

# Unified ML Monitoring Hands On Lab Guide

## Setting Up the Lab Environment

1. Go to this [GitHub Repository](#) and clone or download the contents to your local computer
2. For now, all students will go to the /Azure folder and upload these notebooks to their Databricks Workspace
  - a. To upload these notebooks to your workspace, do the following:
    - i. Open your Databricks Workspace
    - ii. Go to the Workspace tab that can be accessed in the left hand bar of the workspace web interface
    - iii. Create a folder for this lab or leverage an existing folder
    - iv. Click on the folder and right click in the panel of the workspace tab that holds the folder's contents
    - v. Click "Import"
    - vi. Choose one of the notebooks from the cloned / copied repository from GitHub
    - vii. Click "Import"
    - viii. Repeat steps iv through vii for the other notebooks in the /Azure folder of the GitHub repository
3. Create or start a cluster in the Databricks workspace that you are using for the lab and choose Databricks Runtime 7.5 ML
  - a. Make sure the cluster auto-termination is set for 65 minutes (we do not want it timing out before the end of the lab)
4. Create or start a Small SQL Analytics endpoint, if you have SQL Analytics enabled for your workspace that you are using for the lab
  - a. **If you do not have SQL Analytics**, check below for the guide for how to recreate what we do in SQL Analytics so when you talk to your Databricks account team to enable it you will be able to replicate what we did here today.

# SQL Analytics Configuration Instructions

## How to Set Up Main Visuals Seen in the Lab

1. For the MLFlow Query
  - a. Create a new query
  - b. Enter in the following SQL

```
SELECT
  run_id,
  experiment_id,
  start_time,
  end_time,
  ccp_alpha,
  max_depth,
  min_samples_leaf,
  min_samples_split,
  mse,
  n_estimators
FROM
  umlworkshop.experiment_data_bronze
```

- c. Click Save
- d. Create a Visual in the Query Page, and select the following options

Visualization Editor

Visualization Type

Details View

Visualization Name

MLFlow Details

run_id	29f38d7fc07f4c2b86974f500f030934
experiment_id	3718600673158446
start_time	2021-01-29 15:08:40.528
end_time	2021-01-29 15:08:41.752
ccp_alpha	null
max_depth	null
min_samples_leaf	null
min_samples_split	null
mse	0.011233828030526638
n_estimators	null

< 1 2 3 4 5 ... 38 >

i.

## 2. For the Azure Application Insights Query

### a. SQL Query

```
SELECT
timestamp,
requestId,
mappedInputandPrediction.response AS Sensor5_Prediction,
mappedInputandPrediction.input[0].Sensor1 AS Sensor1_Input,
mappedInputandPrediction.input[0].Sensor2 AS Sensor2_Input,
mappedInputandPrediction.input[0].Sensor3 AS Sensor3_Input,
mappedInputandPrediction.input[0].Sensor4 AS Sensor4_Input
FROM
umlworkshop.response_data_silver
ORDER BY timestamp DESC
```

### b. Visualization

#### i. Details View

Visualization Editor

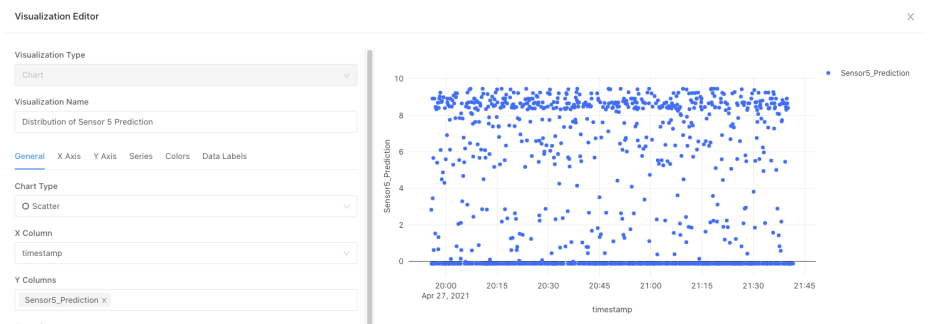
Visualization Type	
Details View	
Visualization Name	
Request View	

timestamp	2021-04-27 21:41:38.894
requestId	aadf5bd3-632c-4771-bdc9-5744ec52e1c8
Sensor5_Prediction	-0.1143052726984024
Sensor1_Input	75.72013092041016
Sensor2_Input	12121.94140625
Sensor3_Input	133.80174255371094
Sensor4_Input	74.3465805053711

< 1 2 3 4 5 ... 1000 >

#### 1.

#### ii. Distribution of Prediction



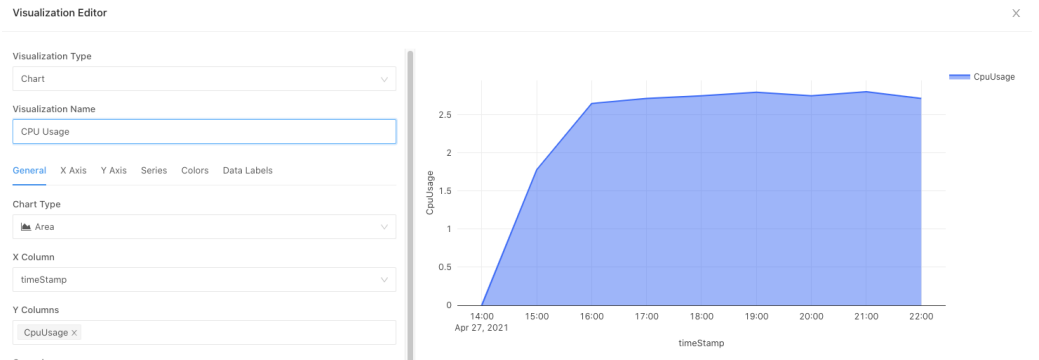
#### 1.

### 3. For the Azure Metrics Query

#### a. SQL Query

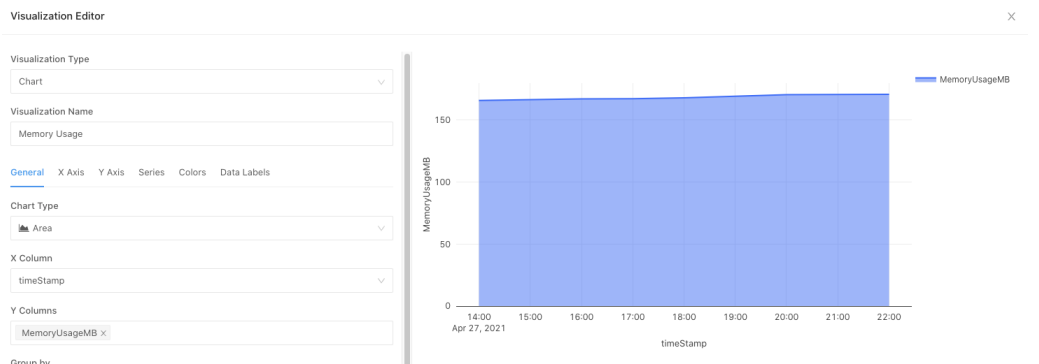
```
SELECT *  
FROM umlworkshop.endpoint_metrics_bronze
```

#### b. CPU Usage



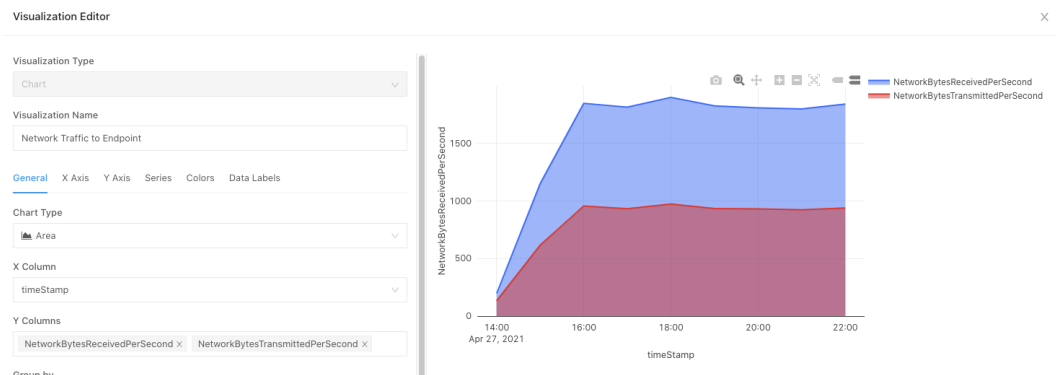
i.

#### c. Memory Usage



i.

#### d. Network Traffic



i.

#### 4. For the Dataset Drift Query

##### a. SQL Query

```
SELECT  
(sensor_data_drift.Sensor1+sensor_data_drift.Sensor2+sensor_data_drift.S  
ensor3+sensor_data_drift.Sensor4)/4 as avg_drfit_across_sensors  
FROM umlworkshop.sensor_data_drift  
ORDER BY _ts DESC
```

##### b. Visualization

Visualization Editor

Visualization Type  
Counter

Visualization Name  
Average Data Drift

General Format

Counter Label  
Average Data Drift

Counter Value Column Name  
counter

Counter Value Row Number  
1

Target Value Column Name  
avg\_drfit\_across\_sensors

Target Value Row Number  
1

☐ Count Rows

(0.672)  
Average Data Drift

i.