

21년 8월 25일 수

---

팀원: DE: 김민형, 이현범(팀장), 한유정

DS: 강주영, 고아름

---

- 과제 이해

기업요구사항 = 고객인터뷰 통한 니즈파악 (근데 그게 없음)

요구사항을 가상으로 만들어야 함. 과제 선정→요구사항 제시

1. (뉴스, 보도자료, 통계자료 → 실태 유추, anonymous: 불특정다수)
2. 우리가 고객 돼서.

- 주제 논의

1. 수산물 가격; 인어교주 해적단
2. 충전차: 수소차, 전기차 (수소차 데이터, 현대제철, 충전소 위치)

(충전소현황데이터, 서울빅데이터: 자동차 등록지수? 세대수, 주차장 개수, 면수, 인구 밀도: 공공데이터에 찾아보는 걸로)

3. 장애인 콜택시 (서울시 공공데이터(O), 운행일자, 시군구동-목적지, 건수  
→ 고객의 의견)

- 각 주제별 데이터 확인 후 내일 중 주제 선정 예정

- 주제 논의

수산물 시세, 수소차 충전소 입지, 장애인 콜택시

→ 수소차 충전소 입지, 장애인 콜택시(API 요청 후 대기 중)

- 수소차 충전소

- 수소차 충전소 관련 고려할 사항

1. 입지조건(feature 선정 중요): 선행연구 참조

- 가스충전소 위치 참고

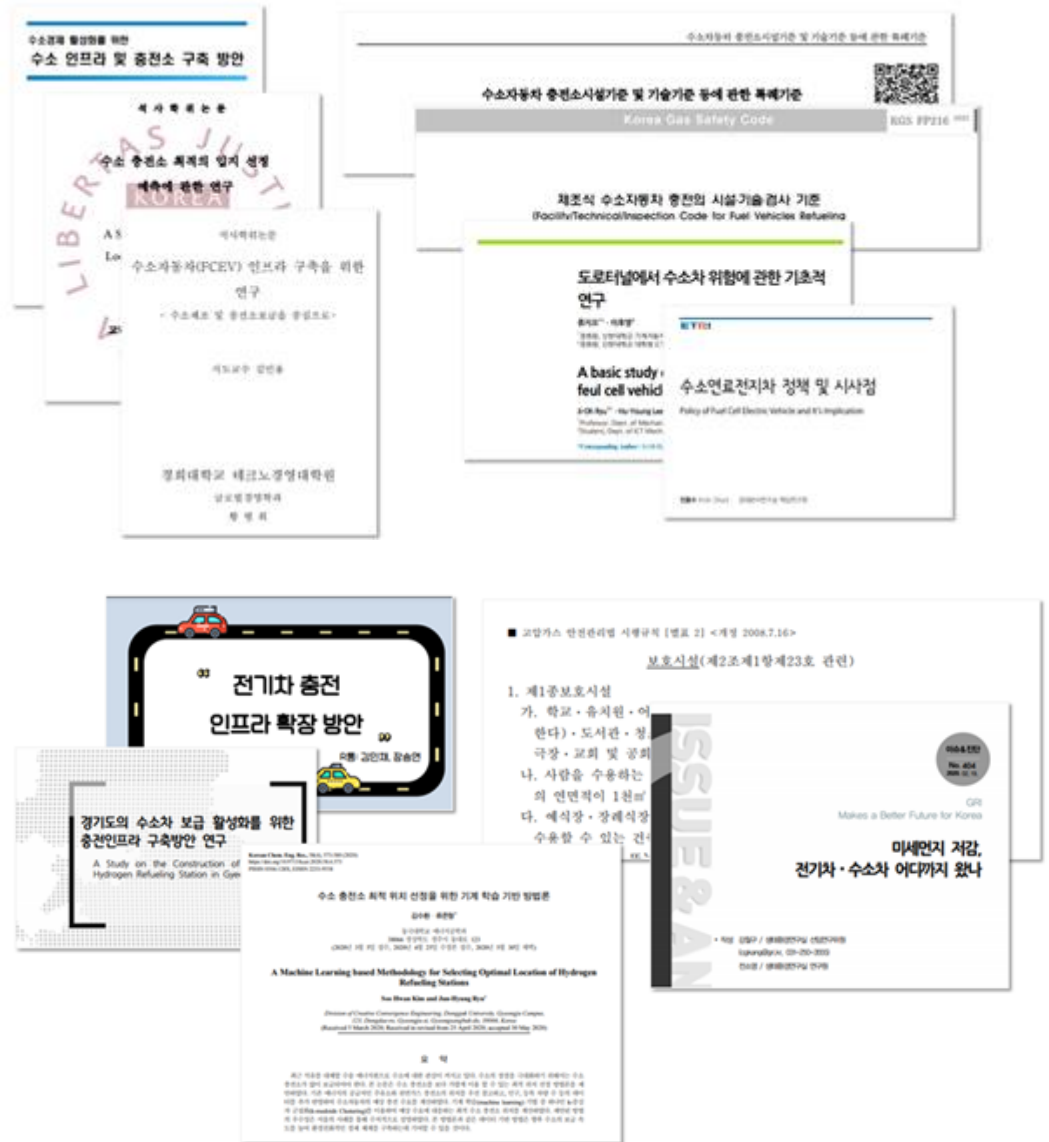
2. 법적 제재 확인 필요

3. 설치 비용

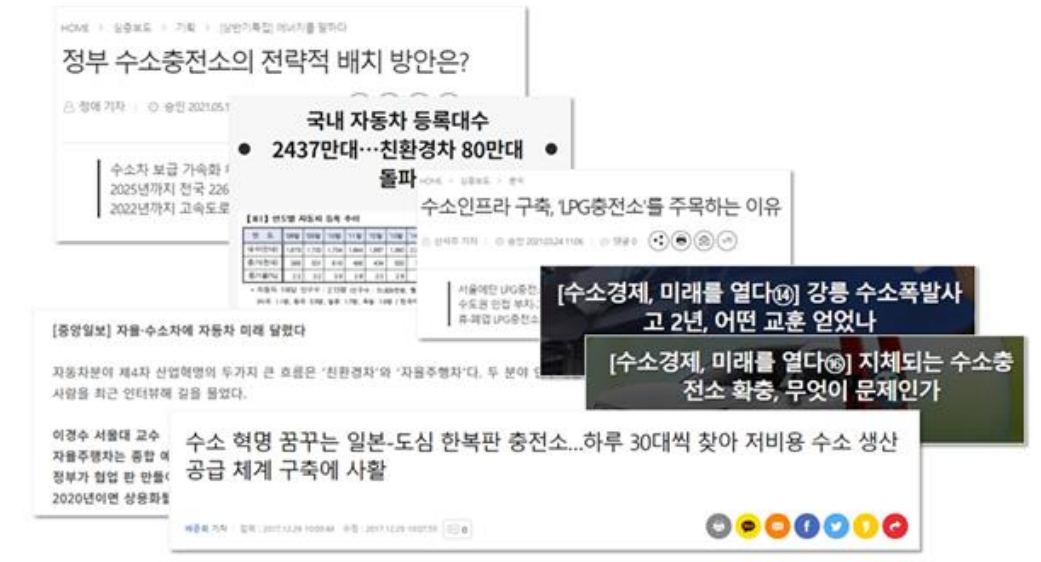
4. 수소차 장점/단점 (전기차 비교)

- 수소차 충전소 관련 논문 및 기사 참조(: feature 선정 근거)

## ○ 논문



○ 기사



● 향후 진행할 내용

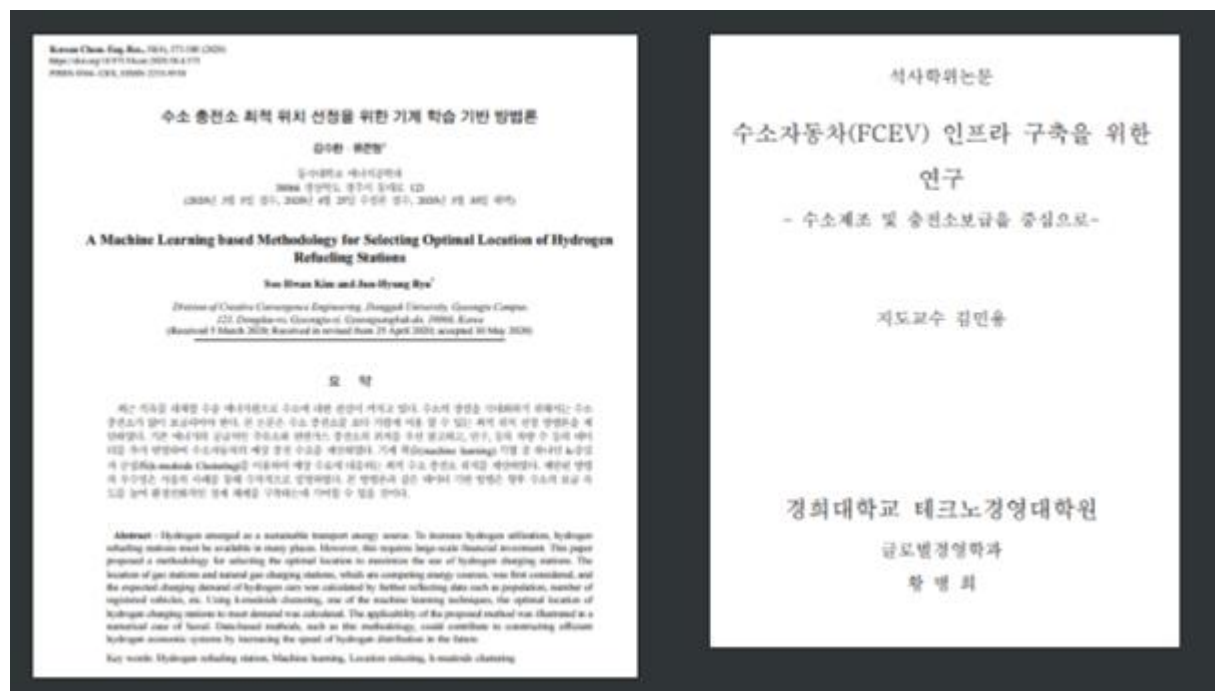
	기간	활동	산출물
PJT 기획	8/25(수) ~ 8/27(금)	<b>프로젝트 기획</b> 프로젝트 주제 최종 확정 및 기획안, WBS 작성	기획안, WBS

- 주제선정
- 기획안, WBS, 수행일지
- 역할분담

21년 8월 27일 금

- 주제선정 :수소 충전소 최적입지 선정
- 선행연구 기반 feature 선정

⇒ LPG 충전소 현황, 1종보호시설(학교, 유치원) 위치, 서울시 동별 화재정보, 수소충전소 위치정보, 수소차 구매 보조금 지급 현황, 서울 자치구별 수소차 등록현황



- 데이터 수집 및 전처리(진행중)



- 기획안, WBS, 수행일지 작성

□ 오전: 멘토링 및 질의응답

□ 오후: 프로젝트 관련 질의응답 및 조언

- 데이터사이언스

1. 레퍼런스와 차별화 차별점을 고민해야 한다.

→ 기존 논문에서 보완할 점을 주는 것도 좋다.

2. EDA 과정 보면서 피쳐의 우선순위를 선정하자.

- 데이터엔지니어

1. 지속성 있는 주제가 아니라 단발성 주제로 데이터 최신화가 필요 없고, 양이 적은 데이터인데 spark, aws를 써야 할까?

⇒ 이거는 보는 시선, 그리고 조원들 간의 소통에 따라 달라질 거 같다.

~이렇기 때문에 사용하지 않았다고 할 수도 있고, ~그럼에도 불구하고 배웠던 도구를 사용해봤다고 할 수 있을 것 같다. (배운 내용 적용, 데이터 늘어날 수 있어서 등)

⇒ 왜 사용했는지가 중요. (면접에서도)

분석에 초점을 맞출지 기술 적용에 초점 맞출지. 어설프게 사용하면 좋지 않다.

2. 태블로로도 충분할 것 같은데 장고를 써야 할까?

⇒ 참고로 다른 조는 장고를 쓰는 것 같다. 그리고 이 부분도 조원 간의 결정이 가장 중요한 부분이다. 태블로로도 구현이 충분히 가능한 부분이다.

프로젝트의 주체는 '우리'이다. 우리가 선택하는 것이 중요하다.

21년 8월 31일 화

---

□ 데이터사이언스

1. 전기차충전소 입지선정 코드 분석
2. 수소충전소 입지선정 중요 데이터 도출  
→ LPG 충전소

□ 데이터엔지니어

3. 데이터사이언스 요청 데이터 수집
4. local의 MySQL에 외부 접속 시도(실패)  
→ 내일 AWS server에 MySQL 설치 후 데이터 저장 및 활용 예정
5. 시각화 논의:  
Tableau 활용 후 시간이 남으면 Django까지 활용
6. AWS 내 MySQL 설치 방법 확인

□ 강사님 피드백

1. 우리조의 데이터는 주로 정적데이터라서 LPG충전소의 매출, 방문 빈도수, 충전량 등의 데이터를 활용해서 이벤트를 넣어주는 것이 어떨까(apache flume & kafka 활용)
2. 입지선정 시 사용한 백데이터 같이 공개하여 인근주민들의 인정 및 수용 필요

21년 9월 1일 수

□ 데이터사이언스

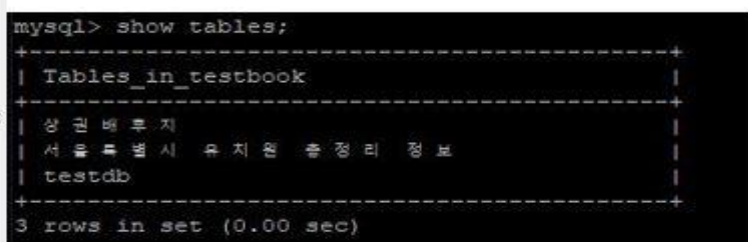
1. 정부의 수소충전소 향후 배치 기준 확인
  - 수소충전소 1기당 수소차 보급대수 180대수 기준
  - 수도권이나, 주요 대도시 기준으로 목표 물량 산정
2. 수소충전소 입지 선정을 위한 LPG 충전소 데이터 활용 근거 마련  
(법적 규제 및 관련 뉴스기사 참고)
3. Clustering 방법론 조사

□ 데이터엔지니어

1. AWS 서버, Local 환경 Mysql 연결
2. AWS DB에 데이터 저장 중
3. 데이터 전처리 작업중 - null 처리, column 정리
4. trouble shooting :
  - 한글 데이터 import 오류 해결



AWS DB workbench



DB 연결 확인



## 21년 9월 2일 목

---

### □ 데이터사이언스

1. Clustering 방법론 학습
2. 지번 주소의 경도, 위도 변환 요청 (EPSG 5174)
3. LPG 충전소 기반 K-means clustering

### □ 데이터엔지니어

1. 데이터 전처리 : 위경도 변환, 적재
2. CM(cloudera manager) 설치 및 공부

□ 데이터사이언스

1. LPG 충전소 위치 기반의 K-medoids clustering을 통한 수소 충전소 입지 선정
2. K-means와 K-medoids clustering 결과 비교

□ 데이터엔지니어

1. 데이터 전처리(중복값 제거)
2. 데이터 엔지니어 적용할 기술 논의 및 공부

□ 공통

1. 9월 1주차 중간보고 후 방향성 논의

a. 서비스의 대상: 수소충전소 설립자 및 운영자

ex. 현대자동차, H2 KOREA

**< 수소충전소 운영자 >**

✓ 지역 주민 반대 등으로 충전소 부지선정이 어려워 대부분 외곽으로 부지를 선정하려 합니다. 기존 충전소 증설이 필요합니다.

✓ 충전소 확충도 중요하지만, 수소차 보급을 확산하기 위해서는 설치된 충전소의 안정적인 운영이 필요합니다.

b. 서비스 제공 내용: 조건별 선정입지

c. 서비스 제공 방법: 태블로 진행 후 시간 여유가 생기면 장고도 진행

21년 9월 6일 월

---

□ DS, DE 공통

- 서비스 대상기업, 요구사항, 서비스 제공 방식 논의

- 대상 기업

수소충전소 특수목적법인(ex. 하이넷, 코하이젠)

- 요구사항

수소충전소 구축 입지 선정 의사결정 지원시스템

- 방법

“수소충전소 구축 입지 선정 의사결정 지원시스템”을 통해 입지조건들을 바탕으로  
최적 입지 추천

- 입지조건 : 고압가스 규제(1종 보호구역, 2종 보호구역 등), 수소차 등록현황, 인구수, 화재발생률 등
- 지역별
- 연도별
- ex. 설치 비용 등도 고려하면 좋을 듯
  - 부지비용, 운송비용이 지역별 차이가 있을 것 같음
- 정부의 수소충전소 구축 목표와 현재 구축 현황 비교
- 수소차 증가 추이

- 수소충전소 하나당 담당하는 수소차 개수를 연도별, 지역별 비교하여 전반적 추이 파악
- 예상화면을 먼저 구상하기
- 기대효과
  - 경제적 효과 등
  - 위치제안 + 알파 효과
  - 활용적인 측면에서 어떤 기능을 제공할 것인가

## 21년 9월 7일 화

---

### □ 데이터사이언스

1. Maximal Covering Location Problem(MCLP) 학습

### □ 데이터엔지니어

1. Django 구현 연습
2. Tableau - Django 연동

### □ 공통

1. 서비스 내용 구상
2. frame image 구상

## H2충전소 입지추천을 통한 의사결정 지원시스템

입지선정

화재, 규제, LPG, 수소차 + 부동산(crawling)

H2 뉴스

1일, 1주, 1달 등 동향(scraping + word cloud)

H2 주식

현황

수소차 증감추이, 정책실현 정도

21년 9월 8일 수

---

□ 데이터사이언스

- 서울시 총인구수(1km X 1km 격자.shp) 데이터를 통한 인구밀도 시각화
- MCLP 분석을 위한 자동차등록대수 현황 및 서울시 행정동 파일 전처리

□ 데이터엔지니어

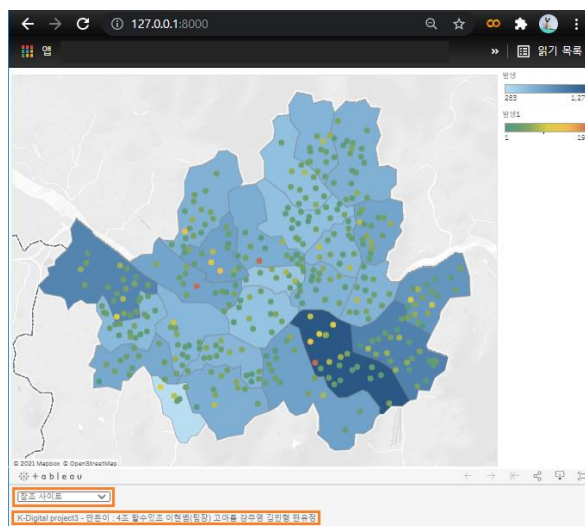
- 연도별 수소차 등록현황 데이터 확인
- AWS 복구 후 mySQL 등록 및 데이터 적재
- 유치원 위경도 시각화 완성 + 속성별 필터 적용
- Django 공부 및 Frame 제작
- Tableau 공부

## □ 데이터사이언스

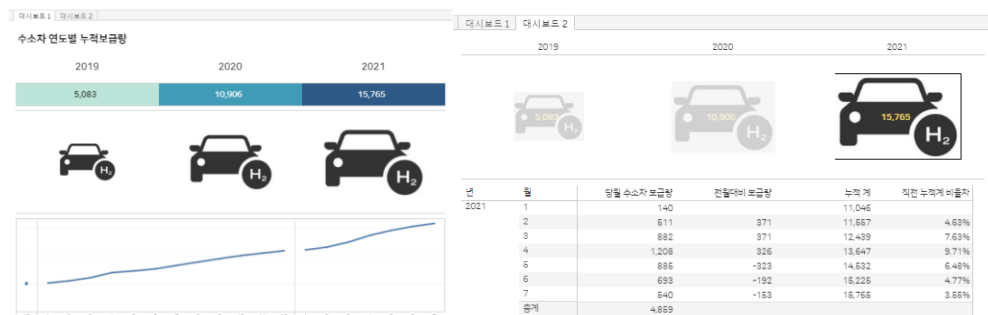
- MCLP 분석을 위한 자동차등록대수 현황 및 서울시 행정동 파일 전처리
- 인구 격자 데이터에 자동차등록대수 정보 할당

## □ 데이터엔지니어

- Django frame 및 templates 구성
- Django를 통해 유관기관 바로가기 select box 생성 및 바닥글 생성



- Tableau를 통한 수소차 보급량 변화추이 시각화



21년 9월 10일 금

---

□ 공통

- 예상되는 서비스 대시보드 구성
- 중간 발표 및 멘토링을 위한 피피티 작성
- 서비스 구현 툴에 대한 논의



□ 공통

- 멘토링 피피티 내용 제작 및 피드백 시간

□ 데이터사이언스

- 1km x 1km 격자 중심점에 LPG 충전소 위치 정보 할당
- MCLP 분석을 위하여 LPG 충전소에 대한 정규화된 인구수 및 정규화된 자동차등록대수에 대한 가중치 산정

□ 데이터엔지니어

- Django frame 구성 논의 및 시각화 자료 검토
- 웹 페이지 구성
- 서비스 구성요소(뉴스, 주식)에 대한 논리적 필요성 토론
- 장고, 태블로 연동, 크롤링 학습

□ 데이터사이언스

- 1km x 1km 격자 중심점에 제 1종 보호시설 및 기존 수소충전소 위치 정보 할당
  - 기존 수소충전소와 1종 보호시설이 격자 중심점 내에 존재하는지의 여부에 따라 존재 = 1, 존재 X = 0을 할당하여 1인 경우만 고려
- 정규화된 인구수 및 정규화된 자동차등록대수에 대한 가중치를 반영한 MCLP 분석 시작

□ 데이터엔지니어

- 장고에 부트스트랩 연동하여 팸플릿 구현
- 장고-태블로 연동 확인
- 뉴스 크롤링 공부
  - i. HTML에서 tag, 상하관계를 통한 데이터 선택 실습 → 참고자료:  
[https://book.coalastudy.com/data\\_crawling/](https://book.coalastudy.com/data_crawling/)
- api 로 데이터 불러오는 방법 구상 및 토론

□ 공통

- 이번으로 된 토지임야 정보의 좌표계 변환 방법 고민

□ 데이터사이언스

- 1km x 1km 격자에서 단위를 줄여 100m x 100m 인구 격자 중심점 분석 시작
- 수소충전소 입지 선정에 토지임야 정보 반영 방법 구상

□ 데이터엔지니어

- 국토교통부\_토지임야정보조회서비스 open api : xml형식으로 저장 및 csv 파일로 변환
  - api주소: <https://www.data.go.kr/iim/api/selectAPIAccountView.do>  
<https://www.data.go.kr/iim/api/selectAPIAccountView.do>
  - open api 받아오는 코드 작성
  - mysql에 자료 올려놓음: 토지임야정보조회 업로드 (총 853031개)
- 토지임야정보조회 주소를 위경도 좌표로 변환 작업
  - 데이터 로컬에 저장: 한글 깨짐 현상 → 변환 오류로 jupyter lab에서 저장한 csv파일 이용
  - 데이터 좌표 변환 시 건수 제한 → 네이버 지도 API-Geocoding 연결 cf. 지오코딩 프로그램 (일 / 1만 건), 구글 API(속도 문제), KAKAO MAP API(일 / 30만 건), **네이버 지도 API-Geocoding (월 / 300만 건)**
  - 네이버 지도 API-Geocoding 사용 코드 실습
- 토지임야정보조회 번지수 에러(날짜, 영어로 인식) → ['] 추가하여 변환 방지
- 뉴스 크롤링 연습

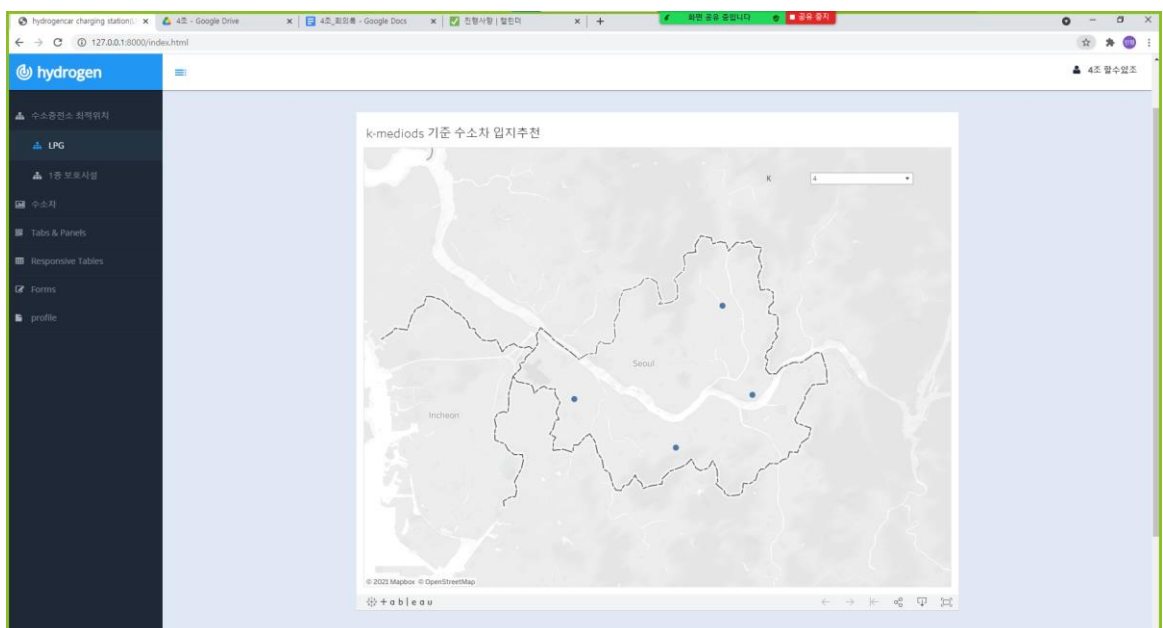
21년 9월 15일 수

□ 데이터사이언스

- 1km x 1km 격자에서 단위를 줄여 100m x 100m 인구 격자 중심점 분석
- AWS 활용하여 100m x 100m 인구 격자 중심점 분석 시작하였으나, jupyter lab 환경설정 문제로 인해 해결중

□ 데이터엔지니어링

1. 경,위도 변환
2. 뉴스 크롤링 실습
3. AWS 상의 python 설치 오류 발견 및 오류 해결
4. 장고 디자인 변환
  - 왼쪽 버튼 수정
  - 색상 수정( 빨간색->파란색)
  - 내용물 전반적인 수정 ( tableau 위치, 제목, head 모양 → 4조 할수있조)
  - footer 삭제
  - 링크명 변경



## 21년 9월 16일 목

### □ 데이터사이언스

- 데엔의 도움을 받아 AWS내 matplotlib import 문제 해결
- putty 접속 오류 해결 중
- AWS 활용하여 MCLP 분석돌리고 있으나, 데이터 수가 많아 오래걸림
- 중간평가 발표를 위한 ppt 작성 중

### □ 데이터엔지니어

- AWS 접속 문제 해결 시도
- Crawling 실습
- 주소 데이터 변환(지번→위경도)
- Django와 Tableau 연동



21년 9월 17일 금

---

□ 공통

- 중간 발표 준비

□ 데이터사이언스

- 

□ 데이터엔지니어

- Crawling
- 지번데이터 → 위경도 데이터 변환

### □ 데이터사이언스

#### - 이경미 강사님 피드백

1. 목적함수 및 공식 제대로 이해하기
2. OR(Operation Research)쪽의 네트워크이론, 최적화 이론 교재 봐보기
3. 사용하는 feature의 예측 모델 고려해보기
  - a. 인구와 자동차 등록대수가 미래에는 다른 값을 가질 것이기 때문
  - b. 예측 모델로는 선형회귀 사용가능 (y값을 인구나 자동차 등록대수를 넣어서 고려) --> 함께 고려해서 서비스를 해주면 좋을 듯
4. LPG 충전소를 Y변수로 놓고 가중치 구할 때, 수소자동차등록대수도 넣고 고려해봐도 됨.

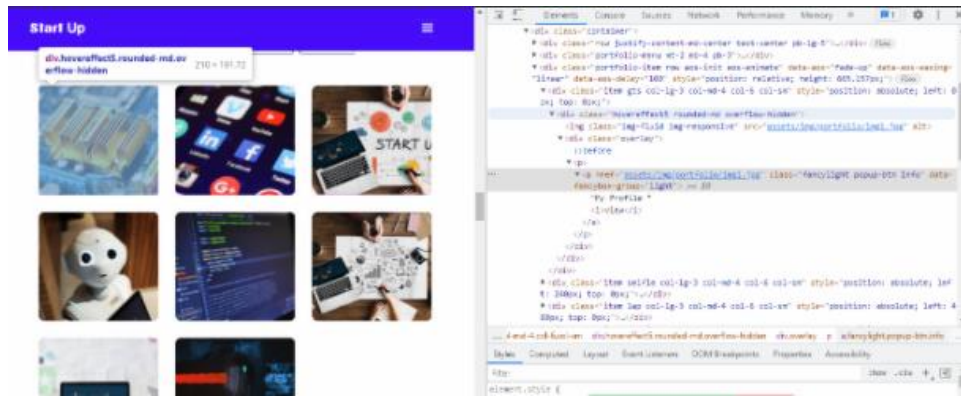
- MCLP 목적함수 정리
- 수소자동차등록현황 정보할당
- 선형회귀 분석을 통한 가중치 산정

### □ 데이터엔지니어

- 뉴스 크롤링(수소 충전)
  - 크롤링 대상 지정: 제목, 사진, 신문사, 링크 (본문x)
  - 업데이트 주기:
- 장고 틀 논의 및 결정, 6칸



- 크롤링 틀 구성 및 코드 구현



- 위경도 데이터 MySQL 저장

21년 9월 24일 금

#### □ 데이터사이언스

- OLS regression 분석을 통한 feature의 p-value 및 다중공선성 확인
- MCLP 2차 분석 (100m x 100m grid)
- 서울시 건물 geojson 데이터 수집

#### □ 데이터엔지니어

- 뉴스 가운데 정렬 성공
  - CSS 수정
  - HTML을 통한 UI 수정 및 대문제작
  - 뉴스 크롤링



21년 9월 28일 화

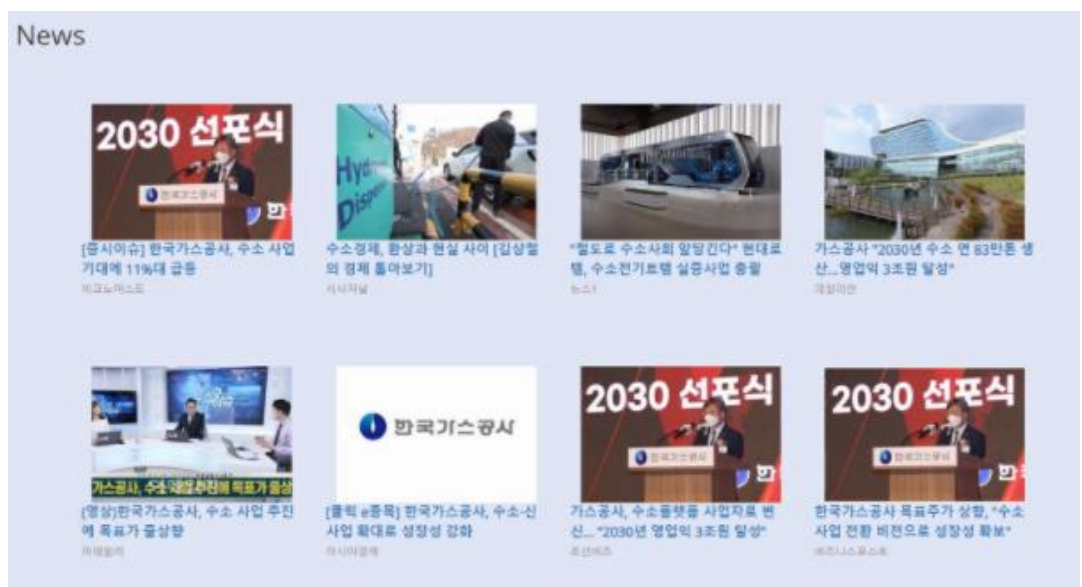
□ 데이터사이언스

- 격자에 기존 LPG 충전소 정보를 할당 후 MCLP 분석
- 각각의 MCLP와 k-medoids 클러스터링 분석이 제안한 입지 비교

□ 데이터엔지니어

news page 내

- Django - crawling data 연동
- 좌우 여백 공간 맞추기(padding)
- 글자 크기 조절, 글자 중간위치로 변경



21년 9월 29일 수

□ 공통

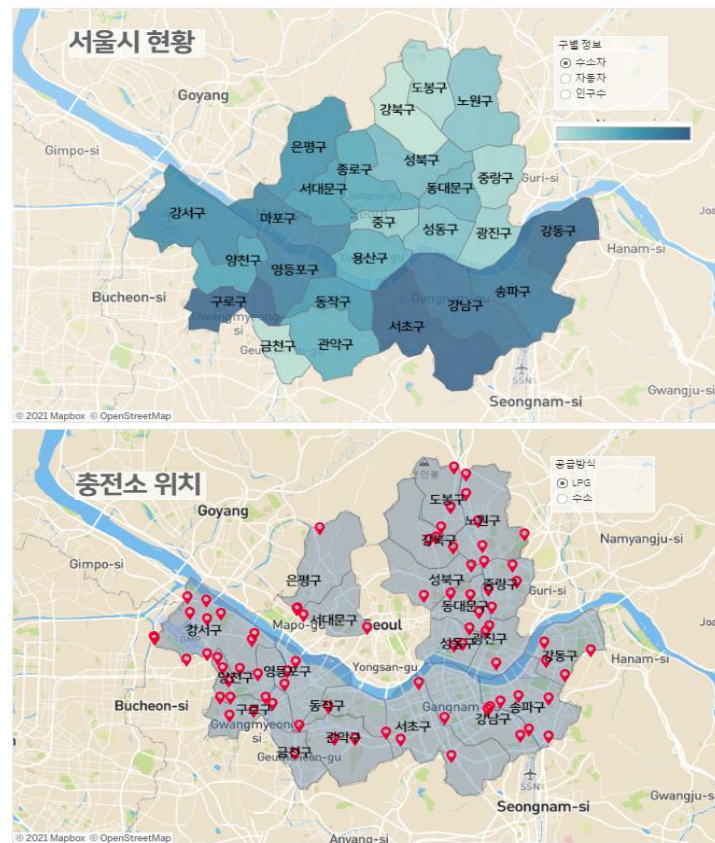
- 기존의 MCLP 분석 결과 도출된 후보지역 좌표 추출
- 서비스 대시보드 디자인 의논
- 최종 프로젝트 1차 리허설 발표자료 작성

□ 데이터사이언스

- 동별 자료(자동차등록대수, 수소차등록대수)를 격자별 자료로 변환

□ 데이터엔지니어

- 태블로 시각화 구현
- 뉴스 창 오류 수정



## 21년 9월 30일 목

---

- 공통

- 발표

- 데이터사이언스

- 동별 데이터(자동차 등록대수, 수소차등록대수) —> 격자형 데이터로 변환
  - 로지스틱 회귀분석을 통한 가중치 재산정
  - 기존 LPG 충전소 위치와 MLCP 결과 추출된 충전소 위치 비교

- 데이터엔지니어

- PPT 주제 선정 이유 스토리 구성 및 작성

21년 10월 01일 금

---

□ 공통

- 마지막 멘토링을 위한 발표자료 준비

□ 데이터사이언스

- 아래의 내용들을 반영하여 새롭게 도출된 후보지 좌표를 데엔팀에 전달 및 시각화 요청
  - 동별 데이터(자동차 등록대수, 수소차등록대수) → 격자형 데이터로 변환
  - 로지스틱 회귀분석을 통한 가중치 재산정

□ 데이터엔지니어

- 추천입지 지도 원 크기 하나로 고정 및 원 안에 우선순위 표시
- 장고 버튼 스타일 수정
- PPT 주제 선정 이유, 흐름도 스토리 구성 및 작성
- 화면설계서 작성

21년 10월 02일 토

---

□ 공통

- 발표 준비(자료 정리 및 대본 수정)
- 피드백 분석 및 질의응답

21년 10월 04일 월

---

□ 데이터사이언스

- MCLP 커버리지 영역 바꾸면서 최적입지 결과물 → DE 전달
- MCLP 분석 결과 정리
- 소스코드 정리

□ 데이터엔지니어

- 좌표 1km/1.5km/2km 좌표 → 주소 변환(완료)
- Usecase Diagram(완료)
- MySQL에 파일 업데이트(완료)
- 요구 명세서 작성중
- 코드 정의서 작성중
- tableau 제작 및 장고 임베딩
- 대문 이미지 수정
- PPT 수정

21년 10월 05일 화

---

□ 공동

- ppt 수정

□ 데이터사이언스

- MCLP Covered Ratio 분석 결과 정리
- 소스코드 정리
- 분석모델 정의서
- 코드정의서
- PPT(Intro - 선행연구 비교, 분석모델 프로세스 구성도)

□ 데이터엔지니어

- 요구 명세서(완료)
- 코드 정의서(완료)
- 데이터 정의서(완료)
- 장고 url 변경

## 21년 10월 06일 수

---

- 공통
  - 피피티 작성 및 강사님 피드백 반영
- 데이터사이언스
  - 커버리지에 따른 MCLP 분석 관련 ppt 추가
- 데이터엔지니어
  - tableau 수정
  - 코드 정의서 수정

## 21년 10월 07일 목

---

- 공통
  - PPT 수정 및 발표연습

## 21년 10월 08일 금

---

- 공통
  - PPT 최종 발표