**빅데이터 활용 프로젝트 기획안**

**2023년 6 월 28 일**

|  |  |
| --- | --- |
| **프로젝트 조** | **이장우 조** |
| **프로젝트 구성원** | **장서영, 장예림, 이채린, 우성진** |
| **프로젝트 주제** | 서울시 골목상권 매출에 영향을 주는 요인과 총 매출 금액 예측 |
| **프로젝트 수행 방향 및 필수 기능**  **(주요 기능 설명)** | **주요 분석 내용**  서울시 골목상권 총 매출액에 영향을 주는 요인이 무엇인지 확인,  서울시 골목상권 총 매출 금액 예측  **1) 종속변수**  **-** 상권\_추정매출  ->서울 시 골목상권\_추정매출  ( 골목상권별 분기당 매출금액 총합 )  **2) 독립변수**  - 상권\_소득소비  -> 식료품 지출 총 비용  -> 교통비 지출 총 비용  - 상권\_생활인구  -> 시간대 (밤새벽,아침점심,오후저녁) > 3개  - 상권\_점포  -> 프렌차이즈 점포 수  -> 총 점포 수  -> 개업/폐업 점포 수 > 2개  -상권\_추정매출  -> 매출 건 수  **예측 기대효과**  1) 서울시 골목상권 총매출액에 영향이 있는 요인들 중 큰 영향을 주는 요인은 교통\_지출\_총금액,식료품\_지출\_총금액이다.  -> 매출액을 상승시키기 위해서는 양의 결과를 갖는  교통이 활발한곳, 점포수가 많은곳, 아침점심 생활인구수가 많은 상권을 선택하는것이 좋다.  2) 상위 25%인 서울시 골목상권은 교통 지출 총 금액에 많은 영향을 받았으므로, 교통이 활발하고 교통편이 좋은 구역임을 추측할 수 있다. 오후저녁보다는 아침점심에 집중해서 운영하는 것을 추천.  3) 하위 25%인 서울시 골목상권은 운영 시간을 아침점심을 피하고 오후저녁에 집중해서 운영하고, 프렌차이즈보다는 개인 사업을 추천.  **데이터 및 관련 기술설명**  **- 사용 데이터** : <http://data.seoul.go.kr/dataList/3/literacyView.do>  서울시 상권분석 서비스  **- 관련 기술**  1) 선형회귀분석  2) 고급회귀분석  3) 의사결정나무  4) 랜덤포레스트  5) lightGBM  6) 군집분석  7) 상관분석 |
|  |
| **프로젝트 수행 도구 및 포함 기술** | **1) 고급회귀분석**  -Ridge  -Lasso  -LinearRegression  - 하이퍼파라미터 튜닝  **2) 시각화**  - 파이차트  - 막대그래프  - 히트맵  - 산점도  - 선형식  - gapminder  **3) 전처리**  - 로그변환  - 표준화  - 스케일링 |
|  |  |

**□ WBS 첨부**