

# 도시 환경 및 인구 데이터를 활용한 지역별 침수 요인 분석과 대비책 제안

Factor analysis of Flood areas and proposal of provision by utilizing Urban Environmental and Demographic data

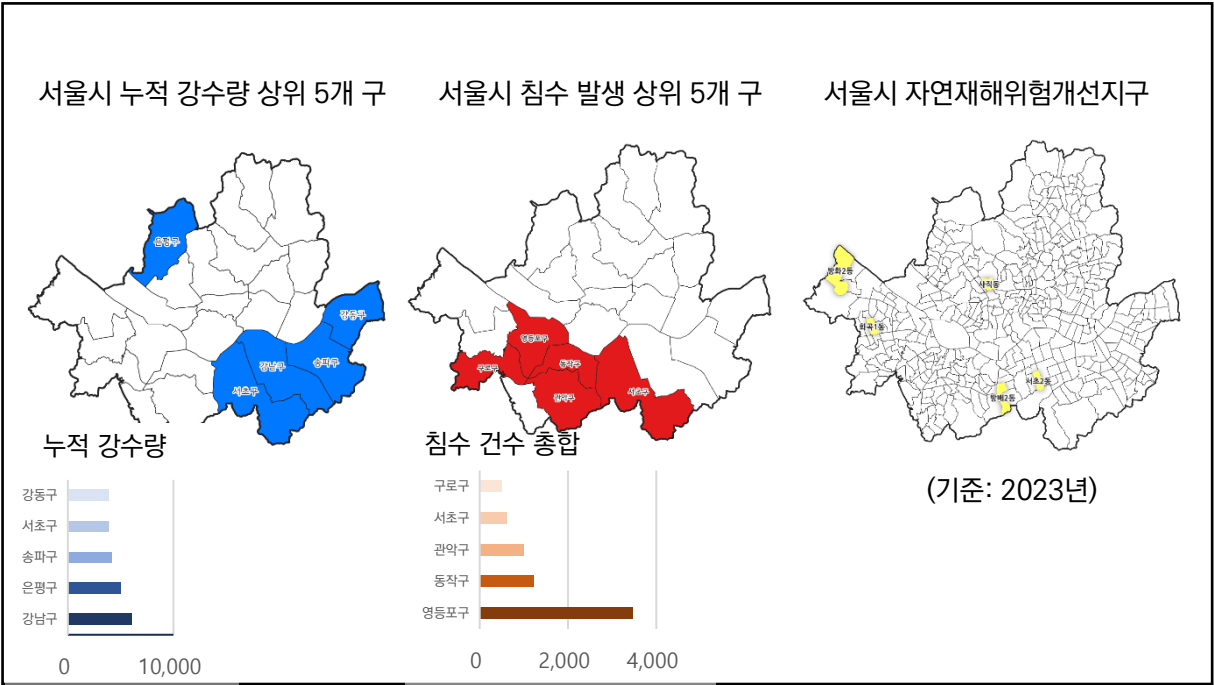
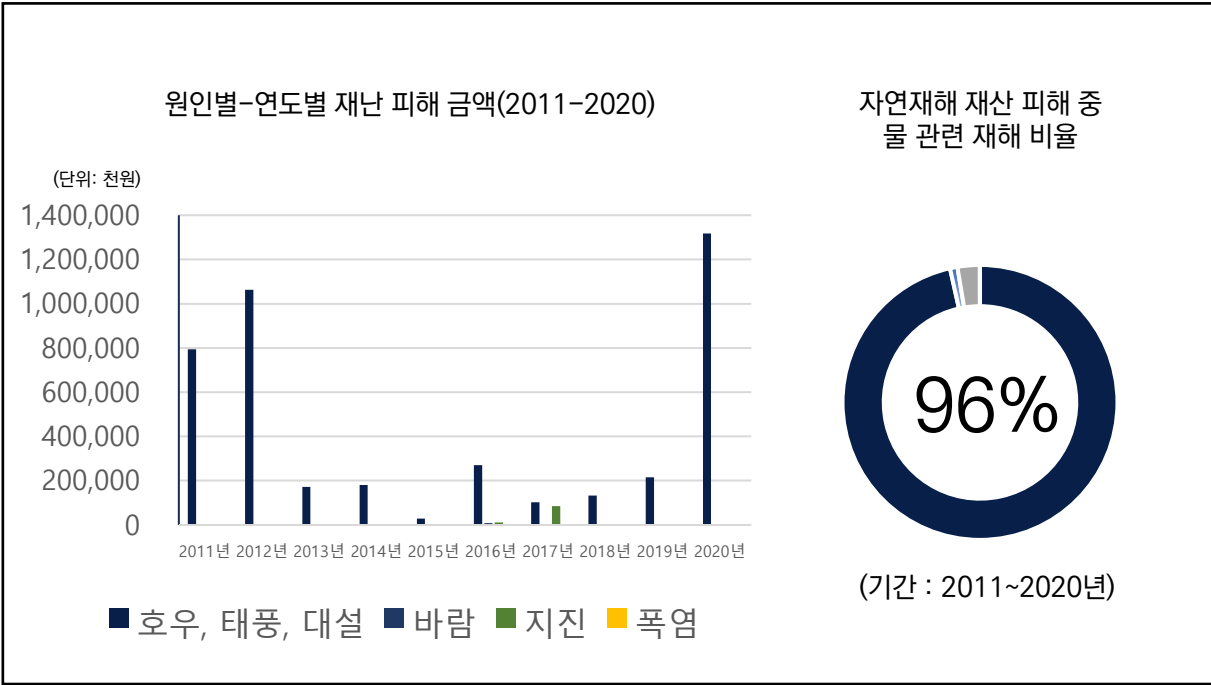
키워드 : #EDA #RandaomForest #SVC #KNN #Logistic Regression #PCA #K-Means #DBSCAN #QGIS

## [ 활용 데이터 ]

강우량, 침수흔적도, 수치표고모델, 배수등급도, 하천경계, 녹지면적, 서울시 도로명주소 기반 건물 공간데이터, 관공서 위치, 서울시 10m단위 도로구간공간, 빗물펌프장, 수문, 암거, 제방, 연속수치지형도, 건축물대장, 인구데이터, 상권분석서비스(소득소비) 등

## [ 주제 선정 배경 ]

- 1. 우리나라 자연재해의 96%는 물 관련 피해이며, 기후변화로 인한 극한호우가 증가하고 있음
- 2. GIS를 사용하여 강수량과 침수 빈도, 서울시 자연재해위험개선지구를 확인한 결과 반드시 일치한다고 볼 수 없음



# [ 분석 프로세스 ]

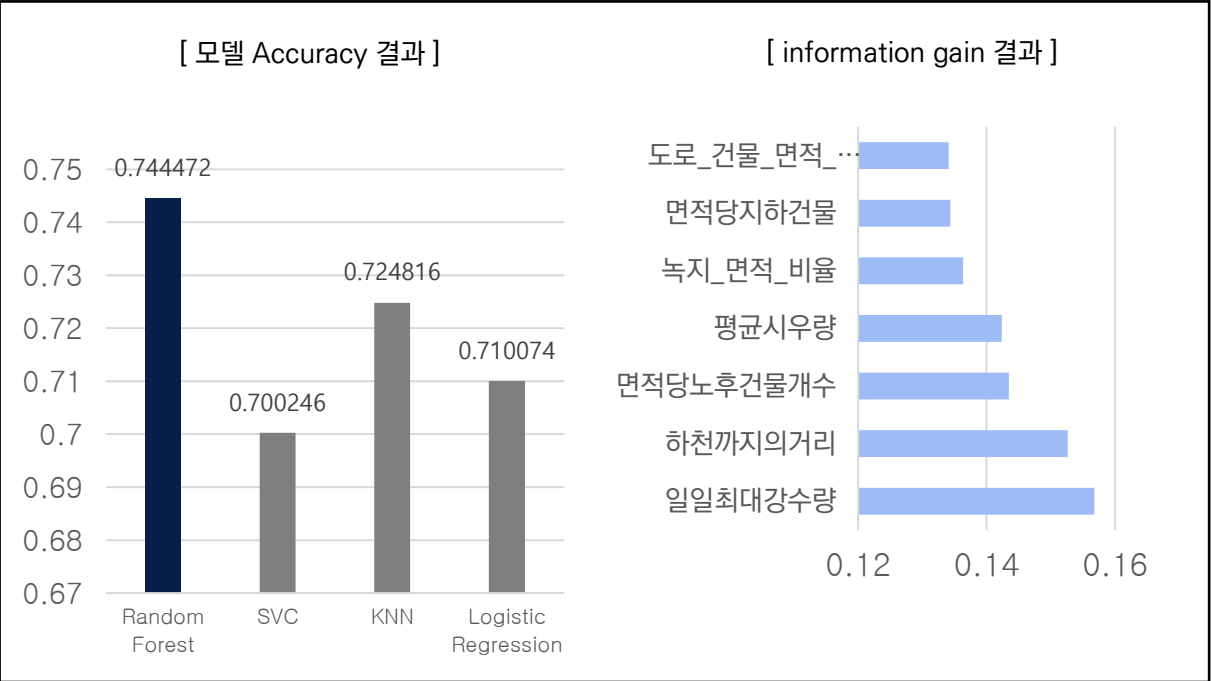
1. 데이터 수집 및 전처리 1) 데이터 수집 - 사회 인프라 데이터셋 - 인구 데이터셋 - 강수량/침수 데이터셋  2) 데이터 전처리 - 구단위로 일일 강수량 집계 - 행정동 단위로 인프라 집계 - 행정동 단위 지형 데이터 집계 - 행정동 단위 사회 데이터 집계	2. 침수 여부 결정 요인 확인 1) 요인별 유의성 검증 - MinMaxScaler - T-test  2) 모델 생성 및 비교 - Random Forest - Logistic Regression - SVC - KNN Classifier	3. 침수 빈도 결정 요인 파악 1) 주요 성분 분석 - MinMaxScaler - 시각화 위해 2차원 축소 - 주요 차원으로 군집분석 실행  2) 군집 분석 - K-means - DBSCAN	4. 군집별 요인의 상관관계 파악 1) 군집 별 회귀분석 - 침수빈도 결정규칙 도출 - 침수빈도 결정 주요 요인 색출 - 침수빈도와 변수 간의 수식 도출	5. 분석 결과 1) 유형별 분석 - 침수 빈도 요인 유형별 군집 시각화 - 소셜 인프라 취약 유형별 군집 분석 - 지역별 해결 방안 탐색
---	--	---	---	---

# [ 침수 발생 결정 요인 및 모델 결정 ]

1. 침수 발생 여부 분석에 사용할 변수를 T-test를 통해 선정함
2. 분석 모델 간 성능을 비교함. 각 모델 별 파라미터는 Random Search 및 Grid Search로 최적화. 분류모델의 정확도 추출 및 평가  
→ 정확도 검사 결과 정확도가 가장 높은 랜덤 포레스트를 사용. 변수 별 information 값을 통해 침수여부와 관련된 주요 변수 추정

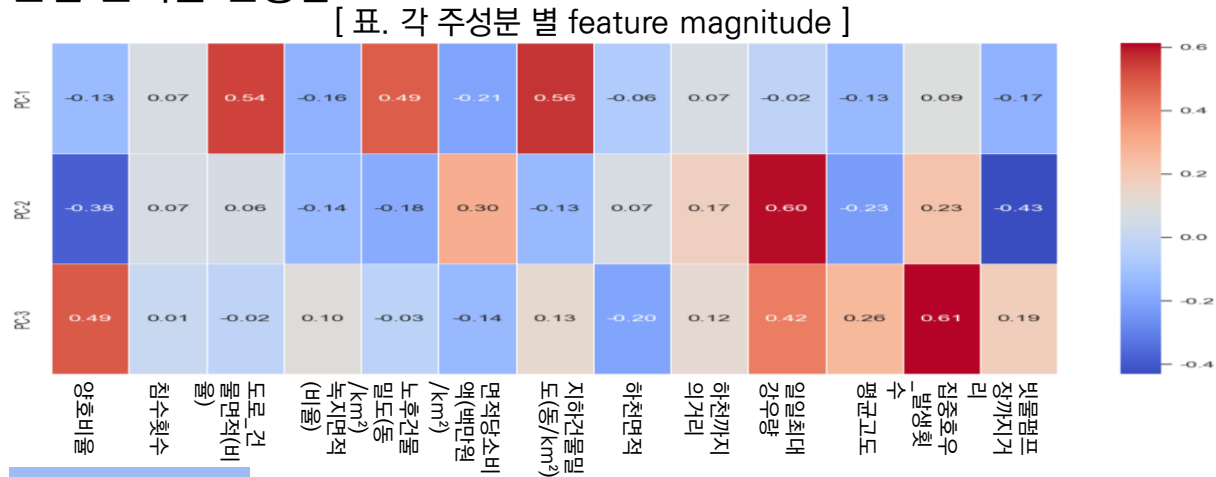
[ 요인별 침수 발생 여부 유의성 검정(T-test)을 통한 사용 변수 선정 ]			
Feature	T-statistic	P-value:	검정결과
녹지_면적_비율	-2.220030	0.027300	유의미함
도로_건물_면적_비율	2.657490	0.008283	유의미함
면적당노후건물개수	3.074329	0.002297	유의미함
하천까지의거리	-2.778680	0.005950	유의미함
면적당지하건물	3.841342	0.000147	유의미함
구단위시우량	3.463020	0.000621	유의미함
일일최대강수량	2.477491	0.013911	유의미함
경사도	0.230574	0.817815	유의미하지않음
인구밀도(명/km²)	0.113898	0.909404	유의미하지않음
평균고도	1.426690	0.154728	유의미하지않음
집중호우_발생횟수	0.238061	0.812040	유의미하지않음
분기평균총지출	-0.655311	0.512981	유의미하지않음
양호비율	1.403122	0.161873	유의미하지않음
빛물펌프장까지의거리	1.098923	0.272600	유의미하지않음
면적당제방	-0.746198	0.456310	유의미하지않음
면적당암거	-1.479084	0.140224	유의미하지않음
면적당수문	1.190880	0.234409	유의미하지않음
하천면적/면적	-1.640116	0.102357	유의미하지않음

구분	변수
종속변수	침수 여부
독립변수	평균 시우량
	일일 최대 강수량
	면적당 노후건물 개수
	면적당 지하건물 개수
	도로건물 면적 비율
	녹지 면적 비율
	하천까지의 거리

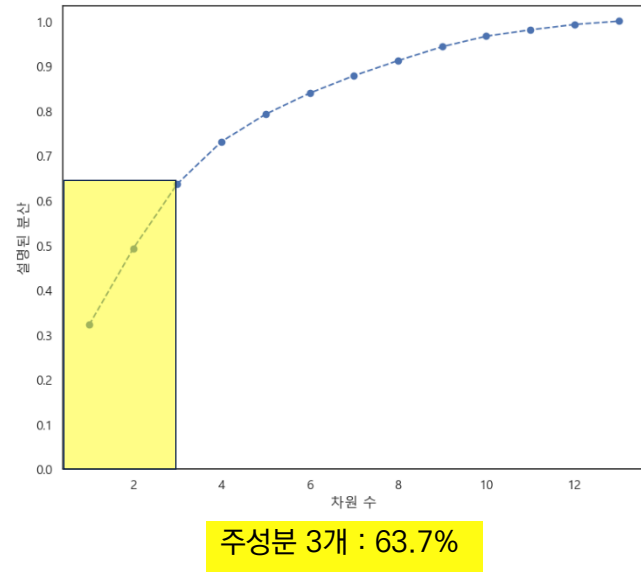


# [ 침수 빈도 결정 요인 ]

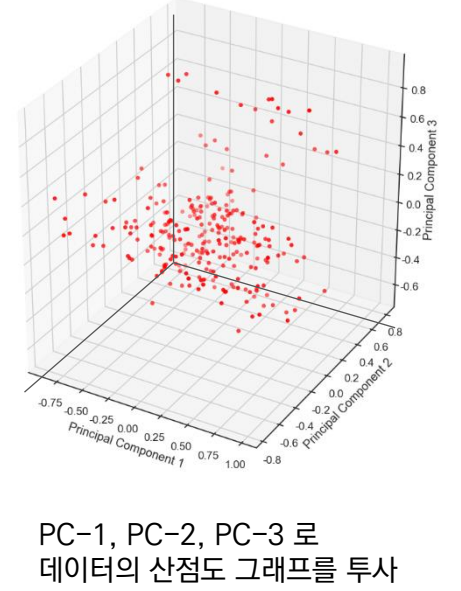
침수 빈도를 추정하기 위한 설명변수가 13개이므로 군집분석에서의 차원의 저주를 피하기 위해 주성분 분석을 통해 주성분을 산출하여 군집 분석을 진행함



## [ 전체에서 해당 주성분의 고윳값이 차지하는 비율 ]

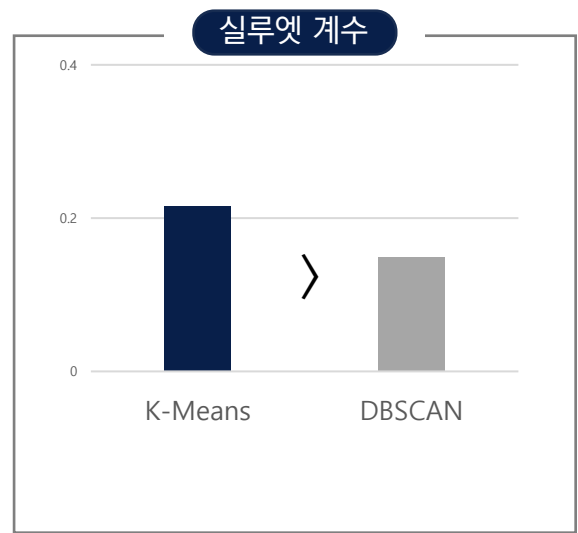
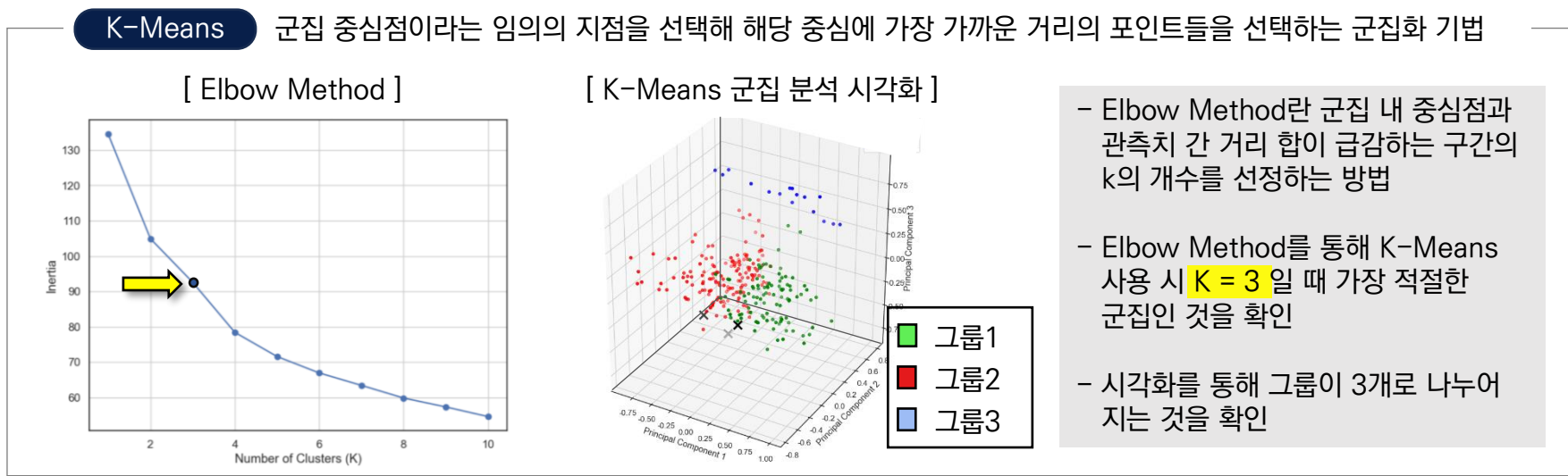


## [ 전체 변동성 PCA 분산 비율 ]



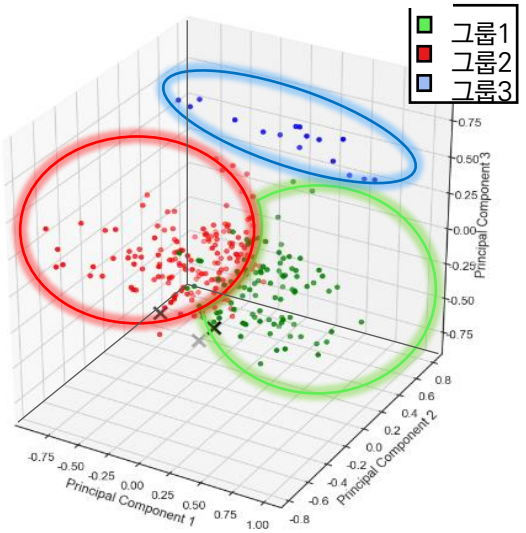
# [ 군집 분석 ]

서울시 전체가 동일한 요소를 가지지 않기 때문에 지역의 유형화를 통해 요인 분석의 정확도를 높여야 한다고 판단하여, 군집 알고리즘인 K-Means와 DBSCAN 분석 실행 후 실루엣 계수가 상대적으로 높은 K-Means 분석을 채택함

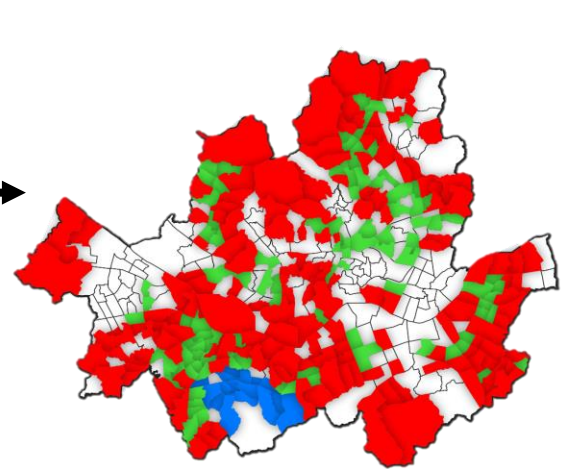


[ 군집분석에 따른 서울시 침수 유형별 군집 지도 ]

[ K-Means 군집 분석 시각화 ]



[ 서울시 침수 유형별 군집 지도 ]



그룹1 - 의사결정나무회귀&다중선형회귀분석

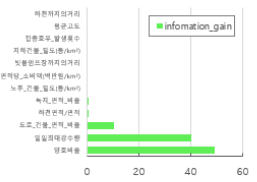
군집 분석으로 묶은 그룹별 특성 - 그룹1

1. 의사결정나무 회귀분석 R² : 0.897

의사결정나무 회귀

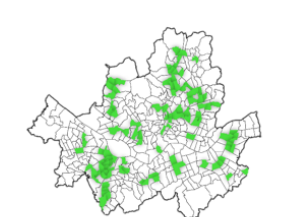


information gain



feature_name	information_gain
배수등급 양호비율	49.2474098
일일최대강수량	40.2233156
도로_건물_면적_비율	10.4492988
하천면적/면적	0.0474043
녹지_면적_비율	0.0325715

[ 그룹1 - 행정동 100개 ]



회귀나무를 이용해 그룹1의 침수 빈도를 결정하는 주요 변수 '배수등급 양호비율, 일일최대강수량, 도로&건물면적 비율, 하천면적 비율, 녹지면적 비율' 도출함

2. 다중선형회귀분석 R² : 0.433

회귀분석 변수 별 계수 및 유의도 결과

	coef	std err	t	P> t
면적당_소배역(백만원/km²)	3.724e+07	2.31e+07	1.611	0.111
지하건물_밀도(동/km²)	0.0671	0.026	2.557	0.012
집중호우_발생횟수	-16.2541	4.475	-3.632	0.000
평균고도	-1.1193	0.688	-1.627	0.107
하천까지의거리	-34.8680	16.627	-2.097	0.039
일일최대강수량	0.3053	0.071	4.296	0.000
도로_건물_면적_비율	-22.0810	146.206	-0.151	0.880
하천면적/면적	-0.0002	0.000	-1.084	0.281
양호비율	-39.1326	44.686	-0.876	0.383

침수횟수 =

14139.518 \* 면적당 지하건물개수 + (-0.02) \* 집중호우발생횟수 + (-0.131) \* 하천까지의거리 + 0.086 \* 일일최대강수량 + (-22.217) \* 도로&건물면적비율 + (-97.338) \* 배수등급양호비율 + 217.7



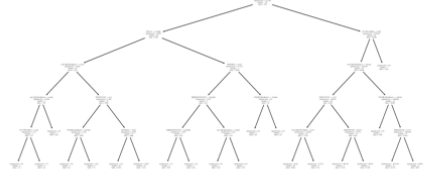
그룹1의 공통적 특징 : **치수 및 배수 시설 정비 문제**

그룹2 - 의사결정나무회귀&다중선형회귀분석

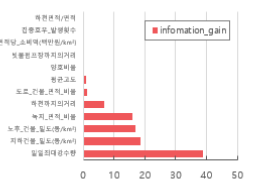
군집 분석으로 묶은 그룹별 특성 - 그룹2

1. 의사결정나무 회귀분석 R² : 0.89

의사결정나무 회귀

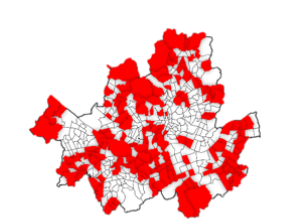


information gain



feature_name	information_gain
일일최대강수량	38.87999054
지하건물밀도(동/km²)	18.51578481
노후건물밀도(동/km²)	16.85658336
녹지_면적_비율	16.01760894
하천까지의거리	6.711371052
도로_건물_면적_비율	1.383381956
평균고도	1.024727273
배수등급 양호비율	0.429718503
빗물펌프장까지의거리	0.180833567

[ 그룹2 - 행정동 150개 ]



회귀나무를 이용해 그룹2의 침수 빈도를 결정하는 주요 변수 '일일최대강수량, 지하건물밀도, 노후건물밀도, 녹지면적 비율, 하천까지의거리, 도로&건물 면적 비율, 고도, 배수등급 양호비율, 빗물펌프장까지의 거리' 도출함

2. 다중선형회귀분석 R² : 0.421

회귀분석 변수 별 계수 및 유의도 결과

	coef	std err	t	P> t
도로_건물_면적_비율	74.8520	35.382	2.116	0.036
집중호우_발생횟수	-4.4153	1.199	-3.681	0.000
일일최대강수량	0.0897	0.018	4.923	0.000
양호비율	-20.2161	13.997	-1.444	0.151
노후_건물_밀도(동/km²)	0.0020	0.020	1.118	0.265
녹지_면적_비율	19.9668	17.545	1.138	0.257
빗물펌프장까지의거리	-0.0004	0.002	-0.166	0.868
평균고도	0.1667	0.088	1.900	0.059
지하건물_밀도(동/km²)	-0.0143	0.020	-0.715	0.476
하천까지의거리	4.9867	4.740	1.052	0.295

침수횟수 =

0.102 \* 일일최대강수량 + (-34368.932) \* 면적당지하건물개수 + (-0.005) \* 집중호우발생횟수 + (-0.161) \* 하천까지의거리 + (-169.079) \* 배수등급양호비율 + 184.52 \* 녹지면적비율 + 374.6



그룹2의 공통적 특징 : **건물 노후, 도시의 과개발**

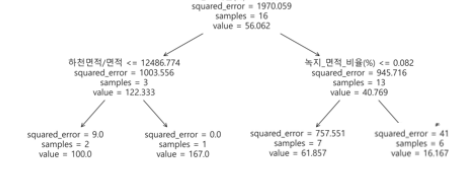
선형회귀를 이용해 도출한 그룹2의 침수빈도와 변수 간의 관계식(선형관계)을 도출함

그룹3 - 의사결정나무회귀&다중선형회귀분석

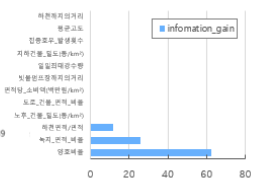
군집 분석으로 묶은 그룹별 특성 - 그룹3

1. 의사결정나무 회귀분석 R² : 0.88

의사결정나무 회귀

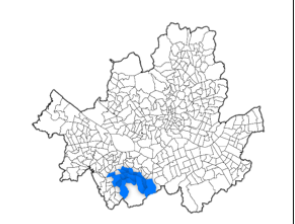


information gain



feature_name	information_gain
배수등급 양호비율	62.4814426
녹지_면적_비율	25.98756606
하천면적/면적	11.53099134

[ 그룹3 - 행정동 16개 ]



회귀나무를 이용해 그룹3의 침수 빈도를 결정하는 주요 변수 '배수등급 양호비율, 녹지면적 비율, 하천면적 비율' 도출함

2. 다중선형회귀분석 R² : 0.726

회귀분석 변수 별 계수 및 유의도 결과

	coef	std err	t	P> t
노후_건물_밀도(동/km²)	0.2344	0.097	2.427	0.046
면적당_소배역(백만원/km²)	1.496e+08	7.78e+07	1.922	0.096
일일최대강수량	0.3564	0.105	3.410	0.011
지하건물_밀도(동/km²)	-0.1631	0.058	-2.802	0.026
집중호우_발생횟수	-4.1802	2.194	-1.905	0.098
평균고도	-0.8331	0.443	-1.879	0.102
하천까지의거리	-42.7403	29.993	-1.425	0.197
양호비율	-67.9697	50.321	-1.351	0.219
녹지_면적_비율	110.2180	297.197	0.371	0.722
하천면적/면적	0.0018	0.001	1.223	0.261

침수횟수 =

75.317 \* 도로&건물면적비율 + (-0.005) \* 집중호우발생횟수 + 0.026 \* 일일최대강수량 + (-50.289) \* 배수등급양호비율 + 33.427 \* 녹지면적비율 + (-3013.34) \* 면적당지하건물개수 + 0.019 \* 하천까지의거리 + 72.0



그룹3의 공통적 특징 : **하수관거 정비와 녹지면적**

선형회귀를 이용해 도출한 그룹3의 침수빈도와 변수 간의 관계식(선형관계)을 도출함

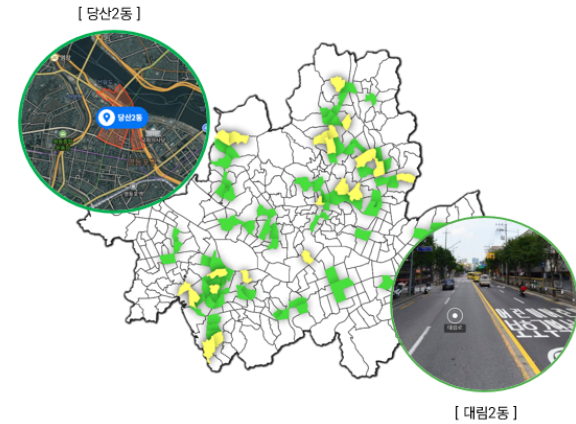


# [ 결론 ]

## 분석 활용방안

침수 빈도 결정 규칙을 통해 지역별 맞춤 개선 방안 제안

### 치수 및 배수 시설 정비 문제



#### 침수 빈도 결정 규칙

- 1) 일일 최대 강수량이 높고 배수 양호 비율이 낮은 지역
- 2) 일일 최대 강수량이 낮지만 녹지 면적 비율이 낮은 지역
- 3) 일일 최대 강수량이 높고 배수 양호시설 비율이 낮고 불투수면 비율이 높은 지역

#### 1. 대림1동, 사당 1동

원인 : 강우 강도가 매우 강하고 배수 시설 정비가 원활 하지 않은 것으로 추정  
해결방안 : 배수체계 개선, 배수능력 제고

#### 2. 쌍문2동, 수유 2,3동, 송천동, 번1동 등 8개 동

원인 : 비의 세기는 상대적으로 약하지만 비 흡수 지형이 부족  
해결방안 : 투수성을 재고하여 (수변,공원,녹지) 등을 확보

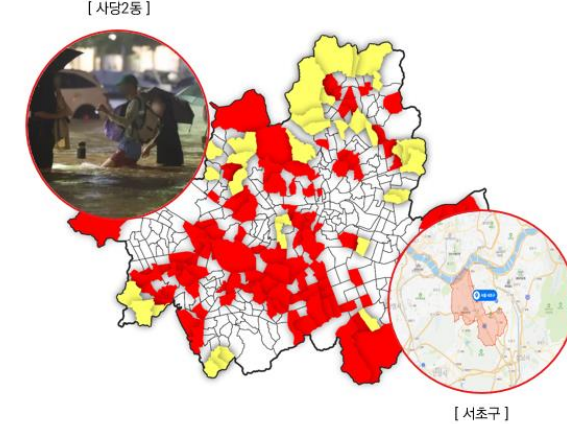
#### 3. 영등포구 일대 (신길,영등포본동,대림) 7개 동

원인 : 도시 개발 진행 속도에 비해 배수 시설 정비가 부족  
해결방안 : 침투시설 증대와 배수 체계 개선 필요

## 분석 활용방안

침수 빈도 결정 규칙을 통해 지역별 맞춤 개선 방안 제안

### 건물 노후, 도시의 과개발



#### 침수 빈도 결정 규칙

- 1) 일일 최대 강수량이 높고 불투수면은 낮지만 녹지 면적 비율이 낮은 지역
- 2) 일일 최대 강수량이 매우 높으며 녹지 면적이 적고 지하건물과 노후 건물의 밀도가 높은 지역

#### 1. 사당 2동 등

원인 : 도로율이 낮지만 면적당 시우량이 높아 물 저장 능력 부족  
해결방안 : 빗물 저류조와 같이 물을 저장할 수 있는 시설 필요

#### 2. 구로2동, 방배2동, 방배본동, 개봉1동 등

원인 : 강수량이 높는데 반해 건물들의 시설물의 노후화와 특색으로 침수 방어가 힘든 지형  
해결방안 : 시설물 점검 및 보수, 저류 시설 확대

#### 3. 가산동 삼전동

원인 : 하천까지와의 거리가 가까우며 불투수면이 높은 지역  
해결방안: 범람에 의한 홍수가 자주 발생하는 지역으로 방재 시설의 증대 필요

## 분석 활용방안

침수 빈도 결정 규칙을 통해 지역별 맞춤 개선 방안 제안

### 하수관거 정비와 녹지면적



#### 침수 빈도 결정 규칙

- 1) 배수시설 양호비율이 낮고 하천 면적 비율이 높은 지역
- 2) 배수시설은 상대적으로 양호하지만 녹지 면적 비율이 낮은

#### 1. 신림동

원인 :하천의 범람을 배수할 수 있는 시설의 노후화  
해결방안 : 하수관거 정비를 통해 노후 관거를 교체하고, 기존 하수관거 용량부족을 해결  
또한 많은 예산이 소요되므로 하수처리구역별 우선순위를 정할 필요가 있음

#### 2. 은천동, 마성동, 삼성동

원인 : 도시의 녹지 면적이 적어 비 흡수 시설이 부족  
해결방안 : 도시의 녹화 산업 증대 필요

# [ 의의 ]

1. 현재 서울시 자연재해위험개선지구에서 선정한 재해 위험지구와 달리, 2022년까지의 데이터를 기반으로 침수피해에 영향을 주는 요인 분석을 통해 실질적 재해 위험지구를 구분함 서울시의 개별 지역이 다른 환경인 점에 착안,각 지역별 침수 피해 원인 파악이 가능함
2. 동일한 유형으로 묶인 그룹 간 침수 피해 데이터를 공유함으로써 정보의 이득이 증가
3. 서울시 침수위험지구에 선정되지 못한 이유인 집값 하락 및 재선의 악영향 우려를 방재 시설 제안을 통해 경제적 효율을 높여 지역 균형 발전, 주민 인식 개선 및 집값 안정화에 도움을 줄 수 있음