### Question 1. Understanding docker first run

Run docker with the python:3.12.8 image in an interactive mode, use the entrypoint bash.

```
What's the version of pip in the image?

24.3.1

24.2.1

23.3.1

23.2.1

Solution

cd c:/users/zhaoy/data-engineering-zoomcamp/2_docker_sql
docker run -it --entrypoint=bash python 3.12.8
```

#### Question 2. Understanding Docker networking and docker-compose

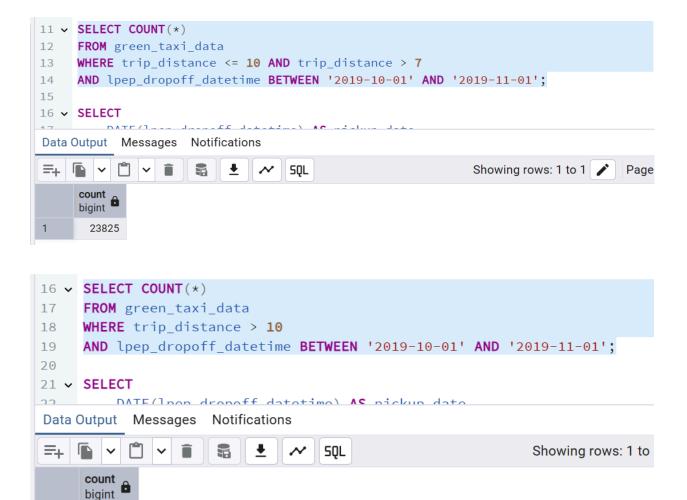
Given the following docker-compose.yaml, what is the hostname and port that pgadmin should use to connect to the postgres database?

```
services:
db:
container_name: postgres
image: postgres:17-alpine
environment:
POSTGRES_USER: 'postgres'
POSTGRES_PASSWORD: 'postgres'
POSTGRES_DB: 'ny_taxi'
ports:
- '5433:5432'
volumes:
```

- vol-pgdata:/var/lib/postgresql/data pgadmin: container name: pgadmin image: dpage/pgadmin4:latest environment: PGADMIN DEFAULT EMAIL: "pgadmin@pgadmin.com" PGADMIN\_DEFAULT\_PASSWORD: "pgadmin" ports: - "8080:80" volumes: - vol-pgadmin data:/var/lib/pgadmin volumes: vol-pgdata: name: vol-pgdata vol-pgadmin data: name: vol-pgadmin\_data postgres:5433 ☐ localhost:5432 ☐ db:5433 postgres:5432 ☑ db:5432 (I know the process, but maybe there are some issues with my dataset. So the answer is based on Excel output, attach is the screenshot of PgAdmin) **Question 3. Trip Segmentation Count** During the period of October 1st 2019 (inclusive) and November 1st 2019 (exclusive), how many trips, respectively, happened: Up to 1 mile In between 1 (exclusive) and 3 miles (inclusive). In between 3 (exclusive) and 7 miles (inclusive), In between 7 (exclusive) and 10 miles (inclusive), Over 10 miles Answers: □ 104,802; 197,670; 110,612; 27,831; 35,281 104,802; 198,924; 109,603; 27,678; 35,189 104,793; 201,407; 110,612; 27,831; 35,281 104,793; 202,661; 109,603; 27,678; 35,189 104,838; 199,013; 109,645; 27,688; 35,202

Solution

```
1 v SELECT COUNT(1)
    FROM green_taxi_data
2
3
     WHERE trip_distance <= 1</pre>
4
    AND lpep_dropoff_datetime BETWEEN '2019-10-01' AND '2019-11-01';
5
6 ▼ SELECT COUNT(*)
7
     FROM green_taxi_data
     WHERE trip_distance <= 3 AND trip_distance > 1
    AND lpep_dropoff_datetime BETWEEN '2019-10-01' AND '2019-11-01';
9
10
11 v SELECT
Data Output Messages Notifications
                                  SQL
                                                             Showing rows: 1 to 1 Page 1
bigint •
      79242
 6 ∨ SELECT COUNT(*)
     FROM green_taxi_data
 7
     WHERE trip_distance <= 3 AND trip_distance > 1
 8
     AND lpep_dropoff_datetime BETWEEN '2019-10-01' AND '2019-11-01';
 9
10
11 v SELECT
Data Output Messages Notifications
=+
    SQL
                                                             Showing rows: 1 to 1
                                                                                   Page
     count bigint
      151599
 6 ▼ SELECT COUNT(*)
     FROM green_taxi_data
     WHERE trip_distance <= 7 AND trip_distance > 3
 9
     AND lpep_dropoff_datetime BETWEEN '2019-10-01' AND '2019-11-01';
10
11 v SELECT
Data Output Messages Notifications
=+ | • | • | • | • |
                                   SQL
                                                             Showing rows: 1 to 1
                                                                                   Page
     count bigint
       89550
```



# Question 4. Longest trip for each day

Which was the pick up day with the longest trip distance? Use the pick up time for your calculations.

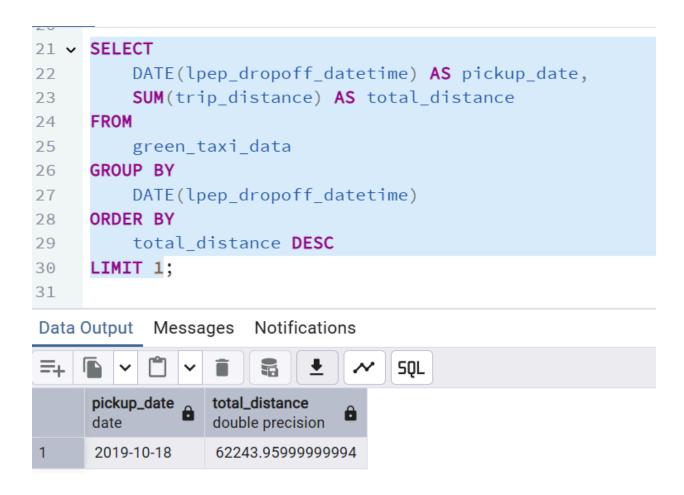
Tip: For every day, we only care about one single trip with the longest distance.

□ 2019-10-11

32002

1

- □ 2019-10-24
- □ 2019-10-26

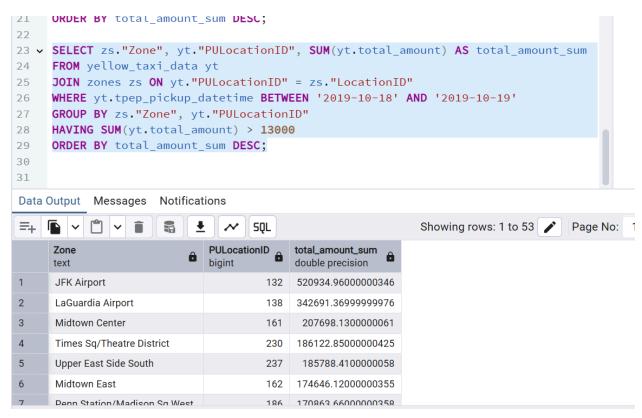


# Question 5. Three biggest pickup zones

Which were the top pickup locations with over 13,000 in total\_amount (across all trips) for 2019-10-18?

Consider only Ipep pickup datetime when filtering by date.

- □ East Harlem North, East Harlem South, Morningside Heights
   □ East Harlem North, Morningside Heights
   ☑ Morningside Heights, Astoria Park, East Harlem South
- ☐ Bedford, East Harlem North, Astoria Park



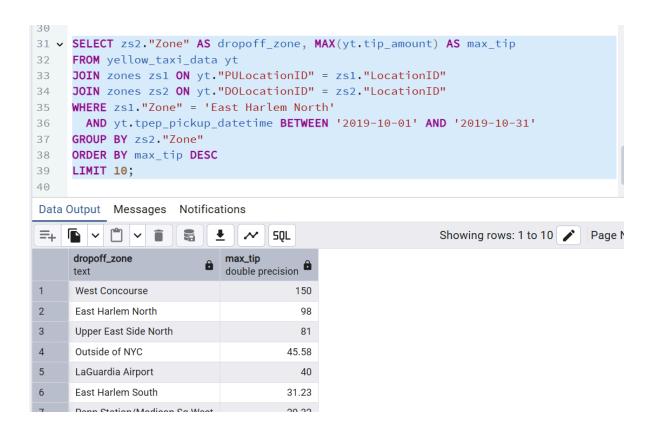
## Question 6. Largest tip

For the passengers picked up in October 2019 in the zone named "East Harlem North" which was the drop off zone that had the largest tip?

Note: it's tip, not trip

We need the name of the zone, not the ID.

- ☐ Yorkville West
- ☐ JFK Airport
- ☐ East Harlem North
- ☑ East Harlem South



Which of the following sequences, respectively, describes the workflow for:

- 1. Downloading the provider plugins and setting up backend,
- 2. Generating proposed changes and auto-executing the plan
- 3. Remove all resources managed by terraform`

## Answers:

	terraform import, terraform apply -y, terraform destroy
	teraform init, terraform plan -auto-apply, terraform rm
	terraform init, terraform run -auto-approve, terraform destroy
$\checkmark$	terraform init, terraform apply -auto-approve, terraform destroy
$\Box$	terraform import, terraform apply -v, terraform rm