주졔 : 세종시 교통약자를 위한 저상버스 정류장 추천 시스템 설계

토의할 것

1. 추천시스템을 어떤 식으로 구성할 것인가
2. 고원식 횡단보도 데이터를 어떻게 얻을 것인가

**수요 예측 모델**

x값:위도, 경도, 편의시설, 유동인구, 거주인구, 주변 도로 교통량, 다니는 버스 수

* 편의시설을 어떻게 정의할지
* 편의시설의 위도, 경도를 가져오자(거리안에 들어와있는지)
* 유동인구 데이터(읍,면,동별 인구이동)가 읍단위로 크게크게 되어 있는데, 이것을 정류장 단위로 어떻게 바꿀 것인가?
* 교통량 데이터는 도로기준으로 나와 있음 -> 정류장 단위로 어떻게 바꾸지?

서울시데이터 로 먼저 학습을 시키면 어떨까?

y값: 이용량(승하차인원)

**정류장 추천 시스템**

정류장 선정 기준 : 장애인들이 자주 가는 목적지와 출발지, 장애인 거주 지역, 일반 수요, 도로 형태, 주변 기존 정류장과의 거리,

1. 새로운정류장을 만들고 가장 가까이 있는 정류장을 버스 수를 가져오자
2. 추천시스템 하나를 만들고 기존의 정류장을 추천하자

<질문, 궁금한 점>

Q. 670개 정도로 모델을 돌릴 수 있나?

1. 정은성 교수님 피셜 -> 600개 정도면 그래도 어느정도는 성능이 나올 것 같다, 부족하다 싶으면 서울시 데이터같이 많이 존재하는 데이터로 먼저 학습을 시켜라

<팁>

1. 전체 가능한 데이터를 수집, 동일한 distribution 을 가지는
2. 가능한 데이터에서 프로젝트 목적/목표에 맞는 feature를 선정
3. 개수가 충분한 지 확인, 최소 1천개
4. 2,3을 통과하지 못하면 프로젝트 목적/목표를 수정하고 1,2,3을 다시 수행

<오늘 진행 상황>

1. 세종시 정류장 약 670개의 위도, 경도 데이터와 지나가는 버스 노선 수 데이터 정리
2. 이제 y값이 될 이용량을 구해야 되는데 이부분이 매우 노가다..
3. 장애인 출발지, 목적지 데이터의 위도 경도를 찾는 크롤링은 부분 성공, 나머지는 수작업할 예정