GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)
ગુજરાતટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંકઃ ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Annexure 1

Enrollment no: 190130111081

STUDENT'S WEEKLY RECORD OF INTERNSHIP

NAME OF STUDENT: Modi Shreshtha Pragnesh

DIARY OF THE WEEK: Dt: _____ **TO** _____

DEPARTMENT: Electronics And Communications Engineering **SEM:** 08

NAME OF THE ORGANISATION: Eternal Soft Solutions

NAME OF THE PLANT/SECTION/DEPARTMENT: Software and Cloud Engineering

NAME OF OFFICER INCHARGE OF THE PLANT/SECTION/DEPARTMENT: Mrs Poonam Patel

DESCRIPTION OF THE WORK DONE IN BRIEF

This week was more project focused as I was focused on creating the dummy dataset which has the same structure of the actual dataset. Fake dummy data can be created using a python library called Faker. I installed the faker library by using pip and then got started creating the dummy data. The dummy data here is called 'customer_data' which is the data collected from a hypothetical e-commerce website. The dataframe contains the columns Name Which has the name of the customers which will be a sensitive column, Gender column has genders of the customers which is insensitive which is not important as it cannot be traced back to the users, Age column contains the ages of the consumers which is a quasidentifying column as it can be used in conjuction with other columns or datasets to trace the data back to the user, Email column is a sensitive column as it can easily be used to reverse search and find not only the name of the user but a lot of other sensitive information of the user, Amount Spent is again a quasiidentifying column and zipcode is a quasidentifying column as well.

I wanted to set the hierarchies manually as arx has limited support for hierarchies. I then spent a day or two setting up and defining the hierarchies for Amount spent , zipcode and Age.
I divided all of the values of amount spent in four columns such as high, low, medium and low-hig for the first level, i generalized the columns even further.
For zipcode, i scraped the zipcodes for a city from the web and got a list of the areas. I then set up the hierarchy to generalize the zipcode values by zipcode-> area->city->state->country->*
I used the interval based hierarchy in pyarxaas to set up a interval based hierarchy for age. Then i used those csv files to set the generalize the privacy levels for the database.

I experimented with and used privacy models kanonymity, tcloseness and ldiversity to anonymize the customer data and came to the conclusion that identifying columns are better anonymized with tcloseness and ldiversity the rest of the quasiidentifying columns. Then performed the risk analysis for the same dataset

GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)

ગુજરાતટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંકઃ ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

TOTAL HOURS: 35

SIGNATURE OF STUDENT

✪ **The above entries are correct and the grading of work done by Trainee is**
EXCELLENT / VERY GOOD / GOOD / FAIR / BELOW AVERAGE / POOR

Signature of Faculty Mentor

Signature of officer-in-charge of Dept. / Section / Plant

Date:

Date:

✪ **Grading of Work, for trainee may be given depending upon your judgement about his Punctuality, Regularity, Sincerity, Interest taken, Work done etc.**

SUPPLEMENTRY NOTES
(add additional sheets if required)