**2023 빅데이터캠퍼스 공모전 분석결과(요약)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 팀명 | 자운 | | 접수번호 | | 미기재 |
| 제목 | 경제적이고 안전을 보장해주는 빅데이터 기반 여성전용 쉐어하우스 안전 서비스 | | | | |
| 추진배경 및 필요성 | **○ 20~30대 1인 여성 가구의 증가 속에서 안전에 대한 걱정이 커짐.**  ▸ 최근 강남 묻지마 살인 사건 등 여성을 대상으로 한 범죄에 대한 사회적 이슈가 대두되면서 20~30대 여성들의 안전에 대한 걱정이 커짐.  ▸ 묻지마 살인 사건 직후 여성 호신용품에 대한 검색량이 크게 늘어난 점에서 여성의 안전 문제는 사회적 문제로 다뤄야 할 필요성이 있음.  **○ 올해 월세의 수요 증가와 월세 가격 상승으로 인한 주거비 압박 문제.**  ▸전세 사기 등의 이유로 인해 올해 월세 수요가 증가하면서 월세 가격도 같이 상승함. 고물가 시대 속에서 월세 가격의 인상은 생활비 압박 문제로 다가옴.  ▸1인 가구가 2인 이상의 가구보다 생활비 비중이 크다는 점에서 1인 가구의 주거비 증가 문제는 해결해야 할 문제임.  **○위의 이유들로 20~30대 독립하려는 여성들이 안전에 대한 걱정과 주거비 문제를 모두 해결해야 할 필요성이 커짐.**  ▸주거비 문제를 해결하기 위해 쉐어하우스를 해결책으로 내세울 수 있음. 추가적으로 안전한 대한 걱정을 해소하기 위해 입주를 고려하는 쉐어하우스 위치가 얼마나 안전한지 확인하기 위한 수단이 필요하기에 빅데이터를 통한 알고리즘 구현할 필요가 있음.  ▸쉐어하우스 입주시 카톡 혹은 전화로 문의할 경우 개인정보 노출에 대한 걱정이 있을 수 있기에 익명을 보장해주는 메신저를 통해 개인 정보 노출을 방지할 필요가 있음. 같은 쉐어하우스 입주한 사람들과도 개인정보 등의 노출로 거부감이 있을 경우 익명 채팅을 통한 개인 정보 노출을 방지할 수 있음.  **○1인 가구의 문제점 중 하나인 외로움이 높은 순위를 차지하고 있음.**  ▸ 쉐어하우스에 입주한 사람들을 위한 폐쇄형 SNS 커뮤니티를 통해 공감대를 형성할 수 있는 집단을 형성하며 외로움을 낮출 수 있을 것으로 기대됨. | | | | |
| 1. 투입 데이터 | 서울시 행정동 단위 거주인구 데이터  서울시 주민등록인구(구별)통계  월세수급동향(종합주택유형) | 종합주택 평균 월세가격  월별 매입자연령대별  여성 호신용품 검색량  서울시 자치구 연도별 방범용 CCTV 운영 현황 | | 서울시 안심이 CCTV 연계 현황  서울시 5대 범죄 발생현황 통계  혼자 생활하면서 곤란한 점  서울시 1인가구(연령별) 통계 | |
| 분석과정 및 방법 | **○ 데이터 수집 및 전처리**   |  |  | | --- | --- | | 서울시 5대 범죄 발생 현황 데이터 서울 열린 데이터 광장에서 다운 | 파생변수 | | 서울시 5대 범죄 발생 현황 | |  |  |  | | --- | --- | --- | | 원본 데이터 | 파생변수 산출방법 | 파생변수 명 | | 서울시 5대 범죄 발생 현황 | read\_csv로 엑셀 불러오기 | crime\_anal | | 서울시 5대 범죄 발생 현황 | 강간,강제추행 검거/ 강간, 강제 추행 발생 | crime\_anal[‘강간, 강제추행 검거율] | | 서울시 5대 범죄 발생 현황 | 살인검거/살인발생 | crime\_anal[‘살인 검거율’] | | 서울시 5대 범죄 발생 현황 | 절도 검거/절도 발생 | crime\_anal[‘절도검거율’] | | 서울시 5대 범죄 발생 현황 | 폭력 검거/폭력 발생 | crime\_anal[‘폭력 검거율’] | | 서울시 5대 범죄 발생 현황 | 강도 검거/ 강도 발생 | crime\_anal[‘강도 검거율’] | | 서울시 5대 범죄 발생 현황 | crime\_anal을 min\_max\_scaling | crime\_anal\_norm | | 서울시 자치구 연도별 방범용 CCTV 운영 현황 | 5대 범죄의 합 | crime\_anal\_norm[‘범죄’] | | 서울시 자치구 연도별 방범용 CCTV 운영 현황 | 5대 범죄 검거율의 합 | crime\_anal\_norm[‘검거’] | | | 자치구 연도별 방범용 CCTV 운영 현황 | | [표1] 서울시에서 제공하는 데이터 | [표2] 파생 변수 목록 |  |  |  | | --- | --- | | 서울시에서 제공하는 데이터 사용 | 파생변수 생성 | | |  |  | | --- | --- | | 데이터 명 | 수집 방법 | | 서울시 안심이 CCTV 연계 현황 | 서울 열린 데이터 광장 | | 서울시 주민등록인구(구별)통계 | 서울 열린 데이터 광장 | | |  |  |  | | --- | --- | --- | | 원본 데이터 | 파생변수 산출방법 | 파생변수 명 | | 서울시 안심이 CCTV 연계 현황 | unnamed column 삭제 후 ‘구분’을 ‘구별’로 변경 | CCTV\_Seoul | | 서울시 주민등록인구(구별)통계 | ‘기간’ colum 삭제 후 ‘자치구’를 ‘구별’로 변경 | pop\_Seoul | |  | CCTV\_Seoul값과 pop\_Seoul을 합하기 | data\_result | | 서울시 안심이 CCTV 연계 현황 | 2023년 cctv 개수/ 지역별 인구 | Data\_result[‘CCTV’] | |  | Polyfit(인구수,2023년 cctv개수) | fp1 | |  | Ploy1d(fp1) | f1 | | 서울시 안심이 CCTV 연계 현황 | Fit\_transform(CCTV지역별설치개수) | Scaled\_CCTV\_COUNT | | | [표3]서울시에서 제공하는 데이터 | [표4] 파생변수 목록 |  |  |  | | --- | --- | |  | 파생변수 생성 | | |  |  | | --- | --- | | 데이터 명 | 수집 방법 | | 서울시 5대 범죄 발생 현황 | 서울 열린 데이터 광장 | | 서울시 안심이 CCTV 연계 현황 | 서울 열린 데이터 광장 | | |  |  |  | | --- | --- | --- | | 원본 데이터 | 파생변수 산출방법 | 파생변수 명 | | 서울시 5대 범죄 발생 현황, 서울시 안심이 CCTV 연계 현황 | crime\_anal\_norm[‘범죄’]- Scaled\_CCTV\_COUNT | 위험지수 | | | [표5]서울시에서 제공하는 데이터 | [표6] 파생변수 목록 |   **○데이터 수집 및 전처리**   |  |  | | --- | --- | | 서울시 안심이 CCTV 연계 현황의 데이터 | 파생변수 생성 | | |  |  | | --- | --- | | 데이터 명 | 수집 방법 | | 서울시 안심이 CCTV 연계 현황 | 서울 열린 데이터 광장에서 데이터 다운 | | |  |  |  | | --- | --- | --- | | 원본 데이터 | 파생변수 산출 방법 | 파생 변수 명 | | 서울시 안심이 CCTV 연계 현황의 데이터 | Cctv위치 데이터중 위도 대입 | lat | | 서울시 안심이 CCTV 연계 현황의 데이터 | Cctv위치 데이터중 경도 대입 | long | | 서울시 안심이 CCTV 연계 현황의 데이터 | 새로운 CCTV 데이터와 기존 CCTV 데이터 사이 거리 | distance | | 서울시 안심이 CCTV 연계 현황의 데이터 | 입력 받은 위치의 위도 | user-latitude | | 서울시 안심이 CCTV 연계 현황의 데이터 | 입력 받은 위치의 경도 | user-longitude | | | [표7] 서울 열린 데이터 광장에서 제공하는 데이터 | [표8] 파생변수 목록 | | | | | |
| 최종결과,  활용방안  및 기대효과 | |  |  | | --- | --- | |  |  | | 구별 5대 범죄검거 비율의 색으로 구분되는 정규화 분포 | 구별 5대 범죄 비율의 색으로 구분되는 정규화 분포 |   그림입니다.   |  |  | | --- | --- | |  | | | 지역별 CCTV의 개수 | 지역별 인구당 CCTV |   그림입니다.   |  |  | | --- | --- | |  |  | | CCTV 데이터 선형회귀 분석 | 스케일링한 CCTV 데이터 산점도 시각화 |  |  | | --- | |  | | 위험지수=Crime(범죄율)점수 - CCTV 점수 | |  | | 도출된 위험지수를 자치구별로 정렬 |  |  | | --- | |  | | 최종 예시 : 원하는 쉐어하우스의 위치를 입력할 경우 반경 500m 이내의 CCTV 개수를 반환한다. 위의 예시에서는 총 81개의 CCTV개가 반경 500m 이내에 있다고 알려주면서 해당 지역이 25번째 구별 중 23번째로 안전하다고 안내해준다. |  1. 쉐어하우스 입주를 결정할 때 해당 쉐어하우스 주변에 CCTV 개수를 반환하여 쉐어하우스 입주에 도움이 될 수 있다. 이를 통한 20~30대 여성과 20~30대 여성들의 부모님의 걱정을 덜 수 있다. 2. 쉐어하우스 문의 혹은 쉐어하우스 입주 후 같은 쉐어하우스 사람들에게 개인정보 노출이 꺼려질 때 익명 채팅을 통한 개인정보 노출 최소화 시킬 수 있다. 3. 20~30대 여성 쉐어하우스 입주자들만을 위한 폐쇄형 커뮤니티를 통해 1인 가구의 문제점 중 하나인 외로움 해결 가능하다. | | | | |
| 분석툴, 참고문헌 | ○ **분석툴** : Python, Pandas, Numpy, Tableau, Matplotlib, Scikit-Learn(sklearn), Seaborn, Requests, Folluim, Geopy  ○ **참고문헌**  빅 데이터 캠퍼스 데이터, 「서울시 행정동 단위 거주인구 데이터」, 2023  서울 열린 데이터 광장, 「서울시 주민등록인구(구별)통계」, 2023  R-ONE 부동산 통계뷰어, 「월세수급동향(종합주택유형) 」, 2023  R-ONE 부동산 통계뷰어, 「종합주택 평균 월세가격」, 2023  R-ONE 부동산 통계뷰어, 「월별 매입자연령대별」, 2023  네이버 데이터랩, 「여성 호신용품 검색량」, 2023  서울 열린 데이터 광장 , 「서울시 자치구 연도별 방범용 CCTV 운영 현황」, 2023  서울 열린 데이터 광장 , 「서울시 안심이 CCTV 연계 현황」, 2023  서울 열린 데이터 광장, 「서울시 5대 범죄 발생현황 통계」, 2023  서울 열린 데이터 광장 , 「혼자 생활하면서 곤란한 점」, 2023  서울 열린 데이터 광장, 「서울시 1인가구(연령별) 통계」, 2023  Science on, 「CCTV 설치가 범죄 발생률과 검거율에 미치는 영향」 : 경산시 영남대 지역을 중심으로」  ○ **참고기사**  1. “외출하기 겁난다. ‘묻지마 범죄’에 대한 호신용품 수요 증가’, 2023.08.06  2. “부산 시내 한복판서 묻지마 폭행, 피해자 의식불명”, 2023.10.27 | | | | |