

Exercise 5

5.1 k -Anonymity and ℓ -Diversity (10pt)

Consider the data set given in `ex05-fake-registrations.csv`. For the following questions, work on an arbitrary subset of about 20 entries from the data set.

- a) Define identifiers, two quasi-identifying attributes, and one sensitive attribute.
- b) Sanitize this data by computing a 3-anonymized version of this data by hand.
- c) Does it become easier or more difficult to derive a 5-anonymized data set?
- d) Sanitize the entries to satisfy 3-diversity.

Note that contrary to the examples in class, not all partitions need to have the same number of entries.