

# [ Data Story ]

음식물 쓰레기가 나오는 양과 코로나 확진자수는 관련이 있을까?  
지역별 음식점 등 주요 상권, 음식물 쓰레기 배출내역, 코로나 확진자 수 및 인구 데이터 분석

음식물쓰레기  
전용수거용기

잠깐!  
노란전용봉투 인가요?  
미사용시  
5만원 과태료 부과됨  
(청소행정과)

잠깐!  
노란전용봉투 인가요?  
미사용시  
5만원 과태료 부과됨  
(청소행정과)

해당 스토리는 데이터 분석에 따른 개인적 견해를 밝혀드립니다.

**“금방 끝날 줄 알았는데”**  
코로나 때문에 다들 집콕한지 벌써 2년.

언택트, 온택트...

미래를 말하는 새로운 마케팅 용어가 난무하지만  
당면해 있는 현실은 아직 희망적이라고 하기 어렵다.

순식간에 택배와 배달이 일상에 굳게 자리잡은 시대  
음식점은 어려움에 처해 있고, “배달” 관련해 많은 사업 모델이 생겨났다.  
소상공인의 위기임은 물론, 쓰레기 발생도 사회 문제로 부각되고 있다.

데이터 분석을 통해 코로나와 쓰레기 발생과의 관계를 규명하고  
그를 통해 이러한 문제를 어떻게 해결할 수 있을지  
전략 도출의 실마리를 찾기 위해 이 시리즈를 기획하였다.

## 사용데이터 소개

### ✓ 소상공인 매출 및 사업자 정보

- 소상공인 및 경기신용보증비대면(모바일)보증 실적 데이터
- <https://www.bigdata-region.kr/#/dataset/266e8af1-d30d-4af0-9c56-cdba6e6c4528>

### ✓ 소상공인 진흥공단 상가(상권) 정보

- 영업중인 전국 상가 업소 데이터(상호명, 업종코드, 업종명, 지번주소, 도로명주소, 경도, 위도 등)
- <https://www.data.go.kr/tcs/dss/selectFileDataDetailView.do?publicDataPk=15083033>

### ✓ 지역별 음식물 쓰레기 배출 추정정보

- 지자체별 음식물 쓰레기 배출내역
- [https://www.bigdata-environment.kr/user/data\\_market/detail.do?id=b5341100-3074-11eb-bc79-3b11eb915d6d](https://www.bigdata-environment.kr/user/data_market/detail.do?id=b5341100-3074-11eb-bc79-3b11eb915d6d)

### ✓ 서울시 코로나 발생 동향

- 서울시 공식 페이지 제공 정보 웹크롤링 (실시간 갱신)
- <https://www.seoul.go.kr/coronaV/coronaStatus.do>

### ✓ 서울시 주민등록인구 구별 통계

- 서울시 열린데이터 광장에서 제공하는 자치구별 주민등록인구수 일반 보고통계
- <https://data.seoul.go.kr/dataList/419/S/2/datasetView.do>

“어느 곳에 어떤 업종이 장사가 잘 될까?”

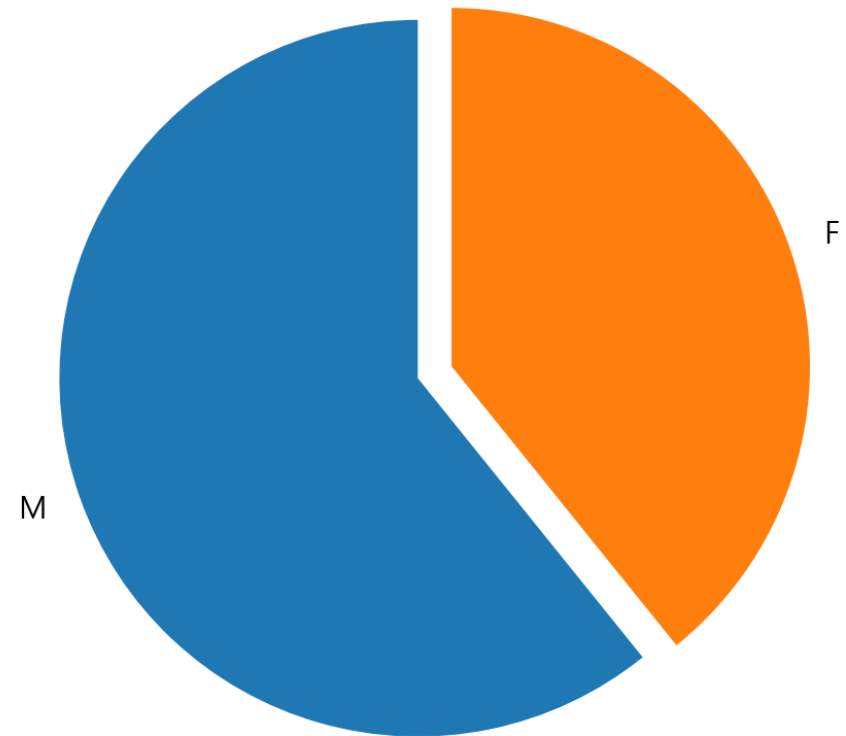
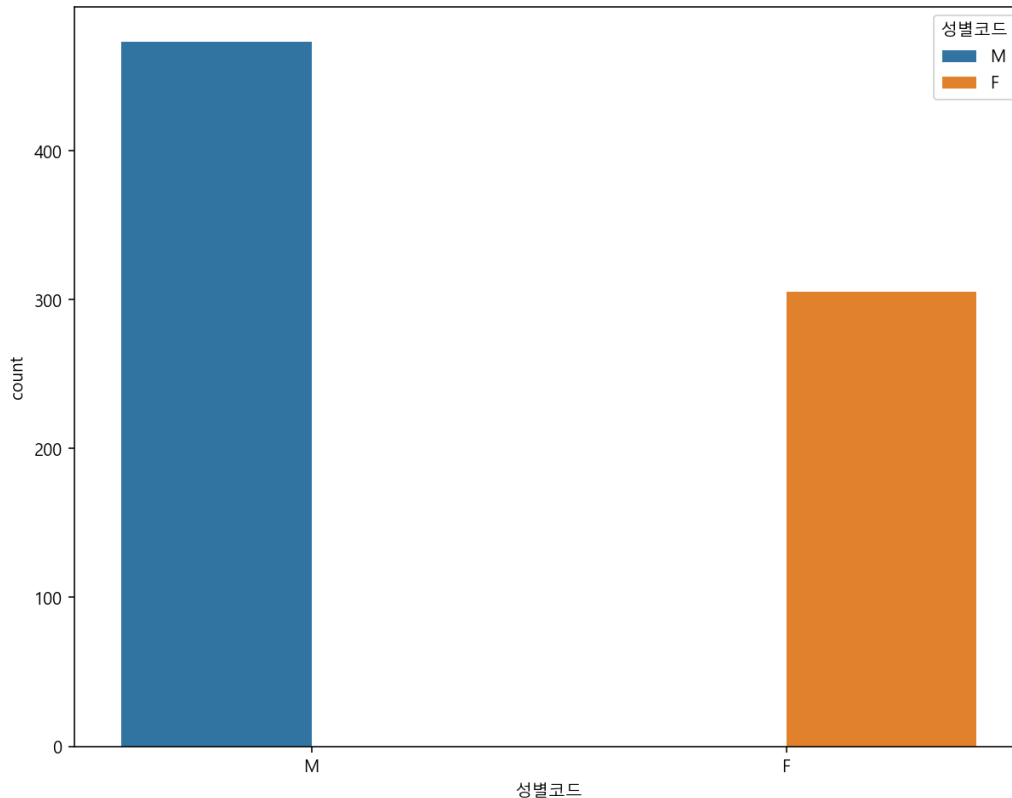
# Story #01

소상공인 매출 현황 및 지역 상권 정보 분석

## 경기도 지역 소상공인 성별 분포

경기신용보증\* 비대면(모바일) 보증 실적 데이터를 분석해본 결과, 경기도에서 비대면으로 보증 신청을 한 소상공인 중 남자는 473명, 여자는 305명으로 남자의 보증 신청이 여자보다 55% 더 많다. 데이터를 좀더 확보하면 “성별 예측”을 하기 적합하다.

경기도 비대면 보증신청 소상공인 성별 비율

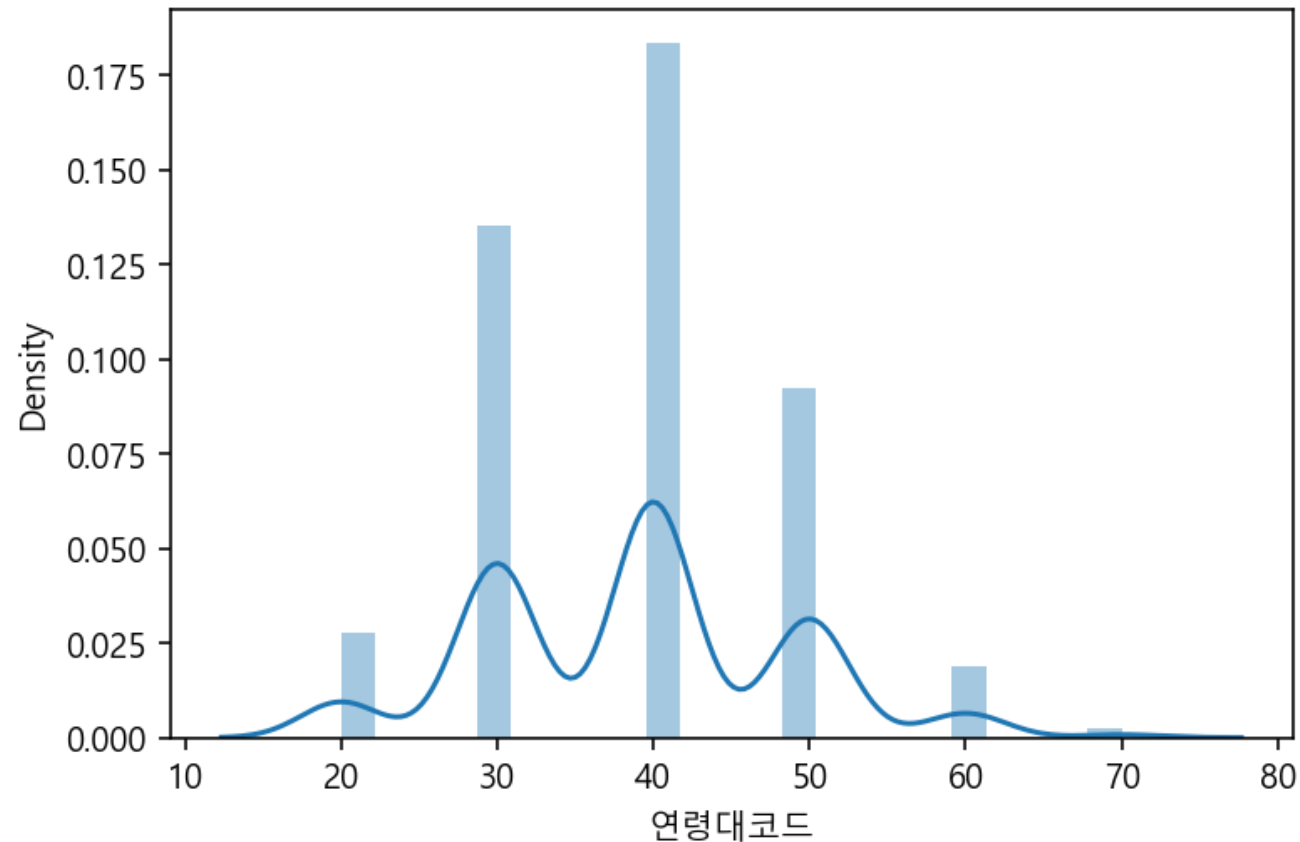


\* 소상공인 및 중소기업 등이 금융기관으로부터 사업자금을 대출받고자 할 때 신용보증기관이 신용을 담보로 금융기관에 보증을 서주는 제도

## 경기도 지역 소상공인 연령대 분포

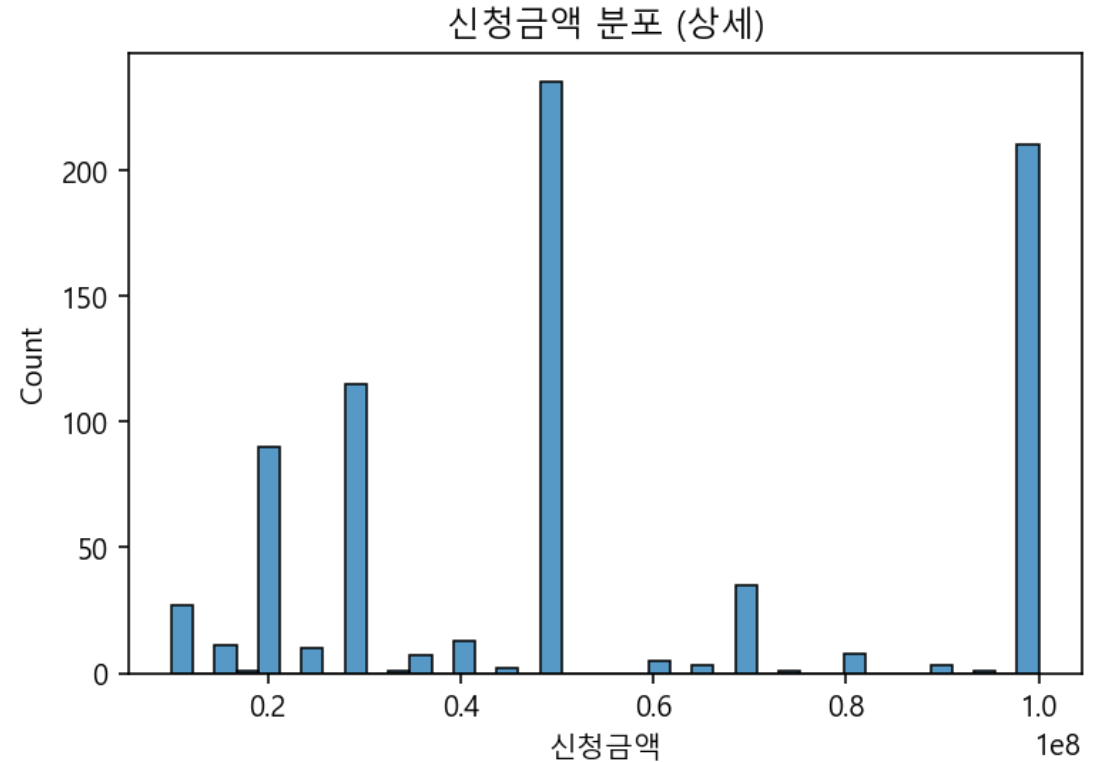
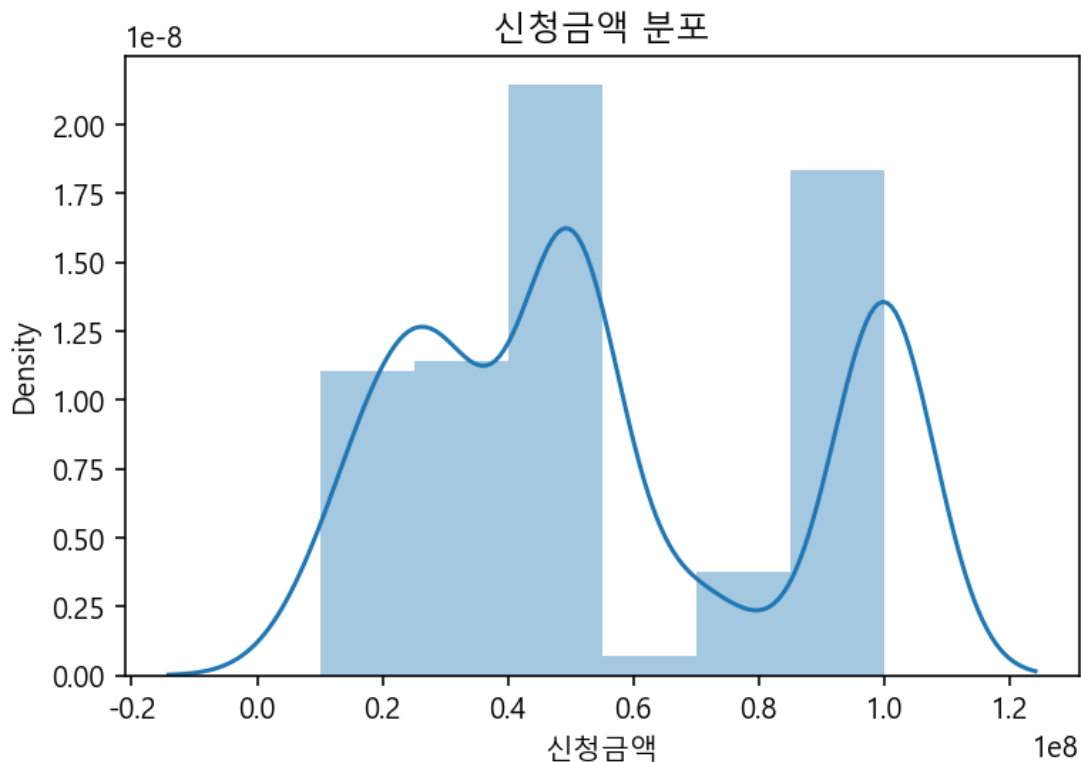
40대가 310명으로 가장 보증 신청을 많이 했고, 30대와 50대 순이다.  
20대의 나이에 벌써 자신의 사업으로 소상공인 활동을 시작한 부분도 눈에 띈다.  
**70대 이후 고령자의 경우, 경제활동인구가 적은 동시에  
모바일 신청에도 익숙하지 않아 신청 인원도 확연히 적은 것이라 판단된다.**

연령대코드 분포



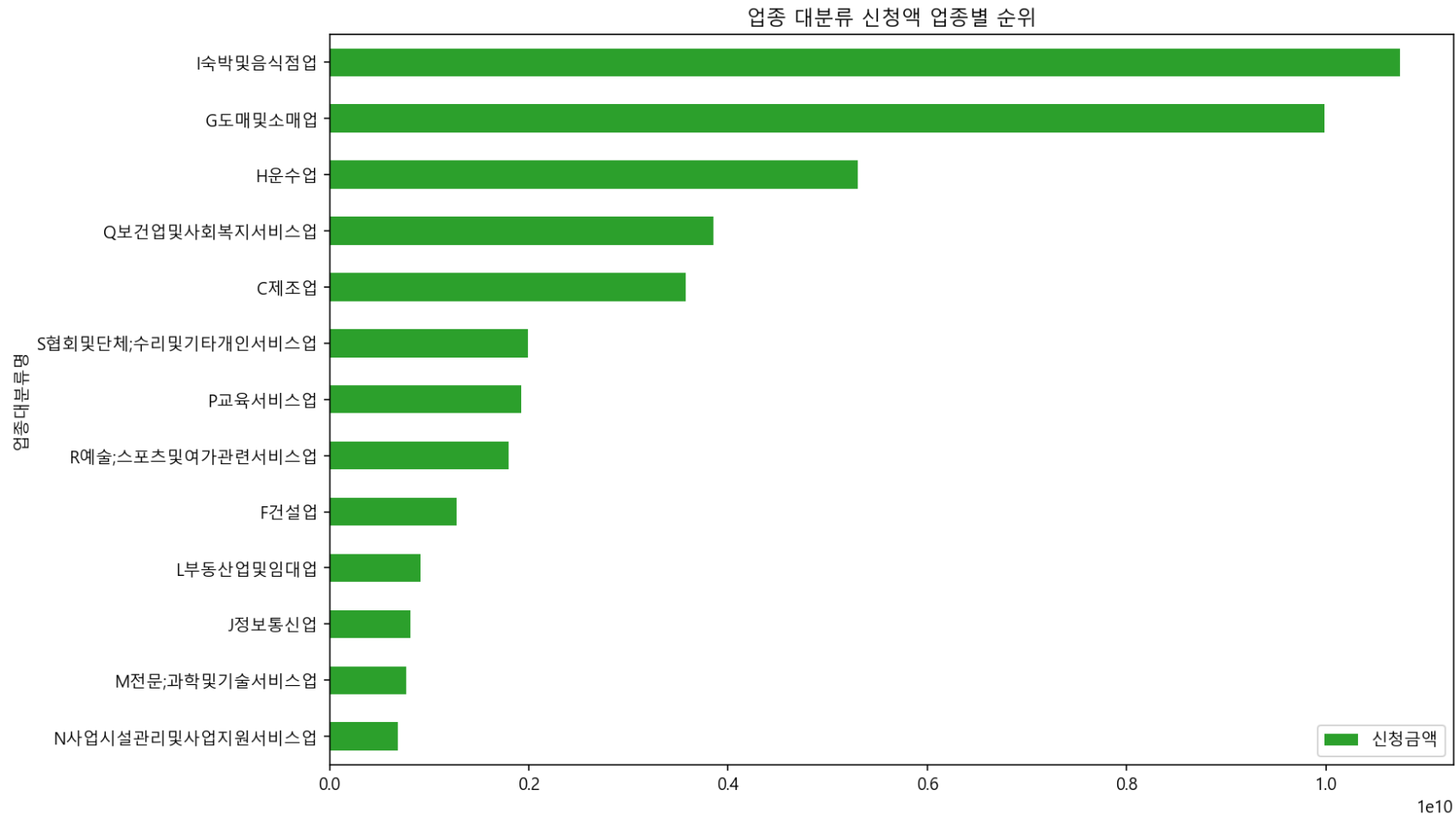
## 보증 신청금액 분포

〈보증 신청금액〉 히스토그램을 살펴보면 5천만원과 1억원의 신청이 가장 많다.  
5천만원 이하 신청액은 일부 있으나, 5천만원과 1억원 사이의 보증 신청은 거의 없다.  
〈5천만원〉 이상 신청에는 추가 절차가 필요해 보증 신청을 부담스러워 하는 것으로 보인다.  
〈1억원〉이 한도이나, 소상공인들의 어려움을 감안해 보증액 증대를 고려할 필요가 있다.



## 〈업종대분류〉 보증 신청액 순위

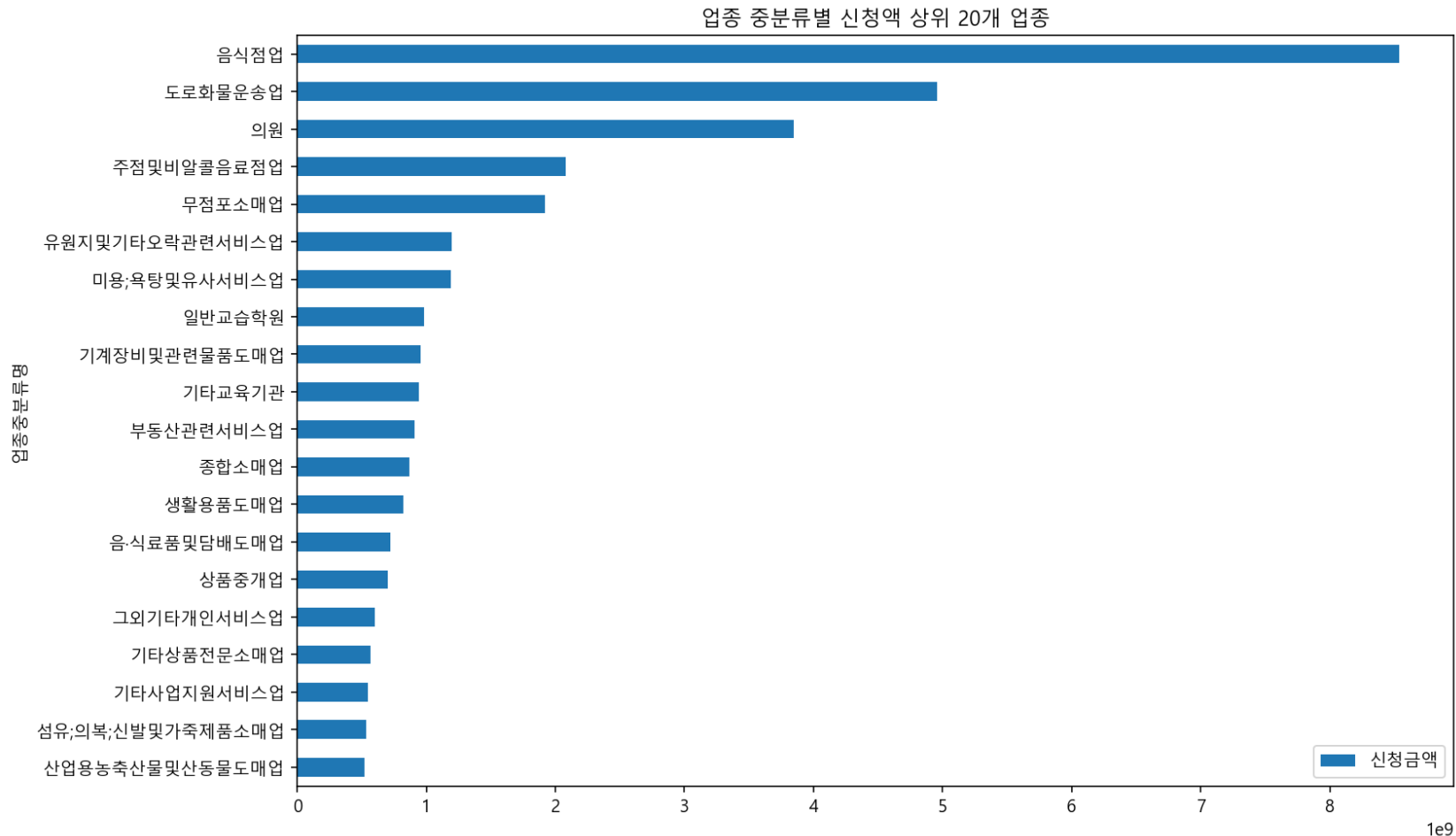
〈숙박 및 음식점업〉, 〈도매 및 소매업〉, 〈운수업〉 순으로 보증 신청액이 많고  
〈정보통신업〉, 〈과학기술 서비스업〉 등의 보증 신청액이 적은 것을 통해  
〈음식점〉이 어려움을 겪는 반면, 〈정보/과학〉업종은 여건이 비교적 나은 편임을 알 수 있다.





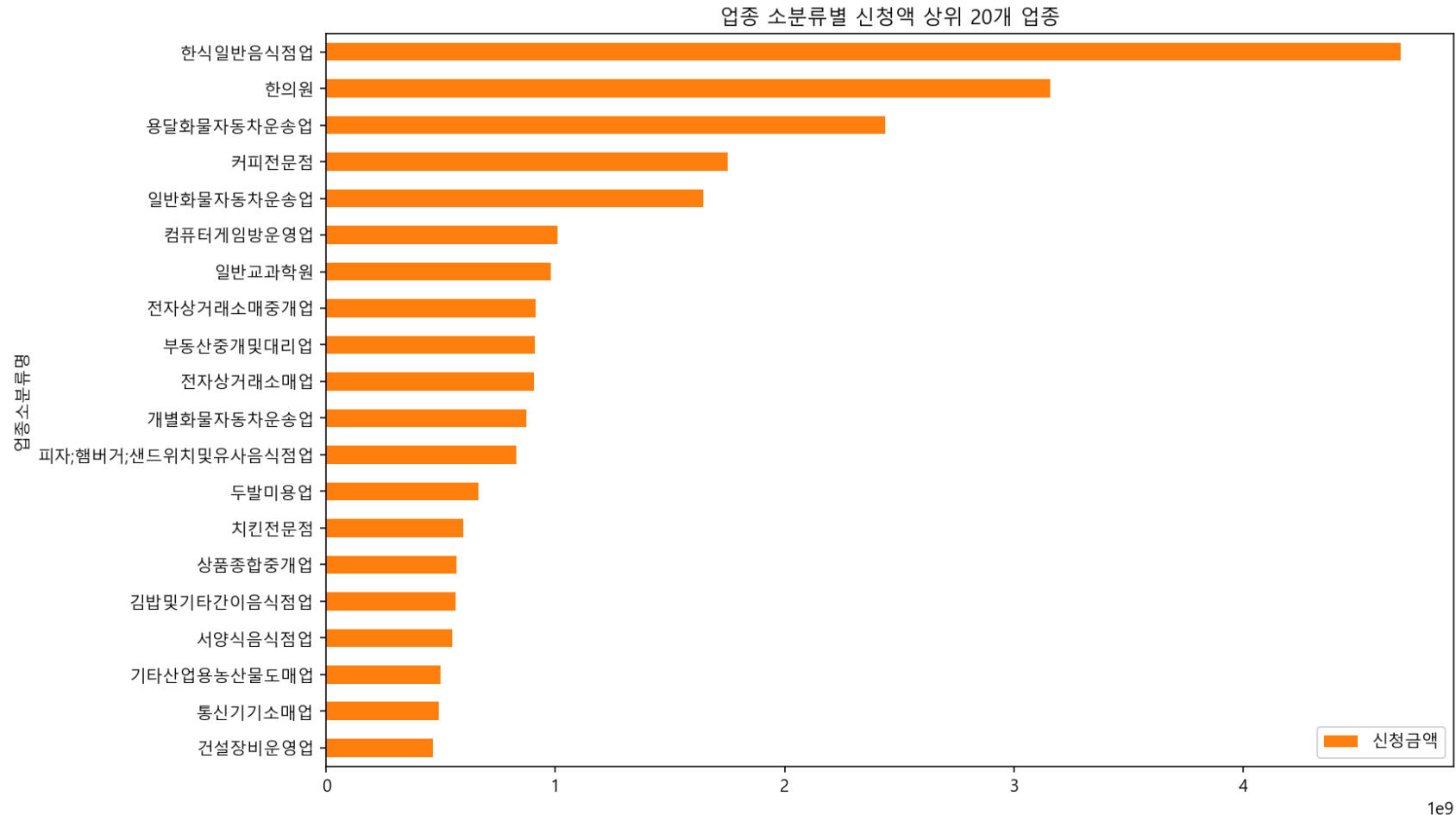
## 〈업종 중분류〉 보증 신청액 순위

〈음식점업〉의 보증 신청액이 가장 많아 해당 업종의 어려움을 알 수 있다.  
〈교육업〉은 안정적이나, 언택트로 온라인 교육이 대세인 현재는 다를 것이다 (코로나이전데이터).  
〈의원〉의 보증 신청이 많은 것을 통해 요즘은 의사도 이전 같지 않다는 것을 알 수 있다.



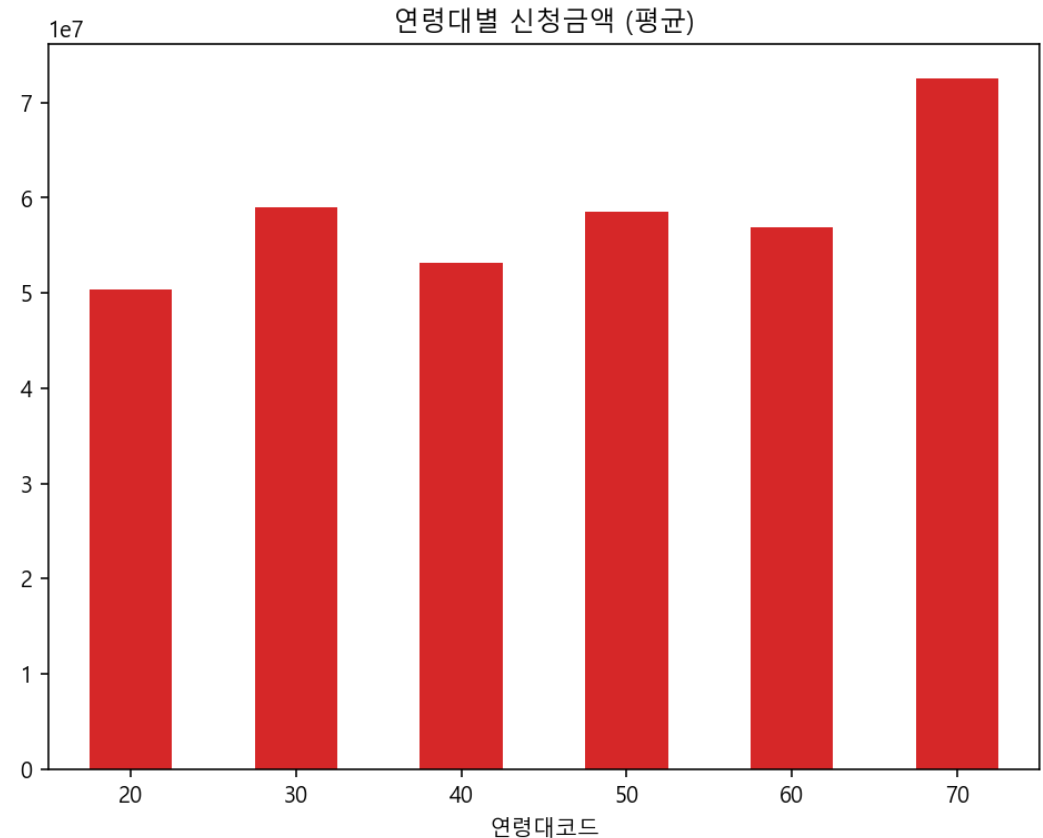
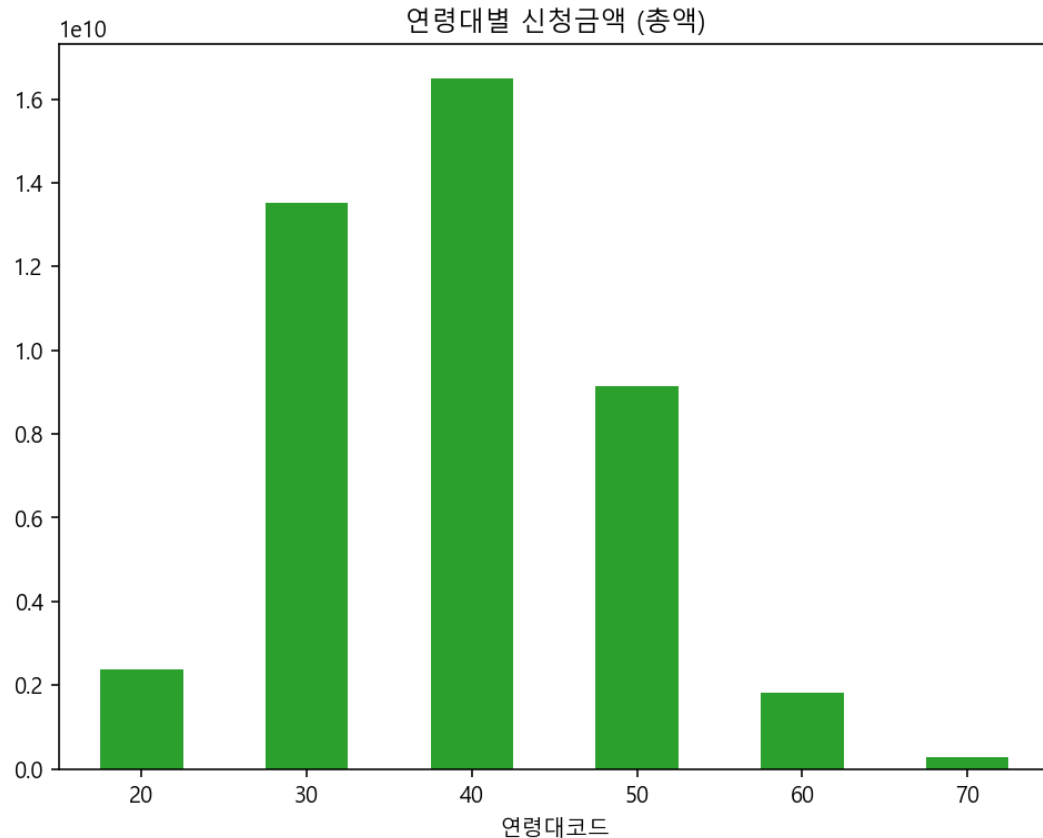
## <업종 소분류> 보증 신청액 순위

<한의원> 보증 신청액이 2위인 점이 의외였다.  
<일반음식점업>에 비해 <치킨 전문점>은 상대적으로 신청액이 작았다. 역시 치맥이다.



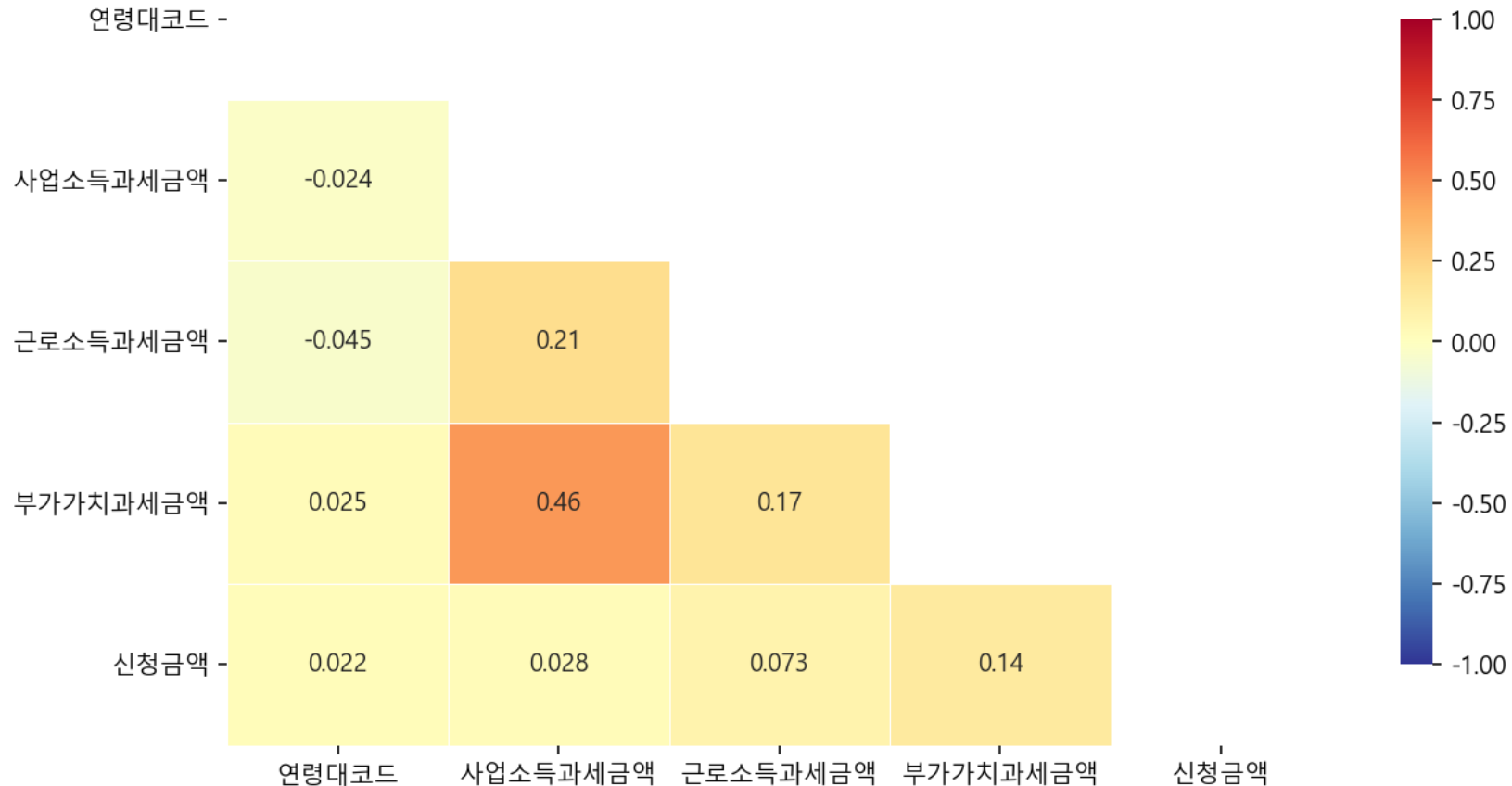
## 연령대별 신청금액 <총액> vs. <평균액> 비교

연령대별 신청금액의 총액은 40대가 가장 많으나, 신청금액 평균은 70대가 가장 많다.  
<모바일 보증 신청> 절차가 어려워 신청 자체는 많이 이루어지지 않지만  
고연령자가 일단 신청하게 될 경우엔 많은 액수를 신청하는 것이다.  
특히 70대 소상공인에게 모바일 보증 신청을 쉽게 만들고, 활용성을 알릴 필요가 있다.



## 〈연령〉과 〈과세금액〉, 〈보증신청금액〉 등과의 상관관계 분석

〈사업소득과세금액〉과 〈근로소득과세금액〉은 함께 증가하는 패턴을 보인다.  
자영업자는 부가가치세에 더해, 사업소득에 대한 소득세도 납부해야 하기 때문이다.  
소상공인에게 다양한 세금 혜택을 줄 수 있는 길이 열려야 한다.



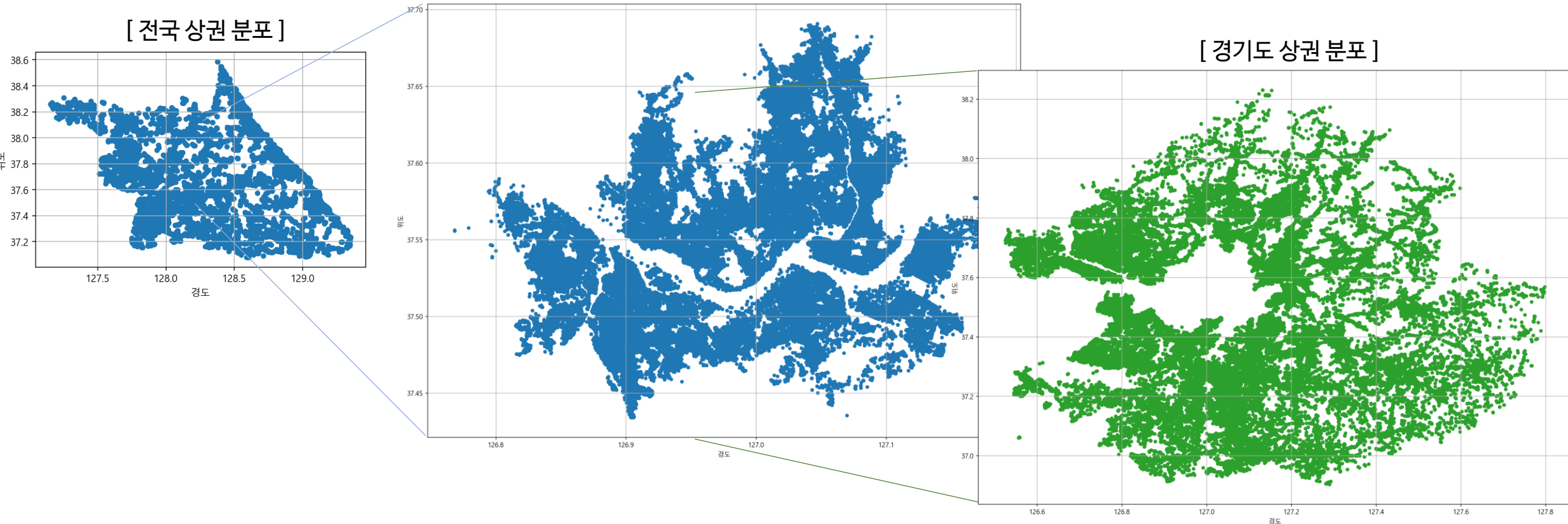
## <전국상권> 데이터에서 <서울>, <경기도> 상권 추출

지금까지 <소상공인>이라는 사람, 즉 “상인”을 분석하였다.  
이제 소상공인이 “활동”하고 있는 지역, 즉 “상권”을 분석해보기로 한다.  
원본인 <전국상권> 데이터에서 <서울>과 <경기도> 상권을 추출하였다.

[ 서울시 상권 분포 ]

[ 경기도 상권 분포 ]

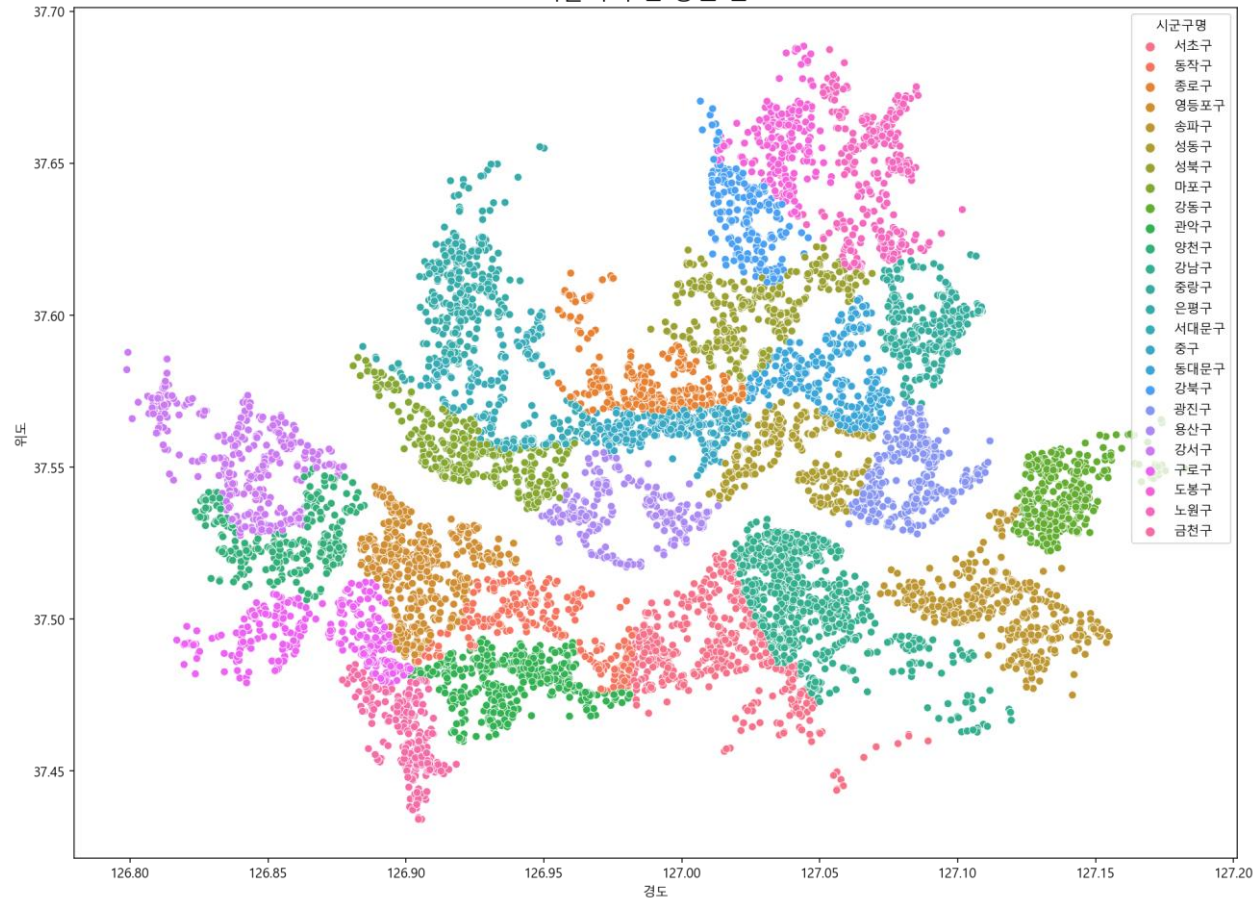
[ 전국 상권 분포 ]



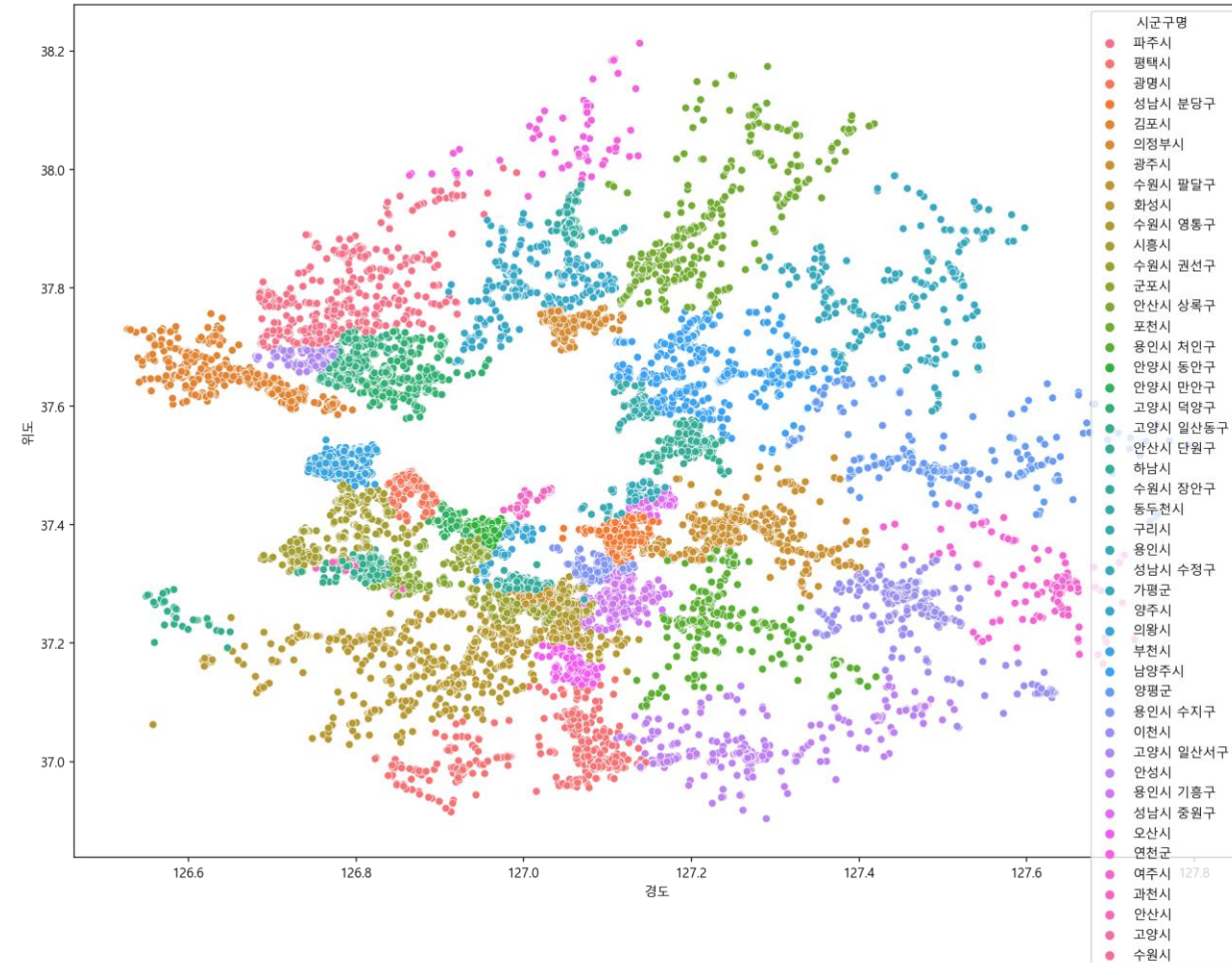
# <서울>, <경기도> 세부 지역별 상권 분포 분석

세분화된 분석을 위해 전국에서 <서울>과 <경기도> 지역만 추출하였다.  
 업종/업태별, 구별, 시군별로 다양하게 상권을 분석해보겠다.

서울시 구별 상권 분포



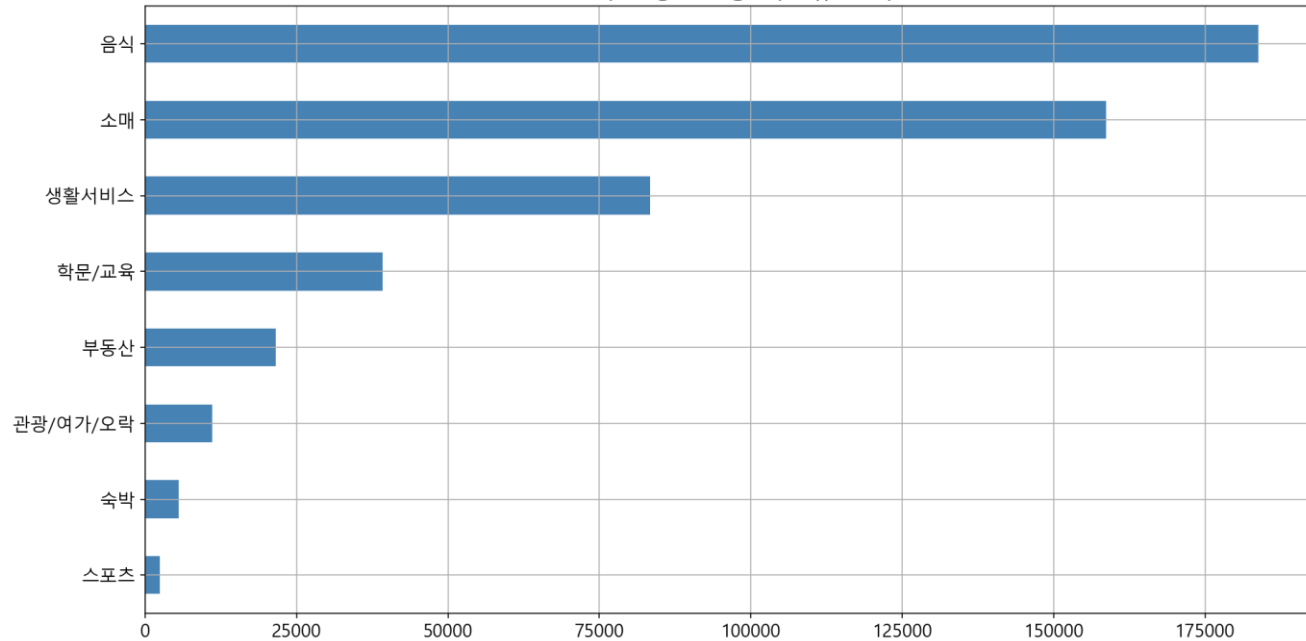
경기도 시군구별 상권 분포



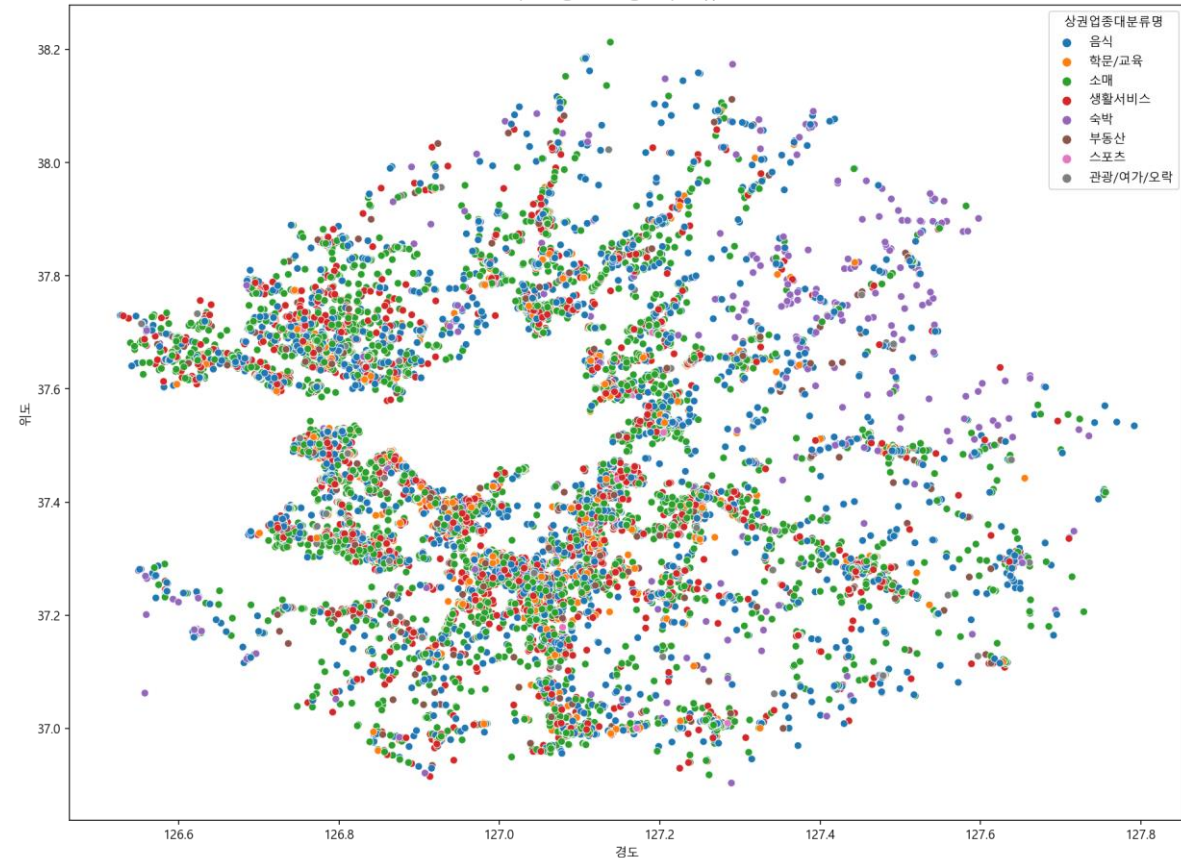
## <경기도> 상권업종대분류 순위

경기도에서 가장 많은 업종은 <음식>이며, 이것은 앞의 “소상공인 분석”에서 확인한 결과와 같다. <숙박> 업종이 다른 업종에 비해 상대적으로 적었으며, <스포츠> 상권은 거의 없음을 알 수 있었다.

경기도 상권업종 대분류 순위



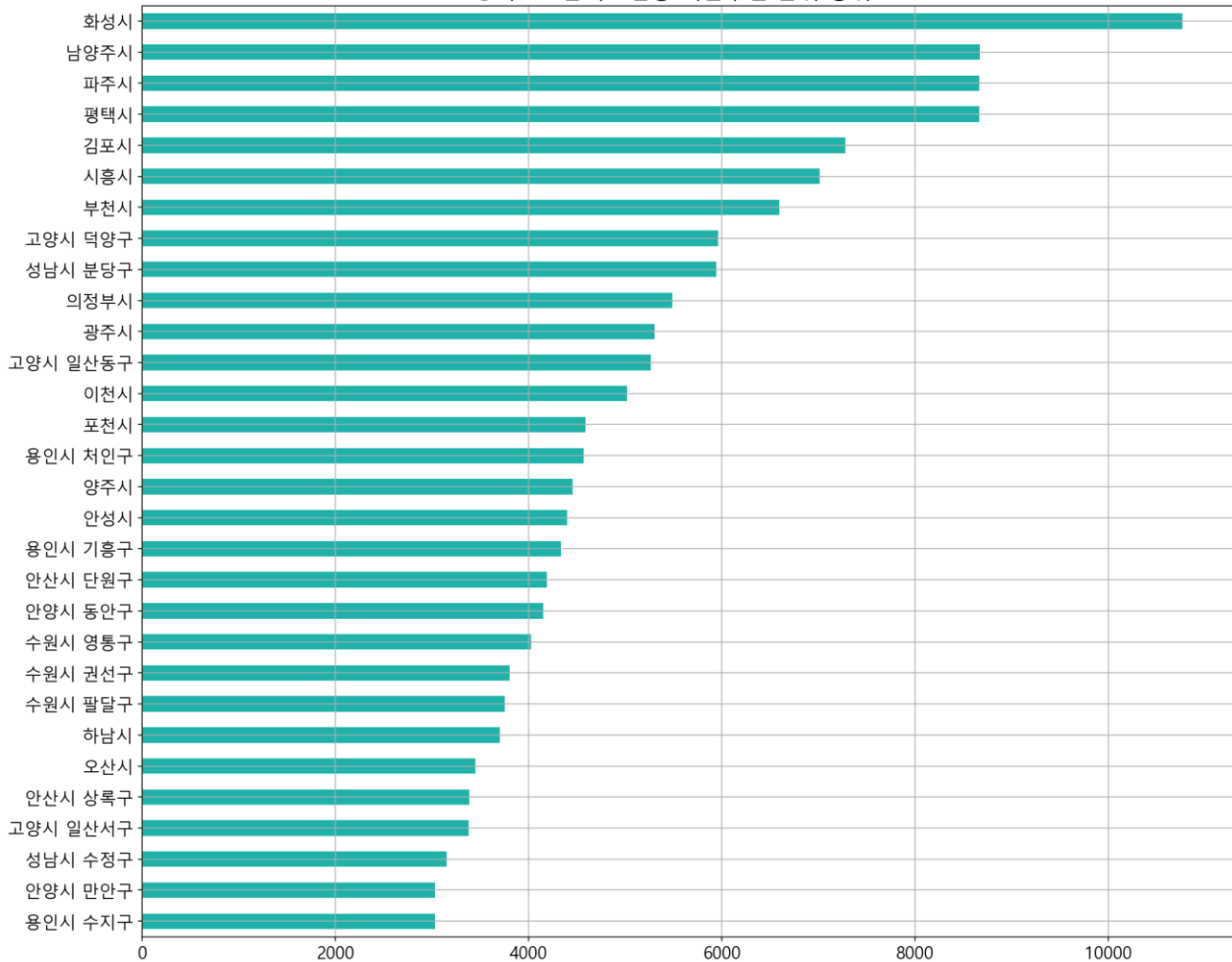
경기도 상권업종 대분류 분포



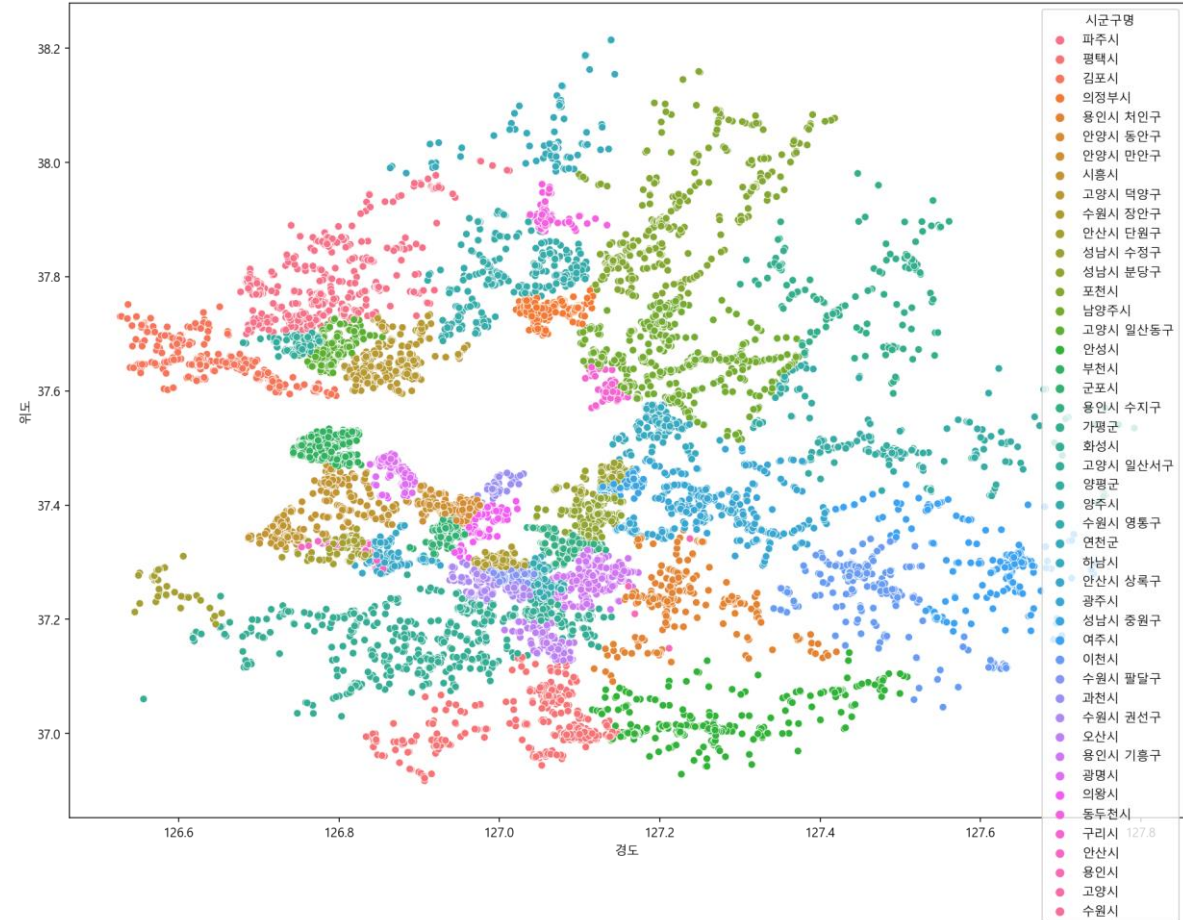
# 경기도 <음식> 업종 상위 30개 시군구

경기도에서 <음식점>이 가장 많은 곳은 <화성시>, 다음은 <남양주시>, <파주시> 순이다.  
 산업시설이 많은 화성시로 출퇴근하는 인구때문에 음식점이 화성시에 가장 많은 것으로 추측해 보았다.

경기도 <음식> 업종 시군구별 순위 상위 30



경기도 <음식> 상권 시군구 분포

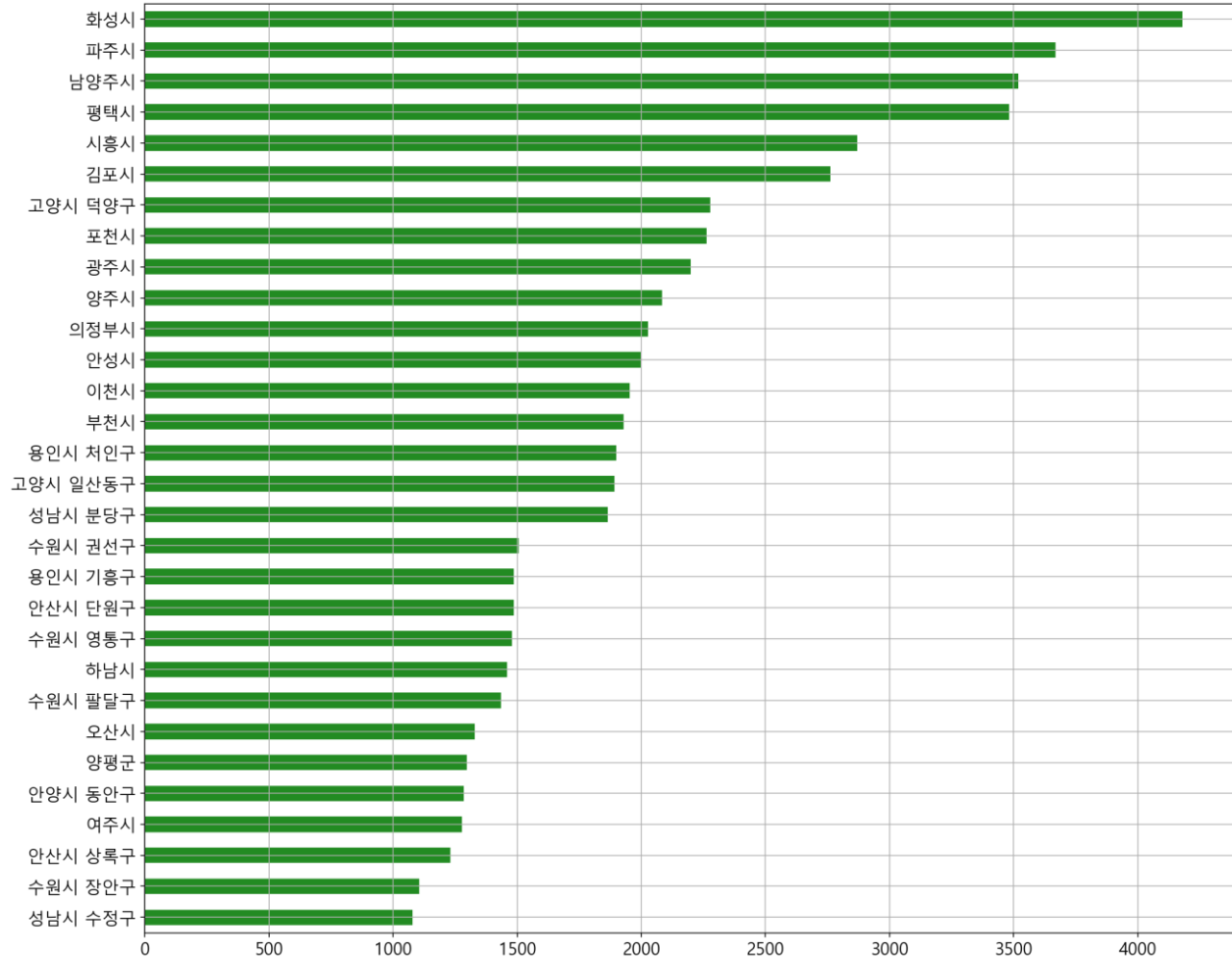




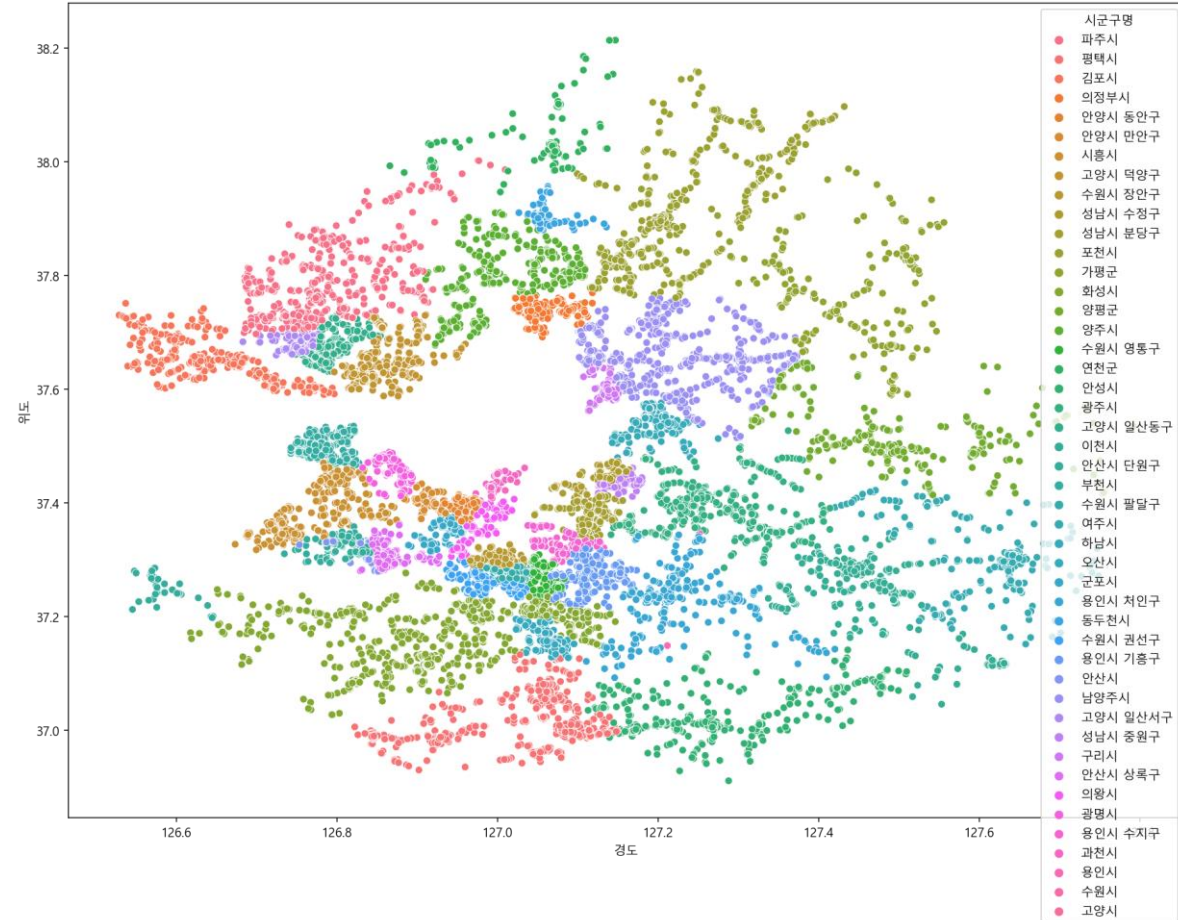
# 경기도 음식업종 중 <한식> 상위 30개 시군구

음식업종 중 <한식>이 가장 많은 지역은 <화성시>, <파주시>, <남양주시> 순이다.  
 전체 음식점은 남양주가 파주보다 많으나, 음식점 중 <한식> 업종은 파주가 더 많다.

경기도 <한식> 업종 시군구별 순위 상위 30

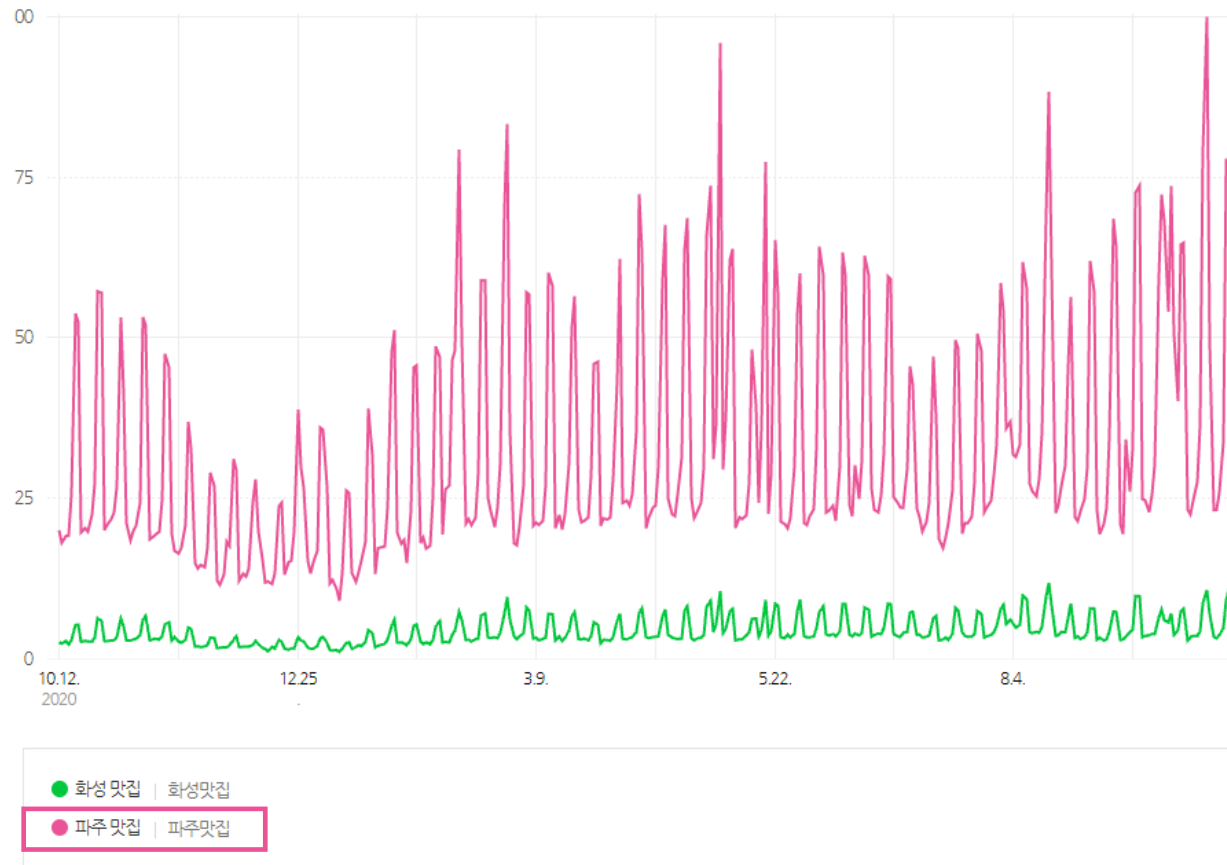


경기도 <한식> 상권 구별 분포



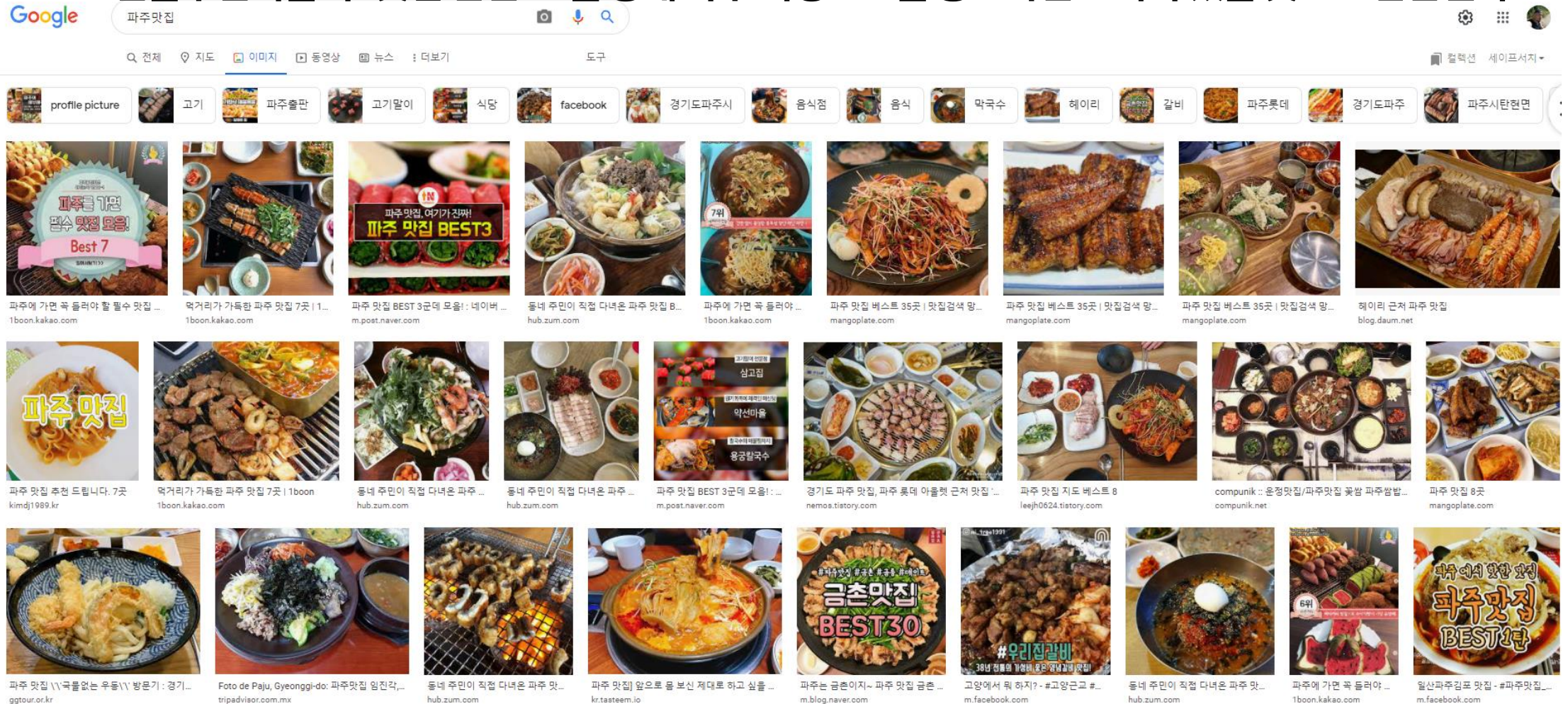
## 〈화성맛집〉과 〈파주맛집〉에 대한 관심도 비교 분석 (네이버 검색 트렌드)

음식점과 한식당의 “개수”는 화성시가 파주보다 많지만,  
〈지역의맛집〉에 대한 “관심”은 파주시가 화성시보다 훨씬 많다.  
맛집 방문을 목적으로 파주시를 찾는 사람이 많다는 사실을 알 수 있다.



# <파주맛집> 관련 콘텐츠 (구글 이미지 검색)

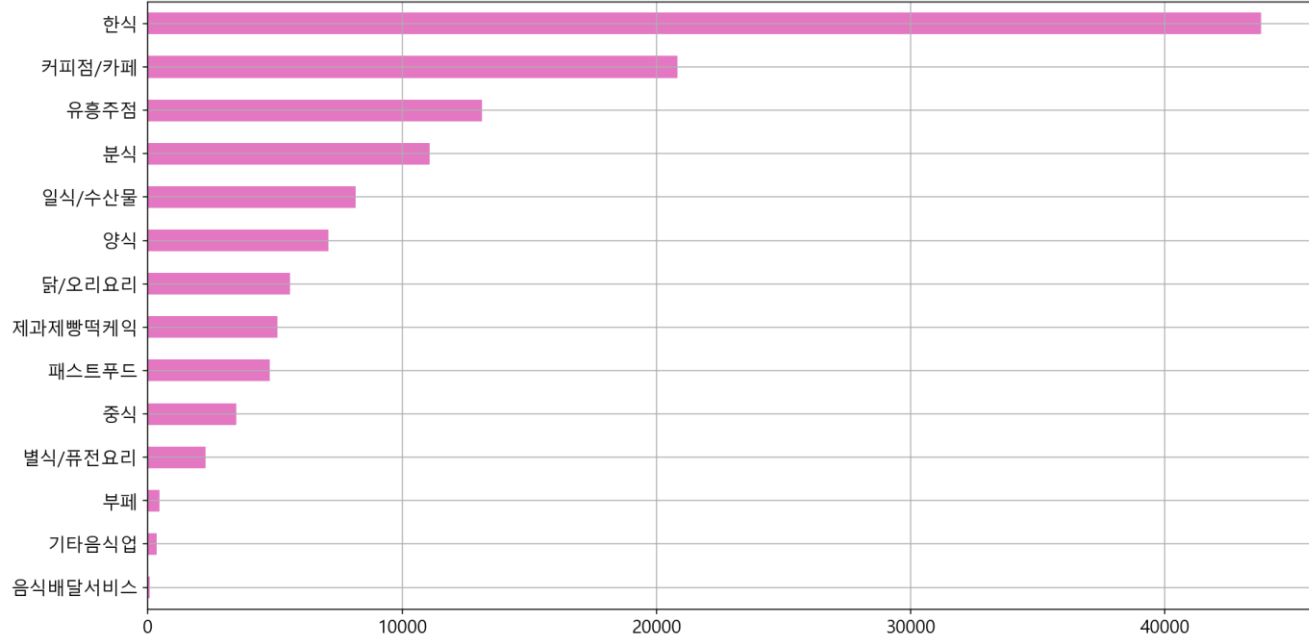
망고플레이트, 페이스북, 카카오, 네이버포스트 등 다양한 채널에  
 “동네 주민이 직접 다녀온 파주 맛집”, “파주는 금촌이지” 등 자발적 맛집 콘텐츠가 생산되고 있다.  
 SNS 인플루언서들의 “맛집 콘텐츠”를 통해 파주 여행코스를 홍보하면 효과가 있을 것으로 판단된다.



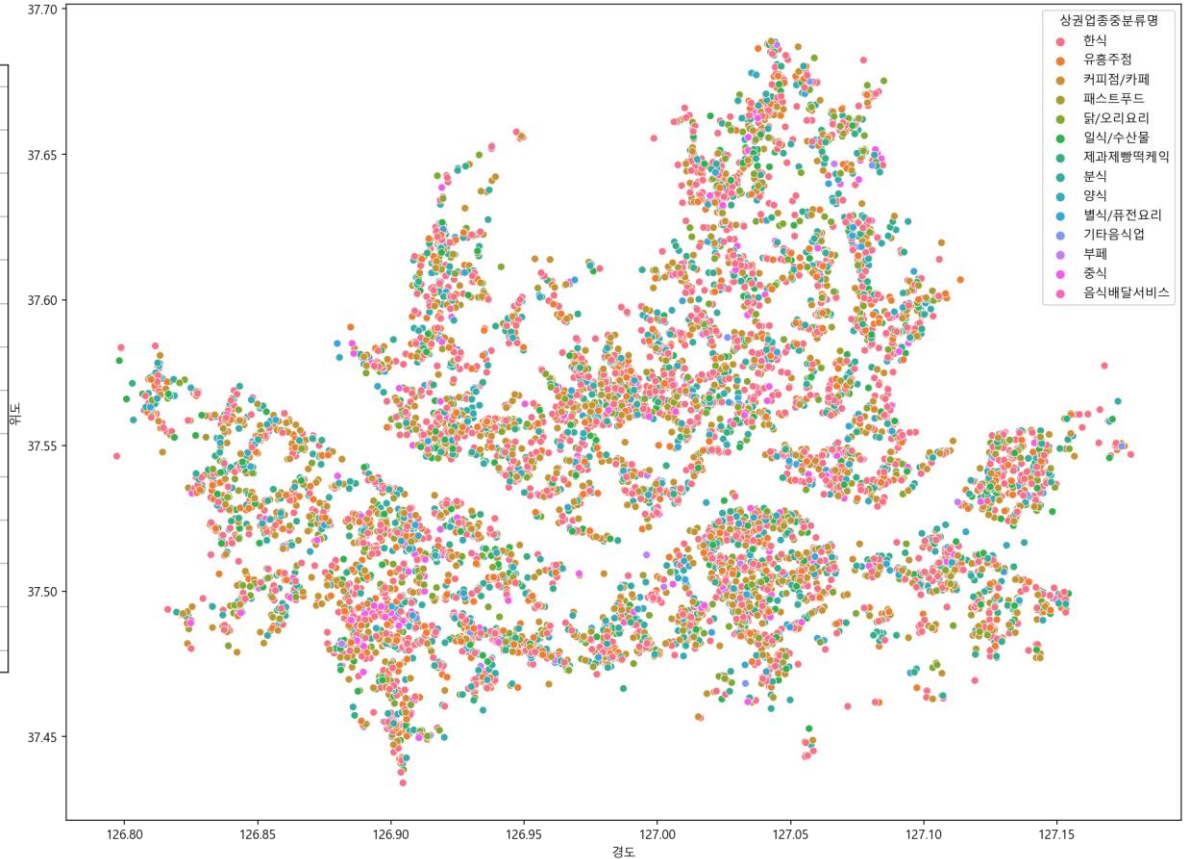
# 서울 음식 업종 중분류 순위

서울의 음식 업종 중 가장 많은 것은 <한식>이며 그다음이 <카페>, <주점> 순이다.  
 <음식배달서비스>는 거의 없는데, 코로나 발생 이전의 데이터이기 때문으로 생각된다.

서울시 <음식> 업종 중분류 순위

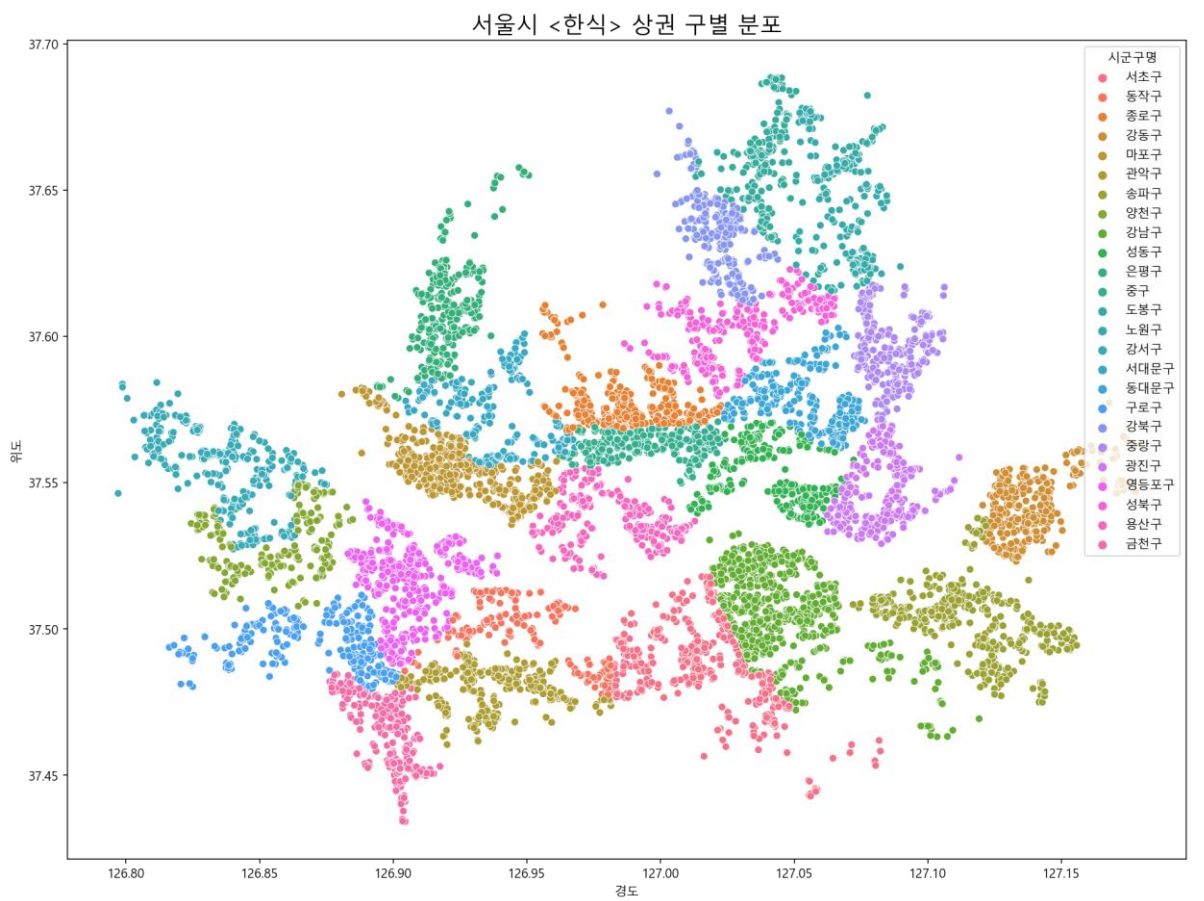
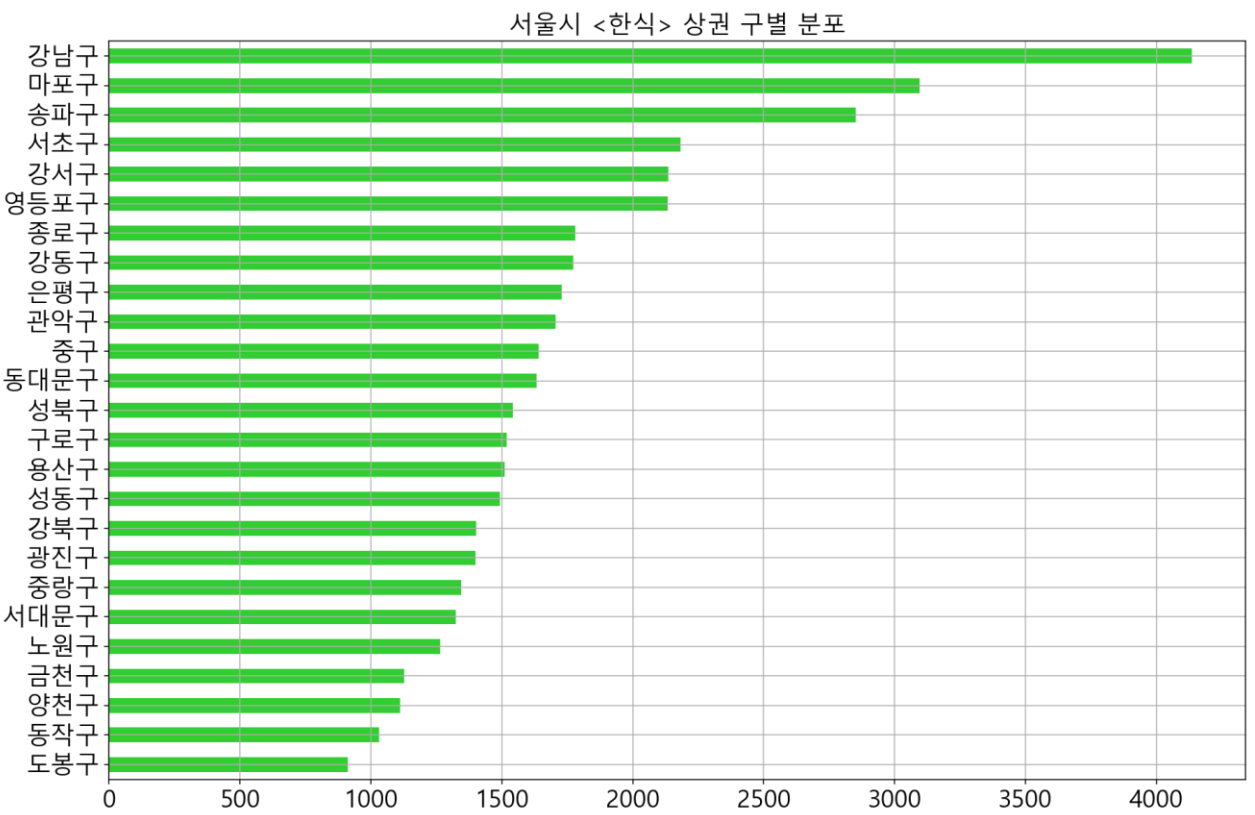


서울시 <음식> 업종 중분류 분포



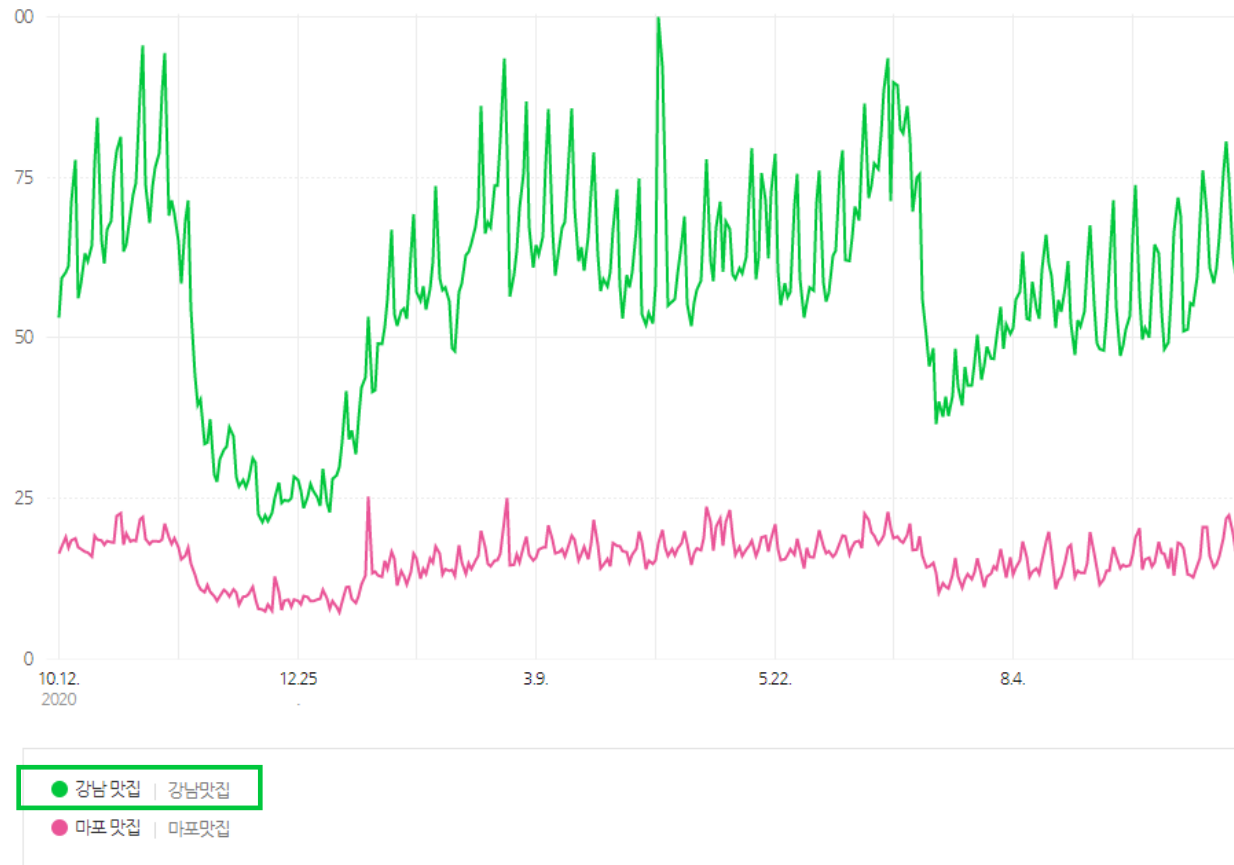
# 서울 음식 중 <한식> 업종 구별 순위

음식 업종 중 <한식>이 가장 많은 구는 <강남구>, <마포구>, <송파구> 순이다.  
 거주 인구가 많거나, 회사가 많은 지역에 한식 업종이 많이 분포한다.  
 면적의 상당 부분을 도봉산이 차지하는 도봉구에는 한식당이 가장 적은 것을 확인할 수 있다.



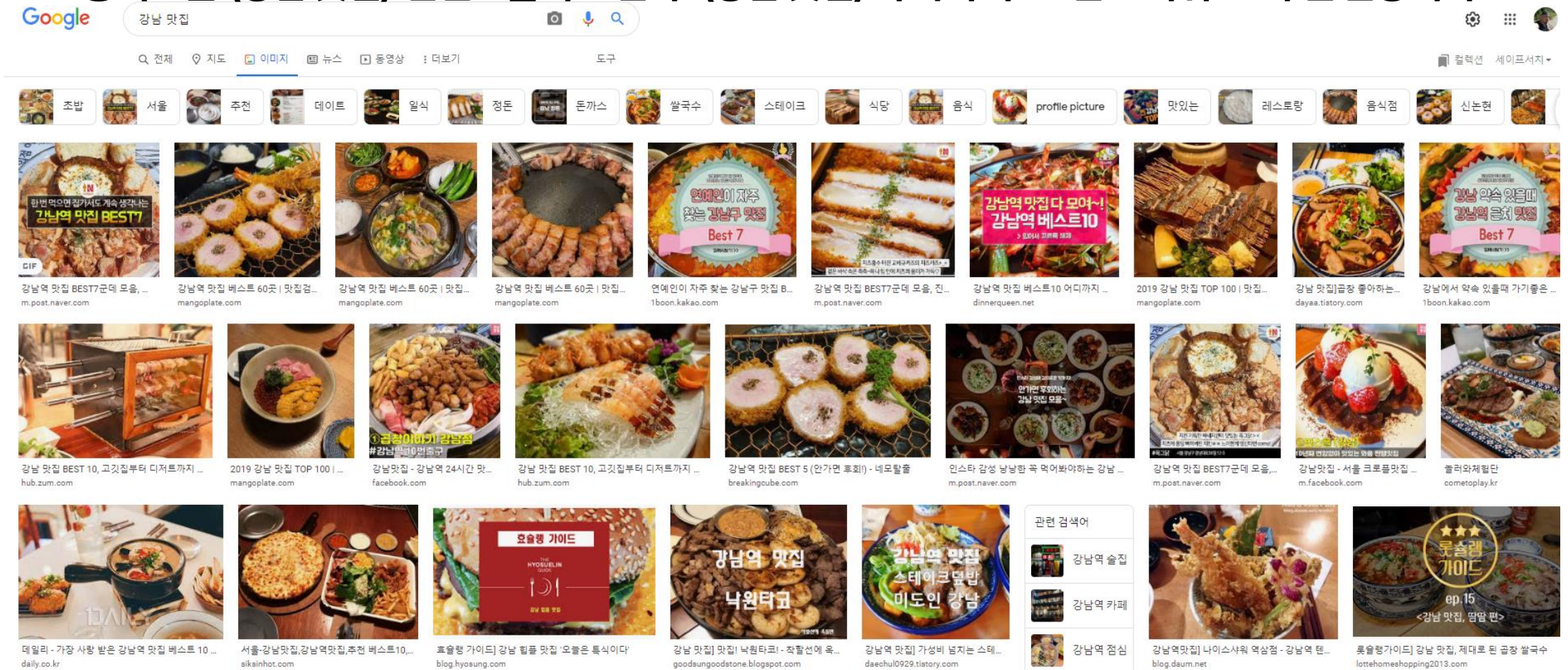
## 〈강남맛집〉과 〈마포맛집〉에 대한 관심도 비교 분석 (네이버 검색 트렌드)

한식당의 갯수는 강남과 마포가 큰 차이가 나지 않지만  
〈동네맛집〉에 대한 “관심”은 강남구가 마포구보다 훨씬 많은 것을 알 수 있다.



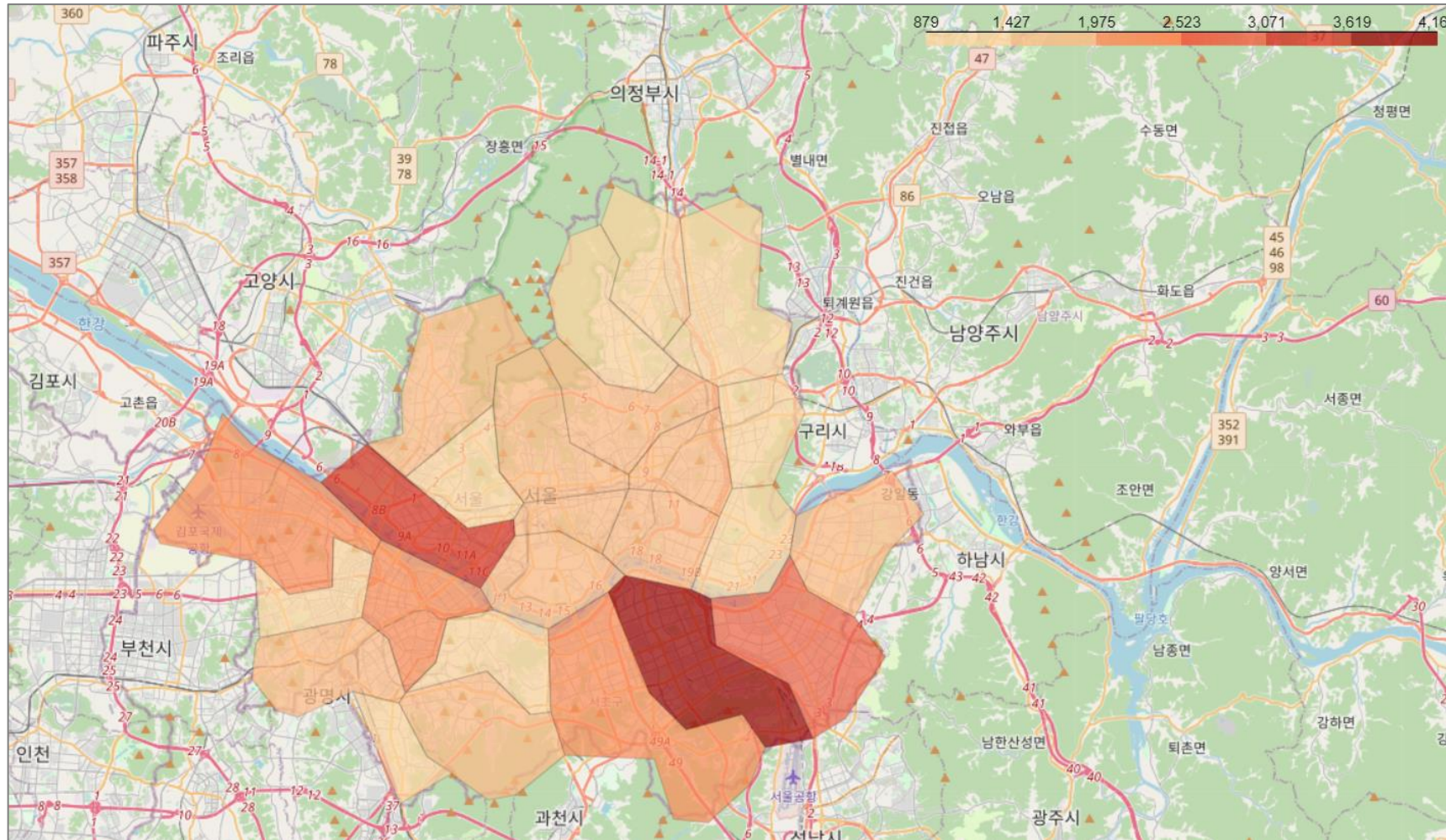
# <강남맛집> 관련 콘텐츠 (구글 이미지 검색)

망고플레이트, 핀터레스트, 유튜브, 네이버포스트 등 다양한 채널에  
 “가장 사랑받은 강남역 맛집 베스트”, “실패 없는 강남맛집” 등  
 흥미로운 <강남맛집> 콘텐츠들이 보인다. <강남맛집>이 하나의 “트렌드 키워드”가 된 현상이다.



## 서울시 구별 한식당 분포 밀도

강남구, 마포구, 송파구, 서초구 등 한식당이 많은 구를 지도상에서 쉽게 찾을 수 있다.  
지금까지 <상권> 정보를 파악하였다. 다음번에는 여러 데이터를 결합해 추가 분석을 시작해보자.





## 1. 곤란을 겪고 있는 소상공인과 업종들

전반적으로 <일반음식점>, <한식> 업종이 어려움을 겪고 있는 상황이다.  
최근의 <배달> 업종의 급성장까지 반영한다면 문제는 더욱 심각해질 것이다.  
보증 신청액 확대, 신청 절차 간편화로 소상공인과 업종을 지원할 필요가 있다.

## 2. 데이터를 통한 상권 활성화 전략 모색 필요

지역 맛집, 동네 맛집에 대한 관심이 큰 것을 확인하였다.  
현 데이터에서는 한계가 있으므로, 이 분석에 맞는 데이터를 수집/분석하여  
화성시 등 “한식당은 많은데 맛집 검색이 적은” 지역의 활성화 전략을 도출해야 한다.

전반적으로 소상공인, 상권, 업종들의 현황과 트렌드를 살펴보았다.  
지금은 현재 상황에서 구할 수 있는 데이터로 다음 분석을 진행하고자 한다.

“음식점이 많은 지역에 음식물 쓰레기도 많이 나올까?”

## Story #02

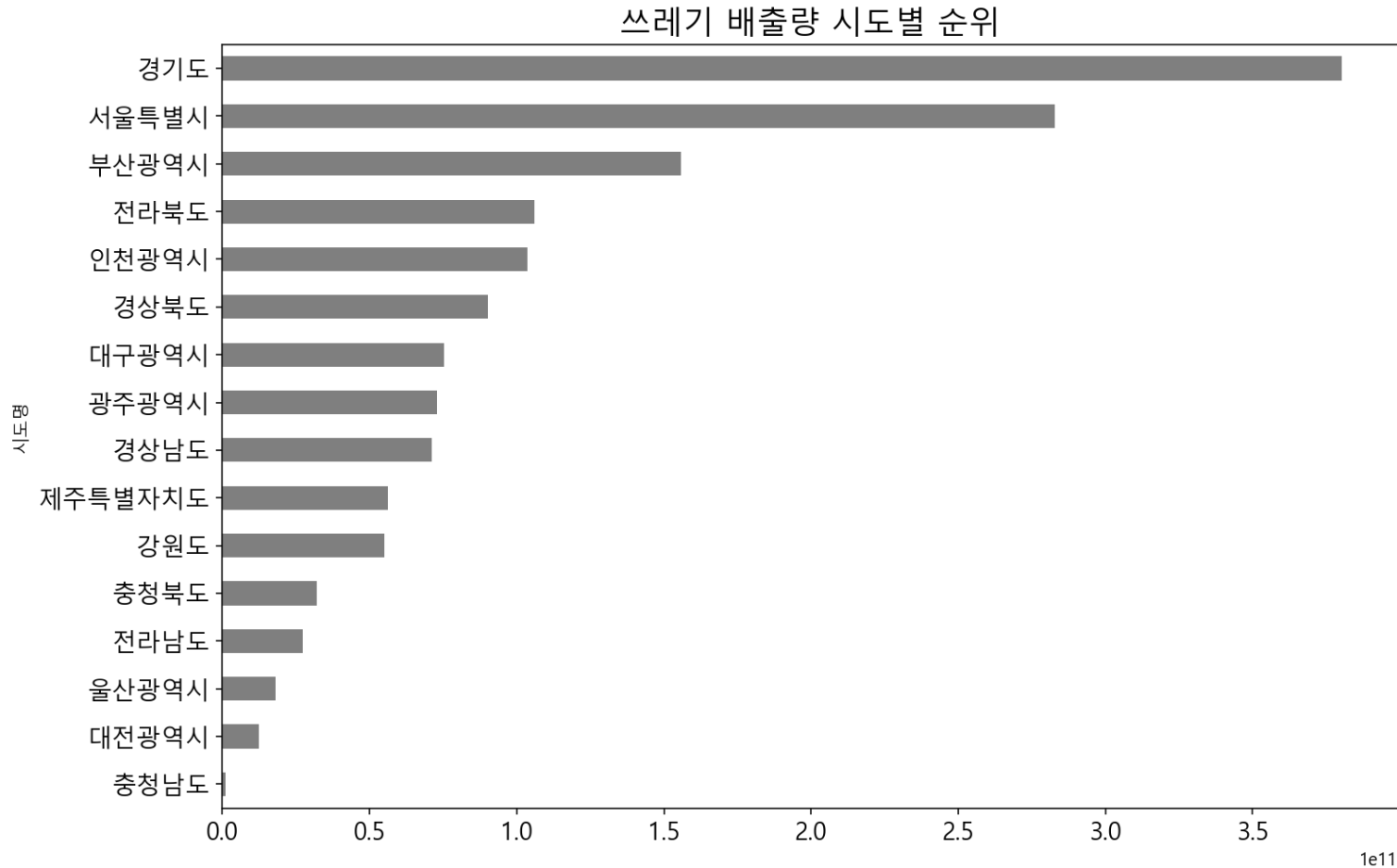
음식물 쓰레기 배출현황 및 음식물 쓰레기와 음식상권과의 연관관계

## 전국 시도별 쓰레기 배출량 순위

**가장 음식물 쓰레기가 많은 지역은 <경기도>, <서울>, <부산> 순이다.**

충청남도에 쓰레기가 나오지 않는 것은 결측치가 의심된다.

경기도에 가장 쓰레기가 많은 것과 <쓰레기매립지>와의 관계는 다른 데이터를 통해 확인해볼 예정이다.

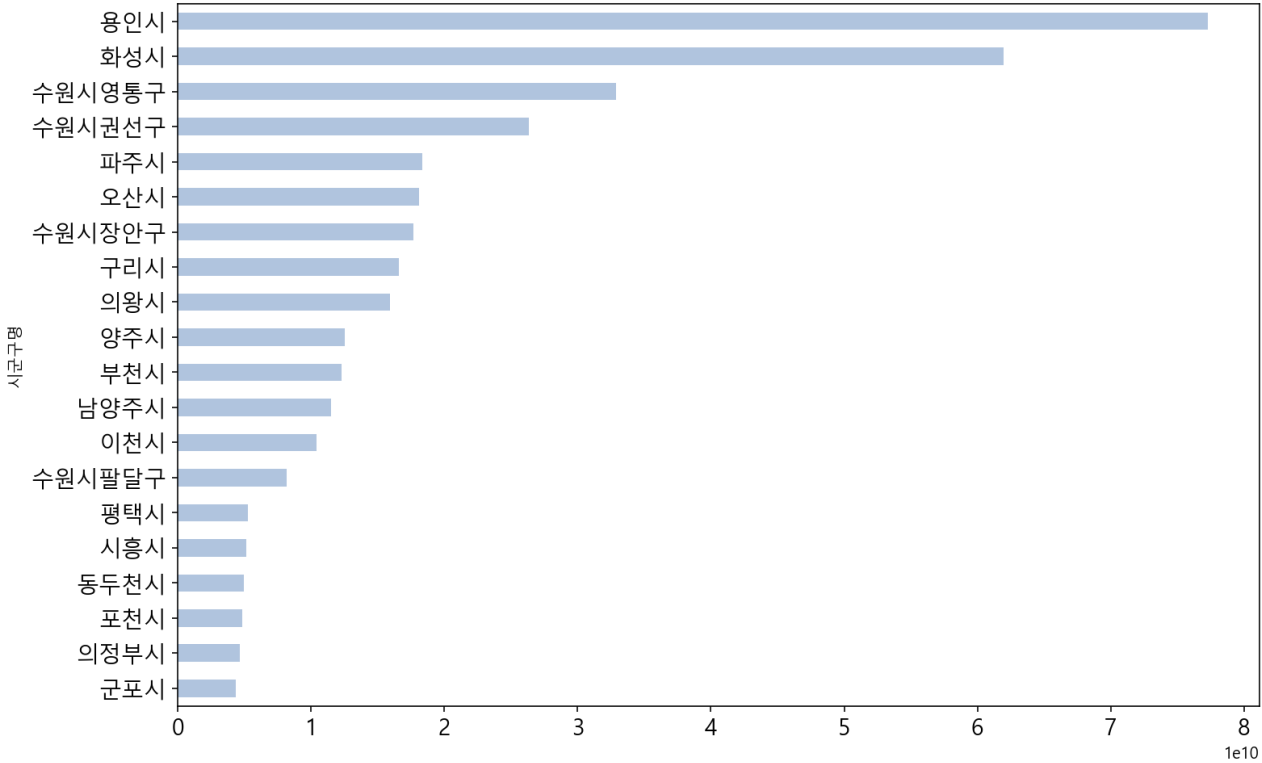


# 〈경기도〉시군별 음식물 쓰레기 배출량 상위 20

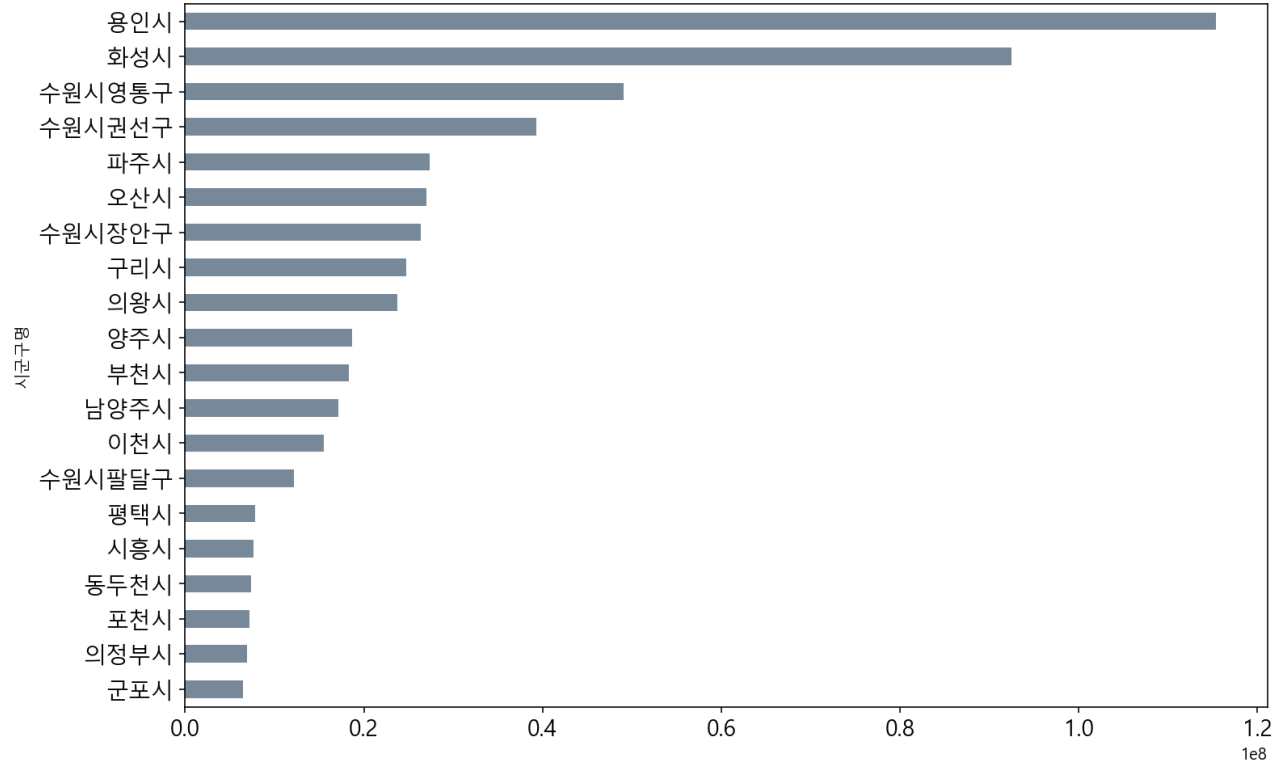
**〈용인시〉, 〈화성시〉 순으로 음식물 쓰레기가 많이 배출되고 있다.**

앞의 “상권데이터”에서는 용인시를 〈용인시 처인구〉, 〈용인시 수지구〉, 〈용인시 기흥구〉 등으로 나누었으나 “쓰레기데이터”는 3개구를 합해 〈용인시〉로 수집해 순위가 달라졌다. 지역명 처리가 필요하나 여기에선 생략한다.

경기도 시군구별 쓰레기배출 총량 상위 20



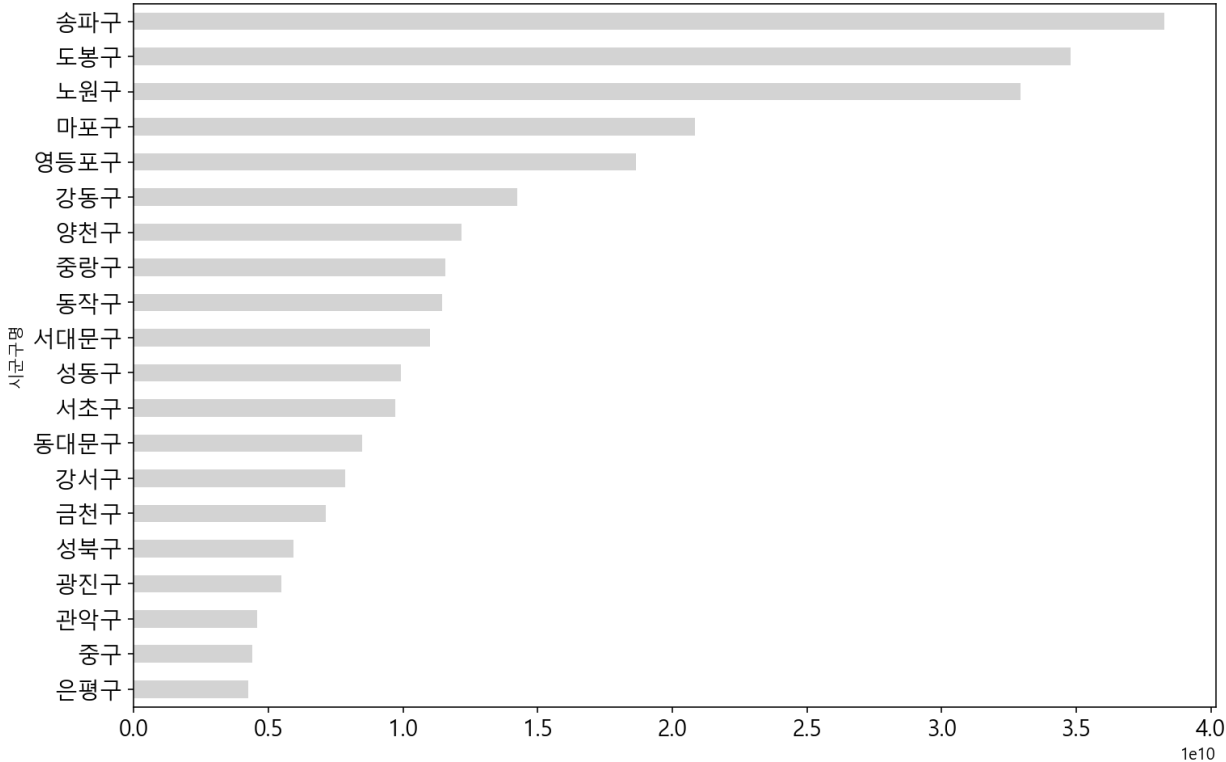
경기도 시군구별 쓰레기배출 평균 상위 20



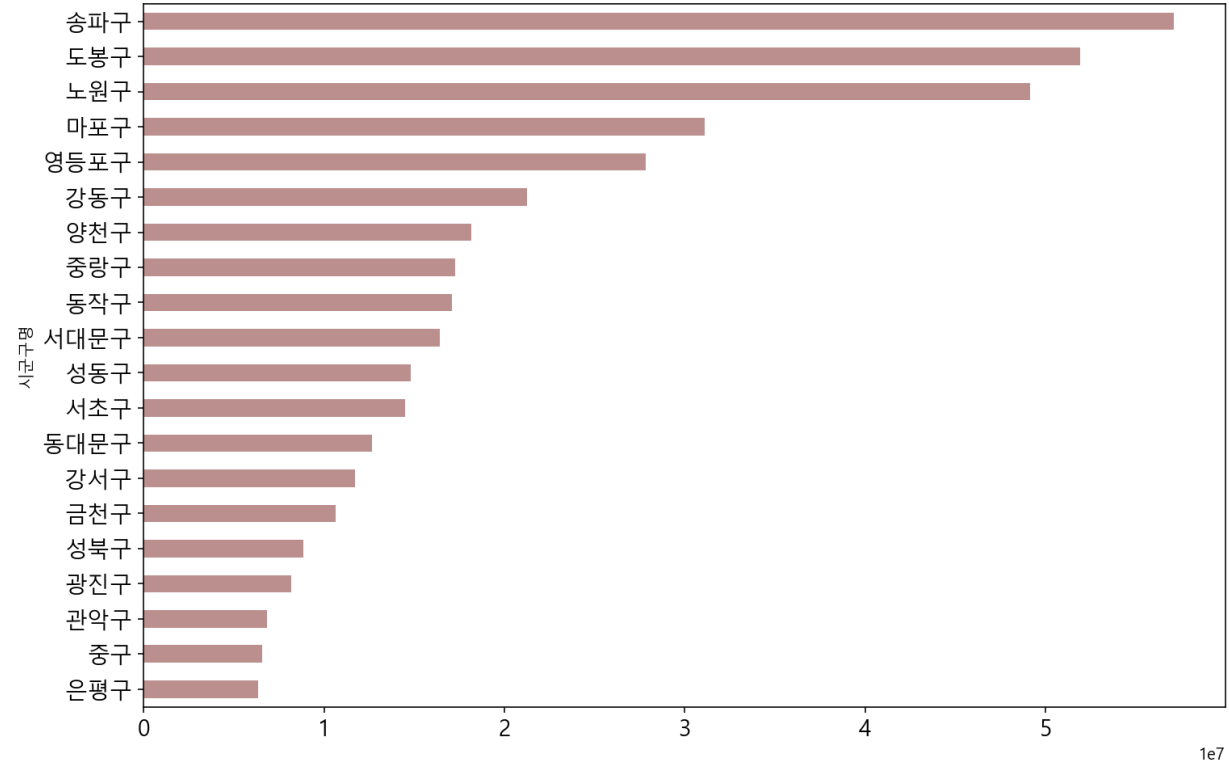
# 〈서울시〉 구별 음식물 쓰레기 배출량 순위

음식물 쓰레기가 가장 많은 지역은 〈송파구〉, 〈도봉구〉, 〈노원구〉 순이며  
 〈관악구〉, 〈중구〉, 〈은평구〉는 상대적으로 쓰레기가 많이 나오지 않는 것을 볼 수 있다.  
 “지역명”이 잘 분류된 서울시 데이터셋으로 추가 분석을 진행한다.

서울시 구별 쓰레기배출 총량 상위 20

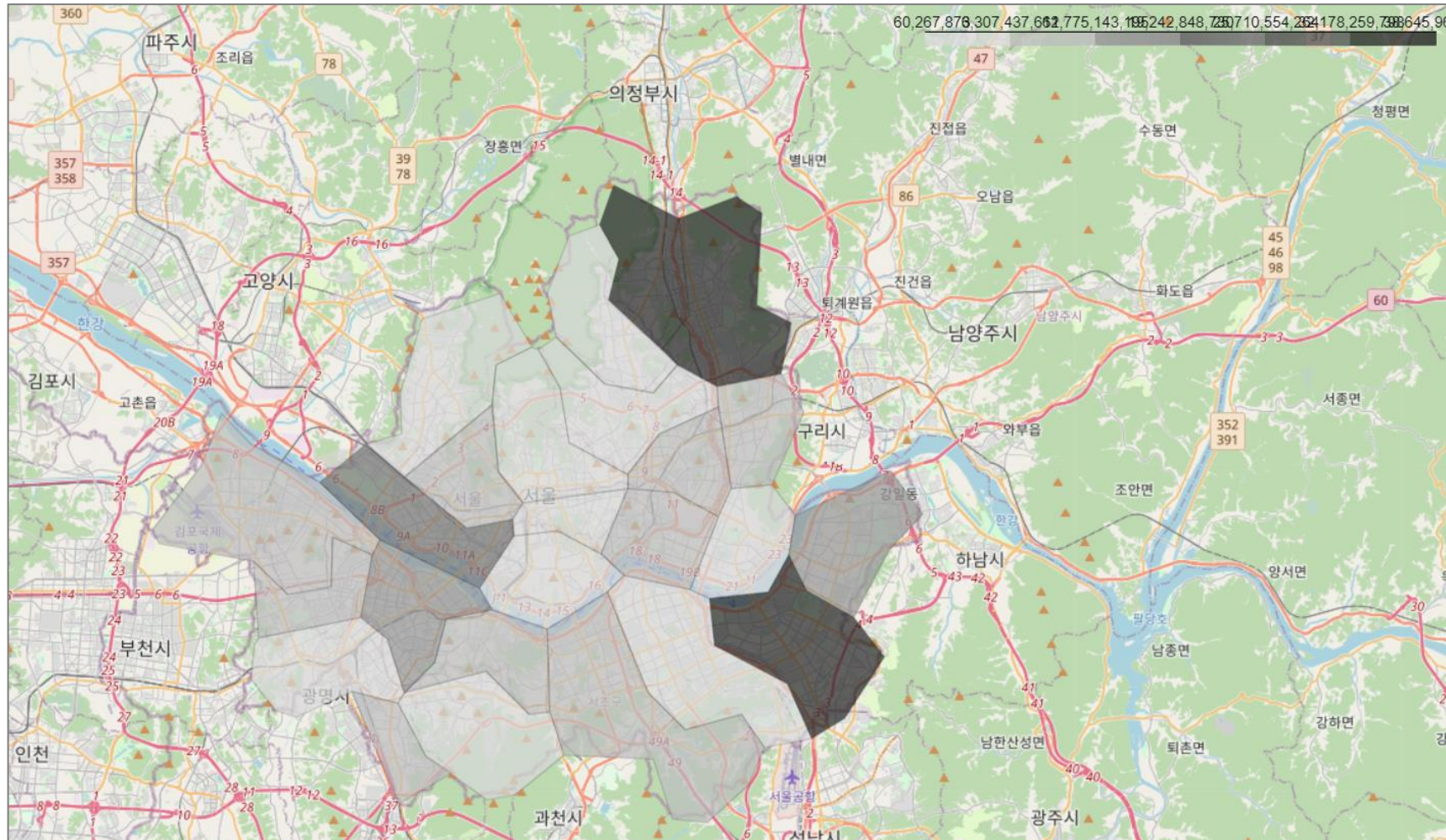


서울시 구별 쓰레기배출 평균 상위 20



# 서울시 구별 음식물 쓰레기 배출량 밀도

음식물 쓰레기가 많은 <송파구>, <도봉구>, <노원구> 등을 지도상에서 쉽게 찾을 수 있다.  
그런데 <음식점 수>와 <음식물 쓰레기 양>은 관계가 있을까?

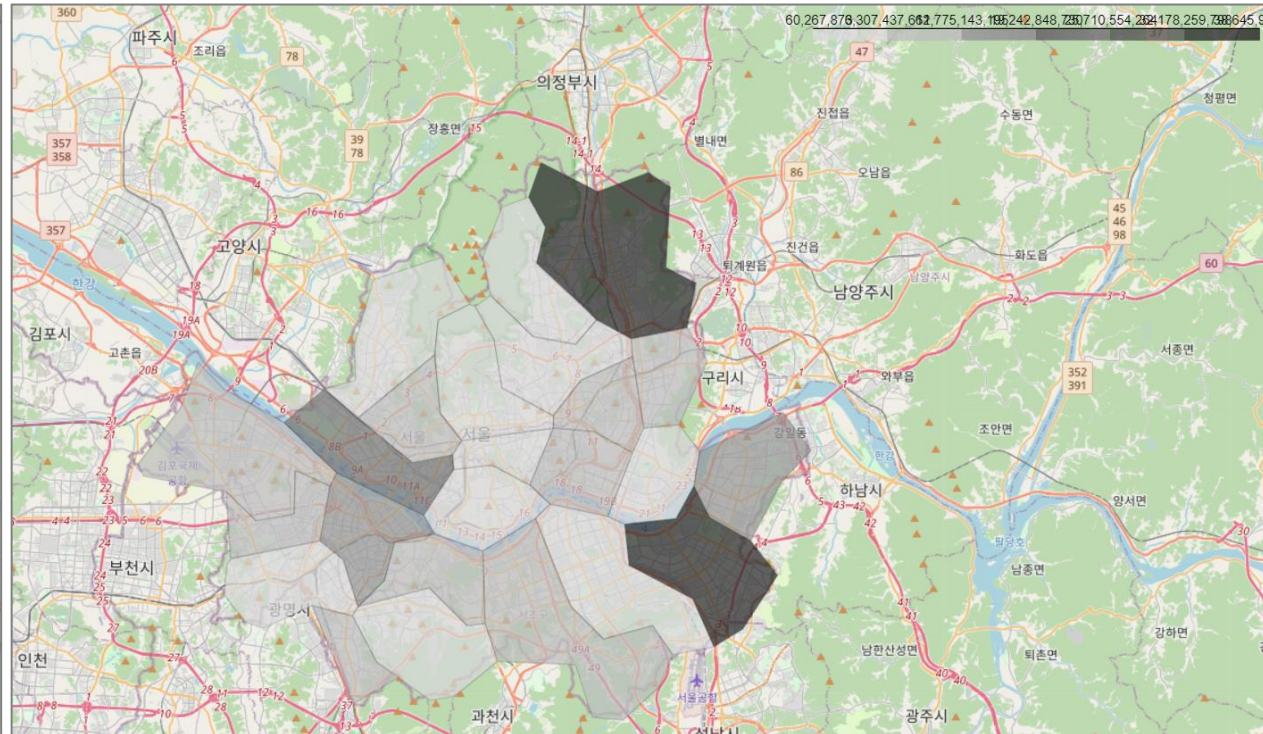
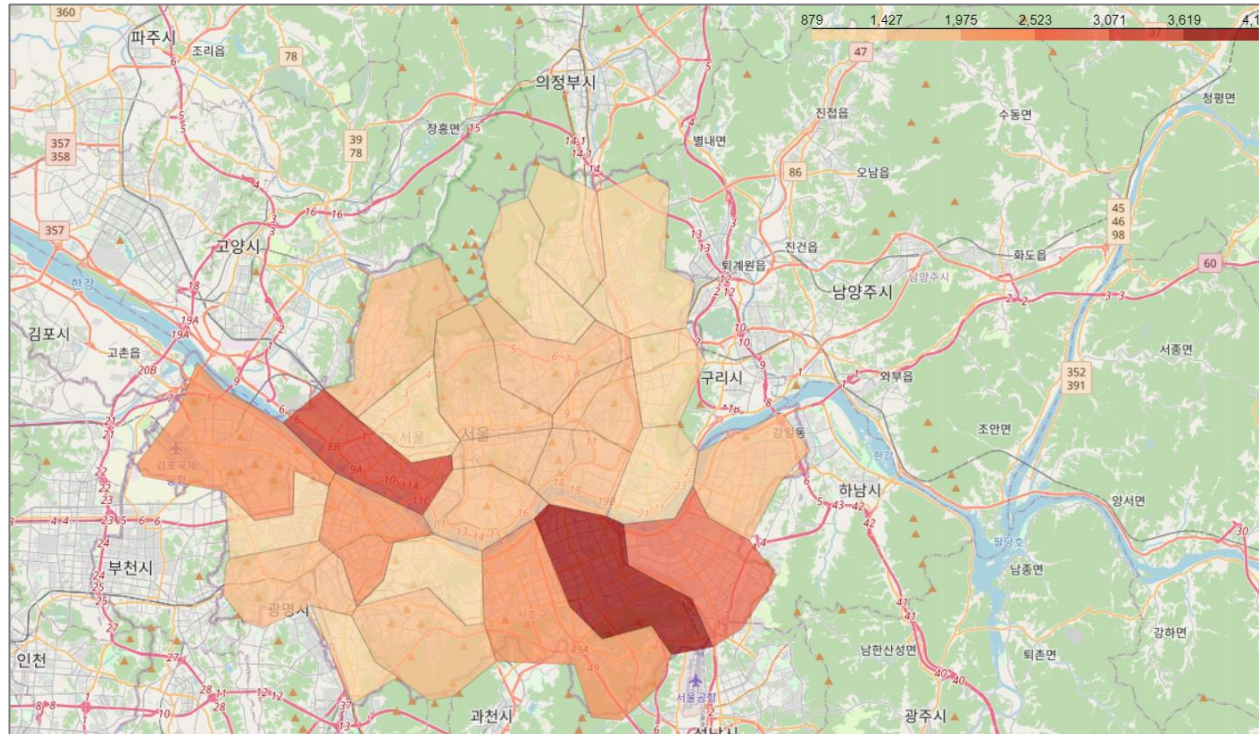


# 서울시 구별 <음식점 수>와 <음식물 쓰레기양> 비교

음식점 수가 가장 적은 <도봉구>가 음식물 쓰레기는 많은 것을 확인할 수 있다.  
<마포구>의 경우 음식점 수도 음식물 쓰레기도 많지만, 전반적으로 두 데이터는 달라 보인다.  
음식점 수와 음식물 쓰레기가 어떤 관계가 있는지 상관분석을 통해 “정확한 수치”로 확인해보자.

[ 서울시 음식점 분포 ]

[ 서울시 음식물 쓰레기 양 ]



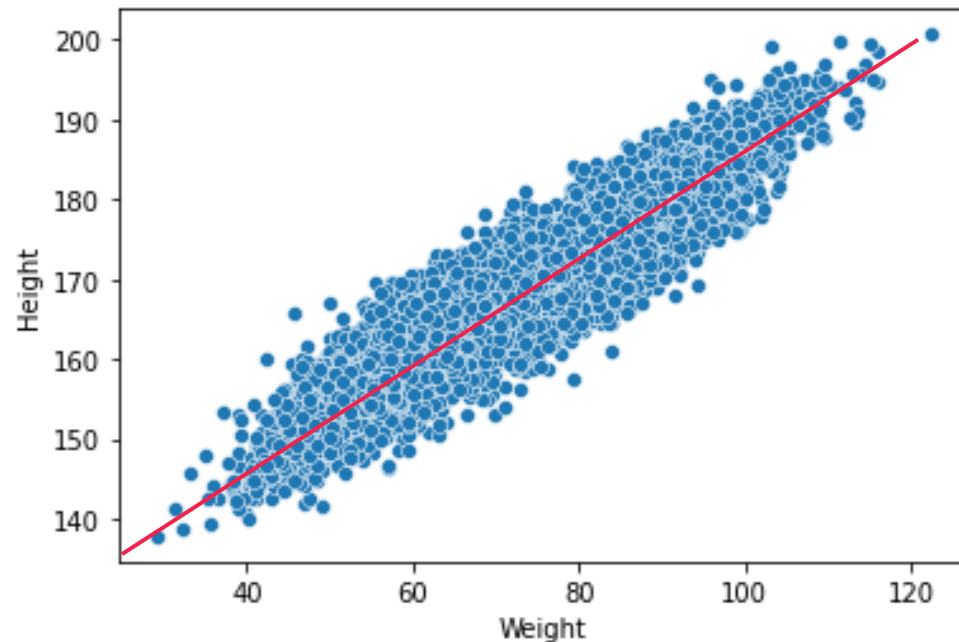
## 심화학습: 데이터 간의 관계와 크기를 파악하는 <상관분석>

일반적으로 사람의 “키”가 커질수록 “몸무게”도 함께 증가한다.

이처럼 <IQ와 수학 성적>, <방의 개수와 집의 가격>, <인구 밀도와 범죄율> 등에 대해 상관분석으로 “데이터 사이에 무슨 관계가 있는지”, “그 관계가 얼마나 큰지 또는 작은지”를 파악한다.

```
[4] sns.scatterplot(x='Weight', y='Height', data=df)
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fde0d083550>



상관분석 :

[https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80\\_%EB%B6%84%EC%84%9D](https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80_%EB%B6%84%EC%84%9D)



## 심화학습: 상관성의 크기 척도인 <피어슨 상관계수>

피어슨 상관계수  $r$ :  $-1 \sim 1$  사이의 값, 상관관계의 강/약 비교

### ✓ 0보다 크면 양(+)의 관계:

하나가 커지면 다른 하나가 같이 늘어나는 관계(키/몸무게)

### ✓ 0보다 작으면 음(-)의 관계:

하나가 커지면 다른 하나가 반대로 줄어드는 관계(겨울철 기온/난방비)

✓ +1에 가까우면 강한 양의 관계, -1에 가까우면 강한 음의 관계

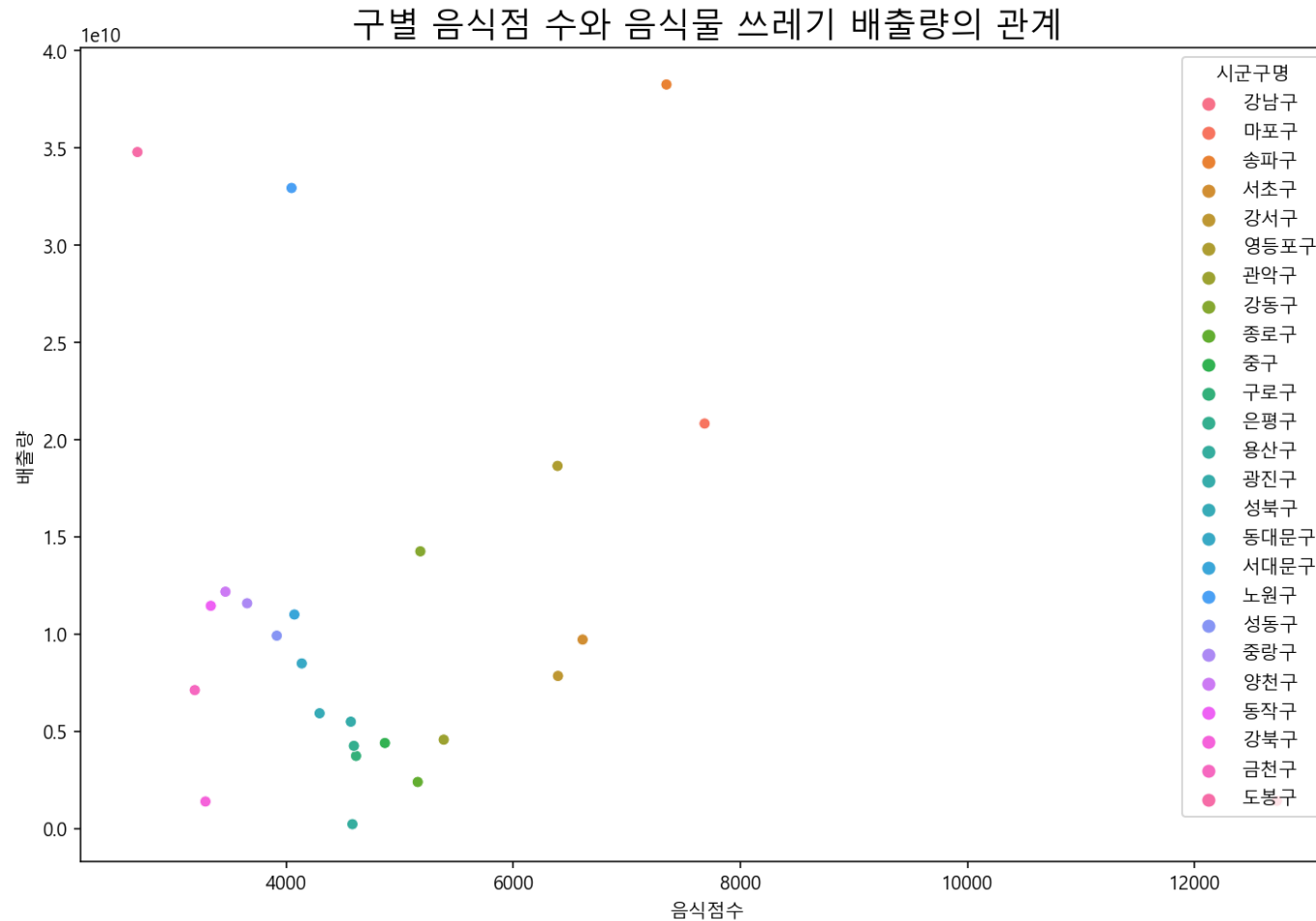
✓ 0에 가까울수록 관계는 약해진다 (0: 상관관계 없음)

피어슨 상관계수 :

[https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80\\_%EB%B6%84%EC%84%9D#%ED%94%BC%EC%96%B4%EC%8A%A8\\_%EC%83%81%EA%B4%80\\_%EA%B3%84%EC%88%98](https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80_%EB%B6%84%EC%84%9D#%ED%94%BC%EC%96%B4%EC%8A%A8_%EC%83%81%EA%B4%80_%EA%B3%84%EC%88%98)

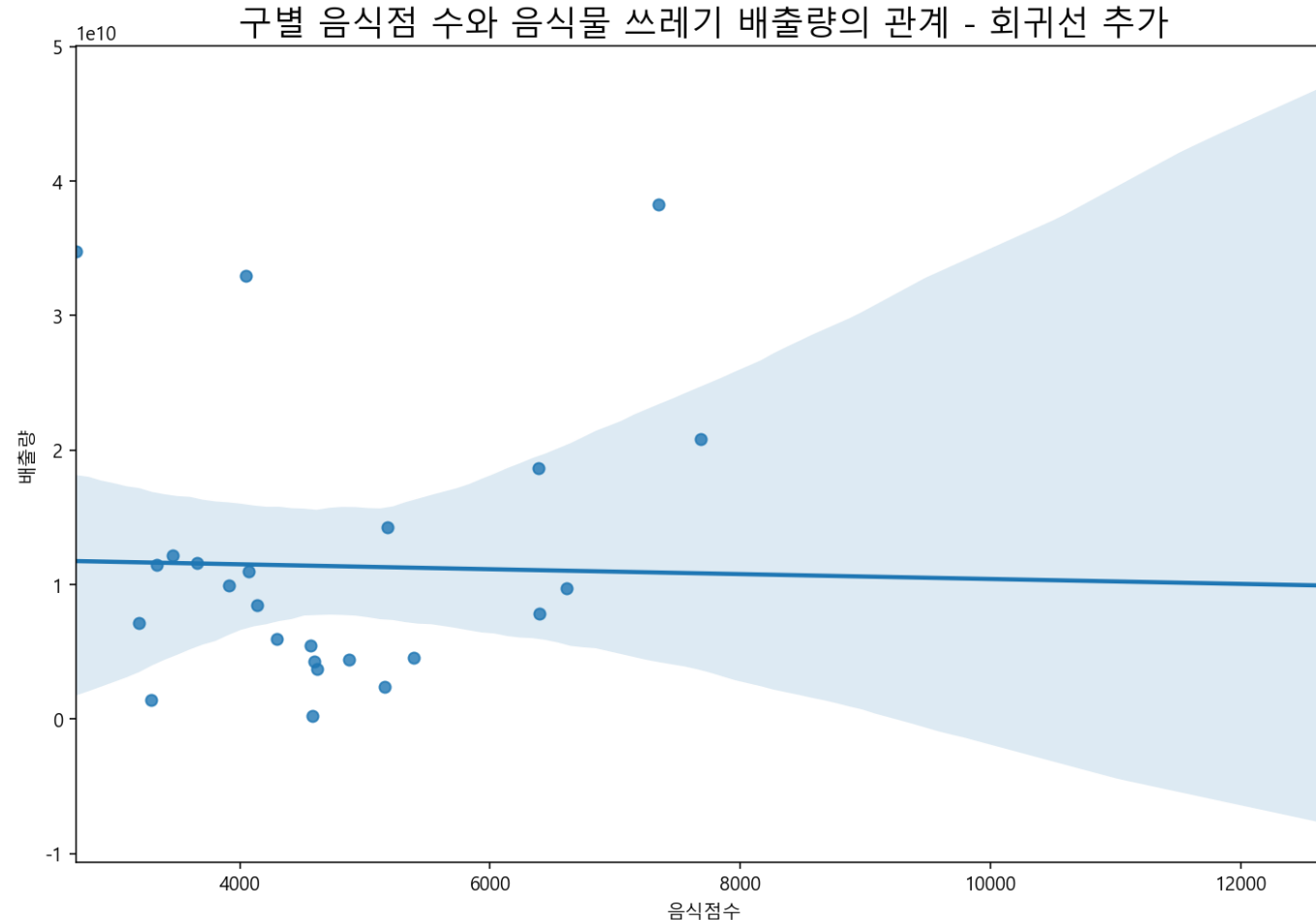
## 구별 <음식점 수>와 <음식물 쓰레기 양> 사이의 관계 - 산포도(scatterplot)

<강남구>, <마포구> 등은 음식점 수도 많고 음식물 쓰레기도 많은 반면 <도봉구>처럼 음식점 수는 적은 반면, 나오는 음식물 쓰레기는 많은 구도 있다.  
**이 시각화에서는 정확한 상관관계 판단이 어려우므로 추가 분석을 실시한다.**



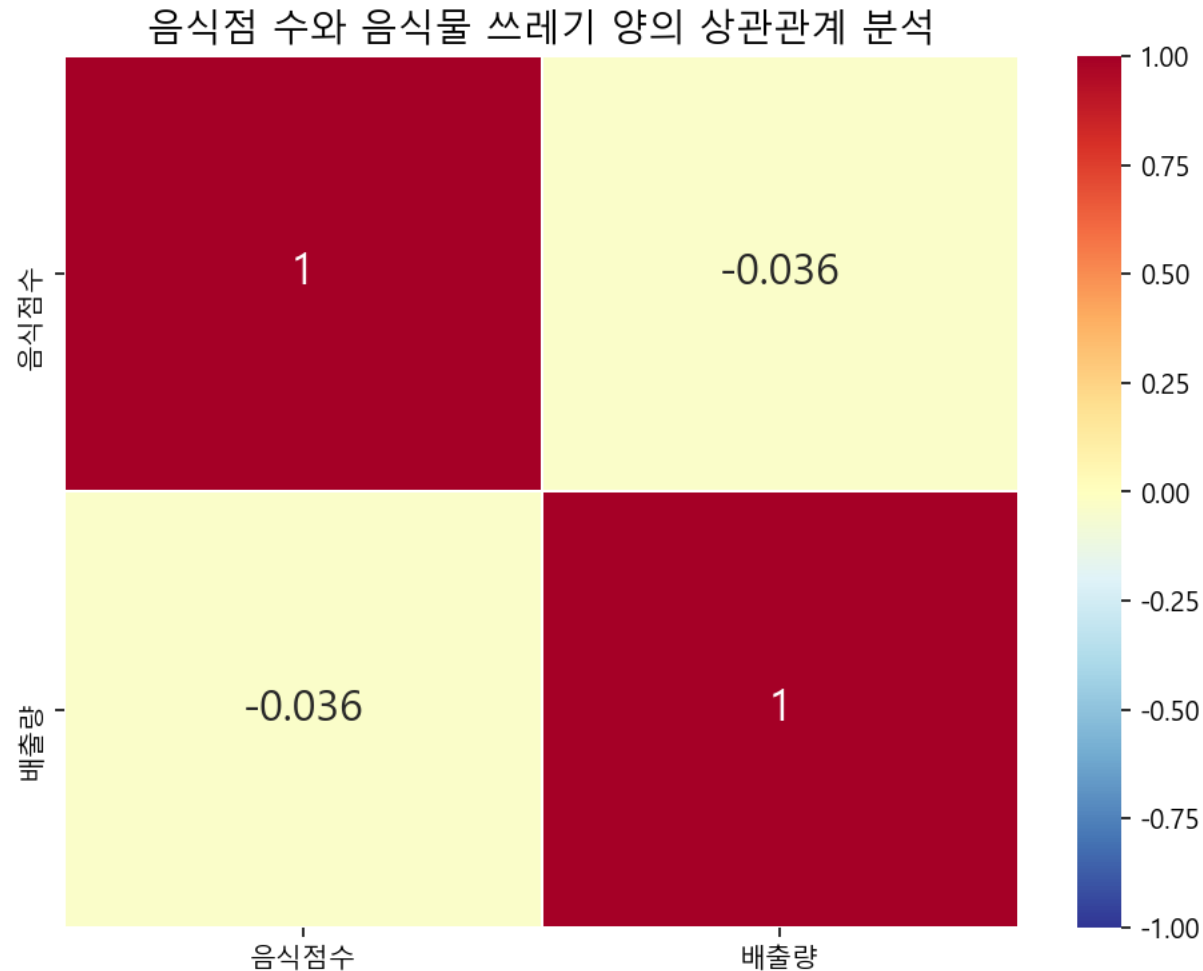
## 구별 <음식점 수>와 <음식물 쓰레기 양> 사이의 관계 - 회귀선 추가

상관관계의 크기와 방향을 나타내는 <회귀선>을 그어본 결과 기울기, 즉 <피어슨 상관계수>는 0에 가까워 음식점 수와 음식물 쓰레기 양은 상관관계가 거의 없다.



## 〈음식점 수〉와 〈음식물 쓰레기 양〉 사이의 상관분석 - heatmap

〈음식점수〉와 〈음식물 쓰레기 양〉의 상관계수는  $-0.036$ 이다.  
이것은 오히려, 음식점 수가 많아질수록 음식물 쓰레기는 조금이나마 줄어드는 것을 의미한다.



〈음식점수가 많을수록 음식물 쓰레기 양은 많을 것이다〉라는 가설 하에  
〈상권 데이터〉와 〈음식물 쓰레기 배출량 데이터〉를 결합,  
다양한 시각화 분석을 거쳐 최종 상관관계 분석까지 진행하였다.

그러나, 해당 분석에서는 예상과 달리 〈음식점 수〉와 〈음식물 쓰레기 양〉은  
상관관계가 별로 없는 것으로 나타났다. 그 이유로는:

**1. 〈음식점〉만 분석 대상으로 삼은 결과**  
〈일반 가정집〉에서 나오는 쓰레기 배출량은 고려되지 않았다.  
〈지역별 가구수〉를 포함한다면 더 정확한 분석 결과가 도출될 것이다.

**2. 코로나로 인해 〈배달〉 트렌드가 대세**  
절대 배달이 되지 않을 것 같았던 “곱창”, “삼겹살” 등도 배달이 되는 요즘이다.  
본 데이터는 코로나 이전 수집된 것으로 최근 상황이 반영되지 않았다.

**3. 〈배달〉로 생기는 쓰레기는 주로 〈재활용 쓰레기〉**  
여기에서 사용한 데이터는 〈음식물 쓰레기〉 데이터로,  
목적대로의 분석을 위해서는 〈재활용 쓰레기〉 데이터가 필요하다.

이상의 문제가 해결되면 현 상황을 정확히 반영한 결과를 도출할 수 있을 것이나,  
지금은 현재 상황에서 구할 수 있는 데이터로 추가 분석을 진행하고자 한다.

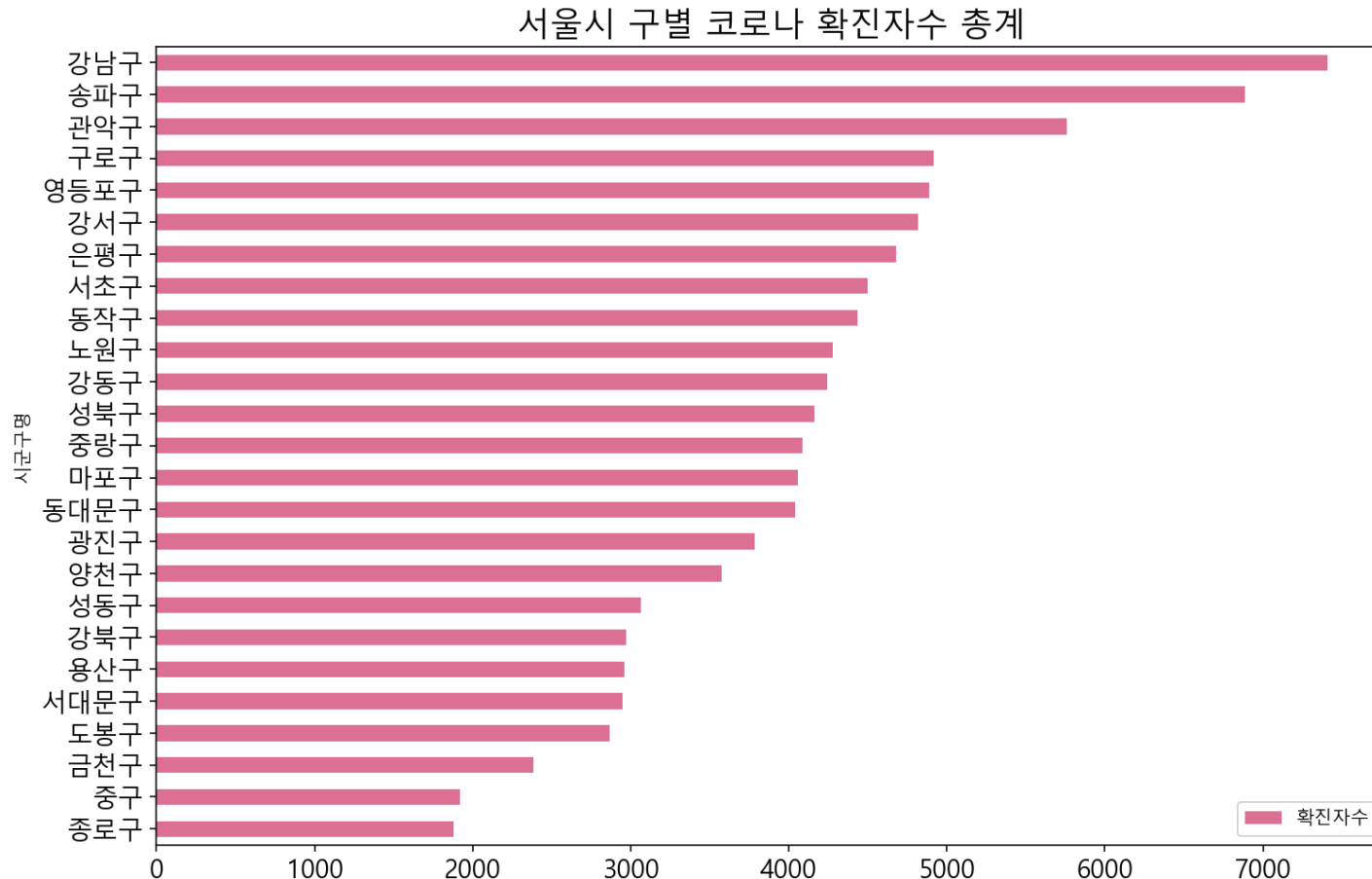
“코로나 확진자가 많으면, 음식물 쓰레기가 많이 나올까?”

## Story #03

음식물 쓰레기 배출현황 및 음식물 쓰레기와 코로나 확진자수의 연관관계

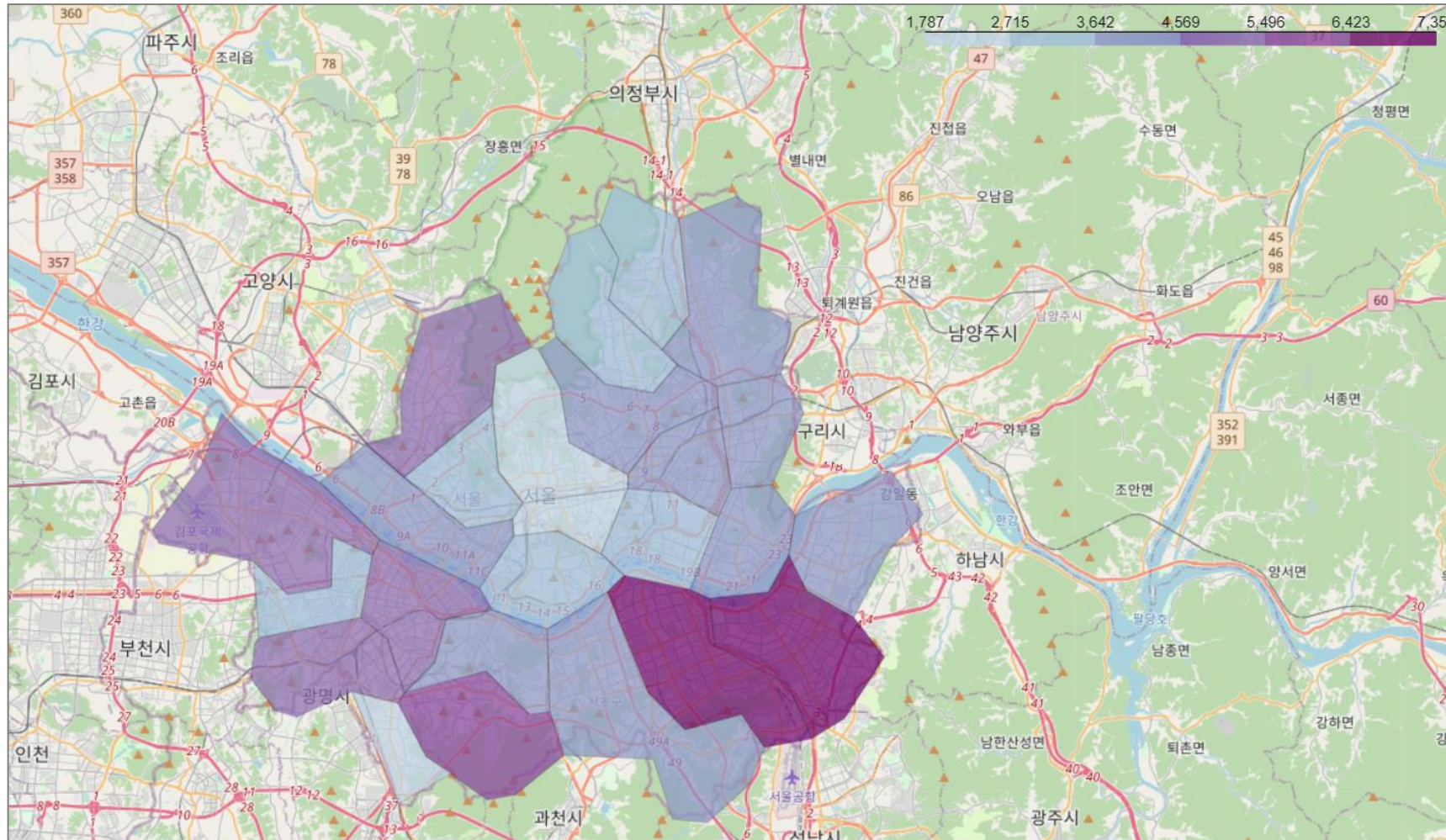
## 〈서울시〉 구별 코로나 확진자수

코로나가 가장 많이 발생한 지역은 〈강남구〉, 〈송파구〉, 〈관악구〉 순이며  
기업들이 다수 위치해 “출퇴근 인구”가 많은 〈중구〉, 〈종로구〉에는 확진자가 적게 발생하였다.  
〈다른 데이터와의 융합〉 등 다양한 추가 분석을 통해 원인을 알아보기로 한다.



## 서울시 구별 코로나 확진자수 밀도

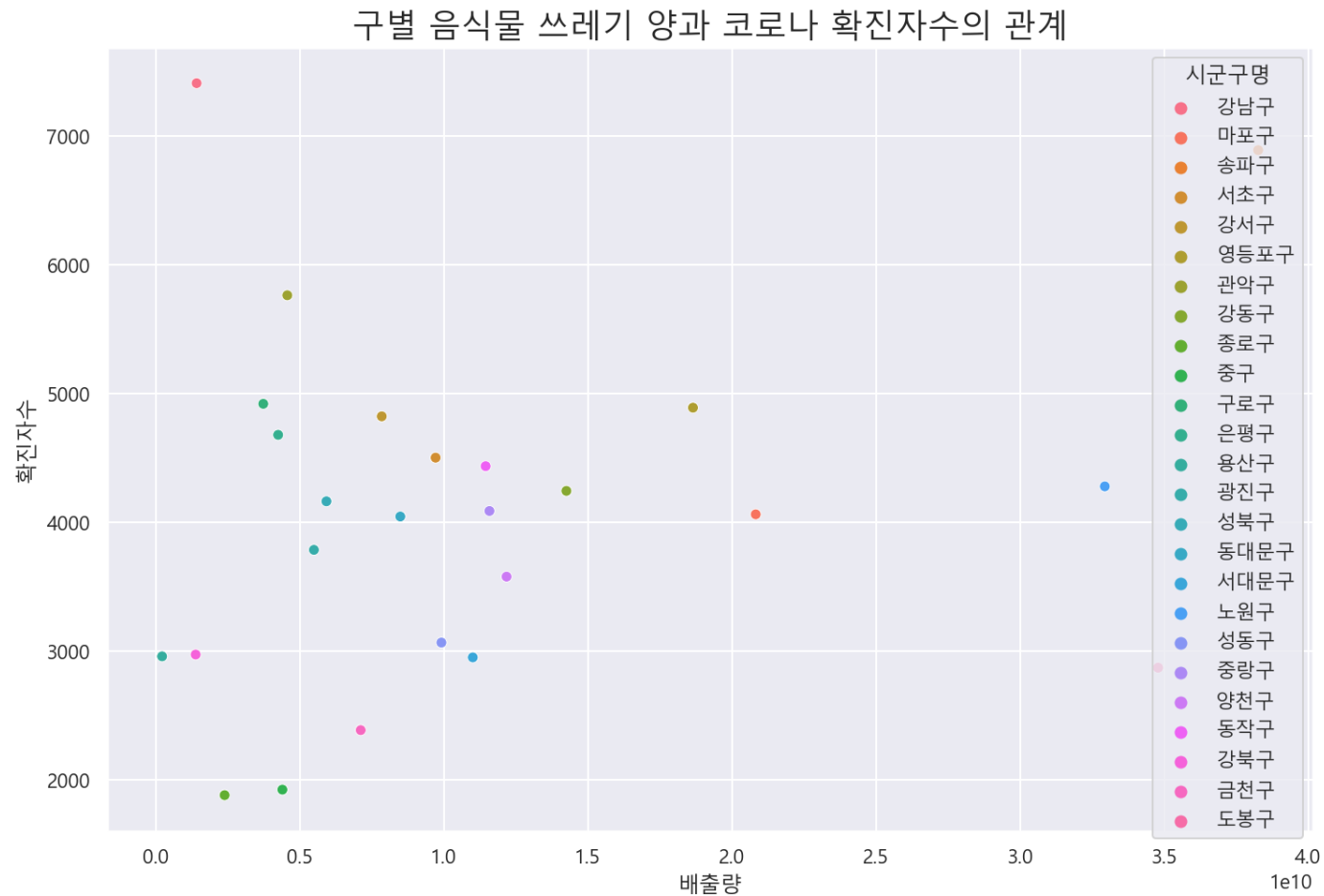
확진자수가 많은 <강남구>, <송파구>, <관악구> 등을 지도에서 색깔로 구분할 수 있다  
그런데 <코로나 확진자 발생수>는 다른 어떤 요소와 관계가 있을까?





## 〈음식물 쓰레기양〉과 〈코로나 확진자수〉 사이의 관계 - 산포도(scatterplot)

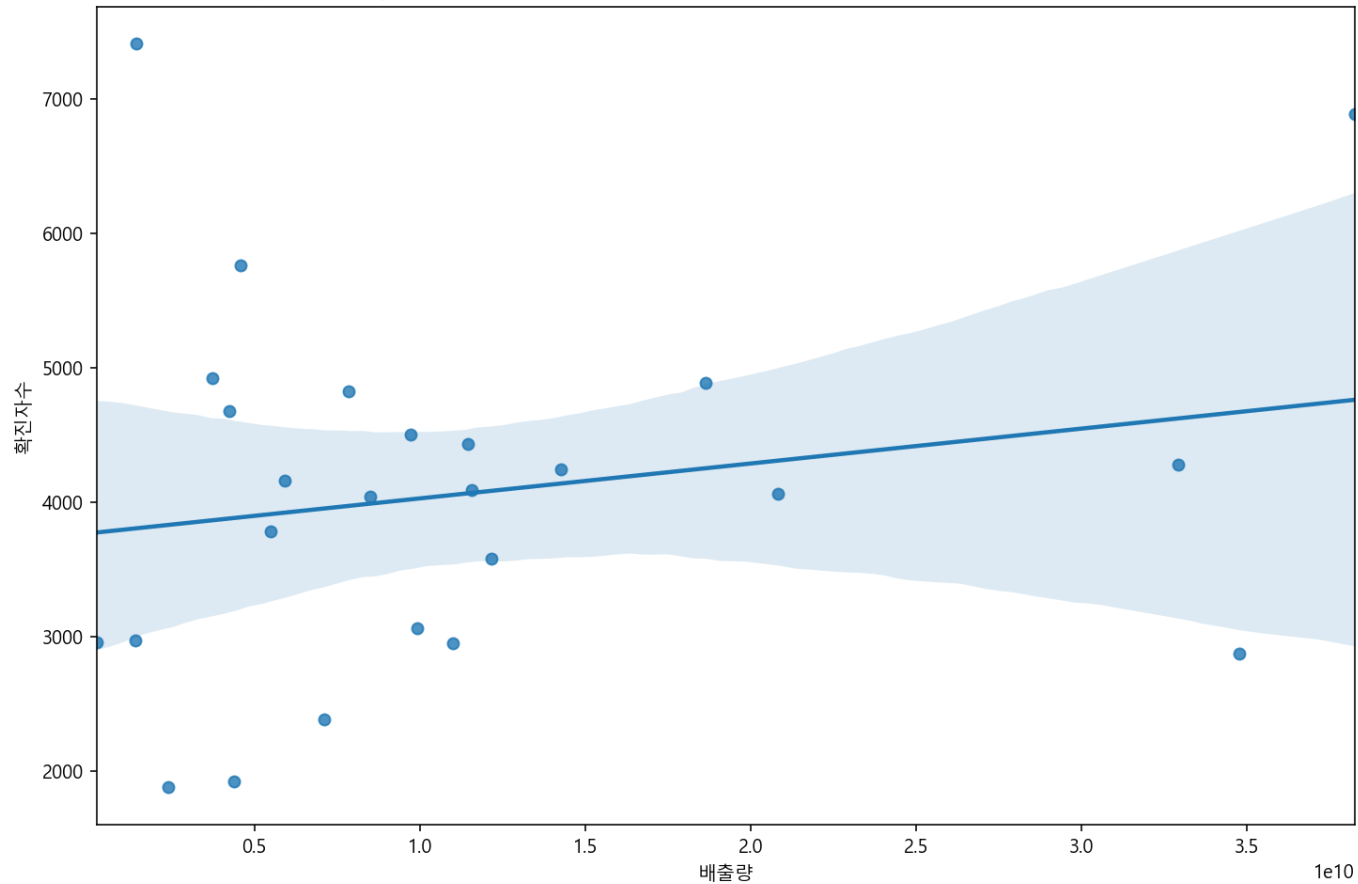
〈음식물 쓰레기양〉과 〈코로나 확진자수〉 사이의 관계에 대해 이 시각화에서는 경향성을 뚜렷이 보기 어려우므로 추가 분석을 실시한다.



## 〈음식물 쓰레기양〉과 〈코로나 확진자수〉 사이의 관계 - 회귀선 추가

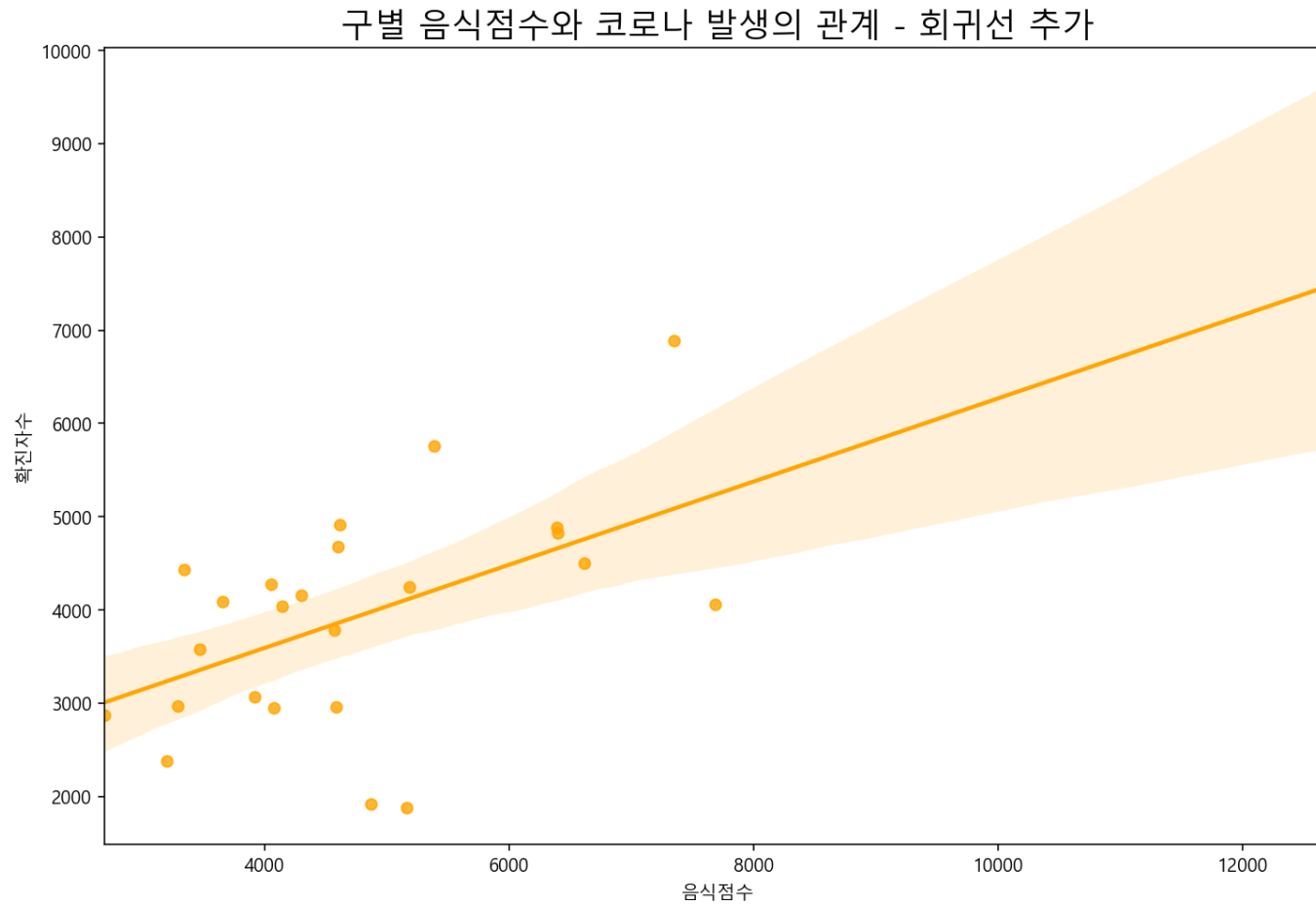
상관관계의 크기와 방향을 나타내는 〈회귀선〉을 그어본 결과  
기울기, 즉 〈피어슨 상관관계수〉는 0보다 약간 커서  
〈음식물 쓰레기양〉과 〈코로나 확진자수〉는 약한 양의 상관관계를 보였다.

구별 음식물 쓰레기 배출량과 코로나 발생의 관계 - 회귀선 추가



## 〈음식점 수〉 과 〈코로나 확진자 수〉 사이의 관계 - 회귀선 추가

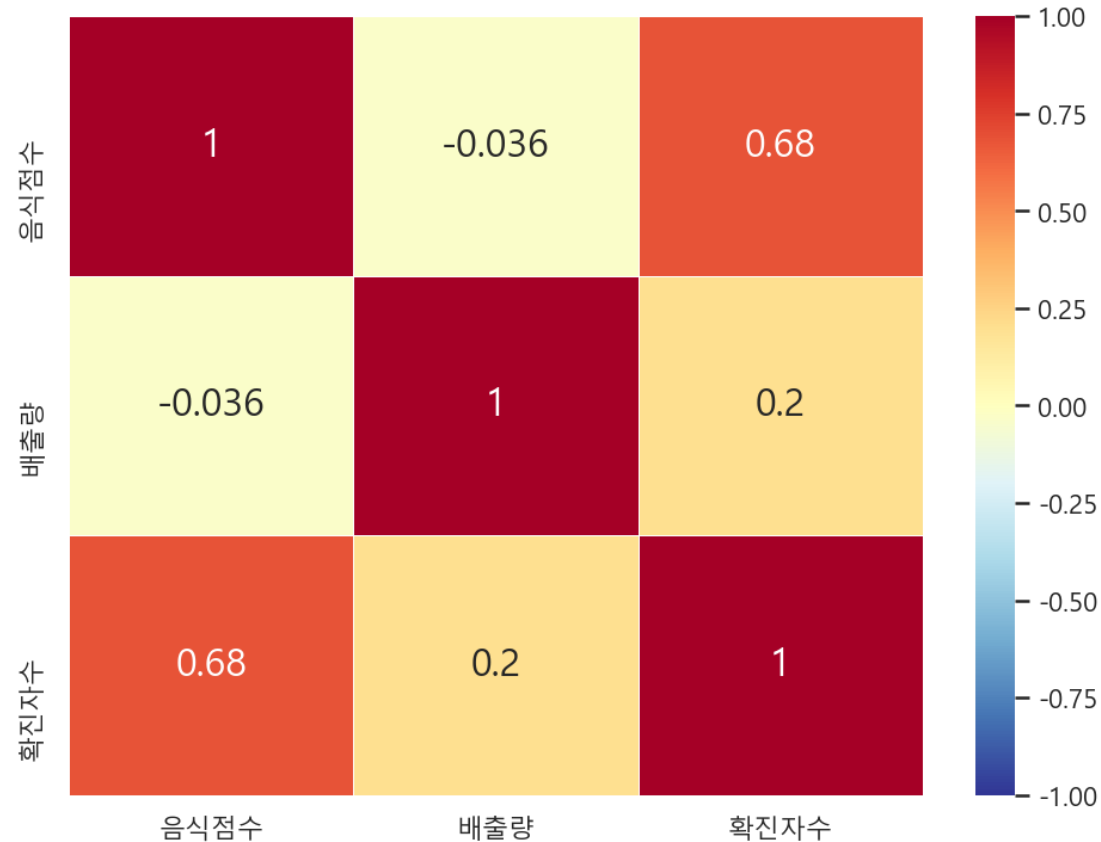
또한, 〈음식점 수〉과 〈코로나 확진자 수〉 역시 양의 상관관계를 보였다.



## 〈음식점 수〉, 〈음식물 쓰레기 양〉, 〈코로나 확진자수〉 간 상관분석 - heatmap

**〈확진자수〉와 〈음식점 수〉의 상관계수는 0.68로 매우 크다. 그런데 이것이 〈음식점〉이란 장소가 코로나의 주요 발생 원인이라는 것을 의미한다고 볼 수 있을까?**  
**〈코로나 발생〉이 〈음식물쓰레기〉에 어떤 작용을 하기에 상관성이 있을까?**

음식점 수, 음식물 쓰레기 양, 코로나 확진자수의 상관관계



데이터 분석가는 <상관관계>와 <인과관계>를 확실하게 구분해야 한다.

- ✓ 상관관계 : 어떤 사건이 (우연일 수도 있지만) 같이 발생한 것
- ✓ 인과관계 : A라는 사건이 원인이 되어 B라는 사건이 발생한 것

상관관계 : 시간적 선후관계는 있지만, 인과적 선후관계는 아니다

- 현상      **아이스크림 판매량이 늘수록  
익사자 수가 증가한다**

✓ 사실      **여름 평균 기온이 증가하면  
피서객이 증가하여  
익사자 수가 증가한다  
(아이스크림이 원인이 아니다)**

상관관계: 시간적 선후관계는 있지만, 인과적 선후관계는 아니다

- 현상

교회 수가 늘수록  
범죄율이 높아진다

✓ 사실

지역 인구가 많아지면  
여러 이해관계/사건사고가 동반되고  
범죄 발생도 함께 증가한다  
(교회가 원인이 아니다)

인과관계: 원인과 그로 인한 결과가 있다

- 현상

물에 열을 가하면  
물이 끓는다

- 특징

자연적, 물리적 법칙에 많으며  
사회 현상에선 명확하게 구분하기 어렵다



## 분석의 결론 :

〈코로나 확진자 수가 많으면 자가격리로 택배/배달을 많이 시키게 되어  
음식물 쓰레기 양이 많아질 것이다〉라는 가설 하에

〈상권 데이터〉와 〈음식물 쓰레기 배출량 데이터〉,  
〈코로나 확진자 데이터〉를 결합,

다양한 시각화를 거쳐 최종 상관관계 분석까지 진행한 결과,

**〈코로나 확진자〉가 늘어날 때 〈음식물 쓰레기〉는 조금씩 증가하고  
〈음식점 수〉는 크게 증가(상관계수 0.68)하는 현상을 확인하였다.**

**〈음식점〉이 코로나가 발생하는 원인이기 때문일까?  
그리고 〈음식물 쓰레기 양〉과 〈코로나 확진자 수〉는 무슨 관계가 있을까?**

## 시사점 및 다음 분석 방향:

〈코로나 확진자 수가 많으면 자가격리로 택배/배달을 많이 시키게 되어  
음식물 쓰레기 양이 많아질 것이다〉라는 가설 하에

〈상관데이터〉와 〈음식물 쓰레기 배출량 데이터〉,  
〈코로나 확진자 데이터〉를 결합, 다양한 시각화를 거쳐  
상관관계 분석까지 진행하였으나 가설을 정확히 확인할 수 없었다. 그 이유로는:

1. 〈음식점 수〉만 분석 대상으로 삼은 결과  
〈택배/음식배달〉에서 나오는 음식물 쓰레기는 고려되지 않았다.  
〈택배/음식배달〉 데이터를 포함한다면 더 정확한 분석 결과가 도출될 것이다.

2. 〈상관관계〉와 〈인과관계〉의 정확한 구별  
〈확진자 수〉가 증가할 때 〈음식물 쓰레기〉도 증가하지만  
〈코로나 확진〉이 원인이 되어 〈음식물 쓰레기〉가 증가했다는 의미는 아니다

마찬가지로, 〈음식점〉이 〈코로나 확진〉의 원인이었다는 의미도 아니다.  
상관관계는 증명되었으나, 인과관계는 확인이 필요하다.

〈코로나 확진자 발생〉에 다른 요소들이 영향을 미친 것인지  
현재 상황에서 구할 수 있는 데이터로 추가 분석을 진행하고자 한다.

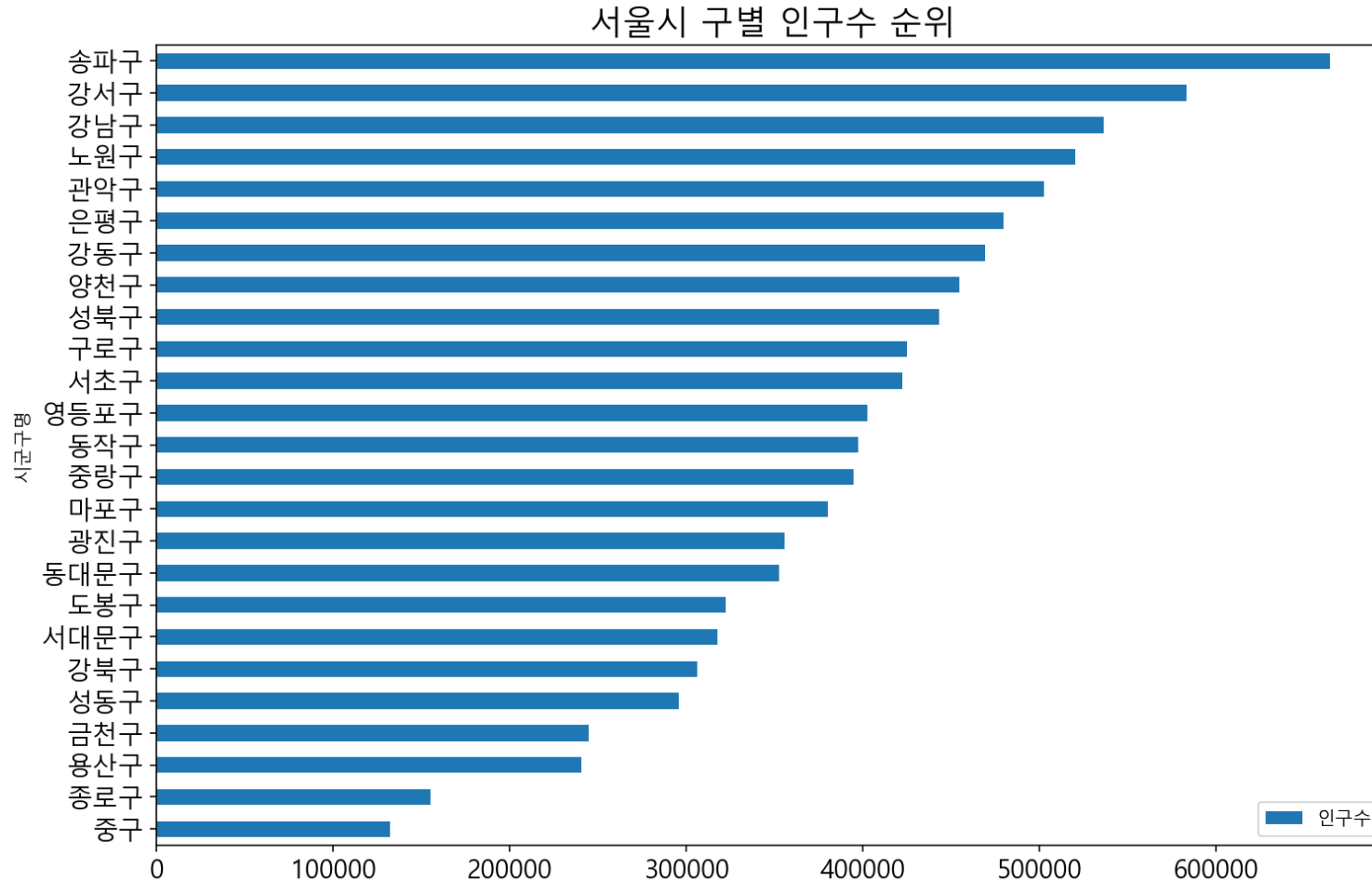
“지역 인구와 코로나 확진자수는 상관 있을까?”

# Story #04

지역별 주민등록인구수와 코로나 확진자수와의 연관관계

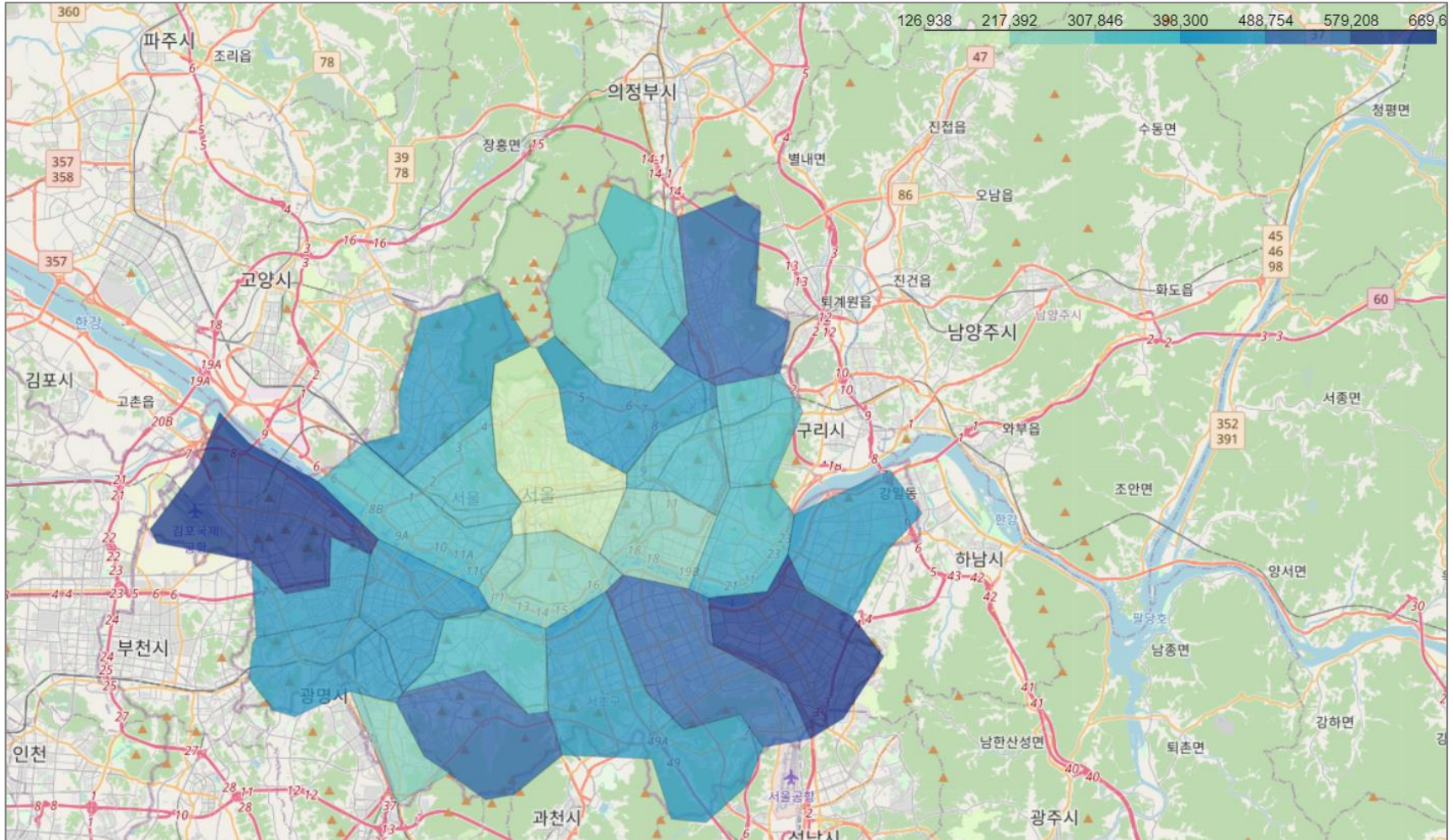
## 〈서울시〉 구별 인구 순위

주요 거주지인 〈송파구〉, 〈강서구〉, 〈강남구〉 순으로 상주 인구가 많고  
“회사”가 많은 〈종로구〉, 〈중구〉는 “출퇴근 인구”가 많은 대신 “거주 인구”는 적은 것을 확인할 수 있다.



## 서울시구별 인구 밀도

인구가 많은 <송파구>, <강서구>, <강남구> 등을 지도상에서 쉽게 찾을 수 있다.  
**그런데 <인구수>와 <코로나 발생>이 관계가 있을까? 데이터 융합 분석으로 알아보았다.**

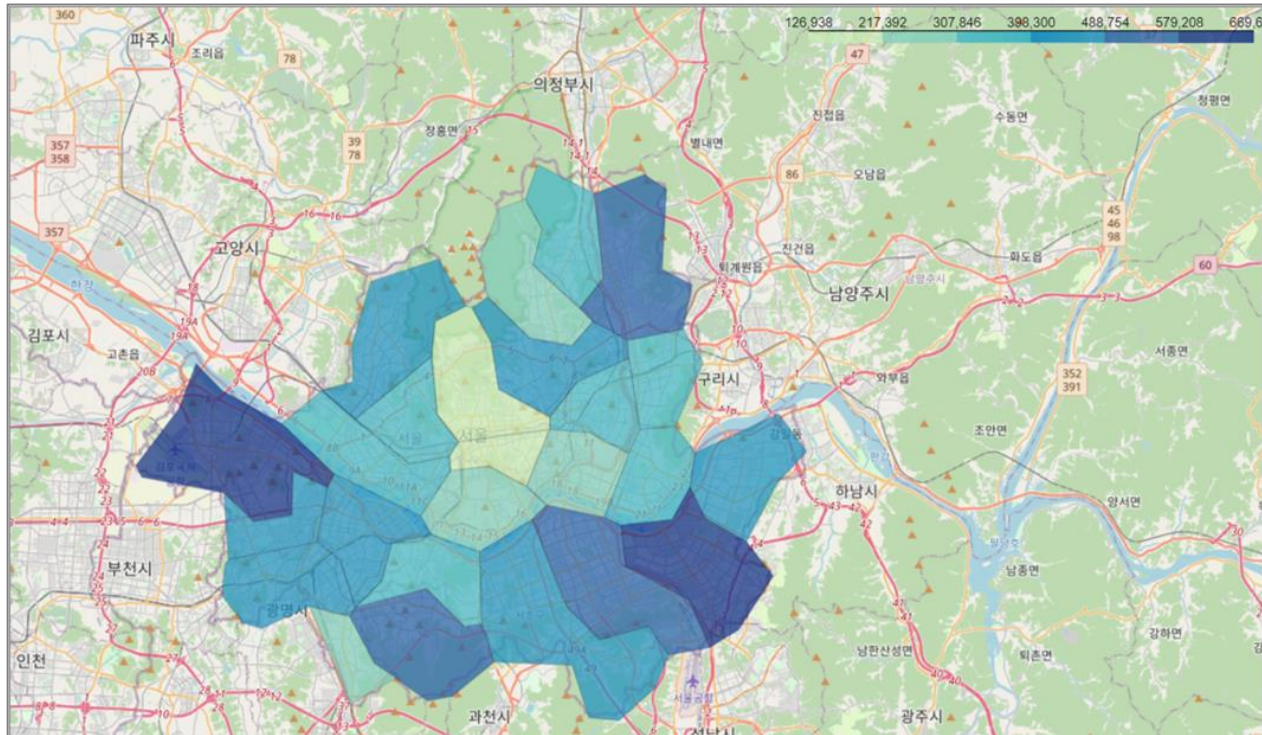
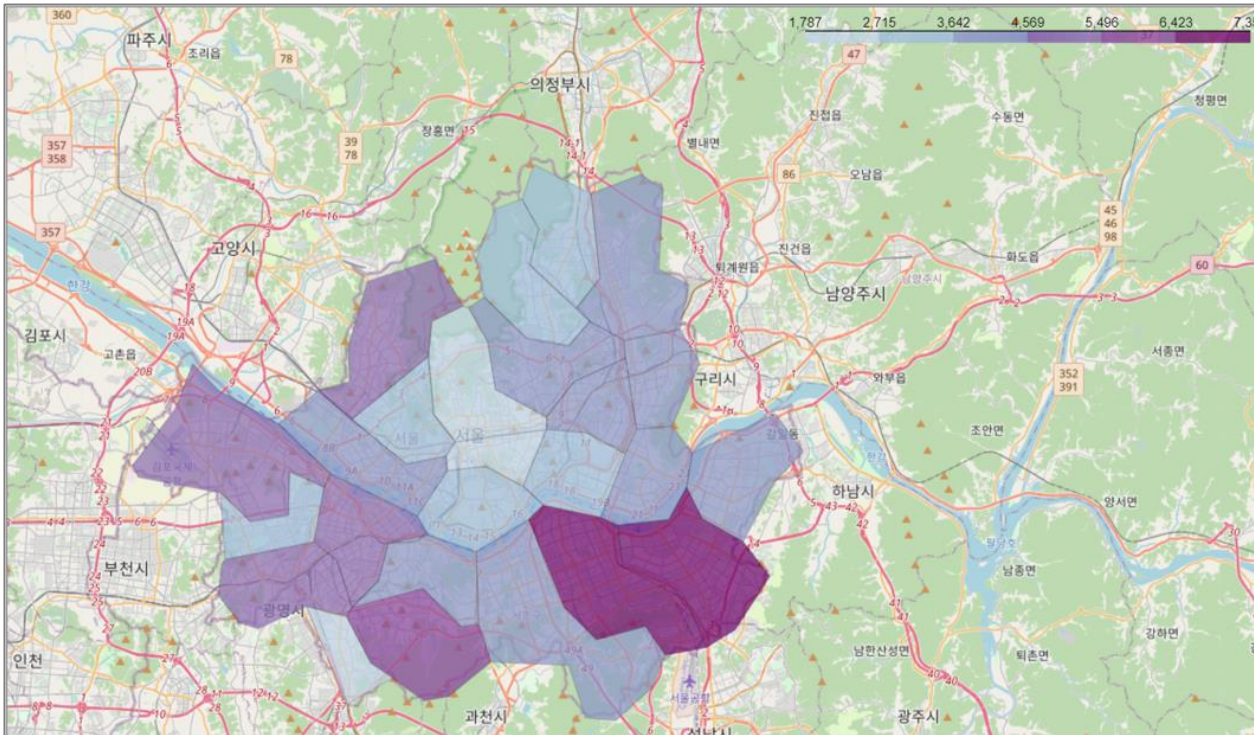


# 서울시 구별 <코로나 확진자수>와 <인구수> 밀도 비교

코로나 발생이 많은 <강남구>, <송파구>, <강서구>가 인구도 많은 것을 볼 수 있다.  
<종로구>는 코로나 환자도 적고 인구도 적은 등 전반적으로 두 데이터는 비슷해 보인다.  
<코로나 환자수>와 <인구>는 어떤 관계가 있는지 상관분석을 통해 “정확한 수치”를 살펴보자.

[ 서울시 코로나 확진자 분포 ]

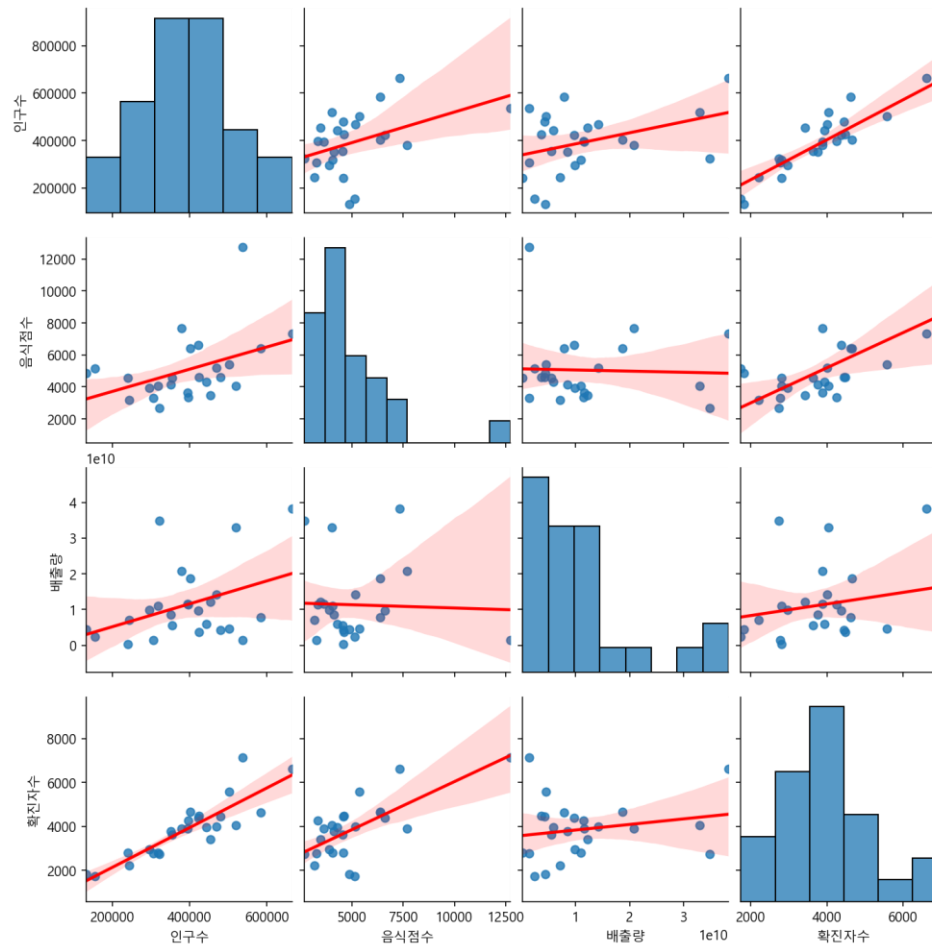
[ 서울시 인구 분포 ]



## <인구>와 <음식점 수>, <음식물 쓰레기>, <코로나 발생>의 관계

회귀선의 각도가 클수록 상관관계가 크다. 히스토그램, 산포도, 회귀선을 동시에 그려 본 결과 <인구>가 많아질 때 <음식점 수>, <음식물 쓰레기>, <코로나 발생>이 동시에 증가하는 것을 확인하였다.

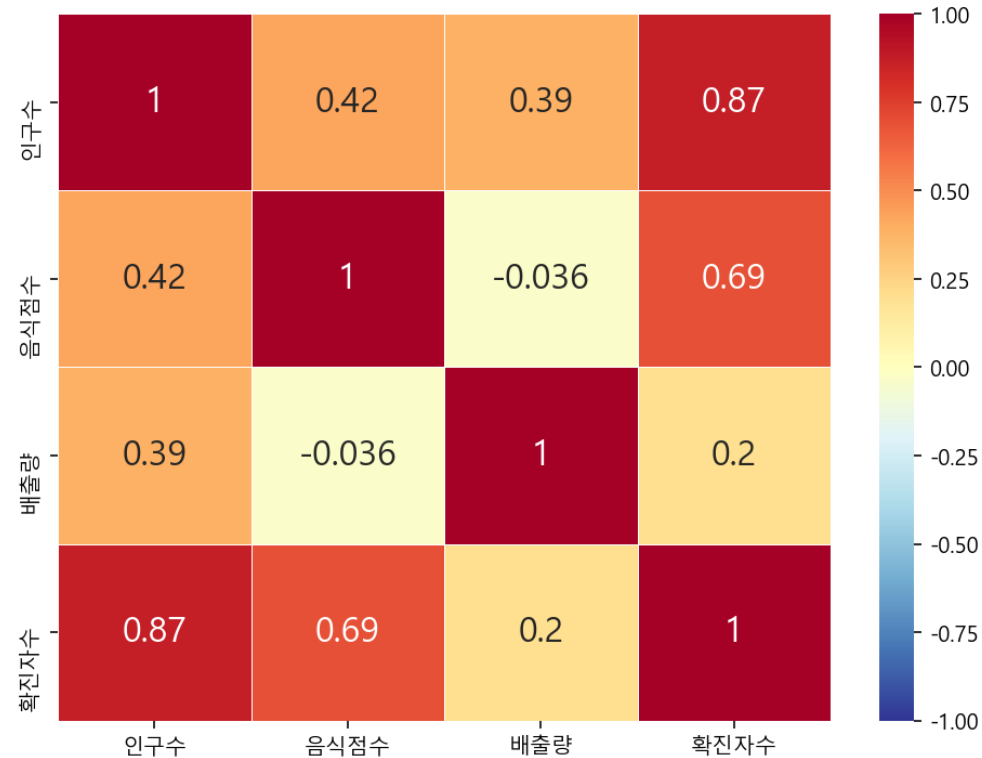
구별 인구, 음식점 수, 음식물 쓰레기 양, 코로나 확진자수의 관계



## 〈인구〉와 〈음식점 수〉, 〈음식물 쓰레기〉, 〈코로나 발생〉의 상관분석 - heatmap

〈인구〉가 많아질수록 자연스럽게 〈음식점 수〉, 〈음식물 쓰레기〉도 함께 늘어난다.  
특히 〈인구〉와 〈확진자 수〉의 상관계수는 0.87로, 〈음식점 수〉와 〈확진자 수〉의 0.69보다 크다.  
〈코로나 확진자 수〉에 가장 큰 영향을 미치는 것은 〈인구 수〉임을 확인하였다.

인구수, 음식점 수, 음식물 쓰레기 양, 코로나 확진자수의 상관관계





## 분석의 결론 :

이전 분석에서 <음식점 수>와 <코로나 확진자 수>의 상관성을 확인하였으나  
<상관관계>와 <인과관계>는 구별해야 하므로  
이것으로 <음식점>이 <코로나 발생>의 근원지라고 결론 내릴 수 없다.

상관관계는 증명되었으나, 인과관계를 확인하기 위해  
<코로나 확진>에 다른 어떤 요소들이 영향을 미친 것인지  
여러 데이터를 융합해 추가 분석을 진행한 결과

<코로나 발생>에 가장 큰 영향을 주는 것은 <인구 수>라고 판단되었으며

인구와 함께 <음식점 수>와 <음식물 쓰레기>도 많아진 효과로  
<음식점 수>에 따라 <코로나 발생>이 늘어나는 것처럼 “보여진” 것이라 결론지었다.

## 1. 추가로 <인구 밀도> 고려

단순히 인구가 많아서 그에 비례해 코로나 환자가 많은 것인지, 지역의 면적까지 고려한 <면적 대비 인구 밀도>가 높음으로 인해 접촉이 많아 코로나가 많이 발생한 것인지 확인할 필요가 있다.

## 2. 데이터 수집의 기준과 시기 일치

여러 데이터를 통합하려면 동일한 기준(수집 시기 등)에 의해 수집된 데이터라야 한다. 각 데이터의 기준과 수집시기가 다르므로 (ex:상권-2020년, 코로나-2021년 실시간) 이번 분석의 의미를 과도하게 확장 해석하는 것은 경계해야 한다.

## 3. <인과관계>의 심화 탐색 필요

“실무적으로” 매우 강한 상관관계가 있을 경우 “인과관계를 추단”한다.  
인구 수와 코로나 발생의 상관관계가 0.87로서 매우 강하여 인과관계를 추단하였으나 잘 설계한 추가 분석에 의해 “추단된 인과관계가 없음이 증명”될 수도 있다.

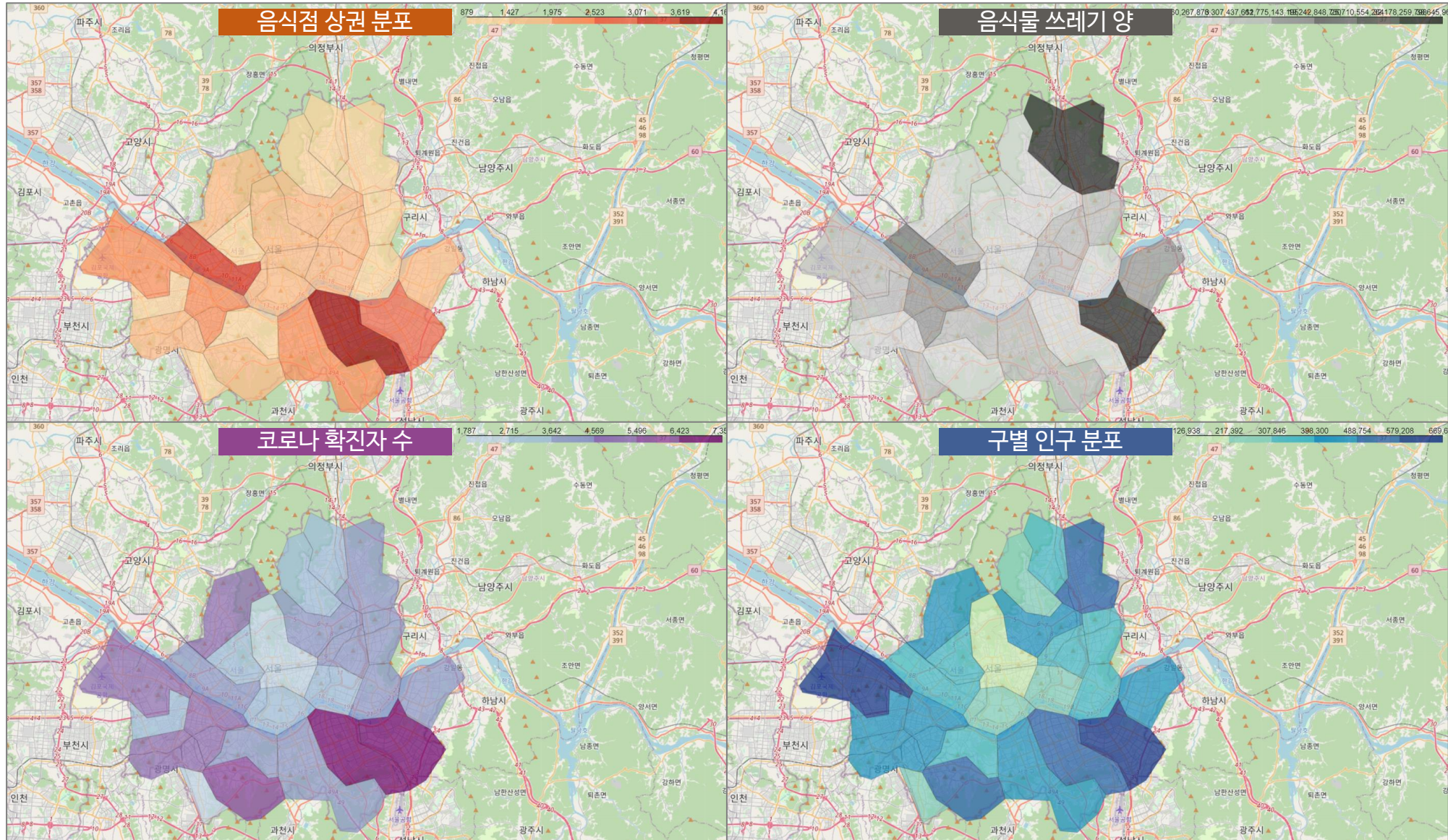
## 데이터 스토리 총정리 및 향후 과제:

데이터를 통해 코로나와 쓰레기 발생과의 관계를 새로이 규명하고  
코로나 시국을 겪는 소상공인의 어려움 등 당면한 문제를  
해결하는데 도움을 주려는 의도에서 이 스토리를 기획하게 되었다.

“바로 현장에 적용가능한 전략”을 도출하기에는  
데이터의 상황 등 여러 한계와 마주쳤으나  
최선을 다해 다양한 시도를 통해 상황을 다각도로 파악하고  
흥미롭고 발전 가능성 있는 인사이트를 이끌어낸 과정이 의미 있었다.

〈코로나〉와 관련된 4편의 데이터 스토리 시리즈를 마무리하며  
향후, 현재의 데이터를 업데이트하고 다양한 추가 데이터를 구하여  
새로운 목적과 그에 따른 분석을 고민하는 것이 다음 과제임을 밝힌다.

# fin. 서울시데이터 요약



END

음식물쓰레기  
전용수거용기

잠깐!  
노란전용봉투 인가요?  
미사용시  
5만원 과태료 부과됨  
(청소행정과)

잠깐!  
노란전용봉투 인가요?  
미사용시  
5만원 과태료 부과됨  
(청소행정과)

