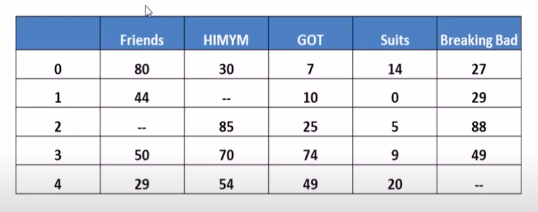
**KNN IMPUTATION FOR MISSING VALUES**

It is the multivariate imputation

DATASET:

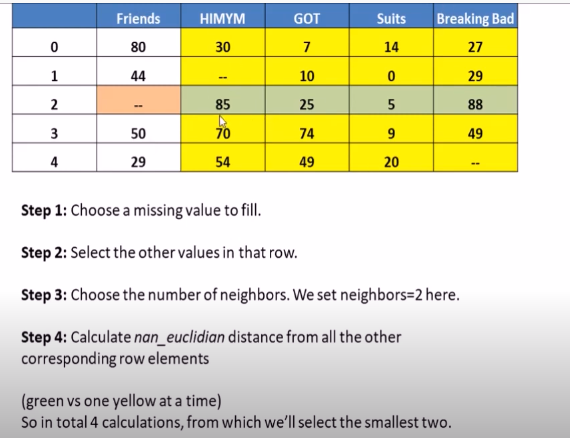


In this dataset we can able to find null values in all the columns. So we want to use KNN imputation

The missing values will be calculated by taking the values of the neighbour of the particular value.

The column with null value will considered as the output feature at that instant.

The null value will be calculated using Euclidian distances



The K represents number of neighbours in KNN.

In this calculation 4 neighbours has been took into reference.

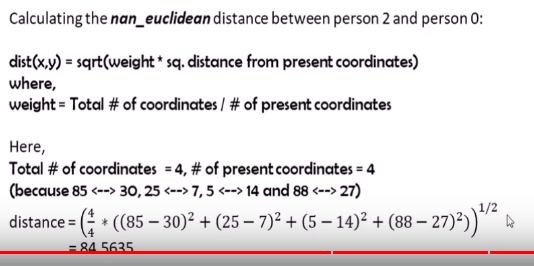
**CALCULATION:-**

**Total # of coordinates** means no of columns considered excluding the column of missing values.

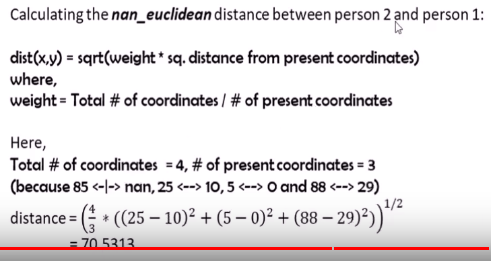
**Total # present coordinates** means during a calculation between a single neighbour row in the instance and row containing missing value.Both the rows should contain value.

If both missing row and neighbour row has a value then it is a present coordinate.

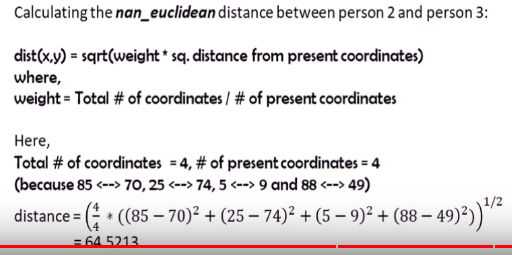
If any one of the value missed in row then it is not present coordinate.

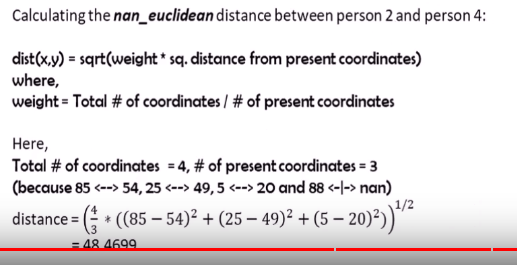


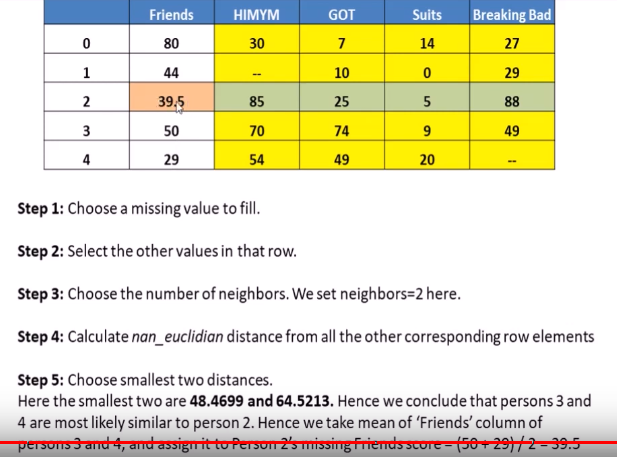
Here total # of coordinates and present coordinates are 4 because in both the rows that is the row of person 0 and person 2 there is no missing value.(excluding missing feature)



Here present coordinates is 3 because there is a missing values in himym column. So we cant take that into reference.







Like wise we want to do it for all the values.

Scikit learn has the KNN imputer.