**상권 분석 보고서**

**전처리 데이터 분석**

**restaurant\_with\_universities.csv 분석**

* hdfs에서 파일을 읽으면 대학교 column에 두 개의 값이 들어있는 경우 "경희대학교, 한국외국어대학교" 형식으로 들어있어 미리 split했음
* trans\_split\_univ.csv

**analysisData.csv 만들기**

* trans\_split\_univ.csv와 university\_info.csv를 hive를 사용해 join 및 변형
* analysisData.csv

**analysisData 분석**

텍스트, 폰트, 영수증, 스크린샷이(가) 표시된 사진

자동 생성된 설명

* analysisData.csv
* 데이터 각 column 데이터 개수 확인

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

* ‘대학명’: 46개, ‘상권업종중분류명’: 10개, ‘상권업종소분류명’: 43개
* 라인, 스크린샷, 그래프, 도표이(가) 표시된 사진

  자동 생성된 설명텍스트, 스크린샷, 디스플레이, 도표이(가) 표시된 사진

  자동 생성된 설명명지대 상권분포 plot
* 라인, 도표, 스크린샷, 그래프이(가) 표시된 사진

  자동 생성된 설명대학별 카페 개수
* 전체 가게 소분류명별 개수

스크린샷, 텍스트, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

* 개수가 가장 많은 카페와 백반/한정식 대학별 비율

스크린샷, 그래프, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

* 삼육대의 비율이 이상하여 삼육대의 전체 소분류명 개수 plot 그려봄

스크린샷, 직사각형, 텍스트, 그래프이(가) 표시된 사진

자동 생성된 설명

* 카페와 백반/한정식 밖에 없음
* 텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

  자동 생성된 설명모든 대학의 중분류명별 개수 plot 그림
* 텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

  자동 생성된 설명모든 대학의 소분류명별 개수 plot 그림
* 모든 대학의 상위 15개의 소분류명 개수 plot 그림

텍스트, 스크린샷, 폰트, 디자인이(가) 표시된 사진

자동 생성된 설명

* 하위 15개는 보고서에서 제외.
* 대학별 소분류명의 개수가 많은 순으로 정렬

텍스트, 문서, 화이트, 패턴이(가) 표시된 사진

자동 생성된 설명

* 명지전문대를 제외한 모든 대학의 1등과 2등이 카페와 백반/한정식임
* 각 중분류명에 속하는 소분류명의 비율

텍스트, 스크린샷, 폰트, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

**재학생을 반경넓이로 나눠 정규화한 밀도를 이용해 분석**

* 밀도 0 방지: 최소값 추가
* 개수 / 밀도 계산, 로그 변환, z-score 표준화 사용

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

* 텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진

  자동 생성된 설명텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

  자동 생성된 설명Plot으로 시각화해보니 로그가 가장 적절해보임

Z-score log

* 종소분류명별 평규을 구하고 평균보다 큰 대학교 plot

텍스트, 스크린샷, 폰트, 디자인이(가) 표시된 사진

자동 생성된 설명

* 가톨릭대학교 \_제3캠퍼스를 제외하고 다시 그림.

텍스트, 스크린샷, 폰트, 평행이(가) 표시된 사진

자동 생성된 설명

* 종소분류명별 평규을 구하고 평균보다 작은 대학교 plot

텍스트, 스크린샷, 평행, 폰트이(가) 표시된 사진

자동 생성된 설명

* 대학별 로그밀도 높은 상위 15개의 소분류명

텍스트, 스크린샷, 폰트, 도표이(가) 표시된 사진

자동 생성된 설명

* 대학별 로그밀도 낮은 하위 15개의 소분류명

텍스트, 스크린샷, 폰트, 평행이(가) 표시된 사진

자동 생성된 설명