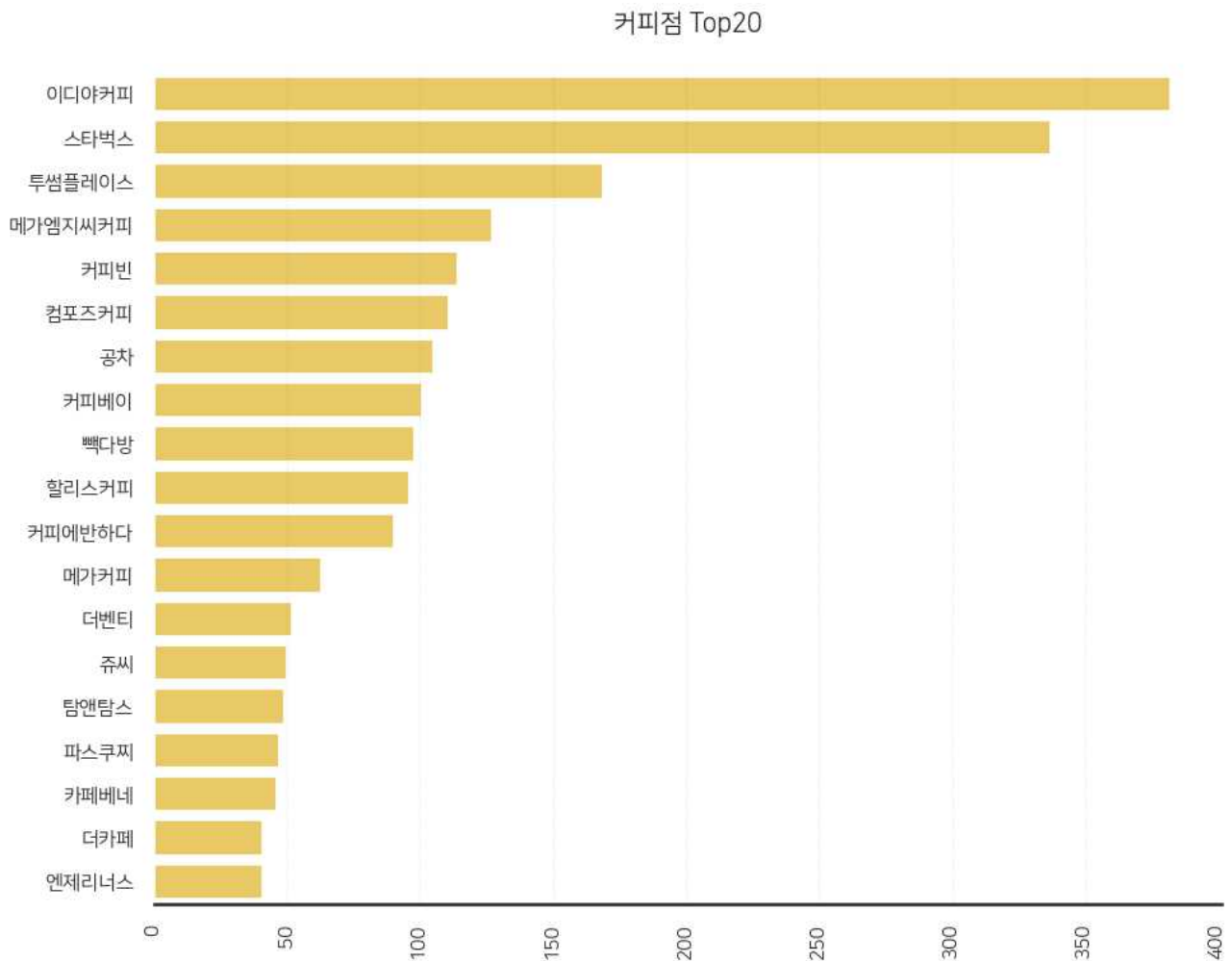
필자는 서울특별시 커피점/카페의 분포에 집중했다.

앞서 구성한 음식 업종 데이터프레임에서 상관업종중분류가 '커피점/카페'인 항목을 추려 cafe_df를 구성했고, shape은 (19702,5)로 19702개의 카페에 대한 데이터가 있음을 확인했다.

- nunique()를 통해 feature별 unique한 카테고리 개수를 확인

- 상호명: 13905개 카테고리
- 상관업종소분류명 : 6개 항목
- 시군구명: 25개 카테고리
- 경도, 위도 : 15546 개 항목

[커피점 Top20 확인]



[분포가 적은 cafe 확인 (unique cafe)]

cafe_df에서 상호명을 value_counts(ascending=True)를 통해 분포가 적은 순으로 배치하여 cafe_value라는 데이터프레임을 구성하였다. 상호명과 개수가 저장된 cafe_value를 unique()를 통해 상호명의 개수를 확인하여 unique cafe의 기준을 설정하였다(3개로 제한).

```
cafe_value.num.unique()
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13,
       14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29,
       30, 31, 32, 34, 35, 36, 37, 38, 41, 46, 47, 49, 50,
       52, 63, 90, 96, 98, 101, 105, 111, 114, 127, 169, 337, 382,
       540])
```

```
# unique 조건 설정
```

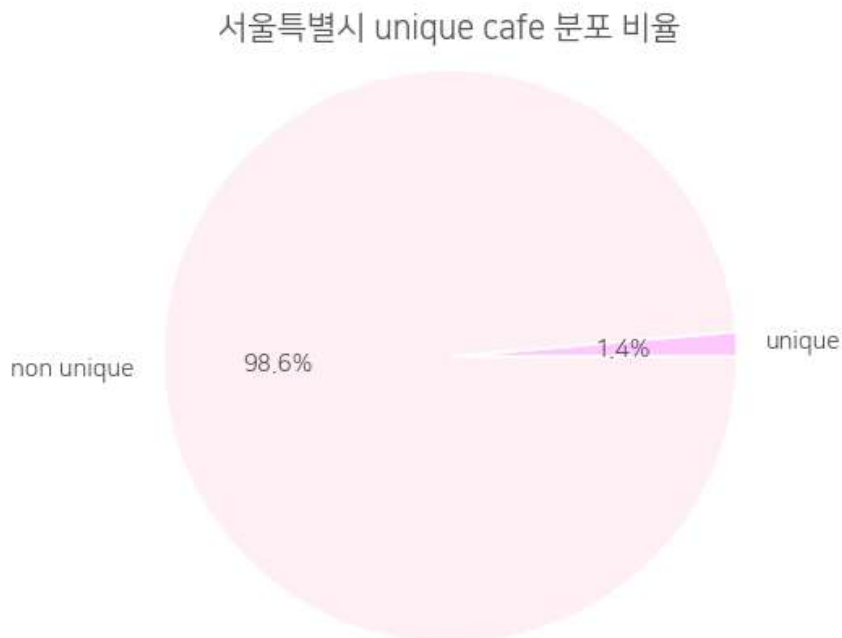
```
unique_condition = (cafe_value['num']==1) | (cafe_value['num']==2) | (cafe_value['num']==3)
```

```
# unique cafe 데이터프레임 형성
```

```
unique_cafe = cafe_value[unique_condition]
```

전체 카페의 길이와 unique_cafe 길이를 사용하여 서울특별시의 카페 중 unique한 카페가 얼마만큼의 비율을 차지하는지 알아보았고, 서울특별시의 unique한 카페는 1.4%를 차지하고 있었다.

- 서울특별시에 3개까지의 분포를 갖고 있는 커피점은 1.4%를 차지



5. [Folium을 통한 cafe 분포별 위치 시각화]

분포에 따라 시각화를 하기 위해 상호명, 위도, 경도만을 사용한 데이터프레임을 형성해주었습니다.

앞서 살펴봤던 cafe top 20에서 우위를 차지하고 있었던 이디야커피, 스타벅스, 투썸플레이스, 메가엠지씨커피, 커피빈에 대해 sub dataframe을 형성해주었고, 해당 데이터프레임에 따라 위도와 경도를 저장해주었습니다.

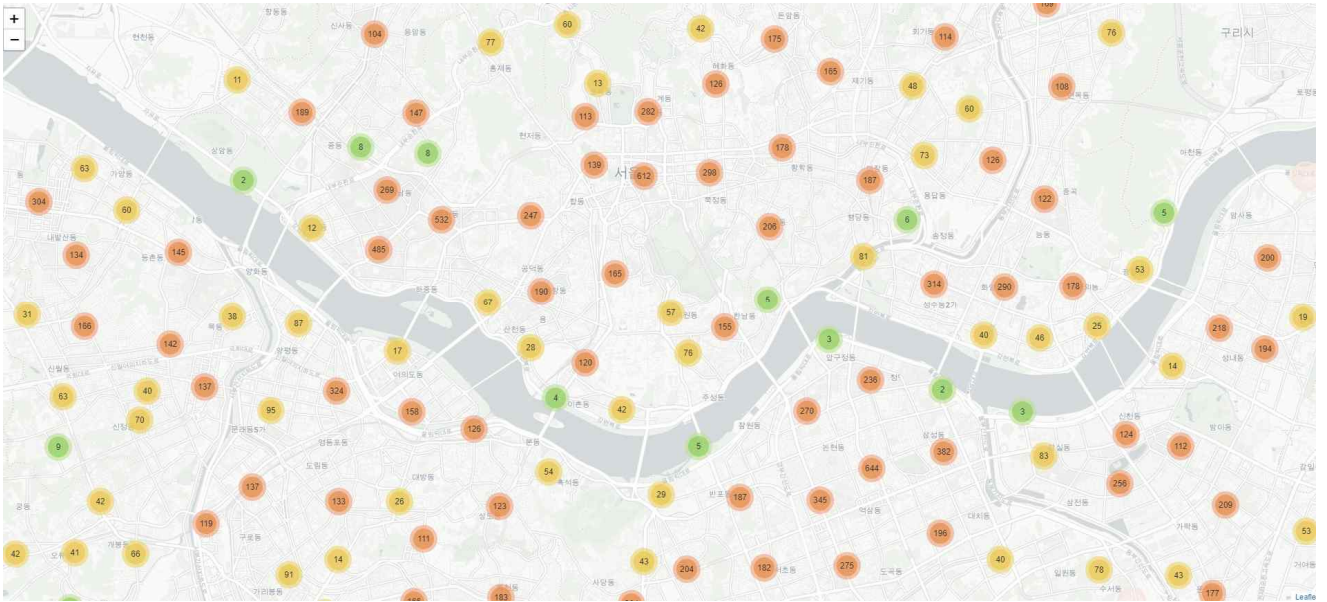
<pre>center = [37.541, 126.986] place = sub_cafe_df[['위도', '경도']] # 카페 별 위도 경도 저장 ediya = ediya_df[['위도', '경도']] star = star_df[['위도', '경도']] twosome = twosome_df[['위도', '경도']] mega = mega_df[['위도', '경도']] bean = bean_df[['위도', '경도']] unique = unique_df[['위도', '경도']] else_cafe = else_df[['위도', '경도']]</pre>	<pre># 중심 저장 # 전체 위도 경도 저장 df # 이디야커피 좌표 # 스타벅스 좌표 # 투썸플레이스 좌표 # 메가엠지씨커피 좌표 # 커피빈 좌표 # unique 카페 좌표 # 기타 카페 좌표</pre>
--	---

지도에 나타낼 중심 좌표를 center로 지정하였고, 전체 위도 경도를 저장한 데이터프레임과 분류할 카페 별로 좌표 정보를 저장하였습니다.

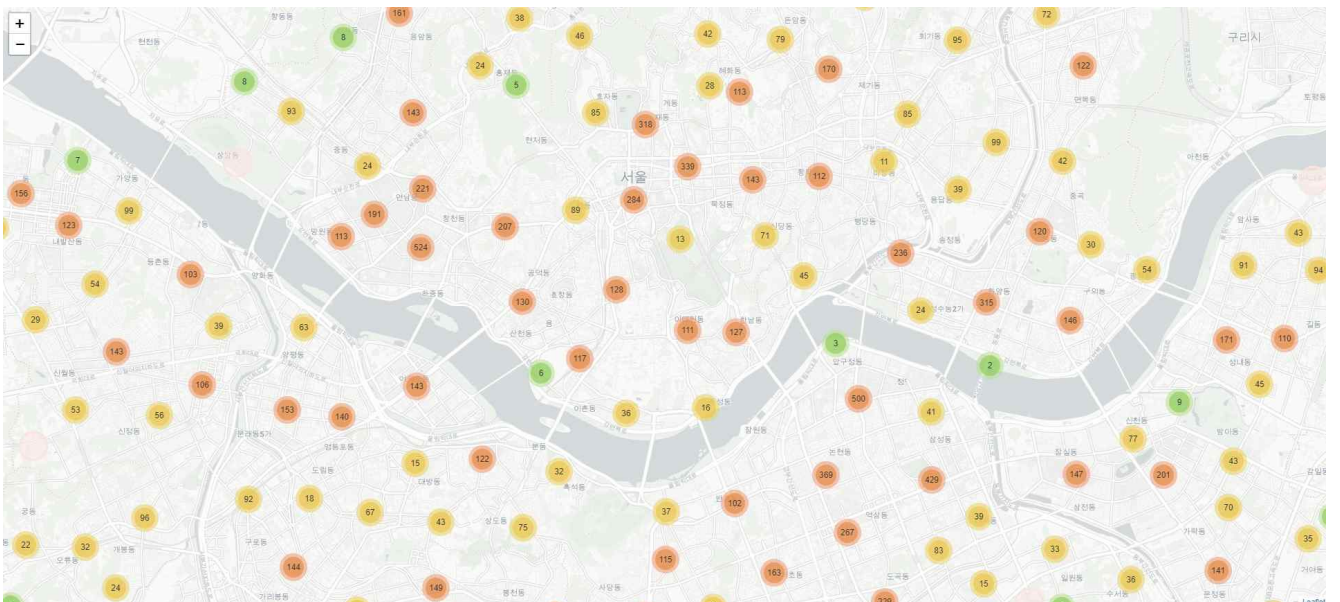
전체 위도를 기반으로 for 문을 통해 주요 카페와 unique cafe, 기타 카페를 나누어 색상화하고자 하였고, for문을 돌렸을 경우 시간복잡도가 높게 나타나 진행이 쉽지 않음을 느끼고 카페에 따라 위치 데이터 프레임을 각각 구성하여 표시하는 방식을 택했습니다.

<pre># 위도와 경도에 따라 지도 객체에 대입 # 이디야커피 for lat, long in zip(ediya['위도'], ediya['경도']): folium.CircleMarker([lat, long], radius=20, color='blue', fill=True, fill_opacity=0.5).add_to(marker_cluster) # 스타벅스커피 for lat, long in zip(star['위도'], star['경도']): folium.CircleMarker([lat, long], radius=20, color='green', fill=True, fill_opacity=0.5).add_to(marker_cluster) # 투썸플레이스 for lat, long in zip(twosome['위도'], twosome['경도']): folium.CircleMarker([lat, long], radius=20, color='red', fill=True, fill_opacity=0.5).add_to(marker_cluster) # 메가엠지씨커피 for lat, long in zip(mega['위도'], mega['경도']): folium.CircleMarker([lat, long], radius=20, color='#FFD228', fill=True, fill_opacity=0.5).add_to(marker_cluster) # 커피빈 for lat, long in zip(bean['위도'], bean['경도']): folium.CircleMarker([lat, long], radius=20, color='#85Bd8e', fill=True, fill_opacity=0.5).add_to(marker_cluster) # unique cafe for lat, long in zip(unique['위도'], unique['경도']): folium.CircleMarker([lat, long], radius=20, color='#ffe4e1', fill=True, fill_opacity=0.5).add_to(marker_cluster) # other cafe for lat, long in zip(else_cafe['위도'], else_cafe['경도']): folium.CircleMarker([lat, long], radius=20, color='black', fill=True, fill_opacity=0.5).add_to(marker_cluster)</pre>	<pre># 이디야커피 blue # 스타벅스커피 green # 투썸플레이스 red # 메가엠지씨커피 yellow # 커피빈 purple # unique cafe pastel pink # other cafe black</pre>
---	--

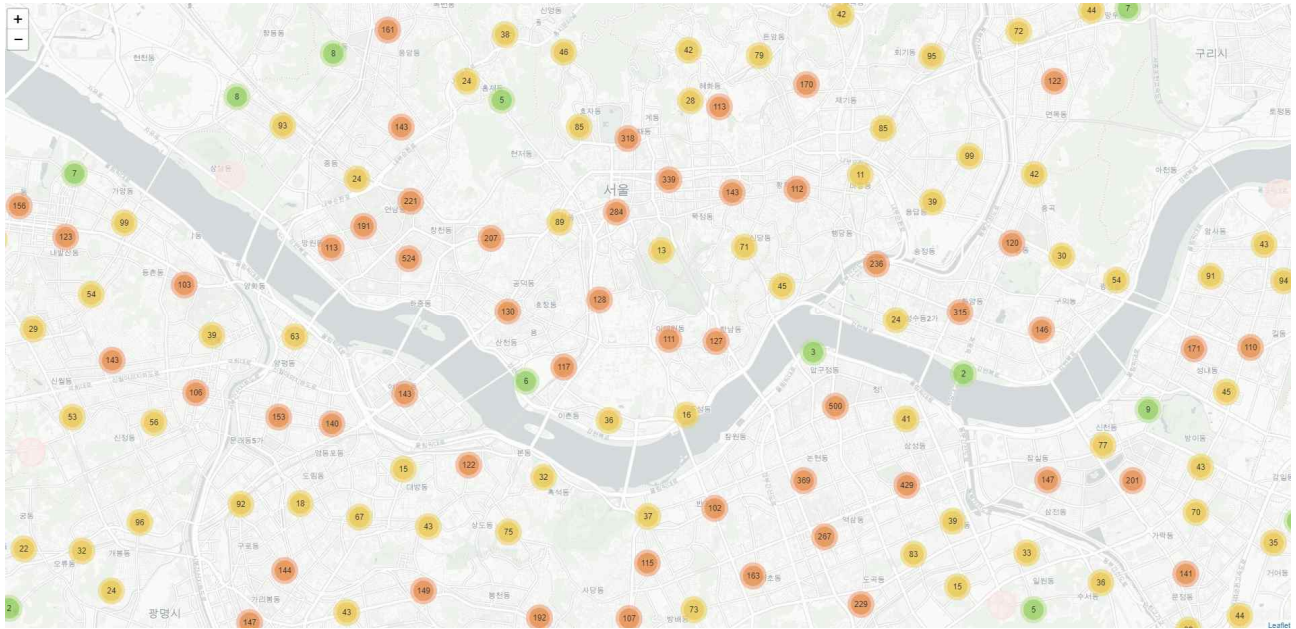
- 전체 cafe 분포



- unique cafe 분포



- twosomeplcace cafe 분포 (대표 예시)



6. 시군구 단위 분석

[시군구별 카페 추이]



시군구별로 카페 추이를 살펴봤을 때 전체적으로 강남구의 카페 분포가 압도적으로 가장 높게 나타난다. 상위 5개로는 강남구, 마포구, 서초구, 송파구, 종로구가 차지하고 있으며, 하위 5개로는 동대문구, 중랑구, 금천구, 도봉구, 강북구로 나타나고 있다.

[시군구 단위 상호명별 카페 추이]

시군구와 상호명으로 묶었을 때 카페의 분포를 확인하고 Top30을 나타낸 것이다.

```
df4 = df4.groupby(['상호명', '시군구명']).size().reset_index(name='count')
non_name_cafe = df4[df4['상호명'] != '카페'] # 상호명이 '카페'인 것 제외
```

