

Boston Crime Multi-variable linear regression Model

Salik Hussaini

Criminal law stems from the fact of criminalization. To criminalize, or in other words, an act of violation of the law, ought not to be done. However, crimes continue to occur in Boston. Our team has begun to analyze the Boston crime data and would like to propose strategies on how to reduce them. We will begin to explore crimes based upon several factors and create a regression model to help reduce them from occurring. Variables that will help build the linear regression model will be a numerical value of the date of when the crime occurred, the reporting area in square miles, the population of the area, the hour in which the crime occurred, and an interaction term in which population is divided by the area of the reporting area where the crime occurred. All these independent variables will be fed into a linear regression model that will help predict the dependent variable of the latitude of where the crime occurred. This model shall be helpful for the Boston crime department by being able to predict potential future crimes.

According to boston.gov, from 2015 to 2019, 393,185 crimes were reported within the Boston Metropolitan Area. Annually, for the year of 2017, 93,278 crimes had been reported, and 91,423 crimes were subsequently reported in 2018. Currently, from January 1 through September 25, 2019, 65,519 crimes have been reported. This data alone shows nothing is being done to slow down the crimes that are continuing to occur in Boston. Further analysis needs to be in place for a reduction in crimes in Boston.

From the original data set that was worked with the team, which included 18 variables; the dataset that was included in this report includes only 4 independent variables, excluding an interaction variable. One can say this model is very specific on the date parameter and the exact location where the crime happened. Both these independent variables stem from the crime information that was reported. Again the dependent variable is the latitude of where the crime occurred. Looking at the linear regression model that was created, it could be said that the model did quite well with the data was given. The r^2 of the model was 31.34. This means 31.34% of the variability of the dependent variable is explained by our model. Looking at the F statistic, it is also significant. It could be said that something in our model is working. Looking at the T-Tests for our model, they all show high significance. This linear regression multicollinearity

was also checked. Multicollinearity was also checked and the VIF's seemed low for the independent variables. This linear regression model does show promise in predicting the latitude on where a potential crime can occur.

Further analyzing the data set, the betas of the independent variables show a wealth of knowledge in analyzing the dependent variable. With each incremental increase, the date variable, the area variable, and also the hour variable report a negative influence on the latitude of the crime. Specifically looking at the hour of the crime, with each increase in hour in time, the latitude of the crime will be decreased by .0000516. Although this is a small value it shows insight on where and when a possible crime could occur.