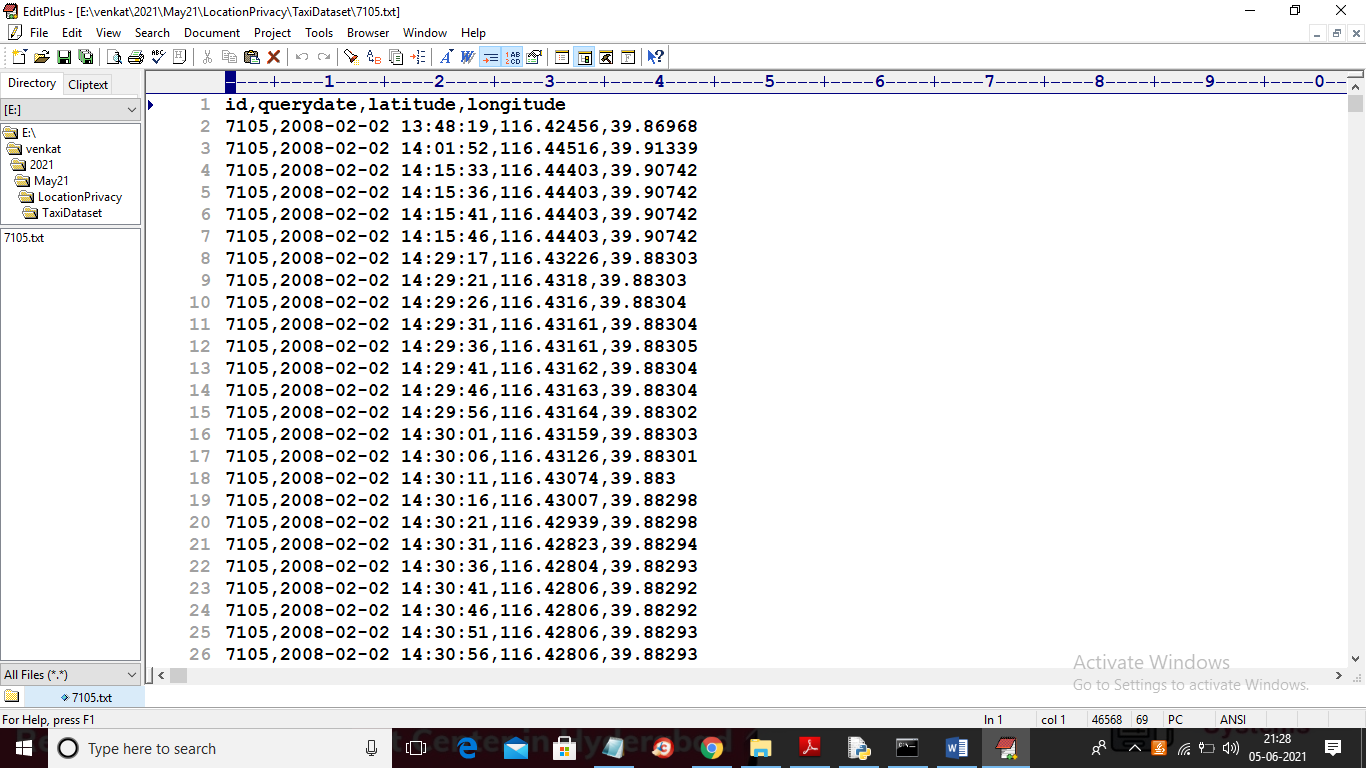
Privacy Preserving Location Data Publishing: A Machine Learning Approach

Now-a-days due to mobile all online applications are recording user locations and then storing them in their app and this location details they can use to track users. Sometime some malicious users can track the user location by knowing their home address and then they can match the home location with other location details to know where user is travelling like bank, hospital or any other locations. To overcome from this problem and to provide security to user location many data anonymization techniques such as K-Anonymity and data perturbation was introduce where data perturbation will add noise to user data and K-Anonymity will adjust data into groups so user location cannot be identified.

But above techniques are not reliable as malicious users can identify how to crack groups and noise data to know user location. To overcome from this problem author has introduce Machine Learning based data privacy preserving technique which consists of 3 models and this 3 models will provide more security and anonymize or generalized which cannot be easily understand or crack.

1. Clustering model: in this model user locations will be clusters by using KMEANS algorithm and then calculate loss value. Loss value indicates difference between correct value and predicted value and the lesser the loss the better is the algorithm. The loss value will be saved to compare with Dynamic Sequence Alignment Loss and this Dynamic Sequence is called as Heuristic Clustering Algorithm.
2. Dynamic Sequence Alignment: In this module or algorithm we will take location form cluster member and then take random locations from original dataset and both this records will be aligned to get location which has minimal loss.
3. Data Generalization: in this module user location will be generalized or anonymised by summing up location with loss values.

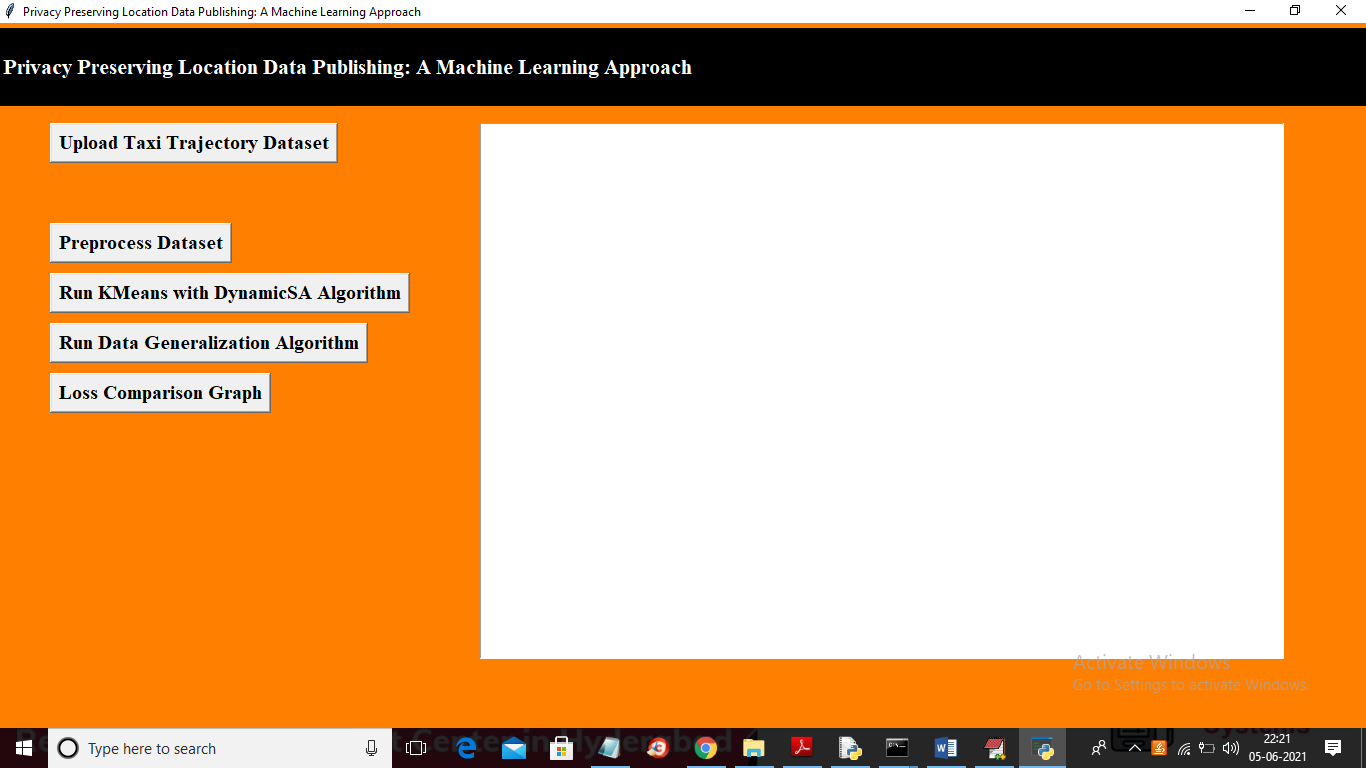
To implement this project we have used Taxi Trajectory (TDRIVE) dataset and below is the dataset details



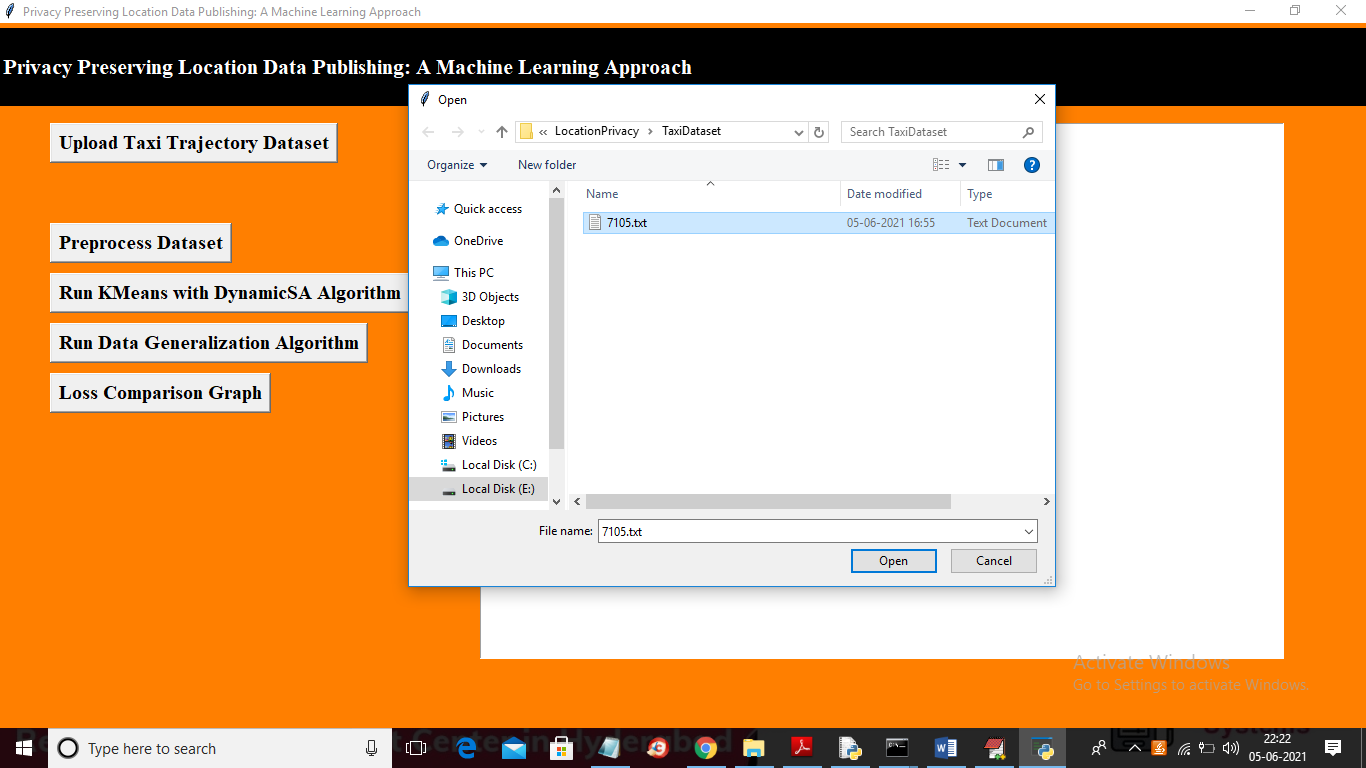
In above screen first rows contains dataset column names and remaining rows contains dataset values and in each row represents users current location latitude and longitude values.

SCREEN SHOTS

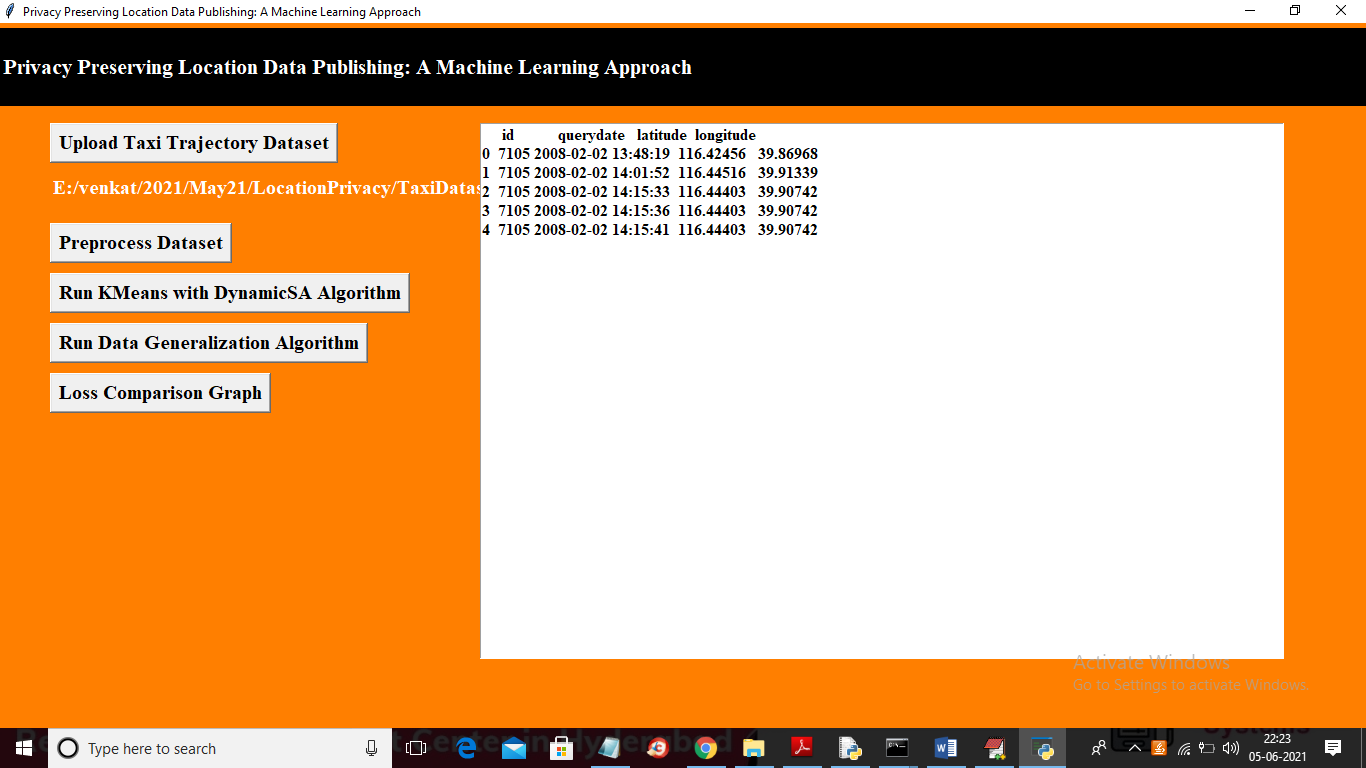
To run project double click on ‘run.bat’ file to get below screen



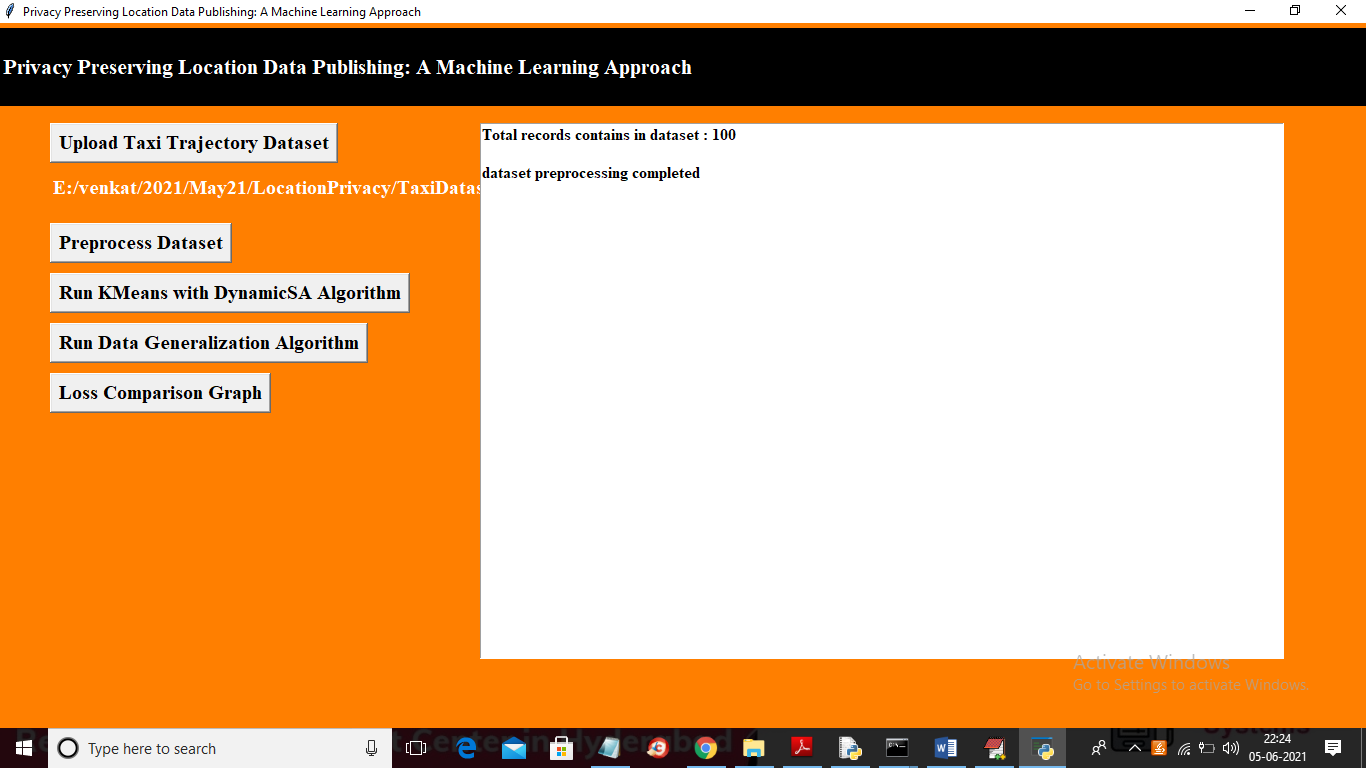
In above screen click on ‘Upload Taxi Trajectory Dataset’ button to upload dataset



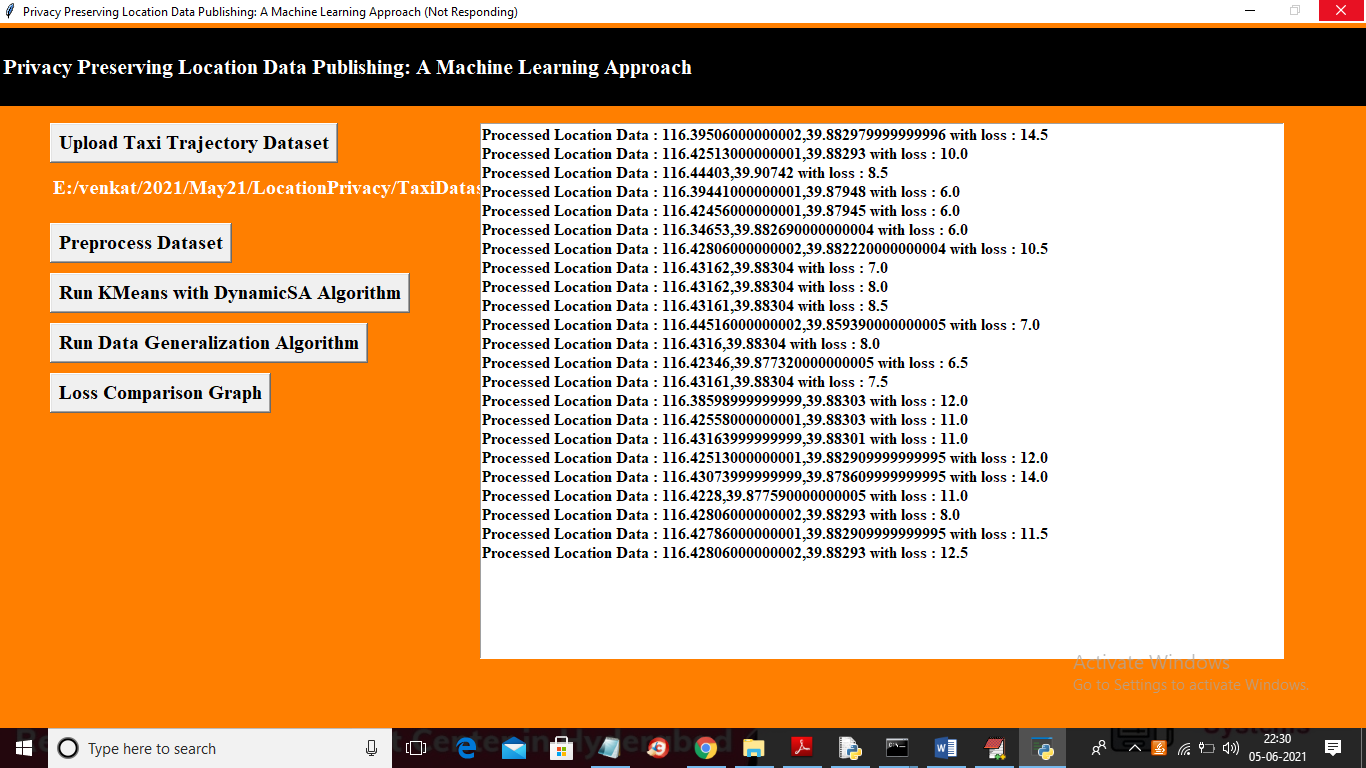
In above screen selecting and uploading taxi trajectory file and then click on ‘Open’ button to load dataset and to get below screen



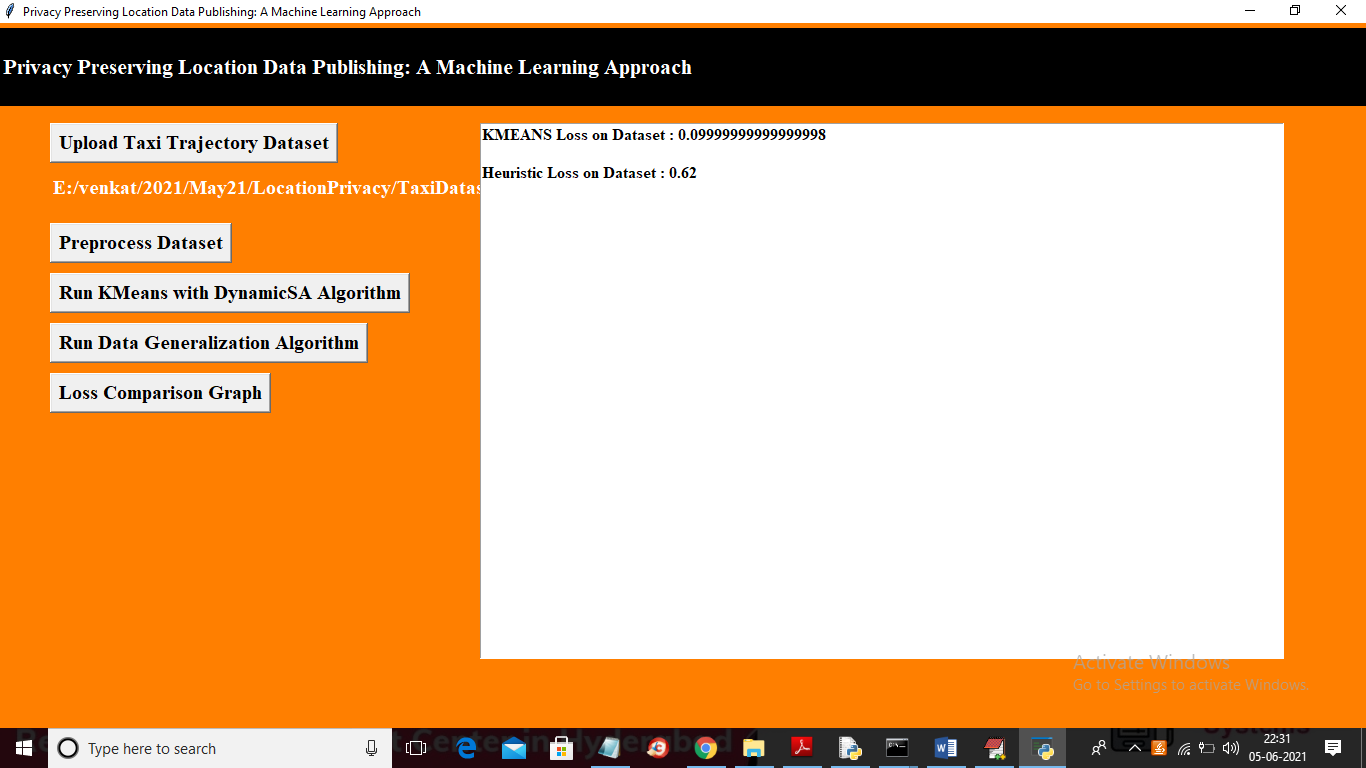
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to remove empty values and then extract latitude and longitude location from above dataset



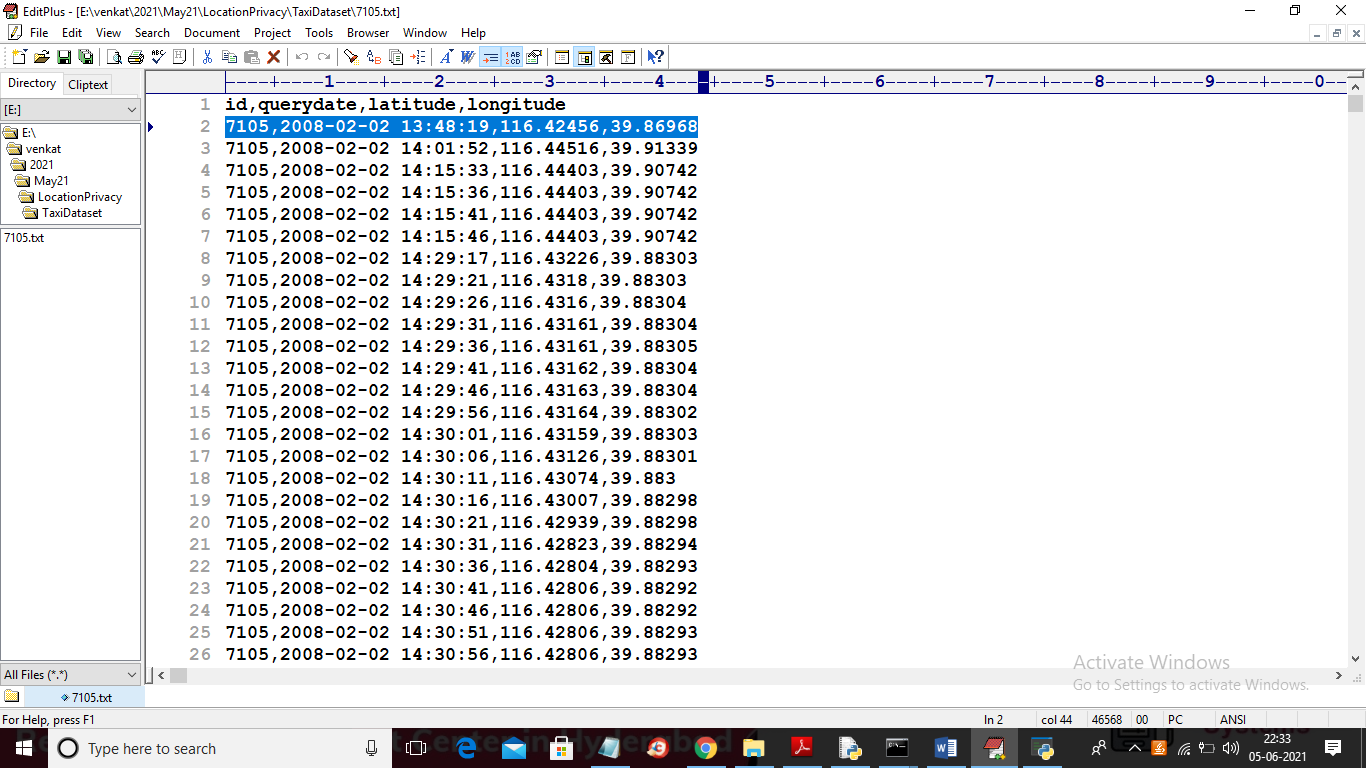
In above screen dataset preprocessing completed and now click on ‘Run KMeans with DynamicSA Algorithm’ button to run KMEANS on dataset with Dynamic SA. This algorithm will group all similar location into same cluster and then perform DYNAMIC SA.



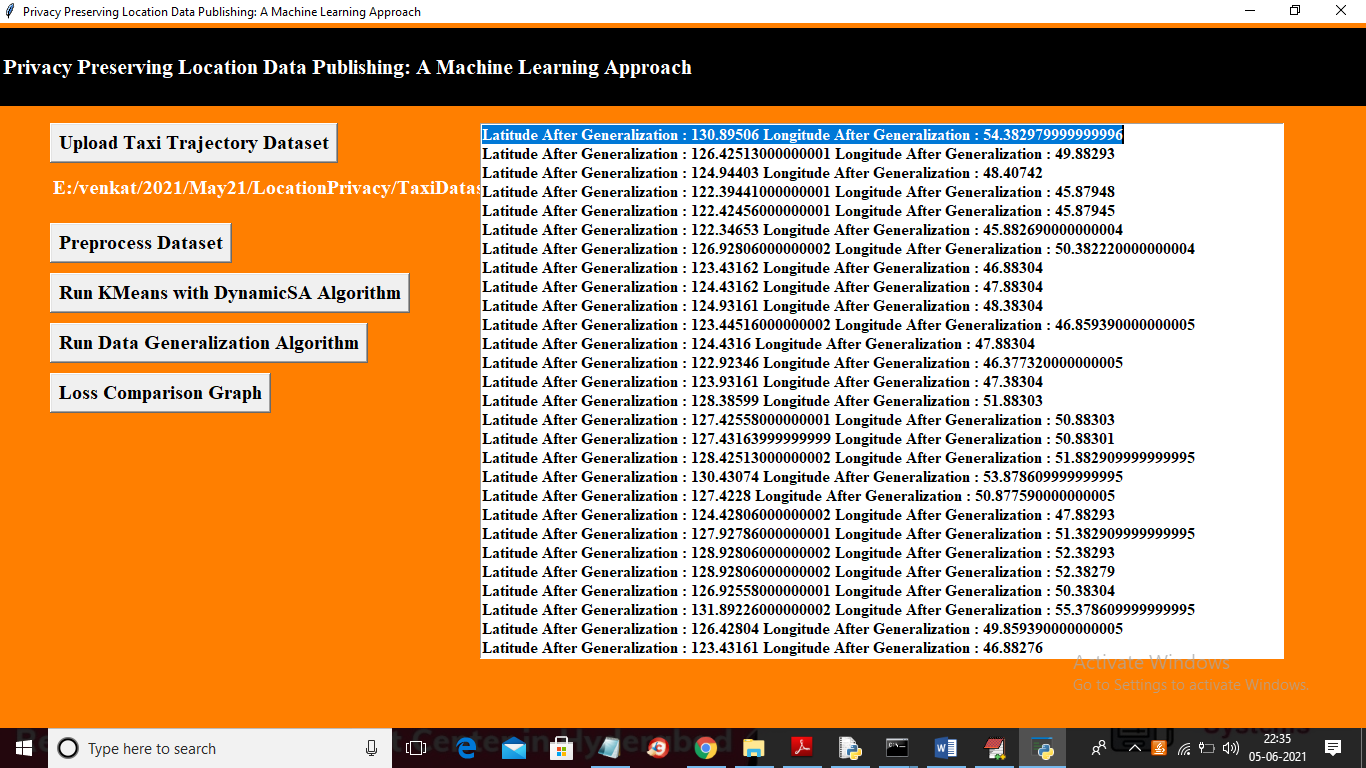
In above screen each location is processed and then calculating loss value with dynamic sequence alignment which align two locations by choosing minimal loss location.



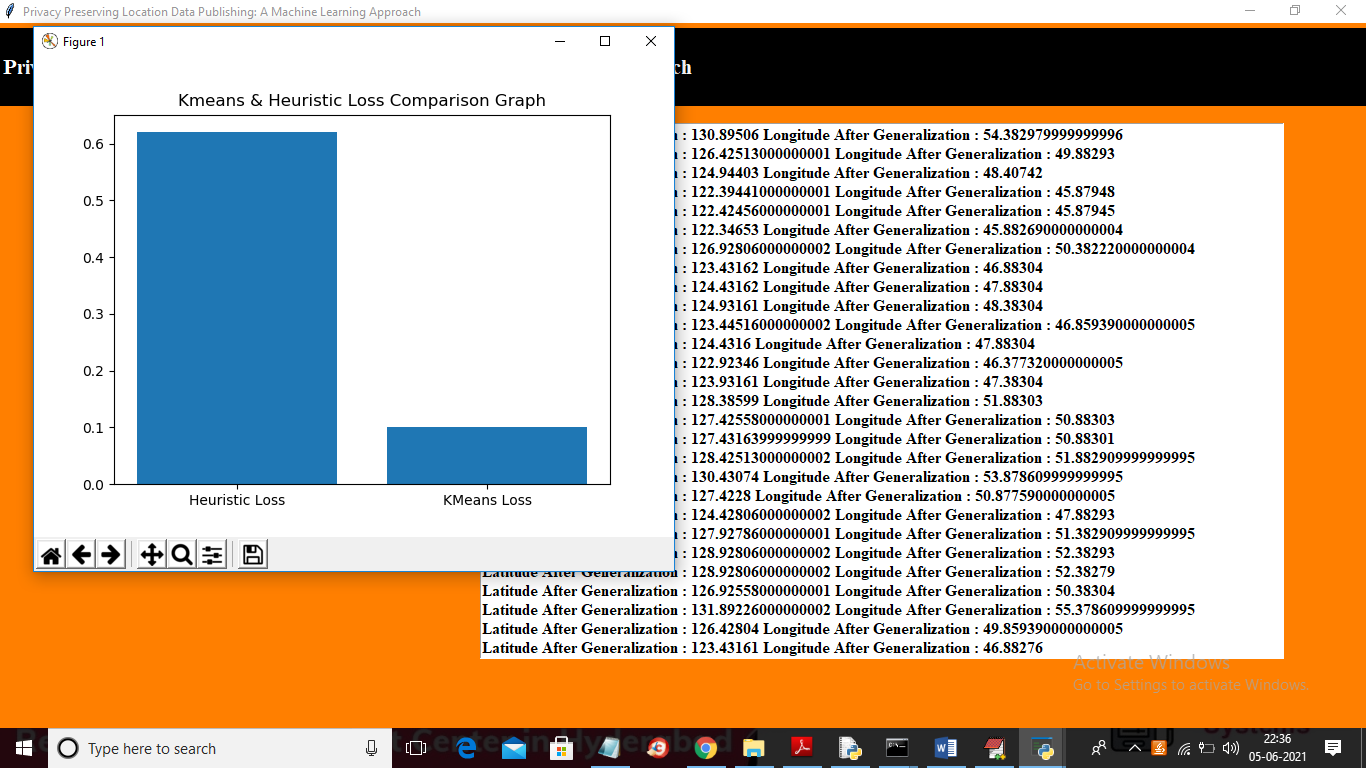
In above screen KMEANS loss is 0.09 and Heuristic Clustering (also known as Dynamic SA) loss is 0.62. Now click on ‘Run Data Generalization Algorithm’ button to generalized data with loss value. In below screen in first record you can see real location values from dataset and in next screen same location was generalized or anonymised with above algorithms



In below screen you can see the same location is generalized with other values



In above screen you can all location values are generalized so no malicious users can understand correct location. Now click on ‘Loss Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm name and y-axis represents loss values generated for that algorithm and in above graph KMEANS got less loss so KMEANS is better in anonymization.