



# HMM 기반 다음 POI 추천 및 경로 시각화

01

# Research Background



01

## GPS

GPS의 대중화로 위치 기반 서비스(LBS) 분야가 성장

02

## LBRS

사용자의 지리 정보와 사용자 특성 정보를 이용하여 구매 물품, 장소, 활동 등을 추천하는 위치 기반 추천 서비스 연구가 활발히 진행됨

03

## HMM

많은 LBRS 연구 논문에서 HMM(은닉 마코프 모델)을 활용하여 추천함

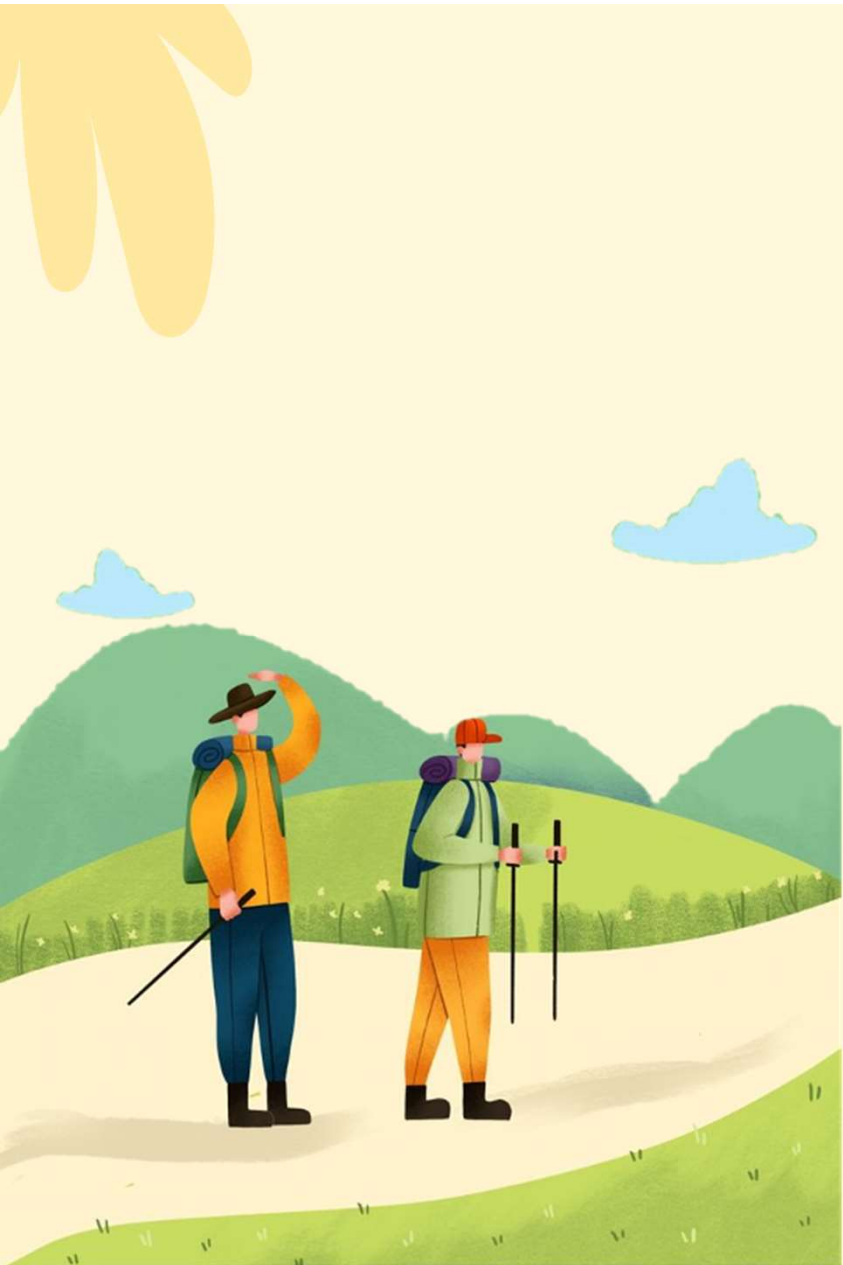
04

## FMM

인용한 LBRS 연구 논문은 시각화 단계에서 단순히 관심 지점을 나타낸 한계가 있음

R

HMM을 모델을 기반으로 사용자 위치를 기반으로 추천 장소 TOP3를 도출하고 이를 FMM을 이용하여 관심 지점(POI) 간의 경로를 효율적으로 나타내고자 함



# **Main Topics**

# Data



[교통, 안전] 월별 렌터카 체류 거점 위치 정보

사용자 이전 로그와 이후 로그의 시간 차이가 20분 이상인 GPS 정보



관심 지점 도출

[교통, 안전] 월별 렌터카 위치 정보

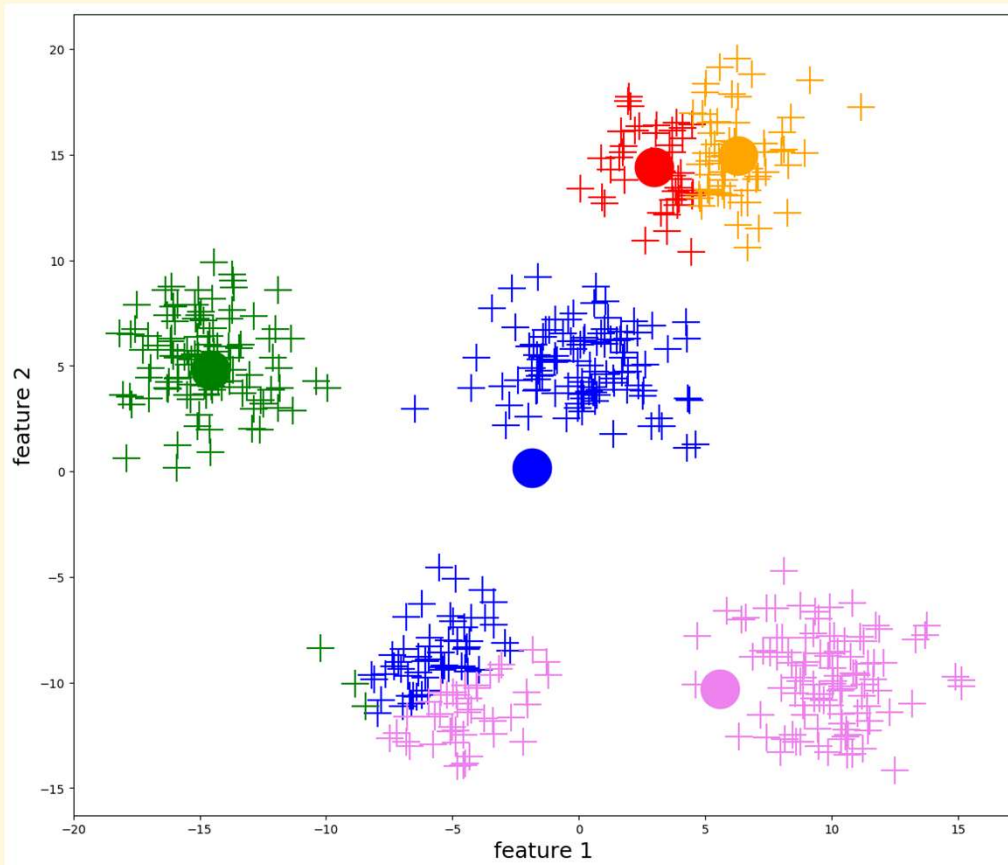
30초마다 수집한 사용자의 GPS 정보



사용자 이동 경로 생성

2020년 1월 ~ 2021년 12월 총 24개월 데이터를 분석에 이용

# K-means++ Clustering



1

## 첫번째 중심점 선택

데이터 포인트 중 하나를 무작위로 선택하여 첫 번째 중심점으로 설정

2

## 다음 중심점 선택

각 데이터 포인트마다 가장 가까운 현재 중심점과의 거리를 계산하고, 이 거리에 비례한 확률로 선택함.  
선택된 새 중심점은 현재 중심점과 거리가 먼 데이터 포인트를 다음 중심점으로 설정

먼 말?

3

## 반복

K개의 초기 중심점을 선택할 때까지 1 ~ 3단계를 반복

4

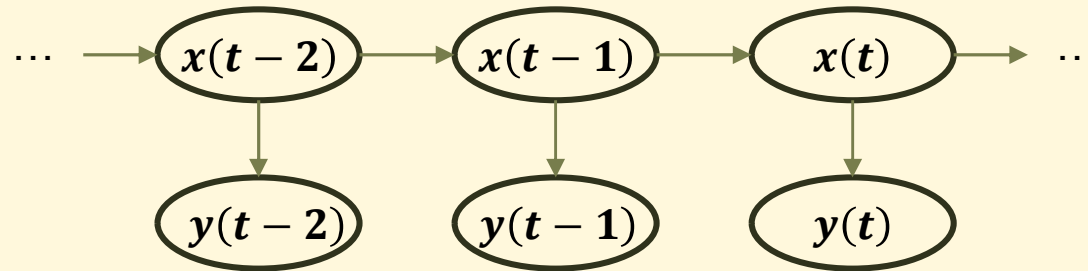
## 클러스터링

선택한 초기 중심점으로 K-Means 클러스터링하여 총 K개의 그룹을 생성

# Hidden Markov Model

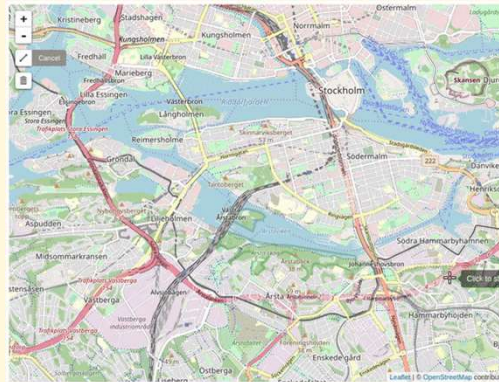
Hidden State  
관측할 수 없는 상태

Observable State  
관측 가능한 상태



HMM(은닉 마코프 모델)은 시간적 순서가 중요한 데이터에서 어떤 관측값이 관측됐을 때 대응하는 숨겨진 상태를 확률적 모델로 추론 가능

## Fast Map Matching



도로망에 노이즈가 많은 GPS 데이터를 매칭하는 데에 효과적인 해결 방식

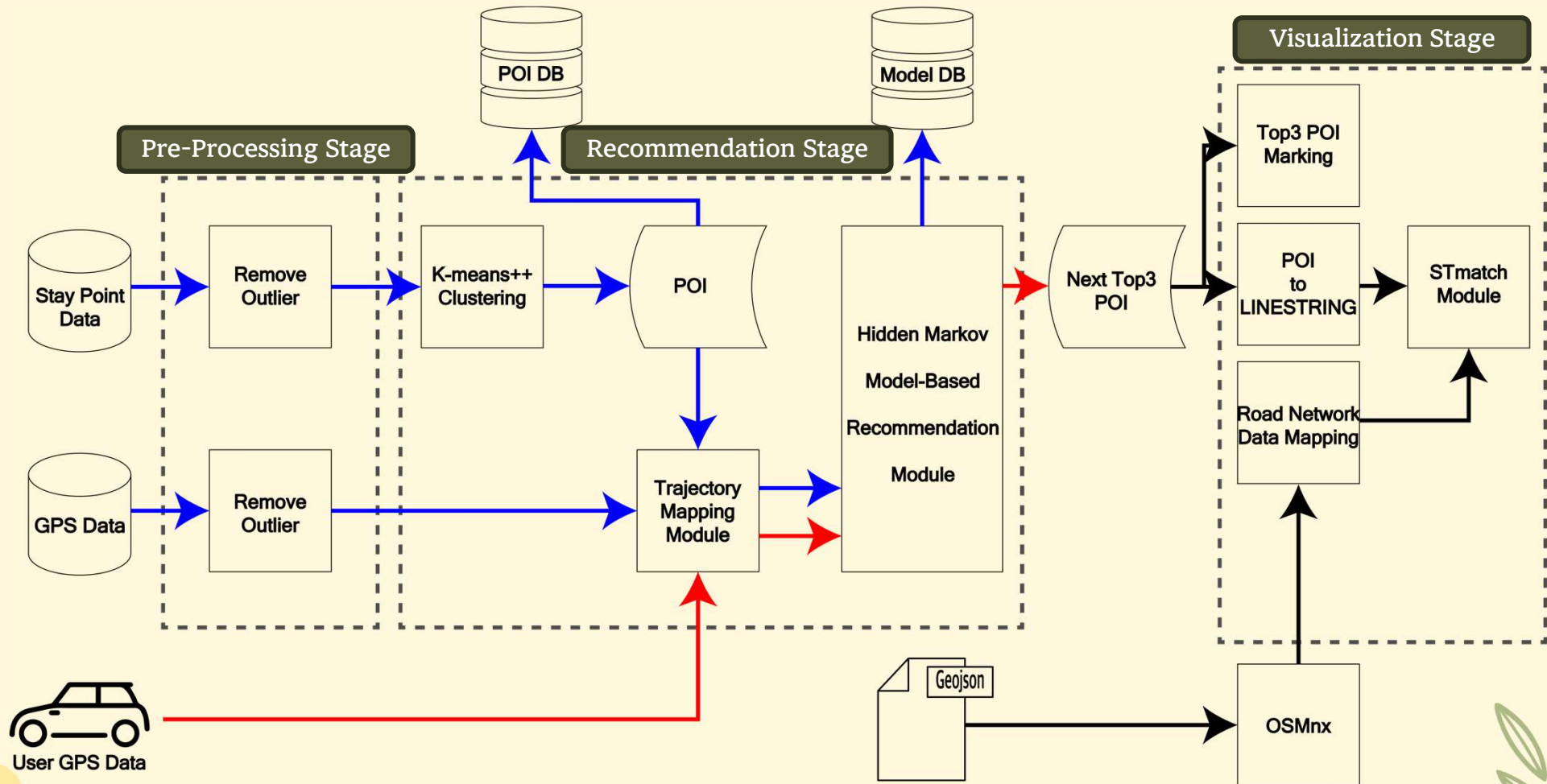


# Research

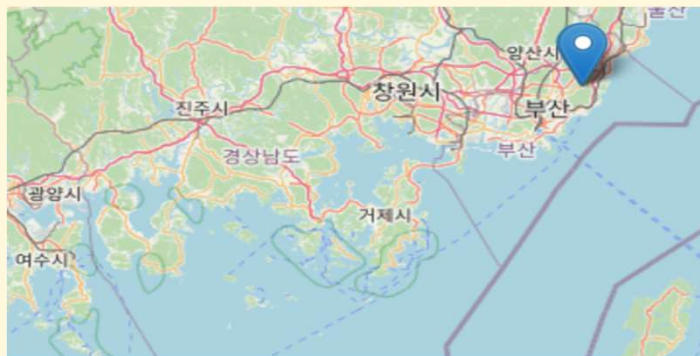
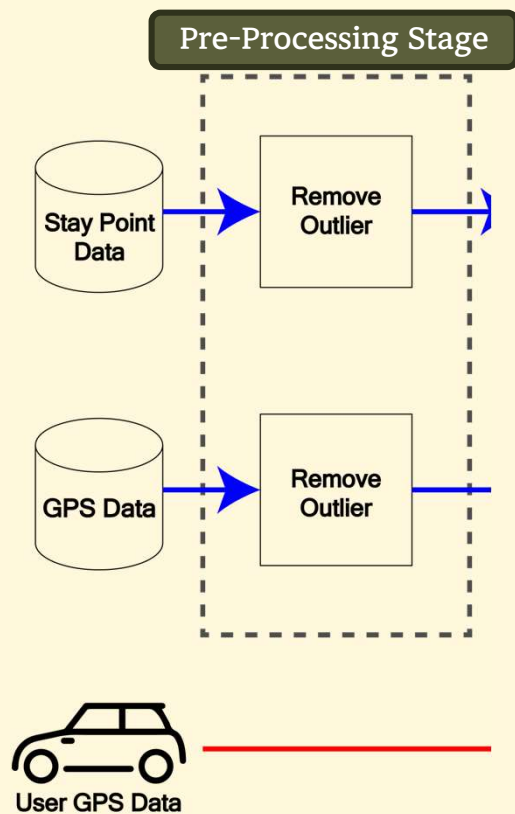




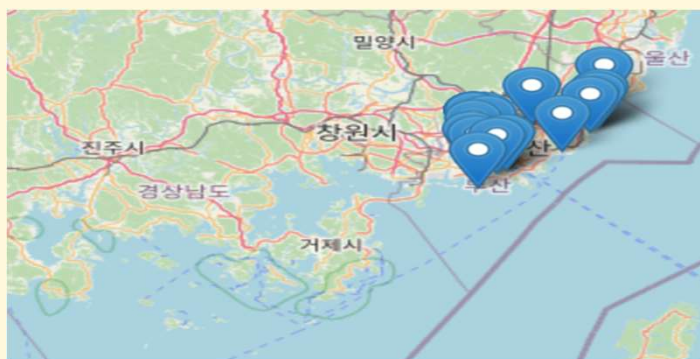
# Entire Diagram



# Pre-Processing Stage



월별 렌터카 위치정보 이상치의 예



월별 렌터카 체류거점 이상치의 예

## Remove Outlier

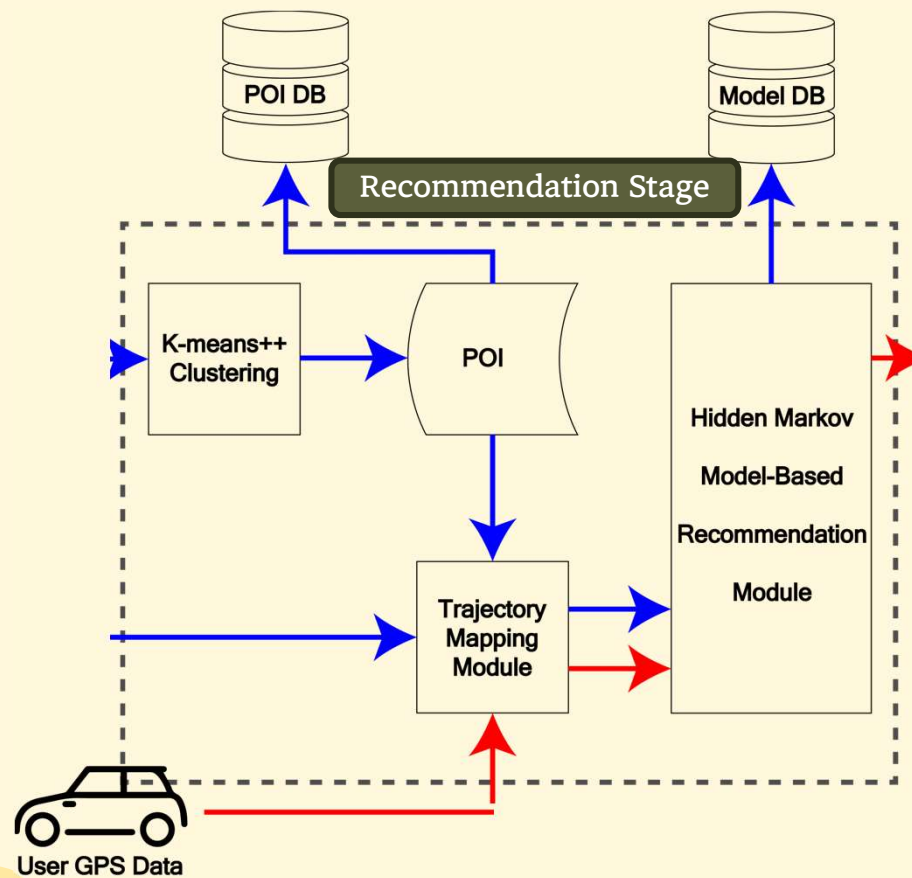
왼쪽 그림은 GPS 정보가 제주도를 벗어난 것으로, 이상치이기 때문에 삭제함

## 슬라이드 10

---

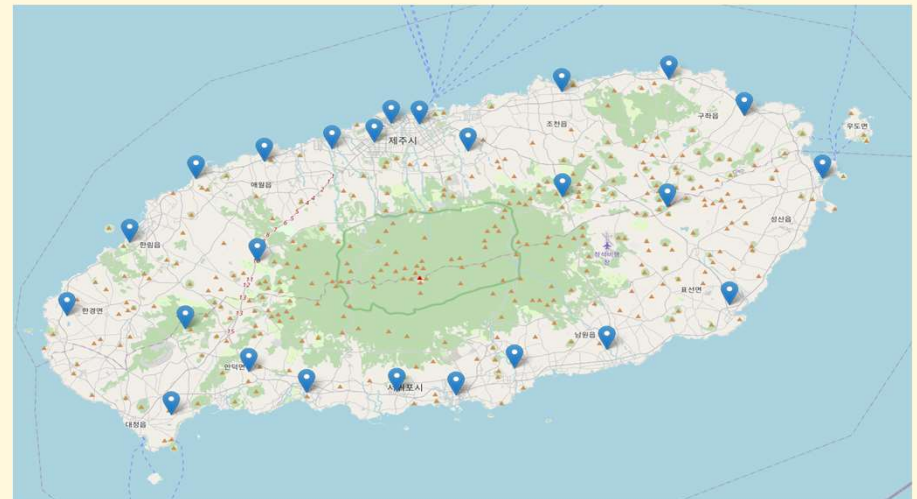
- YH0**      제주도가 하나의 그림 안에 들어올 필요는 없을 것 같습니다. 글 설명을 추가할테니 포인트가 잘 보이도록 캡처를 할 수 있을까요?  
Yeonji Ha, 2023-10-25T09:29:41.380
- YH1**      이 그림 또한 위의 피드백과 같이 수정 부탁드립니다.  
Yeonji Ha, 2023-10-25T09:30:01.517

# Recommendation Stage



## POI List 추출

K-means++ Clustering 을 통해 이상치를 제거한 데이터를 K = 25로 설정하여 Clustering



K = 25 일 때 계절 전체의 Clustering 결과

## 슬라이드 11

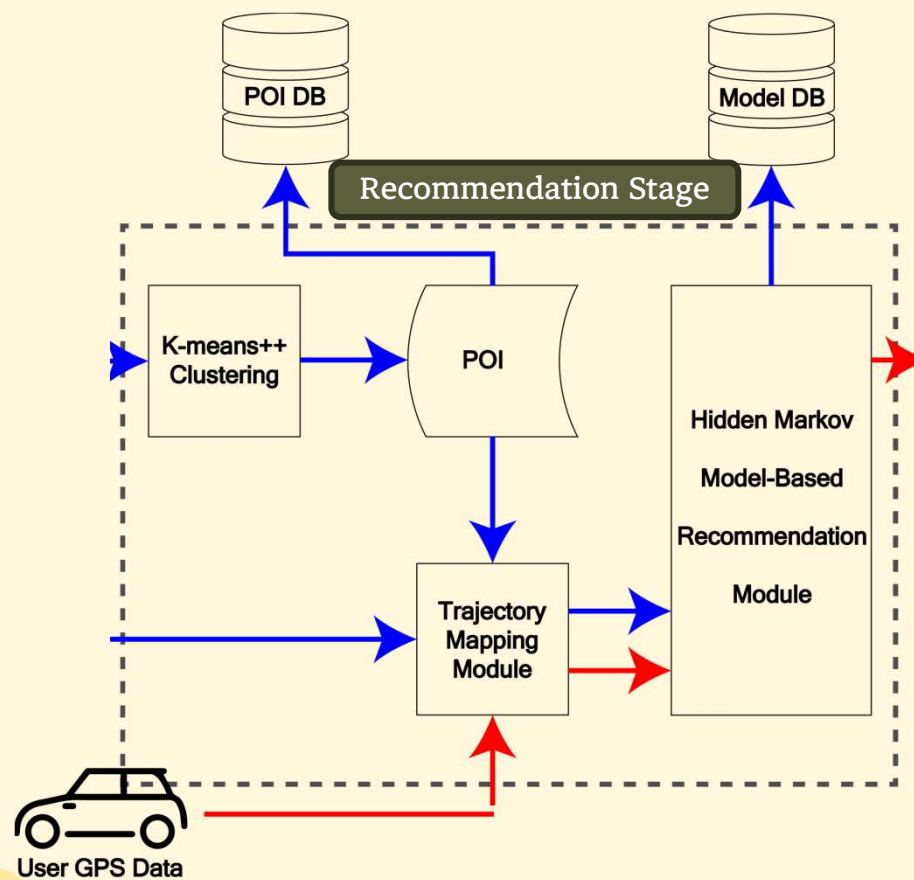
---

성이0

여기에 그냥 Clustering 결과 그림이 나올거같아요

성무 이, 2023-10-25T05:01:41.899

# Recommendation Stage



## Trajectory Mapping

클러스터링한 POI와 유저의 이동경로 데이터를 Mapping하여 Trajectory 생성

oid	collection_dt	longitude	latitude
46100093	20200905142600000	126.2697	33.41523
...			
46100093	20200905162900100	126.3338	33.24673

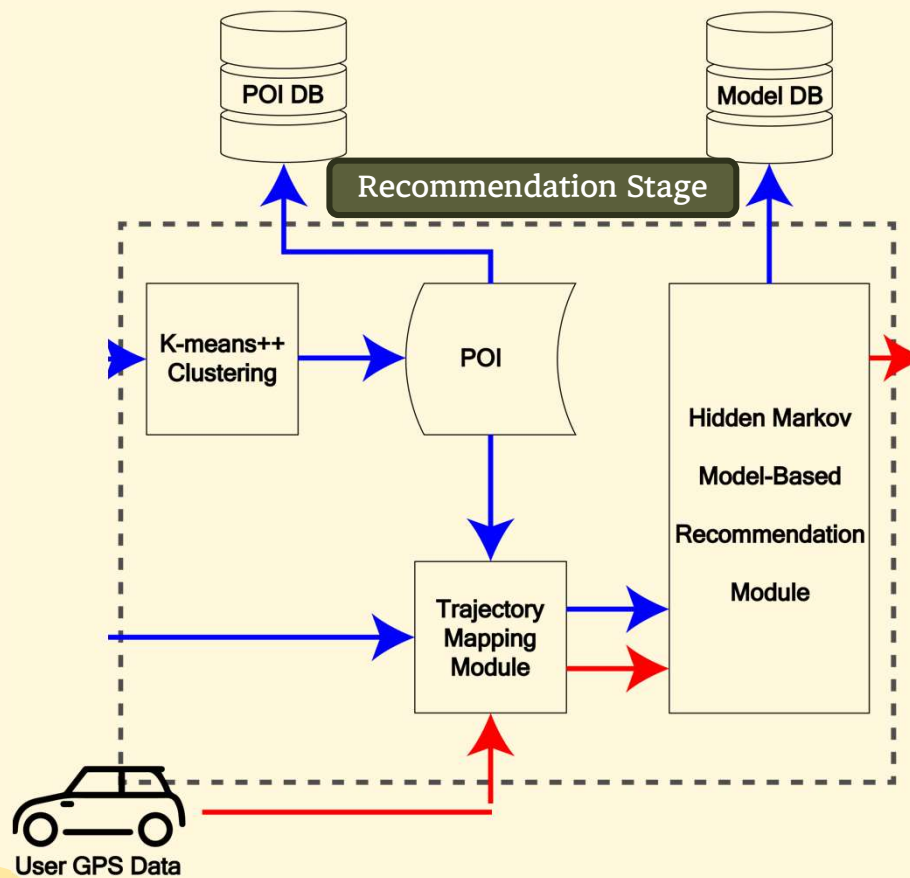
oid	Path	time_period
46100093	POI13-POI4- POI19-POI5- POI19-POI5	afternoon- afternoon- afternoon- afternoon- afternoon

YHO

표로 바꾸기

Yeonji Ha, 2023-10-25T08:23:21.410

# Recommendation Stage



## Hidden Markov

학습 데이터로 HMM 모델을 형성한 후 전 단계에서 Mapping 한 User GPS Data를 모델에 적용하여 다음 관심 지점을 예측

**Hidden State** 관심 지점

$POI_i$  ( $i = 0, \dots, K$ )

**Start Probability**

시작 상태가  $POI_i$  일 확률

$$\pi_i = P(POI_i), 1 \leq i \leq N$$

**Transition Probability**

숨겨진 상태  $S_i$ 에서 상태  $S_j$ 로 전이할 확률

$$a_{ij} = P(POI_{t+1} = S_j | POI_t = S_i), 1 \leq i, j \leq N$$

**Emission Probability**

숨겨진 상태  $S_i$ 일 때 관측값  $O_t$ 를 방출할 확률

$$b_i(O_j) = P(O_t = O_j | POI_t = i), 1 \leq i \leq N, 1 \leq j \leq M$$



## 슬라이드 13

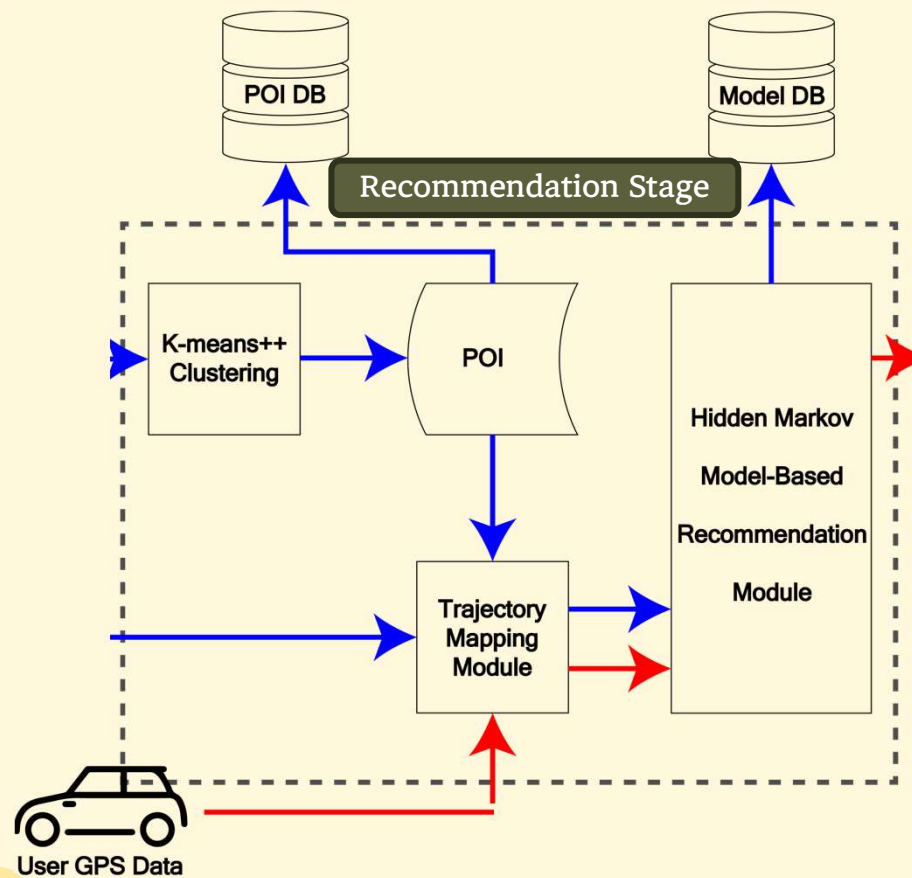
---

성이0

이 다음 슬라이드에 Viterbi 알고리즘에 대한 설명이 있으면 어떨까요??

성무 이, 2023-10-25T05:07:38.307

# Recommendation Stage



## Hidden Markov

학습 데이터로 HMM 모델을 형성한 후 전 단계에서 Mapping 한 User GPS Data를 모델에 적용하여 다음 관심 지점을 예측

## Observable State

관심 지점 + 시간대 정보로 인코딩한 값

시간대는 오전 12시부터 6시간 간격으로 새벽, 오전, 오후, 밤 4개로 구분하며 각각 0, 1, 2, 3으로 대응함. 관심 지점 1개마다 4개의 시간대가 대응되므로, 관심 지점의 개수가 25개일 때 관측값은 총 100개임

$$\text{Observation ID} = 4 * \text{POI ID} + \text{TimePeriodID}$$

## 슬라이드 14

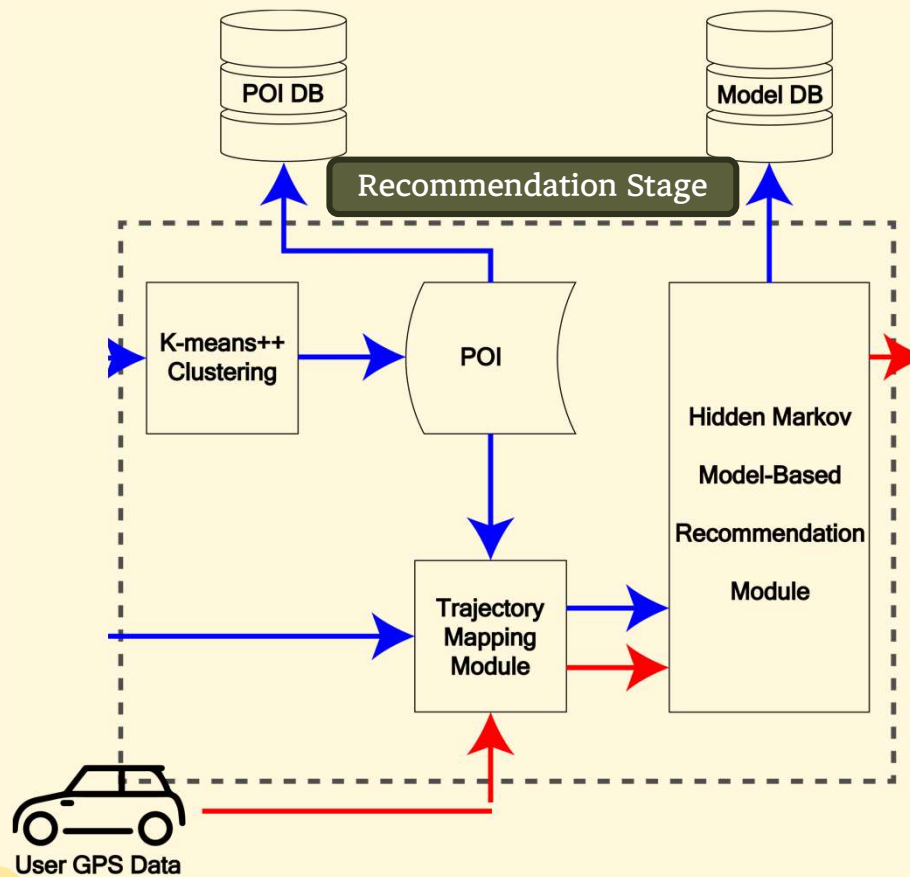
---

성이0

이 다음 슬라이드에 Viterbi 알고리즘에 대한 설명이 있으면 어떨까요??

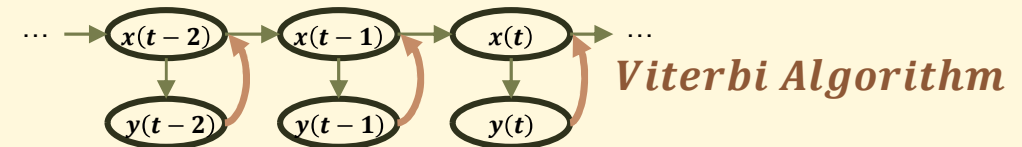
성무 이, 2023-10-25T05:07:38.307

# Recommendation Stage



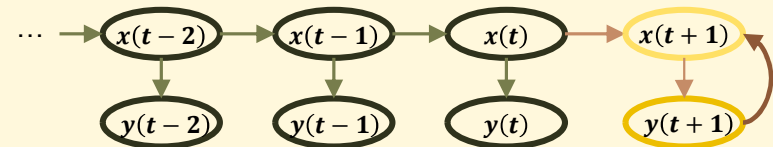
## Viterbi Algorithm

주어진 관측 시퀀스에 대하여 우도 확률이 높은 숨겨진 상태 시퀀스를 찾는 알고리즘

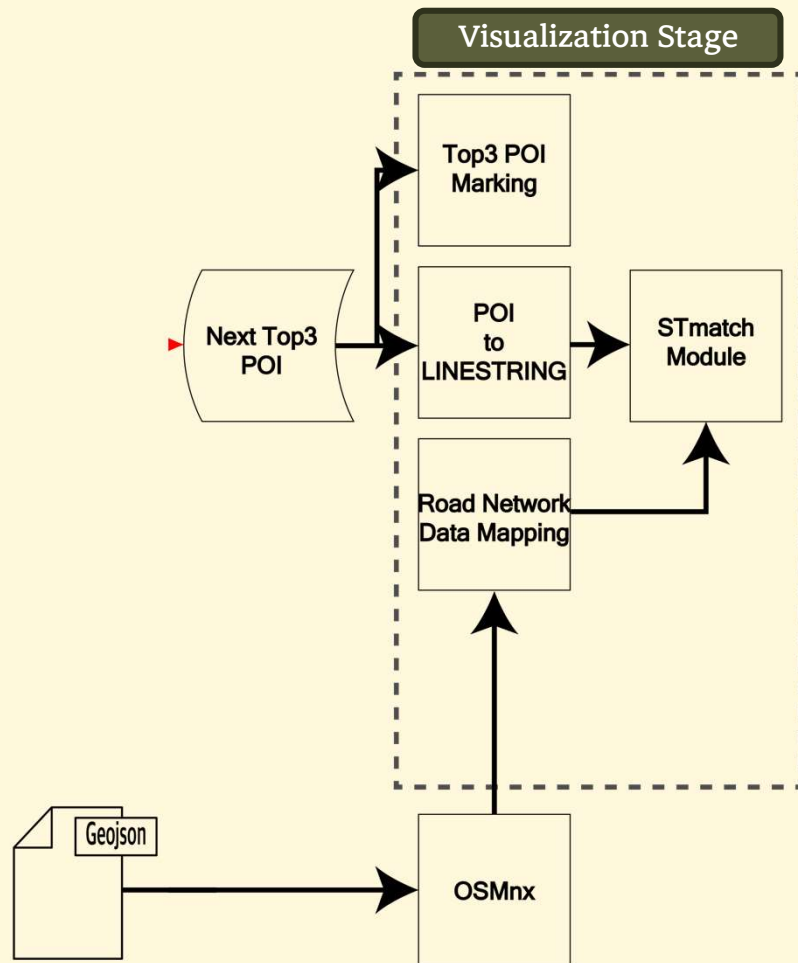


## Recommend TOP 3 POIs

현재까지의 관측치 시퀀스에서 발생할 수 있는 모든 관측치에 대해 Viterbi Algorithm으로 디코딩한 후, TOP3 POI 도출하여 추천함

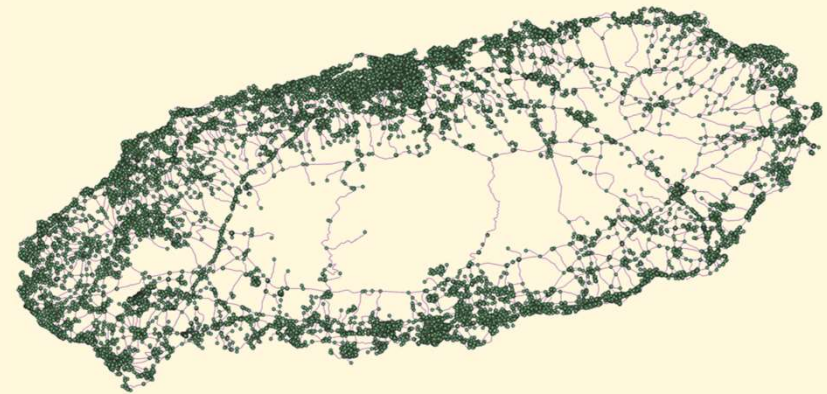


# Visualization Stage

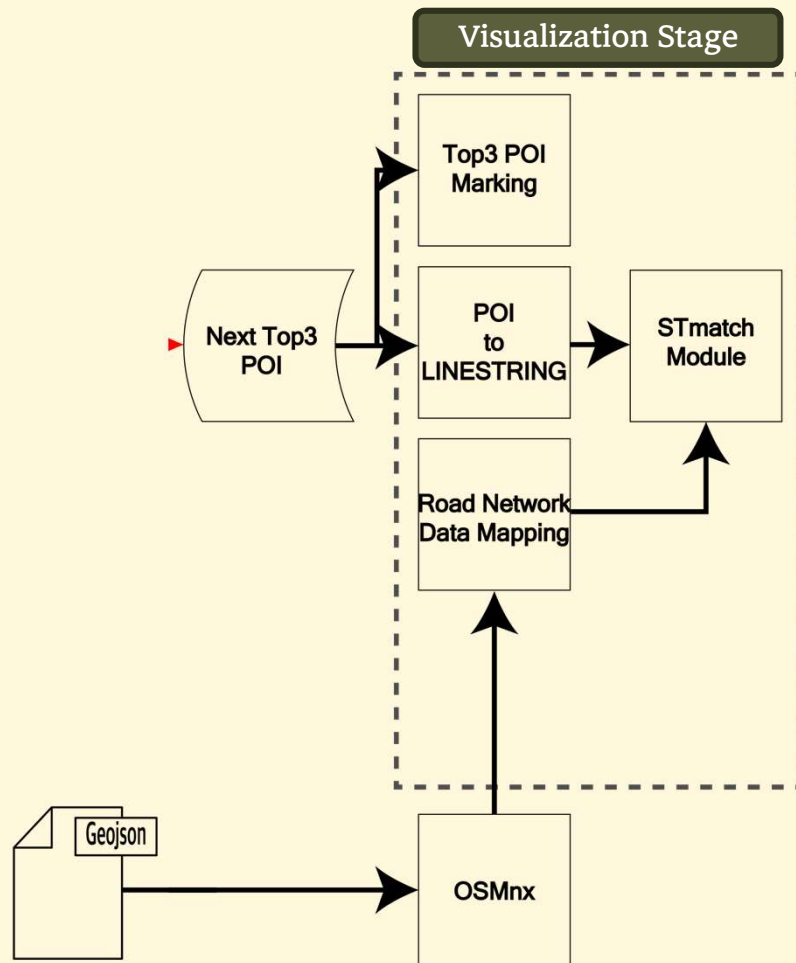


## Get Road Network Data

OSMnx 라이브러리를 이용하여 제주도 Geojson 파일을 Edges 와 Nodes .shp 파일로 변환

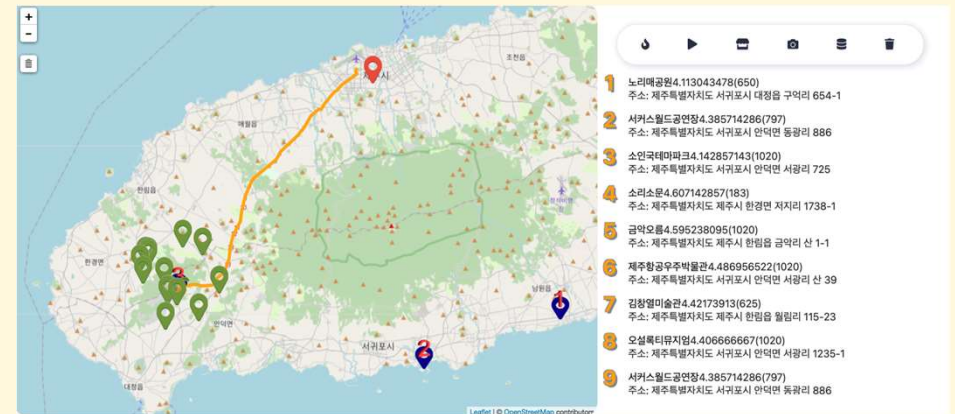


# Visualization Stage



## STmatching

GPU를 활용하여 그래프 패턴 매칭 성능을 가속화 시킨 Stmatching 을 이용하여 시각화에 알맞는 데이터 형태로 변경한 후 가시화 하여 나타냄



# Visualization Demo

because the system used a probability model called HMM, there were cases where the next expected movement point of some users was far from the current location. It could not be applied due to the limitation of the given time, but in the case described above, it is necessary to apply a method to correct it.



Leaflet | © OpenStreetMap contributors



04

**Result**



# Project Result & Conclusion

Seasonal  
Model



Total  
Model



예측 정확도는 계절의 영향이 크지 않다

방문 빈도 수에 따라 관심 지점을 선정하였음.  
계절에 따라 사용자들이 방문하는  
제주도 관광지나 음식점에는  
큰 차이가 없기 때문에 사료됨

K=25 일 때 예측 정확도가 가장 크다

클러스터의 개수가 작을 수록 더 많은 지점이  
뭉이기 때문에, 포괄적인 클러스터가 되므로  
예측 정확도가 가장 높은 것으로 사료됨

## 슬라이드 20

---

성이0

이 그림도 그 포스터에 쓴거로 바꿔야할거 같아요!

성무 이, 2023-10-25T05:02:49.399

성이1

그냥 이거 막대그래프 없애고 왼쪽 그래프를 크게해서 왼쪽꺼로 설명하면 될거같습니다!

성무 이, 2023-10-25T05:03:29.526



05

# Following Subject

# Following Subject based on limitations



Seong-Mu

사용자의 특성까지 고려하여 모델을 구성할 수 없었기에 여행객의 다음 방문 위치 예측에 사용하는 정보는 현재 방문 위치밖에 없다는 한계가 있다. 또한 현재는 제공된 데이터 셋으로 시각화하는 과정만 진행했는데 추후에는 실제 자동차 GPS 데이터 셋을 수집하여 실시간으로 시각화하는 과정을 진행해 보면 좋겠다.

성이0



Yeon-Ji

본 연구에서 적용한 K-means clustering 결과 이동 중인 지점 또한 관심 지점 중 하나로 선택되는 문제가 있다. 이를 개선하기 위하여 Varied-k means clustering으로 장소를 클러스터링하여 의미 없는 장소들은 제외하거나 K-medoids를 적용하여 실제 위치 정보를 중심 지점으로 하는 등을 적용할 수 있다. 또한 현재는 숨겨진 상태를 관심 지점으로 선정하였으나, 특정 관심 지점으로 이동하려는 상태로 정의하여 관측값에 따른 숨겨진 상태 시퀀스를 디코딩하는 방법을 적용하여 정확도를 비교해 볼 수도 있을 것이다.



Jae-Won

다양한 예측 모델과 비교하지 못한 것이 아쉬우며, 제한적인 데이터 또한 아쉽다. 프론트를 개발한 입장으로 이전의 JS, CSS, html 지식이 있었더라면 조금 더 효과적인 웹사이트를 구성해 짜임새 있는 사이트를 구성하지 못한 것이 아쉽다. 다양한 예측 모델과 비교하지 못한 것이 아쉬우며, 제한적인 데이터 또한 아쉽다. 프론트를 개발한 입장으로 이전의 JS, CSS, html 지식이 있었더라면 조금 더 효과적인 웹사이트를 구성해 짜임새 있는 사이트를 구성하지 못한 것이 아쉽다.

성이0

Seong Moo 입니다ㅋㅋㅋㅋㅋ  
성무 이, 2023-10-25T05:04:55.797