

# 서울시 빅데이터 멘토링 1차 멘토링

승하차 인원 -> 혼잡도를 나타내지 않음

환승에 대한 부분이 굉장히 큼

출퇴근 집중(배차 집중)

그때 지하철이 몇대가 있고, 사람이 몇명이 있느냐에 따라 다름

지하철의 용량, 몇명이 내리고 타는 것에 혼잡도가 결정이 됨

머신러닝을 하려면 혼잡도가 필요한데, 혼잡도에 대한 값을 구하기 쉽지 않음

데이터를 좀 더 보완을 해야할 것 임 (안되는부분에 대한)

파이썬 이용

R 시각화

수송인원 (승하차 뿐만 아니라 다른호선에서 내린 사람들)

19년도에 서울교통공사에서 보정해서 구한 데이터임

19년도 데이터와 매칭을 해서 어떤것들이 영향이 있는지

지하철 마다 사람들이 적고 많음이 다른데 그게 어떤 지표와의 연관성

역과 얼마나 떨어져있느냐에 따라 얼마나 영향을 미치냐

나중에 모델을 만들어서 추후 가상의 역을 만들어서 그 주위에 매장에 어떤 영향이 갈까 예측모델을 만들면 깔끔할 것 같다.

+ 500m 이내 역 주변의 시설들(문화시설, 편의시설, 의료시설, 백화점, 쇼핑몰 등), 버스  
-> 얻을 수 있는 곳까지만

서울시에서  
교통 관련 데이터

분석만 하는 사람  
실제 분석을 하고 그 인사이트 도출 -> 분석 한 사람들이  
그 사업을 정확히 알고 있는 사람이 발표

그 사업, 그 내용을 정확히 아는 사람에게 필요한 부분들을 자료를 제공해줘야  
하는  
시각화 이런것들

분석하는 과정에서 서로 많이 얘기를 해야함  
그분이 뭘 필요로 하는지  
그분이 어떤것들을 보고싶어하는지

내관점이 아니라 어떤것을 보고싶어하는지 상의를 하면서 피드백 하는 과정을  
거치게 된다면 이해할 수 있는 결과

혼자 고민하지말고 단계 단계에 따라서 실제 이용하는 분들과 어느 단계가 지  
나가면 어느 부분을 조정하는 단계가 있음 좋을 것 같다.

보고서 : 내 입장이 아닌 보는 사람 입장에서 쓰면 좋음 (내가 말하려는걸 이해  
할 수 있을까? 하는 것을 시간을 들여서 필요없는 정보 다 쳐내고)

내가 하는 버릇대로 하는 것을 버리고 그때 그때 그 사람에게 맞춰서  
한번더 고민을 해보면 전달력이 더 좋아지는것

교통 관련 분석

geopandas : crs 값 변환