

Akilesh K

k.akilesh123@gmail.com

Data engineering - Batch 1

Date: 13-02-24

DAY 17 – Azure Databricks-data visualization

We intend to visualize the data using charts or graphs. we should be able to see the list of charts and graphs supported by Azure Databricks as shown below. Some of the most used charts are very well supported here like Bar chart, Scatter chart, Maps, Line chart, Area chart, Pie chart etc.

Data profiling is the process of collecting statistics and summaries of data to assess its quality and other characteristics.

Read csv file and display dataframe

Untitled Notebook 2024-02-13 11:38:35 Python ☆

File Edit View Run Help Last edit was 2 minutes ago New cell UI: OFF ▾

▶ Run all azuser1079_mm1.local's... Schedule Share

Cmd 1

```
1 sparkDF = spark.read.csv("/databricks-datasets/bikeSharing/data-001/day.csv", header="true", inferSchema="true")
2 display(sparkDF)
```

▶ (3) Spark Jobs

▶ sparkDF: pyspark.sql.dataframe.DataFrame = [instant: integer, dteday: date ... 14 more fields]

Table ▾ + New result table: OFF ▾

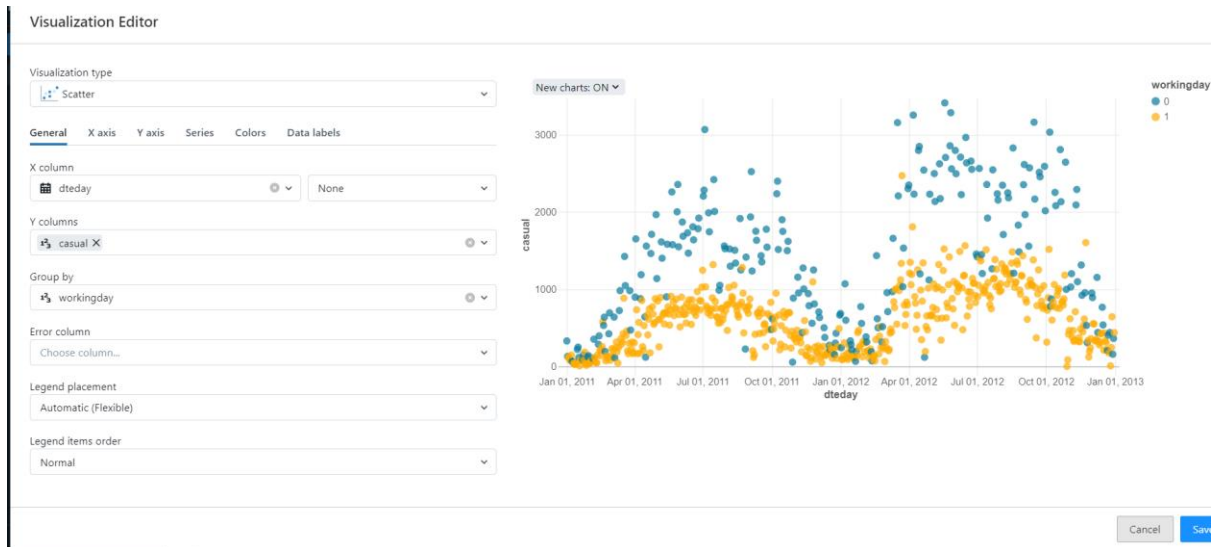
	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum
1	1	2011-01-01	1	0	1	0	6	0	2	0.344167	0.363625	0.81
2	2	2011-01-02	1	0	1	0	0	0	2	0.363478	0.353739	0.67
3	3	2011-01-03	1	0	1	0	1	1	1	0.196364	0.189405	0.41
4	4	2011-01-04	1	0	1	0	2	1	1	0.2	0.212122	0.51
5	5	2011-01-05	1	0	1	0	3	1	1	0.226957	0.22927	0.41
6	6	2011-01-06	1	0	1	0	4	1	1	0.204348	0.233209	0.51
7	7	2011-01-07	1	0	1	0	5	1	2	0.196522	0.208839	0.41

731 rows | 34.46 seconds runtime Refreshed 2 minutes ago

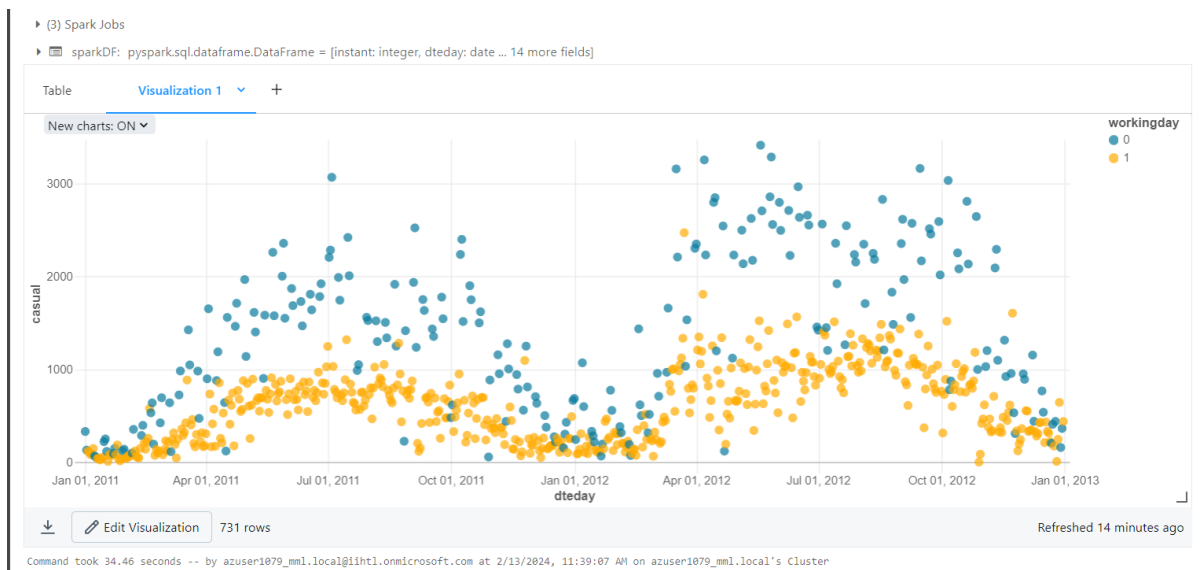
Command took 34.46 seconds -- by azuser1079_mm1.local@i1htl.onmicrosoft.com at 2/13/2024, 11:39:07 AM on azuser1079_mm1.local's Cluster

[Shift+Enter] to run
[Shift+Ctrl+Enter] to run selected text

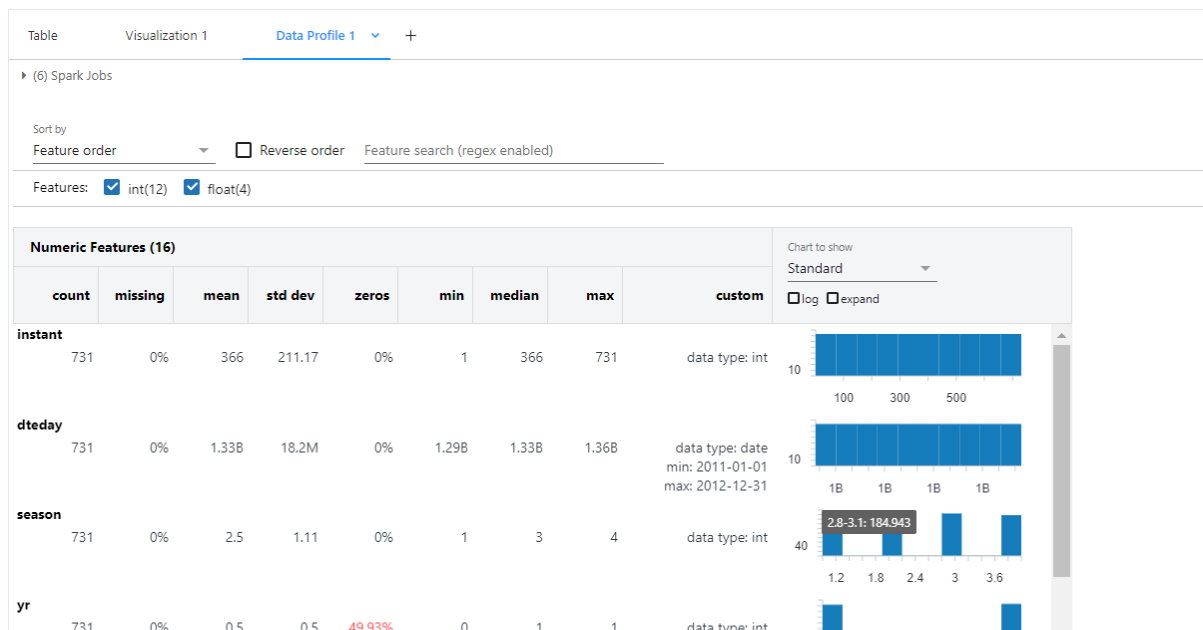
Using scatter plot and choosing its axis



Visualization of the scatter plot



Getting Data profile of the data frame



All the attribute of the table is displayed as data profile

