Security, Privacy and Explainability in Machine Learning

Exercise 1

Ana Terović Fani Sentinella-Jerbić

April 2023

1 Dataset

We chose the first dataset, Adult Census from UCI Machine Learning repository. It consists of sensitive information from 1994 about people's income level, age, education level, marital status, occupation, race, and gender. After the dataset preparation we were left with 30,161 samples and 13 attributes.

2 k-Anonymization

The dataset was anonymized with hierarchies as described in the task. The anonymization parameters were:

- k = 2,
- max. supression = 0%,
- measure: loss,
- aggregate function: geometric mean,
- attribute weights = 0.5.

The generalisation lattice of k-anonymous datasets is shown in Figure 1. The optimal solution is the yellow one marked with a rectangle. Set F is going to be composed of the two datasets made from transformations of the two solutions linked to the optimal one.

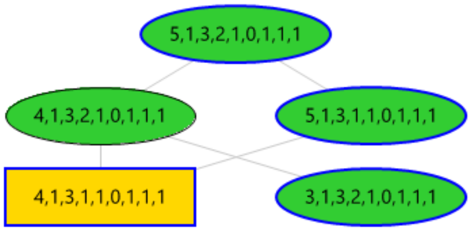


Figure 1: Generalisation lattice of k-anonymous datasets

3 Record-linkage

Given our anonymization resulted in set F composed of two datasets, we will perform two tries of record linkage:

a. on two datasets from F

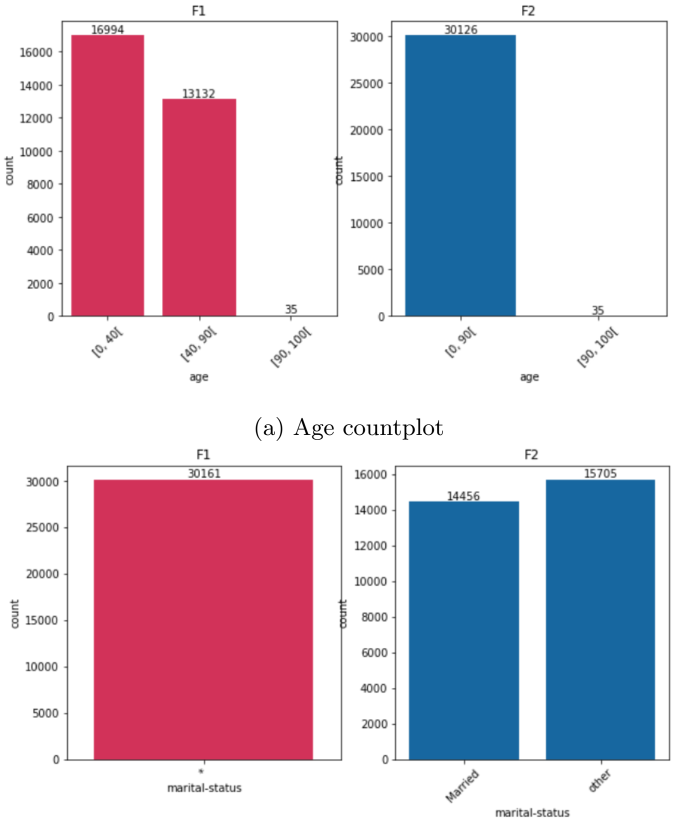
b. on a dataset from F and an arbitrarily anonymized dataset

We are expecting harder linkage of records in the case of **b.** because we will try an even stricter setting.

Our tool of choice is the **RecordLinkage** Python library with respective functionalities of creating possible pairs of records, comparing, and matching them.

a. Record linkage on datasets from F

First we performed exploratory data analysis on the two resulting datasets from set F. As expected, non-QI's contain exact matches of values so when plotted in Figure 3, they show same distributions. For QI's, both datasets have been completely anonymized for attributes sex, workclass, education, occupation, and race. The QI's which have resulted in different anonymizations are marital status and age, as can be seen in the Figure 2.



(a) Age countplot

(b) Marital status countplot

Figure 2: Countplots for differing QI's

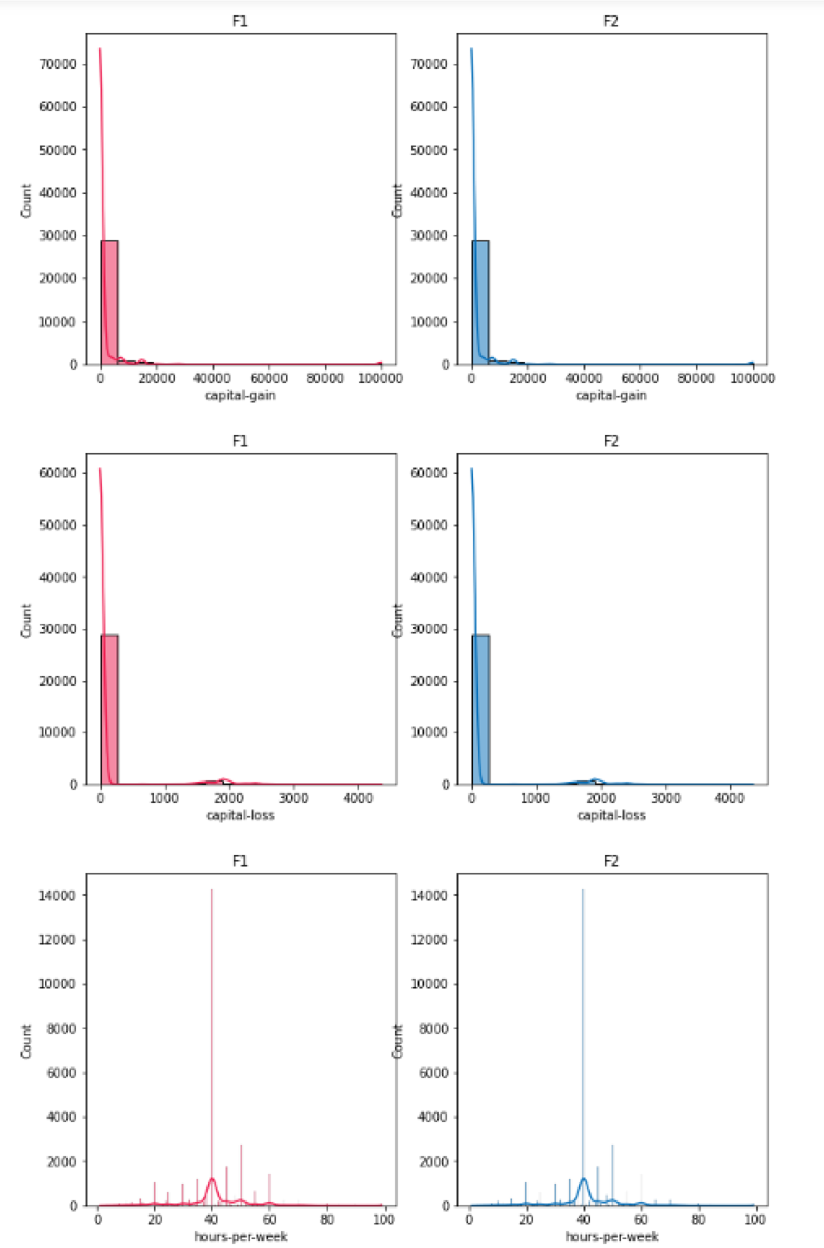


Figure 3: Sanity check for non-QI attributes

Tackling age. For the dataset containing smaller age intervals, we transform it to match the dataset containing bigger age intervals. More specifically, we transform [0, 40] and [40, 90] to [0,90].

Tackling marital status. One dataset contains completely removed marital status information, while the other contains values of 'Married' and 'Other'. However, both datasets contain completely non-anonymized information about relationship from which marital status can be derived. But we will not perform this derivation as we don't need both attributes, the relationship one carries more information for record linkage either way.

After tackling differing QI's, we construct pairs of records to compare. The datasets are large and for checking each pair too much memory would be used, so we opt for blocking a few attributes. Blocking simply means we are considering only some pairings based on values of an attribute. For example if we block for relationship, we are only considering pairs of individuals which have the same status of the relationship. This greatly reduces the number of comparisons required.

Lastly, we define strategies for comparing the attributes and try to match the records. However, for our case of datasets, the anonymization is so strong that comparing doesn't make sense. After blocking, we are left with possible pairs of records, but within most of the groups one guess in matching is just as probable as the other because we are dealing with categorical data and attributes either match or they don't. But we do have some cases where there is only one pairing made and these cases we can consider successful linkage! We have found **1064** of such records, meaning we have managed to find exact matches for 3,53% of all records.

b. Record linkage on dataset from F and an arbitrarily anonymized dataset

We created an arbitrarily anonymized dataset which contains hours per week, capital loss and gain, sex and income group. We want to see how we would perform record linkage in an even harder case, when we don't even have the same relationship attribute, but a somewhat related sex attribute. Our goal is to reveal people's the income levels with this attack.

We try to use information from relationship attribute of the F dataset to match it with the sex attribute in the arbitrary dataset. We map 'Husband' to 'Male' and 'Wife' to 'Female'. With this even more restricted setting we still managed to link a total of **523** records and reveal income levels, which is something a lot of people would consider fairly private information!

4 Conclusions

Based on our results, we conclude non-QI data can be very problematic in data anonymization. Because the values can be very specific, it allows for easy one-to-one mapping, especially for outliers. Another big issue is general or background knowledge. With our second experiment we saw how simple inference of one catgorical variable to another can also allow for easier record linkage.