15CS481 Machine Learning and Data Mining Lab
Exercise -3
Measuring Proximity of Datapoints

1. In this lab we will explore proximity concept of mixed attributes
2. Download the dataset from here
3. The description of dataset is available in UCI Machine Learning Repository
4. Using Python, analyse the dataset and answer the following questions (You need to write the code for each of the questions.
5. Please follow the same question order in the notebook.
6. Give question titles using markdown options (see the sample notebook attached). (or download from here
7. Submit the notebooks with output (rename it with your register number).

Questions
1. Find how many records in the dataset? (1 Mark)
2. Print how many attributes and whether any of attribute have Null value (1 mark)
3. Assign roll number as a new column to the dataset (1 mark)
4. Take 4 columns (Roll Number, G1, G2, and G3). Compare the students based on pairwise Euclidean distance. Print the first 10 students highly similar to Roll number 1 (2 marks).
5. Compute the dissimilarity matrix based on following attributes (age, family size, Fjob, guardian, activities, absences, G1, G2). Please refer to the UCI link for types of attributes. This is a mixed type dataset. So follow the method discussed in Lecture 4.3. (5 Marks)