

## (3주차) 데이터 전처리

✓ 2023.05.08 ~ 05.14 🔗

### ✓ 5.8 기획서 관련 면담 (이상현 교수님)

- 개발 하면서 기획서 내용 보충하기 (API명세와 프론트 페이지 연결해서 추가하기)
- 할 수 있으면 보충 했으면 하는 기능들
  - 추후 어떤 서비스를 더 많이 이용했는지 등 서비스 개선을 위하여
  - 리액트에서 요청했던 정보를 히스토리에 남겨놓기(로그 데이터)
  - 클라이언트가 응답정보 확인하면 신뢰도 체크해서 피드백 할 수 있는 기능

### ✓ 5.8. Data 관련 면담 (한상곤 교수님)

- 전처리 된 데이터 플라토에 업로드
- 폴더 및 파이썬 스크립트, 노트북 파일 정리해서 github 올리기
- 우리가 하고싶은 것 : 시간별로 정체구간 나타내기
- 이번주까지 평균속도 라벨링 하자!

(단위: km/h)

원활	서행	정체	정보없음
25 이상	15 ~ 25	15 미만	-

### ✓ 5.10. 기획서 관련 면담(홍봉희 교수님) → 5/12까지 수정할 것

- 주제 재선정 필요
  - 기존 : 도로정체 예측
  - 변경 : 도로정체 예측 및 차량 운전자 안전운전 판단
- 기획서 내용 수정
  - 내용 전체적으로 개조식으로 수정
  - 선정배경 / 목적 및 필요성 서술
  - 요구사항 분석
  - 전체 일정(상세하게)
  - 구조 → 서비스 설계로 변경
    - 전체 구조도(+기술스펙)
    - 데이터 관련내용
    - API 설계내용
    - 화면 설계내용
    - 각 항목 담당자 더 세분화해서 쓰기

### ✓ 5.11. 전체 Data 라벨링

- 일, 10분단위로 groupby
  - 정보발생일시 10분 단위로 구분
  - 00분, 10분, 20분... 으로 구성된 컬럼 생성 ('10분단위')
  - groupby한 데이터를 mean() 으로 매 10분 마다의 평균속도 추출
- 정체/서행/원활로 라벨링
  - 평균속도를 15km/h 미만, 15km/h 이상 25km/h 미만, 25km/h 로 구분하고 각각에 정체, 서행, 원활로 혼잡도 라벨링
  - 라벨링 후의 데이터 시각화



## 5.12. 회차별 데이터 분석

- 노선
  - 115-1번 버스
  - 중간에 시동이 꺼졌을 때를 판단하여 자름
  - 한 회차당 약 44km
  - 주기가 일정하지 않고, 가스충전소에 다녀오는 듯한 데이터가 포함되어 있어 회차를 구분하기 어려움.
  - 사람별 위험운전행동을 판단하기는 어려워 보이고, 하루당 분석결과를 도출해내는 것이 적당해 보임.
- 홍봉희 교수님, 김태연 교수님 미팅 결과
  - 홍봉희 교수님: 위험운전행동을 분석하고 이상치 탐지를 도출해내라고 하심.
  - anomaly detection 사용하라는 것 같음.
  - 승차감 분석도 해보라고 하셨는데 불가능해보여서 못하겠다고 함.
  - 승차감 분석을 위해서는 최소한 가속도 Z 데이터는 있어야 하는게 아니냐는 것이 우리의 생각.
  - 김태연 교수님: 데이터 라벨링이 어려워보여 clustering 또는 anomaly detection을 해보라고 하심.

## 백엔드 작업

- 스프링부트 3.0.6 → 2.7.11로 버전 다운그레이드해서 프로젝트 재생성

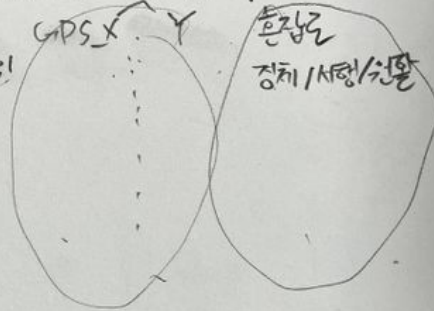
## 평균속도 라벨링

- 데이터 년, 월, 일, 시 로 groupby 해서 평균속도 구하기
- 구간 별 라벨링 하기

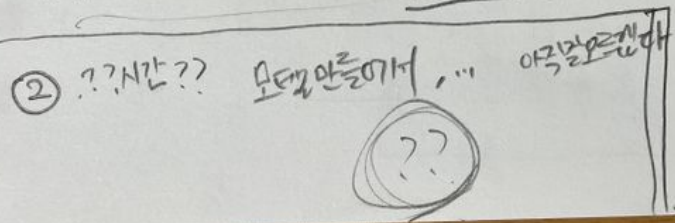
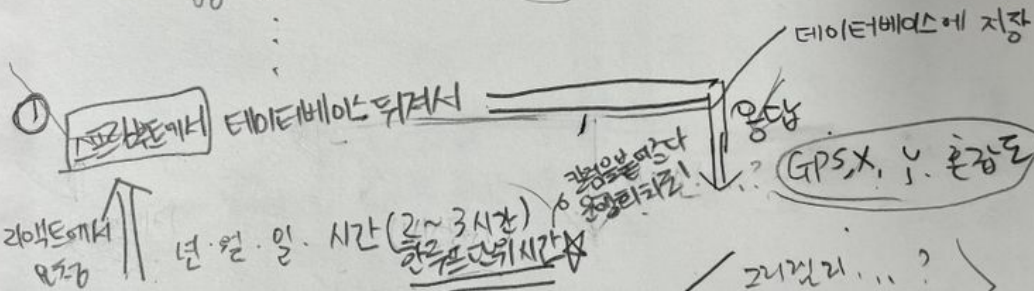
## 기획서 전면 재작성 - 아이템 재선정 및 방향성 관련 회의 (5/10)

약 140만개 : 한차량의 한달지 전체데이터

년 월 일 시 분 평균속도  
 2022 12 1 05 00 00km/h  
 01 00km/h  
 02 00km/h  
 ...  
 59 00km/h  
 06 00 00km/h  
 ...



11



< 거리... ? >

2. 차량 4대

각 차량별 한 회차씩 데이터 권

\* 차량 위치 파악을 위하여

(ii) 속도, 가속도, 방향, 시간, GPS, 주행거리

결과 DB 저장, 차량번호, 날짜 시간, 회차번호, 주행거리, 급가속, 급감속, 급정지, 급가감속

기행선

점속 → 30인 →

년.월.일, 차량번호, 회차번호

시점  
종점

DB 저장

관리자 (인) ~ 배정

2015.10.20

기행선에서 해체되어  
부속으로 분리

기획서 전면 재작성2 - 아이템 재선정 및 방향성 관련 회의 (5/12)

- 한 차당 하루에 얼마나 운행하는지 등 파악하기 위하여
- 정지시간 제외하고 회차별로 분리해보고자 하였으나
- 한 사이클 운행거리(44km) 말고 다른 루트로 가는 (주유? 정비?) 경우도 있어서
- 회차별로 나누기는 쉽지 않아보임
- 우선은 각 차량의 하루 기준 데이터로 분석해보고자함



05:56:44

1854 DAY1

10 07:55:59 (9951)

44km

08:05:07 (10013)

← 이상한데 가는중 (

11:10 10:28:49 (11943)

← 왜가만히있죠?

11:53:11 (12060)

드디어갔네요

44km

3:10 15:10:52 (23192)

2:10 17:33:18 (23491)

← 난또왜??가만히??

1854 DAY2

06:56:01 (0)

44km

10:02:02 (10977)

가만히

10:10:01 (11105)

88km

13:25:05 (21608)

가만히...

13:39:26 (21749)

132km

17:13:24 (33117)

176km

20:39:07 (43752)

220km

22:43:19