

KU-BIG

공공분야 빅데이터 분석

프로젝트 아웃라인

ABOUT US



통계학과 허찬



통계학과 이정진



보건정책관리학부 이나영

ABOUT PROJECT

목표

다양한 종류의 공공 빅데이터의 전처리/분석/시각화를 통해서
사회문제에 대한 인사이트를 얻고, 해결책 모색

데이터

서울시 빅데이터 캠퍼스, 서울 열린데이터 광장, 공공 데이터 포털,
AI HUB 코로나 데이터 등

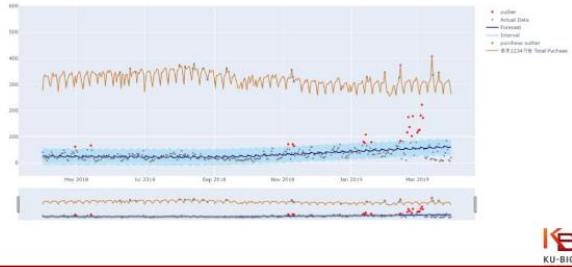
분석도구

PYTHON, R, EXCEL 등

EXAMPLE(미세먼지 프로젝트)

1. ANOVA

- 시각화를 통해 overview
- ANOVA
- 정규성, 동질성, 등분산성 검정
- 종로 1,2,3,4동



KS
KU-BIG

분석

시계열 및 시각화 툴 사용

“미세먼지, 카드매출, 유동인구, 유통 데이터” 분석

2. INSIGHT

- 1) 45세 이후로 미세먼지에 둔감하다
- 2) 미세먼지 농도에 따라 의복 구매에 차이가 있다
- 3) 미세먼지 농도에 따라 40대 여성의 유통 소비에 차이가 있다

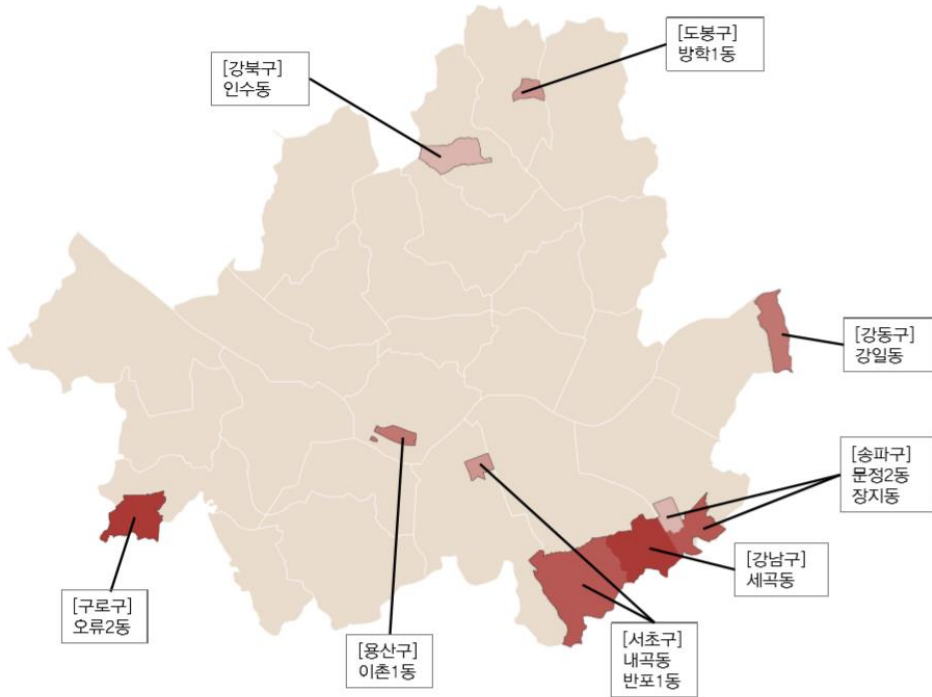
KS
KU-BIG

인사이트

미세먼지에 관한 사회적 인사이트 도출

EXAMPLE(국공립 유치원 분석)

국공립유치원 분석 : 분석결과



분석

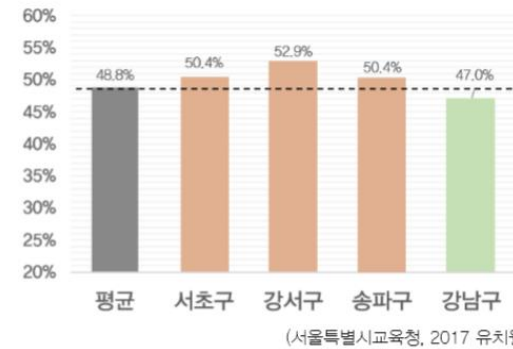
국공립 유치원 설립 시급지역 도출

국공립유치원 분석 : 분석결과

[국공립유치원 설립 시급지역에 고소득층 밀집 지역이 선정된 결과에 대해]

Q. 소득수준이 높은 지역은 사립을 더 선호하지 않을까?

:서울시 평균 국공립유치원 취학수요는 48.8% 이며
강남구 47.0% / 서초구 50.4% / 송파구 50.4% / 강서구 52.9%
강남구를 제외한 3개의 구는 평균보다 높고
강남구도 평균에서 크게 떨어지지 않아
구별 수요의 편차가 많지 않은 편으로 확인됨.
따라서 국공립 유치원 자체에 대한 선호는 소득수준과 관계 없음



인사이트

국공립 유치원 자체에 대한 선호는

소득수준과 관계없다는 인사이트 도출

ABOUT DATA

인증이 완료되었습니다.

사용자 인증이 완료되었습니다.

하단에 표시된 로그인 계정 정보를 통하여 서울시 빅데이터 캠퍼스 이용이 가능합니다.

인증 상태

이름 허찬

인증상태 인증

모델링 제공 데이터 안내

▶ 로밍 데이터

- KT 로밍 이용고객 전체 데이터를 활용하여 생성한 통계 데이터로, 체류 국가의 입국/출국 날짜 및 여행자수 정보 제공

※ 데이터 구조

컬럼명	컬럼설명	샘플
return	로밍 여행자가 한국에 귀국한 날짜	20200128
iso	2자리 국가코드	cn
arrival	iso 국가에 도착한 날짜	20200123
departure	iso 국가에서 떠난 날짜	20200128
count	return 일자에 귀국한 로밍 고객 중 iso국가에 arrival에 도착해서 departure에 떠난 고객 수	161

▶ 감염병 기사 데이터

- 해외 감염병 발병 기사를 수집한 데이터로 WHO, CDC 등 주요기관에서 감염병 발생 확인을 위해 사용하는 데이터와 동일한 수준의 데이터

데이터셋

제공신청

데이터셋

검색어를 입력하세요

🔍 찾고 싶은 데이터를 입력해 주세요.

전체(34,294)

파일데이터(30,816)

전체 34,294건을 찾았습니다.

데이터셋

전체 5,243건을 찾았습니다.

방대한 양의 데이터'는' 확보!!

♥ I WANT YOU ♥



1. 리얼 데이터 분석을 하고 싶으신 분!

정제되지 않은 데이터에
전처리/모델링/분석/시각화까지 모든 통계 기법을
적용해 보고 싶으신 분들 대환영!!

2. 데이터 인사이트 도출에 관심 있으신 분!

데이터 관련 분야에서 가장 핫한 인사이트 도출..!
데이터 분석을 넘어 효과적인 인사이트 도출을
저희와 한학기 동안 같이하실 분들을 찾습니다...ㅎ

Q&A
