

## persistentMetricsRepository

```
entity instance name value dataset_date VY_ML_00 tags
Column nr_cpf[Completeness] 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, ...]
Dataset " " Size 500.0 1635087280875 2021-10-25 [tag1x, tag2iy, ...]
Column dt_init_camp[Completeness] 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, ...]
Column vl_rating_risc Minimum 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, ...]
Column vl_rating_risc Maximum 9.0 1635087280875 2021-10-25 [tag1x, tag2iy, ...]
Column nr_cpf[Completeness] 1.0 1635088521742 2021-10-25 [tag1x, tag2iy, ...]
Dataset " " Size 500.0 1635088521742 2021-10-25 [tag1x, tag2iy, ...]
Column dt_init_camp[Completeness] 1.0 1635088521742 2021-10-25 [tag1x, tag2iy, ...]
Column vl_rating_risc Minimum 1.0 1635088521742 2021-10-25 [tag1x, tag2iy, ...]
Column vl_rating_risc Maximum 9.0 1635088521742 2021-10-25 [tag1x, tag2iy, ...]
Column nr_cpf[Completeness] 1.0 1635087280875 2021-10-24 [tag1x, tag2iy, ...]
Dataset " " Size 500.0 1635087280875 2021-10-24 [tag1x, tag2iy, ...]
Column dt_init_camp[Completeness] 1.0 1635087280875 2021-10-24 [tag1x, tag2iy, ...]
Column vl_rating_risc Minimum 1.0 1635087280875 2021-10-24 [tag1x, tag2iy, ...]
Column vl_rating_risc Maximum 9.0 1635087280875 2021-10-24 [tag1x, tag2iy, ...]
Column nr_cpf[Completeness] 1.0 1635088521742 2021-10-24 [tag1x, tag2iy, ...]
Dataset " " Size 500.0 1635088521742 2021-10-24 [tag1x, tag2iy, ...]
Column dt_init_camp[Completeness] 1.0 1635088521742 2021-10-24 [tag1x, tag2iy, ...]
Column vl_rating_risc Minimum 1.0 1635088521742 2021-10-24 [tag1x, tag2iy, ...]
Column vl_rating_risc Maximum 9.0 1635088521742 2021-10-24 [tag1x, tag2iy, ...]
```

```
entity instance name value dataset_date VY_ML_00 tags
Column nr_cpf Completeness 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Dataset " Size 500.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column dt_init_camp Completeness 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column vl_rating_risc Minimum 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column vl_rating_risc Maximum 9.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
```

```
def fromJson.printSchema()
root
-- analyzerContext: struct (nullable = true)
-- element: struct (containsNull = true)
-- analyzerName: string (nullable = true)
-- metric: struct (nullable = true)
-- metric: struct (nullable = true)
-- entity: string (nullable = true)
-- instance: string (nullable = true)
-- metrics: string (nullable = true)
-- name: string (nullable = true)
-- value: double (nullable = true)
-- resultKey: struct (nullable = true)
-- datasetDate: long (nullable = true)
-- tags: struct (nullable = true)
-- tag: string (nullable = true)
-- tags: string (nullable = true)
-- tag: string (nullable = true)
```

```
analyzerContext resultKey
[[[Completeness,...]]]
[[[Completeness,...]]]
[[[Completeness,...]]]
```

## from Repository to JSON

Transform filtered Dataframe from persistentMetricsRepository into a jsonLikeDataframe. The jsonLikeDataframe can be easily written to HDFS as a repositoryJSON

```
dropColumns = ("entity", "instance", "name", "value")
addMetricChap(filteredRepository.df)
drop(dropColumns).show(truncate=false)

dataset_date tags metricschap
1635087280875 [tag1x, tag2iy, tag3iz] [[Minimum, vl_rating_risc, DoubleMetric, Minimum, 1.0]]
1635087280875 [tag1x, tag2iy, tag3iz] [[Size, ], Dataset, DoubleMetric, Size, 500.0]]
1635087280875 [tag1x, tag2iy, tag3iz] [[Completeness, nr_cpf], [Column, nr_cpf, DoubleMetric, Completeness, 1.0]]
1635087280875 [tag1x, tag2iy, tag3iz] [[Completeness, dt_init_camp], [Column, dt_init_camp, DoubleMetric, Completeness, 1.0]]
1635087280875 [tag1x, tag2iy, tag3iz] [[Maximum, vl_rating_risc], [Column, vl_rating_risc, DoubleMetric, Maximum, 9.0]]
```

## from PyDeequ output to persistentMetricsRepository

Get standard PyDeequ output into persistentMetricsRepository pattern. Both of them are Dataframes, but each one structured in its way

## inputData

```
nr_cpf dt_init_camp dt_init_camp dt_init_camp dt_init_camp dt_init_camp dt_init_camp dt_init_camp dt_init_camp dt_init_camp
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
1635087280875 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00 2021-10-25 00:00:00
```

## verificationSuite()

```
Metrics Repository: Success
entity instance name value
Column nr_cpf[Completeness] 1.0
Dataset " Size 500.0
Column dt_init_camp[Completeness] 1.0
Column vl_rating_risc Minimum 1.0
Column vl_rating_risc Maximum 9.0
```

```
Verification Run Status: Warning
checkLevelCheckStatus constraintStatus constraintMessage
WarningType Warning Warning(HistogramConstraint...) Success
WarningType Warning Warning(HistogramConstraint...) Failure Value: 9.0 does n...
WarningType Warning Warning(HistogramConstraint...) Success
WarningType Warning Warning(HistogramConstraint...) Success
WarningType Warning Warning(HistogramConstraint...) Success
WarningType Warning Warning(HistogramConstraint...) Success
```

```
entity instance name value dataset_date VY_ML_00 tags
Column nr_cpf Completeness 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Dataset " Size 500.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column dt_init_camp Completeness 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column vl_rating_risc Minimum 1.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
Column vl_rating_risc Maximum 9.0 1635087280875 2021-10-25 [tag1x, tag2iy, tag3iz]
```

```
nextKey.resultKey.toString()
ResultKey(1635087280875, Map(tag1 -> x2, tag2 -> y2, tag3 -> z2))
```