MSDS-422-SEC56 Quick Assignment Data Transformation Daniel Zaremba

MinMaxScaler Data Observations

Selected field MORTDUE and normalized it without using MinMaxScaler. The below shows the results are similar to MinMaxScaler.

	IMP_MORTDUE	nor_IMP_MORTDUE	calc_IMP_MORTDUE
0	25860.0	0.119783	0.119783
1	70053.0	0.342229	0.342229
2	13500.0	0.057568	0.057568
3	58936.0	0.286272	0.286272
4	97800.0	0.481894	0.481894

StandardScaler Data Observations

Selected field MORTDUE and normalized it without using StandardScaler. The below shows the results are similar to StandardScaler.

	IMP_MORTDUE	std_IMP_MORTDUE	calc_IMP_MORTDUE
0	25860.0	-1.263155	-1.263035
1	70053.0	0.029760	0.029757
2	13500.0	-1.624761	-1.624606
3	58936.0	-0.295481	-0.295452
4	97800.0	0.841529	0.841448

Notes & Observations

Two traditional ways to scale the data, first one is the MinMaxScale, where we take the min and max and scale everything between 0-1. Second way is to standardize the data, subtract by the mean and divide by the standard deviation everything will be between +- 3 usually. Both methods MinMaxScaler and StandardScaler require the data to be numeric, need to take out the categorical variables. For this assignment it was imperative to handle the outliers prior to normalizing the variables, in this exercise we dropped the outliers that were 3+ standard deviations. The screenshots above show that the manually normalized fields match the calculations performed by StandardScaler and MinMaxScaler.