Primary data set – EDA

* We have 296 rows which means 296 countries information, which means there is information for both countries and sub regions in the countries. So we are filtering the rows which only have country code(which means the rows which have country codes are countries and rest all are sub-regions)
* Removed the rows where there are missing values in the target variable. As algorithms like KNN cannott handle(to my knowledge).
* After doing the above steps we are now ended up having 196 rows (=countries).
* Now, refactored the country and country\_code columns.
* There are some missing values, as the boxplots between those columns showed there are some outliers. Replacing the missing values with median of their respective columns.
* Have looked into the correlation matrix and correlation plot. It seems there is almost no correlation between the predictor values('poverty\_gap','public\_health\_exp','malnutrition\_death\_rates','Infant\_mortality\_rate','GDP\_per\_capita','annual\_health\_care\_per\_capita') to the target -'homicide\_Rate'.