KNN Imputation

Maths	Chemistry	Physics	Computer Science
80	70	NaN	78
90	65	57	89
NaN	58	84	67
92	NaN	78	NaN

Step 1:Choose the first column with the missing value to fill in the data

Column name :Maths

Step 2:Select the values in a row

Maths	Chemistry	Physics	Computer Science
80	70	NaN	78
90	65	57	89
NaN	58	84	67
92	NaN	78	NaN

Step 3: Choose the number of neighbors

K=2

Step 4: Calculate the Euclidean distance from all other data points corresponding to each other in the row.

$$d(X, Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$

where X_i and Y_i are the values of the i-th variable in observations X and Y, respectively.

D(1,3)=SQRT((70-58)^2+(78-67)^2) =16.27

D(2,3)=SQRT((65-58)^2+(57-84)^2 +(89-67)^2) =35.52

D(4,3)=SQRT((78-84)^2) =6

Step 4: Select the smallest k values and take average of them.

$$\hat{Y}O = \frac{1}{k}\sum i = 1^k Y_{S_i}$$

where $\hat{Y}O$ is the imputed value for the missing income variable in observation O, YS_i is the value of the income variable for the i-th observation in S, and k is the number of observations in S.

Imputed_value=1/2*(80+92) =**86**

For Weighted average

D(1,3)=(70-58)^2+(78-67)^2=265

D(2,3)=(65-58)^2+(57-84)^2 +(89-67)^2 =534

D(4,3)=(78-84)^2 =36

Weights

Total number of columns/Total no.of columns having values

Maths	Chemistry	Physics	Computer Science	Weights
80	70	NaN	78	3/2
90	65	57	89	3/3
NaN	58	84	67	
92	NaN	78	NaN	3/1

Weights*Sum of Euclidean distance

SQRT((3/2)* 265)=19.93

SQRT ((3/3)*534)=23.10

SQRT((3/1)*36)=10.39

Imputed_value=1/2*(80+92) =**86**