Privacy Preserving Location Data Publishing: A Machine Learning Approach

In this project as extension I said I will be using BISECTING-KMEANS but this algorithm was not giving better result compare to KMEANS so I have used SPECTRAL CLUSTERING which is novel and efficient and better in performance compare to KMEANS and the LOSS value obtained from spectral clustering is lesser than KMEANS.

In recent years, spectral clustering has become one of the most popular modern clustering algorithms. It is simple to implement, can be solved efficiently by standard linear algebra software, and very often outperforms traditional clustering algorithms such as the k-means algorithm. On the first glance spectral clustering appears slightly mysterious, and it is not obvious to see why it works at all and what it really does.

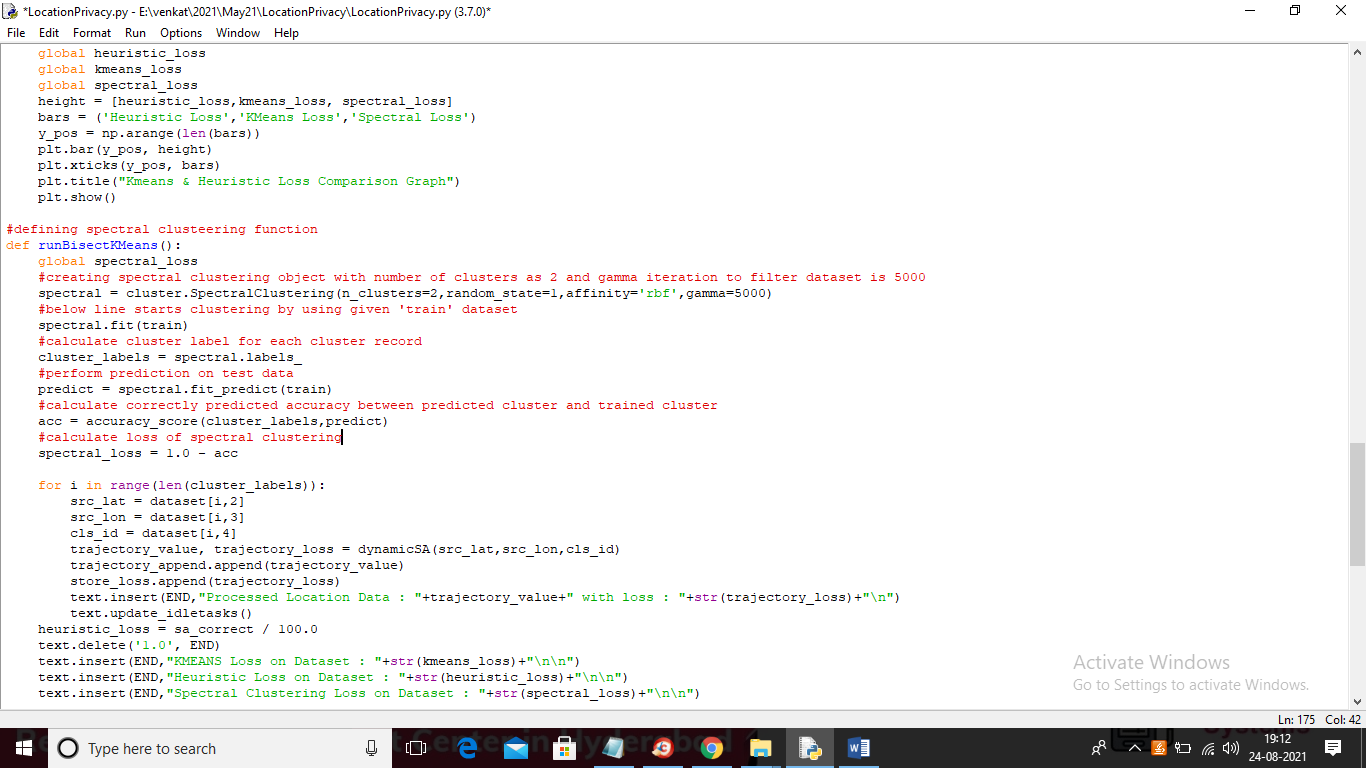
Clustering is one of the most widely used techniques for exploratory data analysis, with applications ranging from statistics, computer science, and biology to social sciences or psychology. People attempt to get a first impression on their data by trying to identify groups of “similar behaviour” in their data.

Spectral clustering is an EDA technique that reduces complex multidimensional datasets into clusters of similar data in rarer dimensions. The main outline is to cluster the all spectrum of unorganized data points into multiple groups based upon their uniqueness “Spectral clustering is one of the most popular forms of multivariate statistical analysis” ‘Spectral Clustering uses the connectivity approach to clustering’, wherein communities of nodes (i.e. data points) that are connected or immediately next to each other are identified in a graph. The nodes are then mapped to a low-dimensional space that can be easily segregated to form clusters. Spectral Clustering uses information from the eigenvalues (spectrum) of special matrices (i.e. Affinity Matrix, Degree Matrix and Laplacian Matrix) derived from the graph or the data set.

Spectral clustering methods are attractive, easy to implement, reasonably fast especially for sparse data sets up to several thousand. Spectral clustering treats the data clustering as a graph partitioning problem without making any assumption on the form of the data clusters.

In this project output I added extra function to apply Spectral clustering on train data and then calculate loss value

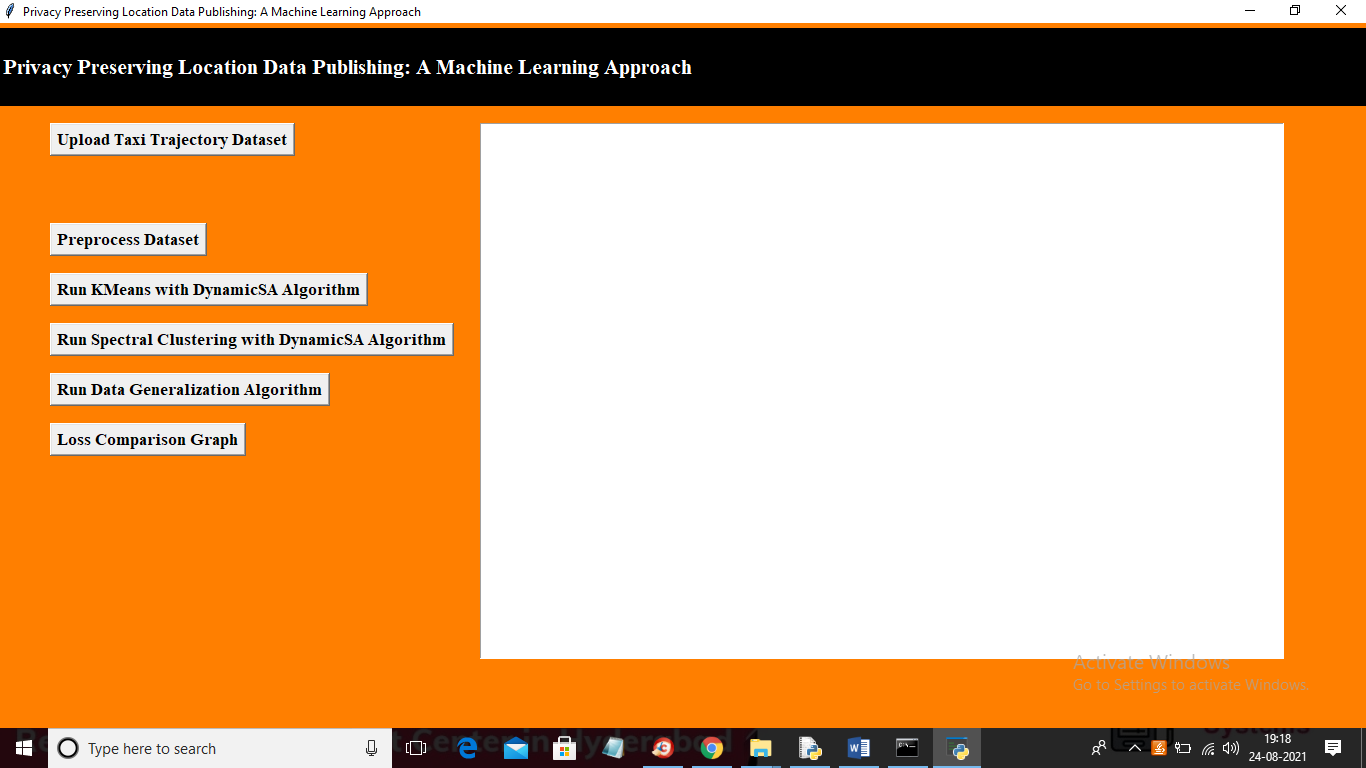
In below screen you can read red colour comments to know about spectral clustering



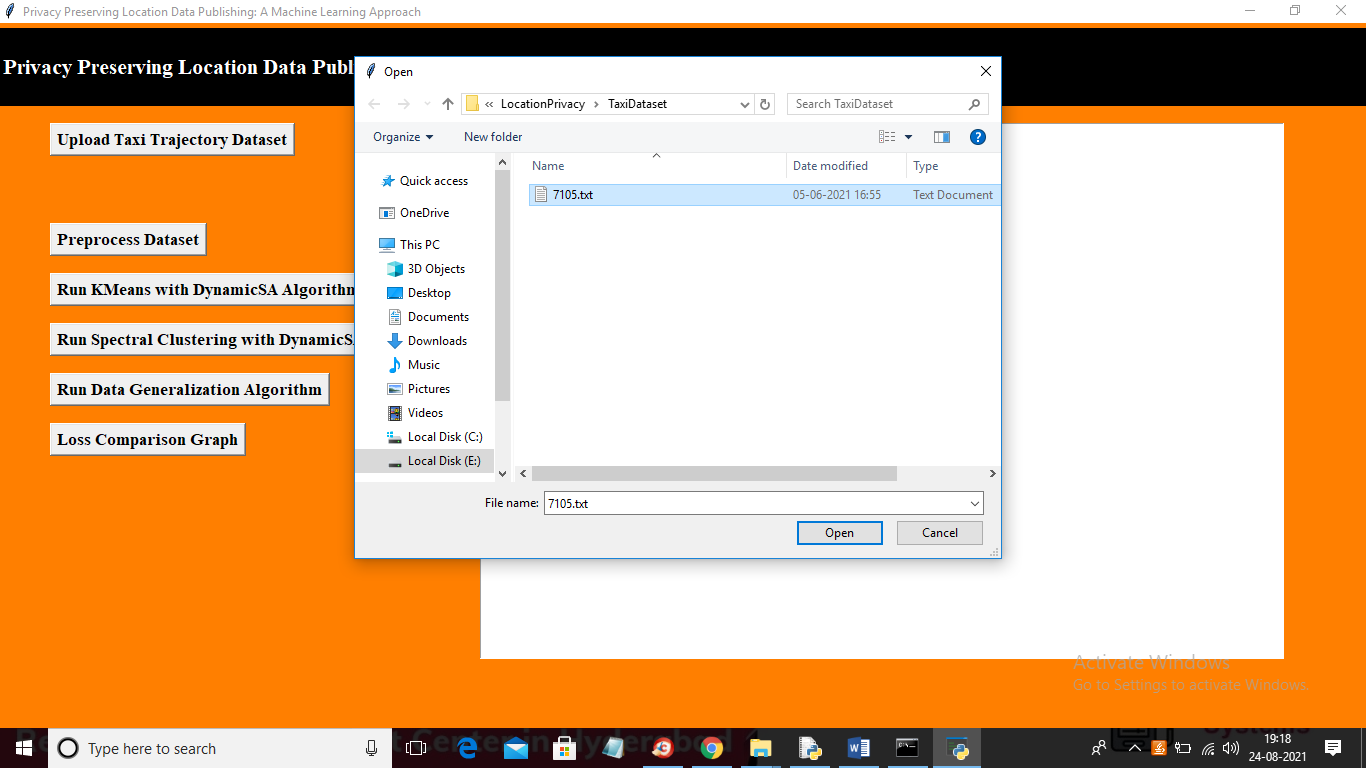
In above screen read red colour comments to know about spectral clustering implementation

SCREEN SHOTS

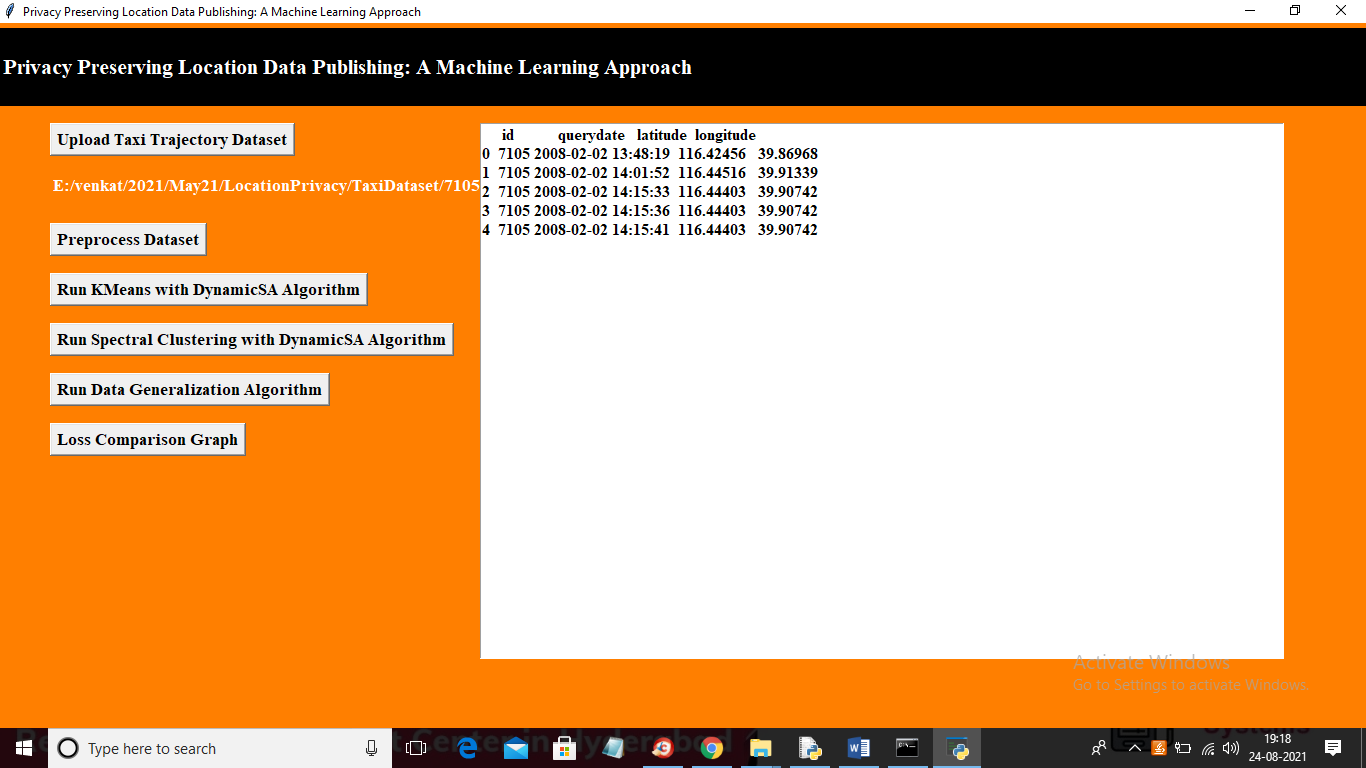
To run project double click on ‘run.bat’ file to get below screen



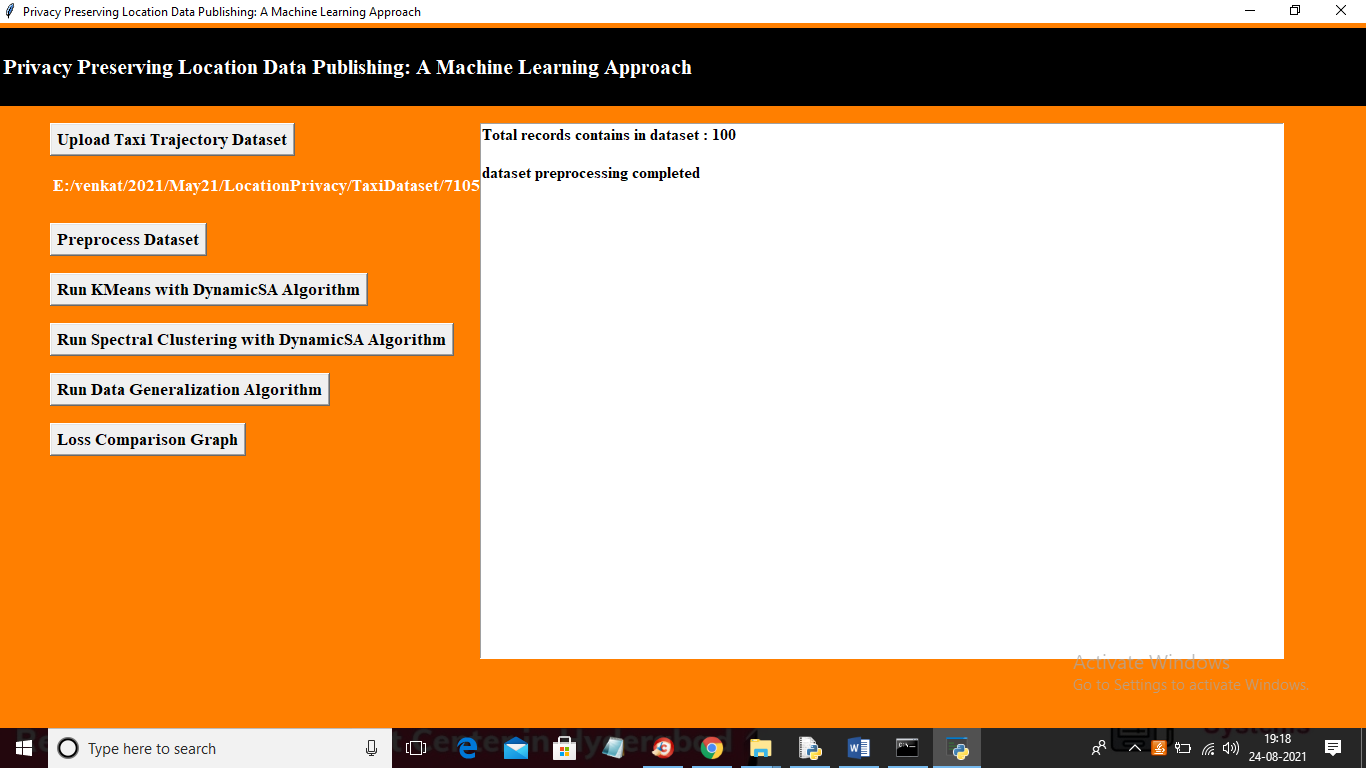
Like old execution upload dataset and run each button one by one



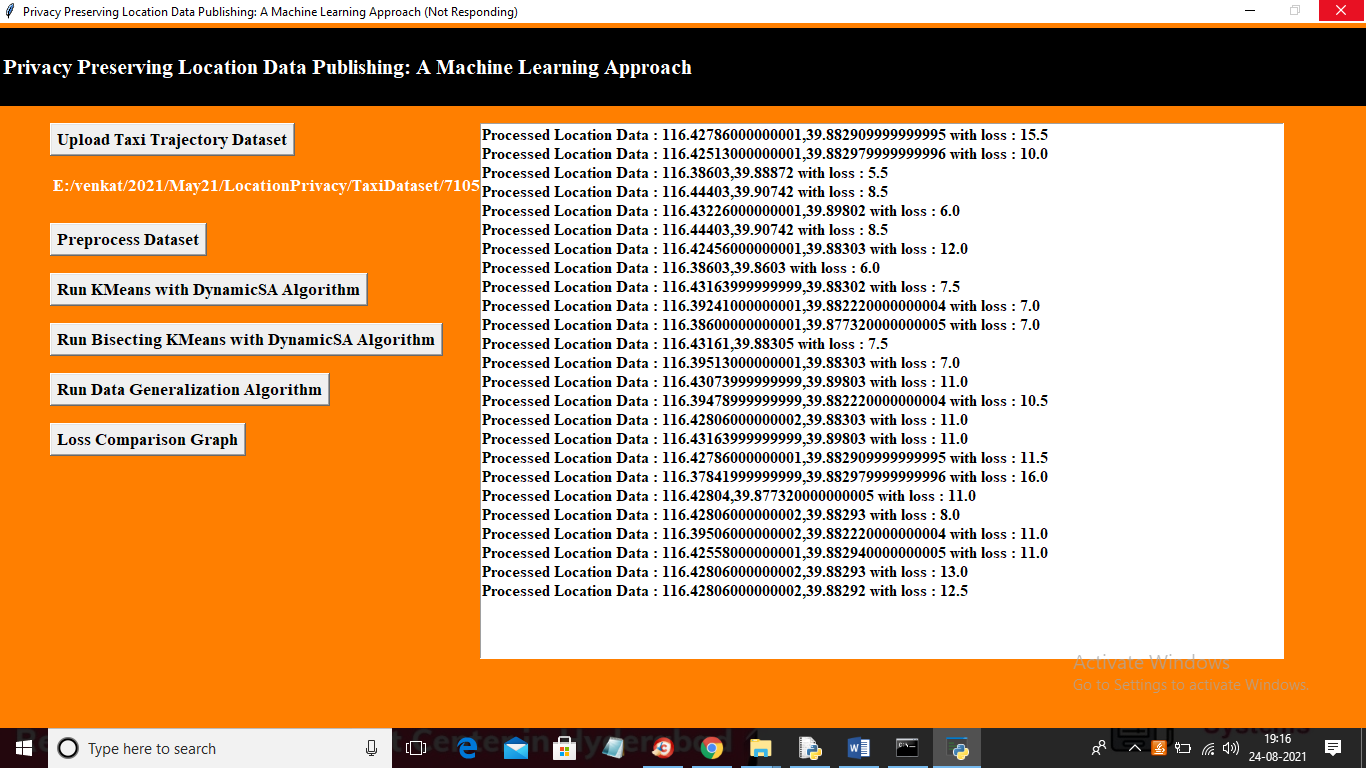
In above screen uploading dataset and after loading dataset will get below screen



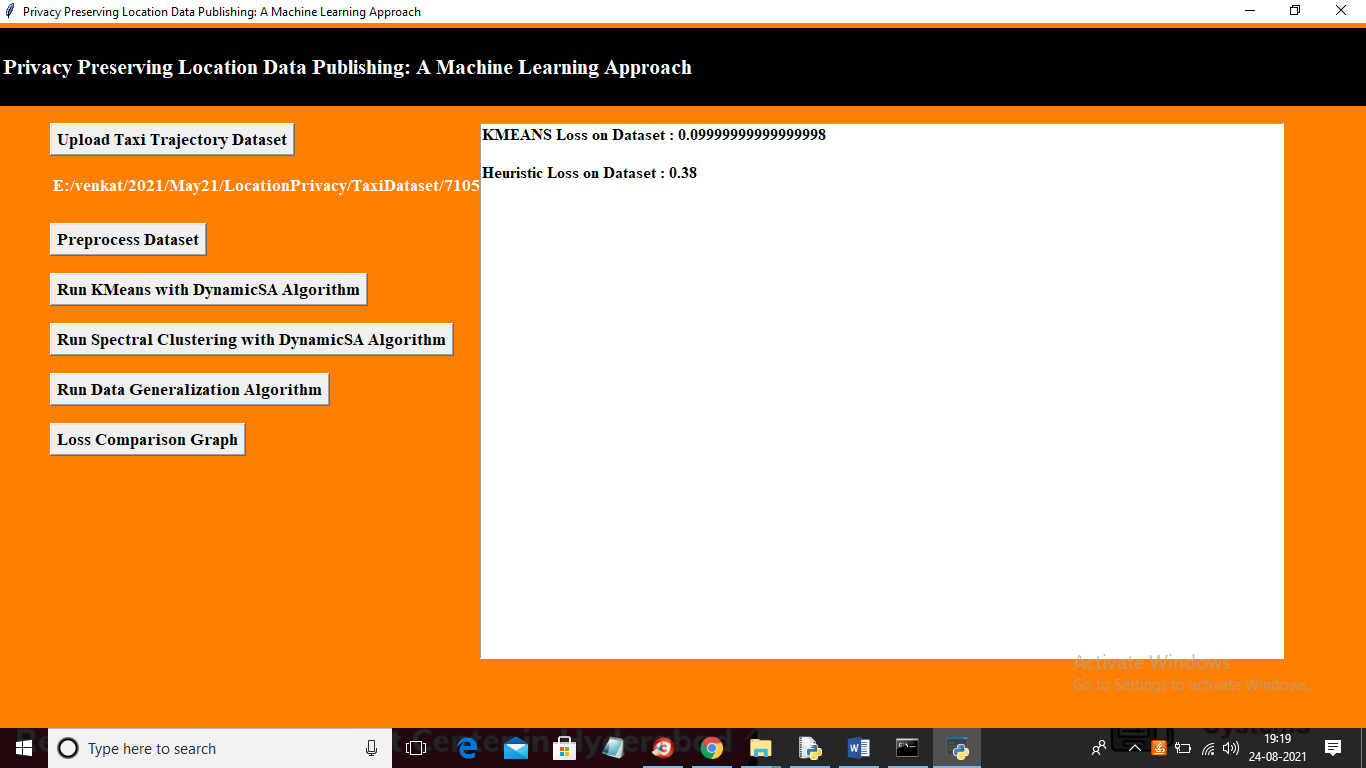
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to read and clean dataset and to get below screen



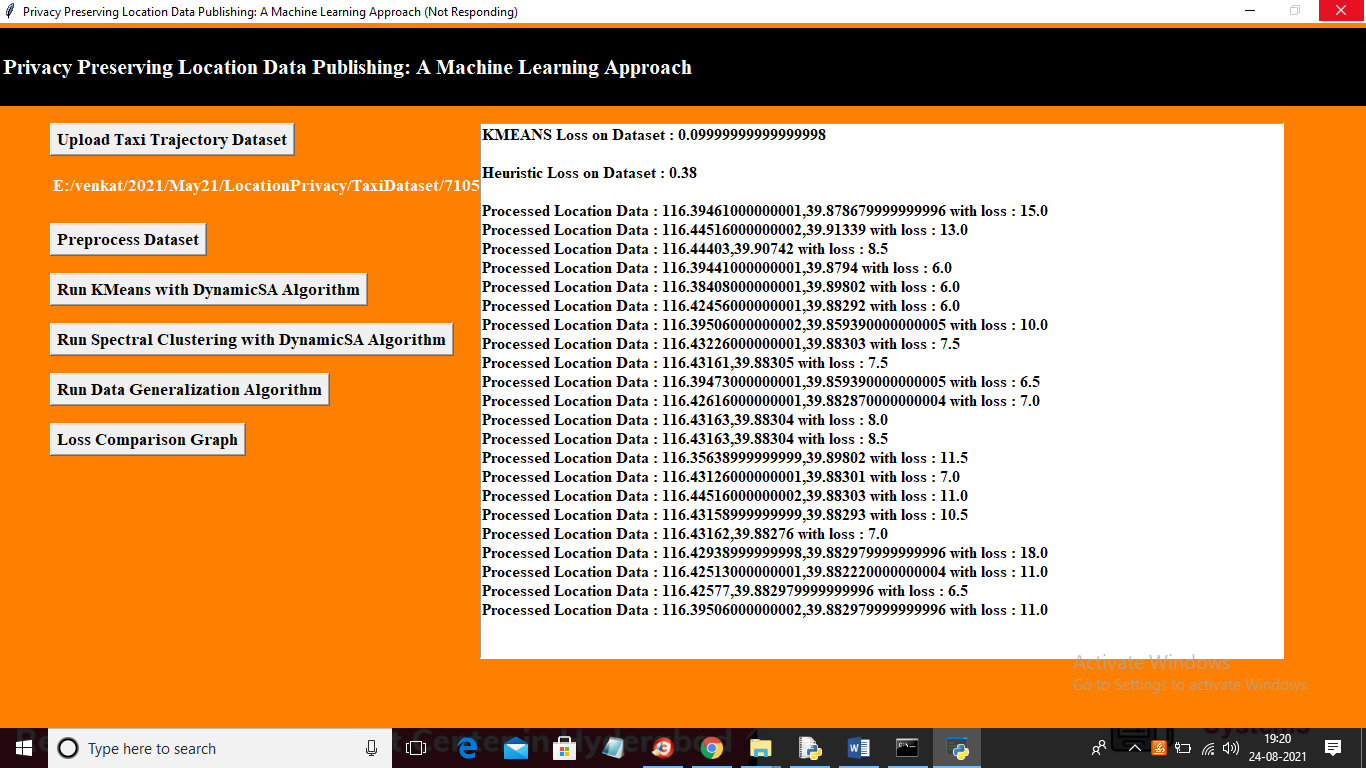
In above screen dataset contains 100 records and now click on ‘Run KMeans with Dynamic SA Algorithm’



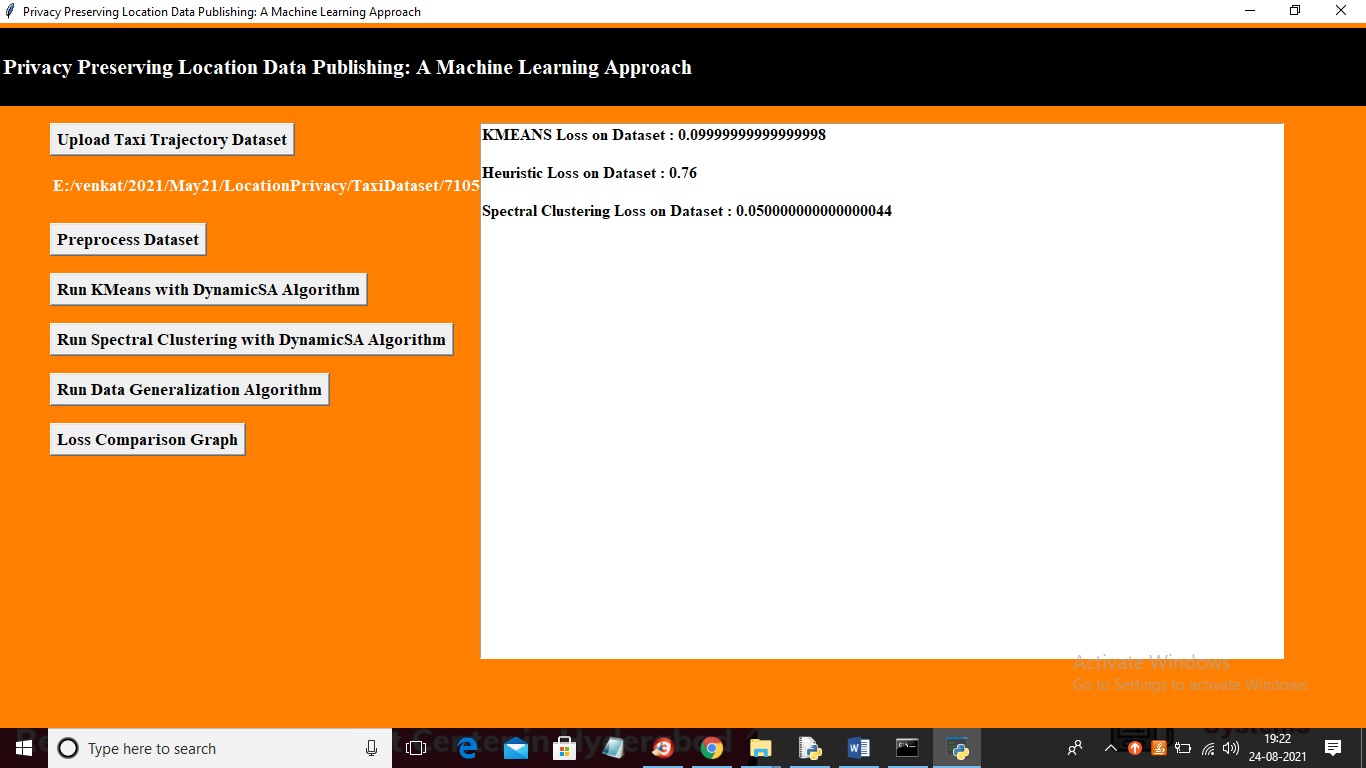
In above screen KMeans predicting location with KMEANS and calculating loss and after predicting all locations will get below output



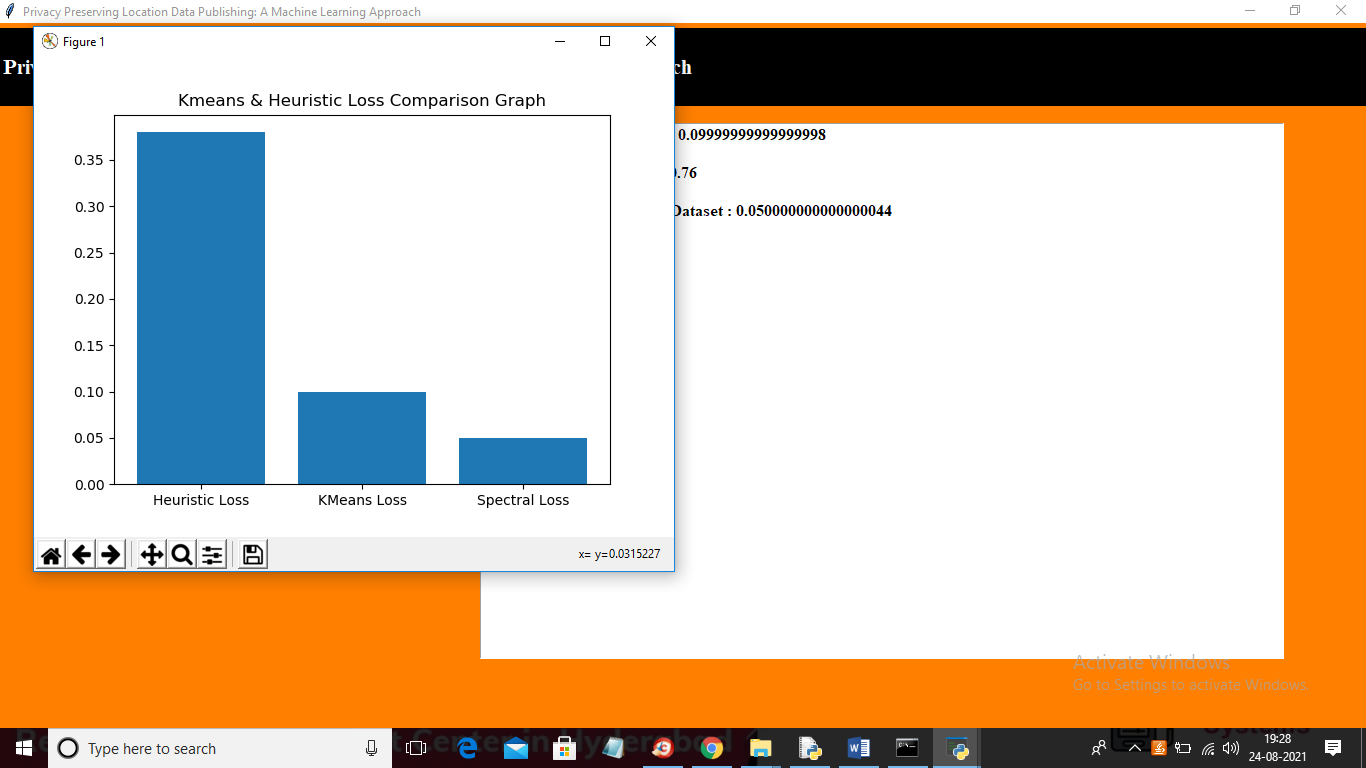
In above screen with Heuristic Algorithm total loss is 0.38 and with KMEASN we got loss value as 0.099 and now click on ‘Run Spectral Clustering with DynamicSA Algorithm’ to calculate loss value using spectral clustering



In above screen spectral clustering location prediction started and in first record u can see spectral loss value is 15.0 and in Kmeans you can see that loss value is 15.5 and after all location prediction will get below output



In above screen with spectral clustering we got loss value as 0.05 which is lesser than kmeans and heuristic algorithm. Now click on ‘Loss Comparison Graph’ button to get below graph



In above graph we can see with spectral we got less loss values which means it has high accuracy