HW2 Report

Q2:

	spread
spread	1.000000
PCR_10	0.212723
PCR_07	0.042114
blood_A_AB	0.037510
PCR_01	0.022668
PCR_06	0.020925
covid	0.014039
PCR_02	0.013176
num_of_siblings	0.010719
household_income	0.008204
PCR_03	0.003534

Figure 1: The 10 most correlated features to `spread`

Q3:

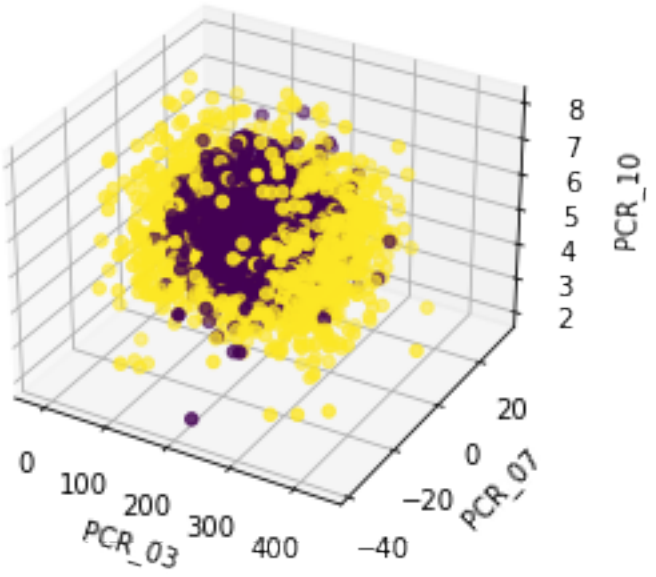

Figure 2: 3D Scatter plot of `PCR_03`, `PCR_07`, and `PCR_10` according to `spread`

Q4:

Q6:

Z-score scaling scales the data of a feature by ensuring that they have zero mean and unit variance. What this means is that features scaled according to this technique have their outliers handled correctly, whereas the scale between different features may vary. This technique would be preferably used on features which already follow a normal or close to normal distribution and or have many significant outliers. On-the-other-hand, the min-max technique involves scaling the data of a feature to a specific range (generally between 0 and 1). Contrary to the Z-score method, this technique guarantees uniform scale across features, but does not handle outliers well. Therefore, it would be preferable to use this technique when the feature in question has no significant outliers and or does not follow a normal distribution.