



API는 내가 돌릴게, **Geojson**은 누가 할래? : 난 갈래

공공데이터 부수려다
우리가 부서질뻔한 이야기



정우주 이선영 이수연 윤한솔 한예은 오수민

주니어 데이터 분석가들의 서울시 지진해일 대피소 공공데이터 분석

분석 결과 한눈에 보기

주어진 데이터

1. 2019년 10월
전국지진해일대피소
표준데이터
2. 서울시 유동인구
3. 거주인구 데이터
(빅캠퍼스 반출데이터)

분석 과제

- 기본 분석
Python 의 Pandas, folium,
matplotlib, numpy library
- 추가 분석
서울 지진대피소의 안전성과
접근성 분석

분석 중점

1. 주어진 공공 데이터
기반으로 분석
2. 서울에 지진대피소가 충분한가?
3. 서울의 지진대피소가
접근 가능한가?

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”

가장 큰 에러는 “휴먼 에러”



미쳤습니까
휴먼?

전처리 고군분투기

1. 도로명 주소 결측치와 오류는

소재지번주소 컬럼을 활용하거나
통계청에 **직접 요청**하여 데이터 받음

2. 위도, 경도로 결측치 처리

(KAKAO API와 LOCATIONIQ API 활용)

3. 지진해일대피소 최대수용인원수

오기입 된 정보는 **비슷한 면적인**
대피소 최대수용인원수 평균으로 대체

4. 거리 주소만 남은 정보는 **구글로 검색**

-> 수기로 **정확한 위치** 기입



결론 : 현실의 데이터 분석..!

“

어려운 분석이나 코드를 어떻게 짜지?

”

예측 모델이나 k-means 클러스터링

같은 통계 모델...? 어떻게 해야하지..?



“

1. 데이터를 수집하고 전처리 하는

과정부터 만만치 않다는 것

2. 데이터 분석에 용이할 수 있는 주제를

잘 선정하는 것도 하나의 역량

3. 전처리를 제대로 해야 의미 있는 분석

4. 커뮤니케이션 또한 중요하단 것

”



Q&A

감사합니다!

더 구체적인 분석 결과 보러 오세요 :)

<https://github.com/canary4651/Earthquake-shelters-project>