The Effectiveness of Differential Privacy in Real-world Settings: A Metrics-based Framework to Help Practitioners Visualise and Evaluate ε

Akasha Shafiq[a], Abhishek Kesarwani[b], Dimitrios Vasilopoulos[c], Paolo Palmieri[a]

[a]*Department of Computer Science, University College Cork, Cork, Ireland*
[b]*Indian Institute of Information Technology, Vadodara, India*
[c]*SnT, University of Luxembourg, Esch-sur-Alzette, Luxembourg*

Abstract

Differential privacy (DP) has emerged as a preferred solution for private data analysis, having been adopted by several leading Internet companies. DP is a privacy-preserving mechanism that protects against re-identification of individuals. *Privacy budget* ε determines the trade-off between privacy and utility. Currently, there are no widely accepted methodologies for selecting ε and choosing the best DP mechanism that offers an optimal trade-off between privacy and utility. We propose using a novel set of metrics and an easy-to-implement, step-by-step framework to facilitate DP deployment and guide the selection of ε based on the desired accuracy vs utility trade-off. To address this gap, we perform experiments using three real-world datasets, aiming to identify the optimal parameters . The analysis focuses on the commonly used queries for each dataset. Based on our experiment results, we observe that using our metric and framework, one can analyse noise distribution of multiple queries, and choose the suitable ε and the DP mechanism for achieving a balance between privacy and utility. Additionally, we show that the optimal ε depends on the particular query, desired accuracy and context in which DP is implemented, which suggests that an arbitrary a prior selection of ε cannot provide adequate results. Our framework prioritises the visualisation of key metrics in the DP analysis as a function of ε and mechanism. Enabling the adoption easy for a wider audience without deep knowledge, providing a road map for DP implementation.

Keywords: Differential Privacy, Privacy Budget, Anonymisation

1. Introduction

As the volume of personal data collected by organisations increases, driven by online interactions and smart devices, so does the challenge of preserving privacy. The benefits of analysing datasets containing personal information are substantial for many organisations, who rely on analysing and releasing statistical information about the data in order to target and personalise products and advertising. The goal of privacy-preserving data analysis is to facilitate secure statistical analysis of databases without compromising the privacy of individual users whose data are included. Specifically, it is essential to mitigate the risk that querying the database could expose information about any individual. The importance of this is underlined by regulations such as the *General Data Protection Regulation* (GDPR) in Europe.

Differential Privacy is a privacy-preserving mechanism that protects against the re-identification of individuals within aggregated datasets. Dwork (14) describe it as: "You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources are available". A DP *mechanism* $\mathcal{M}$ enables the release of statistical information about a database in the form of queries, which help to extract information from a database. Primarily, it involves introducing a calibrated amount of noise to the true query response in a way that it is meaningful and truthful while not revealing individuals' information. More noise refers to more privacy and less usefulness of data, also referred to as utility. This balance is known as the utility versus privacy trade-off, that is governed by two parameters: the privacy budget ε, which controls the amount of noise added to the data, and δ, which is the probability of information leakage.

Problem Overview. In recent years, DP has become a widely accepted solution for privacy-preserving analytics and has been used by major internet companies when publishing or analysing data, including Google (7), Facebook (37), LinkedIn (45) and (4). It has also been deployed by government agencies, such as the U.S. Census Bureau, which uses it to publish statistics related to census (1; 27). One of the major challenges when deploying DP is the lack of consensus on the choice of ε and other parameters, as also highlighted by the NIST report (21), including privacy leakage and the choice of noise mechanism. Moreover, many organisations refrain from disclosing pa-

rameter selections or provide vague and unsatisfactory justification for their choices. This is highlighted by (15) and leads to a communication gap in conveying DP guarantees to engineers and stakeholders. Furthermore, the work in (11) and (49) highlights the significance of visualisation methodologies, revealing that users prioritise privacy and are inclined to adopt DP when aware of its protective benefits. Ngon et al. (43) evaluated DP tools with practitioners, highlighting that they should have easy navigation, improved APIs, and resources to ease adoption.

A number of open source libraries are available (such as PINQ (25), pydp (23), Opacus (50), and opendp (38)), enabling DP to be easily implemented in software (18; 24). For example, the 'diffprivlib' library was developed to facilitate real-world deployment of DP and perform analysis of datasets (28). However, (13) highlighted several key weaknesses, including unclear and inconsistent privacy notions, floating-point vulnerability, and mis-handled edge cases in diffprivlib, which could lead to misleading experiment results for real-world use cases. Moreover, the authors in (13) and (3) critique DP libraries because they often work as 'black boxes', making it difficult for non-expert users to understand how these libraries choose the privacy parameters (ε and δ) which are crucial for the privacy vs utility trade-off. This lack of clarity is problematic when differential privacy is applied in real-world applications. If users do not fully understand the settings, they may mistakenly believe their data is more private than it actually is. As a result, there is a need for a better understanding of the underlying DP mechanisms to ensure greater transparency in DP applications. Users select a mechanism and ε without clear intuition of how these choices align with query sensitivity, dataset scale, and data modality affect the released statistics.

DP research has focused mainly on theoretical privacy guarantees (10; 15), often overlooking the need for comprehensive practical metrics that quantify the privacy achieved and utility retained in real-world implementations (12). Following Nist's guideline (8), no single utility metrics suffices for DP relying on just absolute error or other measures can be misleading across use cases. Our framework, therefore, proposes a portfolio of metrics with query, sensitivity and use case-aware indices to make ε selection evidence-based for the task at hand.

For instance, the same choice of parameters ε, δ, and DP mechanism $\mathcal{M}$ will have different effects on different datasets and as per the use case. Previous efforts to define metrics for DP were focused on privacy requirements such as generalising DP for an arbitrary notion of distance (9) or limited to

specific data types (35). The challenges that emerged with the deployment of DP are mentioned in (10), and most of these challenges stem from real-world use cases (12). We focused on the ones related to the adaptation of central differential privacy.

A recent initiative, 'A Community-Driven Differential Privacy Deployment Registry' by NIST (21), aims to capture real-world deployments and their parameter choices. The NIST report highlighted two practical constraints: (i)convincing stakeholders that DP is feasible and will retain sufficient utility, and (ii)answering open design questions such as how to quantify utility and select parameters that provide adequate privacy and complementing theory with empirical evaluation. The registry enables contributors to share settings and provides practitioners with inspiration, but does not guarantee that the same settings will work for others, depending on their dataset characteristics/workload. Thus, our framework complements this ecosystem.

Therefore, we need standard metrics that are crucial in the practical implementation of DP to analyse the implementation before deciding on parameters.

We outline some of the key challenges that we aim to address in this work.

- There is a need for guidance in selecting DP mechanisms and privacy parameters that ensure a reasonable trade-off between privacy and utility according to the use cases.

- Data holders and practitioners often lack standardised metrics to quantify the errors of DP mechanisms. They need tools that provide intuitive visualisations of these errors.

- Currently, there is no consensus among experts on which error metrics to optimise in algorithm design. How can we create effective tools and visuals to make error analysis accessible and actionable for non-experts.

- The need for end-to-end DP frameworks with intuitive interfaces and seamless parameters$(\varepsilon, \delta, \mathcal{M})$ tuning need to be developed to enable broader deployment of DP in practice.

- Impact of mechanism design on utility vs. privacy trade-off. How can we provide user-friendly methods for selecting ε and explaining the reliability of DP and fairness of output in a real-world setting.

1.1. Real-World Applicability and Use Case Scenario

The proposed framework is designed to serve a practical value. An end-to-end framework that computes sensitivity, sweeps ε across mechanisms and visualises privacy vs utility metrics for a range of stakeholders, such as government agencies, non-profit organisations, researchers, companies and private enterprises who seek to analyse the datasets and publish the statistical insights without compromising individuals' privacy. The work supports the implementation of DP by employing simulation and empirical analysis, thereby enhancing the comprehensibility of DP parameters and simplifying their adjustments to align with the practical applications. Consider, an organisation that has individuals' data and they aim to publish demographic trends related to income, education, gender, health or employment. Using the datasets, the organisation may want to release statistical queries such as count, sum, and average queries. A few examples of such queries are below.

- Number of individuals in the study
- The total capital gain or wealth across different continents
- The average working hours across different demographics
- The percentage of people in a particular income bracket or in a particular employment category

Parameter Selection. Traditionally, implementing DP in real use cases is challenging due to manually setting the sensitivity section 2.4, selecting a suitable noise mechanism, and finding and justifying the value for ε. Our framework simplifies the process of implementing DP by providing empirical evaluation and visualisation to help make informed decisions about these parameters.The proposed framework evaluates the trade-offs using multiple metrics, which are well-known metrics in statistics, and we use these metrics to evaluate the differential private mechanisms and to guide for the selection of ε. The used metrics, such as threshold bounds, entropy relative errors and sensitivity indices, offer more insight and enable the stakeholders to balance privacy vs utility according to their context.

The proposed work helps determine the optimal privacy parameters, and once the parameters are selected, those settings can be reused for future releases of the same or similar datasets. Ensuring the consistent enforcement of privacy policies, even when new data is collected or updated. The ability

to simulate and assess outcomes before deployment empowers organisations, practitioners, or researchers to publish differential private statistics tailored to their objectives confidently. Whether for policy making, economic analysis, or analysis of trends among individuals, without revealing any sensitive information.

Why we need Empirical Evaluation. Choosing optimal parameters for the implementation of a differentially private framework, one can not rely solely on the theoretical guarantees. The optimal parameters will vary from case to case, and we can not generalise them based on theoretical guarantees due to the stochastic nature of the differential privacy, which leads to the need for empirical evaluation in order to choose optimal parameters. DP provides rigorous privacy guarantees but does not provide the utility for a given dataset, query and workload. DP introduces random noise, utility targets are inherently probabilistic and depend on sensitivity, data scale, and implementation choices, which we discuss in detail in subsequent sections.

As part of this study we also explored the use of a simulation-based approach to identify an optimal ε leveraging Dwork's theoretical definition of DP, which compares the probability of obtaining the same output with and without an individual's data. While this approach aligns with the formal guarantee of DP, our empirical findings highlights the limitation when applied in practice. For example, a specific value of ε may be appropriate for one type of count query, but it might not be suitable for an average query. This suggests that such a choice of parameters might compromise the privacy despite theoretically satisfying the ratio condition. Thus, emphasising that the theoretical guarantees alone are not sufficient to assess real-world privacy and utility trade-off, and fall short in capturing different datasets and privacy parameters such as ε, δ, and noise mechanism $\mathcal{M}$ nuances. These factors can differ significantly from one use case to another.

1.2. Contribution

The key contributions of our work are as follows:

- We propose the use of well-established statistical metrics and an overarching novel framework that provides a step-by-step method to analyse the deployment of the DP mechanism in practice over real-world datasets. Specifically, we define metrics like threshold range, accuracy, Type I & II error, relative error, sensitivity index and entropy of a noise DP mechanism as discussed in Section 4. We also provide a structured

approach to do a sensitivity analysis for a given query as it is crucial to scale the amount of noise added to the query's result.

- We validate the effectiveness and usefulness of the metrics with extensive experiments over three datasets, namely, the Adult census dataset (32), Heart attack prediction dataset (6), and the Billionaire's statistics dataset (33), widely used for practical deployment of DP. For a set of commonly used queries as mentioned in Section 3, we conduct experiments and plot the results in a graph using two DP mechanisms, Laplace and Gaussian for the given $\varepsilon \in [0.1, 1]$.

- From the experiment results, we observe that our visualisation methodology helps in choosing suitable ε and the DP mechanism, allowing a balanced trade-off between privacy and utility. We show that the value of ε is a function of the query, the size of the dataset and the desired accuracy required discussed in Section 5.

- Guidance on selecting ε and deploying differential privacy mechanisms while understanding the trade-off between privacy and utility through empirical evaluation for a broader, non-technical audience.

- Lastly, we theoretically calculate the optimal ε (Algorithm 3) (denoted by $\mathcal{O}_\varepsilon$) by taking the ratio between the probability of achieving a specific noisy result with an individual's data and without an individual's data. Our experiment provide evidence that ε value is not one-size-fits-all-queries, as we were unable to find $\varepsilon \in [0.1, 1]$ for the average query for both datasets, highlighting the importance of the correct selection of ε.

In our analysis, we find that the ε value is query-centric. Using a framework like the one proposed in this paper to select a suitable DP mechanism, users can publish more queries with a smaller privacy budget while achieving a reasonable privacy and utility trade-off.

1.3. Organisation of the paper

In Section 2, we provide definitions and the notation of differential privacy relied upon in the following. In Section 3 the set of queries considered in this work. In Section 4, we define a set of metrics for the evaluation of differential privacy in practice. In Section 6 experimental results are presented, validating the usefulness of the proposed metrics and key takeaways.

2. Preliminaries

Here, we provide definitions and notation that will be used throughout the paper. Let $\mathcal{D}$ denote the Universe space of possible datasets, and in particular dataset is denoted by D. Let $\mathcal{O}$ denote the set of possible outputs. A *mechanism* $\mathcal{M} : \mathcal{D} \to \mathcal{O}$ is a randomised function that takes the database D as input and outputs $\mathcal{M}(D)$, usually a random variable. Let $|D|$ denote the cardinality of set D. If $|D| = n$, then the dataset consists of n rows, where each row represents the records of an individual having a set of attributes.

2.1. Neighbourhood Dataset

A dataset D' is called a *neighbouring* dataset of D if they are of the same size and identical except for a single record. The datasets can differ in two possible ways: Either they have the same size and differ only on one record, or one is a copy of the other with one extra record. A DP mechanism is called *unbounded* if achieved by adding or removing one record and it is called *bounded* if achieved by changing only one record. For more details see (34). While conducting our experiments we solely consider unbounded DP and create all the possible neighbourhood datasets by removing one record at a time from the datasets.

2.2. Privacy Parameter ε vs Privacy Utility Trade-off

Parameter ε represents the maximum privacy loss incurred when querying a dataset multiple times. This parameter controls the amount of noise to add; hence, it also determines what can be learned about an individual in a database. Each query contributes to the overall privacy loss, which accumulates as the sum of the ε used across all queries. A smaller ε provides stronger privacy, while a larger ε provides more utility for a dataset protected by a DP mechanism; this is referred to privacy vs. utility trade-off.

2.3. Differential Privacy Mechanisms

Definition 1. *A mechanism $\mathcal{M}$ is (ε, δ)-differentially private or (ε, δ)-DP if, for any two neighbouring databases D and D' and for any subset $S \subseteq \mathcal{O}$ of outputs (17),*

$$\Pr[\mathcal{M}(D) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta \tag{1}$$

where δ in (ε, δ)-DP is sometimes explained as the probability that the privacy loss of the output is larger than e^{ε}.

If $\delta = 0$, we say that $\mathcal{M}$ is ε-differentially private. The (ε, δ)-differential privacy is also called *approximate differential privacy.*

2.4. Sensitivity

Sensitivity depicts how much a change in the dataset will affect the output of a query. Note that, the carefully calibrated noise using $\mathcal{M}$ can mask data and the amount of noise required depends on the sensitivity of the query. Sensitivity takes into account the maximum change in the output for a dataset (say D), where all possible changes are bound to the dataset being queried and not all possible datasets.

Definition 2. *For the given query function $f : \mathcal{D} \leftarrow \mathbb{R}^d$ on a local dataset D, the* sensitivity *of f is*

$$\triangle f = \max_{D' \in \mathcal{D}} \|f(D) - f(D')\| \tag{2}$$

for all neighbouring datasets D' of D.

A query function f is *c-sensitive* if the sensitivity of f is c, that is, $\triangle f = c$.

In this work, we opt for global sensitivity (worst-case) to capture the maximum change in the query's output. This choice is standard and required for the Laplace mechanism; with the Gaussian mechanism, one may use global or local sensitivity (41). We preferred global because of conservative and stronger privacy guarantees [1] Global sensitivity also enables pre-specified bounds without inspecting the data. For instance, a company can cap the salaries at a known maximum and minimum to calibrate noise accordingly. For the calculation of sensitivity, refer to the Algorithm2.

2.5. Selected DP Mechanisms

In this work, we consider two noise mechanisms, namely, Laplace and Gaussian continuous distributions. We now define two differentially private mechanisms that depend on these distributions. The canonical example of a differentially private mechanism is the Laplace mechanism, as proposed by Dwork et al. (16).

[1] Our experimental datasets did not include extreme outliers. In deployments where outliers are expected, we recommend using pre-specified clipping bounds, see (41).

Definition 3 (The Laplace Mechanism). *Let ε and c be a positive real number and f be a c-sensitive query function. For such a function, the Laplace mechanism is defined by $\mathcal{M}_{Lap}(D) = f(D) + v$ where v is drawn from the Laplace distribution $Lap(\mu = 0, b = \frac{c}{\varepsilon})$, that is, with probability density function*

$$F(v) = \frac{\varepsilon}{2c} exp\left(\frac{\varepsilon|v|}{c}\right). \tag{3}$$

This mechanism guarantees $(\varepsilon, 0)$-DP, whereas (ε, δ)-DP is guaranteed by the Gaussian mechanism which is defined below.

Definition 4 (The Gaussian Mechanism). *Let ε, δ be constant positive real numbers however small and c be some positive real number. Consider c-sensitive query function. For such a function, the Gaussian mechanism is defined by $\mathcal{M}_2(D) = f(D) + v$ where v is drawn from the normal distribution $\mathcal{N}(\mu = 0, \sigma^2)$ where variance σ^2 is*

$$\sigma^2 = \frac{2\ \ln(1.25/\delta)c^2}{\varepsilon^2} \tag{4}$$

and probability density function $F(v)$ is

$$F(v) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-v^2}{2\sigma^2}\right). \tag{5}$$

Remark 1. *A common option is to choose a δ that is significantly smaller than $1/N$, where N is the total number of record in the database. In the worst-case scenario, each individual has a δ probability that their data leaks. So, the total odds that someone's data leaks are $\simeq N\delta$: we need to ensure this number is small enough. In this work, similar to (14), we chose $\delta = \frac{1}{N^2}$.*

3. Queries

In this section, we describe the set of queries (Q) that we executed on the real-world datasets. These queries are commonly used in the relevant literature (41; 1).

3.0.1. Count Query

It is the most common type of query that determines how many times a particular value occurs. Noise is introduced to the query's output to ensure differential privacy. Let x be the specific event and Q_c denote the count query of the event, then for a dataset D of size N we have

$$Q_c(D) = \sum_{x \in D} 1. \tag{6}$$

3.0.2. Sum Query

The query calculates the sum of desired events. It is widely used for answering statistical questions. Let x_i denote the attribute value for the ith row of a dataset D with N number of rows, then sum query denoted by Q_s

$$Q_s(D) = \sum_{i=1}^{N} x_i. \tag{7}$$

3.0.3. Average Query

It calculates the mean value of the desired events. It is the factor of sum query by the size of the dataset, that is, the average query denoted by Q_a is

$$Q_a(D) = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{8}$$

In the following, we list some important observations on queries and properties of (ε, δ)-DP, on which we will rely later, followed by short intuitive proofs and additional remarks.

Remark 2. *The sensitivity for the count query is always* 1.

Because the presence or absence of any user in a particular cell can change the outcome by at most 1, the sensitivity of the count query is always 1.

Observation 1. *The noise added in the count query function under the Laplace mechanism is independent of the size of the dataset.*

Note that from remark 2, we have $c = 1$. We know that the noise v is drawn from the Laplace distribution $Lap(\mu = 0, b = \frac{c}{\varepsilon})$, therefore v is just function of scale factor $b = \frac{1}{\varepsilon}$.

Note that this is not true for the Gaussian mechanism, as variance σ^2 is a function of δ, which itself is a function of N as discussed in remark 1.

Observation 2. *(14) Let $(p, e_\varepsilon, \alpha)$ be the triplet denoting the estimated proportion of times the noisy result is correct, the margin of error and the significance level, respectively. Then the number of times an experiment should be performed to achieve the desire accuracy is*

$$\frac{Z^2_{\alpha/2} p(1-p)}{e^2_\varepsilon} \tag{9}$$

where $Z^2_{\alpha/2}$ is the Z-value for the desired confidence level.

Confidence interval is a statistical range which estimates where a population parameter (such as mean or proportion) is likely to fall. Note that for 95% confidence level (40) the significance level $\alpha = 0.05$ and $Z_{\alpha/2} \approx 1.96$. If $p = 0.5$ is maximum uncertainty possible for large dataset with error of margin $e_\varepsilon = 0.05$ then for 95% confidence level and 95% of accuracy one requires 385 times to run the experiments.

Observation 3. *(16) For any ε-DP mechanism to have desired utility and privacy then*

$$\varepsilon \geq \frac{1}{|D|} \tag{10}$$

For the desired utility, the noise added by the mechanism should be relatively small. Specifically, if the noise exceeds the size of the dataset, the output becomes meaningless. Note that, the noise added by the ε-DP mechanism is proportional to $\frac{c}{\varepsilon}$ and we want this noise to be less than or equal to $|D|$ therefore we have

$$\frac{c}{\varepsilon} \leq |D| \quad \text{implies} \quad \varepsilon \geq \frac{c}{|D|}.$$

For simplicity, assume that $c = 1$ (remark 2) this gives $\varepsilon \geq \frac{1}{|D|}$.

In the literature, we find ε-DP put in practice using $0.1 \leq \varepsilon \leq 10$, including (4; 19; 37) with no or little convincing justification. In her work (15) mentioned the choice of ε is a social or an economic question. Choosing a suitable ε is still an open problem. In this paper, we show how, even with small values of ε ($0.1 \leq \varepsilon \leq 1$), one can achieve a meaningful balance between privacy and accuracy working with real-world datasets.

4. Differential Privacy Metrics & Framework

In this section, we define a set of metrics that provides a detailed and systematic evaluation of differential privacy in practice. This facilitates the development of a robust framework for comparing differential privacy mechanisms. For the sake of easiness consider the following notation

$$\texttt{NoisyResult} = \texttt{TrueResult} + \texttt{Noise}. \tag{11}$$

where `Noise` follows a particular distribution Laplace or Gaussian ($\mathcal{M}_{Lap}$ or $\mathcal{M}_{Gau}$). Note that, absolute difference is one of the commonly used metric in literature hence we take this metric from prior work

4.0.1. Absolute Difference

Absolute difference helps in analysing how far the noisy results are from the true result for each given ε. Let $\triangle_\varepsilon$ denote the absolute difference associated with ε. Then we have

$$\triangle_\varepsilon = |\texttt{NoisyResult} - \texttt{TrueResult}|. \tag{12}$$

4.0.2. Threshold range

It provides the range of noisy results generated by the DP mechanisms. To regulate the variability introduced by the noise with a particular mechanism, we calculate the thresholds using a percentile-based approach. The percentile helps determine the values in the data that split the distribution at specific points, providing insight into the spread of noisy results. Threshold range provides the range of noisy results generated by the DP mechanisms. Let m_ε and M_ε denote the lower and upper bound associated with ε. Then for confidence level $(1-\alpha) = 0.95$, the expression for lower and upper bound is given by

$$m_\varepsilon = P\left(\texttt{NoisyResult}, \frac{1-\alpha}{2} * 100\right), \tag{13}$$

$$M_\varepsilon = P\left(\texttt{NoisyResult}, \frac{1+\alpha}{2} * 100\right) \tag{14}$$

where P denotes the percentile. Thus, the threshold range is the interval $[m_\varepsilon, M_\varepsilon]$. We then compute the lower and upper bounds using the 2.5th and 97.5th percentiles, respectively. These bounds represent a 95% confidence interval, which means that 95% of noisy results lie within this range.

4.0.3. Accuracy

The accuracy of a DP mechanism is referred as the percentage of noisy results that lie in the pre-determined threshold range. In other words, a noisy result is said to be accurate if

$$m_\varepsilon \leq \texttt{NoisyResult} \leq M_\varepsilon,$$

where $[m_\varepsilon, M_\varepsilon]$ is the acceptable threshold range. Let C_ε and A_ε denote the total number of correct predictions and the accuracy associated with ε then we have

$$A_\varepsilon = \frac{C_\varepsilon}{1000} * 100. \tag{15}$$

To measure the percentage accuracy of each DP mechanism we have considered the proportion of noisy results that fell within a particular predetermined threshold. The accuracy threshold was dynamically measured based on the distribution of the noisy results, enabling us to quantify how consistently the noisy data approximate the true results.

4.0.4. Type I and Type II Error

Type I and Type II errors quantify the probability that noisy results appear falsely within an acceptable range (false positive) or fall falsely outside of it (false negative). These errors help us evaluate how well the mechanisms trade-off between privacy and utility. In mathematical terms, we have

$$\text{False Positive: } \texttt{NoisyResult} \in [m_\varepsilon, M_\varepsilon], \texttt{TrueResult} \notin [m_\varepsilon, M_\varepsilon],$$
$$\text{False Negative: } \texttt{NoisyResult} \notin [m_\varepsilon, M_\varepsilon], \texttt{TrueResult} \in [m_\varepsilon, M_\varepsilon].$$

The Type I and Type II errors are the proportion of false positives and false negatives, respectively, over the total number of noisy results generated.

4.0.5. Relative Error

In terms of DP the relative error measures the degree to which the noisy result deviates from the true result as a proportion to the true result. It enables to get insights about how much noise is induced by a particular queries and mechanisms. The relative error associated to ε denoted by e_ε is given by

$$e_\varepsilon = \frac{\texttt{Noisy Result -True Result}}{\texttt{True Result}} \tag{16}$$

4.0.6. Sensitivity Index

The sensitivity index in terms of DP allows us to measure how much the noisy result deviates from the true result relative to the privacy parameter ε. This metric helps determine how much noise, relative to ε, affects the accuracy of the query result. The sensitivity index associated to ε denoted by I_ε is given by

$$I_\varepsilon = \frac{\texttt{Noise}}{\varepsilon} \tag{17}$$

4.0.7. Entropy

The entropy in terms of DP is the measure of uncertainty introduced by a particular mechanism, and it serves as the measure of the variability in the noise produced by the mechanism. In this context, higher entropy means higher uncertainty and stronger privacy, while lower entropy indicates less uncertainty and more predictable results. For given ε, let $H_{\varepsilon,\mathcal{M}_{Lap}}$ and $H_{\varepsilon,\mathcal{M}_{Gau}}$ denote the entropy for the Laplace and Gaussian mechanisms, respectively. Then we have

$$H_{\varepsilon,\mathcal{M}_{Lap}} = 1 + \log(2b) \text{and } H_{\varepsilon,\mathcal{M}_{Gau}} = \frac{1}{2}\log(2\pi e_\varepsilon \sigma^2) \tag{18}$$

where $b = \frac{c}{\varepsilon}$ is a scale factor, e_ε is the relative error and σ is a variance given in Equation (4).

5. Experiment Settings

In this section, we provide a detailed analysis for experimentation along with a discussion of the observation and results. In order to demonstrate the effectiveness of the proposed metrics, we conduct extensive experiments over three widely available datasets, the Adult census dataset (32), the Billionaires statistics dataset (33), and Heart attack prediction (6). The attributes and size of the datasets are summarised in Table 1. Notably, these datasets are well established in the literature, allowing us to study the behaviour of differential privacy parameters. In particular, we focus on how changing ε affects the accuracy and utility of the results when employing Laplace ($\mathcal{M}_{Lap}$) and Gaussian ($\mathcal{M}_{Gau}$) mechanisms in real-world scenarios. For each value of ε and query Q, we conduct our experiments 1000 2 times to generate

a set of noisy results. The code and results are available on GitHub for reproducibility of the framework.[2]

Dataset	**Size** N	**Attributes**
Adult D_A	32562	Age, Gender, Native country, Capital Gain, Capital Loss
Billionaires D_B	2641	Age, Gender, Country, Final Worth, GDP
Heart Attack D_H	8764	Age, Gender, Income, BMI, Cholesterol

Table 1: Real-world Datasets

5.1. Research Questions

We aim to bridge the gap between theory and practice by helping practitioners understand and implement differential privacy without relying on complex libraries; see section 7 for library discrepancies. These research questions frame our findings.

1. For count, sum and average queries, how much does the choice of mechanism change the ε required to keep the absolute error $\leq \tau$ (acceptable error threshold)?

2. Given a fixed acceptable error threshold, does the choice of differential private mechanism influence the optimal ε?

3. For a fixed utility(depending on the user's use case), does the choice of noise mechanism matter, and does it affect the ε needed?

4. Can we rely solely on theoretical guarantees for implementing differential privacy? Does the optimal ε provide sufficient privacy in practice?

5. Does the change in privacy budget ($\Delta\varepsilon$), from question 1, stay small when we vary data size, sparsity, and class imbalance?

The framework follows Algorithm 1, takes a database D consisting of numerical attributes, and the set of queries that you want to execute. To perform differential private analysis on numerical attributes start by inputting

[2] `https://github.com/Akasha-Shafiq/Evaluating-Differential-Privacy-Frameworks-`

Algorithm 1: Publishing Differential Private Statistics

Input : Any centralised database D

Output: Differential private statistics and candidate for Optimal epsilon($\mathcal{O}_\varepsilon$)

1 Define the set of queries Q,
2 Input a range of ε,
3 Calculate sensitivity for each Q based on the neighbourhood datasets,
4 calculate $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$ scale using calculated sensitivity for desired query as mentioned in 3 and 1,
5 Tune in δ for $\mathcal{M}_{Gau}$ at least $\frac{1}{N}$,
6 Add noise to the desired Q,
7 Evaluate the choice of $\mathcal{M}$ and ε for optimal parameters using metrics in 4

the desired range for ε, the procedure outlined in Algorithm 2 creates all possible neighbourhood datasets and calculates the sensitivity of the input query. The amount of noise added in the query is determined by the calculated sensitivity and given ε. In case of $\mathcal{M}_{Gau}$ user needs to input the value of δ. Finally, evaluate the affect of the choice of mechanism and ε using metrics in 4.

Algorithm 2: Sensitivity Analysis

Input : A dataset D of size N

Output: $Q's$ sensitivity (global sensitivity)

1 Select a numeric column C for a desired $Q's$ sensitivity analysis
2 Compute the original sum or average for column C
3 For each record $i \in \{1, 2, 3, ...n\}$ perform:
4 Construct and neighbouring dataset D' by removing record i, considering all possible neighbouring datasets
5 Compute the sum or average for D and D'
6 Compute sensitivity for the desired Q, maximum change in the output

Sensitivity Analysis: To ensure a rigorous analysis of DP mechanisms, we perform a sensitivity analysis for each query type. Sensitivity depicts how much a change in the dataset affects a query's output. From remark 2, we

know that the sensitivity of the count query is always 1; however, we need to calculate the sensitivity for sum and average queries. Sensitivity is calculated by considering the neighbourhood dataset, each differing by a single record from the original dataset. In case of sum queries, the Δf depends on the range of the numerical values within the column or dataset. The difference between the dataset D and its neighbourhood D' was calculated, and the maximum change was recorded as the sensitivity of that particular query. In the same way, Δf was calculated for average queries; however, in this case the maximum change was recorded by the maximum change in the mean value of D and D' of the particular column. Sensitivity captures the worst-case scenario of individual record contribution; hence, it is the most significant parameter in the privacy-preserving mechanism.

Resource Constraint: To ensure robust privacy guarantees, we opted for global sensitivity in all our calculations. Sensitivity was calculated for each query type. The count query is fixed to 1; for the sum and average query, it was calculated as the global maximum change. The runtime for the sensitivity calculation for the sum query with a dataset size of 32000 records was 1 minute and 22 seconds, and for the dataset with 2600 records, it took 33.61 seconds. For the average query with a larger dataset, the time taken for the sensitivity analysis is 1 minute and 49 seconds, and 60 seconds for the rest of the analysis. Thus, the time taken by the smaller dataset for sensitivity analysis was 1.17 seconds, and the total execution time for the average query was 35.23 seconds. The time complexity for calculating is sensitivity is $O(n)$, as a result, the computation scales linearly with the size of D.

5.1.1. Optimal ε

Now, the natural question arises: For a fixed query in a dataset how to choose an ε in a DP mechanism that balances privacy and utility. If such an ε exists, we call it *optimal epsilon*, denoted by $\mathcal{O}_\varepsilon$. As we will see in section 6, a smaller value of ε provides a stronger privacy guarantee, while a large value of ε provides more accuracy. Thus, the $\mathcal{O}_\varepsilon$ is a value that provides a satisfactory trade-off between privacy and utility. Note that in a DP mechanism, the probability of an individual's presence or absence in the dataset differs by a factor of e^ε. This concept is known as *varepsilon*-indistinguishability. We calculate the optical epsilon using the following ratio and algorithm 3:

Algorithm 3: Calculate the optimal epsilon for a given query in a dataset

Input : Dataset(D), Query(Q), Number of simulations(s), Mechanism($\mathcal{M}$)
Output: Optimal epsilon($\mathcal{O}_\varepsilon$)

for ε **in** $[0.1, 0.2, \ldots, 1]$ **do**
  for $r = 1$ **to** s **do**
    Randomly form neighbourhood dataset D' from D ;
    Generate $\texttt{NoisyResult}_r$ and $\texttt{NoisyResult}_r^{'}$ from D and D', respectively ;
  end
  if $\frac{\Pr[\mathcal{M}(D) \in S]}{\Pr[\mathcal{M}(D') \in S]} \leq e^{\varepsilon}$ **then**
    $\mathcal{O}_\varepsilon = \varepsilon$;
    Break;
  end
end

$$\frac{Pr[\mathcal{M}(D) \in S]}{Pr[\mathcal{M}(D') \in S]} \leq e^{\varepsilon}$$

6. Results

During the experiment analysis, we first considered generic queries of interest Section 3 and described the results for each dataset below. All experiments were conducted using Python 3.11.4, with data processing, visualisation and implementation using standard libraries such as Numpy and Pandas. Each noise mechanism was implemented using custom Python functions, without relying on any available libraries for differential privacy, to maintain transparency and gain greater control over each step of the privacy analysis. The experiments were conducted on a Dell system using running Windows 11,12th Gen Intel(R) Core(TM) i7-1260P CPU operating at 2.10 GHz and 32 GB RAM. The Algorithm 1 describes the step-by-step procedure which we followed to publish and evaluate the implementation of DP mechanisms.

6.1. Adult Dataset

In this section, we examine the Adult dataset, denoted D_A, which contains a substantial amount of records as shown in Table 1. For this dataset, we implemented three queries each: for the count, sum, and average. These queries are the number of participants in the study or the total records (Q_c), the sum of capital-gain (Q_s), and the average age of the participants (Q_a). We consider two DP mechanisms, $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$, for values of ε ranging from 0.1 to 1. For each of the queries mentioned, we conducted experiments and plotted the results for each metric as a graph. Note that for each metric presented in Section 4, we have plotted graphs by varying ε and the mechanisms. Blue and orange lines are used to distinguish between the Laplace and Gaussian mechanisms, respectively.

Regarding count query (Q_c), (46) mentioned that small epsilon values, such as 0.1, provide maximum privacy and values up to 1 are considered private. However, more privacy indicates that more noise is added to the data, resulting in a significant deviation of the noisy result from the actual result. This raises concerns because, if too much noise is added, we cannot trust the accuracy of the noisy results. Our findings on Q_c for D_A are detailed below.

1. We observe that for $\varepsilon = 0.1$, $\mathcal{M}_{Lap}$ provides a significant trade-off between privacy and utility, whereas $\mathcal{M}_{Gau}$ offers more privacy compared to $\mathcal{M}_{Lap}$ (Fig 1). Similarly, for $0.3 \leq \varepsilon \leq 1$, $\mathcal{M}_{Lap}$ provides more accuracy and less privacy, whereas $\mathcal{M}_{Gau}$ provides more privacy than $\mathcal{M}_{Lap}$. Consequently, $\mathcal{M}_2$ offers greater accuracy at smaller values of ε in contrast to $\mathcal{M}_{Lap}$.

2. $\mathcal{M}_{Gau}$ exhibits a wider threshold range compared to $\mathcal{M}_{Lap}$. This implies that $\mathcal{M}_{Lap}$ gives less variability and more consistent results. Whereas, a narrower threshold for $\mathcal{M}_{Lap}$ indicates that the noise produced by $\mathcal{M}_{Lap}$ is tightly controlled, leading to more accurate results.

Next, we consider the sum query Q_s, the sum of capital-gain in the Adult dataset. We observe that the sensitivity of this query is $\Delta f = 4356$, which is much larger than that of the count query ($\Delta f = 1$).

Additionally, based on the experimental results, we can make the following observations in relation to the sum query Figure 2).

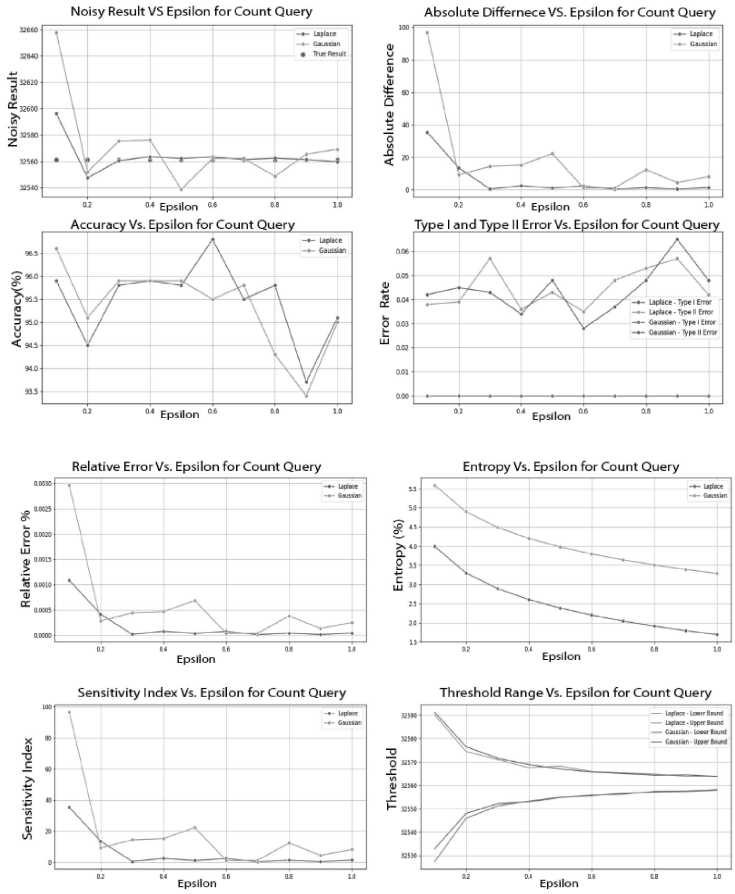


Figure 1: Results for count query, the number of participants in the study or the total records (Q_c) in Adult dataset (D_A)

1. $\mathcal{M}_{Lap}$ provides more stability, consistency, and lower entropy compared to $\mathcal{M}_{Gau}$ for high-sensitive sum query, whereas in count query result were opposite.

2. The threshold range $[m_{\varepsilon,M_\varepsilon}]$ for $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$ converges for $0.1 \leq \varepsilon \leq 1$.

3. In this query, the threshold is wider compared to the sum query, particularly for small values of ε. As ε increases, there is a sharp decline in the threshold range, indicating that the noise is tightly controlled for a large ε. Hence, it has more utility since the noisy result is accurate

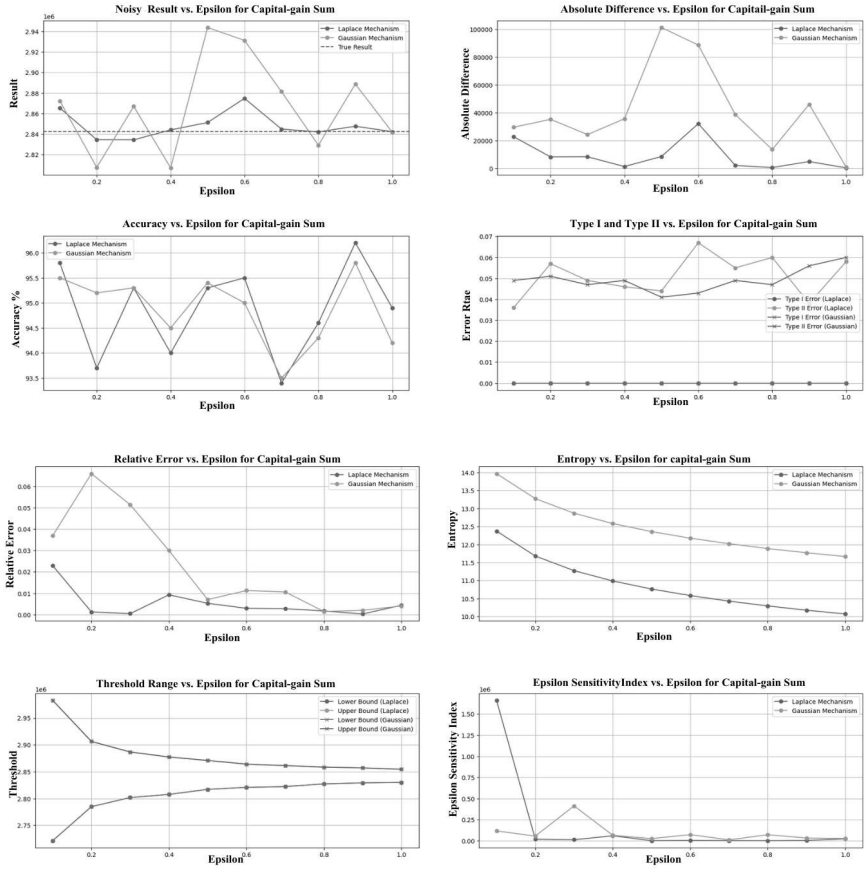


Figure 2: Results for sum query, the sum of capital-gain (Q_s) in Adult dataset (D_A)

and close to the true result.

4. $\mathcal{M}_{Lap}$ has better accuracy than $\mathcal{M}_{Gau}$. We note a sharp decline of I_ε for $\mathcal{M}_{Lap}$ in contrast to $\mathcal{M}_{Gau}$ (sensitivity index vs epsilon, Figure 2). The decline in I_ε and thresholds shows that the noise is tightly controlled at large ε.

Now we define the analysis for the average query Q_a, the average age of participants, note that the sensitivity for this query is $\Delta f = 0.01579$ which is smaller than the count sensitivity. We can make the following observations on Q_a on D_A (detailed graphs in Figure 3.

1. In the context of D_A or any significantly large dataset, the sensitivity

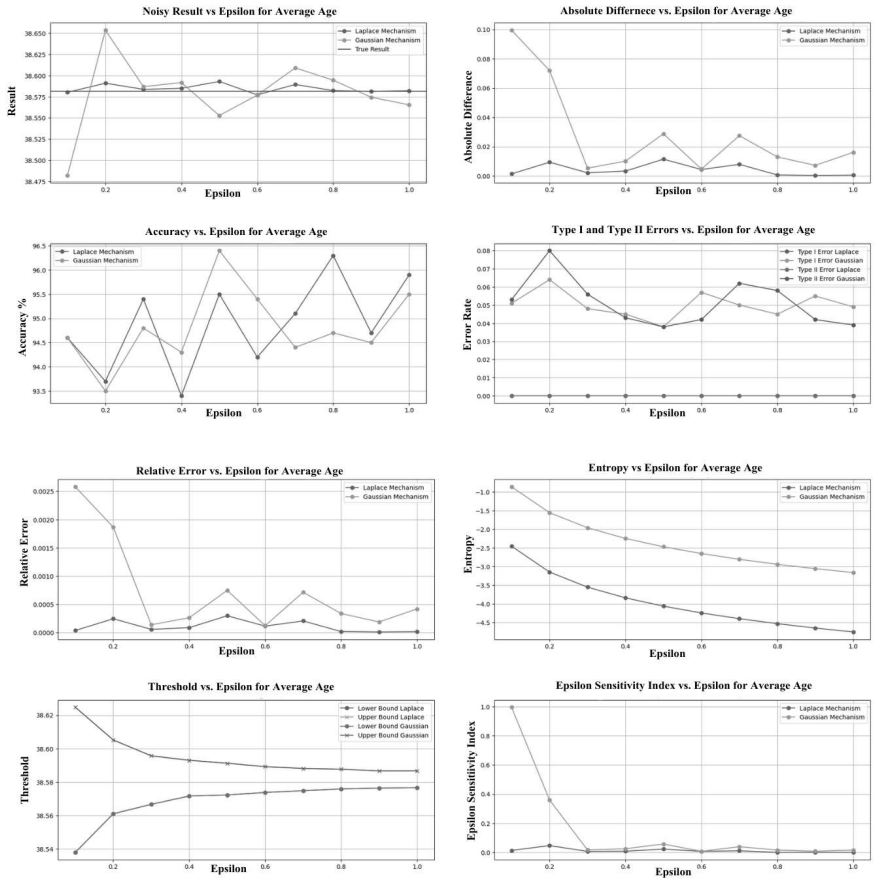


Figure 3: Results for average query, the average age of the participants (Q_A) in Adult dataset (D_A)

for 'Age' tends to be quite low. This is attributed to the minor influence each individual exerts on the cumulative outcome. In fact, the age can be bound between 0 and 122. In our case, for D_A, the maximum age is 90 and the minimum is 17. If one already knows the participant's age, they can set the upper and lower bounds.

2. We observe that the DP framework achieves high accuracy with both mechanisms $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$. The threshold value is nearly identical $[m_\varepsilon, M_\varepsilon]$. Hence, the choice of mechanism has little impact on the query's output.

3. For this query, from the accuracy vs. epsilon graph, we observe that both mechanisms have the same accuracy except for $\varepsilon \in [0.4, 0.6]$ where $\mathcal{M}_{Gau}$ demonstrates greater accuracy.

4. $\mathcal{M}_{Gau}$ is more sensitive for $\varepsilon \in [0.1, 0.5]$ but stabilises as ε increases, whereas $\mathcal{M}_{Lap}$ shows a stable sensitivity index throughout the ε range.

6.2. Billionaires Dataset

In this section, we discuss the Billionaire dataset denoted by D_B, which contains significantly fewer records compared to the Adult dataset D_A Table 1.

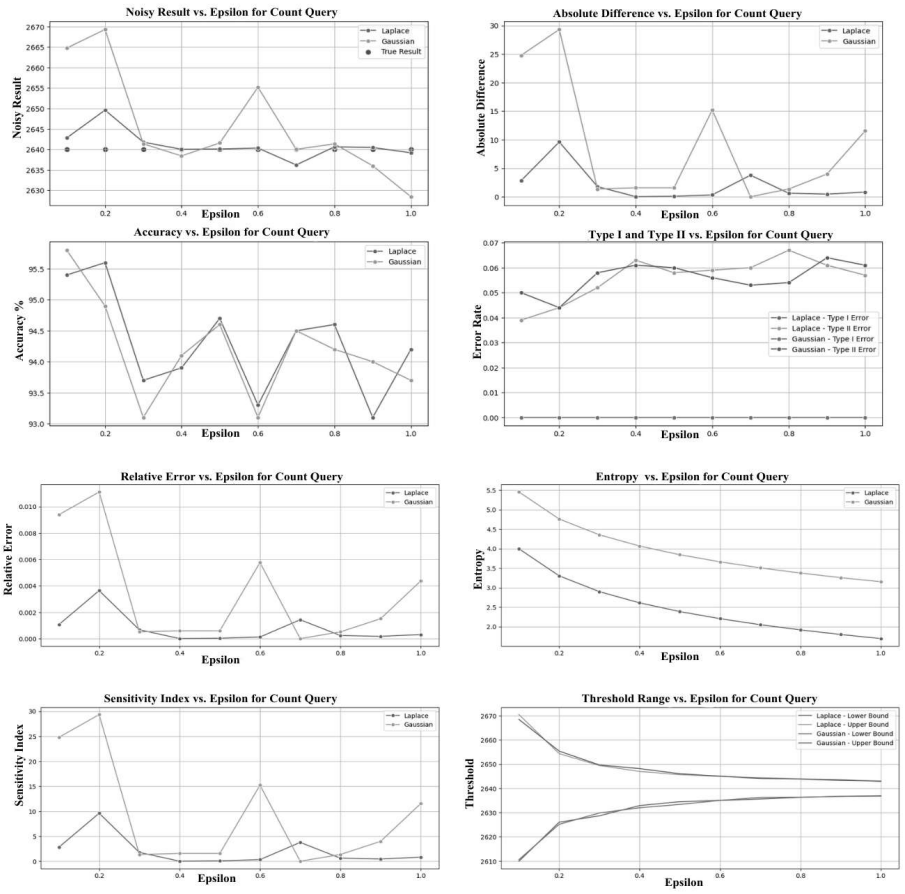


Figure 4: Result for Count Query, the total number of participants in the study or the total records (Q_c) in dataset (D_B)

In this dataset, we implement three queries from each category: count, sum, and average. These queries include the number of participants in the study or the total records (Q_c), the sum of Billionaires' final-worth (Q_s), and the average age of the participants (Q_a).

For each query, we conducted experiments and plotted the results, varying ε and the mechanisms for each metric mentioned in section 4.

We have the following observations for the count query (Q_c) in Billionaire dataset (D_B), fig. 4.

1. Our key observation is that for $\varepsilon \in [0.1, 0.4]$ $\mathcal{M}_{Gau}$ is more sensitive, whereas $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$ behave similarly for larger ε values, that is, for $\varepsilon \in [0.5, 1]$.

2. For the count query in D_B, we observe that $\mathcal{M}_{Lap}$ exhibits more stable accuracy compared to $\mathcal{M}_{Gau}$ for all $0.1 \leq \varepsilon \leq 1$. This shows that the mechanism $\mathcal{M}_{Lap}$ is the preferred choice when predictable accuracy is required.

Next, we consider the sum query, the sum of Billionaires' final-worth in Billionaire dataset. We observe that the sensitivity of this query is $\Delta f = 211000$, which is much larger than that of the count query. We have the following observation, fig. 5).

1. In this query, our key observation is that the mechanism $\mathcal{M}_2$ fluctuates significantly at $\varepsilon = 0.1$. This occurs due to the inconsistency of noise in the mechanism with a smaller size dataset and higher sensitivity.

2. In the case utility is a priority, we recommend using $\varepsilon > 0.5$. Conversely, a smaller value may be preferred when prioritising privacy.

3. Lastly, we observe that both noise mechanisms behaves similarly for $\varepsilon > 0.3$.

Here we discuss the result for Q_a, the average age of the participants in the Billionaire dataset. Note that the sensitivity for this query is $\Delta f = 0.01831$, which is slightly higher than the Adult dataset. We have the following observations Fig 6).

1. In this query, the mechanism $\mathcal{M}_{Gau}$ demonstrates a suitable trade-off between privacy and utility at $\varepsilon = 0.1$.

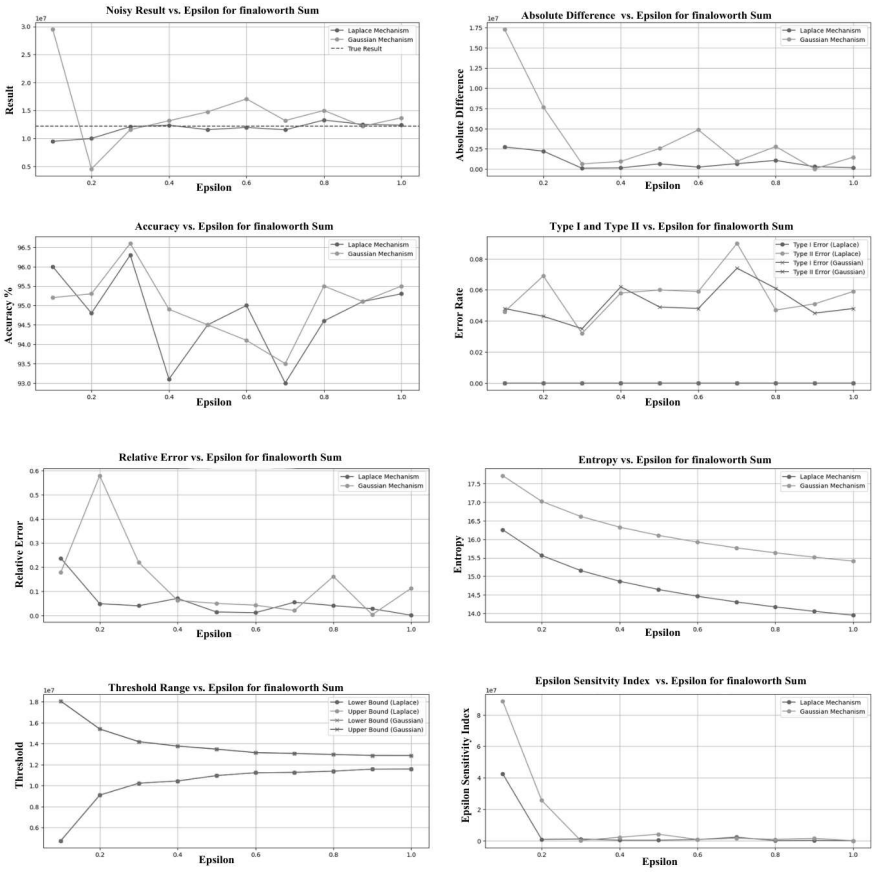


Figure 5: Results for sum query, the sum of Billionaires final-worth (Q_s) in Billionaires dataset (D_B)

2. We observe that $\mathcal{M}_{Lap}$ adds less noise and produces results very close to the true result across all values of ε. This shows no privacy.

3. Note that for all queries, the noisy result always falls under the threshold range, hence we likely get Type I error only, whereas both mechanisms exhibits Type II error. In fact, $\mathcal{M}_{Lap}$ has a marginally lower error rate due to the tighter distribution of noise around $(0, b)$ where b is a scale factor dependent on sensitivity and epsilon.

4. $\Delta f = 0.01831$, with this low sensitivity, we observe that the $\mathcal{M}_{Lap}$ mechanism provides consistently low error rates, stabilisation, and tighter

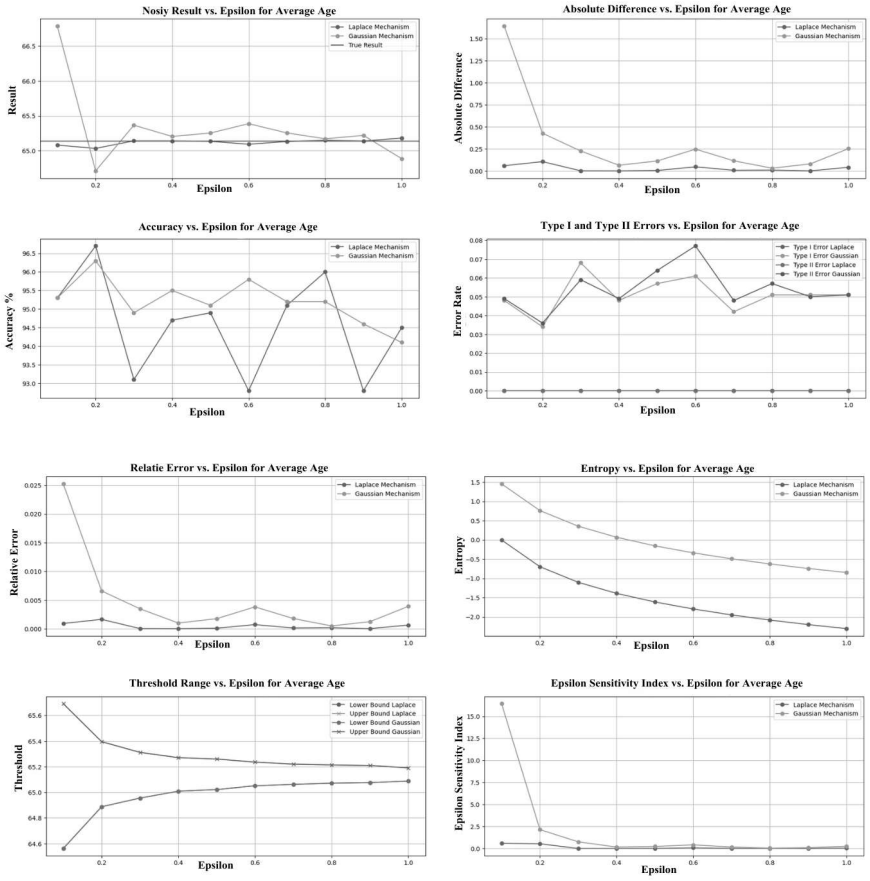


Figure 6: Results for average query, the average age of participants (Q_a) in billionaires dataset (D_B)

thresholds.

6.3. Heart Attack Dataset

In this section, we discuss the results for the Heart Attack prediction dataset denoted by D_H, which contains a significant amount of records as mentioned in 1. The queries executed for this dataset are the number of participants Q_c, the sum of income of individuals Q_s and the average income of the participants Q_a.

For the count query, the sensitivity is 1 and the total execution time for analysing Q_c is 36.89 seconds, the sensitivity for Q_s is $\Delta f = 299954$ and total

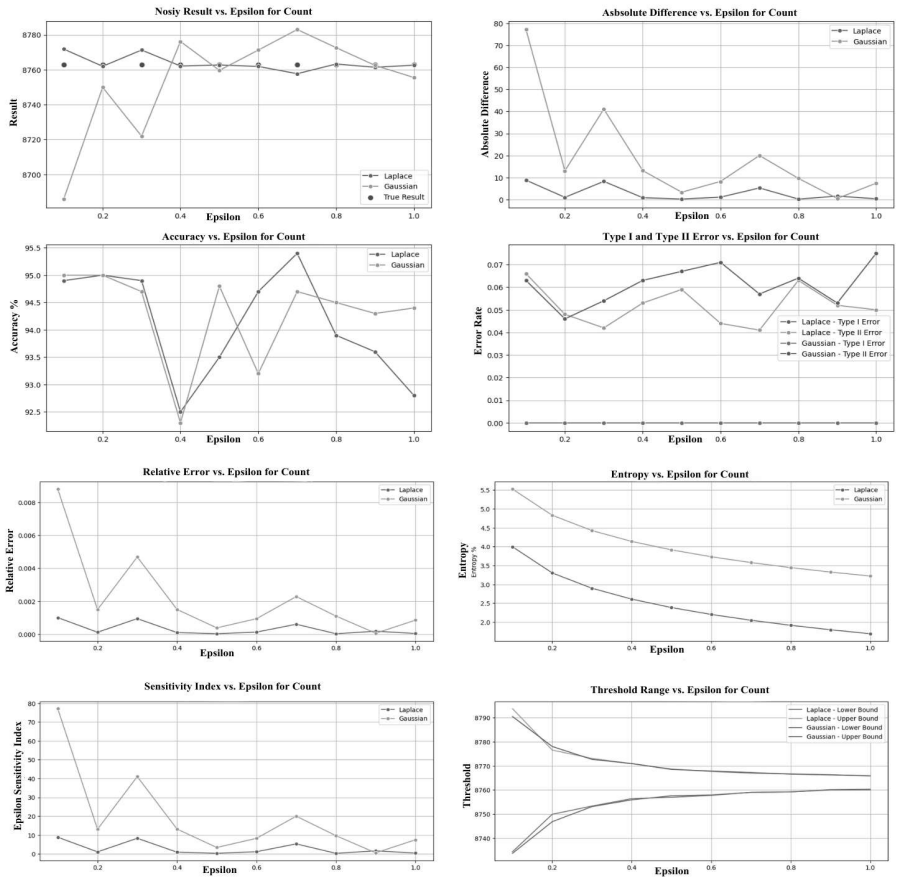


Figure 7: Results for count query, total number of participants (Q_C) in heart attack dataset (D_H)

analysis time was 49.89 seconds. The sensitivity for Q_s is $\Delta f = 16.17105$, and the total analysis time taken is 50.91 seconds.

The figures attached show the analysis of the heart attack dataset, demonstrating how the DP mechanism and the required privacy budget vary across use cases and datasets.

1. The figures illustrate that smaller datasets require less privacy budget to achieve the same level of privacy compared to larger datasets in case of Q_c and Q_s. However, the noise thresholds for smaller datasets are much wider than those for larger datasets. Additionally, the threshold declines sharply for Q_c, Q_a, and Q_S, as shown in Fig. 7, 8, and 9.

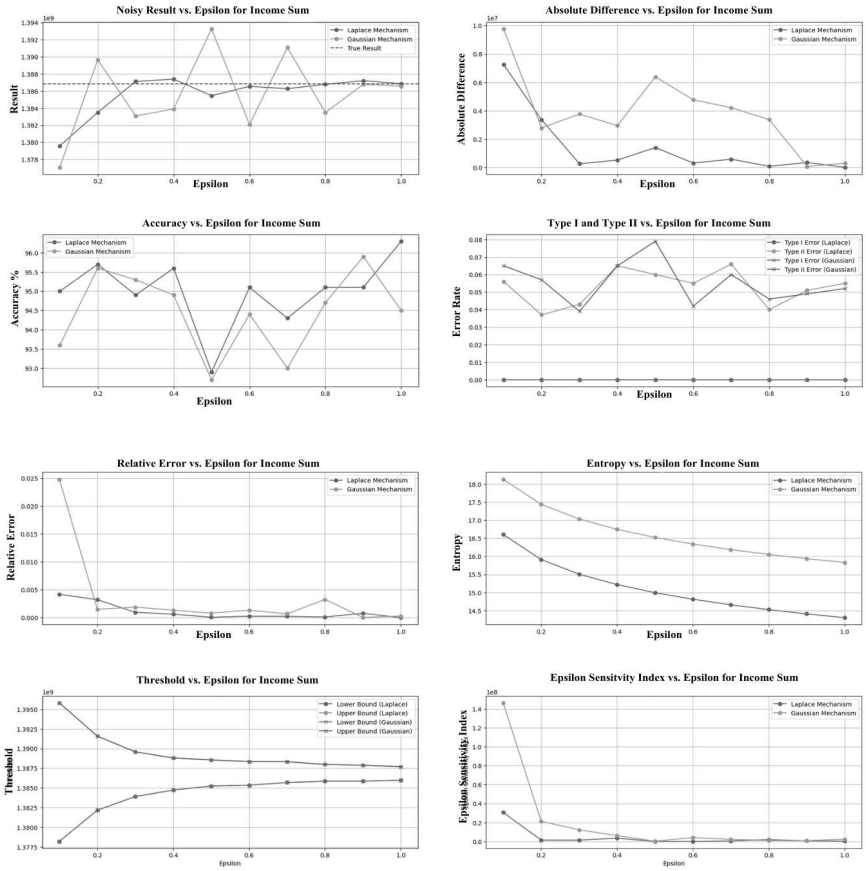


Figure 8: Results for sum query, the sum of income (Q_s) in heart attack dataset (D_H)

2. In the case of Q_a in Fig. 9, we observe that the $\mathcal{M}_{Lap}$ offers almost no privacy in this scenario. While the acceptable error range (95% accuracy) from $\mathcal{M}_{Lap}$ offers high utility, it comes at the expense of a high privacy budget and privacy loss.

3. Though with slightly higher Δf and larger D size DP mechanisms offer an acceptable trade-off as compare to D_B but still ε-DP offers almost no privacy.

4. The impact of the DP mechanism selection diminishes as the dataset size decreases, with the value of ε for switching $\mathcal{M}$ reducing to between 0.0 and 0.1.

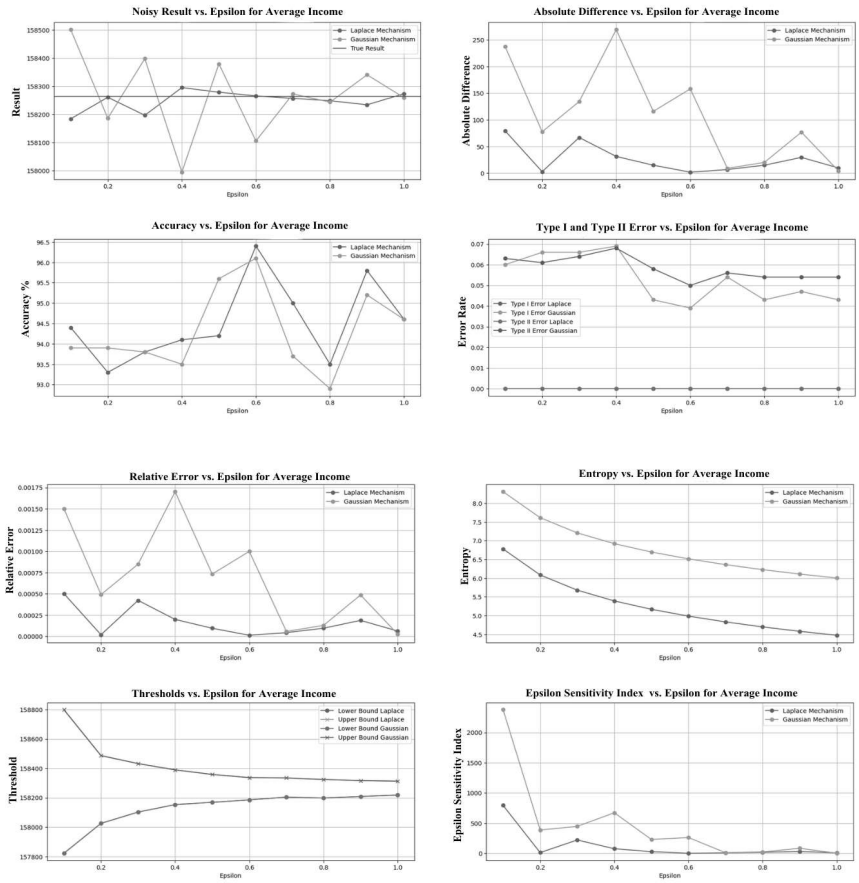


Figure 9: Results for average query, the average income of participants (Q_a) in heart attack dataset (D_H)

5. As the D's size decreased and with a change in Δf, we observed the privacy budget required to achieve 95% accuracy increases but the other supporting evaluation metrics show that relying on the theoretical assumptions is not reliable the practitioners sometimes might over or underestimate the amount of noise produces by the DP mechanisms.

6. The theory suggests that as the size of D decreases, the differential privacy (DP) mechanism adds more noise. However, our findings indicate that this is not always the case. It is evident that the thresholds decline sharply in this scenario. The only value that aligns with this intuition is $\varepsilon = 0.1$.

6.4. Key Takeaways

We answer the research questions (section 5.1) by using the 5% relative error cut off practised by statistical agencies and DP practitioners as mentioned in (5), (42), (44). However, this acceptable threshold can vary for different use cases. For each query we measure the relative error both under $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$, systematically varying $\varepsilon = \{0.1, 0.2, ..., 1.0\}$. All experiments involved 1000 simulations to ensure the statistical robustness, and results are reported as the mean with a 95% confidence interval.

- Mechanism impact is negligible. The accuracy and threshold bounds clearly indicate that, in this use case, the mechanism choice is immaterial, even though there is a slight difference in relative error, suggesting minimal impact. In our case, according to eq. (15), the smallest ε for which both mechanisms cross 95% and meet the 5% error requirement is the minimum required privacy budget. However, in case of a few Q's to meet this threshold requirement, a larger ε is required. Although other metrics indicate that those big ε satisfy this condition, but might not be acceptable according to the use case, as results were so close to the true values. Thus, implying the practitioner may need to deviate from the established accuracy threshold.

- The Gaussian mechanism achieves the 95% accuracy threshold at any ε value $\approx$0.2 lower than Laplace, irrespective of the dataset, sensitivity and query type. The $\approx$0.2 offset appears for all the queries. This uniform change in $\Delta\varepsilon$ that the Gaussian mechanism consistently offers a small, fixed privacy budget advantage over Laplace when targeting the same accuracy.

- Although theoretical (ε, δ) treats Laplace and Gaussian distribution under different norms, our empirical results show that the numerical effect tends to shift the curves slightly. The ε-DP theoretically adds more noise as compared to (ε, δ)-DP due to the small probability of privacy leakage (δ), meaning systematically the Gaussian should be adding less noise for the same value of privacy budget. It is quite evident from our evaluation that the Gaussian mechanism adds more noise as compared to Laplace at small ε, opposite to what theory predicts as indicated by the entropy calculation and noisy results.

- The choice of mechanism will always influence the selection of ε. However, based on experimental results under similar use cases, the influence is not significant. When switching from the Laplace mechanism to the Gaussian mechanism, you can allocate a small mechanism cost of $\Delta\varepsilon \approx 0.1$ to 0.2, i.e consider this change as mechanism cost to switch from $\mathcal{M}_{Lap}$ to M_{Gau}. This offset persists regardless of the data scale, and the query sensitivity practitioner can consider this as the Gaussian cost when targeting the same error bound rates.

- In our experiment, the finalworth of the billionaires is an example of a heavily skewed distribution net worth of the individuals follows a Pareto-like tail. This skew will fundamentally alter how the DP mechanism adds noise to the data due to the high sensitivity, thus requiring more privacy budget to achieve the same accuracy compared to the other balanced queries (count and average) and datasets. In this scenario, it still does not shift how much extra privacy budget you pay when you switch mechanisms. However, due to the high sensitivity and skewness of the data, the noise introduced by small ε is higher as compared to other datasets. Thus, we recommend using a slightly higher ε.

- The Algorithm 3 highlights the key strength of our multi-metric empirical calibration framework. Relying solely on theoretical guarantees can overshoot the actual minimal ε needed for the desired utility vs privacy trade-off. Thus, combining the simulation-based checks with theoretical guarantees, we can pinpoint ε that satisfies both privacy and utility constraints and avoid wasting privacy budget on negligible utility or privacy improvements.

- Currently, there are not many algorithms in the literature (8; 29) that faciltate finding the ε. We analyse an Algorithm 3 to find an $\mathcal{O}_\varepsilon$. Note that we set the number of simulations s at 1000, that is in line with observation 2, which states that for 95% of accuracy, we need at least 385 simulations. We summarise the results in Table 2. Due to the specified range of ε, we could not find $\mathcal{O}_\varepsilon$ for average queries on both datasets, which we pointed out in Table 2 by a question mark. This shows that $\varepsilon > 1$ may be effective, as no given ε satisfies the ratios for Q_a. We run the experiment for all datasets considering $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$.

Dataset	Mechanism	Queries	Optimal epsilon
D_A	$\mathcal{M}_{Lap}/\mathcal{M}_{Gau}$	Q_c	0.8/1.0
		Q_s	1.0/1.0
		Q_a	?
D_B	$\mathcal{M}_{Lap}/\mathcal{M}_{Gau}$	Q_c	0.9/1.0
		Q_s	1.0/1.0
		Q_a	?
D_H	$\mathcal{M}_{Lap}/\mathcal{M}_{Gau}$	Q_c	1.0/1.0
		Q_s	1.0/1.0
		Q_a	1.0/1.0

Table 2: Optimal Epsilon value for Real-world Datasets, where '?' Means no Optimal Value within given Range

The findings suggest that the implementation parameters for privacy can not be generalised for all sorts of datasets and use cases. The analysis also depicts how just relying on theoretical guarantees is not enough due to the stochastic nature of DP, for the implementation of differential privacy in real-world scenarios. The literature relies heavily on the relative error measure to assess the utility of the implemented mechanism or library. The use of proposed metrics to analyse the utility and selection of parameters promotes the understanding, adaptability of differential privacy. The proposed work bridges the gap between theory and practice, analyses DP guarantees without relying on DP libraries, and provides a framework for practitioners to implement DP and make informed decisions about its implementation.

7. Discussion

Optimal Epsilon $\mathcal{O}_\varepsilon$. We show that the optimal ε depends on the particular query and context in which DP is implemented, which suggests that an arbitrary, a-priori selection of the optimal value cannot provide adequate results.

(31), mentions that the high sensitive queries require a high noise adding mechanisms ($\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$), whereas our observation shows that both mechanisms $\mathcal{M}_{Lap}$ and $\mathcal{M}_{Gau}$ have a similar threshold and lead to a better utility vs privacy trade-off for ε in the standard range, that is, $0.1 \leq \varepsilon \leq 1$. We observe that high sensitivity and dataset size are the two responsible

factors along with the choice of ε. We also observe that with low sensitivity query, one can achieve enough utility at smaller values such as $\varepsilon = 0.1$. Moreover, the impact of high sensitivity also seems to diminish due to the large size of the dataset. Also, in (26) the authors stated the open problems for an analyst to publish maximum queries in a given fixed privacy budget. Our experiments using the Adult dataset indicate that a utility-privacy trade-off can be maintained even with $\varepsilon = 0.1$, allowing analysts to publish more queries. For instance, with a privacy budget of 1, it is possible to answer ten queries in a differentially private manner. For low-sensitivity queries, we observe that analysts have the flexibility to choose ε values smaller than 0.1, which helps to conserve the privacy budget.

Libraries Dependency Practitioners generally rely on libraries instead of implementing mechanisms from scratch. Each library, however, has its own conventions for parameters like scale, δ, sensitivity, mechanism, and numeric precision. Consequently, we assess the gap between theoretical and practical applications without relying on DP libraries.

Using IBM's diffpriv tools (28) for data analysis, such as count non zero, sum mean, and all the high-level functions, seems to take both parameters ε and δ, but they only use ε and if you pass δ it drops it and you get a compatibility warning. These functions use bounds which automatically pick min and max value from the data which lead to adding more noise if you have outliers in your data. Passing an entire data frame to the count function merges all columns into a single list and counts all non-zero cells rather than records. To obtain a Gaussian mechanism (ε, δ), one should use 'diffprivlib.mechanism' and specify the correct sensitivity. The issue with the mechanism is that the analyst or user has to enter the sensitivity manually. The user also has to set a budget accountant, which keeps track of the epsilon of the total budget and how much one spent on each query. One can not fix a budget without evaluating what works for the desired use case of DP implementation. Moreover, when we implemented the same queries as mentioned in section 6 and section 6.2, the library added more noise for the same values of ε. In the case of the adult dataset, when we executed the count query Q_c using the library, it added more noise than the theoretical guarantees, and strangely, sometimes returned the noisy result exactly as the true result. The Gaussian mechanism adds more noise compared to Laplace, but when we executed the same queries with a library specifically for a smaller dataset, even the Laplace mechanism added more noise, which was more than our results of the Gaussian mechanism.

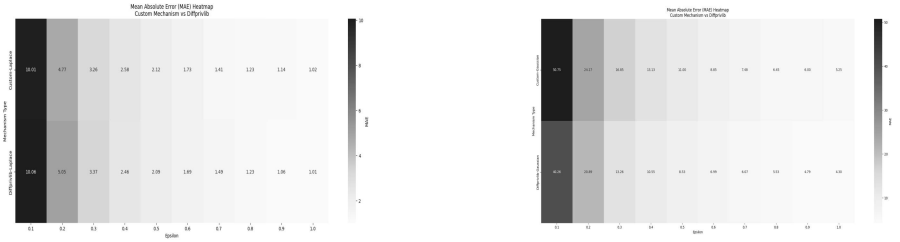
(3) gave an overview of three different libraries where he compared benchmark libraries IBM's Python diffpriv library, Google's C++ differential privacy, and Rubinsten's R diffpriv library on a synthetic dataset, measuring differential private mean behaviour as you vary the dataset and ε. The libraries behave as expected, meaning a larger ε yields lower error as the noisy results approach the true result. But the author demonstrated that Google's library does something odd when you pick a very small ε instead of adding more noise at a smaller ε it clips to the midpoint of the bounds and returns a value without noise, which violates the DP behaviour. Moreover, IBM and R's libraries do not exhibit the same behaviour but they do produce unbounded error resulting into adding arbitrarily large errors for smaller ε. He also demonstrated that all the libraries show larger noise on smaller datasets, but Google's library behaves oddly and shows it's noise variance peaks at the smaller dataset (snapping mechanism collapses noise to zero at smaller ε) but for IBM's it grows linearly. Also, the R's library erratic error patterns, adding less noise yet producing high errors for the larger dataset, suggest hidden sensitivity assumptions that can mislead users. The study by (47) explores and highlights the usability and trust issues surrounding DP libraries, illustrating a conceptual and mental model of DP libraries. Highlighting how the users often rely on open source transparency and institutional backing to assess the trustworthiness of the libraries, and the libraries do not always facilitate the understanding of the DP concepts. Thus, the DP libraries often struggle to help users implement correct solutions.

Moreover, these libraries can not be used without the expert knowledge of the domain.

8. Comparison Analysis/Interoperability

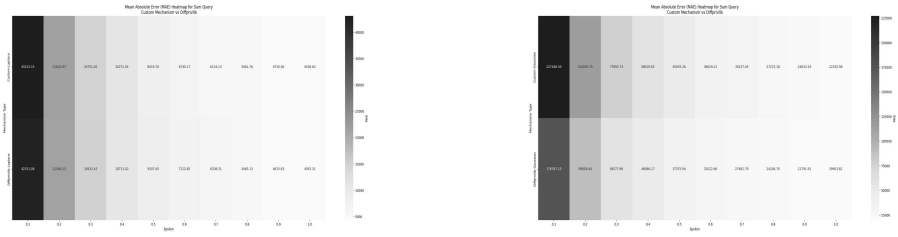
To evaluate the generalisation and real-world applicability of our framework, we analysed the noisy results of our implementation versus IBM's differential privacy library (28). This comparison enables practitioners to evaluate how closely third parties aligned with controlled empirically validated implementation. For a fair evaluation, we used the same parameter values as in our implementation; however, the library still encounters the issues mentioned in section 7. Instead of letting the library automatically select the maximum value from the dataset, we manually adjusted the sensitivity and other parameters.

We used Mean Absolute Error(MAE) (8) as a distance metric to compare the results across the range of ε. MAE is the absolute value of the difference between the true and noisy result. The graphs show the heatmaps generated for both implementations, for count, sum, and average queries, under both mechanisms.



(a) Laplace (b) Gaussian

Figure 10: MAE for Count Query

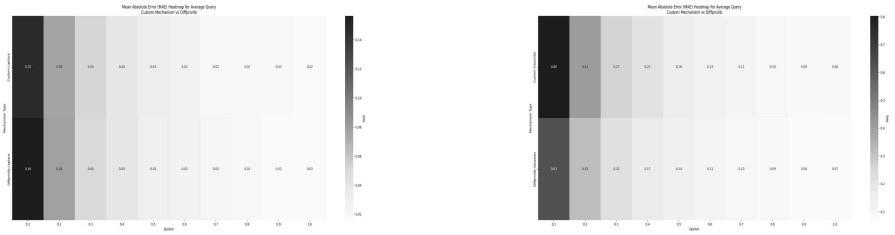


(a) Laplace (b) Gaussian

Figure 11: MAE for Sum Query

As evident from heatmaps both implementations exhibit similar MAE across range of ε value, depicting the correctness of the implementation. At smaller values of ε our implementation shows a marginally lower or similar MAE as compared to library. The heatmaps shows that the choice of mechanism has minimum impact on overall error for all the queries, reinforcing the earlier conclusions from our framework's metrics .

The tables 3 to 5 summarises the standard deviation of our implementation versus libraries for each query over a range of ε. The mean of noisy outputs assesses the bias introduced by the privacy mechanism. The mean closer to the true value shows that the mechanism maintains high accuracy



(a) Laplace (b) Gaussian

Figure 12: MAE for Average Query

and is unbiased on average. Whereas, standard deviation measures the spread or variability of the noisy outputs. A lower standard deviation means the noise outputs are tightly concentrated around the mean, ie., the noise is more stable and predictable. As ε increases both implementations demonstrates the decrease in noise, reflected by the reduction in standard deviation and MAE values. Our implementation mechanisms closely match the Diffprivlib in terms of accuracy and statistical consistency, specifically at mid to high values of ε.

Epsilon	Our STD *Lap*	Our Mean *Lap*	DiffPriv STD *Lap*	DIffpriv Mean *Lap*	Our STD *Gau*	Our Mean *Gau*	Diffpriv STD *Gau*	Diffpriv Mean *Gau*
0.1	13.66	32560.26	14.34	32560.81	63.37	32564.47	49.58	32563.54
0.2	6.96	32560.81	6.93	32560.88	32.23	32560.15	26.40	32561.25
0.3	4.92	32561.11	4.56	32560.91	21.43	32561.06	17.52	32561.09
0.4	3.43	32561.08	3.47	32560.94	16.88	32560.62	13.15	32561.29
0.5	2.92	32561.15	2.71	32561.04	13.01	32561.28	10.93	32560.67
0.6	2.22	32561.06	2.30	32560.95	10.79	32561.13	9.18	32561.60
0.7	2.09	32561.06	2.03	32560.93	8.91	32561.26	7.82	32561.38
0.8	1.74	32561.05	1.78	32561.01	8.41	32560.98	6.80	32560.67
0.9	1.55	32560.97	1.59	32560.98	7.19	32560.71	5.73	32561.10
1.0	1.41	32561.00	1.39	32560.94	6.45	32561.09	5.59	32561.25

Table 3: Standard Deviation and Mean for Count Query Under Gaussian and Laplace Mechanism

For instance, at $\varepsilon = 0.3$, the difference between the STD and the mean of our implementation and diffprivlib is significantly small, which is due to the stochastic nature of DP. Also, this small difference reflects the consistency and the practical utility of our framework to validate outputs.

Furthermore, we evaluated the two most commonly used libraries for publishing differentially private statistics (38) and (28), following the instructions given by the libraries. The fig. 13 shows the implementation of the average age query using the diffpriv library; this huge noise is the result of just taking the maximum value in the dataset. To conduct a fair comparison with our implementation, we made some adjustments to execute the average query

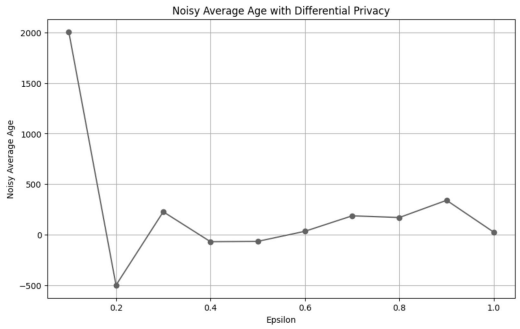
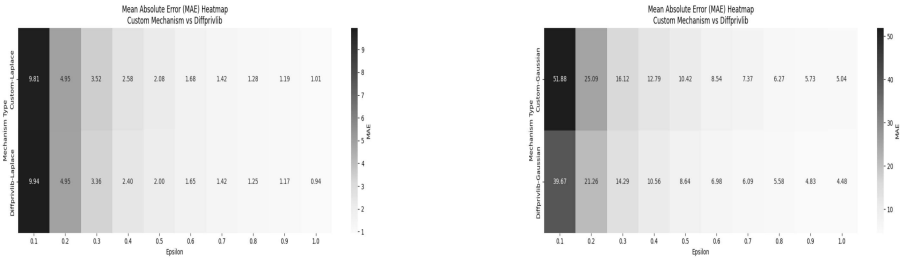


Figure 13: Noisy Age

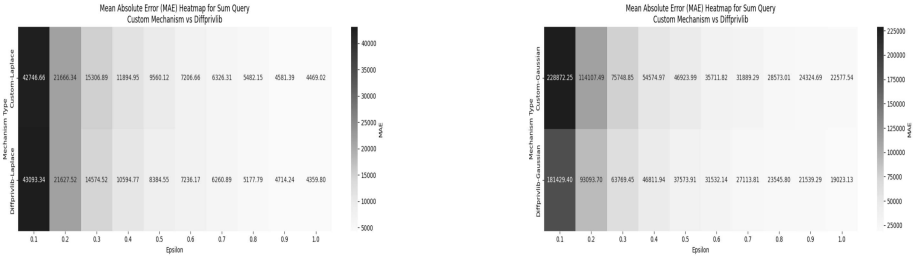
using the Diffpriv and opendp libraries. The datasets used in our experimentation do not contain outliers. As a result, the libraries produce optimal noisy outputs that remain in reasonable bounds for utility. However, this does not generalise to the high-variance datasets. In scenarios where significant outliers exist, Diffprivlib, rely on the maximum to minimum bound to calibrate noise. This leads to the addition of excessive noise that can degrade utility if the bounds are not set correctly, thereby violating DP guarantees.



(a) Laplace (b) Gaussian

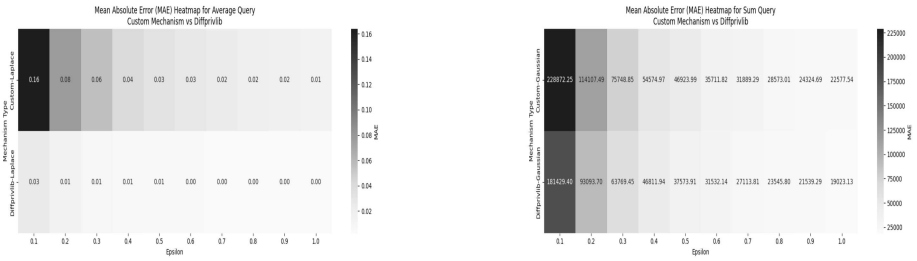
Figure 14: MAE for Count Query using Diffpriv Library

The figs. 14 to 16 and fig. 17 shows the comparison for three different queries. We compared our custom implementation with the libraries, as seen in the figures. It is evident from the comparison of both libraries that our Gaussian mechanism exhibits more MAE than the libraries. However, the experiment section shows that our implementation, even with added noise, still offers an acceptable privacy vs. utility trade-off. This also indicates that libraries provide way less privacy (less noise) for the same value of ε, focusing more on utility. While implementing DP with libraries, the user



(a) Laplace (b) Gaussian

Figure 15: MAE for Sum Query using Diffpriv Library

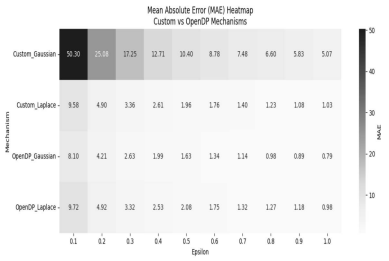


(a) Laplace (b) Gaussian

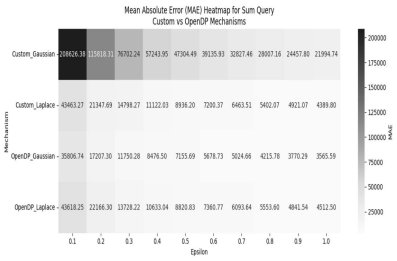
Figure 16: MAE for Average Query using Diffpriv Library

has little control over parameters, such as sensitivity or δ, which results in a low error rate at small ε. While our implementation provides flexibility and customisability, this indicates the need for frameworks to evaluate these libraries' implementations and highlights the case where mechanisms may require additional tuning.

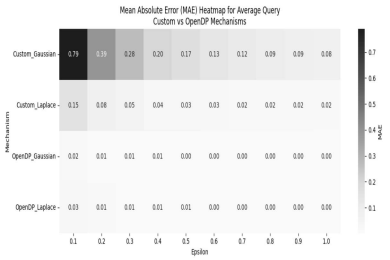
Our empirical analysis also reveals that several implementations return extremely low error in some cases zero, over a wide range of ε, including very low privacy budget. This suggests that the libraries are introducing insufficient noise, raising concerns about weak privacy guarantees. According to the DP definition, a lower ε should correspond to higher noise, thereby compromising the DP guarantee. This discrepancy observed suggests either incorrect sensitivity, setting, overly relaxed bounds or internal optimisations that compromise the theoretical guarantees for practical usability. These findings also align with the literature (48; 15; 13; 44), which indicates that differential privacy libraries do not enforce robust privacy by default and may lead to overly optimistic utility at the cost of privacy. This also high-



(a) Count Query



(b) Sum Query



(c) Average Query

Figure 17: MAE for Queries using OpenDP

lights the need for transparency in sensitivity assumptions, noise calibration and internal safeguards when using off-the-shelf libraries. Moreover, these libraries present high usability barriers, specifically OpendDp, which requires users to define explicit domains, bounds, and transinformation, which are non-trivial for practitioners without expertise in differential privacy. This makes libraries less accessible for data scientists and engineers seeking to integrate DP mechanisms into real-world systems without detailed theoretical knowledge.

Users can confidently use our framework to evaluate and understand the DP implementation, to evaluate ε values and mechanisms selection before applying a library, ensuring not just theoretical but backed by the accuracy and error trade-off.

Our proposed work is to enhance usability and complement DP implementations for all types of users, enabling critical insights before deployment by: empirically validating ε choices specific to the dataset and queries, and quantifying expected noise and utility using intuitive metrics. Practitioners can use the framework in their validation pipelines to test different privacy budgets and mechanisms to use with existing libraries in advance, thereby reducing the guesswork and improving deployment confidence in privacy-

Epsilon	Our STD *Lap*	Our Mean *Lap*	DiffPriv STD *Lap*	DIffpriv Mean *Lap*	Our STD *Gau*	Our Mean *Gau*	Diffpriv STD *Gau*	Diffpriv Mean *Gau*
0.1	60461.18	2843730.26	61084.61	2844966.47	288189.38	2841833.47	216349.35	2843075.33
0.2	30511.92	2842107.87	31604.04	2842014.47	150892.08	2846731.68	110099.50	2840885.89
0.3	20637.80	2843048.43	19631.88	2842796.97	96200.56	2840438.28	74957.56	2844546.82
0.4	15198.58	2843407.29	14298.33	2842685.25	69430.28	2838344.97	60383.22	2845049.30
0.5	11512.80	2843070.56	12282.31	2842397.23	56190.34	2843414.33	48579.64	2841966.74
0.6	9966.65	2842855.82	10352.96	2842414.84	48001.74	2843215.03	38200.00	2840769.04
0.7	8784.18	2842642.59	8579.63	2842390.77	41074.18	2842086.19	33782.26	2843260.53
0.8	7642.02	2842669.30	7559.30	2842780.69	35791.67	2842166.02	29438.48	2843813.44
0.9	6644.54	2842701.56	6634.90	2842843.64	32183.74	2841427.87	26647.34	2843336.81
1.0	6341.40	2842608.54	6369.58	2842719.71	29193.27	2842056.57	24129.36	2842786.48

Table 4: Standard Deviation and Mean for Sum Query Under Gaussian and Laplace Mechanism

Epsilon	Our STD *Lap*	Our Mean *Lap*	DiffPriv STD *Lap*	DIffpriv Mean *Lap*	Our STD *Gau*	Our Mean *Gau*	Diffpriv STD *Gau*	Diffpriv Mean *Gau*
0.1	0.2277	38.5781	0.2213	38.5744	1.0267	38.5661	0.8134	38.5894
0.2	0.1143	38.5767	0.1156	38.5838	0.5256	38.5755	0.4007	38.5704
0.3	0.0738	38.5859	0.0741	38.5760	0.3413	38.5732	0.2708	38.5815
0.4	0.0580	38.5785	0.0555	38.5801	0.2565	38.5901	0.2034	38.5874
0.5	0.0452	38.5813	0.0442	38.5809	0.1974	38.5732	0.1693	38.5798
0.6	0.0353	38.5836	0.0360	38.5834	0.1791	38.5794	0.1423	38.5789
0.7	0.0319	38.5826	0.0312	38.5826	0.1425	38.5834	0.1259	38.5821
0.8	0.0284	38.5823	0.0281	38.5815	0.1231	38.5767	0.1035	38.5818
0.9	0.0248	38.5806	0.0260	38.5821	0.1099	38.5824	0.0960	38.5842
1.0	0.0223	38.5818	0.0212	38.5817	0.1074	38.5784	0.0875	38.5843

Table 5: Standard Deviation and Mean for Average Query Under Gaussian and Laplace Mechanism

preserving systems.

9. Prior Work

The theoretical foundations of differential privacy are well established. The concept of differential privacy was proposed by (16). The most important question related to the implementation of differential privacy, how to choose an ε is not addressed fully. There are no clearly defined rules on how to choose ε. (15) in her work mentions the choice of ε as a social question and proposed a registry where experts can expose their ε. (36) highlighted the strong theoretical guarantees offered by differential privacy and the complex practical applications, as the privacy risk depends on domain attributes such as dataset, sensitivity, and query type. They proposed selecting the value of ε based on the risk of disclosure and the adversary's ability to infer an individual's presence in the dataset. Their proposed method controls privacy risk but lacks the modelling of utility loss, in the proposed work, they only assume Laplace mechanism and mean query it also lacks evaluation on real- world dataset. (39) proposed using interval estimation; they used a confidence interval and a confidence level instead of ε, which provides a more intuitive picture of how much an individual is protected. (30) proposed an economic

method for choosing ε. Their model involves two parties: the data contributors, who seek protection, and the data analysts, who need accuracy. Their proposed model considers four parameters: accuracy, measure of accuracy, study budget and individual's expected cost if they do not participate. By plugging the values of these parameters into their model, one can determine if the study is feasible or not, and what the optimal value of ε is. proposed work focuses on the economic perspective only focusing on the losses that can be quantified monetarily, lacking empirical evaluation and limited use cases. he applicability of the proposed model is limited to the Laplace mechanism and lacks an explanation of the proposed method in terms of utility.ty. (22) explored the application of differential privacy for ECG data analysis. The authors demonstrated that careful selection of parameters is required when dealing with sensitive data. Ghazarian et al. demonstrated that the model proposed by (30) for choosing an optimal ε is not applicable in a real-world scenario. The privacy and utility requirements may vary based on the real use case. (40) proposed methods to explain ε for non-experts, two methods based on odds framing and one based on sample reports, and validated their method based on a vignette study to improve risk comprehension and data sharing decisions. The study suggests that communicating DP guarantees, including utility and privacy implications, can empower people to make more informed decisions. The proposed study explains the ε parameter in the context of the individual participating in the study.

(2) proposed GRAM DP, a basic DP framework using three libraries: IBM's $DiffPrivlib$ (28), Opendp's $smartNoise$ (38) and OpenMined $PyDP$ (23). The framework was introduced to simplify the application of DP for non-experts, allowing them to implement DP without requiring the underlying knowledge of libraries. However, the proposed framework has several limitations; the users are required to input the sensitivity manually or input 'None', prompting the framework to pick the maximum value from the desired column. This can lead to excessive noise specially if the columns contain outliers. Additionally, the user is asked to select a choice of ε, which influences the privacy vs utility tradeoff. The framework's use of broad privacy categories such as 'very high,' 'high,' 'moderate,' 'low,' and 'very low,' lacks clarity because these categories do not correspond to specific ε values, making it difficult for users to understand the actual privacy guarantees being applied. Furthermore, the proposed framework is limited only to the Laplace mechanism for adding noise, which might be useful in many scenarios but may not be suitable for all datasets or contexts. While the proposed

framework simplifies the process of selecting privacy parameters, it falls short in providing adequate explanation or guidance on the implications of the selected privacy levels, leaving the user unclear about the trade-offs.

While the theoretical guarantees are robust and well documented but a significant gap remains in translating these guarantees into effective real-world implementation. Our work addresses this gap by providing a comprehensive empirical framework that simulates and visualises the effects of DP parameters on privacy and utility trade-off's. Through extensive simulation and visualisation, we demonstrate how practitioners can interpret metrics like accuracy, relative error, entropy and bounds to make informed choices. One of the important parameters is the choice of ε, even when implementing the libraries discussed above, the issue with the libraries is that they behave as black boxes. How can we make an informed decision about the choice of ε, noise mechanism and other parameters involved in the implementation of differential privacy. We provide an empirically evaluated framework which complements existing theoretical models and addresses a notable gap in applied DP, making it more accessible, transparent and tunable for real-world applications. Our proposed framework is designed to bridge the gap between theoretical and practical DP, particularly during the design and parameter selection phase, where understanding the effect of ε, sensitivity, and choice of noise mechanism is critical. The author in (20) highlighted a few future works for privacy researchers to improve guidance on selecting ε and communicating the behaviour of the mechanism to non-experts, demonstrating how the mechanism design affects utility.

10. Conclusion

This work introduced a set of metrics and an overarching framework designed to empirically evaluate the implementation of Differential Privacy in practical applications involving real-world datasets. We introduced key evaluation metrics to guide parameter selection, visualised them through graphical representations, and validated our approach on three well-known datasets: the Adult and Billionaire datasets, and the Heart attack prediction dataset using count, sum, and average queries. We performed the experiment considering the Laplace and Gaussian mechanisms. Our results show that the selection of ε depends on the query type, the size of the dataset, and the desired privacy utility trade-off, and the context in which DP is implemented, emphasising that no single ε in the specified range of 0.1 to 1 is suitable for

all queries. The experiments also suggest that the arbitrary prior selection of parameters can not guarantee adequate results, even when using well-known libraries. Furthermore, we analysed an algorithm to compute the optimal epsilon, which optimises the privacy-utility trade-off.

Our work bridges the gap between theory and practice, guiding the non-experts to understand and implement differential privacy. Through experimental analysis, we demonstrate how a differential private framework can be implemented and evaluated while offering reasonable privacy and utility trade-off.

11. Acknowledgement

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland (grant number 18/CRT/6223). Dimitrios Vasilopoulos and Paolo Palmieri received additional support from Meta Platforms, Inc. through the Empower Science Foundation Ireland Spoke (grant number 20/SP/8955).

References

[1] Abowd, J.M.: The us census bureau adopts differential privacy. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2867–2867 (2018)

[2] Aitsam, M.: Differential privacy made easy. In: 2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE). pp. 1–7. IEEE (2022)

[3] Amiet, N.: Differential privacy: a comparison of libraries. kudelski Website (march 2020), https://research.kudelskisecurity.com/2020/03/11/differential-privacy-a-comparison-of-libraries/

[4] Apple: Learning with privacy at scale. privacy (2017), https://machinelearning.apple.com/research/learning-with-privacy-at-scale

[5] Australian Bureau of Statistics: Retail trade, australia methodology (2024), https://www.abs.gov.au/methodologies/retail-trade-australia-methodology/jun-2024#reliability-of-estimates

[6] Banerjee, S.: Heart attack dataset (2024), https://www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset

[7] Bavadekar, S., Boulanger, A., Davis, J., Desfontaines, D., Gabrilovich, E., Gadepalli, K., Ghazi, B., Griffith, T., Gupta, J., Kamath, C., et al.: Google covid-19 vaccination search insights: Anonymization process description. arXiv preprint arXiv:2107.01179 (2021)

[8] Bowen, C.M.: Utility metrics for differential privacy: No one-size-fits-all (2021), https://www.nist.gov/blogs/cybersecurity-insights/utility-metrics-differential-privacy-no-one-size-fits-all

[9] Chatzikokolakis, K., Andrés, M.E., Bordenabe, N.E., Palamidessi, C.: Broadening the scope of differential privacy using metrics. In: Privacy Enhancing Technologies - 13th International Symposium, PETS. LNCS, vol. 7981, pp. 82–102. Springer

[10] Cummings, R., Desfontaines, D., Evans, D., Geambasu, R., Jagielski, M., Huang, Y., Kairouz, P., Kamath, G., Oh, S., Ohrimenko, O., et al.: Challenges towards the next frontier in privacy. arXiv preprint **1** (2023)

[11] Cummings, R., Kaptchuk, G., Redmiles, E.M.: " i need a better description": An investigation into user expectations for differential privacy. In: Proceedings of the 2021 ACM SIGSAC. pp. 3037–3052 (2021)

[12] Desfontaines, D.: A list of real-world uses of differential privacy. https://desfontain.es/blog/real-world-differential-privacy.html (10 2021), ted is writing things

[13] Desfontaines, D.: https://desfontain.es/blog/diffprivlib.html (2025), ted is writing things

[14] Dwork, C., Roth, A.: The Algorithmic Foundations of Differential Privacy. Foundations and trends in theoretical computer science, Now Publishers (2014)

[15] Dwork, C., Kohli, N., Mulligan, D.: Differential privacy in practice: Expose your epsilons! Journal of Privacy and Confidentiality **9**(2) (2019)

[16] Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography:

Third Theory of Cryptography Conference, Proceedings 3. pp. 265–284. Springer (2006)

[17] Dwork, C., McSherry, F., Nissim, K., Smith, A.D.: Calibrating noise to sensitivity in private data analysis. In: Theory of Cryptography, TCC, Proceedings. LNCS, vol. 3876, pp. 265–284. Springer (2006)

[18] Gaboardi, M., Hay, M., Vadhan, S.: A programming framework for opendp. Manuscript, May (2020)

[19] Garfinkel, S.: Differential privacy and the 2020 us census (2022)

[20] Garrido, G.M., Liu, X., Matthes, F., Song, D.: Lessons learned: Surveying the practicality of differential privacy in the industry. arXiv preprint arXiv:2211.03898 (2022)

[21] Gary Howarth, e.a.: A community-driven differential privacy deployment registry. NIST Internal Report 8588 (Initial Public Draft) (2025), https://csrc.nist.gov/pubs/ir/8588/ipd

[22] Ghazarian, A., Zheng, J., Rakovski, C.: Privacy-preserving ecg data analysis with differential privacy: A literature review and a case study. arXiv preprint arXiv:2406.13880 (2024)

[23] Google: Pydp (2023), https://pydp.readthedocs.io/en/latest/introduction.html

[24] Guevara, M.: Enabling developers and organizations to use differential privacy (2019)

[25] Haeberlen, A., Pierce, B.C., Narayan, A.: Differential privacy under fire. In: Proceedings of the 20th USENIX Conference on Security. p. 33. SEC'11 (2011)

[26] Harvard: Differential privacy, harvard university privacy tools project (2024), https://privacytools.seas.harvard.edu/differential-privacy

[27] Hawes, M.: Differential privacy and the 2020 decennial census. In: APHA's 2020 VIRTUAL Annual Meeting and Expo APHA (2020)

[28] Holohan, N., Braghin, S., Mac Aonghusa, P., Levacher, K.: Diffprivlib: the IBM differential privacy library (2019)

[29] Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B.C., Roth, A.: Differential privacy: An economic method for choosing epsilon. In: 2014 IEEE 27th Computer Security Foundations Symposium. pp. 398–410 (2014)

[30] Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B.C., Roth, A.: Differential privacy: An economic method for choosing epsilon. In: 2014 IEEE 27th Computer Security Foundations Symposium. pp. 398–410. IEEE (2014)

[31] Joseph Near, David Darais, K.B.: Differential privacy for privacy-preserving data analysis: An introduction to our blog series (2020), https://rb.gy/p0ecam

[32] Kaggle: Adult income dataset (Nov 2016), https://www.kaggle.com/datasets/wenruliu/adult-income-dataset

[33] Kaggle: Billionaires stats dataset (Nov 2016), https://www.kaggle.com/code/nelgiriyewithana/an-introduction-to-billionaires-statistics-dataset

[34] Kifer, D., Machanavajjhala, A.: No free lunch in data privacy. In: Proceedings of the 2011 ACM SIGMOD. pp. 193–204 (2011)

[35] Laud, P., Pankova, A., Pettai, M.: A framework of metrics for differential privacy from local sensitivity. Proc. Priv. Enhancing Technol. **2020**(2), 175–208 (2020)

[36] Lee, J., Clifton, C.: How much is enough? choosing ε for differential privacy. In: Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings 14. pp. 325–340. Springer (2011)

[37] Messing, S., DeGregorio, C., Hillenbrand, B., King, G., Mahanti, S., Mukerjee, Z., Nayak, C., Persily, N., State, B., Wilkins, A.: Facebook privacy-protected full urls data set. Version DRAFT VERSION (2020)

[38] Microsoft: Opendp (2023), https://github.com/opendp

[39] Naldi, M., D'Acquisto, G.: Differential privacy: An estimation theory-based method for choosing epsilon. CoRR **abs/1510.00917** (2015), http://arxiv.org/abs/1510.00917

[40] Nanayakkara, P., Smart, M.A., Cummings, R., Kaptchuk, G., Redmiles, E.M.: What are the chances? explaining the epsilon parameter in differential privacy. In: 32nd USENIX Security Symposium. pp. 1613–1630 (2023)

[41] Near, J.P., Abuah, C.: Programming Differential Privacy, vol. 1 (2021)

[42] Near, J.P., Darais, D., Lefkovitz, N., Howarth, G., et al.: Guidelines for evaluating differential privacy guarantees. Tech. rep., National Institute of Standards and Technology (2023)

[43] Ngong, I.C., Stenger, B., Near, J.P., Feng, Y.: Evaluating the usability of differential privacy tools with data practitioners. In: Twentieth Symposium on Usable Privacy and Security (SOUPS 2024). pp. 21–40 (2024)

[44] Pujol, D., Desfontaines, D.: Open problem - better privacy guarantees for larger groups. DifferentialPrivacy.org (2023), https://differentialprivacy.org/open-problem-better-privacy-guarantees-for-larger-groups/

[45] Rogers, R., Cardoso, A.R., Mancuhan, K., Kaura, A., Gahlawat, N., Jain, N., Ko, P., Ahammad, P.: A members first approach to enabling linkedin's labor market insights at scale. arXiv:2010.13981 (2020)

[46] som: Differential privacy while sharing sensitive data, https://rb.gy/2llksg

[47] Song, P., Sarathy, J., Shoemate, M., Vadhan, S.: " i inherently just trust that it works": Investigating mental models of open-source libraries for differential privacy. Proceedings of the ACM on Human-Computer Interaction **8**(CSCW2), 1–39 (2024)

[48] Song, P., Sarathy, J., Shoemate, M., Vadhan, S.: "i inherently just trust that it works": Investigating mental models of open-source libraries for differential privacy. Proc. ACM Hum.-Comput. Interact. **8**(CSCW2)

(Nov 2024). https://doi.org/10.1145/3687011, https://doi.org/10.1145/3687011

[49] Xiong, A., Wu, C., Wang, T., Proctor, R.W., Blocki, J., Li, N., Jha, S.: Using illustrations to communicate differential privacy trust models: an investigation of users' comprehension, perception, and data sharing decision. arXiv:2202.10014 (2022)

[50] Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., Nguyen, J., Ghosh, S., Bharadwaj, A., Zhao, J., Cormode, G., Mironov, I.: Opacus: User-friendly differential privacy library in pytorch. CoRR **abs/2109.12298** (2021)