Observational data was collected on towns (i.e. subarbs) of Boston, Massachusetts. The data is

in ‘boston\_project1.csv’. The variables from the study and a description of each variable is in

the table below.

Table

Description automatically generated



General Research Questions of Interest

1. Conduct some exploratory data analysis (visualization or numeric summaries) on the

variables above and how they relate or don’t relate to crime rate (crime\_cat)

2. Researchers want to know if the category of crime rate (crime\_cat) can be predicted

from the data collected and quantifying how well they would be able to expect their

model to work on new observations (i.e. you should use cross-validation). Lastly, they want to know what accuracy they can expect your chosen model to have on new observations and want a final model fitted on all of the data.

Notes

* You may want to consider transforming some variables after doing 1) if needed.
* Tidymodels functions for LDA, QDA, Naïve Bayes
  + LDA - discrim\_linear()
  + QDA – discrim\_quad()
  + Naïve Bayes – naive\_Bayes()
* You should pick the model to use by comparing performance using cross-validation
* You should make a final fit of your optimal model on all the data. You can report coefficients if you want.