Safeguard Customer Privacy Without Sacrificing Analytical Capability with **Diffprivlib**

University of Minnesota
Driven to Discover℠

Team Number 4

Abstract:

While data offer tremendous business opportunity, they also confer great risk to firms and their customers, the data points. Differential privacy, which strives to eliminate the risk of privacy breaches [1], is an evolving and critical asset that allows data scientists to uphold a wide range of analytical capabilities. The Python package, Diffprivlib, has the capability to preserve simple descriptive statistics and complicated machine learning algorithms but not risk the privacy act violation. Companies like Apple, Google, and Uber have adopted this technology, and the 2020 census data will also use differentially privatized data. New tools like Diffprivlib allow for simple and business friendly adoption.

Storing data comes with financial, reputation, and regulation risk

$8.64 Million

Average cost of a data breach in US [2]

280

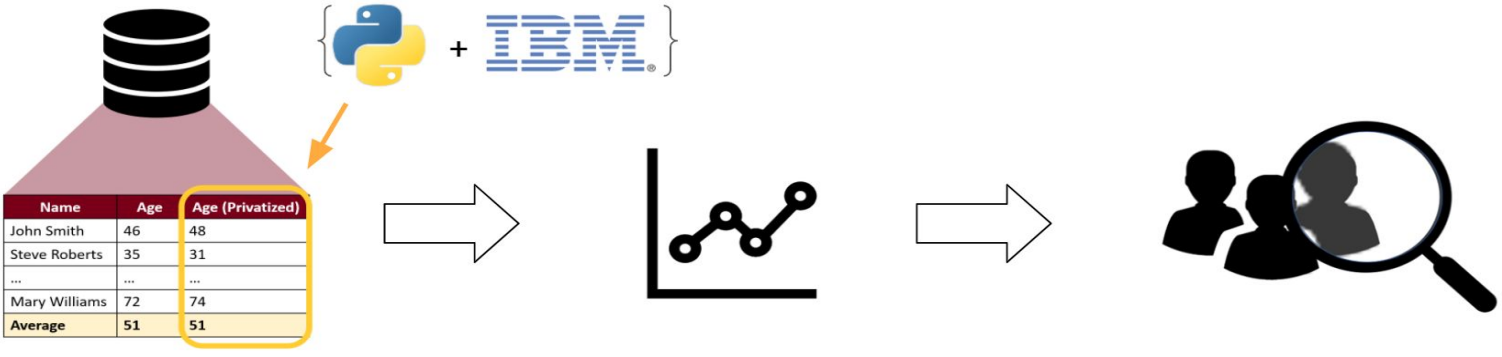
Average time (in days) to identify and address a breach [2]

$$$?

Company reputation is not quantifiable

29

States enacted, enforced, or considered privacy legislation in 2019 [3]



Diffprivlib **adds noise** to confidential data while **maintaining statistical integrity**

Users can conduct **accurate analyses** and build **valid models** with privatized data

Gain population or customer **insights** while **safeguarding privacy**

Diffprivlib

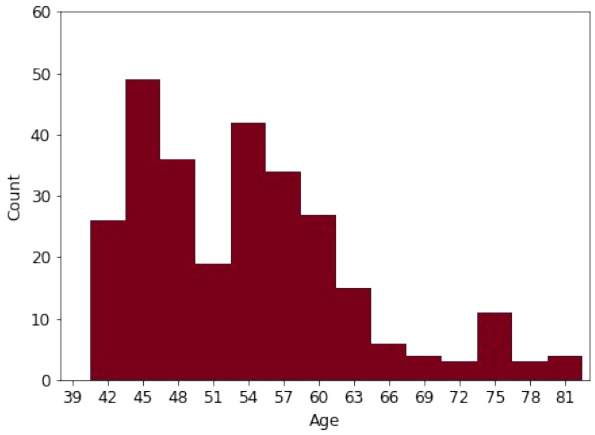
Team Number **4**

Encryption is not sufficient

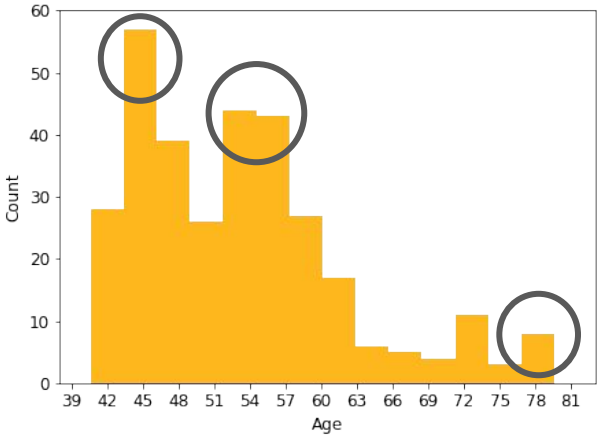
- Encrypted business data can be combined with other public data to extract user information
- Little information is needed to obtain an individual's identity
  - Almost 90 percent of the US population has a unique combination of 5-digit zip code, gender, and date of birth [3]

Diffprivlib adds noise to your private data and retains the business value

Age Distribution **(Original Dataset)**



Age Distribution **(Masked Dataset)**



Ample and Diverse Business Value

Reduce risk of sharing data externally

kaggle NUMERAI

Crowd-source analytics solutions

Expand pool of potential **partners**

Avoid **legal repercussions**

Laura Cattaneo (catta008@umn.edu)

Harshitha Kuriminisetty (kurim006@umn.edu)

Casey Easterday (easte060@umn.edu)

Tzu-Hsuan Lin (lin00491@umn.edu)

Check out our GitHub