Final Project Guidelines

General outline

The final project in the course are meant to give you the opportunity to further explore an aspect of differential privacy and gives you the experience of formulating, carrying out, and presenting an interesting, short-term independent project that is similar to an experience you might have in a career as an applied data scientist confronting privacy issues.

1. Implement and experimentally evaluate differentially private algorithms or attacks on real-life datasets. For example, identify a dataset that resembles a sensitive data use case and a type of statistical analysis that would be useful on such a dataset, implement and tune a differentially private algorithm for that analysis.

2. Critically evaluate an existing system for privacy protection, identify potential vulnerabilities and propose or demonstrate improvements using techniques.

3. Explore how the "noisy" results from differential privacy can be properly.

Specific Problem:

Data Set:
https://github.com/microsoft/USBroadbandUsagePercentages/tree/master/dataset

The Broadband Usage Percentages Data Set is derived from aggregated and anonymized data. Given the zip code-level data set provides a more granular view of broadband usage percentages by households, we took the additional step to ensure data privacy guarantees by utilizing differential privacy.

Differential privacy is a technique that adds noise to the data aggregations, preventing leakage about the presence of specific individuals in the data set.

1. Implement differential privacy through the SmartNoise platform, a first-of-its-kind opensource platform for differential privacy co-developed by OpenDP initiative.
2. Estimate broadband usage by combining privatized data from multiple services.

3. The data privatization is done using differential privacy mechanisms on top of count queries, with a privacy parameter epsilon = 0.1 and 0.2.

4. Empirically calculate the error range caused by differential privacy

Submission guidelines

- Please submit the python notebook and a one page write up of your analysis