**BONUS QUESTION ANSWER: when doing the clustering we have to find the nearest boundary values. Non null values are good to process the data when compared to the null values. so, mean values with non zero values are better than the null values.**