CS306 Data Analysis and Visualization Winter, 2019

Project Yes Bank Big Data Analysis

1. For this project you have a dataset yes_bank.csv. It is a large dataset with over 10 million entries of yes bank customers. Entries consist of two main sectional details of current address and details of permanent address. In the project you have to achieve two goals …

1. Visualize the customers as heat map on a global world map.
2. Classify the customers as household and business based on the data and then visualize them as different heat maps.

In your report clearly identify the following data analysis steps and your solutions to them.

1. Data cleaning and handling missing data.
2. Data Normalization.
3. Outlier Detection and removal of them.
4. Data Augmentation.
5. Method/Algorithm used for classification and clustering.

In the report clearly describe your solution for each of the above task and clearly display your results. (At least 4-5 images.)

1. Heat map based on permanent addresses.
2. Heat map based on current addresses.
3. Heat map for household customers.
4. Heat map for business.

How many outliers your method has detected ?

👉 (Submit your solution via mail on cs306.daiict@gmail.com before .)