hmwk5

June 28, 2023

[2]:
```
!pip install evidently
```

```
Collecting evidently
  Downloading evidently-0.3.3-py3-none-any.whl (2.0 MB)
     |                  | 2.0 MB 9.5 MB/s
Collecting nltk>=3.6.7
  Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)
     |                  | 1.5 MB 106.9 MB/s
Requirement already satisfied: pandas>=1.3.5 in
/opt/conda/lib/python3.10/site-packages (from evidently) (2.0.2)
Requirement already satisfied: numpy>=1.19.5 in /opt/conda/lib/python3.10/site-
packages (from evidently) (1.24.3)
Requirement already satisfied: statsmodels>=0.12.2 in
/opt/conda/lib/python3.10/site-packages (from evidently) (0.14.0)
Requirement already satisfied: scipy>=1.5.4 in /opt/conda/lib/python3.10/site-
packages (from evidently) (1.10.1)
Requirement already satisfied: requests>=2.19.0 in
/opt/conda/lib/python3.10/site-packages (from evidently) (2.31.0)
Requirement already satisfied: pydantic<2,>=1.9 in
/opt/conda/lib/python3.10/site-packages (from evidently) (1.10.9)
Collecting plotly>=5.5.0
  Downloading plotly-5.15.0-py2.py3-none-any.whl (15.5 MB)
     |                  | 15.5 MB 107.3 MB/s
Requirement already satisfied: scikit-learn>=0.24.0 in
/opt/conda/lib/python3.10/site-packages (from evidently) (1.2.2)
Requirement already satisfied: PyYAML>=5.1 in /opt/conda/lib/python3.10/site-
packages (from evidently) (6.0)
Collecting umap-learn>=0.5.3
  Downloading umap-learn-0.5.3.tar.gz (88 kB)
     |                  | 88 kB 19.6 MB/s
  Preparing metadata (setup.py) … done
Requirement already satisfied: joblib in /opt/conda/lib/python3.10/site-
packages (from nltk>=3.6.7->evidently) (1.2.0)
Requirement already satisfied: click in /opt/conda/lib/python3.10/site-packages
(from nltk>=3.6.7->evidently) (8.1.3)
Requirement already satisfied: regex>=2021.8.3 in
/opt/conda/lib/python3.10/site-packages (from nltk>=3.6.7->evidently) (2023.6.3)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-packages
```

```
(from nltk>=3.6.7->evidently) (4.65.0)
Requirement already satisfied: tzdata>=2022.1 in /opt/conda/lib/python3.10/site-
packages (from pandas>=1.3.5->evidently) (2023.3)
Requirement already satisfied: python-dateutil>=2.8.2 in
/opt/conda/lib/python3.10/site-packages (from pandas>=1.3.5->evidently) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.10/site-
packages (from pandas>=1.3.5->evidently) (2023.3)
Collecting tenacity>=6.2.0
  Downloading tenacity-8.2.2-py3-none-any.whl (24 kB)
Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-
packages (from plotly>=5.5.0->evidently) (23.1)
Requirement already satisfied: typing-extensions>=4.2.0 in
/opt/conda/lib/python3.10/site-packages (from pydantic<2,>=1.9->evidently)
(4.6.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evidently)
(3.1.0)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-
packages (from requests>=2.19.0->evidently) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evidently)
(2023.5.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evidently)
(1.26.15)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/conda/lib/python3.10/site-packages (from scikit-learn>=0.24.0->evidently)
(3.1.0)
Requirement already satisfied: patsy>=0.5.2 in /opt/conda/lib/python3.10/site-
packages (from statsmodels>=0.12.2->evidently) (0.5.3)
Requirement already satisfied: numba>=0.49 in /opt/conda/lib/python3.10/site-
packages (from umap-learn>=0.5.3->evidently) (0.57.0)
Collecting pynndescent>=0.5
  Downloading pynndescent-0.5.10.tar.gz (1.1 MB)
     |                  | 1.1 MB 118.1 MB/s
  Preparing metadata (setup.py) … done
Requirement already satisfied: llvmlite<0.41,>=0.40.0dev0 in
/opt/conda/lib/python3.10/site-packages (from numba>=0.49->umap-
learn>=0.5.3->evidently) (0.40.0)
Requirement already satisfied: six in /opt/conda/lib/python3.10/site-packages
(from patsy>=0.5.2->statsmodels>=0.12.2->evidently) (1.16.0)
Building wheels for collected packages: umap-learn, pynndescent
  Building wheel for umap-learn (setup.py) … done
  Created wheel for umap-learn: filename=umap_learn-0.5.3-py3-none-any.whl
size=82816
sha256=2a9db83427b3cb37cff3330cf490d19479e41805f7ab3854d1245581e7c5ca5c
  Stored in directory: /home/jovyan/.cache/pip/wheels/a0/e8/c6/a37ea663620bd5200
ea1ba0907ab3c217042c1d035ef606acc
```

```
  Building wheel for pynndescent (setup.py) … done
  Created wheel for pynndescent: filename=pynndescent-0.5.10-py3-none-
any.whl size=55622
sha256=29f3ceeed98ee0633187ed6f87e1006c69e6df74e142ef74ca6d1c7813f4303b
  Stored in directory: /home/jovyan/.cache/pip/wheels/4a/38/5d/f60a40a66a9512b7e
5e83517ebc2d1b42d857be97d135f1096
Successfully built umap-learn pynndescent
Installing collected packages: tenacity, pynndescent, umap-learn, plotly, nltk,
evidently
Successfully installed evidently-0.3.3 nltk-3.8.1 plotly-5.15.0
pynndescent-0.5.10 tenacity-8.2.2 umap-learn-0.5.3
WARNING: You are using pip version 21.3.1; however, version 23.1.2 is
available.
You should consider upgrading via the '/opt/conda/bin/python3.10 -m pip install
--upgrade pip' command.
```

[34]:
```
import requests
import datetime
import pandas as pd

from evidently import ColumnMapping
from evidently.report import Report
from evidently.metrics import ColumnDriftMetric, DatasetDriftMetric,␣
 ↪DatasetMissingValuesMetric, ColumnQuantileMetric, RegressionErrorNormality

from joblib import load, dump
from tqdm import tqdm

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error
```

[9]:
```
files = [('green_tripdata_2023-03.parquet', './tripdata'),␣
 ↪('green_tripdata_2022-03.parquet', './tripdata')]

print("Download files:")
for file, path in files:
    url=f"https://d37ci6vzurychx.cloudfront.net/trip-data/{file}"
    resp=requests.get(url, stream=True)
    save_path=f"{path}/{file}"
    with open(save_path, "wb") as handle:
        for data in tqdm(resp.iter_content(),
                         desc=f"{file}",
                         postfix=f"save to {save_path}",
                         total=int(resp.headers["Content-Length"])):
            handle.write(data)
```

```
Download files:

green_tripdata_2023-03.parquet: 100%|      | 1730999/1730999 [00:08<00:00,
198205.67it/s, save to ./tripdata/green_tripdata_2023-03.parquet]
green_tripdata_2022-03.parquet: 100%|      | 1615562/1615562 [00:08<00:00,
200619.69it/s, save to ./tripdata/green_tripdata_2022-03.parquet]
```

[40]:
```
mar_data = pd.read_parquet('tripdata/green_tripdata_2023-03.parquet')
```

[41]:
```
mar_data.describe()
```

[41]:

	VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	RatecodeID	PULocationID	DOLocationID	passenger_count	trip_distance	fare_amount	extra	mta_tax	tip_amount	tolls_amount	ehail_fee
count	72044.000000	72044	72044	67507.000000	72044.000000	72044.000000	67507.000000	72044.000000	72044.000000	72044.000000	72044.000000	72044.000000	72044.000000	0.0
mean	1.865443	2023-03-16 08:02:42.689842432	2023-03-16 08:20:56.525747968	1.145703	98.437413	137.613556	1.286844	12.449511	17.018203	0.878109	0.576929	2.157651	0.183268	NaN
min	1.000000	2023-02-23 16:46:29	2023-02-23 16:55:25	1.000000	1.000000	1.000000	0.000000	0.000000	-115.000000	-5.000000	-0.500000	-1.100000	0.000000	NaN
25%	2.000000	2023-03-08 13:55:49.500000	2023-03-08 14:15:59.750000128	1.000000	74.000000	74.000000	1.000000	1.100000	9.300000	0.000000	0.500000	0.000000	0.000000	NaN
50%	2.000000	2023-03-16 08:05:27.500000	2023-03-16 08:22:27	1.000000	75.000000	138.000000	1.000000	1.840000	13.500000	0.000000	0.500000	1.600000	0.000000	NaN
75%	2.000000	2023-03-23 21:51:25.500000	2023-03-23 22:12:31.500000	1.000000	129.000000	216.000000	1.000000	3.300000	20.000000	1.000000	0.500000	3.360000	0.000000	NaN
max	2.000000	2023-04-01 00:09:27	2023-04-01 17:32:21	99.000000	265.000000	265.000000	9.000000	92064.680000	477.000000					
std	0.341252	NaN	NaN	1.225735	60.481870	76.169705	0.923652	641.094653	13.832399					

```
max       12.500000      4.250000    270.270000     36.050000        NaN
std        1.297642      0.384129      3.136215      1.139159        NaN

       improvement_surcharge  total_amount  payment_type     trip_type  \
count           72044.000000  72044.000000  67507.000000  67500.000000
mean                0.961604     22.292310      1.373680      1.028681
min                -1.000000   -116.000000      1.000000      1.000000
25%                 1.000000     12.980000      1.000000      1.000000
50%                 1.000000     18.100000      1.000000      1.000000
75%                 1.000000     26.730000      2.000000      1.000000
max                 1.000000    478.000000      5.000000      2.000000
std                 0.185185     15.852047      0.512307      0.166911

       congestion_surcharge
count          67507.000000
mean               0.714837
min               -2.750000
25%                0.000000
50%                0.000000
75%                2.750000
max                2.750000
std                1.206435
```

[42]:
```
# q1 72044
mar_data.shape
```

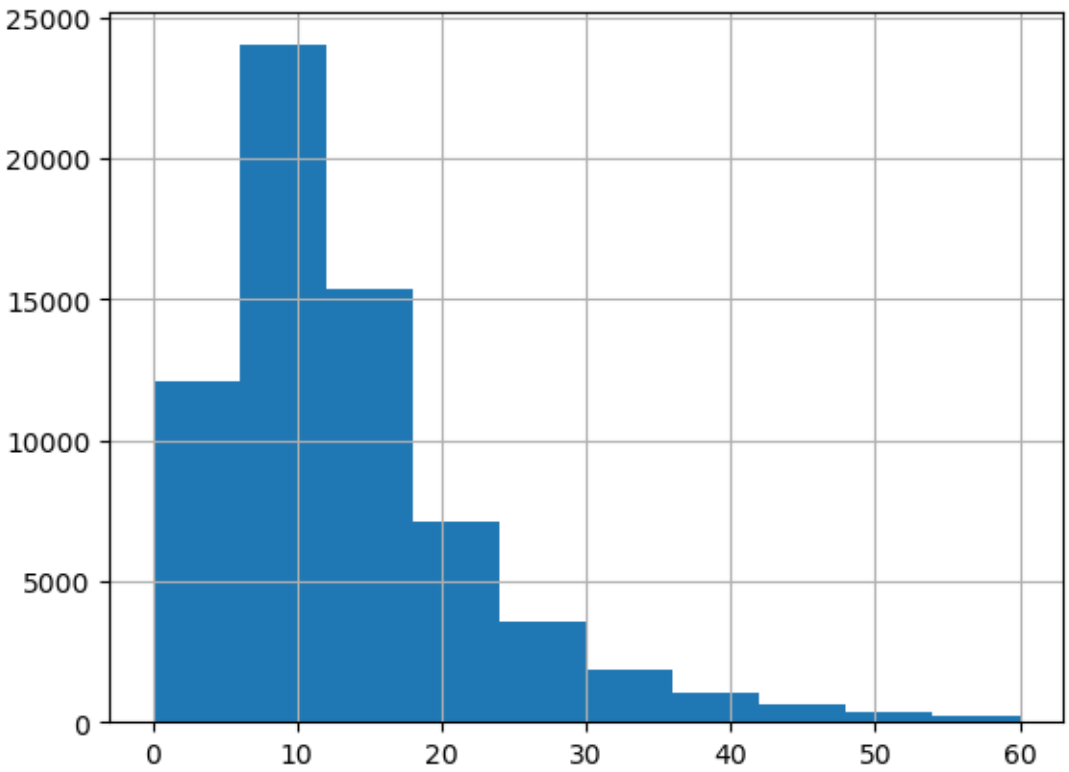
[42]:
```
(72044, 20)
```

[43]:
```
# create target
mar_data["duration_min"] = mar_data.lpep_dropoff_datetime - mar_data.
 ↪lpep_pickup_datetime
mar_data.duration_min = mar_data.duration_min.apply(lambda td : float(td.
 ↪total_seconds())/60)
```

[44]:
```
# filter out outliers
mar_data = mar_data[(mar_data.duration_min >= 0) & (mar_data.duration_min <=␣
 ↪60)]
mar_data = mar_data[(mar_data.passenger_count > 0) & (mar_data.passenger_count␣
 ↪<= 8)]
```

[45]:
```
mar_data.duration_min.hist()
```

[45]:
```
<Axes: >
```



[46]:
```
# data labeling
target = "duration_min"
num_features = ["passenger_count", "trip_distance", "fare_amount",␣
 ↪"total_amount"]
cat_features = ["PULocationID", "DOLocationID"]
```

[47]:
```
mar_data.shape
```

[47]: (66255, 21)

[48]:
```
train_data = mar_data[:30000]
val_data = mar_data[30000:]
```

[49]:
```
model = LinearRegression()
```

[50]:
```
model.fit(train_data[num_features + cat_features], train_data[target])
```

[50]: LinearRegression()

[51]:
```
train_preds = model.predict(train_data[num_features + cat_features])
train_data['prediction'] = train_preds
```

[52]:
```
val_preds = model.predict(val_data[num_features + cat_features])
val_data['prediction'] = val_preds
```

[53]:
```
print(mean_absolute_error(train_data.duration_min, train_data.prediction))
print(mean_absolute_error(val_data.duration_min, val_data.prediction))
```

```
3.5641759571064893
3.646951695051196
```

Dump model and reference data

[54]:
```
with open('models/hmwk5_lin_reg.bin', 'wb') as f_out:
    dump(model, f_out)
```

[55]:
```
val_data.to_parquet('tripdata/reference.parquet')
```

Evidently Report

[60]:
```
column_mapping = ColumnMapping(
    target='duration_min',
    prediction='prediction',
    numerical_features=num_features,
    categorical_features=cat_features
)
```

[61]:
```
# q2: RegressionErrorNormality
report = Report(metrics=[
    ColumnDriftMetric(column_name='prediction'),
    DatasetDriftMetric(),
    DatasetMissingValuesMetric(),
    RegressionErrorNormality(),
    ColumnQuantileMetric(column_name='fare_amount', quantile=0.5)
])
```

[62]:
```
report.run(reference_data=train_data, current_data=val_data,␣
 ↪column_mapping=column_mapping)
```

[63]:
```
report.show(mode='inline')
```

[63]:
```
<IPython.core.display.HTML object>
```

[64]:
```
result = report.as_dict()
```

[65]:
```
result
```

[65]:
```
{'metrics': [{'metric': 'ColumnDriftMetric',
   'result': {'column_name': 'prediction',
    'column_type': 'num',
```

```
  'stattest_name': 'Wasserstein distance (normed)',
  'stattest_threshold': 0.1,
  'drift_score': 0.021912744737353492,
  'drift_detected': False,
  'current': {'small_distribution': {'x': [-29.976938687069985,
     -6.360433663687498,
     17.25607135969499,
     40.87257638307747,
     64.48908140645996,
     88.10558642984245,
     111.72209145322493,
     135.33859647660742,
     158.95510149998992,
     182.57160652337242,
     206.1881115467549],
    'y': [9.343432451141208e-06,
     0.03443054858245535,
     0.007515623477886711,
     0.00034570700069222476,
     2.335858112785302e-05,
     8.175503394748558e-06,
     5.839645281963255e-06,
     1.1679290563926504e-06,
     2.335858112785301e-06,
     1.1679290563926519e-06]}},
  'reference': {'small_distribution': {'x': [-44.794580269945996,
     -20.705700872911244,
     3.383178524123508,
     27.472057921158253,
     51.56093731819301,
     75.64981671522777,
     99.7386961122625,
     123.82757550929726,
     147.91645490633204,
     172.0053343033668,
     196.09421370040155],
    'y': [1.522140821178265e-05,
     0.00010654985748247855,
     0.03960887169582424,
     0.0016729711389132016,
     7.610704105891324e-05,
     2.4907758892007995e-05,
     2.767528765778663e-06,
     2.767528765778661e-06,
     1.3837643828893314e-06,
     1.3837643828893314e-06]}}}},
{'metric': 'DatasetDriftMetric',
```

```
 'result': {'drift_share': 0.5,
  'number_of_columns': 8,
  'number_of_drifted_columns': 0,
  'share_of_drifted_columns': 0.0,
  'dataset_drift': False}},
{'metric': 'DatasetMissingValuesMetric',
 'result': {'current': {'different_missing_values': {'': 0,
    -inf: 0,
    None: 36257,
    inf: 0},
   'number_of_different_missing_values': 1,
   'different_missing_values_by_column': {'VendorID': {'': 0,
     -inf: 0,
     None: 0,
     inf: 0},
    'lpep_pickup_datetime': {'': 0, -inf: 0, None: 0, inf: 0},
    'lpep_dropoff_datetime': {'': 0, -inf: 0, None: 0, inf: 0},
    'store_and_fwd_flag': {'': 0, -inf: 0, None: 0, inf: 0},
    'RatecodeID': {'': 0, -inf: 0, None: 0, inf: 0},
    'PULocationID': {'': 0, -inf: 0, None: 0, inf: 0},
    'DOLocationID': {'': 0, -inf: 0, None: 0, inf: 0},
    'passenger_count': {'': 0, -inf: 0, None: 0, inf: 0},
    'trip_distance': {'': 0, -inf: 0, None: 0, inf: 0},
    'fare_amount': {'': 0, -inf: 0, None: 0, inf: 0},
    'extra': {'': 0, -inf: 0, None: 0, inf: 0},
    'mta_tax': {'': 0, -inf: 0, None: 0, inf: 0},
    'tip_amount': {'': 0, -inf: 0, None: 0, inf: 0},
    'tolls_amount': {'': 0, -inf: 0, None: 0, inf: 0},
    'ehail_fee': {'': 0, -inf: 0, None: 36255, inf: 0},
    'improvement_surcharge': {'': 0, -inf: 0, None: 0, inf: 0},
    'total_amount': {'': 0, -inf: 0, None: 0, inf: 0},
    'payment_type': {'': 0, -inf: 0, None: 0, inf: 0},
    'trip_type': {'': 0, -inf: 0, None: 2, inf: 0},
    'congestion_surcharge': {'': 0, -inf: 0, None: 0, inf: 0},
    'duration_min': {'': 0, -inf: 0, None: 0, inf: 0},
    'prediction': {'': 0, -inf: 0, None: 0, inf: 0}},
   'number_of_different_missing_values_by_column': {'VendorID': 0,
    'lpep_pickup_datetime': 0,
    'lpep_dropoff_datetime': 0,
    'store_and_fwd_flag': 0,
    'RatecodeID': 0,
    'PULocationID': 0,
    'DOLocationID': 0,
    'passenger_count': 0,
    'trip_distance': 0,
    'fare_amount': 0,
    'extra': 0,
```

```
 'mta_tax': 0,
 'tip_amount': 0,
 'tolls_amount': 0,
 'ehail_fee': 1,
 'improvement_surcharge': 0,
 'total_amount': 0,
 'payment_type': 0,
 'trip_type': 1,
 'congestion_surcharge': 0,
 'duration_min': 0,
 'prediction': 0},
'number_of_missing_values': 36257,
'share_of_missing_values': 0.0454570529456752,
'number_of_missing_values_by_column': {'VendorID': 0,
 'lpep_pickup_datetime': 0,
 'lpep_dropoff_datetime': 0,
 'store_and_fwd_flag': 0,
 'RatecodeID': 0,
 'PULocationID': 0,
 'DOLocationID': 0,
 'passenger_count': 0,
 'trip_distance': 0,
 'fare_amount': 0,
 'extra': 0,
 'mta_tax': 0,
 'tip_amount': 0,
 'tolls_amount': 0,
 'ehail_fee': 36255,
 'improvement_surcharge': 0,
 'total_amount': 0,
 'payment_type': 0,
 'trip_type': 2,
 'congestion_surcharge': 0,
 'duration_min': 0,
 'prediction': 0},
'share_of_missing_values_by_column': {'VendorID': 0.0,
 'lpep_pickup_datetime': 0.0,
 'lpep_dropoff_datetime': 0.0,
 'store_and_fwd_flag': 0.0,
 'RatecodeID': 0.0,
 'PULocationID': 0.0,
 'DOLocationID': 0.0,
 'passenger_count': 0.0,
 'trip_distance': 0.0,
 'fare_amount': 0.0,
 'extra': 0.0,
 'mta_tax': 0.0,
```

```
  'tip_amount': 0.0,
  'tolls_amount': 0.0,
  'ehail_fee': 1.0,
  'improvement_surcharge': 0.0,
  'total_amount': 0.0,
  'payment_type': 0.0,
  'trip_type': 5.5164804854502824e-05,
  'congestion_surcharge': 0.0,
  'duration_min': 0.0,
  'prediction': 0.0},
 'number_of_rows': 36255,
 'number_of_rows_with_missing_values': 36255,
 'share_of_rows_with_missing_values': 1.0,
 'number_of_columns': 22,
 'columns_with_missing_values': ['ehail_fee', 'trip_type'],
 'number_of_columns_with_missing_values': 2,
 'share_of_columns_with_missing_values': 0.09090909090909091},
'reference': {'different_missing_values': {'': 0,
  -inf: 0,
  None: 30000,
  inf: 0},
 'number_of_different_missing_values': 1,
 'different_missing_values_by_column': {'VendorID': {'': 0,
   -inf: 0,
   None: 0,
   inf: 0},
  'lpep_pickup_datetime': {'': 0, -inf: 0, None: 0, inf: 0},
  'lpep_dropoff_datetime': {'': 0, -inf: 0, None: 0, inf: 0},
  'store_and_fwd_flag': {'': 0, -inf: 0, None: 0, inf: 0},
  'RatecodeID': {'': 0, -inf: 0, None: 0, inf: 0},
  'PULocationID': {'': 0, -inf: 0, None: 0, inf: 0},
  'DOLocationID': {'': 0, -inf: 0, None: 0, inf: 0},
  'passenger_count': {'': 0, -inf: 0, None: 0, inf: 0},
  'trip_distance': {'': 0, -inf: 0, None: 0, inf: 0},
  'fare_amount': {'': 0, -inf: 0, None: 0, inf: 0},
  'extra': {'': 0, -inf: 0, None: 0, inf: 0},
  'mta_tax': {'': 0, -inf: 0, None: 0, inf: 0},
  'tip_amount': {'': 0, -inf: 0, None: 0, inf: 0},
  'tolls_amount': {'': 0, -inf: 0, None: 0, inf: 0},
  'ehail_fee': {'': 0, -inf: 0, None: 30000, inf: 0},
  'improvement_surcharge': {'': 0, -inf: 0, None: 0, inf: 0},
  'total_amount': {'': 0, -inf: 0, None: 0, inf: 0},
  'payment_type': {'': 0, -inf: 0, None: 0, inf: 0},
  'trip_type': {'': 0, -inf: 0, None: 0, inf: 0},
  'congestion_surcharge': {'': 0, -inf: 0, None: 0, inf: 0},
  'duration_min': {'': 0, -inf: 0, None: 0, inf: 0},
  'prediction': {'': 0, -inf: 0, None: 0, inf: 0}},
```

```
'number_of_different_missing_values_by_column': {'VendorID': 0,
 'lpep_pickup_datetime': 0,
 'lpep_dropoff_datetime': 0,
 'store_and_fwd_flag': 0,
 'RatecodeID': 0,
 'PULocationID': 0,
 'DOLocationID': 0,
 'passenger_count': 0,
 'trip_distance': 0,
 'fare_amount': 0,
 'extra': 0,
 'mta_tax': 0,
 'tip_amount': 0,
 'tolls_amount': 0,
 'ehail_fee': 1,
 'improvement_surcharge': 0,
 'total_amount': 0,
 'payment_type': 0,
 'trip_type': 0,
 'congestion_surcharge': 0,
 'duration_min': 0,
 'prediction': 0},
'number_of_missing_values': 30000,
'share_of_missing_values': 0.045454545454545456,
'number_of_missing_values_by_column': {'VendorID': 0,
 'lpep_pickup_datetime': 0,
 'lpep_dropoff_datetime': 0,
 'store_and_fwd_flag': 0,
 'RatecodeID': 0,
 'PULocationID': 0,
 'DOLocationID': 0,
 'passenger_count': 0,
 'trip_distance': 0,
 'fare_amount': 0,
 'extra': 0,
 'mta_tax': 0,
 'tip_amount': 0,
 'tolls_amount': 0,
 'ehail_fee': 30000,
 'improvement_surcharge': 0,
 'total_amount': 0,
 'payment_type': 0,
 'trip_type': 0,
 'congestion_surcharge': 0,
 'duration_min': 0,
 'prediction': 0},
'share_of_missing_values_by_column': {'VendorID': 0.0,
```

```
            'lpep_pickup_datetime': 0.0,
            'lpep_dropoff_datetime': 0.0,
            'store_and_fwd_flag': 0.0,
            'RatecodeID': 0.0,
            'PULocationID': 0.0,
            'DOLocationID': 0.0,
            'passenger_count': 0.0,
            'trip_distance': 0.0,
            'fare_amount': 0.0,
            'extra': 0.0,
            'mta_tax': 0.0,
            'tip_amount': 0.0,
            'tolls_amount': 0.0,
            'ehail_fee': 1.0,
            'improvement_surcharge': 0.0,
            'total_amount': 0.0,
            'payment_type': 0.0,
            'trip_type': 0.0,
            'congestion_surcharge': 0.0,
            'duration_min': 0.0,
            'prediction': 0.0},
           'number_of_rows': 30000,
           'number_of_rows_with_missing_values': 30000,
           'share_of_rows_with_missing_values': 1.0,
           'number_of_columns': 22,
           'columns_with_missing_values': ['ehail_fee'],
           'number_of_columns_with_missing_values': 1,
           'share_of_columns_with_missing_values': 0.045454545454545456}}},
       {'metric': 'RegressionErrorNormality', 'result': {}},
       {'metric': 'ColumnQuantileMetric',
        'result': {'column_name': 'fare_amount',
         'column_type': 'num',
         'quantile': 0.5,
         'current': {'value': 12.8},
         'reference': {'value': 12.8}}}]}
```

[26]:
```
#prediction drift
result['metrics'][0]['result']['drift_score']
```

[26]:
```
0.021912744737353492
```

[27]:
```
#number of drifted columns
result['metrics'][1]['result']['number_of_drifted_columns']
```

[27]:
```
0
```

[28]:
```
#share of missing values
result['metrics'][2]['result']['current']['share_of_missing_values']
```

[28]:
```
0.0454570529456752
```

[67]:
```
#share of missing values
print(result['metrics'][4]['result']['current']['value'])
print(result['metrics'][4]['result']['reference']['value'])
```

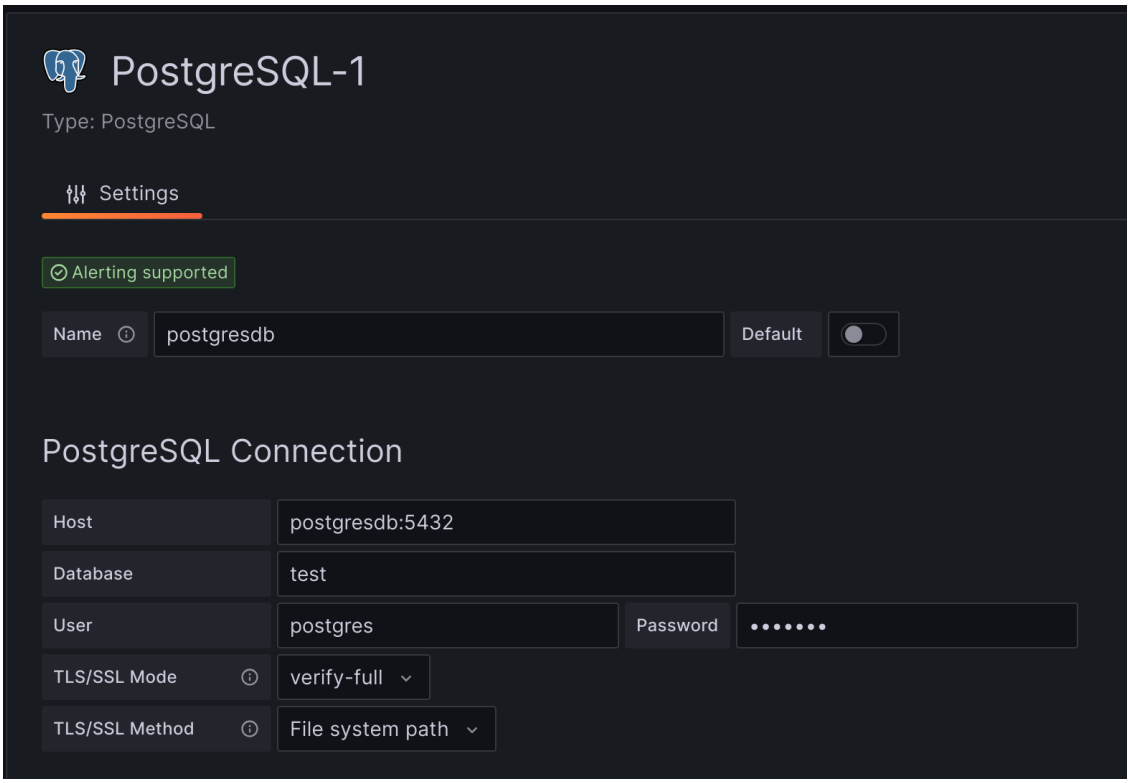
```
12.8
12.8
```

[30]:
```
#q3: @task(retries=2, retry_delay_seconds=5, name="calculate metrics")
!pip install psycopg psycopg_binary
```

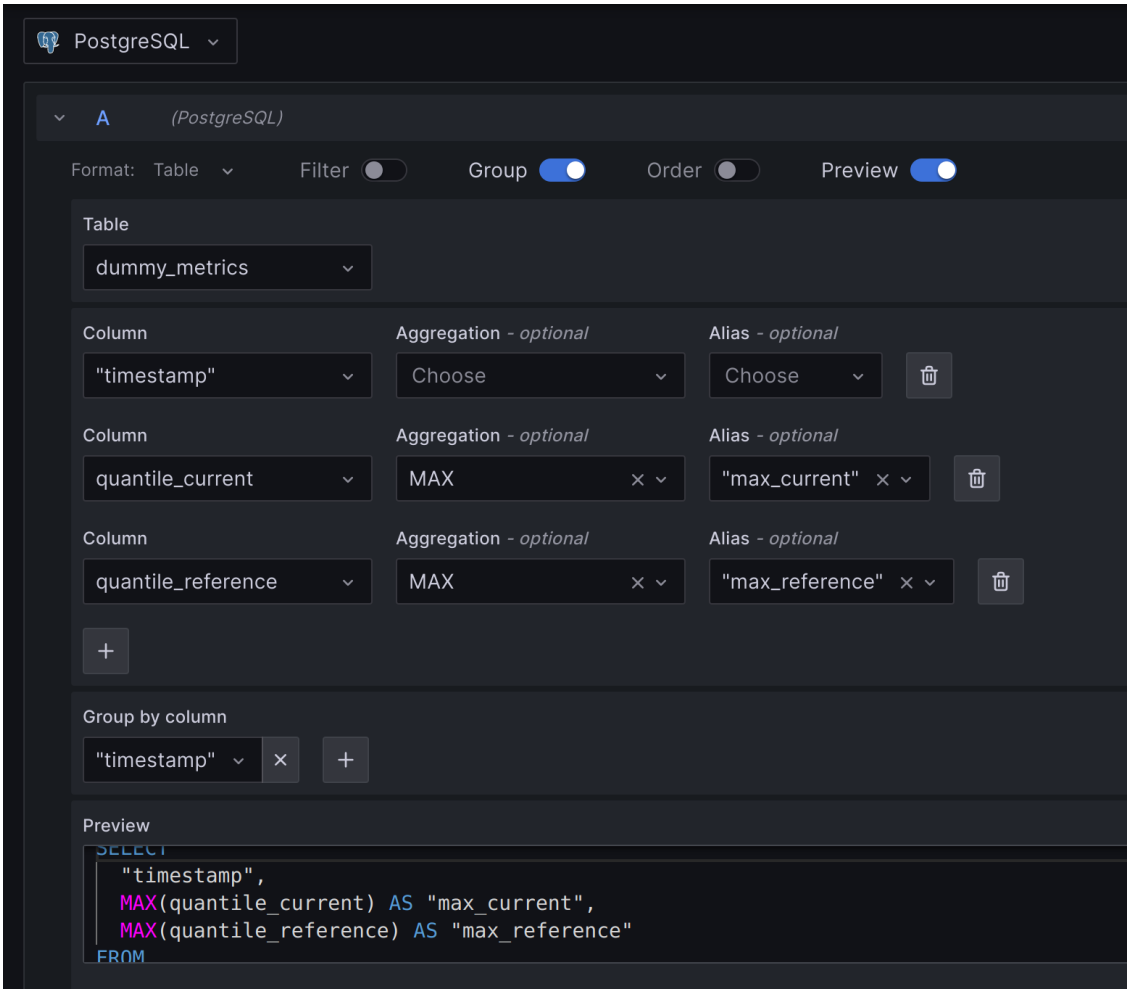
```
Collecting psycopg
  Downloading psycopg-3.1.9-py3-none-any.whl (167 kB)
     |                  | 167 kB 14.9 MB/s
Collecting psycopg_binary
  Downloading
psycopg_binary-3.1.9-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(3.3 MB)
     |                  | 3.3 MB 107.2 MB/s
Requirement already satisfied: typing-extensions>=4.1 in
/opt/conda/lib/python3.10/site-packages (from psycopg) (4.6.2)
Installing collected packages: psycopg-binary, psycopg
Successfully installed psycopg-3.1.9 psycopg-binary-3.1.9
WARNING: You are using pip version 21.3.1; however, version 23.1.2 is
available.
You should consider upgrading via the '/opt/conda/bin/python3.10 -m pip install
--upgrade pip' command.
```

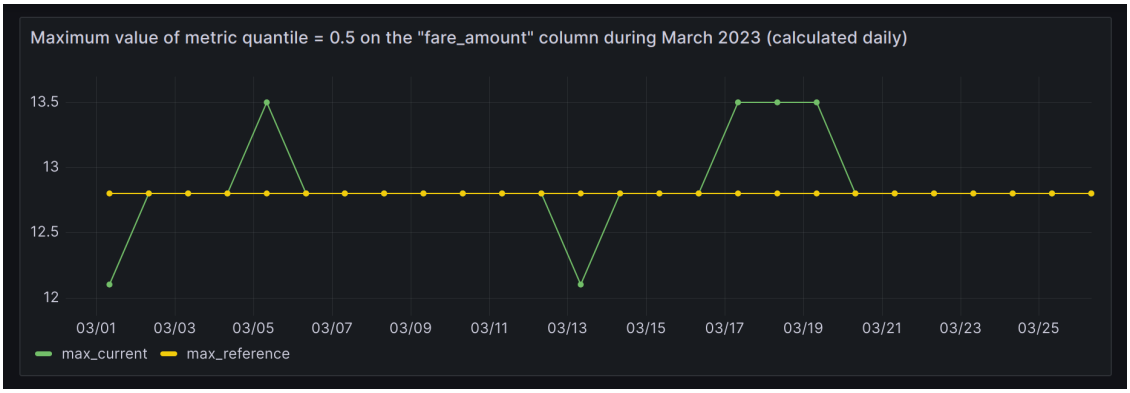
[31]:
```
# open File -> new -> terminal, run the command to have proper access to␣
 ↪postgres container
#!python3 evidently_metrics_calculation.py
```

[32]:
```
# add postgresdb connector to grafana (postgresdb - container name, postgres -␣
 ↪default user and the password set in docker-compose (disable SSL)
```



[ ]:
```
# check sql query before adding to a dashboard
```





[ ]:
```
# maximum value of metric quantile = 0.5 on the "fare_amount" column during␣
↪March 2023 (calculated daily) is 13.5
```

[ ]:
```
# json config of the dashboard is saved under project_folder/dashboards
```