Data Analysis - Explanations

May 21, 2024

To assess collinearity, we use correlation plots. One plot shows correlations using age bins, where age is grouped into ranges, and the other uses average age as a continuous variable. Both plots help identify how age, whether binned or averaged, relates to other variables, revealing potential collinearity issues. Here, we see that there are no collinearity issue between the different variables. Only the age relate to one another as they represent bins. To continue, to
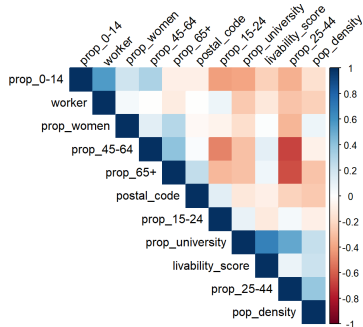

Figure 1: Age Bins Correlation Plot

Figure 2: Age Average Correlation Plot

assess collinearity, we use Variance Inflation Factor (VIF) values. VIF measures how much the variance of a regression coefficient is inflated due to collinearity with other predictors. High VIF values indicate a high degree of collinearity, which can affect model accuracy and interpretability. Once again, only the age bins have high values. When discared, all values lie below 4 showing very low collinearity potential. Finally, to assess the normality of residuals in a regression

	postal_code	prop_0.14	prop_15.24	prop_25.44	prop_45.64
	1.173359	153.568434	264.299072	553.670692	185.667452
prop_university	worker	prop_women	pop_density		
1.505215	1.967125	1.401917	1.460526		

Figure 3: Age Bins VIF values

postal_code	Weighted_Average_Age	Standard_Deviation_Age		prop_university
1.181432	2.433321	3.983387		3.208942
worker	prop_women	pop_density		Aid
1.809524	1.224057	1.648782		1.378689
Income	single_family_housing			
2.201357	2.073348			

Figure 4: Age Average VIF values

model, we use a Q-Q plot (Quantile-Quantile plot). A Q-Q plot compares the distribution of residuals to a normal distribution. If the residuals follow a normal distribution, the points on the Q-Q plot will lie approximately along a straight

line. Deviations from this line indicate departures from normality. Both graph show that the data is fairly normal. The age average graph has a slightly disturbance at the extremes, it is however insignificant for the model.
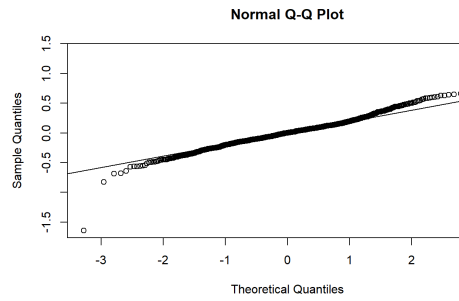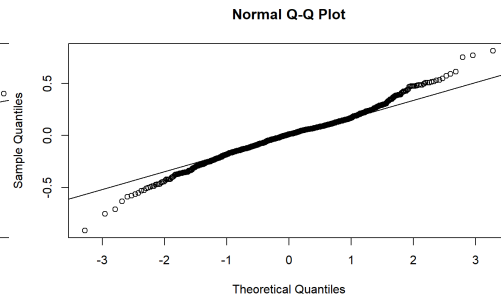

Figure 5: Age Bins Q-Q Plot

Figure 6: Age Average Q-Q Plot