**5. EXPERIMENTS**

Here Age, Gender and Zip code are the Quasi Identifiers (QI).

5 different conditions are used for plotting multiple line graphs.

**1) Marital Status:** **2) Marital Semantic Tree (One):**

Primary Sensitive - Marital Status Primary Sensitive - Marital Parent

Secondary Sensitive - Education Secondary Sensitive - Education

Tertiary Sensitive - Employment Tertiary Sensitive - Employment

Quaternary Sensitive - Race Quaternary Sensitive - Race

**3) Marital Semantic Tree (Two):** **4) Relationship:**

Primary Sensitive - Marital Parent Primary Sensitive - Relationship

Secondary Sensitive - Education Secondary Sensitive - Education

Tertiary Sensitive - Employment Tertiary Sensitive - Employment

Quaternary Sensitive - Race Quaternary Sensitive - Race

**5) Disease Semantic Tree:**

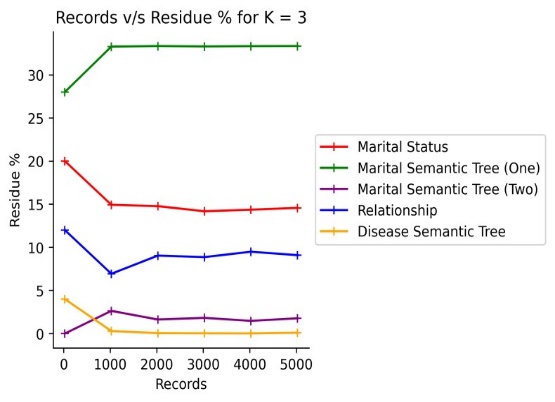
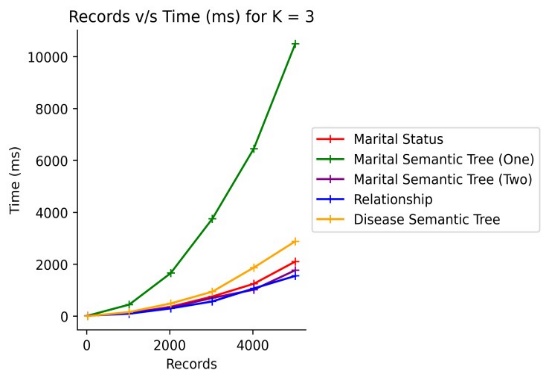
Primary Sensitive - Disease Parent

Secondary Sensitive - Education

Tertiary Sensitive - Employment

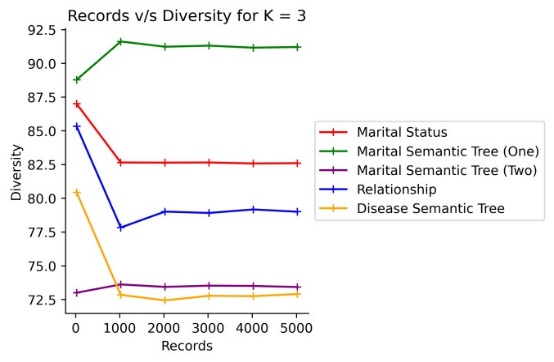
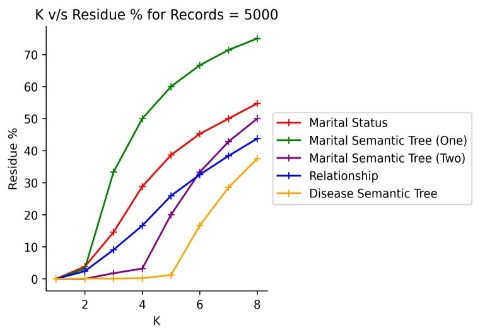
Quaternary Sensitive - Race

Numerous graphs for various parameters and constant conditions are plotted

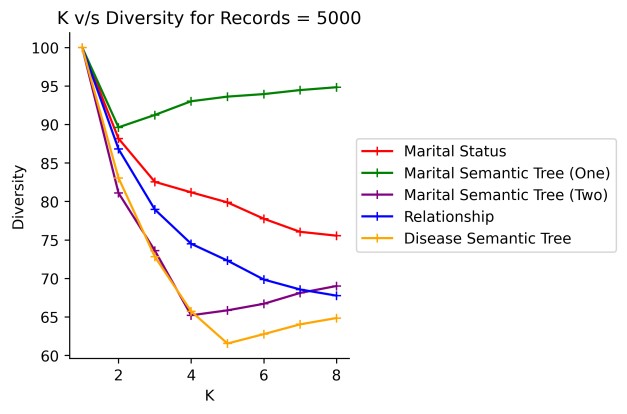
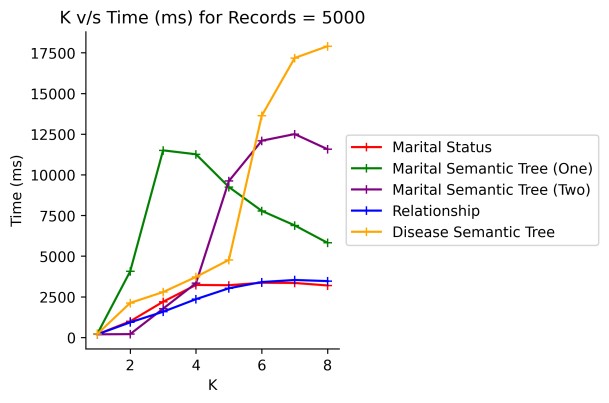
**Fig 5.1 Fig 5.2**

**Records v/s Residue % for K=3 Records v/s Time (ms) for K=3**

**Fig 5.3 Fig 5.4**

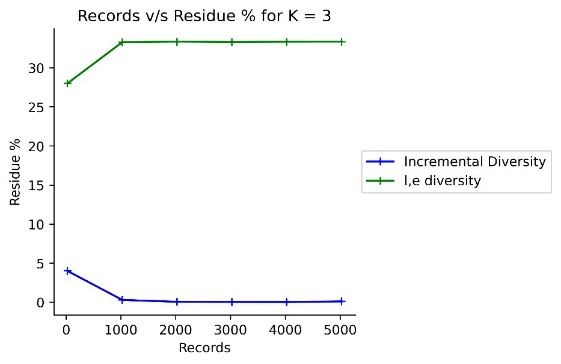
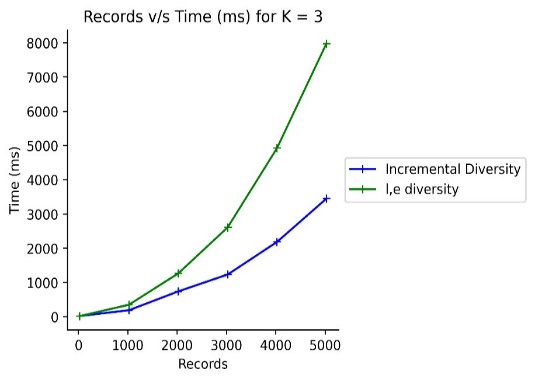
**Records v/s Diversity for K=3 K v/s Residue % for Records=5000**



**Fig 5.5 Fig 5.6**

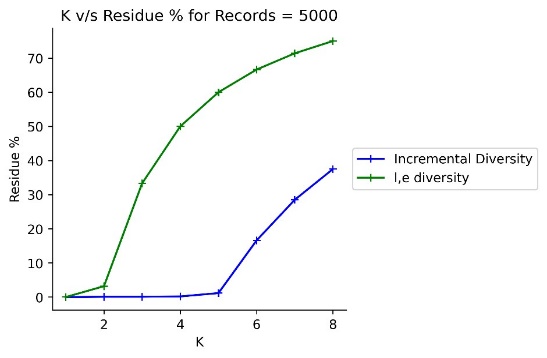
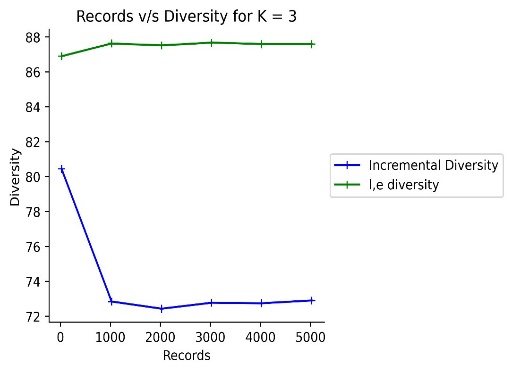
**K v/s Time (ms) for Records=5000 K v/s Diversity for Records=5000**

Comparison of l, e Diversity and Incremental Diversity: -



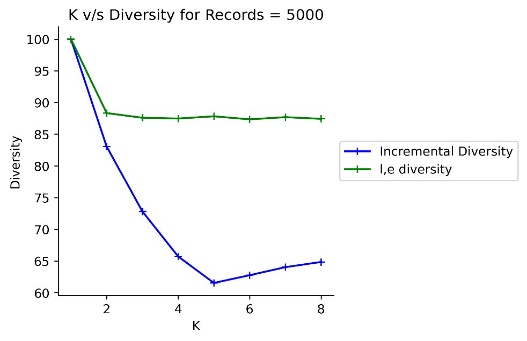
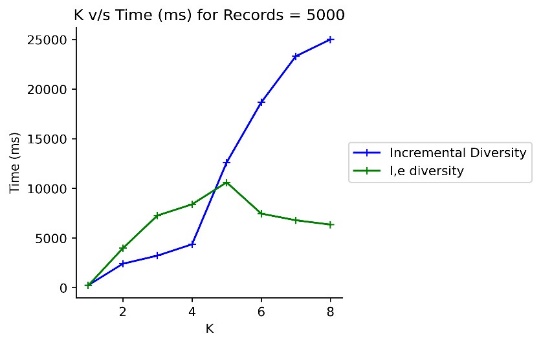
**Fig 5.7 Fig 5.8**

**Records v/s Time (ms) for K=3 Records v/s Residue % for K=3**



**Fig 5.9 Fig 5.10**

**Records v/s Diversity for K=3 K v/s Residue % for Records=5000**



**Fig 5.11 Fig 5.12**

**K v/s Time (ms) for Records=5000 K v/s Diversity for Records=5000**

**Software Tools used are: -**

1) Python Language

2) CSV Files Contains Dataset

3) Math Module Mathematical Operations

4) Time Module Time taken for Execution

5) Matplotlib.pyplot Module Plot Graphs

6) Tabulate Module Tabulate Function

7) Copy Module Deepcopy() function

**6. RESULTS AND DISCUSSIONS**

**1. No. of Records v/s Residue %:** From Fig 5.1, the best possible case for arrangement of records will be based on unique Diseases (condition 5) in each equivalence group as the residue % for it is very less for larger records since number of unique Diseases are more than the number of unique Marital Parents.

**2. No. of Records v/s Time (ms):** In Fig 5.2, since the uniqueness of the Marital Parent is very low, large number of residues will be obtained in the residue list. Therefore, more time is spent iterating through the residue list multiple times to replace records.

**3. No. of Records v/s Diversity:** In Fig 5.3, for condition 2, the diversity increases as the number of records increases because most of the records will be sent to residue list hence less repetition in the modified microdata table and more diversity.

**4. K v/s Residue %:** In Fig 5.4, there is a steady increase in residue % for all conditions because when the K value increases there will be more repetition which leads to higher residue %.

**5. K v/s Time (ms):** From Fig 5.5, the time taken for execution increases as the value of K increases, since the program iterates through larger residuals and then swaps the datasets leading to increase in the time taken for execution.

**6. K v/s Diversity:** In Fig 5.6, the diversity is very high for smaller values of K because the no. of unique values is more relative to K values, so less repetition.

**Comparison between l, e diversity algorithm and incremental diversity algorithm**

l, e diversity algorithm employs the use of a primary sensitive attribute with lesser no. of parents in its semantic hierarchical tree (Eg: Marital Status) having one common parent for the sensitive value in the equivalence class.

Incremental diversity algorithm makes use of a primary sensitive attribute with more no. of parents in its semantic hierarchical tree (Eg: Disease) allowing only non-repeating parents to be present in each equivalence class while also performing the incremental diversification for secondary, tertiary and quaternary sensitive attributes. It outperforms l, e diversity in terms of faster time performance and overall decrease in residue records percentage. From the Fig 5.7 and Fig 5.8 we can infer that incremental diversity produces lesser residue records and runs faster in spite of performing diversification for multiple sensitive attributes. Fig 5.9, producing lesser residue records is a double-edged sword which leads to lesser diversity in the records present in the equivalence classes of the table which is a drawback of the incremental diversity algorithm as compared to l, e diversity algorithm.

**8. Conclusion**

Choosing Primary Sensitive Attribute**:** The choice of the right primary sensitive attribute can make or break the diversification and privacy of a table. There will be a trade-off between residue records produced and diversity of the final private table. For example, Disease as the primary sensitive attribute (algorithm 5) produces lesser no. of residue records but at the same time diversity of the records is hampered. On the other hand, Marital Status as the primary sensitive attribute (algorithm 1) produces highly diverse table but it comes at the cost of more no. of residues produced and the loss of precious data. It depends on the use case as to whether diversity needs to be sacrificed in order to produce lesser residue records and prevent huge data loss or, the diversity of the data is of utmost importance and it is okay to overlook the production of enormous residue records.

**REFERENCES**

1. Boyu Li, Yanheng Liu, Minghai Wang, Geng Sun and Bin Li. Local anatomy for personalized privacy protection, 2021.
2. Widodo, Murien Nugraheni and Irma Permata Sari. Simple Distribution of sensitive values for multiple sensitive attributes in privacy preserving data publishing to achieve anatomy, 2021.
3. Khan Razaullah, Tao Xiaofeng, Anjum Adeel, Haider Sajjad and Saif ur Rehman Malik. Privacy Preserving for Multiple Sensitive Attributes against Fingerprint Correlation Attack Satisfying c-Diversity, 2020.
4. Yuelei Xiao and Haiqi Li. Privacy Preserving Data Publishing for Multiple Sensitive Attributes Based on Security Level, 2020.
5. Jassma N Vanasiwala and Nirali R Nanavati. Privacy Preserving Data Publishing of multiple Sensitive Attributes by using Various Anonymization Techniques, 2020.
6. W Widodo and A Wahyudin, A preliminary phase on anatomyzing multiple sensitive attribute by determining main sensitive attribute, 2020.
7. Widodo Widodo, Eko K. Budiardju, Wahyu Catur Wibowo and Harry Achson. An approach for distributing sensitive values in k – anonymity, 2019.
8. Rong Wang, Yan Zhu, Tung-Shou Chen and Chin-Chen Chang. Privacy-Preserving Algorithms for Multiple Sensitive Attributes Satisfying t-Closeness, 2018.
9. Ozgu Can. Personalised Anonymity for micro data release, 2018.
10. Edwin M. Knorr, Giulio Valentino Dalla Riva and Orlin Vakarelov. Anatomy of a new data science course in privacy, ethics and security, 2018.
11. Widodo and Wahyu Catur Wibowo. A distributional model of sensitive values on p – sensitive in multiple sensitive attributes, 2018.
12. Wantong Zheng, Zhongyue Wang, Tongtong Lv and Yong Ma. K – Anonymity Algorithm based on improved clustering, 2018.
13. Junjie Jia and Luting Chen. (l,m,d) – Anonymity : A Resisting similarity attack model for multiple sensitive attributes, 2017.
14. Haiyuan Wang, Jianmin Han and Jivi Wang. (l,e) – diversity A privacy preserving model to resist semantic similarity attack, 2014.
15. Hua Jin, Shang – Cheng Liu and Shi-Guang Ju. Privacy preserving for multiple sensitive attributes based on l – coverage, 2014.
16. Michael Kern. Anonymity: A Formalization of privacy – l – diversity, 2013.
17. Sunyong Yoo, Moonshik Shin and Doheon Lee. An Approach to Reducing Information Loss and Achieving Diversity of Sensitive Attributes in k-anonymity Methods, 2012.
18. Xiang Min Ren, Jing Yang, Jian Pei Zhang and Zong Fu Jia. Uncertain Data Privacy Protection Based on K-Anonymity via Anatomy, 2012.
19. Vijayarani Mohan. K-Anonymity Techniques – A Review, 2010.
20. Khaled El Emam and Fida Kamal Dankar. Protecting Privacy Using k-Anonymity, 2008.
21. Alina Campan and T.M. Truta. Data and Structural k-Anonymity in Social Networks, 2008.
22. Ashwin Machanavajjhala, Daniel Kifer and Johannes Gehrke. L -diversity, 2007.
23. V. Ciriani, S. De Capitani di Vimercati, S. Foresti and P. Samarati. k-Anonymity, 2007.
24. Xioakui Xiao and Yufei Tao. Anatomy: Privacy and Correlation Preserving Publication, 2006.
25. T.M. Truta and Bindu Vinay. Privacy Protection: p-Sensitive k-Anonymity Property, 2006.
26. M. Nergiz and Chris Clifton. Thoughts on k-Anonymization, 2006.
27. Xioakui Xiao and Yufei Tao. Anatomy: Simple and Effective Privacy Preservation, 2006.
28. Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer and Muthuramakrishnan Venkitasubramaniam. ℓ-Diversity: Privacy Beyond k-Anonymity, 2005.
29. Wei Jiang and Chris Clifton. Privacy-Preserving Distributed k-Anonymity, 2005.
30. L. Sweeney. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression, 2002.
31. Latanya Sweeney. k- anonymity: A model for protecting privacy, 2001.