

7월 15일 세미나 정리

예산 집행 보고서 초안 확인

### 1. 크롤링반

서울시 초미세먼지 시간 별 데이터 확인 및 크롤링 중 -> 데이터 6년 + 데이터 양 24배 증가

어려워, 찡동은 공개된 웹페이지 화면에서만 크롤링이 가능하기때문에, 필요한 데이터가 없다고 판단. 크롤링 진행 안하기로 하였음.

서울시 공기오염 데이터 분석 결과, AQI의 98%는 초미세먼지(PM2.5)로 인해서 선정되었음.

PM2.5와 PM10의 AQI값 분석 결과, PM2.5가 두 배 가까이 더 높게 나왔음.

심층신경망(DNN)을 이용하여, PM2.5(1996~2017) 추정. 실측값(2014~2017)과 추정값이 0.95의 상관계수로, 높은 상관관계를 갖고 있는 것으로 나타남. 추정 데이터의 추정오차는 2.699로 나타남.

### 2. 돌고래반

기본 SQL문: 변수의 사용, 데이터 형식과 형 변환, MySQL 내장 함수, 날짜 및 시간 함수, 조인 (join), SQL 프로그래밍, SQL 오류처리

클라우드 서버 : 6개월 60원, 해외에서도 방문 가능. 다음 모임 전까지 구매, 사용가능 여부 확인 및 사용법 숙지

월 데이터와 일 데이터를 비교하는 방법? 일 데이터에 월을 입력하여 관계를 맺어준다

### 3. 무지개반

박스플롯(box plot): 함수 box\_plot(filename), 중앙값, 평균값, 1분위 수(25%), 3분위수(75%), 데이터 전체 범위

상관분석: Pandas 형식 DataFrame (Dictinoary 형식의 리스트 필요)