


# 제주 도로 교통량 예 측 AI 경진대회

팀명:제주 치킨

# 목차

## contents

- 
1. Intro
  2. Idea
  3. EDA
  4. Feature Engineering
  5. Modeling
  6. Outro

# 1. Intro

## 배경

- 제주도내 주민등록인구는 2022년 기준 약 68만명으로, 연평균 1.3% 정도 매년 증가
- 외국인과 관광객까지 고려하면 전체 상주인구는 90만 명을 넘을 것으로 추정
- 제주도민 증가와 외국인의 증가로 현재 제주도의 교통 체증이 심각한 문제

## 주제

- 제주도 도로 교통량 예측 AI 알고리즘 개발

## 설명

- 제주도의 교통 정보로부터 도로별 평균속도 회귀 예측

## 제공 데이터

- `train.csv` : 2022년 8월 이전 데이터만 존재하며 날짜, 시간, 교통 및 도로구간 등의 정보와 도로의 차량 평균 속도(`target`) 정보 포함
- `test.csv` : 2022년 8월 데이터만 존재하며 날짜, 시간, 교통 및 도로구간 등의 정보 포함

## 외부 데이터

- 날씨 데이터 : 경민님 적으세요.

## 2. IDEA

출발-도착지 거리 피쳐 추가,  
bg 정규화

데이터 가중치 (6, 7월)

날씨 외부데이터

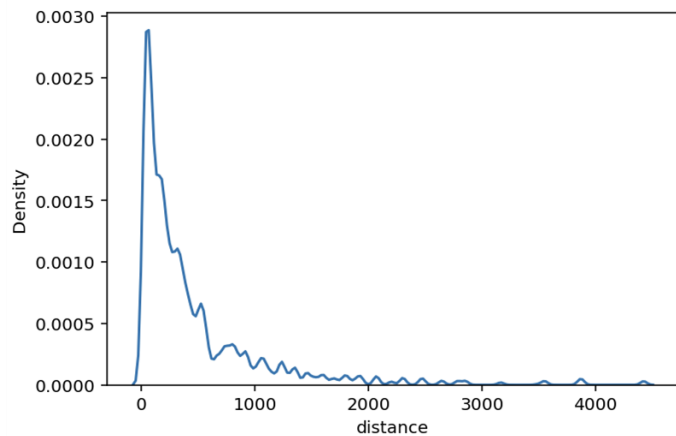
위경도 기반 clustering

시간대별 등급 구분

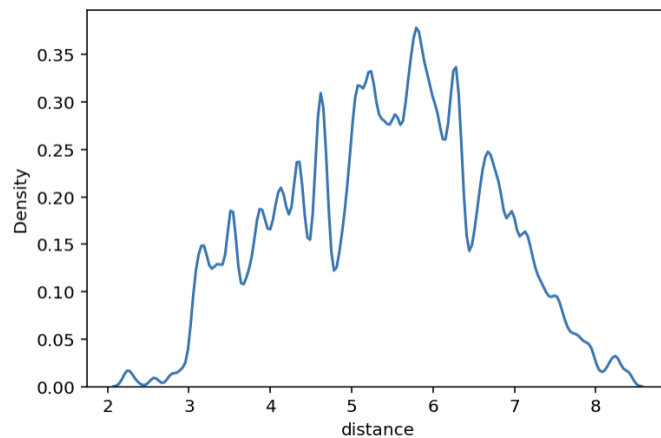
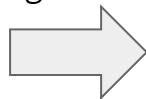
연월 병합

# 3. EDA

## 1) 출발지에서 도착지까지 거리 정규화

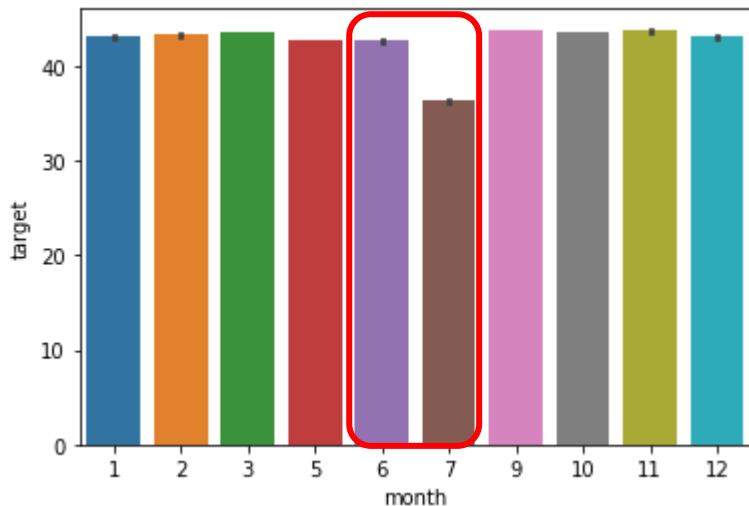


lg 변환



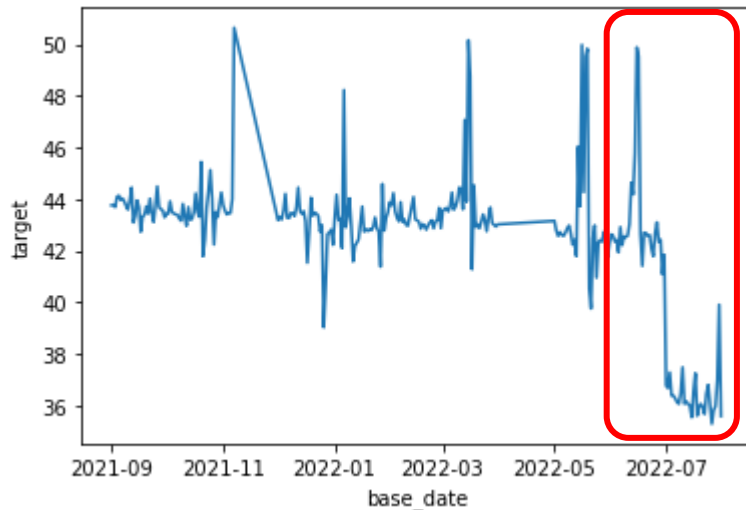
# 3. EDA

## 2) 월별 교통량 추이



- 6,7월이 최근접 데이터
- 6,7월에 속력이 내려가는 변화 보임

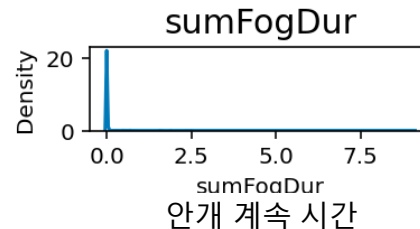
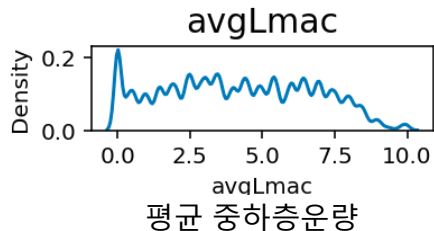
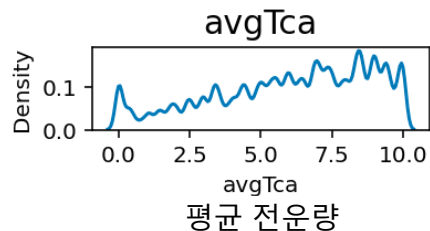
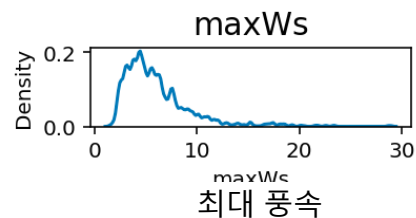
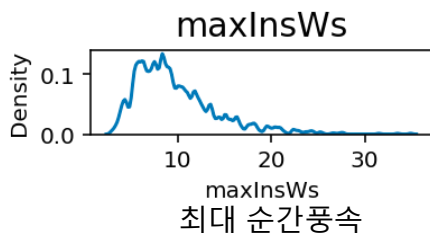
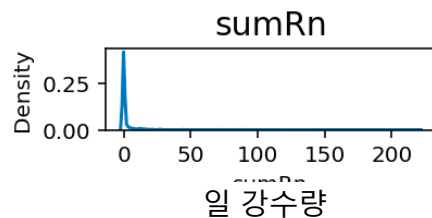
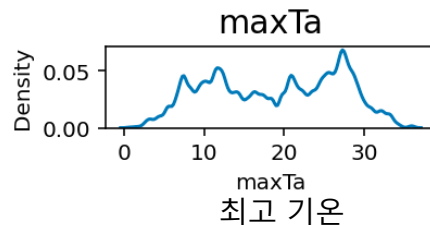
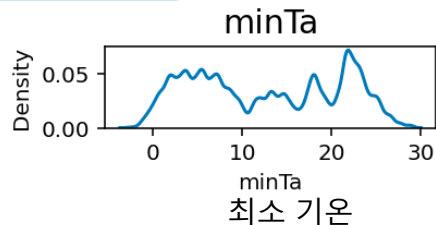
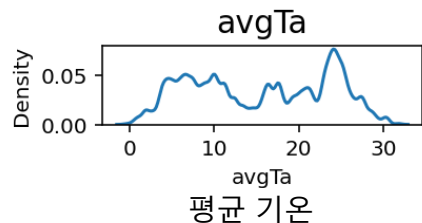
## 연월 병합



6, 7월 데이터에 큰 가중치 부과

# 3. EDA

## 3) 날씨

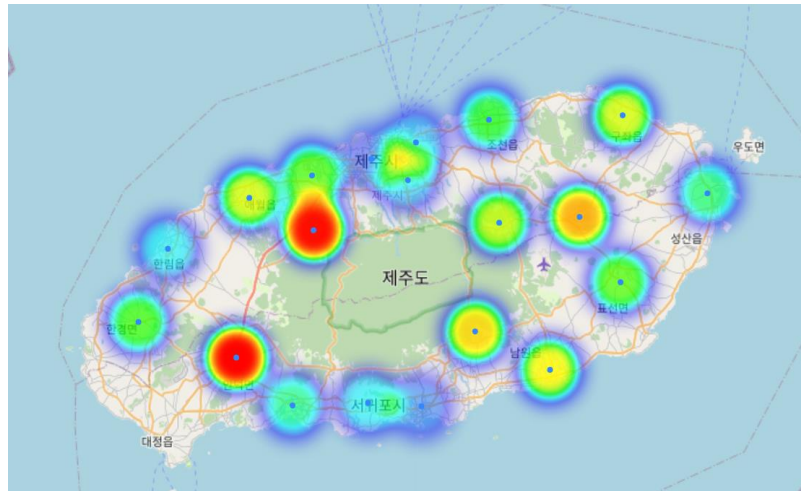


# 3. EDA

## 4) 위경도 기반 Clustering



<20개 group으로 clustering>



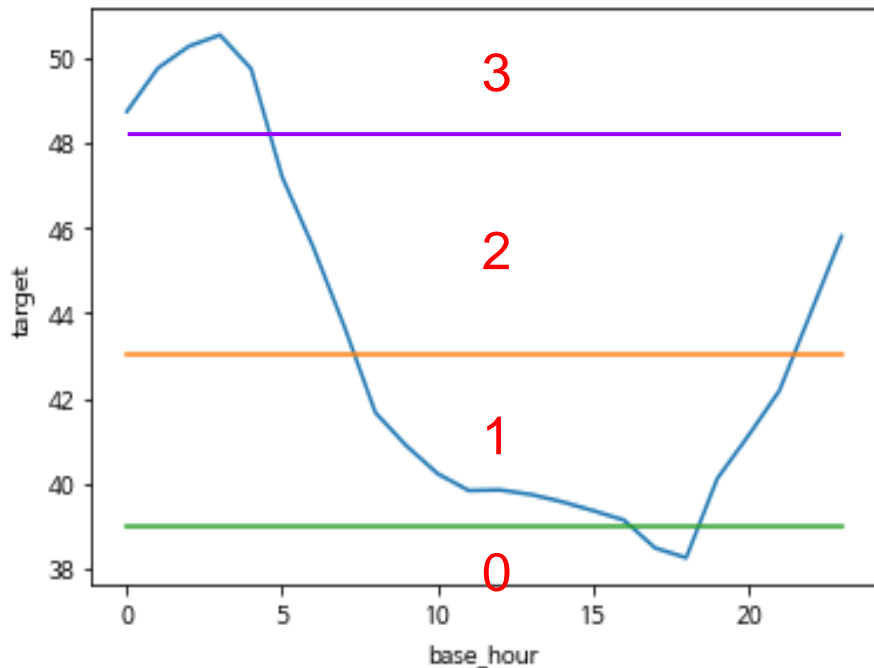
<group별 평균속도>

- 도착지점 위경도를 기준으로 clustering
- clustering group 별로 평균속도 heat map



# 3. EDA

## 5) 시간대별 평균속도

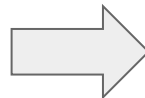


- 4개의 구간으로 나누어 속도에 따른 점수로 수치화
- 0~5시 : 3
- 6~7, 22~23시 : 2
- 8~17, 17~21시 : 1
- 17~18시 : 0

# 4. Feature Engineering

- 파생변수 생성
  - base\_date -> yeamonth
  - day\_of\_week, base\_hour -> time\_category

base_date	day_of_week	base_hour
20220623	목	17
20220728	목	21
20211010	일	7
...	...	...

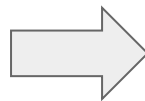


year_month	time_category
202206	0
202207	1
202110	2
...	...

# 4. Feature Engineering

- 파생변수 생성
  - latitudes & longitudes → distance, cluster\_num

start_latitude	start_longitude	end_latitude	end_longitude
33.427747	126.662612	33.427749	126.662335
33.500730	126.529107	33.504811	126.526240
33.279145	126.368598	33.280072	126.362147
...	...	...	...



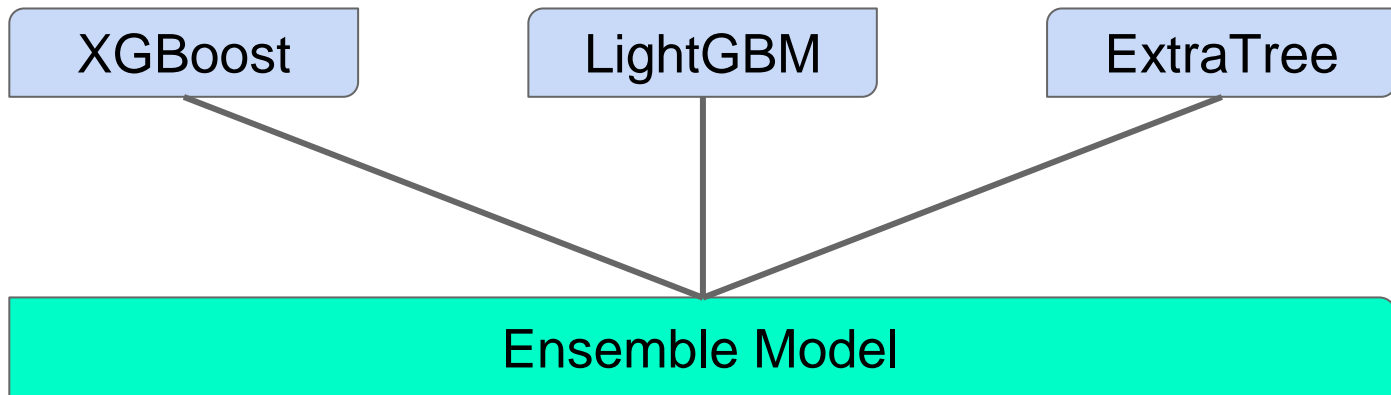
distance	cluster_num
0.000277	8
0.004987	1
0.006516	2
...	...

# 4. Feature Engineering

- 최종 Feature

이름	설명	이름	설명
yearmonth	연월	start_turn_restricted	시작지점의 회전제한 유무
day_of_week	요일	end_latitude	도착지점의 위도
base_hour	시간	end_longitude	도착지점의 경도
weight_restricted	통과제한하중	end_turn_restricted	도착지점의 회전제한 유무
lane_count	차로수	distance	시작지점-도착지점간 거리
maximum_speed_limit	최고속도제한	time_category	시간에 따른 그룹
road_type	도로유형	cluster_num	위치 클러스터 번호
road_rating	도로등급	stnNm	날씨별 지역
road_name	도로명	avgTa	평균 기온
start_node_name	시작지점명	avgTca	평균 전운량
end_node_name	도착지점명	minTa	최소 기온
start_latitude	시작지점의 위도	maxTa	최고 기온
start_longitude	시작지점의 경도	maxInsWs	최대 순간 풍속

# 5. Modeling



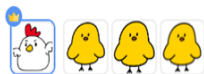
- $0.8 \times \text{XGBoost} + 0.15 \times \text{LightGBM} + 0.05 \times \text{ExtraTree}$
- 각 모델을 FLAML을 활용해 하이퍼 파라미터 튜닝
- 결과 - 평가지표(MAE)

#

팀

팀 멤버

최종점수



# 6. Outro



김경민

## Feature Engineering

- 날씨 data 추가

## Modeling

- AutoML, DNN



김태종

## EDA

- Visualization

## Feature Engineering

- time\_category  
- clustering  
- 출발-도착지  
거리 정규화



박수진

## Feature Engineering

- DBSCAN

## Modeling

- catboost,  
ExtraTree,  
Pycaret



조근혜

## Feature Engineering

- yearmonth 피쳐 추가  
data 가중치 추가

## Modeling & Tuning

- HistGB, FLAML