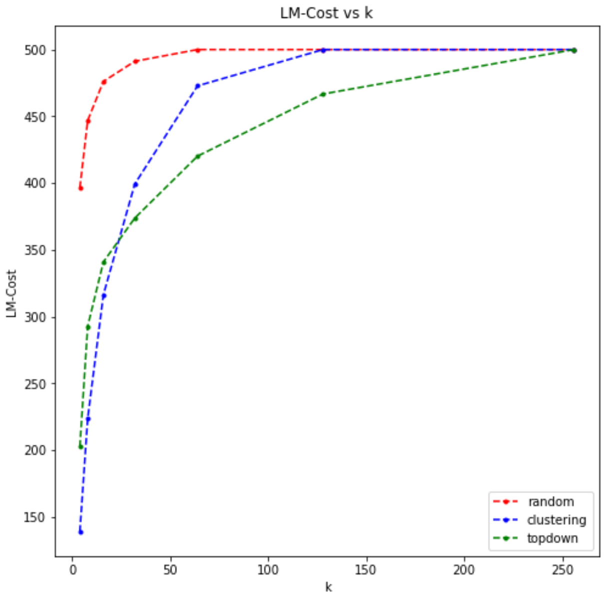
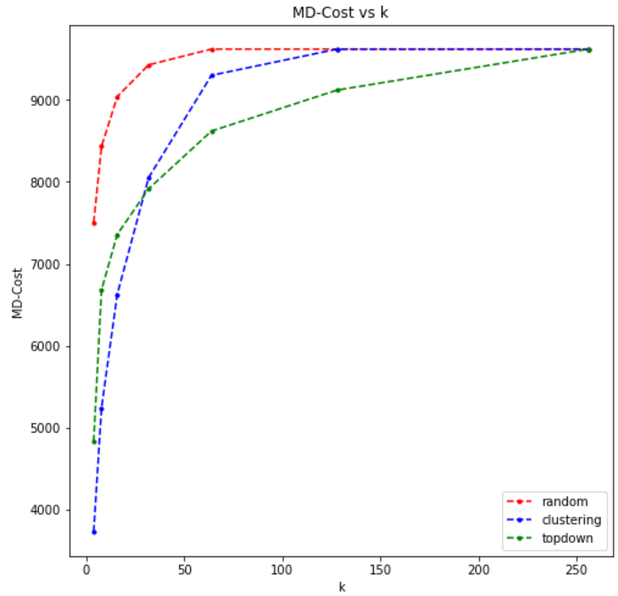
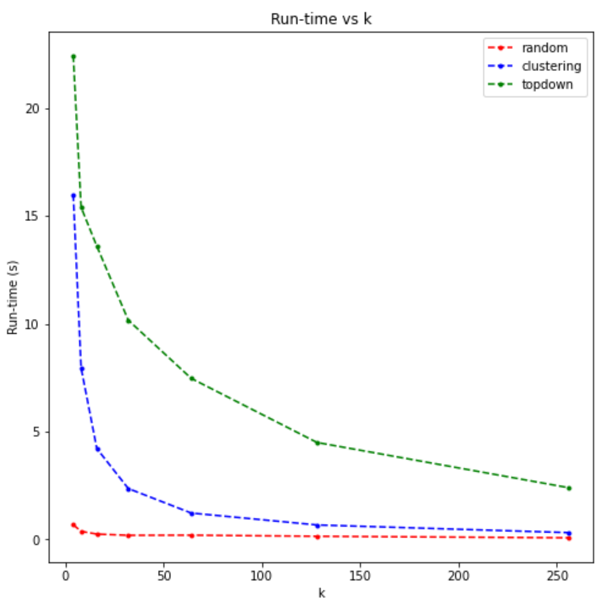
Alp Kural - 71959
Computer Engineering

Comp 430 - Homework 01 Project Report

Graphs are constructed from runtime results of varying anonymizers with differing values of k listed as “k = 4, 8, 16, 32, 64, 128, 256” and the respective dataset used to be “mini-adult1.csv” from the provided test cases.







Trade-offs:

There exists a trade-off between runtime and distortion and loss metrics.

Higher runtime algorithms result less utility loss.

Random anonymizer is fastest but incurs higher distortion and loss metrics.

Clustering and topdown anonymizers offer better utility but have longer runtimes.

Besides, increasing k values reduce runtime, however, lead to higher distortion and loss metrics.

Observations:

Balancing utility and runtime are crucial; choice depends on specific use case requirements.

Preferences Based on Use Case:

Random anonymizer is better in terms of speed and could be used in the cases where less significance is assigned to utility loss.

Clustering anonymizer is good for a balance between utility loss and runtime duration.

Topdown anonymizer can be used for maximizing utility with higher computational resources.

Ps. Runtime results are fitted my initial expectations from the outputs of the implemented anonymizers.