MODULE 5 QUIZ **(Mayur Brijwani)**

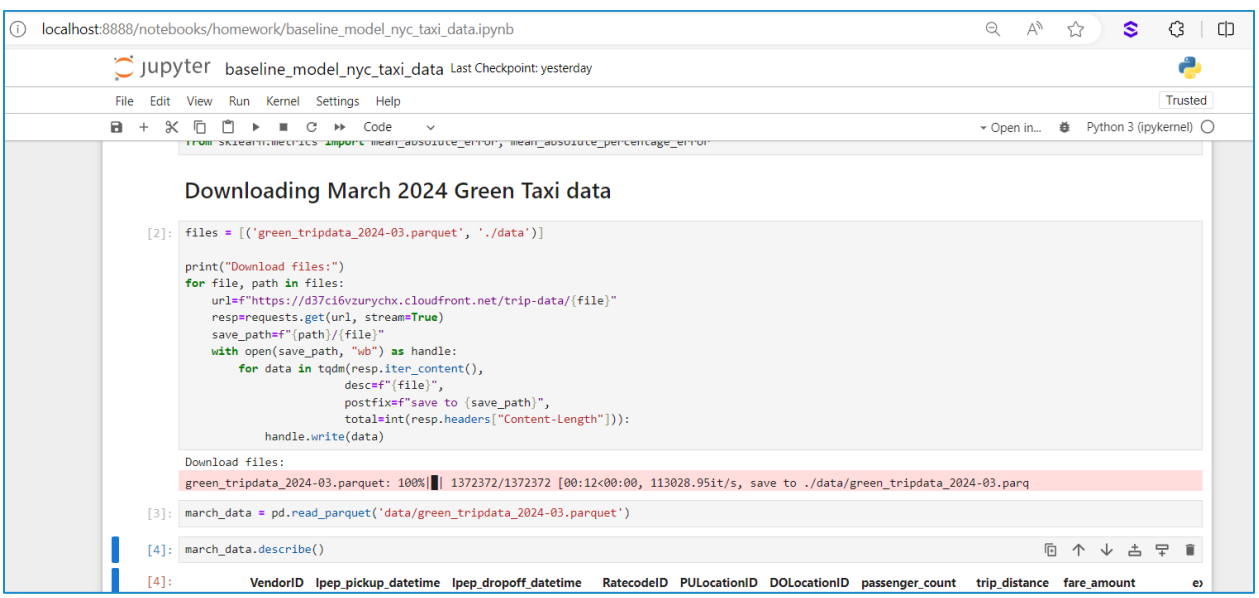
Q1. Prepare the dataset

Start with `baseline_model_nyc_taxi_data.ipynb`. Download the March 2024 Green Taxi data. We will use this data to simulate a production usage of a taxi trip duration prediction service.

What is the shape of the downloaded data? How many rows are there?

- 72044
- 78537
- 57457
- 54396

ANS: 57457



Q1. Prepare the dataset

Start with `baseline_model_nyc_taxi_data.ipynb`. Download the March 2024 Green Taxi data. We will use this data to simulate a production usage of a taxi trip duration prediction service.

What is the shape of the downloaded data? How many rows are there?

- 72044
- 78537
- 57457
- 54396

ANS: 57457

```
[7]: march_data.shape
[7]: (57457, 20)
```

Q2. Metric

Let's expand the number of data quality metrics we'd like to monitor! Please add one metric of your choice and a quantile value for the `"fare_amount"` column (`quantile=0.5`).

Hint: explore evidently metric `ColumnQuantileMetric` (from `evidently.metrics import ColumnQuantileMetric`)

What metric did you choose?

ANS: ColumnQuantileMetric

Q2. Metric

Let's expand the number of data quality metrics we'd like to monitor! Please add one metric of your choice and a quantile value for the `"fare_amount"` column (`quantile=0.5`).

Hint: explore evidently metric `ColumnQuantileMetric` (from `evidently.metrics import ColumnQuantileMetric`)

What metric did you choose?

ANS: ColumnQuantileMetric

```
21]: report = Report(metrics=[
        ColumnDriftMetric(column_name='prediction'),
        DatasetDriftMetric(),
        DatasetMissingValuesMetric(),
        ColumnQuantileMetric(column_name='fare_amount', quantile=0.5),
    ]
)
```

Q3. Monitoring

Let's start monitoring. Run expanded monitoring for a new batch of data (March 2024). What is the maximum value of metric `quantile = 0.5` on the `"fare_amount"` column during March 2024 (calculated daily)?

- 10
- 12.5
- 14.2
- 14.8

ANS: 14.2

Q3. Monitoring

Let's start monitoring. Run expanded monitoring for a new batch of data (March 2024).

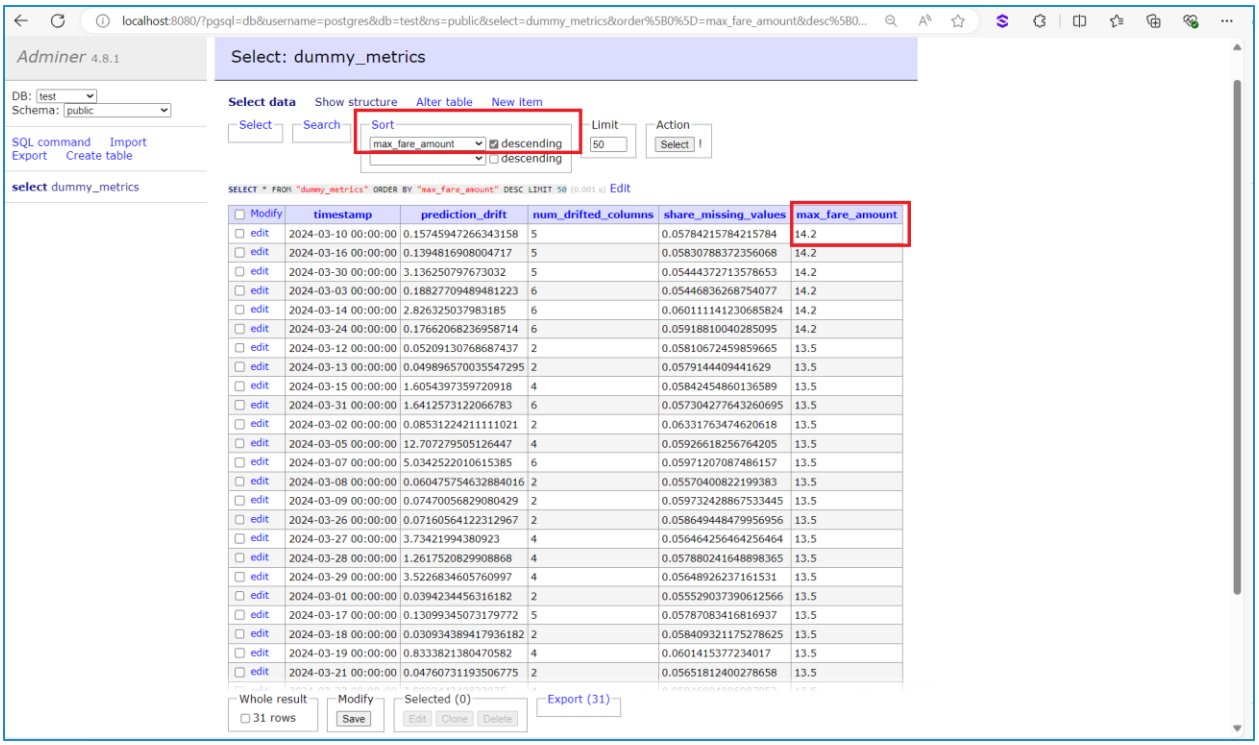
What is the maximum value of metric `quantile = 0.5` on the `"fare_amount"` column during March 2024 (calculated daily)?

- 10
- 12.5
- 14.2
- 14.8

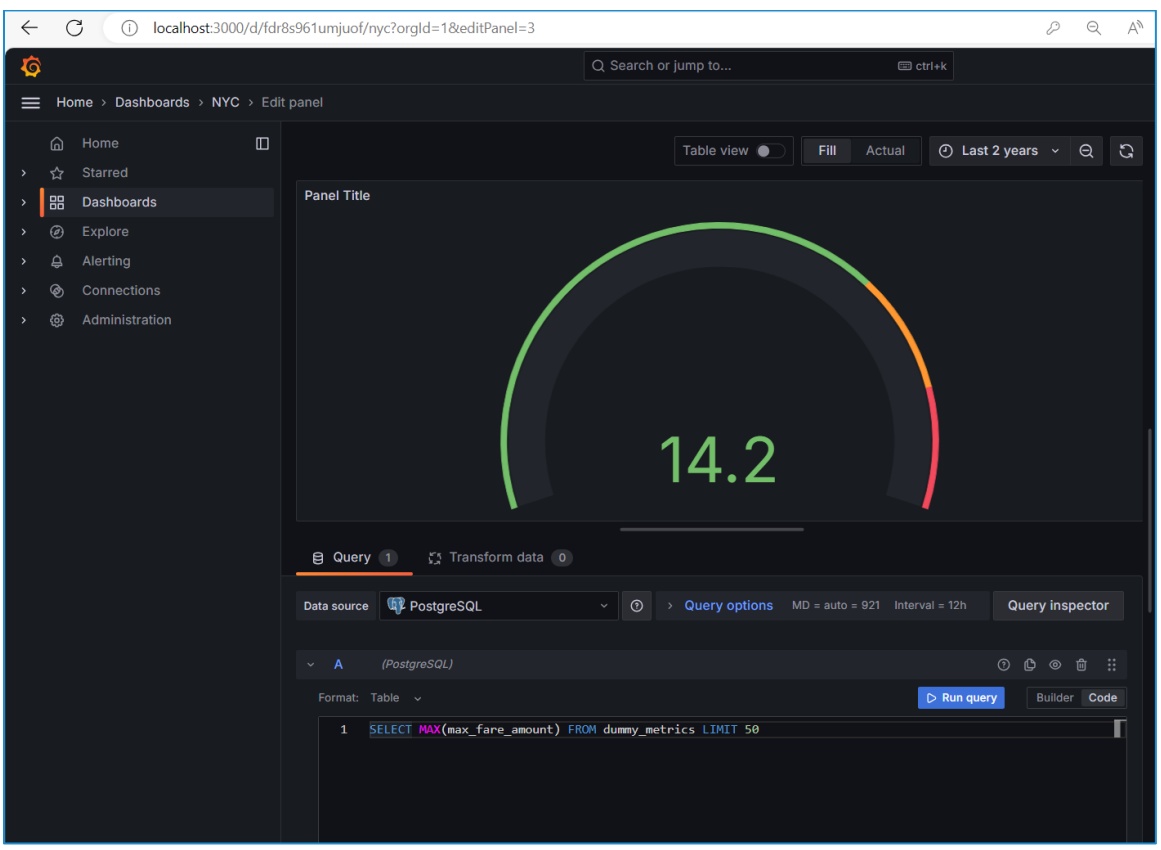
ANS: 14.2

[20]:
```
maxFare = float('-inf')
for i in range(1,31):
    max_fare_report = Report(
        metrics=[
            ColumnQuantileMetric(column_name='fare_amount', quantile=0.5),
        ],

    )

    max_fare_report.run(reference_data=None,
                        current_data=march_data.loc[march_data.lpep_pickup_datetime.between(f'2024-03-{i}', f'2024-03-{i+1}', inclusive="left")],
                        column_mapping=column_mapping)

    result = max_fare_report.as_dict()
    maxFare = max(maxFare,result['metrics'][0]['result']['current']['value'])

print(f"Maximum value is {maxFare}")
```

```
Maximum value is 14.2
```



Modify	timestamp	prediction_drift	num_drifted_columns	share_missing_values	max_fare_amount
edit	2024-03-10 00:00:00	0.15745947266343158	5	0.05784215784215784	14.2
edit	2024-03-16 00:00:00	0.1394816908004717	5	0.05830788372356068	14.2
edit	2024-03-30 00:00:00	3.136250797673032	5	0.05444372713578653	14.2
edit	2024-03-03 00:00:00	0.18827709489481223	6	0.05446836268754077	14.2
edit	2024-03-14 00:00:00	2.826325037983185	6	0.060111141230685824	14.2
edit	2024-03-24 00:00:00	0.17662068236958714	6	0.05918810040285095	14.2
edit	2024-03-12 00:00:00	0.05209130768687437	2	0.05810672459859665	13.5
edit	2024-03-13 00:00:00	0.049896570035547295	2	0.0579144409441629	13.5
edit	2024-03-15 00:00:00	1.6054397359720918	4	0.05842454860136589	13.5
edit	2024-03-31 00:00:00	1.6412573122066783	6	0.05730427764326069 5	13.5
edit	2024-03-02 00:00:00	0.08531224211111021	2	0.06331763474620618	13.5
edit	2024-03-05 00:00:00	12.707279505126447	4	0.05926618256764205	13.5
edit	2024-03-07 00:00:00	5.0342522010615385	6	0.05971207087486157	13.5
edit	2024-03-08 00:00:00	0.060475754632884016	2	0.05570400822199383	13.5
edit	2024-03-09 00:00:00	0.07470056829080429	2	0.059732428867533445	13.5
edit	2024-03-26 00:00:00	0.07160564122312967	2	0.05864944847995695 6	13.5
edit	2024-03-27 00:00:00	3.73421994380923	4	0.05646425646425464	13.5
edit	2024-03-28 00:00:00	1.2617520829908868	4	0.05788024164889836 5	13.5
edit	2024-03-29 00:00:00	3.5226834605760997	4	0.05648926237161531	13.5
edit	2024-03-01 00:00:00	0.039423445631618 2	2	0.055529037390612566	13.5
edit	2024-03-17 00:00:00	0.13099345073179772	5	0.05787083416816937	13.5
edit	2024-03-18 00:00:00	0.030934389417936182	2	0.05840932117527862 5	13.5
edit	2024-03-19 00:00:00	0.8333821380470582	4	0.0601415377234017	13.5
edit	2024-03-21 00:00:00	0.04760731193506775	2	0.05651812400278658	13.5



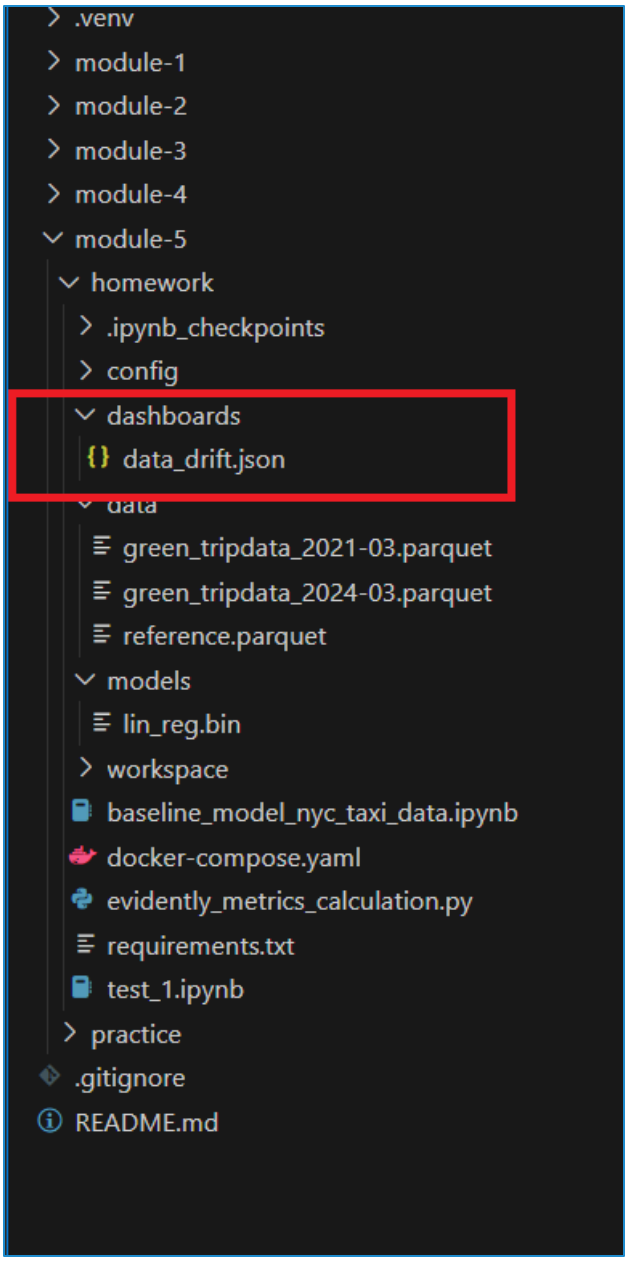
Q4. Dashboard

Finally, let’s add panels with new added metrics to the dashboard. After we customize the dashboard let's save a dashboard config, so that we can access it later. Hint: click on “Save dashboard” to access JSON configuration of the dashboard. This configuration should be saved locally.

Where to place a dashboard config file?

- `project_folder` (05-monitoring)
- `project_folder/config` (05-monitoring/config)
- `project_folder/dashboards` (05-monitoring/dashboards)
- `project_folder/data` (05-monitoring/data)

ANS: project_folder/dashboards (05-monitoring/dashboards)


.venv
module-1
module-2
module-3
module-4
module-5
homework
.ipynb_checkpoints
config
dashboards
data_drift.json
data
green_tripdata_2021-03.parquet
green_tripdata_2024-03.parquet
reference.parquet
models
lin_reg.bin
workspace
baseline_model_nyc_taxi_data.ipynb
docker-compose.yaml
evidently_metrics_calculation.py
requirements.txt
test_1.ipynb
practice
.gitignore
README.md