Data Privacy Challenge

Unmasking the Private: Adversarial Attacks on Differential Privacy

In our world, where data is used increasingly, it is essential to protect people's private information. Privacy protection methods are designed to hide individual details and keep sensitive information in large collections of data safe. These methods try to preserve useful patterns that allow researchers to study the data without compromising privacy. However, ensuring that these privacy protections are strong enough to withstand skilled attackers is still a major problem.

The data

In this competition, participants will try to find ways to break the privacy protections of a dataset. You will be given two versions of the data:

- **The Original Data**: A realistic collection of information that includes sensitive details.
- **Protected Data**: The same collection of information, but the data has been altered to hide private details, either by hiding some of the information or slightly altering them. You will not know exactly how those protections were applied.

The data contains details of the income and expenditure of 20,000 individuals. Each row is identified by *Name* and an *Identifier* - which are encrypted in the protected data. Several columns contain information characterizing the individual: their *Age*, *Occupation*, *City_Tier*, number of *Dependents*. Finally, their total monthly *Income* is provided along with a breakup of the individual's expenses: *Rent*, *Loan_Repayment*, *Insurance*, *Groceries*, etc. You may expect all or some of this information to be subtly altered.

Your challenge

1. Your task is to try and figure out the hidden private information in the protected data, by any method you choose to employ. You can use the original data as a reference to help you, along with any research you undertake.
2. You must try to match each row in the private data to the corresponding row in the original data. Each row you correctly identify scores you a point.
3. Finally, you must explain your methods, your results, and your analysis of the privacy problems clearly and simply.

Evaluation Criteria

- **Effectiveness of the Attack**: The accuracy and extent to which sensitive information is recovered.
- **Novelty and Creativity**: The originality and innovation of the attack methodologies.
- **Analysis and Insight**: The depth and clarity of the analysis of the privacy vulnerabilities.
- **Clarity and Presentation**: The quality of the presentation of the findings and methodology.