The dataset for this project (Global shark attacks) from kaggle required a lot of cleaning.The missingno library was used to analyse and look at the percentage of missing values.Some columns had to be dropped since they couldnt be imputed at all eg:Time. Age column had to be imputed with the mean of ages. They were no outliers as such since there was only one numeric data point.There was some data extracting which had to be done like extracting month from case number which is demonstarted in Data Storytelling-2 python notebook. Plus the activity column had to be cleaned and replace with proper entries.
Column names had to be altered for easy use age.