

Question 1 : Which tag has the following text? - *Automatically remove the container when it exits*

Code: `` docker run --help ``

Ans: --rm

Question 2. Understanding docker first run

Run docker with the python:3.9 image in an interactive mode and the entrypoint of bash. Now check the python modules that are installed (use pip list). What is the version of the package wheel ?

Code: `` docker run -it --entrypoint=bash python:3.9 `` , `` pip list ``

Ans: 0.42.0

Question 3. Count records

How many taxi trips were totally made on September 18th 2019?

Tip: started and finished on 2019-09-18.

Code: `` pgcli -h localhost -dp 5432 -u root -d ny\_taxi `` ,

```
`` SELECT COUNT(*) AS trip_count
FROM green_taxi_data
WHERE DATE(lpep_pickup_datetime) = '2019-09-18'
AND DATE(lpep_dropoff_datetime) = '2019-09-18'; ``
```

Ans: 15612

Question 4. Largest trip for each day

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

Code: SELECT DATE(lpep_pickup_datetime) AS pickup_day,
MAX(trip_distance) AS max_trip_distance
FROM green_taxi_data
GROUP BY pickup_day
ORDER BY max_trip_distance DESC
LIMIT 1;

Ans: 2019-09-26

Question 5. Three biggest pick up Boroughs

Consider lpep_pickup_datetime in '2019-09-18' and ignoring Borough has Unknown

Which were the 3 pick up Boroughs that had a sum of total_amount superior to 50000?

Code:

```
import pandas as pd
# Assuming df_iter is your DataFrame
# Assuming 'lpep_pickup_datetime' is in string format
# Convert 'lpep_pickup_datetime' to datetime format
df_result['lpep_pickup_datetime'] = pd.to_datetime(df_result['lpep_pickup_datetime'])
# Filter rows based on conditions
filtered_df = df_result[(df_result['lpep_pickup_datetime'].dt.date ==
pd.to_datetime('2019-09-18').date()) & (df_result['Borough'] != 'Unknown')]
# Group by 'Borough' and calculate the sum of 'total_amount'
borough_totals = filtered_df.groupby('Borough')['total_amount'].sum()
# Filter Boroughs with sum of 'total_amount' greater than 50000
boroughs_over_50000 = borough_totals[borough_totals > 50000]
# Select the top 3 Boroughs
top_3_boroughs = boroughs_over_50000.nlargest(3)
# Display the result
print(top_3_boroughs)
```

Ans: "Brooklyn" "Manhattan" "Queens"

Question 6. Largest tip

For the passengers picked up in September 2019 in the zone name Astoria which was the drop off zone that had the largest tip? We want the name of the zone, not the id.

Code:

```
import pandas as pd
# Assuming df_result is your DataFrame
# Assuming 'lpep_pickup_datetime' is already converted to datetime format
# Filter rows based on conditions
filtered_df = df_result[
    (df_result['lpep_pickup_datetime'].dt.year == 2019) &
    (df_result['lpep_pickup_datetime'].dt.month == 9) &
    (df_result['Zone'] == 'Astoria')
]
# Display the resulting DataFrame
df_irene = filtered_df[['lpep_pickup_datetime', 'DOLocationID', 'tip_amount']]
```

```
# Merge with df_iter
df_result2 = pd.merge(df_irene, df_iter, how='left', left_on='DOLocationID',
right_on='LocationID')
# Find the drop-off zone with the largest tip
max_tip_zone = df_result2.loc[df_result2['tip_amount'].idxmax(), 'Zone']
```

Ans: JFK Airport

Question 7. Creating Resources

After updating the main.tf and variable.tf files run:

```
terraform apply
```

Code:

Ans: