Dataset: fitness_app_data.csv
Date: 30th October, 2025
Tool: ARX Privacy Framework

1. Input Specifications

Attributes & Transformations

Before anonymizing, I reviewed every column and grouped them based on the type of risk they create.

Identifying attribute:

- Name is a direct identifier, so it was removed.

Quasi-identifiers

- Age, Location, and Gender were treated as quasi-identifiers because, when combined, they can be used to single people out.

To reduce that risk, the following transformations were applied:

- Age was generalized into wider ranges (hierarchy level 3)

- Location was transformed using multi-layer character masking. Each level hides more characters in the "<city>, <country>" text, which makes rare locations less identifiable while still retaining some regional features.

- Gender used ordering (3 values → 1) when needed. In small groups, all three values collapsed into a single generalized category to prevent inference.

Sensitive attributes

Health-related and behavioral attributes like Weight, Height, BMI, Workout Frequency, Workout Duration, Workout Type, Interaction, Content Preference, and Allergies can reveal personal details.

These were protected using distinct ℓ-diversity, mostly set to 3. Allergies used $\ell = 5$ because medical information is more sensitive.

Insensitive attributes:

Values like Progress, Fitness Goals, Features Used, Dietary Preference, and App Time do not meaningfully increase re-identification risk, so they were not transformed.

Configuration

Based on the certificate:

- Suppression limit = 0.0. Meaning no rows were deleted; everything stays in the final dataset.

2. Output Properties

Output Data

After anonymization, the certificate shows:

- All 500 rows remain.
- All 18 attributes are still present.
- No suppression was needed.

This means the dataset is still complete, just less detailed.

Solutions

ARX explored part of the search space (536 possible transformations, 19 materialized) and found a valid solution that satisfied all privacy models. This means the anonymization was efficient and feasible.

Transformations

The final output shows:

- Age is now displayed in ranges instead of exact values.
- Location has characters replaced with "*" depending on the privacy level required for that record. It required maximum generalization.
- Gender remained at level 0, meaning the gender did not change

These transformations remove uniqueness but retain the dataset's usefulness for pattern analysis.

Privacy Models

The certificate confirms:

- **k-Anonymity (k = 10)** was satisfied. Every record now looks like at least nine others.

- **ℓ-Diversity** was applied to sensitive attributes to prevent attribute disclosure inside an anonymized group.

Combining both models protects against identity disclosure and inference attacks.

Conclusion

Using age ranges, location masking, and attribute diversity, the final dataset prevents identity disclosure and reduces sensitive inference. All rows remain, privacy risks stay below the thresholds, and the data is still usable for analysis.

The ARX certificate confirms that the anonymization meets the required privacy conditions.