MLOps

1강

Database

목차

01.Overview

02.DB서버 생성

03.테이블 생성

04.데이터 삽입

05.데이터 삽입 Loop

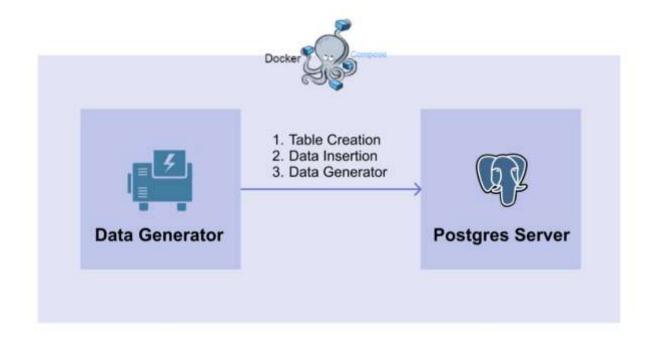
06.데이터 생성 on Docker

07.데이터 생성 on Docker Compose

Overview

- 1.Docker 를 이용하여 DB server 를 생성
- 2.psycopg2 패키지를 이용하여 테이블 생성 및 데이터 삽입
- 3.Dockerfile 과 Docker Compose 파일을 생성
- 4.Docker 컨테이너 안에서 계속해서 데이터를 생성하는 서비스를 구축

Overview



1.docker run 명령어를 사용하면, 간단한 옵션들을 통해 DB 서버를 생성 2.docker ps 명령어를 통해, DB 서버가 잘 동작하는지 확인

```
$ docker run -d \
 --name postgres-server \
 -p 5432:5432 \
 -e POSTGRES_USER=myuser \
 -e POSTGRES PASSWORD=mypassword \
 -e POSTGRES_DB=mydatabase \
 postgres:14.0
```

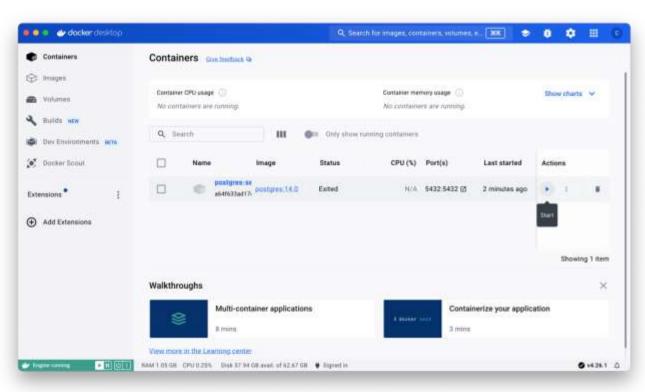


\$ docker ps

```
...
                           □ 데스크탑 — -bash — 82×29
choemin-uui-MacBookAir:~ chaiminwoo8223$ cd desktop
choemin-uui-MacBookAir:desktop chaiminwoo@223$ docker images
REPOSITORY TAG
                      IMAGE ID CREATED SIZE
choemin-uui-NacBookAir:desktop chaiminwoo@223$ docker run -d \
> --name postores-server \
> -p 5432:5432 \
   -e POSTGRES USER=myuser \
  -e POSTGRES PASSWORD=mypassword \
  -e POSTGRES D8=mydatabase \
> postgres:14.8
Unable to find image 'postgres:14.8' locally
14.0: Pulling from library/postgres
a9eb63951c1c: Pull complete
31b94a816ae6: Pull complete
83e887b8d485: Pull complete
ba77da1e8ef6: Pull complete
d76e0c23ff0f: Pull complete
b5a464f946a5: Pull complete
887b48f941ed: Pull complete
5daeef133ea7: Pull complete
Baef96543a56: Pull complete
d8828c898e6b: Pull complete
2579e9abcb64: Pull complete
4b76875b894d: Pull complete
f5a9ab29fcd4: Pull complete
Digest: sha256:db927beee892dd02fbe963559f29a7867708747934812a80fB3bff406a0d54fd
Status: Downloaded newer image for postgres:14.0
a64f633ad17dd173b9144cccc2386321ccf51c516eb8df29dae86f14bf936d9d
choemin-uui-MacBookAir:desktop chaiminwoo@223$ docker ps
```

```
■ 데스크탑 — -bash — 82×29
> -p 5432:5432 \
> -e POSTGRES USER=myuser \
> -e POSTGRES PASSWORD=mypassword \
> -e POSTGRES D8=mvdatabase \
> postgres:14.8
Unable to find image 'postgres:14.8' locally
14.0: Pulling from library/postgres
a9eb63951c1c: Pull complete
31b94a816ae6: Pull complete
03e007b8d405: Pull complete
ba77da1eBef6: Pull complete
d76e0c23ff0f: Pull complete
b5a464f946a5: Pull complete
887b48f941ed: Pull complete
5daeef133ea7: Pull complete
Baef96543a56: Pull complete
d0828c098e6b: Pull complete
2579e9abcb64: Pull complete
4b76875b894d: Pull complete
f5a9ab29fcd4: Pull complete
Digest: sha256:db927beee892dd02fbe963559f29a7867708747934812a80f83bff406a0d54fd
Status: Downloaded newer image for postgres:14.0
a64f633ad17dd173b9144cccc2386321ccf51c516eb8df29dae86f14bf936d9d
choemin-uui-MacBookAir:desktop chaiminwop@223$ docker ps
CONTAINER ID IMAGE
                              COMMAND
                                                        CREATED
                                                                         STATUS
                                NAMES
       PORTS
a64f633ad17d postgres:14.8
                              *docker-entrypoint.s_*
                                                      16 minutes ago
                                                                       Up 16 min
utes 0.0.0.8:5432->5432/tcp postgres-server
choemin-uui-MacBookAir:desktop chaiminwoo@223$
```







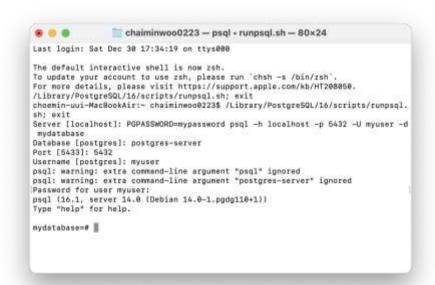
1.psql 실행

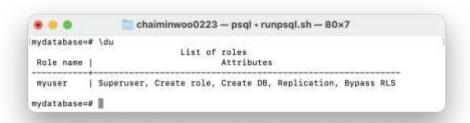
2.psql 명령어를 통해, DB 서버가 잘 동작하는지 확인

PGPASSWORD=mypassword psql -h localhost -p 5432 -U myuser -d mydatabase

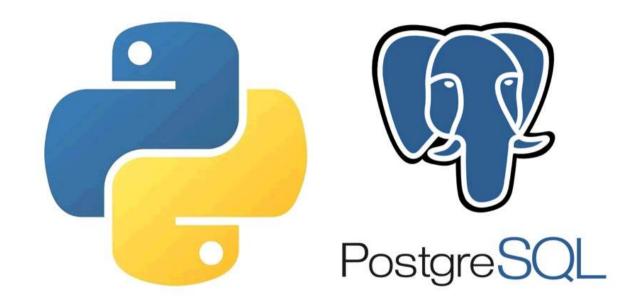
- PGPASSWORD=: 접속할 유저의 비밀번호를 입력합니다.
- -h: 호스트를 지정합니다.
- *-p*: 포트를 지정합니다.
- -U: 접속할 유저의 이름을 입력합니다.
- -a: DB 의 이름을 입력합니다.

- 1.접속 성공
- 2.\du 를 통해 DB 의 role name 과 attributes 을 확인





- 1.패키지 설치: pip3 install pandas psycopg2-binary scikit-learn
- 2.Python을 이용하여, PostgreSQL DB 서버에 접근하는 코드를 구현(psycopg2 패키지)
- 3.DB Connection → Table Creation Query → Query 실행







- 1) DB Connection
- psycopg2 를 이용하여 DB 에 접근하기 위해서는 connect 함수를 사용
- DB 에 연결할 때 user, password, host, port, database 의 총 5가지 정보가 필요

```
import psycopg2

db_connect = psycopg2.connect(
    user="myuser",
    password="mypassword",
    host="localhost",
    port=5432,
    database="mydatabase",
)
```

- 2) Table Creation Query
 - PostgreSQL에서는 float64, int 64를 지원하지 않음
 - PostgreSQL에서는 float8, int를 사용

```
create_table_query = """
CREATE TABLE IF NOT EXISTS iris_data (
   id SERIAL PRIMARY KEY,
   timestamp timestamp,
   sepal_length float8,
   sepal_width float8,
   petal_length float8,
   petal_width float8,
   target int
);"""
```

- 3) Query 실행
 - Visual Studio Code 터미널에서 python table_creator.py를 입력
 - 또는, 오른쪽 상단의 <mark>실행 버튼</mark>을 클릭

```
$ python table_creator.py

CREATE TABLE IF NOT EXISTS iris_data (
    id SERIAL PRIMARY KEY,
    timestamp timestamp,
    sepal_length float8,
    sepal_width float8,
    petal_length float8,
    petal_width float8,
    target int
);
```



- 1.psql 실행
- 2.\d를 통해 생성된 테이블들의 목록을 확인
- 3.iris_data 테이블의 데이터 전체를 확인: select * from iris_data;

- 1.생성한 테이블에 iris 데이터 한 줄 삽입
- 2.iris 데이터 불러오기 → Data Insertion Query → Query 실행



- 1) iris 데이터 불러오기
- 생성된 테이블의 Column 이름과 일치하도록 수정

```
import pandas as pd
from sklearn.datasets import load_iris

X, y = load_iris(return_X_y=True, as_frame=True)
df = pd.concat([X, y], axis="columns")
rename_rule = {
    "sepal length (cm)": "sepal_length",
    "sepal width (cm)": "sepal_width",
    "petal length (cm)": "petal_length",
    "petal width (cm)": "petal_width",
}
df = df.rename(columns=rename_rule)
```

- 2) Data Insertion Query
 - 삽입 순서 중요

```
insert_row_query = f"""
INSERT INTO iris_data
  (timestamp, sepal_length, sepal_width, petal_length, petal_width, target)

VALUES (
    NOW(),
    {data.sepal_length},
    {data.sepal_width},
    {data.petal_length},
    {data.petal_width},
    {data.target}
);"""
```

- 3) Query 실행
 - Visual Studio Code 터미널에서 python data_insertion.py를 입력
 - 또는, 오른쪽 상단의 <mark>실행 버튼</mark>을 클릭

```
$ python data_insertion.py

INSERT INTO iris_data
    (timestamp, sepal_length, sepal_width, petal_length, petal_width, target)

VALUES (
    NOW(),
    6.2,
    2.9,
    4.3,
    1.3,
    1.0
);
```

1.psql 실행

2.iris_data 테이블의 데이터 전체를 확인: select * from iris_data;

- 1.생성된 테이블 안에 데이터를 계속해서 추가
- 2.Loop 추가 → Query 실행

- 1) Loop 추가
 - while True
 - 너무 빠른 시간에 데이터가 추가되면, DB에 과부하가 생길 수 있음
- 데이터를 삽입 후 잠시 대기하는 시간을 추가: time.sleep(1)

```
import time

def generate_data(db_connect, df):
    while True:
        insert_data(db_connect, df.sample(1).squeeze())
        time.sleep(1)
```

- 2) Query 실행
 - Visual Studio Code 터미널에서 python data_insertion_loop.py를 입력
 - 또는, 오른쪽 상단의 실행 버튼을 클릭

2) Query 실행

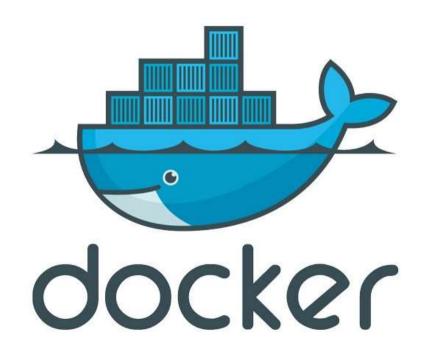
```
$ python data insertion loop.pv
   INSERT INTO iris data
       (timestamp, sepal_length, sepal_width, petal_length, petal_width, target)
       VALUES (
           NOW(),
           6.5.
           2.8.
           4.6.
           1.5.
           1.0
   INSERT INTO iris_data
        (timestamp, sepal_length, sepal_width, petal_length, petal_width, target)
       VALUES (
           NOW(),
           5.5.
           4.2.
           1.4.
           0.2.
           0.0
       );
```

1.psql 실행

2.iris_data 테이블의 데이터 전체를 확인: select * from iris_data;

id timestamp orget	sebat_tength	seber_ergtm	petal_length	petal_width t
1 2023-12-30 11:03:42.166353	1 6	1 2,7	5,1	1.6
1 2 2023-12-30 11:50:15,371534	7.2	3,2	1 6	1.8 [
3 2023-12-30 11:50:16.389618	5.5	3.5	1.3	0.2
4 2023-12-30 11:50:17.401493	5.7	1 4.4	1.5	8.4
6 5 [2023-12-30 11:50:18.412507	4.9	3.1	1 5.1	2.3 [
6 2023-12-38 11:58:19,42266	5.1	1 3.8	1 1.6	0.2
7 2023-12-38 11/58/20,43241	1 4.9	1 9	1.1	9.1
8 2023-12-30 11:50:21.443966	5.7	3	4.2	1.2 [
9 2823-12-38 11:58:22.451867	6.2	2.2	1 4.8	1.5
10 2023-12-30 11:50:23.461263	5.4	3.9	1.7	0.6 1
11 2023-12-30 11:50:24.472526	5.2	4.5	1.5	0.1
12 2023-12-30 11:50:25.480016	7.2	3.2	4 1	1.8 1
13 2023-12-30 11:50:26.491447	6.3	1 2.7	1 4.9	1.8
14 2023-12-30 11:50:27.503513	6.4	3.2	5.3	2.3
15 2023-12-30 11:50:38.515577	0.1	1 2	4.9	1.8 [

- 1.앞서 작성했던 코드를 Docker 컨테이너 안에서 실행: Dockerfile
- 2.Docker 컨테이너 간의 네트워크를 연결하여, DB에 데이터를 계속해서 삽입
- 3.Data Generator → Dockerfile → Docker Network



1) Data Generator

```
parser = ArgumentParser()
parser.add_argument("--db-host", dest="db_host", type=str, default="localhost")
args = parser.parse_args()
```

- 2) Dockerfile
- Visual Studio Code에서 Docker 확장자를 설치



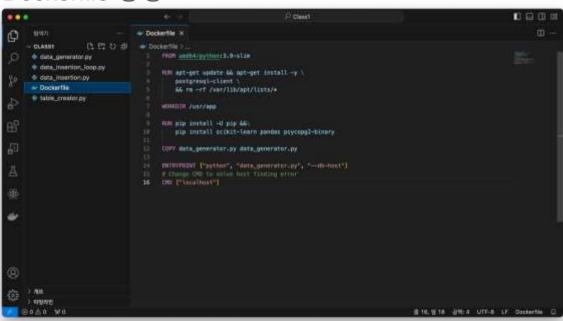
2) Dockerfile

- Dockerfile 생성

```
FROM amd64/python:3.9-slim
RUN apt-get update && apt-get install -y \
    postgresgl-client \
   && rm -rf /var/lib/apt/lists/*
WORKDIR /usr/app
RUN pip install -U pip &&\
    pip install scikit-learn pandas psycopg2-binary
COPY data_generator.py data_generator.py
ENTRYPOINT ["python", "data_generator.py", "--db-host"]
# Change CMD to solve host finding error
CMD ["localhost"]
```

2) Dockerfile

- Dockerfile 생성

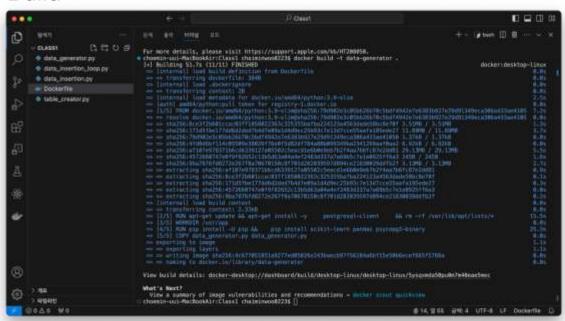


- 2) Dockerfile
 - Build

\$ docker build -t data-generator.

2) Dockerfile

- Build

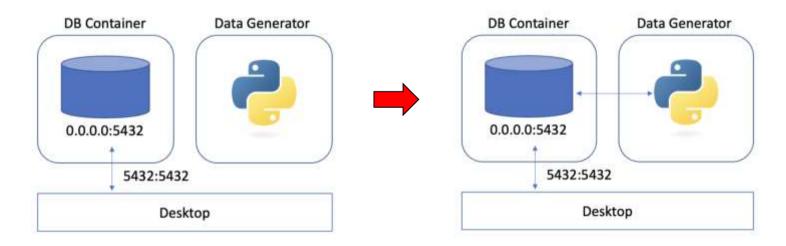


- 2) Dockerfile
 - 이미지 생성 확인

\$ docker image Is

```
• choemin-uui-MacBookAir:Class1 chaiminwoo0223$ docker image ls
REPOSITORY TAG IMAGE ID CREATED SIZE
data-generator latest 4c677051851a 4 minutes ago 592MB
postgres 14.0 01b2dbb34042 2 years ago 354MB
• choemin-uui-MacBookAir:Class1 chaiminwoo0223$ ■
```

3) Docker Network



- 3) Docker Network
 - 네트워크 생성
 - \$ docker network create my-network
 - 네트워크 연결
 - \$ docker network connect my-network postgres-server

- 3) Docker Network
 - data-generator 컨테이너 실행

```
$ docker run -d \
    --name data-generator \
    --network "my-network" \
    data-generator "postgres-server"
```

1.psql 실행

2.iris_data 테이블의 데이터 전체를 확인: select * from iris_data;

mydatabase=# select * from iris_(id timestamp arget		sepel_width	petal_length	petal_width t
1 2023-12-30 11:03:42.166353	1 6	1 2.7	5,1	1.6
2 2023-12-30 11:60:15.371534	7,2	3,2	6	1.8
3 2023-12-30 11:50:16.309618	5.5	1 3.5	1.3	0.2
4 2023-12-30 11:50:17.401693	5.7	1 4.4	1.5	8.4
5 2823-12-38 11:58:18.412587	1 4.9	3.1	1 5.1	2.3 (
6 2023-12-38 11:58:19,42266	1 5,1	1 3.6	1 2-6 1	9.2
7 2023-12-38 11/58/20,43241	1 4.9	1 9	1.1	8.1
B 2023-12-30 11:50:21.443966	5.7	3	1 4.2	1.2
9 2023-12-30 11:50:22.453067	1 6.2	1 2.2	1 4.5	1.5
10 2073-12-30 11:50:73.461763	5.4	3.9	1.7	0.6 1
11 2023-12-38 11:58:24.472526	5.2	[4.5	1.0	0.1 [
12 2023-12-30 11:50:25.480016	7,2	1 3.2	1 6 1	1.8 1
2 13 2023-12-30 11:60:26.491447	1 6.3	1 2.7	1 4.9	1.8
14 2023-12-38 11:50:27.503513	1 6.4	3.2	5.3	2.3
15 2023-12-30 11:50:38.515577	0.1) 3	1 4.9	1.8 [
2 :skipping				

- 1.DB 컨테이너와 Data Generator 컨테이너를 함께 띄우기 위한 Docker Compose 파일을 작성
- 2.Docker Compose를 이용하기 위해, DB 서버와 Data Generator를 종료
 - \$ docker rm --force postgres-server data-generator
- 3.Docker Compose → Docker Compose Healthcheck → Docker Compose Network

- 1) Docker Compose
- docker-compose.yaml 생성
- postgres server 가 띄워진 뒤에 곧바로 Data Generator 가 띄워짐 → X

```
version: "3"

services:
  postgres-server:
  data-generator:
    ...
```

- 2) Docker Compose Healthcheck
 - postgres server 가 사용 가능한 상태가 되어있는지 체크하고,
 - Data Generator 를 띄움

```
healthcheck:
    test: ["CMD", "pg_isready", "-q", "-U", "myuser", "-d", "mydatabase"]
    interval: 10s
    timeout: 5s
    retries: 5
```

- 2) Docker Compose Healthcheck
 - Compose 파일 실행
 - \$ docker compose up -d
 - Compose 파일 종료
 - \$ docker compose down -v

- 2) Docker Compose Healthcheck
 - Compose 파일 실행

```
P Chart
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     D C C D
                                                                                                         docker-compose verril ×

    docker-compose sunt

- data igenerator py

    statu insertion loop w

 - Doctorfile
                                                                                                                                                 container Hamil postgres-server
· table preatonay
                                                                                                                                                      POSTGRED PACOVOICE INVESTMENT
                                                                                                                                                      POUTGRED THE HUNDERSHAP
                                                                                                                                                       tests ("OUP", "pg_investy", "-e", "-it", "moser", "-et", "montaine").
                                                                                                                                                                                                                                                                                                                                                                                                                                                                p task Ⅲ 歳 → ペ××
                                                                                                         # channin-auj-MacSonkAir:Classi chaiminwoo@2225 docker compose up -d
                                                                                                             [+] Bullding 2.8s (11/11) FINESHID
                                                                                                                                                                                                                                                                                                                                                                                                                                                                     dockersdesktop-Linux

    Data governing internal! load hold definition from Opchertile
or transferring industrials 1888.

                                                                                                               - Itemforing center) 25
- (att-powertor interes) Lead menders for doctor. Leading typical. 1-110
- (att-powertor interes) Lead mender to doctor. Leading typical. 1-110
- (att-powertor interes) Leading typical. 1-1-110
- (att-powertor 1/2) FRM maker Leading the 1-1-110
- (att-powertor 1/2) FRM maker Leading typical. 1-1-110
- (att-powertor 1/2) FRM maker Leading ty
```

- 3) Docker Compose Network
 - 네트워크 확인

\$ docker network Is

```
choemin-uui-MacBookAir:Class1 chaiminwoo0223$ docker network ls
NETWORK ID
              NAME
                               DRIVER
                                         SCOPE.
5f45cbe2641b
               bridge
                               bridge
                                         local
0984a5bde7a4
               host
                                         local
                               host
601603b2d8fc
               mlops-network
                               bridge
                                         local
88ccda5f4e8f
                               null
                                         local
               none
choemin-uui-MacBookAir:Class1 chaiminwoo0223$
```

- 3) Docker Compose Network
 - 생성된 네트워크 확인

\$ docker network inspect [생성된 네트워크 이름]

```
networks:

default:

name: mlops-network
```

- 3) Docker Compose Network
 - 생성된 네트워크 확인

```
choemin-uui-MacBookAir:Class1 chaiminwoo8223$ docker network inspect mlops-network
          "Name": "mlops-network",
         "Id": "601603b2d8fc5a5bcc0b8d3309f533e527412421cf229d717f00d3f94a56b94b",
         "Created": "2024-01-09T08:39:39.936612919Z",
         "Scope": "local",
         "Driver": "bridge",
         "EnableIPv6": false.
         "IPAN": {
             "Driver": "default",
             "Options": null,
             "Config": [
                      "Subnet": "172.18.0.0/16",
                      "Gateway": "172.18.0.1"
         "Internal": false,
         "Attachable": false.
         "Ingress": false,
         "ConfigFrom": {
             "Network": ""
         "ConfigOnly": false,
         "Containers": {
             "ae83f6a9c458a921f80a45f1626471b67bc8f581e143f1ed13b60803f051a184": (
                  "Name": "data-generator",
                  "EndpointID": "6e2835c9679d85a23d5237a948b587789367c1b7dbe2588523bdc8c4451832ba",
                  "MacAddress": "02:42:ac:12:00:03",
                  "IPv4Address": "172.18.0.3/16",
                  "IPv6Address": ""
```

1.psql 실행

2.iris_data 테이블의 데이터 전체를 확인: select * from iris_data;

id erget	timestamp	1 *	epsl_length	!	sspel_width	1	petal_length	petal	width	1 3
1 2023-12	2-38 11:83:42.16635	L.		ı	2,7	į	5,1	ry.	1,6	1
2 2023-1	1-30 11:50:15.371534	1	7.2	Ţ	3,2	ţ	6	6)	1.8	ŧ
3 2023-11	1-38 11:58:16.389618	1.	5.5	ŧ	3.5	t	1.3		0.2	į
4 2023-12	1-38 11:58:17,481691	1.1	5.7	ı	4.4	ı	1.5		8.4	i
5 [2023-13	7-38 11:58:18.412587	1	4.9	1	3.1	1	5.1		2.3	ı
6 2023-12	-38 11:58:19.42266	1	5.1	1	3.8	1	1.6	i i	8.2	i
7 2023-13	2-38 11:68:29.43241	1	4.9	t	9	t	1.1	j)	8.1	i
8 2023-12	2-38 11:58:21.443966	61	5.7	ī	3	1	4.2	Ú.	1.2	I
9 2823-13	1-38 11:58:22,45306)	90	6.2	ı	2.2	1	4.5	ri.	1.5	ŧ
10 2073-13	2-30 11:50:23,461263	GI.	6.4	T	3.9	1	1.7	Vi.	0.4	1
11 2023-12	2-38 11:58:24.472526	cT.	9.2	Ľ	4.1	ï	1.0	n.	0.1	ı
12 2023-12	2-38 11:50:25.488016	1.1	7.2	ı	3.2	1			1.8	i
13 2023-11	1-38 11:58:26.491447	1	6.3	1	2.7	1	4.9		1.8	i
14 2023-12	r-30 11:50:27.503513	1.1	6.4	1	5.2	1	5.3		2.3	ı
15 2823-13	2-38 11:58:28.515577	1	0.1	1	3	1	4.9		1.8	ı

참고자료

- https://mlops-for-mle.github.io/tutorial/
- https://www.youtube.com/playlist?list=PLuHgQVnccGMDeMJsGq2O-55Ymtx0ldKWf
- https://khj93.tistory.com/entry/Docker-Docker-%EA%B0%9C%EB%85%90
- ChatGPT 4

Thank You