

DATA MINING ASSIGNMENT 1

Welcome to the world of Data Mining. This is your first assignment in this course. This will help you to build your thinking ability in the field of data. Try to explore more and report more analysis in your report file.

Instructions: You are allowed to discuss but the final answer should be your own. Any instance of cheating will be considered as academic dishonesty and penalty will be applied.

- This assignment is to be done individually.
 - Any help taken from any outside source must be reported in your report.
 - Institute's policy on plagiarism will apply.
 - In Q1, don't follow the kernels blindly as there might be some wrong/misinterpreted observations.
-

Q1 - Please go through the below kaggle link -

<https://www.kaggle.com/c/recruit-restaurant-visitor-forecasting>

Read the problem statement carefully and do exploratory data(given in the link) analysis. You can take the help of kernels provided in the link but you are supposed to do your exploration in Tableau. All the observations, data visualization, and analysis graphs must be submitted through your tableau sheet. You may want to submit a report as well as describing your inferences. [40 marks]

Q2 - Generate a random dataset with 1000 data points and 10 classes (classification problem). Perform sampling (with and without repetitions) on the dataset and find the proper number of samples required such that there should be at least one representative point (k) from each class.

Plot a probability v/s sample size graph. The probability represents the chances of getting at least one representative point from each class.

Perform the above experiment with $k = 1, 2, 3, 4$.

Plot a graph representing the frequency of each proper sample size obtained.

Submit a report along with your code and inferences, comparisons, and graphs. Your analysis and the detailed process followed should be reported neatly. [10 marks]

ALL THE BEST !!!