[서비스 산업 데이터를 활용한 빅데이터 분석]

비정형 데이터 활용 분석 프로젝트 기획안

**기획안 작성일자 : 2022. 03. 29**

|  |  |
| --- | --- |
| 조 명 | OKKK |
| 프로젝트 주제  및 개요 | **강서구 1인가구를 위한 거주지역추천** |
| 프로젝트수행  방향 | **● API활용 강서구 각 행정동 별 관련 정보 제공**  ▪ 전국적으로 1인 가구의 비율은 매년 지속해서 늘어가는 추세이며, ‘2021 강서구 사회조사 보고서’에 따르면, 강서구의 전체 가구 중 1인 가구 비율은 전체 서울시 중 2위로 나타난다. 이러한 가구형태의 변화과정에서 ‘1인 가구가 살기 좋은 강서구’ 로 브랜딩하여 거주환경이 좋은 곳을 추천해 줌과 동시에, 나아가 강서구를 1인 가구 살기 좋은 환경을 만드는 것을 목적으로 한다.  **● 데이터소개**  ▪ 2020년~2021년 인구 및 가구현황DB, 강서구 테마지도 카테고리 크롤링 및 소상공인진흥공단 상가(상권) DB 등을 활용, 강서구 행정동 별로 약 40여 개 항목에 대한 자료를 추출하여 데이터를 구축했다.  ▪ SOC시설 및 상권 데이터는 행정동별 거주환경에 대한 가중치 부여 및 추후 주거 환경 선호도에 대한 조사를 통해 강서구 내 균형 발전과 신규 진입 및 거주 인구에 대한 지역추천에 활용한다.   |  |  |  |  | | --- | --- | --- | --- | | 유형 / 구분 | 변수명 | 활용 내용 및 변수 정의 | 활용자료명 | |  | 1인 가구 | 1. 행정동 기준 1인 가구  2. 행정동 기준 성별 1인가구  3. 행정동 기준 연령별 1인가구 | 인구주택총조사  SGIS통계지리정보서비스 | | **공공시설<10개 변수>** | | | | | 환경 및  문화시설 | 체육시설 | 각 행정동 체육시설의 위치 및 수 | 강서구 테마지도 | | 도서관 | 각 행정동 도서관의 위치 및 수 | 강서구 테마지도 | | 문화시설 | 각 행정동 문화시설의 위치 및 수 | 강서구 테마지도 | | 공원 | 각 행정동 공원의 위치 및 수 | 1. 강서구 테마지도  2. 서울시 주요 공원현황  (서울 열린데이터 광장) | | 복지시설 | 장애인 복지시설 | 장애인복지시설 위치 및 개수 | 강서구 테마지도 | | 어르신 돌봄시설 | 어르신 돌봄 시설 위치 및 개수 | 강서구 테마지도 | | 경로당 | 경로당 위치 및 개수 | 강서구 경로당 현황 및 지원사항 - | | 사회복지시설 | 위치데이터 및 시설 수 | 강서구 사회복지시설 목록  (서울 열린데이터 광장) | | 공공  지원시설 | 구청, 동주민센터, 보건소 | 각 행정기관 위치 데이터 및 수 | 강서구 테마지도 | | 공영주차장 | 위치데이터 및 수 | 강서구 테마지도 | | **상권시설<9개 변수>** | | | | |  | 전통시장 | 업종별 지역 내  위치데이터 및 사업체 수 | 소상공인시장진흥공단  상가(상권) DB | | 백화점 | | 대형마트&슈퍼마켓 | | 편의점 | | 세탁소 | | 식당 | | 카페 | | 헬스클럽 | | 영화관 | | **대중교통 및 공공자전거 대여소<2개 변수>** | | | | |  | 버스정류장 | 지역 내 버스정류장 위치데이터 | 서울시 버스정류소 좌표데이터  (서울 열린데이터 광장) | | 공공자전거대여소 | 공공자전거 대여소 위치 및 수 | 강서구 공공자전거 대여소 등록 현황  (서울 열린데이터 광장) | | **생활안전 및 건강<5개 변수>** | | | | |  | 파출소 | 지역 내 위치데이터 및 수 | 서울시 시내 주요 기관 통계  (서울시 열린데이터 광장) | | 여성  안심 택배함 | 강서구 테마지도 | | 심리상담센터 | 강서구 테마지도 | | 병원 | 서울특별시 병원 인허가 정보  (서울시 열린데이터 광장) | | CCTV | 강서구 CCTV(안심이) 설치 현황  (서울시 열린데이터 광장) | | **반려동물 관련 데이터<3개 변수>** | | | | |  | 강아지 등록 수 | 강아지 등록 현황 | 서울특별시 강서구 반려동물 등록 현황  (서울시 열린데이터 광장) | | 고양이 등록 수 | 고양이 등록 현황 | | 동물병원 | 행정동별 동물병원  위치 및 사업체 수 | 서울특별시 동물병원 인허가 정보  (서울시 열린 데이터 광장) |   **● 수행도구**  ▪ Python (Pandas, NumPy, Folium)  ▪ Google Maps Platform (https://developers.google.com/)  ▪ Kakao Developers (https://developers.kakao.com/)  ▪ R (psych, PerformanceAnalytics, corrgram)  ▪ Excel  **● 1인가구 데이터 간 상관관계 분석**  ▪ 강서구의 전체 1인 가구 수와 수집한 항목들의 상호관련성을 파악하기 위해 상관 분석을 진행하였다.    ▲ pearson correlation matrix  ▪ 편의점, 식당, 대형마트와 슈퍼마켓, 카페 등의 상권 항목들은 1인 가구 수와 대체로 높은 상관관계를 보였으며 특히 20세~34세 사이의 청년층 1인 가구에서 이러한 특징이 두드러졌다. 심리상담센터는 청년층 1인 가구에서 높은 상관관계를 보였으며 노년층으로 갈수록 관련도가 급격히 떨어지는 것을 볼 수 있다.  ▪ 1인가구는 정신적으로 다인 가구보다 취약한 환경이며, 정신 건강이라는 측면에서 1인가구와 심리상담센터가 보이는 높은 상관계수는 시사점이 있고, 심리상담센터의 주 이용층이 1인 가구임을 추측해 볼 수 있다.  **● 1인 가구 지역추천**    ▪ 데이터 스케일링  데이터 스케일링을 통해 수집한 데이터의 범위와 분포를 동일하게 설정하였다. 또한 음수가 없는 수집된 데이터의 특성상 최솟값이 0으로 스케일링 되기 때문에 이후 가중치 행렬을 이용한 알고리즘 적용이 불가능하다. 때문에 중윗값과 사분위 수를 이용하여 극단치의 영향을 최소화할 수 있는 robust scaling 기법을 사용하였다.  **● 가중치 설정**  ▪ 통계청, 『2021 통계로 보는 1인가구』 <1인가구의 어려움>정보를 통해 각 시설 선호도 파악    ▲<2021 통계로 보는 1인가구, 통계청>  (1) 성별 및 연령  ▪ 남성 1인가구는 생활•가사영역에 필요성을 요구하는 시설에 가중치를 부여하였으며, 여성은 안전과 건강 영역에서 남성보다 관련 시설 필요성이 높아 CCTV, 경찰서 및 여성안심택배함 항목에 높은 가중치를 두어 결과를 산출해 보았다.  ▪ 청년 1인가구는 편의점, 식당, 세탁소 등 생활 편의시설에 가중치를 높게 두었고 노년 1인가구는 경로당, 복지시설 등에 높은 가중치를 두었다.  ▪ 위의 항목을 바탕으로 연령별•성별 선호도를 파악하여 강서구 내에서 살기 좋은 동 추천을 진행하였다.  (2) 개인화 가중치  임의의 가중치 부여를 통해 개인의 특성에 맞는 지역을 추천하는 것에 활용하였다. 이 경우, 전세를 살며 차량을 소유하고 건강에 관심이 있는 중년층을 가정하여 가중치 부여했다.    ▲변수 별 가중치 부여 결과  **● 추천지역 선정 결과**      ▪ 결과 해석  (1) 여성 청년 1인 가구 추천지역  마곡지구가 자리잡고 있는 가양1동이 1순위로 추천되었으며, 생활편의시설과 상권이 밀집된 지역이 주로 추천되었다. 또 다른 특징으로는 버스정류장, 자전거 대여소와 같은 대중교통이 활성화되어 있는데, 이는 본인 소유의 차량이 없는 청년을 기준으로 가중치를 높게 부여한 결과가 반영되었다. 결과적으로 이러한 지역이 타 지역과 큰 차이를 보이며 우선적으로 추천되었다.  (2) 노인 1인 가구 추천지역  전반적으로 복지 및 사회기반시설에 상대적으로 높은 가중치를 부여했다. 그 결과 어르신 돌봄 시설, 경로당 및 사회복지시설 등의 분포가 높게 되어 있는 발산1동과 방화1, 2동이 추천되었다.  (3) 개인화 추천지역  가중치 적용 과정에서 경제적으로 안정된 중년층으로 상정하였다. 백화점, 세탁소 등 개인의 편의성에 관련된 업종에 높은 가중치를 부여하였고, 반려동물을 기른다고 추가로 가정하여 강아지 등록 수와 동물병원에 높은 가중치를 부여했다. 이에 따라 관련 점수가 높은 화곡1동이 우선되어 추천되었으며 이어 가양1동, 염창동 순으로 추천되었다.  ▪ 시사점  노인 1인 가구 추천지역 결과 추천지역은 발산1동과 방화1, 2동 순으로 나타난다. 반면 통계청의 실제 노인 1인가구의 거주 지역 순위에 의하면 등촌3동, 가양2동, 화곡1동이 전체 30%를 차지하며, 발산 1동의 경우 12위에 불과하다. 이렇게 추천 결과가 상반되게 나타나는 이유는 행정동의 고령자 이용률이 높다고 예상되는 시설을 위주로 가중치를 설정했기 때문이다. 가양2동과 화곡1동의 경우 많은 고령 1인가구가 거주하지만, 경로당과 어르신 돌봄 시설 등의 수가 상대적으로 적게 존재하는데, 강서구 홈페이지에서 수집한 결과에 의하면 가양2동과 화곡1동의 경로당 및 어르신 돌봄 시설의 수는 각 6곳, 7곳에 불과하며, 이는 높은 순위로 추천된 방화1동과 비교했을 때 약 4배 가까이 차이를 보인다.  이러한 생활기반시설의 불균형은 고령자뿐만 아니라 전 연령대에 유사하게 나타난다. 실제 가장 많은 1인가구가 거주하는 가양1동과 화곡1동 일대 역시 공원, 도서관 등 공공시설의 수는 타 행정동에 비해 유사하거나 적게 나타난다. 강서구를 1인가구가 살기 좋은 지역으로 브랜딩 하기 위해서는 상업 시설 외에도 아니라 이러한 공공생활 시설을 인구수를 비례하여 확충하는 것이 도움이 될 것이라고 생각한다.  그럼에도 불구하고 추천 결과가 완벽하다고는 할 수 없다. 본 분석에서는 강서구 내에서의 시설 개수, 집값 등과 같은 양적변수만을 사용하였다. 편의시설에 대한 실제 접근성, 지역에 대한 구민들의 인식, 만족도와 같이 의사결정에 영향을 끼치는 질적 변수들을 고려하지 못하는 한계가 있다.  **● 활용방안 및 기대효과**  ▪ 결과예시  개인의 라이프 스타일 및 선호에 맞춘 최적의 동네를 추천하며, 강서구 거주를 희망하는 1인가구가 활용할 수 있는 종합정보를 어플리케이션을 통해 제공한다.  **<시나리오 앱>**        **1. 개인의 선호도가 반영된 지역 추천**  ▪ 공공데이터를 기반으로 개인의 선호를 반영한 거주지역을 추천한다.    (1) 각 동의 특징을 확인할 수 있으며, 개인별로 가중치를 높게 부여한 항목들의 실제 위치를 지도 상으로 확인할 수 있다.  (2) 각 공공•상업 시설에 대한 위치•연락처•운영시간 등 다양한 정보를 확인할 수 있다.  **2. 커뮤니티 활성화**    (1) 추천된 지역의 1인가구 정책과 프로그램 리스트를 제공한다.  ▪ 기존 1인 가구 지원센터, 강서구 지역자치회관 및 주거복지센터 등 파편화 되어 있는 정보를 종합하여 제공하며, 다양한 강서구 1인 가구 정책사업을 홍보하는 수단으로 활용할 수 있다.  **3. 개인화 시스템 데이터 축적을 통한 머신러닝 활용**  (1) 머신러닝 학습을 통해 다양한 변수 활용을 통한 추천리스트를 만들 수 있을 것으로 예상한다.  (2) 1인가구가 늘어날 지역 등을 예측하고 그에 맞는 정책 및 시설 확충 계획을 세울 수 있다. |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  조원: 오건재, 전인수, 정은원, 조윤주 |
| 프로젝트  추진 일정 | ● 일정  **▪ 3월1일 ~ 3월 2일:** 주제선정 및 일정 수립  **▪ 3월3일 ~ 3월8일:** 데이터수집 및 주간 회의  **▪ 3월 9일 ~ 3월 15일:** 데이터셋 정리 및 데이터 분석  **▪ 3월 16일 ~ 3월 23일:** 데이터 분석 및 결과보고서 작성  **▪ 3월 24일:** 최종 보고서 제출 |