

데이터 과학

L00: Orientation

Kookmin University

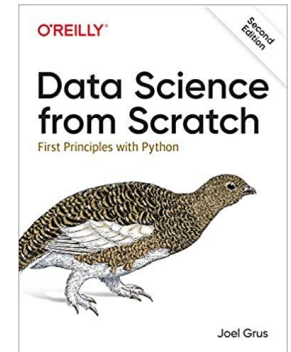
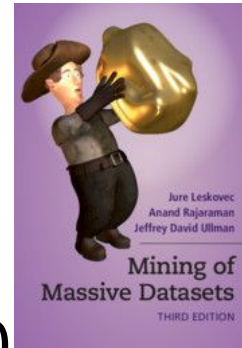
교수 소개

- 이름: 박하명
- Office: 미래관 403호
- 이메일: hmpark@kookmin.ac.kr
- 홈페이지: <http://hmpark.me>
- 데이터마이닝 연구실: <https://kmudmlab.github.io>
 - 데이터마이닝 알고리즘 최적화 (kNN Search 등)
 - 빅데이터 처리 및 분석
 - 분산 컴퓨팅 (Hadoop, Spark 등)
 - 그래프 알고리즘 (연결요소 찾기, 삼각형 찾기 등)
 - 그래프 뉴럴 네트워크

강의 개요

- 강의 시간

- 1분반: 월/수 9:00 ~ 10:15
- 2분반: 월/수 10:30 ~ 11:45
- Office Hours: 월/수 13:00 ~ 14:30



- 강의 홈페이지

- eCampus 활용 (공지, 수업자료, 과제 등)

- 교재

- (교재1) Mining of Massive Datasets 3/E, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2020
- (교재2) Data Science from Scratch 2/E, Joel Grus, O'REILLY, 2019

강의 개요

- **점수 배점 (변경 가능)**
 - 중간고사 또는 중간 프로젝트 20%
 - 기말고사 또는 기말 프로젝트 20%
 - 과제 50%
 - 출석 10%

과제 부정행위 발각시 해당과제 포함 모든 과제 0점 처리

*** 코드 유사도 검사 프로그램 활용 예정 ***

강의 일정 (변경 가능)

주차	내용
1주차	강의소개, 데이터시각화
2주차	기초선형대수, 기초확률통계
3주차	회귀 (Regression) 1
4주차	회귀 (Regression) 2
5주차	비슷한 데이터 찾기 1
6주차	비슷한 데이터 찾기 2
7주차	중간고사

주차	내용
8주차	추천시스템 1
9주차	추천시스템 2
10주차	군집화 (Clustering)
11주차	k-Nearest Neighbors
12주차	차원축소 (PCA)
13주차	네트워크 분석 1
14주차	네트워크 분석 2
15주차	기말고사

Overview

❖ 데이터 과학 사례

데이터 과학자

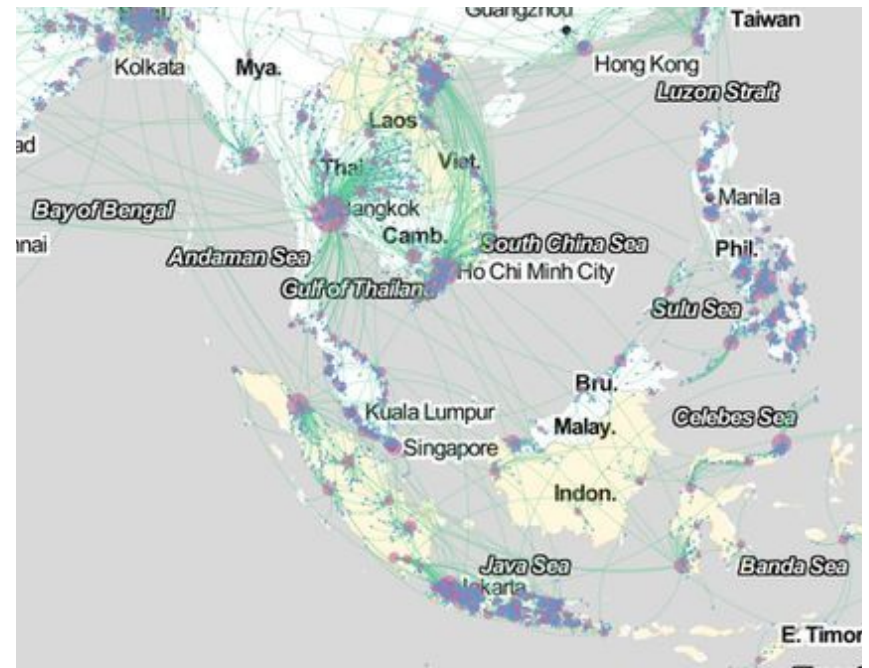
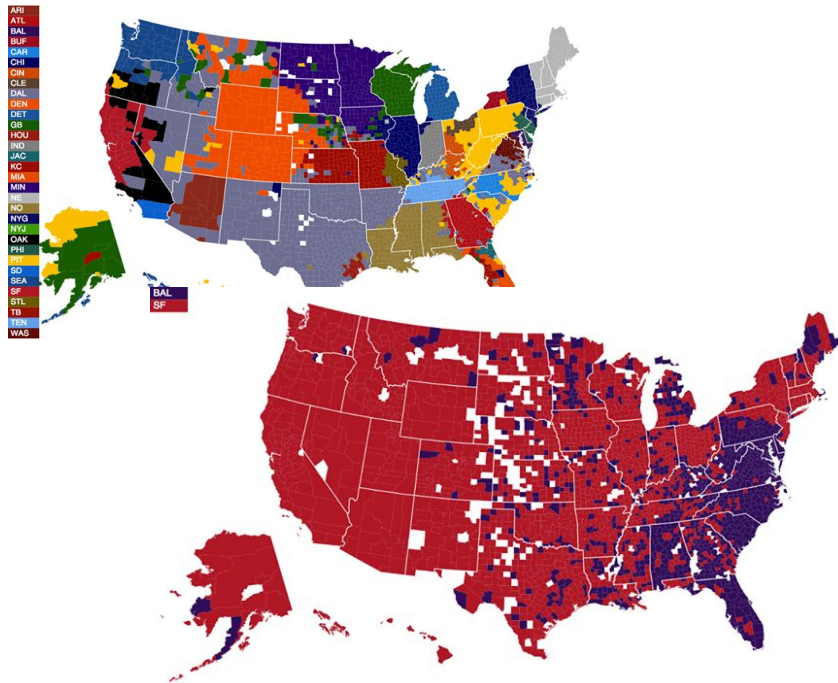
- 데이터 과학자?

“데이터로부터 통찰(Insight)을 발견하는 사람”



데이터 과학 활용 사례

Facebook은 사용자 위치정보를 이용하여 ‘국제 이주 패턴’
이나 ‘미식축구팀 팬들의 거주 지역’ 등을 분석함



<http://on.fb.me/1EQTq3A>

<http://on.fb.me/1EQTvn0>

데이터 과학 활용 사례

2016년 트럼프 대선 캠프는 온라인 광고 데이터를 분석하여 좋은 캠페인 광고를 선별하여 대선을 승리로 이끔

Coby says, the campaign was running 40,000 to 50,000 variants of its ads, testing how they performed in different formats, with subtitles and without, and static versus video, among other small differences. ... The more variations the team was able to produce, Coby says, the higher the likelihood that its ads would actually be served to Facebook users.



<https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news/>

데이터 과학 활용 사례

Target(미국의 마트)은 구매 내역 데이터로부터 임신한 고객을 추정하여 아기용품 광고

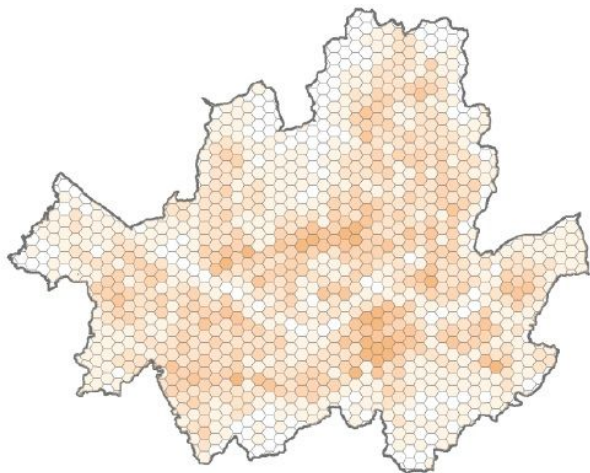


<http://nyti.ms/1EQTznL>

데이터 과학 활용 사례

서울시는 휴대전화 통화 이력 데이터를 활용하여 심야버스 노선을 결정함 → 9개 노선, 하루 평균 이용자 수 6079명

유동인구 밀집도 시각화



노선별 시뮬레이션 및 최적화



최종 심야버스 노선도



<https://seoulsolution.kr/ko/content/빅데이터를-이용한-교통계획-심야버스와-사고줄이기>

Questions?