An Empirical Approach to Protect Data with Differential Privacy

Prepared By:

Shivansh Srivastava (2018BCS-053)

@sastava007

Overview

- The amount of sensitive data that is being digitally recorded is increasing rapidly. This new data collection creates ever increasing ways to violate privacy.
- But it also creates exciting opportunities to make better analysis and improve user experience, if made available to the right data scientists and researchers.
- There is a natural tension between protecting the privacy of individuals and obtaining better analytical results. Differentially private algorithms are a promising technical solution that could ease this tension, allowing analysts to perform aggregate analysis while guaranteeing meaningful protection of each individual’s privacy.

Need of Anonymization

- **Netflix** in 2006, announced a 1 million dollar challenge for improving their recommendation engine, together with releasing 100 million anonymized movie ratings.

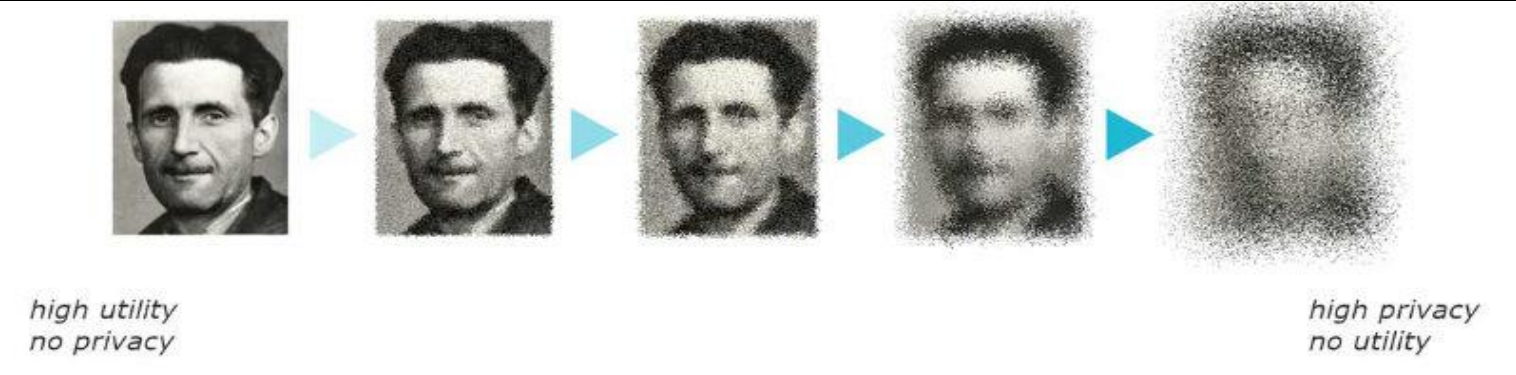
- Although the data sets were constructed to preserve customer privacy, two researchers from The University of Texas, Austin were able to identify individual users by matching the data sets with film ratings on the IMDb.

NETFLIX

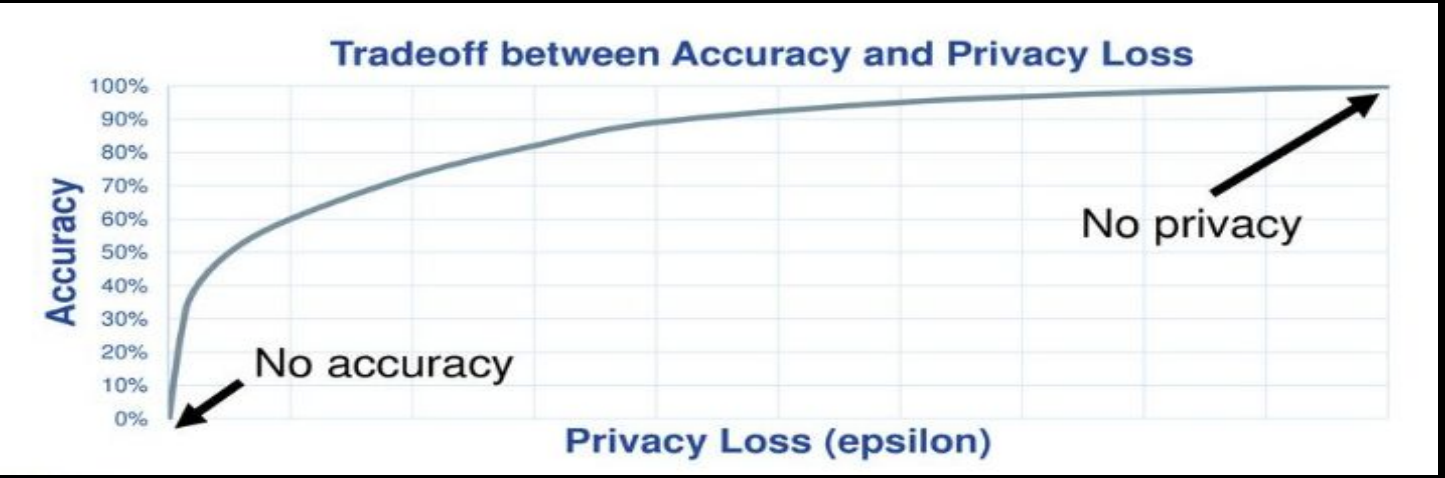
Differential Privacy

- Differential privacy is a data anonymization technique that allows the researchers to build accurate models without sacrificing the privacy of individual subjects by introducing randomness into the process of data retrieval.
- Differential Privacy is not just an algorithm, but rather “an entire scientific field at the intersection of computer science, statistics, economics, law and ethics”.
- It guarantees the that the attacker can learn virtually nothing more about an individual than they would learn if that person’s record were absent from the dataset.

Privacy v/s Accuracy



Data Privacy is about proper usage, collection, retention, deletion, and storage of data. **Data Accuracy** is all about how true and accurate the data is.



Case Study

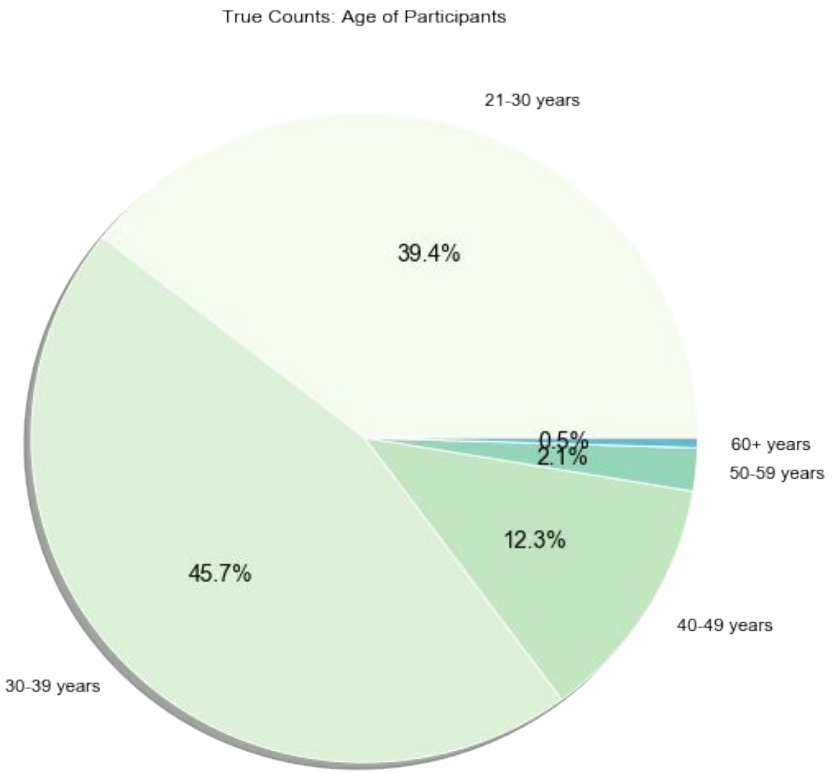
An analysis of Mental Health in Tech Survey Data with/without preserving data privacy

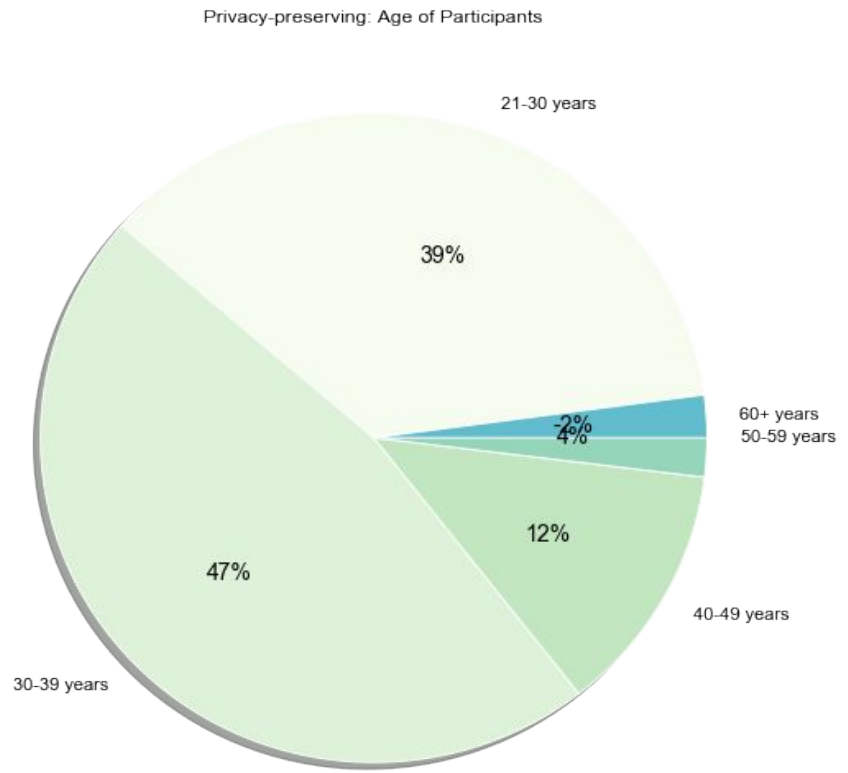
- The purpose of this demo is to showcase the utility of OpenDP differential privacy framework by making statistical queries to data with and without privacy-preserving mechanisms.
- This data set is released by OSMI which consists of 27 questions, answered by 1,259 volunteers including attributes like (age, gender, country, remote_work, family_history and treatment).
- After comparing the query results side-by-side, we reached to the conclusions about the data are almost similar in both the settings.

Now, we will make statistical queries on different variables to generate a comparative analysis of results obtained from data with & without differential privacy mechanisms.

1: Age

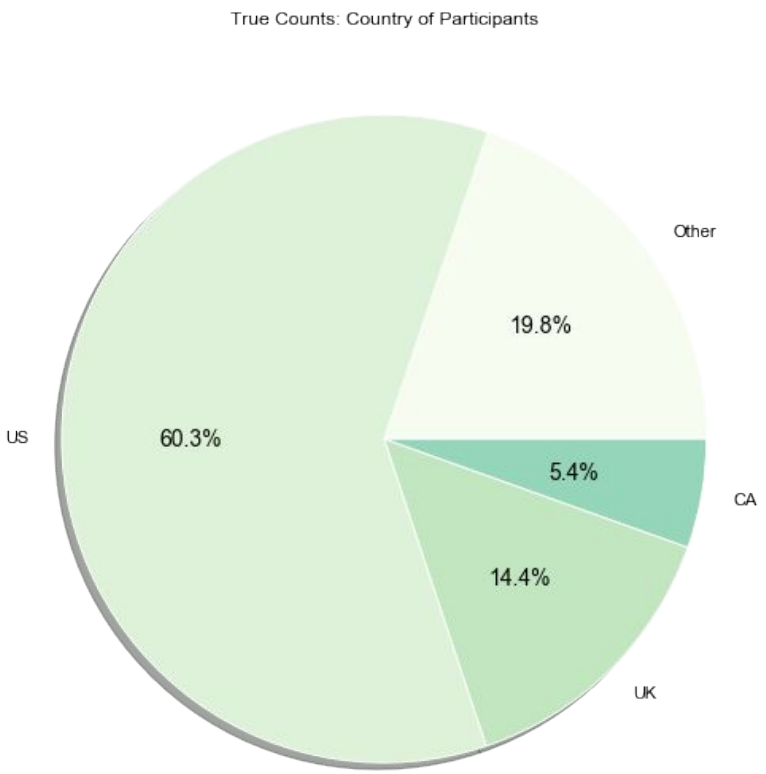
- The true age distribution is: **[478, 554, 149, 26, 6]**
- The age histogram obtained using DP is: **[458, 555, 148, 48, 25]**

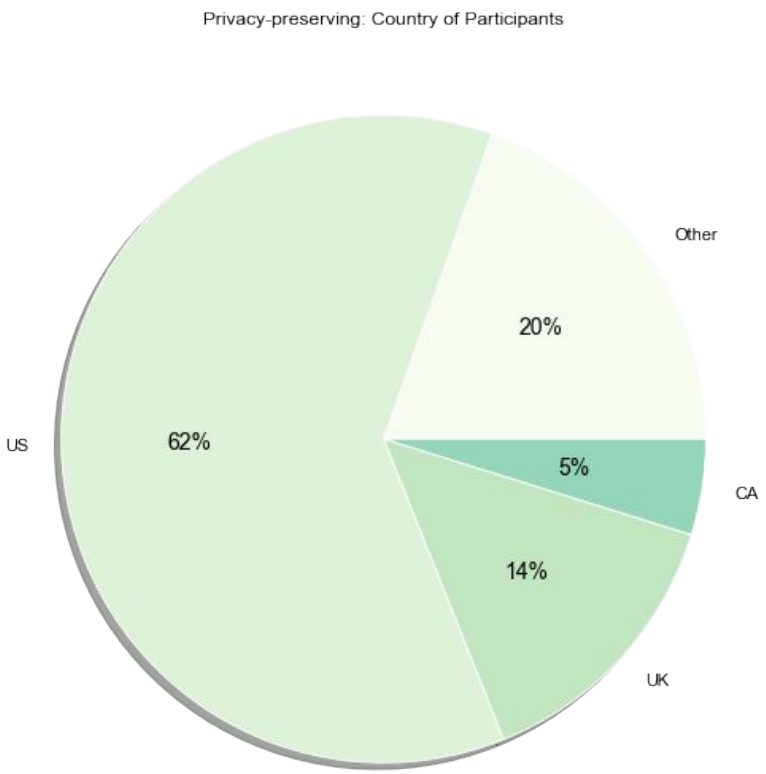




2: Country

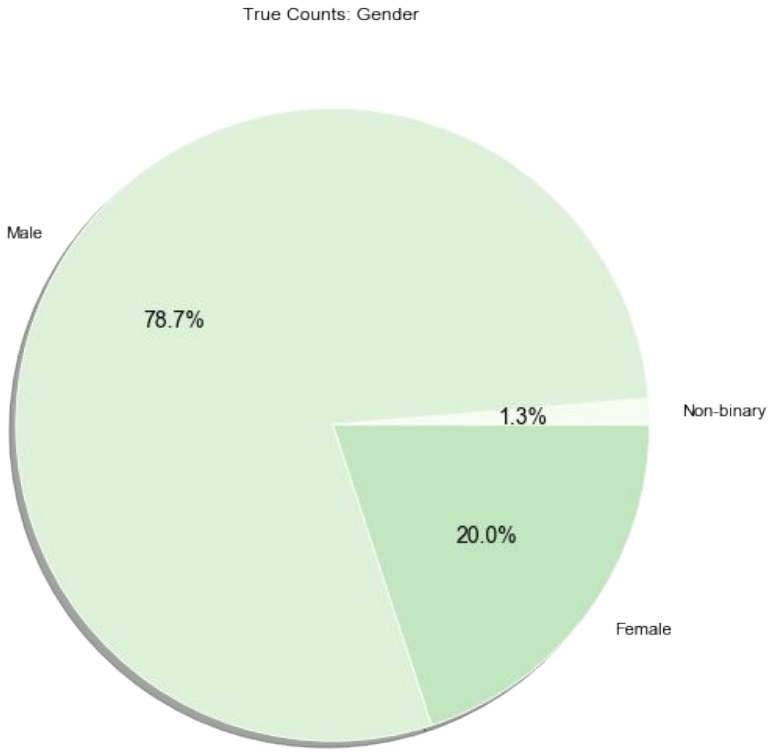
- The true country distribution is: **[240 732 175 66]**
- The country histogram obtained using DP is: **[257 810 186 63]**

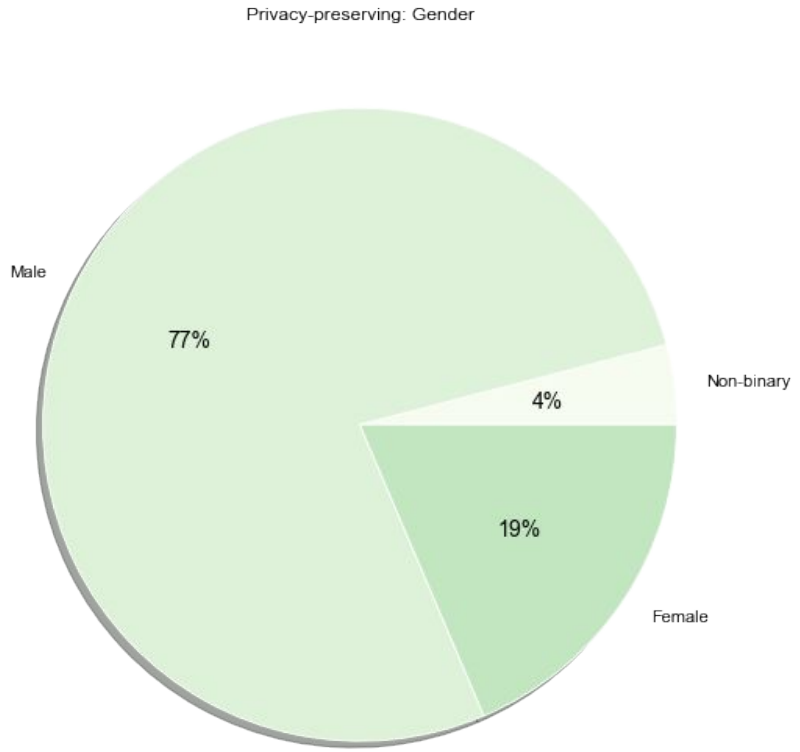




3: Gender

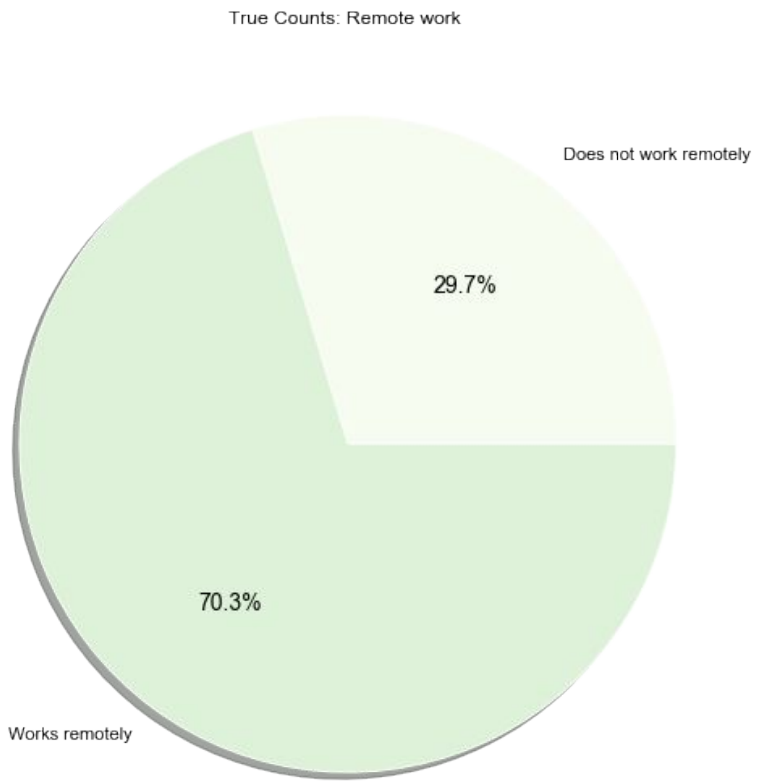
- The true gender distribution is: **[16 955 242]**
- The gender histogram obtained using DP is: **[52 990 238]**

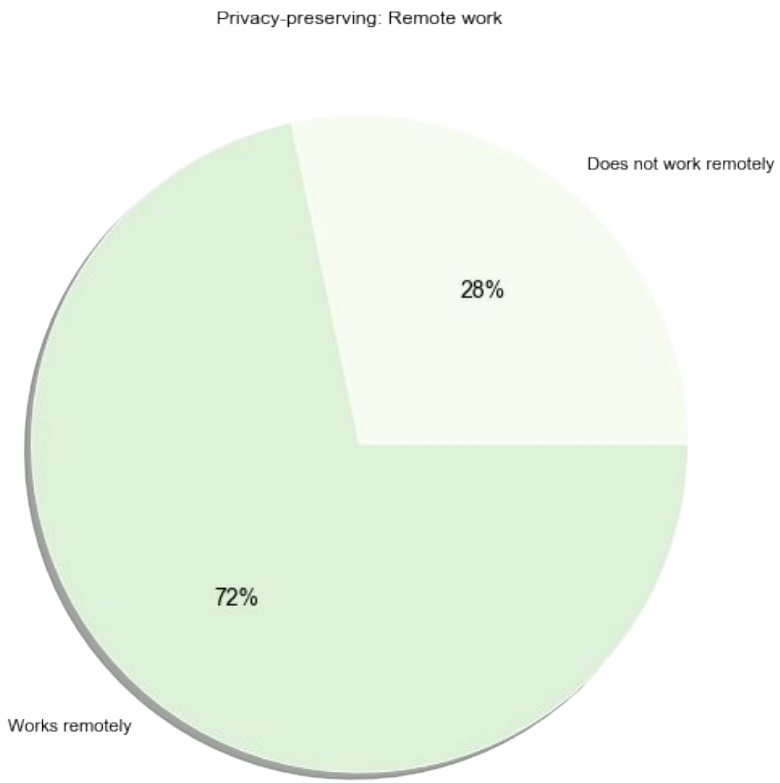




4: Remote Work

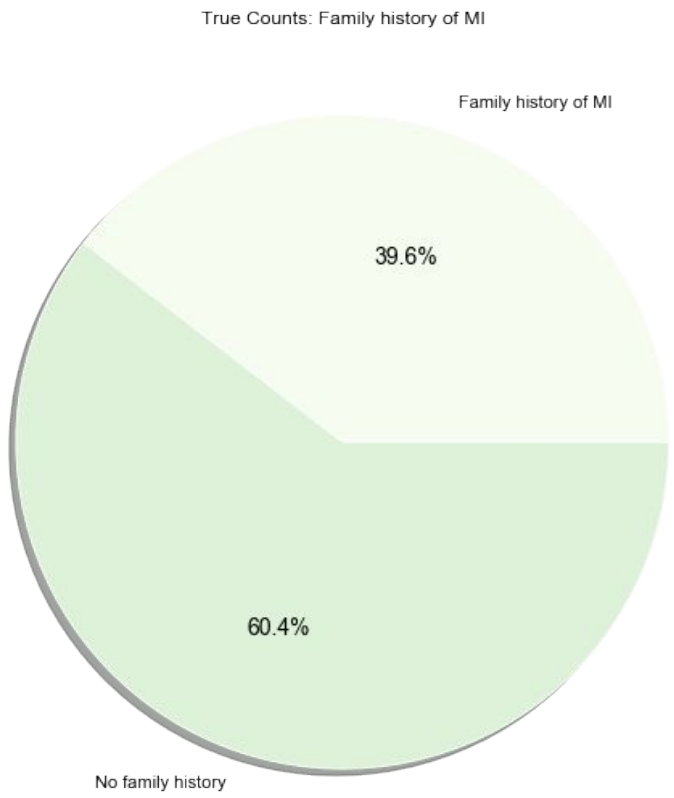
- The true remote work distribution is: **[360 853]**
- The remote work histogram obtained using DP is: **[355 898]**

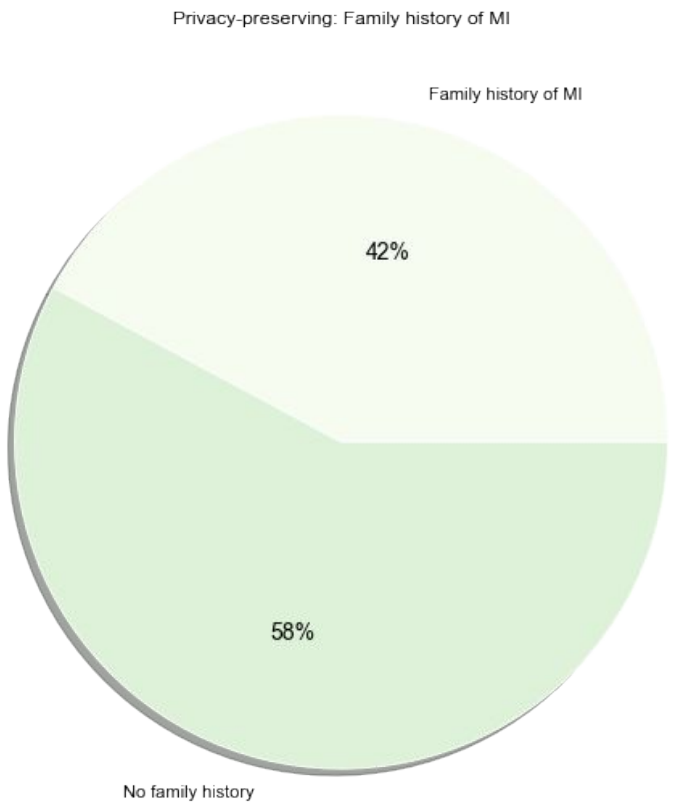




5: Family History

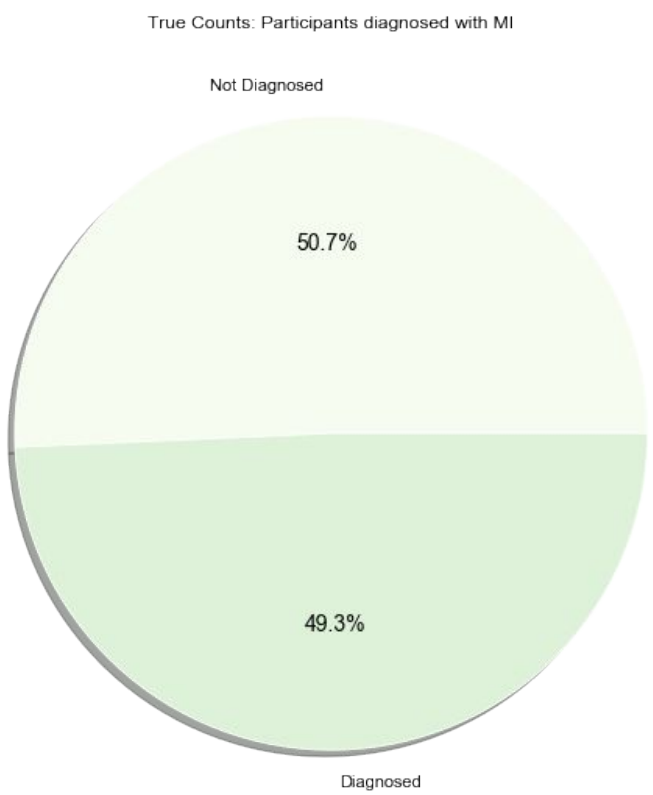
- The true count of participants with a family history of mental illness is: **[480 733]**
- The family history histogram obtained using DP is: **[517 709]**

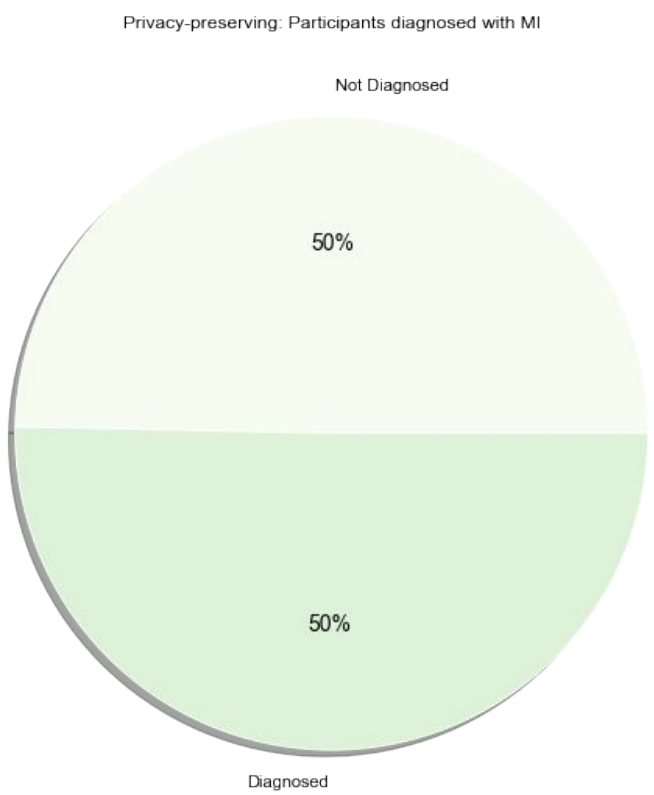




6: Treatment

- The true count of participants diagnosed with mental illness is: **[619 598]**
- The mental illness obtained using DP is: **[604 612]**





Usefulness of Differential Privacy in handling the attack on an individual’s data

- In this demo, we will examine perhaps the simplest possible attack on an individual's private data and what the differential privacy can do to mitigate it.

- We are considering a dataset of 10,000 people having attributes like (name, sex, age, education, income, married, race).

```
person of interest:

sex           0.0
age          45.0
educ          6.0
income     6000.0
married       1.0
race          1.0
Name: 0, dtype: float64
```

- Consider an attacker who knows everything about the data except for the person of interest's (POI) income, which is considered private.

  They can back out the individual's income very easily, just by asking for the mean overall mean income.

```
POI_income = overall_mean * n_obs - known_mean * known_obs
```

- But if the attackers were made to interact with the data through differential privacy and was given a privacy budget of ε = 1.

  Upon executing the code in the attached Jupyter notebook the result we got are as follows:

```
Known Mean Income: 26886.001600160016
Observed Mean Income: 26883.930944271226
Estimated POI Income: 6179.4427122677835
True POI Income: 6000.0
```

Conclusion

- We've reviewed the theory of differential privacy and seen how it can be used to quantify privacy.
- Through the first simulation on mental health data, we've showed that DP guarantees that anyone seeing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis.
- In the second case study of finding the income of POI, we've seen that DP provides a provable guarantee of privacy protection against a wide range of privacy attacks include (differencing attacks, linkage attacks, and reconstruction attacks).

Thank You!