



Predict Mosquito Status

서울특별시 모기 예보 지수를 사용한 머신러닝모델 웹 어플리케이션

2021년 10월 11일 이윤수

목차

1. 서비스 소개

서비스 개요, 데이터 선정, 데이터 베이스 선정, 모델 소개

2. 서비스 시연

3. 개선점 및 보완점



서비스 소개

Predict Mosquito Status

서울시 주거지 모기예보지수 예측 서비스

서울특별시 시민건강국 감염병관리과에서 제공하는 모기예보지수 정보와 기상청의 평균기온, 최저기온, 최고기온, 강수량을 바탕으로 기계학습모델을 만든 서비스입니다. 웹페이지에 평균기온, 최저기온, 최고기온, 강수량을 입력하면 입력된 값을 기계학습모델에 넣어 예상되는 모기예보 지수를 사용자에게 보여주는 웹 어플리케이션 서비스로 현재는 로컬 PC환경에서만 작동되는 상태입니다.

Predict House Mosquito Status in Seoul, S.Korea

서울 주거지 모기지수 예측 모델

Do you hate mosquitoes? I don't like mosquitoes too. And I always worry that I might see one of them when I open the windows.
One of government organization in Seoul Metropolitan Government called '시민건강국'(Public-health-institute, I can't find official English name.)
Anyway, I made this machine learning model to predict mosquito status in house to avoid opening my windows for too long.
Hope this model useful and I wish you luck not to get mosquito bites.
Thank you:D

Please insert ONLY numbers

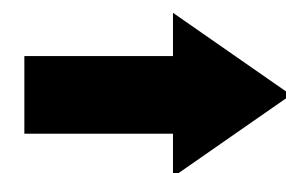
Average Temperture(°C):

Minumum Temperture(°C):

Maximum Temperture(°C):

Precipitation(mm):

PREDICT!



PREDICTION :

Mosquito Status in House is [44.6835703]

[Go back to main page](#)

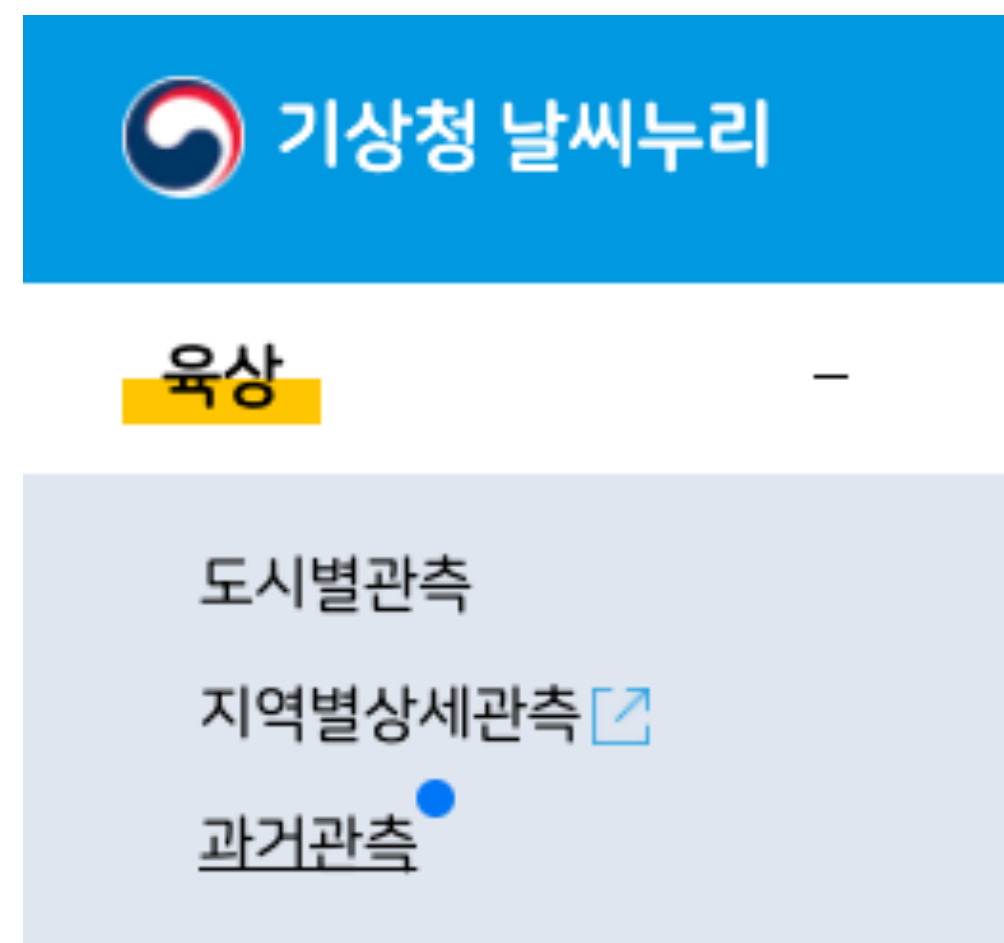
서비스 소개

데이터 선정



서울시 모기예보제 정보*

- 서울지역 모기발생 상황을 지수화하여 모기발생 단계별 시민행동요령을 알려주는 일일 모기발생 예보서비스에서 제공하는 데이터로 2016년부터 현재까지 매일 업데이트 되고 있음.
- 매년 5월 1일부터 10월 31일까지만 모기 활동 지수가 측정됨.
- 2019년까지는 지리타입을 구분하지 않고 서울시 평균 모기 개체수를 제공하였으나 2020년부터는 수변부, 주거지, 공원으로 지리타입을 구분하여 모기활동지수를 제공함.
- 2020년 4월 14일부터 서울시 내 총 50개의 DMS 장비를 사용하여 모든 지리타입의 데이터를 수집함.



기상청 날씨누리 과거관측**

- 기상청 날씨누리의 서울지역 육상 과거 관측기록에서 평균기온, 최저기온, 최고기온, 강수량데이터를 수집함.
- 기상청에서 제공하는 데이터이므로 서울시 감염병관리과에서 사용한 기상데이터와 일치하지 않을 가능성이 있음

*<https://news.seoul.go.kr/welfare/mosquito>

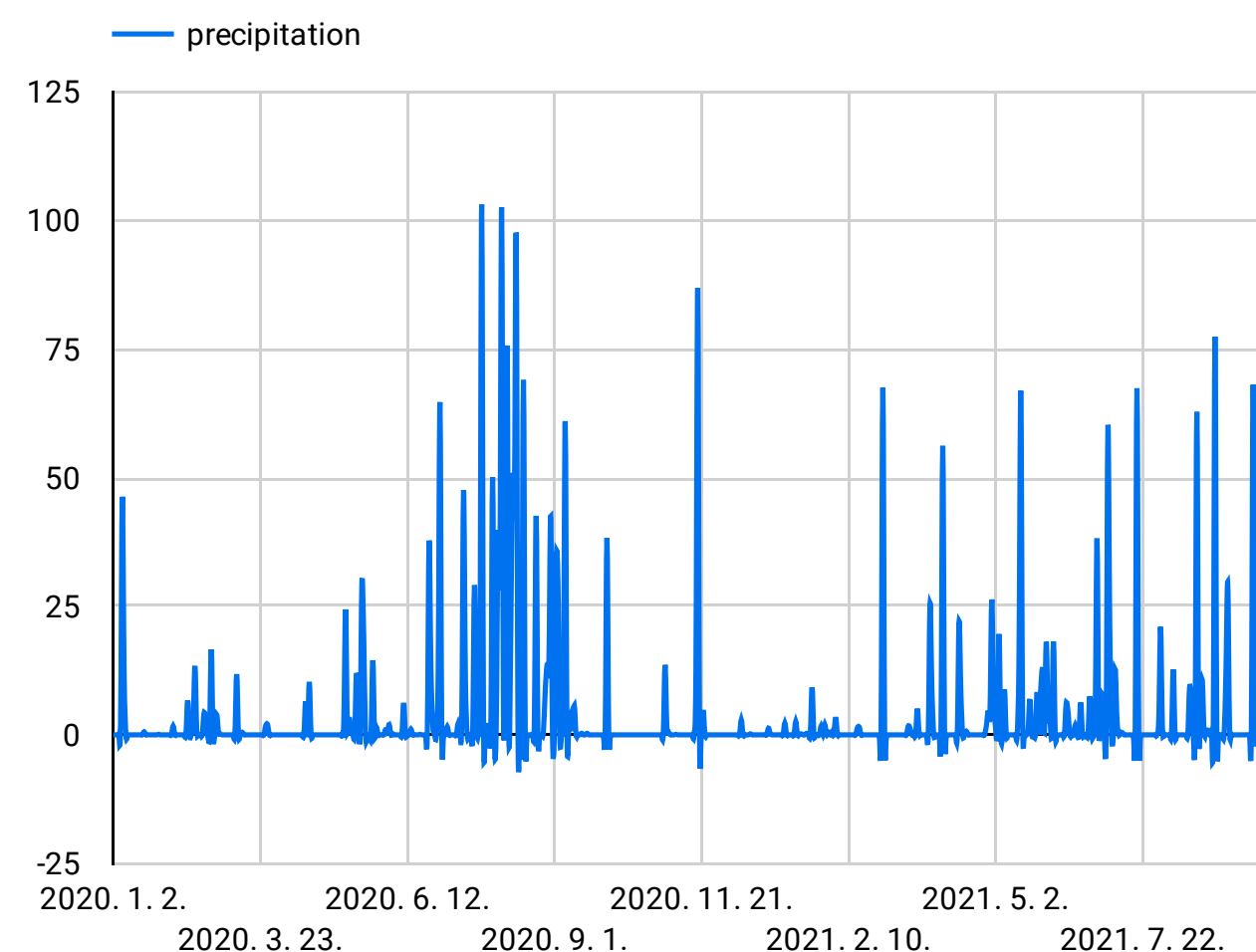
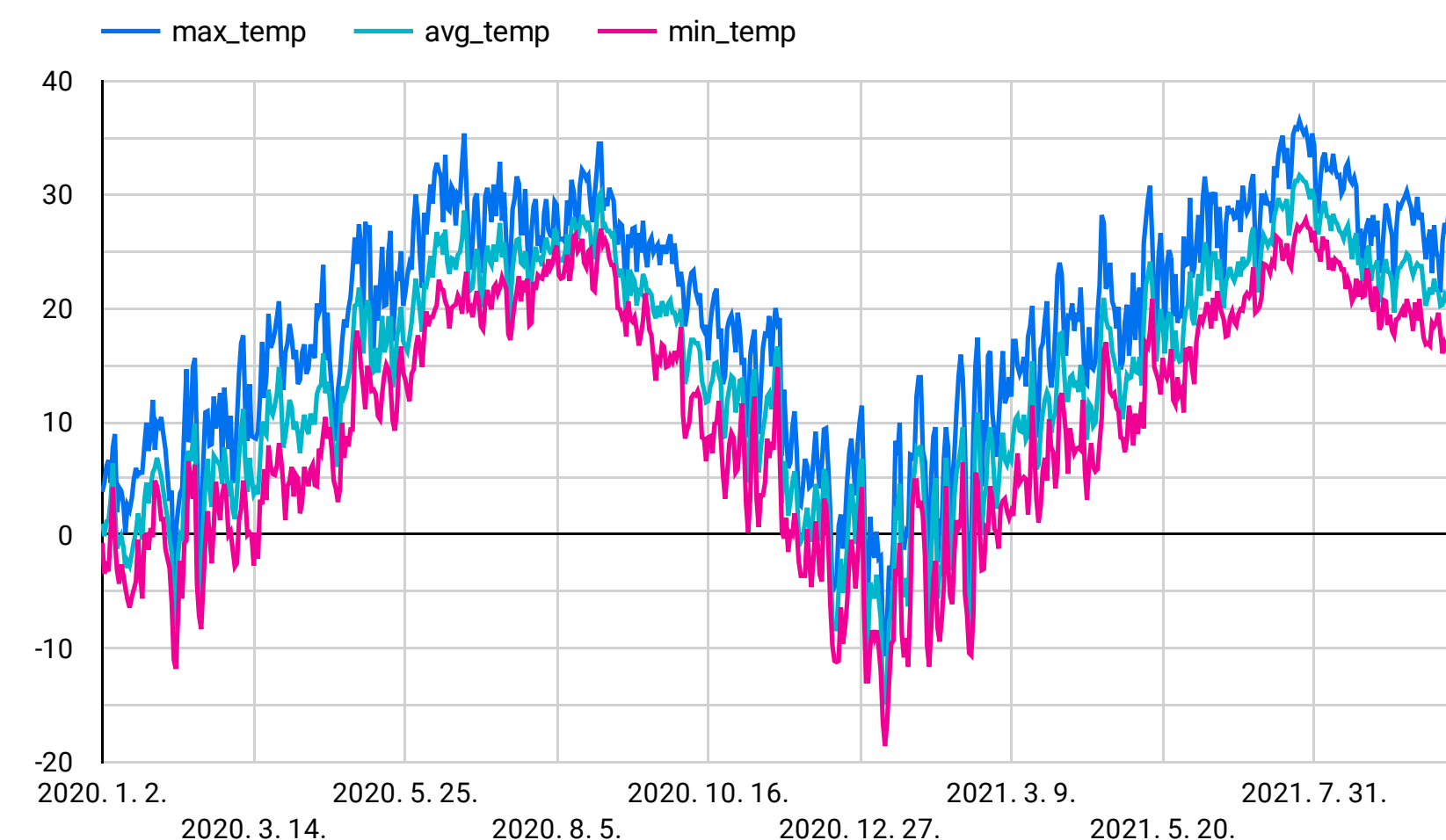
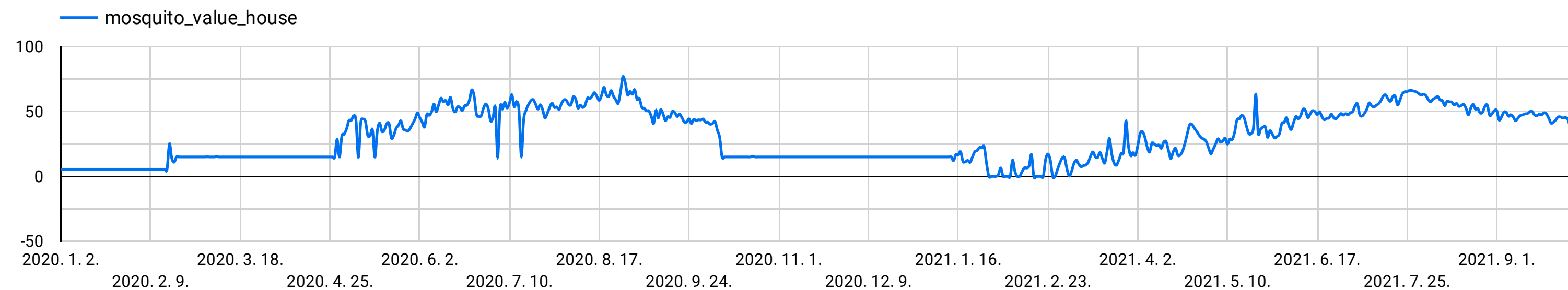
**<https://www.weather.go.kr/w/obs-climate/land/past-obs/obs-by-day.do?stn=108&yy=2017&mm=5&obs=1>

House Mosquito Status in Seoul, South Korea

since 2020.01.01~2021.10.07
data from <https://www.data.go.kr/data/15071887/fileData.do>

date
645

mosquito_value_house
30.94



Google Data Studio Dashboard

- Google DataStudio에서 postgresSQL을 지원하기 때문에 대시보드를 만들고자 데이터를 postgresSQL의 데이터베이스에 저장한 뒤 이를 시계열 자료들로 구성하여 데이터의 전체적인 흐름을 파악해보았습니다.

서비스 소개 데이터베이스 선정

MongoDB

DATABASES: 1 COLLECTIONS: 1

+ Create Database

Q NAMESPACES

▼ myFirstDatabase

- mosquito_collection

myFirstDatabase.mosquito_collection

COLLECTION SIZE: 72.66KB TOTAL DOCUMENTS: 1 INDEXES TOTAL SIZE: 36KB

Find Indexes Schema Anti-Patterns 0 Aggregation

FILTER { field: 'value' }

QUERY RESULTS 1-1 OF 1

```
_id: ObjectId("61614bdb03d160de65c1568")
> avg_temp: Object
> min_temp: Object
> max_temp: Object
> precipitation: Object
> mosquito_value_ho... : Object
```

PostgreSQL

Server fanny.db.elephantsql.com (fanny-01)

Region amazon-web-services::us-east-1

Created at 2021-10-10 17:29 UTC+00:00

User & Default database eubqnbco

Reset

Password Ueqjvs...

Reset


URL postgres://eubqnbco:***@fanny.db.elephantsql.com/eubqnbco

Reset

Current database size 136 KB

Max database size 20 MB


Active Plan



Tiny Turtle

Upgrade Instance

기계학습 모델에 사용한 데이터의 형식이 csv로 NoSQL 형식의 데이터베이스에서 관리하기가 더 적합하다 판단하여 MongoDB에 저장했으며 구글 데이터스튜디오를 이용하기 위해 구글 데이터스튜디오에서 연결을 지원하는 PostgreSQL에 csv파일을 복사하여 저장하였다.



서비스 소개 모델 소개

모기에보제의 모기발생 산식은 환경요인과 모기 개체수로 도출된다고 한다.
각 지리타입별 모기에보지수 산식은 아래와 같다.

수변부 : 최저기온*5.552+12.157

주거지: 평균기온***3.346** - 최고기온***1.114** + **0.273**

공원: 최저기온*3.290+최고기온*0.442-강수량*0.082-21.349

모기에보제가 수치형 지수이며 예측하고자 하는 대상인 주거지의 모기에보지수가 선형적인 관계로 표현될 수 있다는 점에서 선형회귀모델들을 비교하여 가장 성능이 좋은 **다중선형회귀모델(Multiple Regression)**을 선택하였다.

Ridge Regression

MAE : 6.74

R2 : 0.05

Multiple Regression

MAE : 6.70

R2 : 0.75

OLS

MAE : 7.56

R2 : 0.75

*MAE는 0에 가까울수록 모델의 성능이 좋음을 나타내는 지표이다.

*R2는 최소 0, 최대 1 사이의 수치로 표현되는 지표이며 1에 가까울수록 모델의 성능이 좋음을 나타내는 지표이다.

서비스 소개

서비스 구조도

```
seoul_mosquito_status_project
├── requirements.txt
├── __init__.py
├── model.py
├── __pycache__
│   └── __init__.cpython-38.pyc
├── templates
│   ├── home.html
│   └── results.html
├── model
│   └── model_211007.pickle
├── data
│   ├── seoul_mosquito_status.csv
│   ├── seoul_mosquito_status.json
│   ├── data_to_NoSQL.py
│   ├── data_to_progreSQL.py
│   └── raw_data
│       ├── forcast_data.csv
│       ├── rain.csv
│       ├── rn_20211008204612.csv
│       └── 서울시 모기예보제 정보.json
```

<http://127.0.0.1:5000/>

Predict House Mosquito Status in Seoul, S.Korea

서울 주거지 모기지수 예측 모델

Do you hate mosquitoes? I don't like mosquitoes too. And I always worry that I might see one of them when I open the windows.
One of government organization in Seoul Metropolitan Government called '시민건강국'(Public-health-institute, I can't find official English name.)
Anyway, I made this machine learning model to predict mosquito status in house to avoid opening my windows for too long.
Hope this model useful and I wish you luck not to get mosquito bites.
Thank you:D

Please insert ONLY numbers

1

Average Temperture(°C):

Minumum Temperture(°C):

Maximum Temperture(°C):

Precipitation(mm):

2

PREDICT!

평균기온, 최저기온, 최고기온, 강수량을 각각 입력한뒤 “PREDICT!” 버튼을 클릭한다.

서비스 소개

서비스 구조도

seoul_mosquito_status_project

requirements.txt

__init__.py

model.py

__pycache__

└─ __init__.cpython-38.pyc

templates

└─ home.html

└─ results.html

model

└─ model_211007.pickle

data

└─ seoul_mosquito_status.csv

└─ seoul_mosquito_status.json

└─ data_to_NoSQL.py

└─ data_to_progreSQL.py

└─ raw_data

└─ forecast_data.csv

└─ rain.csv

└─ rn_20211008204612.csv

└─ 서울시 모기예보제 정보.json

http://127.0.0.1:5000/predict

PREDICTION :

Mosquito Status in House is **[39.58700506]**

[Go back to main page](#)

입력된 기후에서 예측되는 모기예보지수는 대괄호 안의 수치이다.
다른 예측을 하고 싶다면 “[Go back to main page](#)”를 클릭하여 이전 화면으로 이동할 수 있다.

서비스 시연



서비스 시연은 웹페이지에서 직접
실행하는 것으로 보여드리겠습니다.



개선점 및 보완점

시계열데이터의 특성이 강한 자료이지만 시계열데이터의 특성을 반영한 머신러닝 모델을 구축하여 경향성이나 계절성 등을 고려하지 못했다.

데이터의 수가 적어 과적합의 위험이 크다.

관계형데이터베이스를 최종적으로 사용한 만큼 데이터를 업데이트하고 재구성하기 적합하게 스키마를 구성하면 좋을 것이다.

구글 데이터스튜디오에서 비관계형데이터베이스를 사용하는 방법을 찾아보면 더 유용할 것이다.

CSS와 JavaScript를 추가하여 최종사용자에게 편리한 웹페이지로 만들어줄 수 있을 것이다.

웹 어플리케이션에서 주의 정도를 표현하는 계기판과 행동요령을 넣으면 사용자경험을 더 높힐 수 있을 것이다.

웹 어플리케이션을 배포하지 못하여 로컬 환경에서만 작동한다.

감사합니다.