

Avatarization report for Example_report

Octopize MD

2025-12-23

Contents

Objective of this document	2
Global metrics	2
Applied anonymisation technique	2
Privacy metrics	2
aids	3
Signal metrics	4
aids	4
Definitions	8
Privacy metrics	8
Singling-out	8
Hidden rate (Target: > 90%)	8
Local cloaking (Target: > 5)	8
Distance to closest (Median target: > 70%, outlier percentile target: > 90%)	8
Closest distances ratio (Target: > 90)	9
Closest rate (Target: > 90%)	9
Column direct match protection (Target: > 50%)	10
Row direct match protection (Target: > 90%)	10
Knowledge protection rate (Target: > 95%)	10
Linkability	10
Linkability protection rate (Target: > 95%)	10
How it works	11
Interpretation	11
Requirements	11
Inference	11
Unified attribute inference (Target: > 70%)	11
Signal metrics	12
Hellinger mean (Target: < 0.1)	12
Correlation difference ratio (Target: < 10%)	12
Metadata	12
aids	12

Objective of this document

This automatically generated report allows a technical evaluation of both statistical quality and the residual reidentification risk associated with the related avatarized datasets.

Note: This report **SHOULD** be interpreted alongside the Privacy Impact Assessment (PIA) document. It **DOES NOT** constitute a privacy guarantee on its own. The anonymous status of the generated data must be validated by a Data Protection Officer (DPO) or an equivalent authority within your organization and should depend on both this technical evaluation and the specific context of data usage.

Global metrics

- Privacy global meta-metric: **87.46**
- Signal global meta-metric: **97.7**

Applied anonymisation technique

The original data has been anonymized using **Avatarization** technique. It transforms the original data into synthetic data (avatars) by generating new records that statistically resemble the original ones while ensuring individual privacy. Its core algorithm relies on four main steps publicly described:

1. individuals are projected in a multidimensional space
2. pairwise distances are computed to find the k nearest neighbors in a reduced space
3. synthetic coordinates are pseudo-randomly generated for each individual in the subspace defined by the neighbors
4. privacy metrics are evaluated and synthetic coordinates are inverse transformed back to the original space.

More information about Avatarization technique can be found in Octopize documentation.

Privacy metrics

The metrics used to evaluate privacy and described below measure the **three criteria** defined by the GDPR:

- **Singling out**, which is the risk to identify an individual in a dataset.
- **Linkability**, which is the ability to link individuals with another dataset.
- **Inference**, which is the possibility to deduce, with significant probability, the value of an individual using the anonymized dataset.

Each criteria is evaluated by reproducing **state-of-the-art attacking scenarios** like described in the example below:

- The attacker has both datasets: the original dataset and the anonymized one. Note: this is the worst-case scenario and is **highly improbable in practice**.
- The attacker uses a *distance-based attack* and tries to link each individual with its most resembling synthetic data.
- The individuals usually most vulnerable to this type of attack are those with remarkable characteristics, also called *outliers*.

For easier interpretation, we provide **target values** for each metric. These targets are deliberately conservative and tailored to the context of open data use (highest-risk scenario). Targets **SHOULD NOT** be considered as absolute thresholds but rather as **guidelines** to help interpret the results. The acceptability of the privacy metrics depends on the specific context of data usage and should be evaluated by a Data Protection Officer (DPO) or an equivalent authority within your organization.

aids

STANDALONE Privacy metrics computed on the table.

- **Meta metric: 87.46**
- local_cloaking: 7.0
- hidden_rate: 89.06
- distance_to_closest: 82.85
- distance_to_closest_outlier: 95.75
- closest_distances_ratio: 95.36
- closest_distances_ratio_outlier: 46.49
- closest_rate: 87.46
- column_direct_match_protection: 85.88
- row_direct_match_protection: 100.0
- knowledge_protection_rate: 86.0
- linkability_protection_rate: 99.77
- inference_accuracy: 18.62
- inference_accuracy_ratio: 108.5

STANDALONE Privacy metrics computed on the table.

- **Meta metric: 87.46**
- local_cloaking: 7.0
- hidden_rate: 89.06
- distance_to_closest: 82.85
- distance_to_closest_outlier: 95.75
- closest_distances_ratio: 95.36

- closest_distances_ratio_outlier: 46.49
- closest_rate: 87.46
- column_direct_match_protection: 85.88
- row_direct_match_protection: 100.0
- knowledge_protection_rate: 86.0
- linkability_protection_rate: 99.77
- inference_accuracy: 18.62
- inference_accuracy_ratio: 108.5

Signal metrics

aids

STANDALONE : Signal metrics computed on the table

- **Meta metric: 97.7**
- hellinger_mean: 0.1
- hellinger_std: 0.1
- hellinger_dict:
 - symptom: 0.06
 - age: 0.2
 - strat: 0.02
 - days: 0.27
 - oprior: 0.0
 - drugs: 0.06
 - str2: 0.01
 - cd80: 0.26
 - z30: 0.01
 - offtrt: 0.04
 - homo: 0.06
 - preanti: 0.19
 - cens: 0.05
 - treat: 0.0
 - cd420: 0.24
 - cd40: 0.27

- hemo: 0.03
- gender: 0.04
- wtkg: 0.24
- r: 0.04
- arms: 0.01
- cd820: 0.26
- karnof: 0.05
- race: 0.04
- correlation_difference_ratio: 2.35
- contribution_score:
 - age: 3.01
 - wtkg: 3.17
 - hemo: 3.06
 - homo: 2.95
 - drugs: 2.71
 - karnof: 3.03
 - oprior: 3.06
 - z30: 2.91
 - preanti: 3.07
 - race: 3.12
 - gender: 3.38
 - str2: 10.28
 - strat: 3.42
 - symptom: 3.15
 - treat: 2.85
 - offtrt: 2.81
 - cd40: 3.38
 - cd420: 6.67
 - r: 3.39
 - cd80: 3.44
 - cd820: 2.53
 - cens: 2.62
 - days: 3.04

- arms: 10.3
- multivariate_hellinger_distance_mean: 0.02
- multivariate_hellinger_distance_std: 0.01
- multivariate_hellinger_distance_dict:
 - Dim. 1: 0.12
 - Dim. 2: 0.1
 - Dim. 3: 0.11
 - Dim. 4: 0.09
 - Dim. 5: 0.14

STANDALONE : Signal metrics computed on the table

- **Meta metric: 97.7**
- hellinger_mean: 0.1
- hellinger_std: 0.1
- hellinger_dict:
 - symptom: 0.06
 - age: 0.2
 - strat: 0.02
 - days: 0.27
 - oprior: 0.0
 - drugs: 0.06
 - str2: 0.01
 - cd80: 0.26
 - z30: 0.01
 - offtrt: 0.04
 - homo: 0.06
 - preanti: 0.19
 - cens: 0.05
 - treat: 0.0
 - cd420: 0.24
 - cd40: 0.27
 - hemo: 0.03
 - gender: 0.04

- wtkg: 0.24
- r: 0.04
- arms: 0.01
- cd820: 0.26
- karnof: 0.05
- race: 0.04
- correlation_difference_ratio: 2.35
- contribution_score:
 - age: 3.01
 - wtkg: 3.17
 - hemo: 3.06
 - homo: 2.95
 - drugs: 2.71
 - karnof: 3.03
 - oprior: 3.06
 - z30: 2.91
 - preanti: 3.07
 - race: 3.12
 - gender: 3.38
 - str2: 10.28
 - strat: 3.42
 - symptom: 3.15
 - treat: 2.85
 - offtrt: 2.81
 - cd40: 3.38
 - cd420: 6.67
 - r: 3.39
 - cd80: 3.44
 - cd820: 2.53
 - cens: 2.62
 - days: 3.04
 - arms: 10.3
- multivariate_hellinger_distance_mean: 0.02

- multivariate_hellinger_distance_std: 0.01
- multivariate_hellinger_distance_dict:
 - Dim. 1: 0.12
 - Dim. 2: 0.1
 - Dim. 3: 0.11
 - Dim. 4: 0.09
 - Dim. 5: 0.14

Definitions

Privacy metrics

Singling-out

Hidden rate (Target: > 90%)

The **hidden rate** refers to the singling out criteria. This metric indicates the median percentage of individuals for whom the generated avatar is not the closest-matching one.

Note: a hidden rate below 90% **does not mean** that 10% of the individuals are not protected.

For more in-depth information about the Hidden Rate metric, please visit the Octopize documentation.

Local cloaking (Target: > 5)

The **local cloaking** refers to the singling out criteria. This metric indicates the median number of avatars that are closer to an individual than the avatar generated for that individual.

Note: the higher the local cloaking, the better.

For more in-depth information about the Hidden Local cloaking, please visit the Octopize documentation.

Distance to closest (Median target: > 70%, outlier percentile target: > 90%)

Distance to closest compares the distribution of distances between one half of the original dataset and its avatarized version against the distribution of distances between the two halves of the original dataset. For interpretability, it uses the natural variability of the data as a reference.

For both comparisons, it retrieves: - **Median_ratio**: Percentage change for original sample between avatar median distance to closest and reference median distance to closest - **Outlier_ratio**: Percentage change for original sample between avatar

97.5th percentile distance to closest and reference 97.5th percentile distance to closest

Ratios are then calculated as: $\text{avatar_distance} / \text{original_distance} \times 100$

Interpretation: The result estimates the percentage of the reference distance obtained through avatarization. - the closer to 100, the more avatars match the natural variability of the dataset. - Values < 100 indicate that original data are closer to their avatars than to other original data (less private) - Values > 100 indicate that original data are farther from their avatars than from other original data (more private)

Because this metric does not use the links between original and synthetic records, it can be used to evaluate all kinds of synthetic datasets.

For more in-depth information about this metric, please visit the Octopize documentation.

Closest distances ratio (Target: > 90)

A closest distances ratio is defined as the distance between a record and its closest neighbor divided by the distance to its second closest neighbor. It allows to assess the natural density of the data around each record. This metric compares the distribution of closest distances ratios between one half of the original dataset and its avatarized version against the distribution of closest distances ratios between the two halves of the original dataset. For interpretability, it uses the natural variability of the data as a reference.

For both comparisons, it retrieves: - **Median Ratio:** Percentage change for original sample between avatar median ratio and reference median ratio. - **2.5th Percentile Ratio:** Percentage change for original sample between avatar 2.5th percentile ratio and reference 2.5th percentile ratio.

Ratios are then calculated as: $(\text{avatar_ratio} - \text{original_ratio}) / \text{original_ratio} \times 100$

Interpretation: The result estimates the percentage of the reference ratio obtained through avatarization. - the closer to 100, the more avatars match the natural density of the dataset. - Values < 100 indicate that original data are more isolated with their avatars than with other original data (less private) - Values > 100 indicate that original data are less isolated with their avatars than with other original data (more private)

Because this metric does not use the links between original and synthetic records, it can be used to evaluate all kinds of synthetic datasets.

For more in-depth information about this metric, please visit the Octopize documentation.

Closest rate (Target: > 90%)

The Closest Rate metric combines the results of the Distance to Closest and Closest Distances Ratio metrics to determine the percentage of individuals whose most

similar avatar is not too close, and whose original data/avatar pair is not easily distinguishable from the rest of the dataset.

For more in-depth information about this metric, please visit the Octopize documentation.

Column direct match protection (Target: > 50%)

The **column direct match protection** measures how unlikely each column in the data is to be used as a direct identifier. This is to account for the fact that variables with low cardinality will be more likely to be matched than variables with high cardinality, but that this match does not represent a risk of re-identification. The level of protection is calculated for each column, with only the minimum value reported here, as it represents the variable most vulnerable to direct identification.

Note: A **column direct match protection** of 0 means that a direct identifier has been kept in the data and that the data is **not** anonymized. A **column direct match protection** below the target does not necessarily mean that the data is at risk but it is recommended to check the quality of the avatarization with respect to the variable(s) showing the smallest direct match protection value(s).

For more in-depth information about this metric, please visit the Octopize documentation.

Row direct match protection (Target: > 90%)

The row direct match protection measures the percentage of original unique individuals that are not replicated in the anonymous dataset. The higher, the better.

For more in-depth information about this metric, please visit the Octopize documentation.

Knowledge protection rate (Target: > 95%)

The **knowledge protection rate** refers to the linkability criteria. It is the probability that an attacker makes a mistake when linking an individual with its most similar synthetic individual based on a subset of **known variables**.

Note: a knowledge protection rate below 95% **does not mean** that 5% of the individuals are not protected.

For more in-depth information about the knowledge protection rate metric, please visit the Octopize documentation.

Linkability

Linkability protection rate (Target: > 95%)

The **linkability protection rate** is a comprehensive metric that measures how well avatarized data protects against linkability attacks where attackers try to link records across different data splits using known variables.

How it works

The metric works by:

1. **Splitting known variables:** The set of known variables is randomly divided into two approximately equal groups
2. **Creating data splits:** Both original records and avatars are split using these variable groups, resulting in 4 datasets
3. **Distance-based matching:** For each variable group, the metric finds the closest avatar to each record using Gower distance
4. **Cross-validation:** It counts how many records point to the same avatar across both variable splits
5. **Protection calculation:** The final score is $(1 - \text{linkability_rate}) \times 100$, where `linkability_rate` is the proportion of records consistently pointing to the same avatar

Interpretation

- **Higher values (closer to 100%)** indicate better protection against linkability attacks
- **Lower values** suggest that attackers could more easily link individuals across different data views

Requirements

- At least 2 known variables are required for meaningful analysis
- Uses Gower distance to handle mixed data types (numeric and categorical)
- Results may vary based on random seed due to variable splitting

For more information about linkability protection and privacy metrics, please visit the Octopize documentation.

Inference

Unified attribute inference (Target: > 70%)

The **unified attribute inference** metric evaluates the risk of an attacker successfully predicting sensitive **target** values using avatar data based on known **variables**. This metric implements a holdout strategy where avatars attempt to predict target values, and their performance is compared against a reference baseline.

For both predictions, it retrieves: - **Accuracy Ratio:** Percentage representing reference accuracy relative to avatar accuracy - **Avatar Accuracy:** Percentage of correct predictions using avatar data (0-100%)

The ratio is calculated as: $(\text{reference_accuracy} / \text{avatar_accuracy}) \times 100$

Interpretation: The result estimates if avatars improve inference precision. - **Accuracy Ratio:** Values around 100 indicate better privacy (reference performs relatively similarly compared to avatars) if the ratio is significantly below 100, you can

refer to the Avatar Accuracy for further insights: - **Avatar Accuracy**: Higher values indicate highly predictable targets (more sensitive). If this value is close to 100 check the accuracy ratio for privacy assessment.

Avatars are consider at risk if they have both a low accuracy ratio and a high avatar accuracy.

Key Features: - **Unified approach**: Works consistently for both continuous and categorical targets - **Tolerance handling**: For continuous targets, predictions within 5% of true values are considered correct

For more in-depth information about attribute inference metrics, please visit the Octopize documentation.

Signal metrics

Hellinger mean (Target: < 0.1)

Distance between the original and the avatar distribution is mathematically computed using Hellinger distance. The metric returns the mean of Hellinger distance computed for each variable.

Note: We consider that a Hellinger distance lower than represents a very good signal retention while distances between and represent tolerable signal retention.

For more in-depth information about this metric, please visit the Octopize documentation

Correlation difference ratio (Target: < 10%)

The **Correlation Difference Ratio** measures the relative percentage change in correlation for each pair of numerical variables. It quantifies how much the correlation structure of the transformed dataset deviates from the original dataset.

Note: We consider that a correlation difference ratio lower than 10% indicates good correlation preservation. While correlation difference ratio between 10% and 30% indicates that some correlation structures are distorted.

For more in-depth information about this metric, please visit the Octopize documentation

Metadata

- job creation: 2025-12-23 13:41:50.873126+00:00
- Versions : {'api': '2.39.0', 'web': '0.48.0'}

aids

- original:
 - *hash*: a8439d866dfa9415a824bb9047efe346d430cdf1834d59ad3d08f78d76cdc8

- *number of lines* : 2139
- *number of columns* : 24
- *number of dimensions* : 45
- *number of batches* : 1
- avatars:
 - *hash* : b040f74f7690cd9fa1e48e84953ff3172038d03fa931a906e293d141aeca9761
 - *number of lines* : 2139
 - *number of columns* : 24
 - *number of dimensions* : 45
 - *number of batches* : 1
- avatarization parameters :
 - *k*: '20
 - *ncp*: 30
- privacy metrics parameters :
 - *ncp*: 30
 - *target*: days
 - *known variables*:
 - * age
 - * wtkg
 - * homo
 - * race
 - * gender
- signal metrics parameters :
 - *ncp*: 30