

2021 통계데이터 분석활용대회

**서울특별시**  
**‘여성 1인가구 안심홈세트’ 정책**  
**분배적절성 분석**



# Index :

---

1. 서론	분석배경 및 목적
2. 본론 – R 분석	데이터 선정 분석 코딩북 데이터 전처리 데이터 분석 각 변수의 분포 시각화
3. 본론 – Python 분석	Kmeans 군집화를 통한 군집유형 분석 분석결과 및 해석 정책분석 및 평가
4. 결론 - 분석활용전략	기대효과 방향제시

# 1. 서론

여성안심홈세트란? 여성 1인 가구의 안전한 주거생활을 지원하기 위해 창문잠금장치, 긴급벨 등을 지원하는 서비스

모르는 사람이 있어서 위에 내용 추가필요

## 분석배경 및 목적

### 1인가구의 증가로 인한 연구 필요성 대두

서울시의 1인가구 수는 2000년 50만명에서부터 2015년 100만명을 넘기고, 2019년에는 약 130만명에 이를만큼 꾸준히 증가하고 있다. 또한 주된 가구 구성방식이 핵가족 단위에서 1인가구 단위로 바뀌어가고 있다. 학계에서도 많은 연구가 이루어지고 있기에 지자체에서도 이에 주목하고, 구체적 정책을 제시해야 한다. 서울특별시 1인가구 대책정책 연구(서울시, 2014)에서는 1인가구 정책을 '주거주택 부문', '사회적안전망 부문', '공유사회 부문' 세 가지로 나누고 있는데, 본 연구에서는 '사회적안전망 부문'에 주목해 보았다.



### 여성 1인가구에 대한 정책지원 미비

1인가구 밀집지역은 비밀집지역에 비해 2~3배 높은 범죄발생율을 보인다. 또한 주요 5대 강력범죄가 주거지에서 가장 많이 발생하며, 여성 피해자의 수도 꾸준히 증가세를 보이고 있다. 빅카인즈를 통해 2019년 5월 '신림동 강간미수 사건' 전후로 '1인가구 and 여성' 키워드로 연관어 분석을 해보았을 때, 주거공간에서의 범죄불안 및 지원정책에 대한 수요가 증가했음을 알 수 있었다. 따라서 본 연구에서는 정책지원 필요성이 크게 증가했음을 인지하고, 이에 대해 구체적인 지원이 이루어져야 하기에 여성 1인가구에 대한 정책을 찾아 분석했다.



### 서울시 안심홈세트 지원 서비스 정책 분석

2019년부터 전국 지자체에서 여성 1인가구를 대상으로 한 안심지원사업을 시범적으로 운영하기 시작했고, 2021년 6월부터 서울시에서는 사업을 더 많은 자치구로 확대했다. 정책에 대한 만족도는 높았으나, 여성 1인가구 수에 비해 지원은 턱없이 부족했다. 이는 시범운영이 이루어지던 시기였기에 차차 지원범위를 확대해나가야 한다. 따라서 본 연구자들은 서울시 안심홈세트 지원정책을 활성화하기 위해 R, Python 등을 이용해 서울시의 행정구별 여러 변수를 분석하고, 시각화하여 지자체에서 '우선지원분배 행정구 선택방법'을 제안하고자 한다.



## 2. 본론 – R 분석



### 데이터 선정

데이터명	출처
1인가구 변화 – 서울시(2019)	통계지리정보서비스
범죄발생건수(2019)	
서울시 1인가구(연령별) 통계(2019)	서울시 열린데이터광장
서울시 5대범죄발생현황통계(2019)	
서울시 도로시설물통계(2020)	
서울특별시 (안심이)CCTV설치현황(2021)	
서울시 인구밀도(동별) 통계(2020)	
서울시 시내주요기관(경찰, 소방관서) 통계(2019)	
전국여성안심지킴이집 표준데이터(2021)	공공데이터포털



### 분석 코딩북

변수명	개요	척도
지역	서울시 자치구	명목 (character)
구별면적	서울시 자치구 면적	연속 (number)
여성1인가구수	서울시 여성1인가구수	연속 (number)
면적당여성1인가구수	자치구 면적당 서울시 여성1인가구수	연속 (number)
면적당범죄발생건수	자치구 면적당 서울시 범죄발생건수	연속 (number)
면적당가로등수	자치구 면적당 서울시 가로등수	연속 (number)
면적당CCTV수	자치구 면적당 서울시 CCTV수	연속 (number)
면적당치안센터수	자치구 면적당 서울시 치안센터수	연속 (number)
면적당여성안심지킴이집수	자치구 면적당 서울시 여성안심지킴이집수	연속 (number)

## 2. 본론 – R 분석

### 데이터 전처리

#### 1) 결측치 처리: 결측치 없음

#### 2) 데이터 변환 및 파생변수 생성

원본 데이터에서 필요한 열만 추출하여 변수 생성 후, 구별면적값을 기반으로 '면적당 변수의 개수'를 파생변수로 생성하였다.

'구별면적', '여성1인가구수',  
'면적당여성1인가구수',  
'면적당범죄발생건수',  
'면적당가로등수', '면적당CCTV수',  
'면적당치안센터수',  
'면적당여성안심지킴이집수'를 만들어  
result2 데이터셋을 생성하였고, 이를  
이용해 데이터 분석을 실시했다.

```
> result2 <- result[, c('지역', '구별면적', '여성1인가구수', '면적당여성1인가구수',  
+                        '면적당범죄발생건수', '면적당가로등수', '면적당CCTV수',  
+                        '면적당치안센터수', '면적당여성안심지킴이집수')]  
> str(result2)  
'data.frame': 26 obs. of 9 variables:  
 $ 지역                : chr  "합계" "종로구" "중구" "용산구" ...  
 $ 구별면적            : num  605.23 23.91 9.96 21.87 16.86 ...  
 $ 여성1인가구수       : num  688226 12729 11401 18944 20963 ...  
 $ 면적당여성1인가구수 : num  1137 532 1145 866 1243 ...  
 $ 면적당범죄발생건수  : num  171 161 434 151 149 ...  
 $ 면적당가로등수      : num  469 369 858 358 372 ...  
 $ 면적당CCTV수        : num  111.3 55.6 106.8 68.5 66.4 ...  
 $ 면적당치안센터수    : num  0.729 1.046 2.008 0.594 0.949 ...  
 $ 면적당여성안심지킴이집수: num  1.47 1.25 3.71 1.55 0.89 ...
```

## 2. 본론 – R 분석



### 데이터 분석

#### 전체 변수의 요약통계량

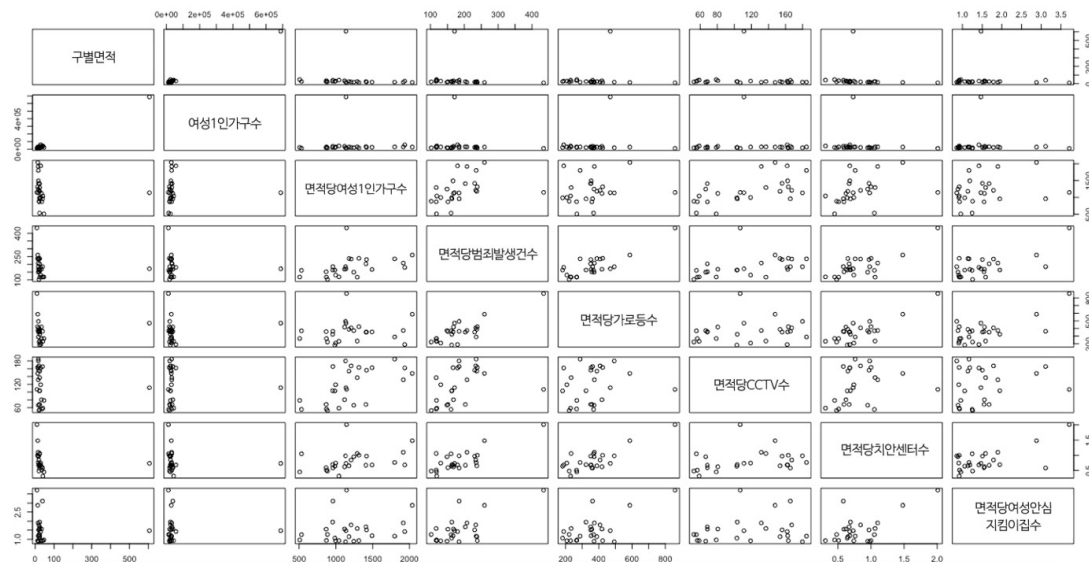
```
> summary(result2)
```

지역	구별면적	여성1인가구수	면적당여성1인가구수	면적당범죄발생건수
Length:26	Min. : 9.96	Min. : 11401	Min. : 508.2	Min. :102.2
Class :character	1st Qu.: 17.46	1st Qu.: 21541	1st Qu.: 965.9	1st Qu.:149.6
Mode :character	Median : 23.73	Median : 27602	Median :1140.9	Median :169.8
	Mean : 46.56	Mean : 52940	Mean :1206.3	Mean :185.7
	3rd Qu.: 29.68	3rd Qu.: 31611	3rd Qu.:1383.8	3rd Qu.:225.0
	Max. :605.23	Max. :688226	Max. :2037.7	Max. :434.4

면적당가로등수	면적당CCTV수	면적당치안센터수	면적당여성안심지킴이집수
Min. :182.2	Min. : 52.78	Min. :0.3137	Min. :0.8616
1st Qu.:266.8	1st Qu.: 71.12	1st Qu.:0.6239	1st Qu.:0.9832
Median :356.3	Median :115.17	Median :0.7100	Median :1.3397
Mean :359.6	Mean :116.96	Mean :0.8161	Mean :1.5230
3rd Qu.:402.7	3rd Qu.:160.42	3rd Qu.:0.9756	3rd Qu.:1.6545
Max. :858.4	Max. :184.47	Max. :2.0080	Max. :3.7149

#### 전체 변수 간 산점도 매트릭스



#### 전체 변수 간 상관관계 결과

```
> rcor <- cor(result2[, 2:9])
```

```
> rcor
```

	구별면적	여성1인가구수	면적당여성1인가구수	면적당범죄발생건수	면적당가로등수	면적당CCTV수	면적당치안센터수	면적당여성안심지킴이집수
구별면적	1.00000000	0.997044911	-0.075078726	-0.09028179	0.1161873	-0.05389536	-0.1062488	-0.03375407
여성1인가구수	0.99704491	1.000000000	-0.003388931	-0.06219662	0.1249303	-0.01246680	-0.0819464	-0.01943884
면적당여성1인가구수	-0.07507873	-0.003388931	1.000000000	0.38522815	0.1474372	0.58156455	0.3489519	0.22372659
면적당범죄발생건수	-0.09028179	-0.062196617	0.385228146	1.000000000	0.8237299	0.40920285	0.7872628	0.66431591
면적당가로등수	0.11618728	0.124930264	0.147437237	0.82372988	1.000000000	0.16584021	0.7945510	0.66370108
면적당CCTV수	-0.05389536	-0.012466796	0.581564547	0.40920285	0.1658402	1.000000000	0.2623354	0.17090479
면적당치안센터수	-0.10624882	-0.081946400	0.348951894	0.78726279	0.7945510	0.26233544	1.000000000	0.61703106
면적당여성안심지킴이집수	-0.03375407	-0.019438840	0.223726588	0.66431591	0.6637011	0.17090479	0.6170311	1.000000000

## 2. 본론 – R 분석

### 데이터 분석

#### 면적당범죄발생건수에 대한 피어슨 상관분석 결과

상관분석 변수	cor	p-value	t값	유의여부
면적당여성1인가구수	0.3852281	0.05196	2.0451	X
면적당가로등수	0.8237299	2.342e-07	7.1174	O
면적당CCTV수	0.4092029	0.03791	2.197	O
면적당치안센터수	0.7872628	1.827e-06	6.2547	O
면적당여성안심지킴이집수	0.6643159	0.0002147	4.3541	O

R을 통한 면적당범죄발생건수와 다른 변수들 간의 상관분석 결과, 유의하지 않은 면적당여성1인가구수를 제외하고  
면적당가로등수, 면적당치안센터수, 면적당여성안심지킴이집수, 면적당CCTV수 순으로  
면적당범죄발생건수와 유의한 상관관계를 가지고 있는 것으로 나타났다.



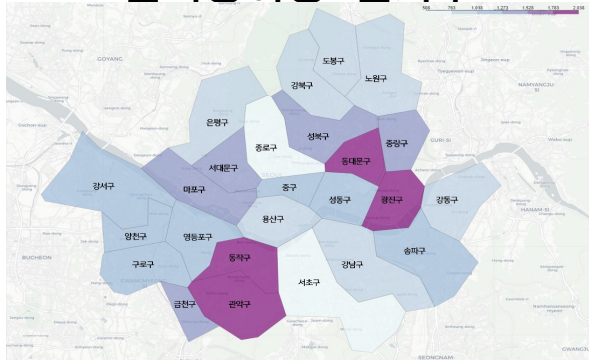
## 2. 본론 - R 분석



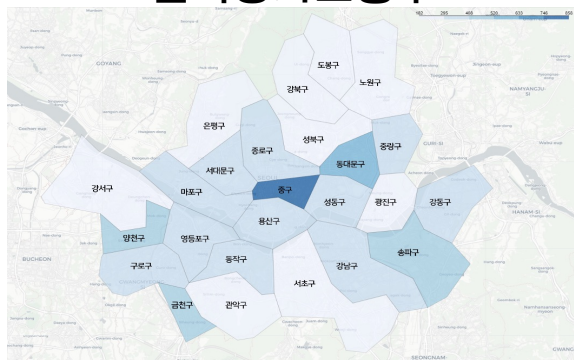
### 각 변수의 분포 시각화

가로등, CCTV의 좌표데이터와 여성1인가구수, 범죄발생건수 등을 Python의 Folium을 이용해 시각화하였다. 지도에 표시된 색을 단계별로 인지하고, 등급을 설정하여 각 '자치구별 분포등급'을 부여했다.

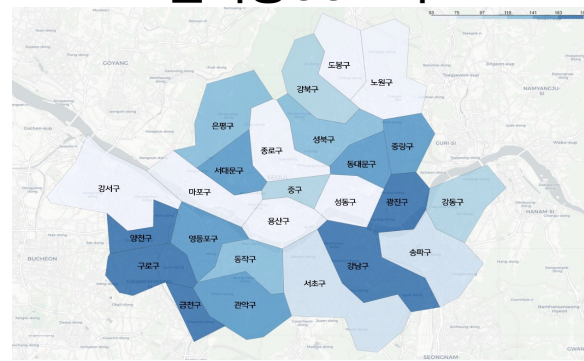
면적당여성1인가구



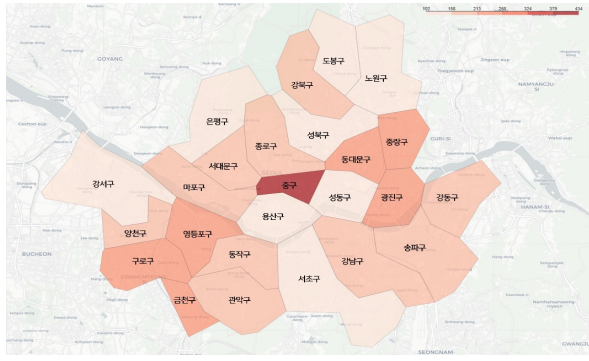
면적당가로등수



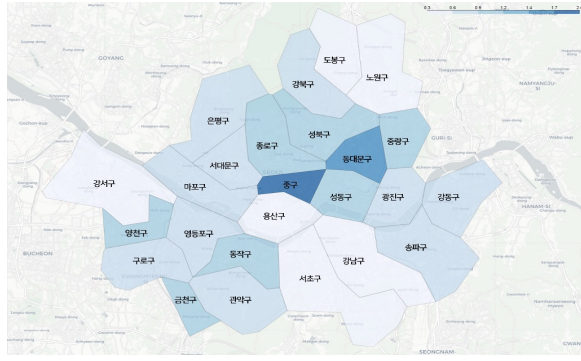
면적당CCTV수



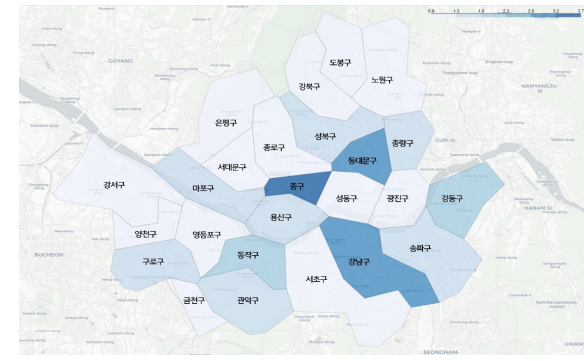
면적당범죄발생건수



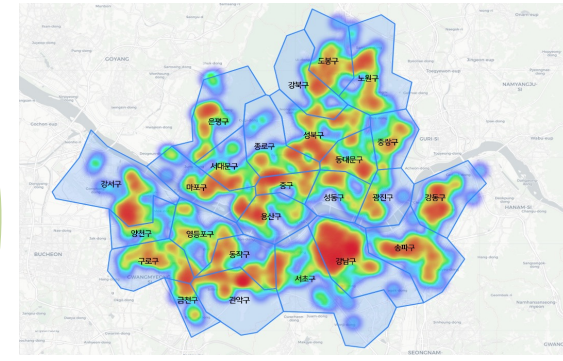
면적당치안센터수



면적당여성안심지킴이집



여성안심지킴이집분포히트맵



위 슬라이드와 아래 슬라이드는 연관성이 있는 슬라이드여서

위 슬라이드에서 \*변수, correlation, 유의여부만 활용하고 이를 통해 시각화를 아래와 같이 했다~ 라는 흐름이 보이게 필요

아래 그림 중 예시는 2가지 정도만 선택적으로 활용



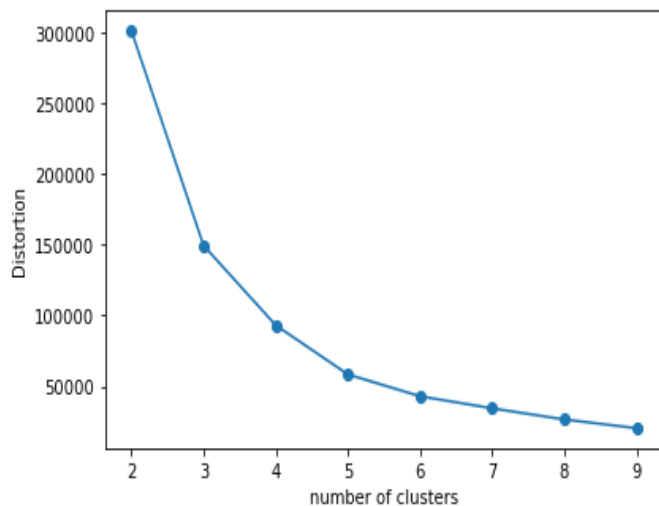
### 3. 본론 – Python 분석



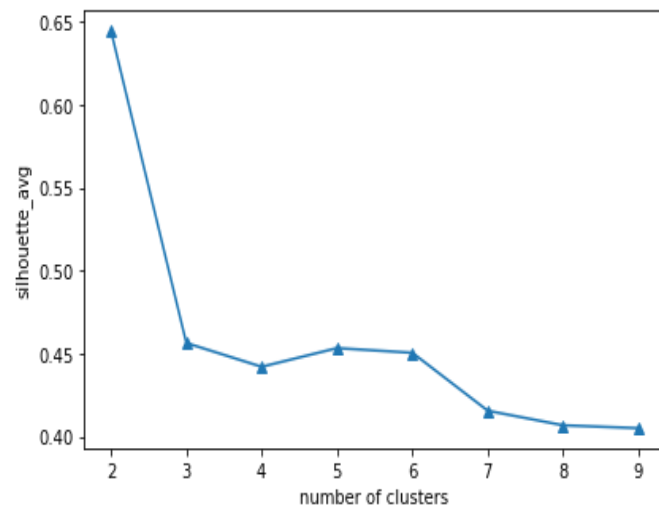
#### Kmeans 군집화를 통한 군집유형 분석

최적의 클러스터를 찾기 위한 기법 사용 및 최적클러스터의 산점도

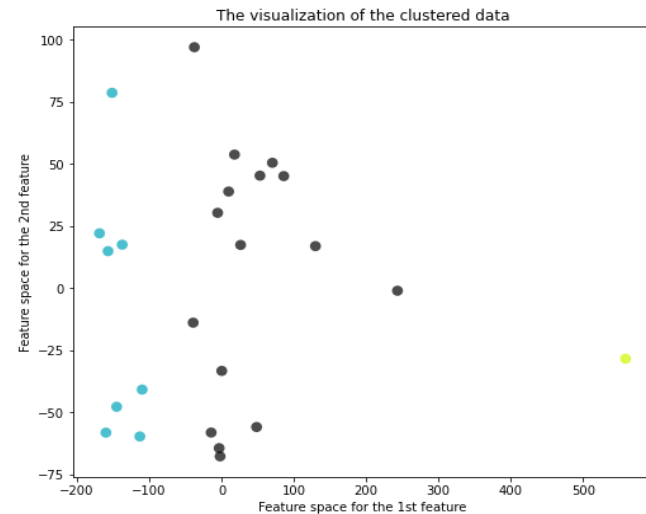
Elbow Method



Silhouette Method



n clusters = 3 산점도



Elbow Method와 Silhouette Method를 통하여 최적의 클러스터를 '3'으로 선정했다.

면적당범죄발생건수와 상관관계를 시각화하고자 PCA를 통해 면적당범죄발생건수와 그 외 피처를 2차원으로 각각 축소하였다.

군집 유형은 아래와 같으며, 군집만큼 '클러스터 가중치'로 적용하였다.

군집 유형	군집0(16개)	군집1(1개)	군집2(8개)
자치구	종로구, 용산구, 성동구, 광진구, 동대문구, 중랑구, 서대문구, 마포구, 양천구, 구로구, 금천구, 영등포구, 동작구, 강남구, 송파구, 강동구	중구	성북구, 강북구, 도봉구, 노원구, 은평구, 강서구, 관악구, 서초구

### 3. 본론 – Python 분석



#### 분석 결과 및 해석 (1/2)

- 종합치안범죄지수 및 여성1인가구 위험노출도 도출  
및 연도별 각 자치구 정책시행현황(2019. ~ 2021.)

$$* \text{종합치안범죄지수: } R = (C + K) / \frac{1}{n} \sum_{a=1}^n S^* w$$

S: 치안요소분포등급, w: 상관계수, C: 면적당범죄건수 분포등급,  
k: 클러스터 가중치, n: 치안요소 갯수

$$* \text{여성1인가구 위험노출도: } D = \log(R^* P)^2$$

P: 여성1인가구 분포등급

구 분	종합치안범죄지수	여성1인가구 위험노출도	위험노출 순위	2019년 지원여부	2020년 지원여부	2021년 지원여부
종로구	2.361042794	0.7462077175	23	-	-	-
중구	2.150711079	1.619406654	11	-	-	○
용산구	1.91736171	1.167468092	16	-	-	○
성동구	1.574028529	1.348267709	13	-	○	-(안심점포)
광진구	2.899811512	2.48104204	1	-	○	○
동대문구	1.270012621	1.763918574	9	-	-	○
종랑구	2.166964624	1.875823626	4	-	-	○
성북구	0.6503324825	0.8303908756	22	-	-	○
강북구	1.864801865	1.143325381	17	-	○	○
도봉구	1.490090896	0.9484855137	19	-	-	-
노원구	1.490090896	0.9484855137	19	-	○	○
은평구	0.8512087164	0.4621321156	24	-	○	-
서대문구	2.022892399	1.816065548	6	-	○	○

### 3. 본론 – Python 분석



#### 분석 결과 및 해석 (2/2)

- 종합치안범죄지수 및 여성1인가구 위험노출도 도출  
및 연도별 각 자치구 정책시행현황(2019. ~ 2021.)

$$* \text{종합치안범죄지수: } R = (C + K) / \frac{1}{n} \sum_{a=1}^n S^* w$$

S: 치안요소분포등급, w: 상관계수, C: 면적당범죄건수 분포등급,  
k: 클러스터 가중치, n: 치안요소 갯수

$$* \text{여성1인가구 위험노출도: } D = \log(R * P)^2$$

P: 여성1인가구 분포등급

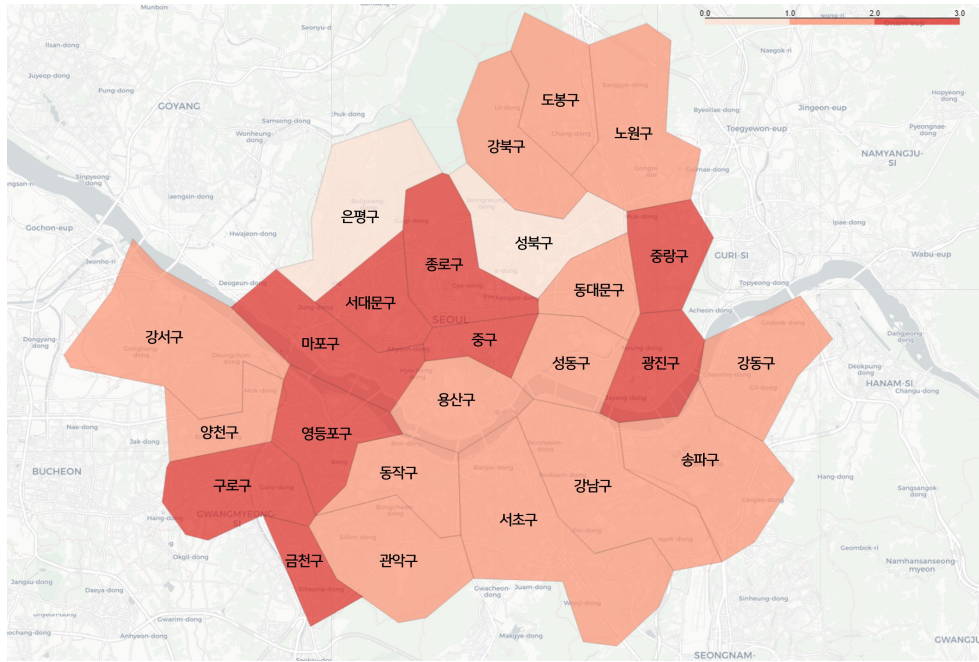
구 분	종합치안범죄지수	여성1인가구 위험노출도	위험노출 순위	2019년 지원여부	2020년 지원여부	2021년 지원여부
마포구	2.419549964	1.971589172	2	-	-	○
양천구	1.509016373	1.311630413	14	○	○	-
강서구	1.490090896	1.300668032	15	-	○	○
구로구	2.283887176	1.671591801	10	-	-	-
금천구	2.01202183	1.81138536	8	-	-	○
영등포구	2.697189865	1.816065548	6	-	-	-
동작구	1.570947936	1.948626085	3	-	○	○
관악구	1.385833319	1.839724498	5	○	○	○
서초구	1.292991983	0.2231916645	25	-	-	○
강남구	1.461400753	0.9315986447	21	-	-	-
송파구	1.937828018	1.528872971	12	-	-	-
강동구	1.862688791	1.142340594	18	-	○	○
합계	-	-	-	2개구	11개구	16개구

### 3. 본론 – Python 분석

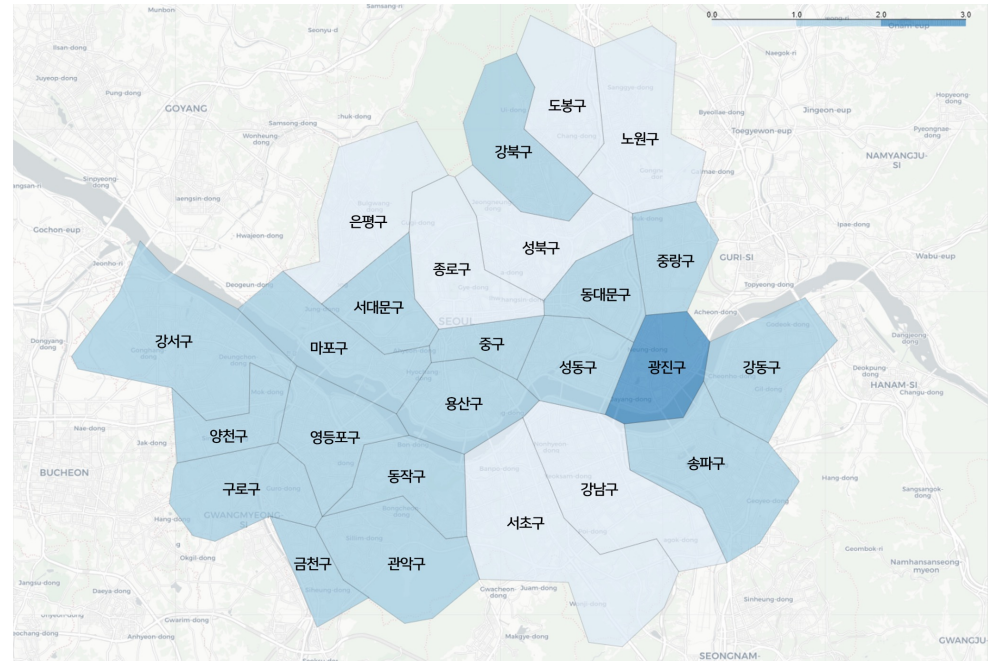


#### 분석 결과 및 해석 (시각화)

종합치안범죄지수 시각화



여성1인가구 위험노출도 지도 시각화



마찬가지로 군집분석을 통해 아래와 같이 나왔다 흐름이 보이게 간단하게만

앞에는 언급을 안했는데 R로는 상관관계(correlation)로 분석해본거고 Python로는 군집유형분석을 한거라서 둘이 구분되게만 디자인 필요

### 3. 본론 – Python 분석



#### 정책 분석 및 평가

- 여성1인가구 위험노출수준별 지역구분 및 정책지원여부 반영

구 분	고위험노출 지역	저위험노출 지역
정책지원 2019년 (2개구)	관악구 등 1개구	양천구 등 1개구
정책미지원 2019년 (23개구)	광진구, 마포구, 동작구, 중랑구, 서대문구, 금천구, 영등포구, 동대문구, 구로구, 중구, 송파구 등 11개구	성동구, 강서구, 용산구, 강동구, 강북구, 도봉구, 노원구, 성북구, 종로구, 은평구, 강남구, 서초구 등 12개구
정책지원 2020년 (11개구)	광진구, 서대문구, 동작구, 관악구 등 4개구	노원구, 강북구, 은평구, 강서구, 성동구, 양천구, 강동구 등 7개구
정책미지원 2020년 (14개구)	구로구, 중랑구, 금천구, 중구, 동대문구 마포구, 송파구, 영등포구 등 8개구	성북구, 도봉구, 서초구, 강남구, 용산구, 종로구 등 6개구
정책지원 2021년 (16개구)	마포구, 광진구, 서대문구, 중랑구, 금천구, 동작구, 중구, 동대문구, 관악구 등 9개구	성북구, 강북구, 노원구, 강서구, 서초구, 용산구, 강동구 등 7개구
정책미지원 2021년 (9개구)	영등포구, 구로구, 송파구 등 3개구	종로구, 성동구, 강남구, 도봉구, 은평구, 양천구 등 6개구



### 3. 본론 – Python 분석



#### 정책 분석 및 평가

- 서울시 정책평가: 지원분배 적절성 \* 지원분배적절도 = 서비스지원구 / 고위험지역구 \* 100

구 분	2019년	2020년	2021년
지원분배적절도	8.33%	33.33%	75.00%

- 우선지원자치구 도출: 정책지원이 다소 부족한 자치구

구 분	치안지수	여성1인가구 위험노출도	2019년 지원여부	2020년 지원여부	2021년 지원여부
영등포구	2.697189865	1.816065548	-	-	-
구로구	2.283887176	1.671591801	-	-	-
송파구	1.937828018	1.528872971	-	-	-

## 4. 결론 - 분석활용전략

### 기대효과

#### 행정공학적 접근

정책입안 시 여타 변수가 아닌, 정책지원 대상자의 생명과 재산을 우선 보호하기 위한 실효성있는 정책을 시행하기 위한 적절한 분석을 적용하였다.

#### 추가 연계정책 기대

해당 모델을 통해 여성1인가구를 지원중인 관련 여성안심정책 '여성안심특별시 3.0'(안심귀가스카우트, 안심택배, 안심차킴이집 등)에도 적용 가능할 것이라 기대한다.

#### 모델 확장성 기대

해당 모델은 종합범죄치안지수를 측정한 후 취약 계층의 인구분포를 반영하였기 때문에 정책 지원시 보다 정밀하게 지원지역을 추릴 수 있다. 이 모델은 독거노인, 아동 등 여타 취약 계층에도 적용이 기대되며, 행정단위라는 점에서 타도 시에도 적용할 수 있을 것이라 기대한다.

## 4. 결론 - 분석활용전략

### 방향제시

#### Column 추가 확보

Column을 추가적으로 확보하고, 다양한 변수를 통해 상관관계를 파악한다면 더욱 정교한 모델을 구축할 수 있을 것으로 기대된다.

#### Row 추가 확보

각 행정기관의 데이터가 행정동 단위가 아닌 25개의 자치구 수로만 제공된다는 점에서 절대량이 부족하다. 분류 등 지도학습에는 부적절하다고 볼 수 있다. 다만 전국단위 혹은 행정동 단위 데이터를 확보할 수 있다면, 더욱 다양한 면이나 데이터의 밀도 등에 있어 유의미한 분석을 기대할 수 있을 것이다.

#### 정책관련 데이터 추가확보

서울시 여성안심정책과로부터 정책현황 데이터를 직접 요청했으나, 보안 등의 문제로 요구 수준의 구체적인 내용을 얻을 수 없었다. 해당 과와 직접 연계하여 분석할 수 있다면 더욱 정밀하고 유의미한 분석을 기대할 수 있을 것이다.

## 정책 의사결정과정 의견피력 및 반영

전문가 심사결과를 근거로 서울시의회 및 서울시 여성안심정책과에 적극 검토요청하여  
분석결과가 예산편성, 의식공유, 분배결정 등 정책결정 과정에 반영하도록 행동을 취할 것이다.  
이를 통해 시민으로서의 행동을 실천하고,  
데이터 기반의 의사결정 문화 확산이라는 대회 기획의도와 통계분석의 본질을 달성할 것이다.

