|  |
| --- |
| Research Question: In pursuit of optimizing our model to predict interest rate |
|  | using independent variables (quantitative/qualitative), we keep a sharp eye on such |
|  | issues as normality, heteroscedasticity, linearity, and multicollinearity. How do we |
|  | properly address, if need be, these issues? How do we select the, in a sense, most |
|  | effective model to predict interest rate? Simply put, what could be better and do better |
|  | than this? |
|  |  |
|  | Call: |
|  | lm(formula = my.data$int\_rate ~ my.data$loan\_amnt + my.data$home\_ownership + |
|  | my.data$annual\_inc + my.data$application\_type + my.data$emp\_length) |
|  |  |
|  | Coefficients: |
|  | (Intercept) my.data$loan\_amnt |
|  | 1.228e+01 1.059e-04 |
|  | my.data$home\_ownershipMORTGAGE my.data$home\_ownershipNONE |
|  | -4.970e-01 1.633e+00 |
|  | my.data$home\_ownershipOTHER my.data$home\_ownershipOWN |
|  | 4.684e-01 -1.186e-01 |
|  | my.data$home\_ownershipRENT my.data$annual\_inc |
|  | 3.163e-01 -8.603e-06 |
|  | my.data$application\_typeJOINT my.data$emp\_length1 year |
|  | 1.432e+00 8.871e-02 |
|  | my.data$emp\_length10+ years my.data$emp\_length2 years |
|  | 2.032e-01 9.079e-02 |
|  | my.data$emp\_length3 years my.data$emp\_length4 years |
|  | 9.012e-02 1.400e-01 |
|  | my.data$emp\_length5 years my.data$emp\_length6 years |
|  | 2.064e-01 3.465e-01 |
|  | my.data$emp\_length7 years my.data$emp\_length8 years |
|  | 3.853e-01 2.066e-01 |
|  | my.data$emp\_length9 years my.data$emp\_lengthn/a |
|  | 2.124e-01 3.196e-01 |
|  |  |
|  |  |
|  | A quick take away is that, |
|  | 1. The more you borrow, the higher the interest rate. |
|  | 2. Mortgage borrowers are favored. |
|  | 3. The more you make each year, the less interest you have to pay. |
|  | 4. Joint loan applications get robbed. This is probably because only people with bad credit |
|  | are getting joint loan apps. |
|  | 5. It seems that 0-3 years loans have the lowest interest rate. After that, it just gets randomly shot up. |
|  |  |
|  | A couple of thoughts. |
|  | 1. Maybe we should consider using a quadratic approach on variables such as loan amount and annual income. |
|  | 2. Use interaction terms between home ownership and loan maturity. |