

프로젝트기반 빅데이터 분석 과정 3기

# 프로젝트 계획서

2022. 5. 24

프로젝트 명	공공 데이터를 활용한 서울특별시 ‘여성 1인 가구’ 안전성 확인 및 해결 방안 탐색	
개발 기간	2022.05.18 ~ 2022.06.08	
팀 장	김해인	eeheainee@gmail.com
팀 원	조가윤	sangsari2221@gmail.com
	주은정	hahory0@gmail.com

# 목 차

## I. 프로젝트 개요

1. 프로젝트 개요 .....	1
2. 추진배경 및 필요성 .....	1
3. 개발범위 .....	2
4. 기대효과 .....	4

## II. 추진 방안

1. 추진목표 .....	5
2. 추진전략 .....	5
3. 추진체계 및 역할 .....	6
4. 추진일정 .....	6

## III. 개발 내용

1. 개발내용 개요 .....	8
2. 상세 요구사항 .....	8

## IV. 평가방법

1. 평가방법 개요 .....	27
2. 세부 평가 항목 .....	30
3. 점수 방법 .....	30

## 1 프로젝트 개요

- 프로젝트 명: 공공 데이터를 활용한 서울특별시 ‘여성 1인 가구’ 안전성 확인 및 해결 방안 탐색
- 개발 기간: 2022.05.18 ~ 2022.06.08 : 프로젝트 제안서 작업
- 주요 개발 내용
  - 수집·정제된 데이터 분석
  - 각 변수의 분포 시각화

## 2 추진배경 및 필요성

- 1인 가구의 증가로 인한 연구 필요성 대두

2000년도 50만 가구였던 서울시 1인 가구 수는 2015년 100만 가구를 넘어 2020년 기준 139만 가구로 증가하였다.
- 여성 1인가구의 안전성 문제 대두

형사정책연구원에 따르면 여성 청년 1인 가구에 주거침입 피해 가능성은 남성에 비해 11.226배에 달하며 이러한 통계결과들은 여성 1인 가구 안전을 위한 정책이 필요하다는 것을 알 수 있다.
- 낮은 정책 인지도 및 범죄율 증가

안심택배, 여성안심귀가스카우트 등 여성 1인가구를 위한 범죄 예방정책이 나왔지만 20% 미만의 낮은 인지도를 갖고 있다.

경찰청에 따르면 2014년 1만 건도 되지 않던 주거침입 범죄로 검거된 사람은 2019년 1만 5천여 건을 넘어섰다.

### 3 개발범위

---

#### ① 빅데이터 자료 수집 방법

1. 서울시 자치구 연도별 CCTV 설치 현황
2. 서울시 구별 여성 1인 가구 현황
3. 서울시 관서별 지구대, 파출소, 치안센터 현황
4. 서울시 구별 유흥업소 수
5. 서울시 관서별 5대 범죄 발생검거 현황
6. 서울시 구별 여성안심지킴이집 점포 수
7. 서울시 구별 가구형태별 가구 및 가구원 통계

#### ② 빅데이터 자료 정제 방법

1. 데이터 정규화를 통해 인구수 맞춰주기
2. 결측치 확인 및 제거하기
3. 이상치 확인 및 제거하기
4. 데이터 변환 및 파생변수 생성

#### ③ 분석 방법 및 구현 방법

1. 위험요인과 예방요인 간의 피어슨 상관분석
2. Kmeans 군집화를 통한 군집유형 분석  
(Elbow Method와 Silhouette Method를 통해 최적클러스터값 도출)
3. 각 변수의 분포 시각화

## 4 기대효과

---

### ○ 실효성 있는 여성 안전 정책 기대

범죄발생 변수와 예방변수들에 따른 안전성을 파악할 수 있으며 이를 토대로 여성 1인가구를 위한 행정정책 등을 기대 할 수 있다.

### ○ 안전 지역 정보의 자유

서울시 거주를 원하는 예비 여성 1인가구들이 공공데이터를 분석한 자료를 바탕으로 안전성이 높은 거주지를 선택할 수 있다.

### ○ 주거 안전 보장

1인 가구 여성 뿐만 아니라 1인 가구 남성, 더 나아가 해당 구에 거주하고 있는 거주민들의 안전성을 확인 할 수 있는 데이터를 제공한다.

## 1 추진목표

- 서울시 공공데이터를 이용하여 범죄 발생 요인(여성 1인 세대 수, 범죄율, 유흥업소비율)과 범죄 예방 요인(CCTV비율, 여성안심지킴이집 비율, 치안시설비율)의 서울시 지역별 현황을 분석하고, 이 결과를 바탕으로 안전한 거주 환경을 위한 정책 대안을 제시하고자 한다.

## 2 추진전략

- KDD(Knowledge Discovery in Databases) 분석 방법론 과정을 따라 진행함으로써 체계화된 절차와 방법을 따르고자 한다.

## 3 추진체계 및 역할

분석대상 비즈니스 이해 (과제 정의)	-요구사항및이슈파악 -과제관련개념및이론정리 -목적정의및수행방안설계	-팀원 전원
데이터셋 선택	-분석에필요한데이터수집및선택 -데이터마이닝에필요한목표데이터를구성	-팀원 전원
데이터 전처리	-데이터재처리를통한데이터셋정제 -추가로요구되는데이터셋이필요한경우데이터 선택프로세스를재실행	-데이터수집및탐색적자료분석:김해인 -데이터셋정제:주은정
데이터 변환	-분석목적에맞게변수를생성및선택 -효율적인데이터마이닝을위한데이터변경	-데이터변환:조가운
데이터 마이닝	-데이터마이닝기법선택및실행 -R을이용한분석실행 -Python을이용한분석실행 -데이터시각화	-R분석:주은정 -Python분석:김해인 -데이터시각화:조가운
데이터 마이닝 결과 평가	-결과에대한해석및평가 -분석목적과의일치성확인	-팀원 전원
결과 보고	-결과보고서작성 -최종보고회	-발표자:조가운

## 4 추진일정

○ 개발 기간: 2022.05.18 ~ 2022.06.08

구 분	5월											6월						
	18	19	20	23	24	25	26	27	30	31	1	2	3	6	7	8		
1. 분석대상 비즈니스의 이해(과제 정의)																		
▪ 요구사항 및 이슈 파악	●	●	●															
▪ 과제 관련 개념 및 이론 정리	●	●	●															
▪ 목적 정의 및 수행방안 설계			●	●														
▪ 프로젝트 계획서 작성 및 제출				●	●	●												
2. 데이터셋 선택																		
▪ 데이터 수집 및 선택						●	●											
▪ 목표 데이터를 구성						●	●											
3. 데이터 전처리																		
▪ 데이터셋 정제							●											
4. 데이터 변환																		
▪ 변수 생성 및 선택							●											
▪ 데이터 변환							●											
5. 데이터 마이닝																		
▪ R을 이용한 분석								●	●	●	●	●	●	●				
▪ Python을 이용한 분석								●	●	●	●	●	●	●				
▪ 데이터 시각화								●	●	●	●	●	●	●				
6. 데이터 마이닝 결과 평가																		
▪ 결과에 대한 해석 및 평가														●	●			
7. 결과 보고																		
▪ 결과보고서 작성													●	●	●			
▪ 최종보고회																●		

※ 세부 일정은 추진상황에 따라 변경 가능

## 1 개발내용 개요

- 본 계획서 「공공 데이터를 활용한 서울특별시 ‘여성 1인 가구’ 안전성 확인 및 해결 방안 탐색은 분야별 계획사항을 충분히 검토 후 요구 기능이 포함된 계획서를 작성하여야 함
- 계획서에는 본 사업에 대한 이해를 전제로 구체적인 방안이 제시되어야 하며, 본 계획서에서 추구하는 목표달성을 위해 추가해야 할 사항이 있을 경우 추진방안 등을 명확히 제시하여야 함

## 2 상세 요구사항

분류		번호	요구사항 명칭
컨설팅 요구 사항	분석과제 파악	CNR-001	과제정의 및 요구사항 분석
		CNR-002	분석 데이터 확보 및 정의
	데이터 분석	CNR-003	R을 활용한 데이터 분석
		CNR-004	Python을 활용한 데이터 분석
		CNR-005	분석 결과 시각화
	분석결과 검증	CNR-006	분석 결과 검증



<b>요구사항 고유번호</b>		<b>CNR-001</b>
<b>요구사항 분류</b>		분석 과제 파악
<b>요구사항 명칭</b>		과제정의 및 요구사항 분석
<b>요구사항 상세설명</b>	<b>정의</b>	분석 대상 정의 및 분석 방법 등 세부 시나리오 정의
	<b>세부내용</b>	<ul style="list-style-type: none"> <li>• <b>분석과제 목표 및 범위 설정</b> <ul style="list-style-type: none"> <li>- 각 과제의 추진 목표 및 추진 방향 설정</li> <li>- 시간적·공간적·내용적 과업 범위</li> <li>- 분석 방법 및 기법 정의</li> </ul> </li> </ul>
<b>산출물</b>		분석 과제 정의서

<b>요구사항 고유번호</b>		<b>CNR-002</b>
<b>요구사항 분류</b>		분석 과제 파악
<b>요구사항 명칭</b>		분석 데이터 확보 및 정의
<b>요구사항 상세설명</b>	<b>정의</b>	분석 대상 데이터 확보, 세부 현황 및 분석 범위, 분석 흐름도 작성
	<b>세부내용</b>	<ul style="list-style-type: none"> <li>• <b>과제 수행에 필요한 분석 데이터 확보 및 정의</b> <ul style="list-style-type: none"> <li>- 통계청 및 서울열린데이터광장 등 공공 데이터 현황 파악</li> <li>- 분석 필서 데이터 확보</li> </ul> </li> <li>• <b>분석대상 데이터 목록</b> <ol style="list-style-type: none"> <li>1. 서울시 자치구 연도별 CCTV 설치 현황</li> <li>2. 서울시 구별 여성 1인가구 현황</li> <li>3. 서울시 관서별 지구대, 파출소, 치안센터 현황</li> <li>4. 서울시 구별 유흥업소 수</li> <li>5. 서울시 관서별 5대범죄 발생검거 현황</li> <li>6. 서울시 구별 여성안심지킴이집 점포 수</li> <li>7. 서울시 구별 가구형태별 가구 및 가구원 통계</li> </ol> </li> <li>• <b>데이터 정제 전처리</b> <ul style="list-style-type: none"> <li>- R 프로그램을 활용하여 데이터 정제</li> </ul> </li> </ul>
<b>산출물</b>		분석 데이터 정의서

요구사항 고유번호		CNR-003
요구사항 분류		R을 활용한 데이터 분석
요구사항 명칭		서울특별시의 범죄 위험 요인에 따른 범죄 예방 요인의 분석
요구사항 상세설명	정의	데이터 수집 및 정의, 통계적 모형 개발, 분석 결과 활용, 시각화
	세부내용	<ul style="list-style-type: none"> <li>• 위험 요인의 특성에 따라 예방 요인을 예측해보는 전략 <ul style="list-style-type: none"> <li>- 범죄의 위험 요인과 예방 요인을 각각 3가지로 보고, 발생 요인에 속하는 여성 1인 세대수, 범죄율, 유흥업소 비율이 높으면 CCTV 비율, 안심지킴이집 비율, 치안 시설 비율이 낮다고 봄.</li> </ul> </li> <li>• 범죄 발생 요인과 예방 요인과의 교차 분석 <ul style="list-style-type: none"> <li>- 여성 1인 가구 수와 CCTV 비율간의 관계 분석</li> <li>- 범죄율과 CCTV 비율간의 관계 분석</li> <li>- 유흥업소비율과 CCTV 비율간의 관계 분석</li> <li>- 여성 1인 가구 수와 안심지킴이집 비율간의 관계 분석</li> <li>- 범죄율과 안심지킴이집 비율간의 관계 분석</li> <li>- 유흥업소비율과 안심지킴이집 비율간의 관계 분석</li> <li>- 여성 1인 가구 수와 치안시설 비율간의 관계 분석</li> <li>- 범죄율과 치안시설 비율간의 관계 분석</li> <li>- 유흥업소비율과 치안시설 비율간의 관계 분석</li> </ul> </li> </ul>
산출물		분석결과 보고서, 시각화 구현 결과물

요구사항 고유번호		CNR-004
요구사항 분류		Python을 활용한 데이터 분석
요구사항 명칭		서울특별시의 범죄 위험 요인에 따른 범죄 예방 요인의 분석
요구사항 상세설명	정의	데이터 수집 및 정의, 통계적 모형 개발, 분석 결과 활용, 시각화
	세부내용	<ul style="list-style-type: none"> <li>• 위험 요인의 특성에 따라 예방 요인을 예측해보는 전략 <ul style="list-style-type: none"> <li>- 범죄의 위험 요인과 예방 요인을 각각 3가지로 보고, 발생 요인에 속하는 여성 1인 세대수, 범죄율, 유흥업소 비율이 높으면 CCTV 비율, 안심지킴이집 비율, 치안 시설 비율이 낮다고 봄.</li> </ul> </li> <li>• 범죄 발생 요인과 예방 요인과의 교차 분석 <ul style="list-style-type: none"> <li>- R을 활용한 데이터 분석 내용과 동일</li> <li>- 위 조건으로 부족한 경우, 추후 보완을 위해 위험요인에 주거형태를 추가하여 분석을 통해 새로운 모델을 만들 수 있음.</li> <li>- 같은 조건으로 R과 Python으로 동일 분석을 하여 서로의 결과를 검토하여 보완해 봄.</li> </ul> </li> </ul>
산출물		분석결과 보고서, 시각화 구현 결과물

<b>요구사항 고유번호</b>		<b>CNR-005</b>
<b>요구사항 분류</b>		데이터 분석 및 시각화
<b>요구사항 명칭</b>		분석 결과 시각화
<b>요구사항 상세설명</b>	<b>정의</b>	활용 가능한 형태의 분석 결과 시각화 제작
	<b>세부내용</b>	<b>분석 결과 시각화</b> - 산출된 분석 결과 중 필요한 현황을 시각화하여 활용할 수 있는 형태로 제작하여 제공 - 구 단위로 시각화 - 차트, 그래프, GIS 등 다양한 방법으로 제시
<b>산출물</b>		시각화 대상 데이터 및 시각화 구현 결과물

<b>요구사항 고유번호</b>		<b>CNR-006</b>
<b>요구사항 분류</b>		분석결과 검증
<b>요구사항 명칭</b>		분석 결과 검증
<b>요구사항 상세설명</b>	<b>세부내용</b>	• 분석 결과 검증 - 예측모형으로 도출된 회수 점수·등급과 이에 따른 실질적인 회수 활동 결과와 비교 검증 - 검증 후 추가 수정/보완사항 반영
<b>산출물</b>		검증결과 보고서

## IV

## 평가 방법

## 1 평가 방법 개요

- 데이터 활용, 데이터 분석, 프로젝트 관리 등의 기준을 세우고 알맞게 진행되었는지 평가한다.

## 2 세부 평가 항목

평가부문	평가항목	평 가 요 소	배점
데이터 활용	데이터 요구사항 (1)	방법론 및 분석 도구를 통하여 구체적인 내용으로 분석되고 구현 방안이 구체적으로 기술되어 있는가를 평가하고, 제안한 방안 및 기술이 적용 가능한지를 평가한다.	40
	데이터 요구사항 (2)	데이터 전환 계획 및 검증 방법, 예러 데이터 처리 방법에 대해 구체적인 내용을 제시하고 있는가를 평가하며, 데이터 전환을 위해 책임 조직이 투입되는가를 평가한다.	
데이터 분석	분석 목표의 적정성	분석의 적절한 목표를 제안하며 제시된 목표가 구체적으로 기술되어 있고, 적정한가를 평가한다.	40
	분석 방안의 적정성	분석의 적절한 방안을 제안하며 제시된 방안이 구체적으로 기술되어 있고, 적정한가를 평가한다.	
	공정별 세부계획의 적정성	공정별 세부계획을 제안하며 제시된 세부계획이 구체적으로 기술 되었는가를 평가하고, 실행 가능여부 및 적정성을 평가한다.	
프로젝트 관리	일정계획	사업수행에 필요한 활동을 도출하여 정확한 활동 기간의 산정과 도출된 활동 간의 배열이 합리적인지, 중간목표가 적절하게 제시되어 있는지, 각 활동에 적합한 자원이 적절히 할당되어 있는지 등을 평가한다.	20
합 계			

## 3 검수 방법

- 정해진 스케줄대로 프로젝트가 진행 되었는가?
- 적절한 분석방법을 통해 신뢰성 있는 데이터를 도출했는가?
- 추진 목표에 알맞은 데이터가 도출되었는가?