

## Module 2 Homework

1. Within the execution for Yellow Taxi data for the year 2020 and month 12: what is the uncompressed file size (i.e. the output file `yellow_tripdata_2020-12.csv` of the extract task)?

- ☒ 128.3 MB
- ☐ 134.5 MB
- ☐ 364.7 MB
- ☐ 692.6 MB

Overview Gantt Logs Topology **Outputs** Metrics Audit Logs

Outputs

vars vars

exitCode 0

outputFiles 1 items >

yellow\_tripdata\_2020-12.csv Inte

yellow\_tripdata\_2020-12.csv

Debug Outputs

```
{{ outputs["extract"]
["outputFiles"]
["yellow_tripdata_2020-12.csv"] }}
```

Debug Outputs

Download Preview (128.3 MiB)

2. What is the rendered value of the variable `file` when the inputs `taxi` is set to `green`, `year` is set to `2020`, and `month` is set to `04` during execution?

- ☐ `{{inputs.taxi}}tripdata{{inputs.year}}-{{inputs.month}}.csv`
- ☒ `green_tripdata_2020-04.csv`
- ☐ `green_tripdata_04_2020.csv`
- ☐ `green_tripdata_2020.csv`

3. How many rows are there for the Yellow Taxi data for all CSV files in the year 2020?

- ☐ 13,537,299
- ☒ 24,648,499
- ☐ 18,324,219
- ☐ 29,430,127



- ☐ 1,428,092
- ☐ 706,911
- ☒ 1,925,152
- ☐ 2,561,031

Untitled query

RUN

SAVE

Query completed

```
1 SELECT COUNT(*)
2 FROM `dataengineer.zoomcamp.yellow_tripdata_2021_03`
3
```

Press Alt+F1 for Accessibility Options

Query results

SAVE RESULTS

OPEN IN

<

JOB INFORMATION

RESULTS

CHART

JSON

EXECUTION DETAILS

>

Row	f0_
1	1925152

6. How would you configure the timezone to New York in a Schedule trigger?
- ☐ Add a timezone property set to EST in the Schedule trigger configuration
  - ☒ Add a timezone property set to America/New\_York in the Schedule trigger configuration
  - ☐ Add a timezone property set to UTC-5 in the Schedule trigger configuration
  - ☐ Add a location property set to New\_York in the Schedule trigger configuration