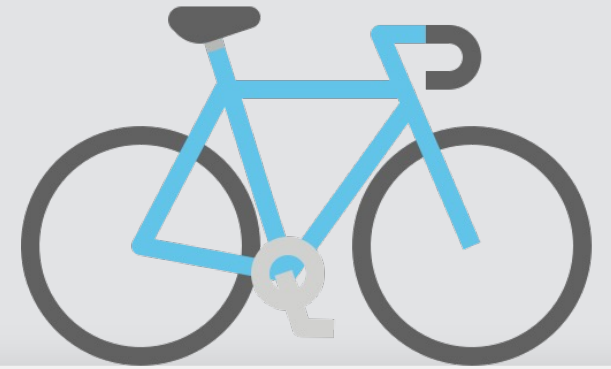


2022 UOS 빅데이터 알고리즘 경진대회



발표자 - 최지혁

신주연

박은비

김진호



재벌집 막내 조



2022 UOS 빅데이터 알고리즘 경진대회

주제: 서울시 지역구별 따릉이 대여량 예측 모델 개발

설명: 2022년의 서울시 지역구별 일별 따릉이 대여량 회귀 예측

데이터: 시계열 데이터

평가 산식: MAE(평균 절대 오차)



데이터 소개

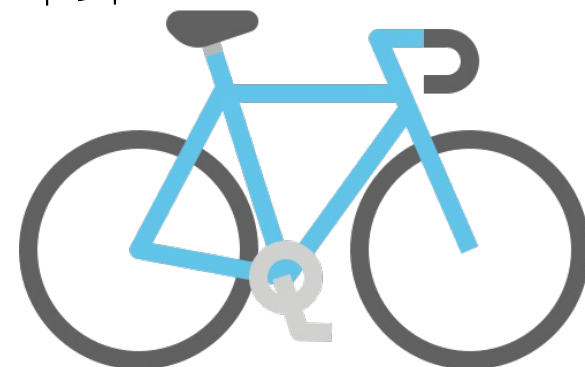
	일시	광진구	동대문구	성동구	종량구
0	20180101	0.592	0.368	0.580	0.162
1	20180102	0.840	0.614	1.034	0.260
2	20180103	0.828	0.576	0.952	0.288
3	20180104	0.792	0.542	0.914	0.292
4	20180105	0.818	0.602	0.994	0.308
5	20180106	0.618	0.522	0.696	0.228
6	20180107	0.598	0.478	0.606	0.238
7	20180108	0.742	0.528	0.848	0.202
8	20180109	0.652	0.488	0.838	0.206
9	20180110	0.500	0.406	0.636	0.160

train.csv[파일]

2018년 01월 01일 ~ 2021년 12월 31일까지의 지역구별 따릉이 대여량

- 일시 : 년월일
- 광진구 : 광진구의 일별 따릉이 대여량 (천명 단위)
- 동대문구 : 동대문구의 일별 따릉이 대여량 (천명 단위)
- 성동구 : 성동구의 일별 따릉이 대여량 (천명 단위)

<<< 간단한 데이터 형식으로 구성
따라서 외부 데이터 활용 예정
예) 날씨, 공휴일, 평균 기온 등 변수 추가





대회선정 이유

1. 시계열 데이터를 핸들링 해볼 수 있다.
2. 금융 데이터와 같이 주제 자체에 대한 심도있는 학습을 필요로 하지 않는다.
3. 다뤄보지 않은 시계열 예측 모델 RNN, GRU, LSTM 등을 손쉽게 적용 가능하다.
4. Prophet과 같은 시계열 예측 라이브러리 등을 사용하면서 새로운 경험을 할 수 있다.
5. Smoothing 기법(데이터 평활 기법) 등 시계열에서만 사용가능한 새로운 데이터 전처리 방법을 배울 수 있다.
6. 외부 데이터를 활용하여 변수 추가 등 아이디어 도출 가능





다음에 할 일!

EDA



데이터 전처리



외부데이터 탐색

