

Analyzing Crowded Housing in Chicago Community Areas

Erica Castillo, Shane Gerami, Abdallah Hasan, Cindy Rocha, Samin Sohrabi





Descriptive Statistics

What is the mean, median, first, second, and third quartile of your assigned variable?

Mean: 4.91

Median: 4.2

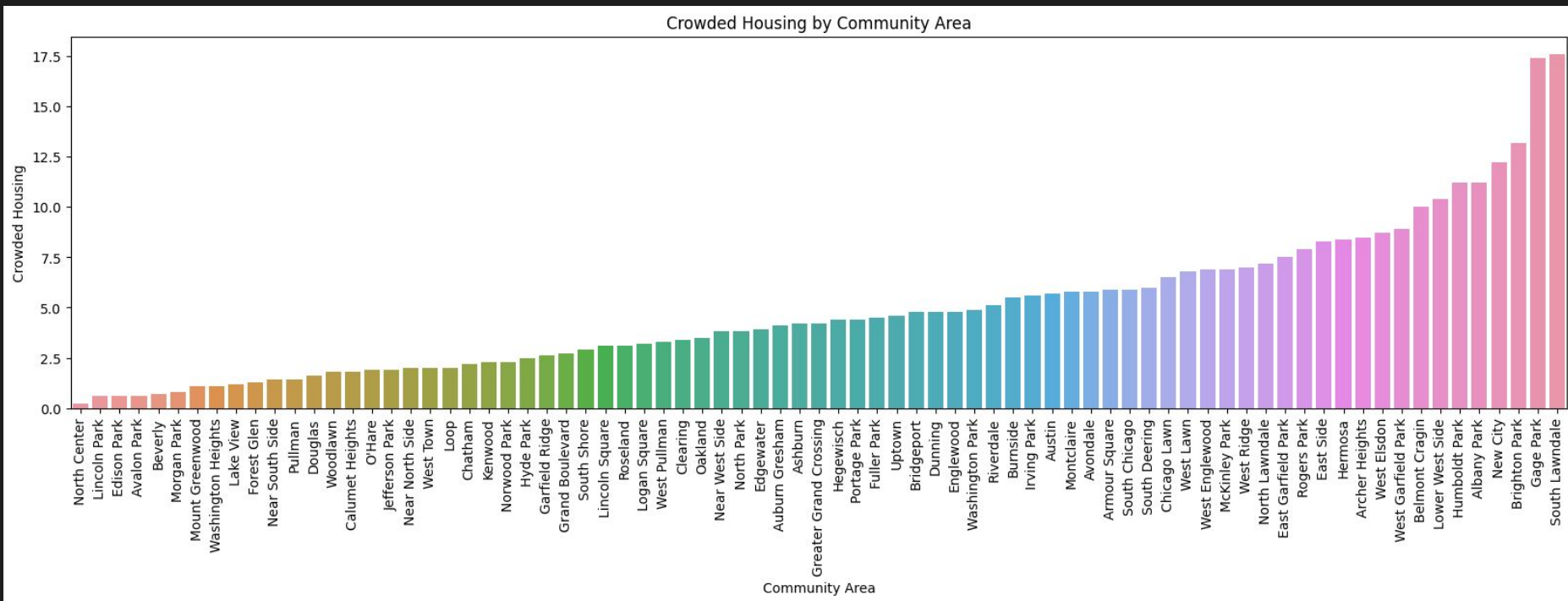
First quartile: 2.0

Second quartile: 4.2

Third quartile: 6.8

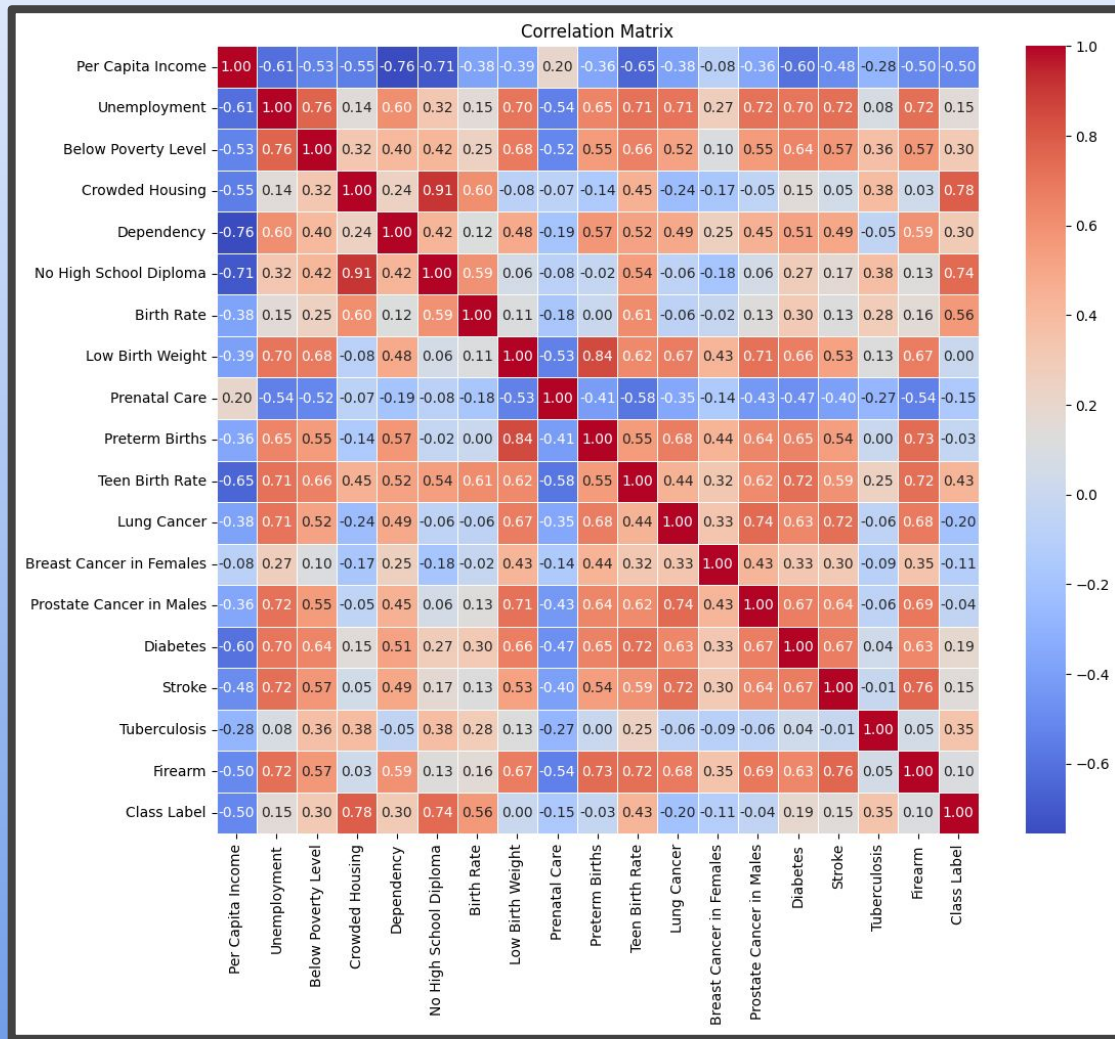


Data Visualization





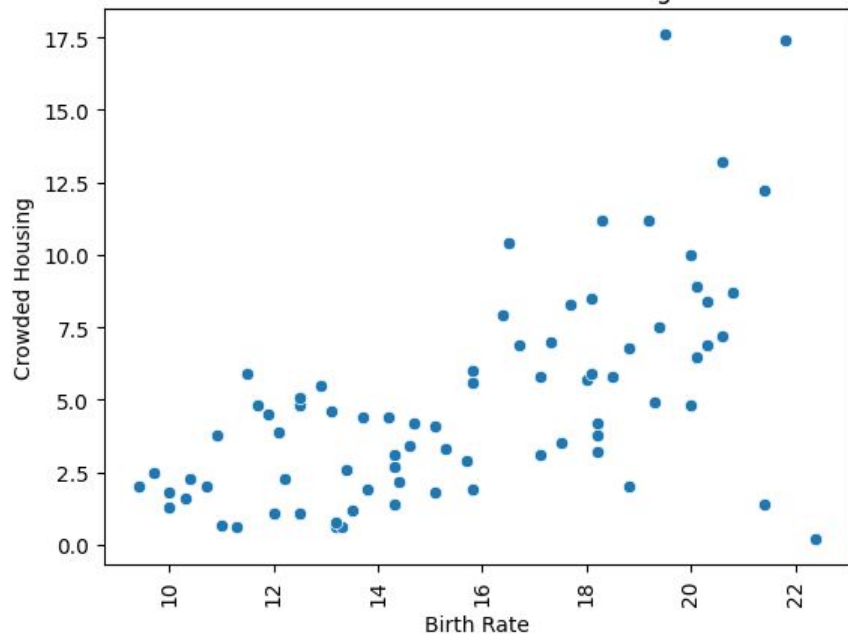
Correlation Matrix



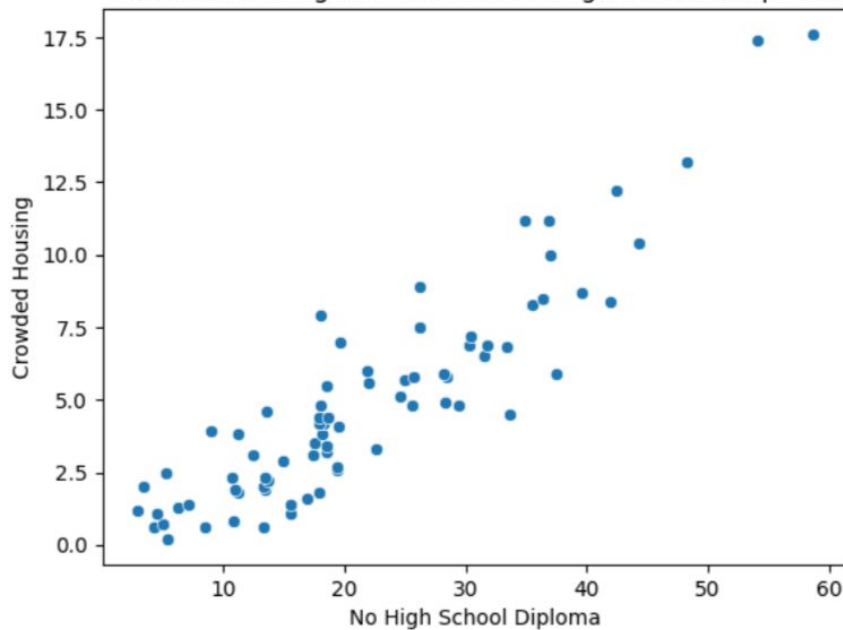


Data Visualization

Birth Rate Vs Crowded Housing

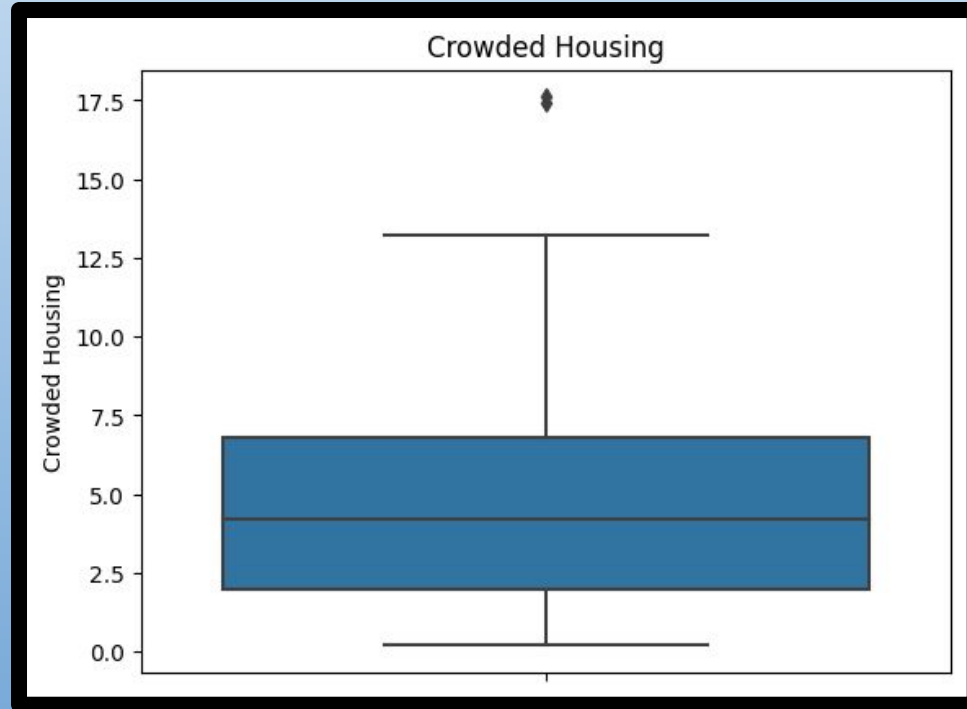


Crowded Housing as a Result of No High School Diploma





Data Visualization





Classification

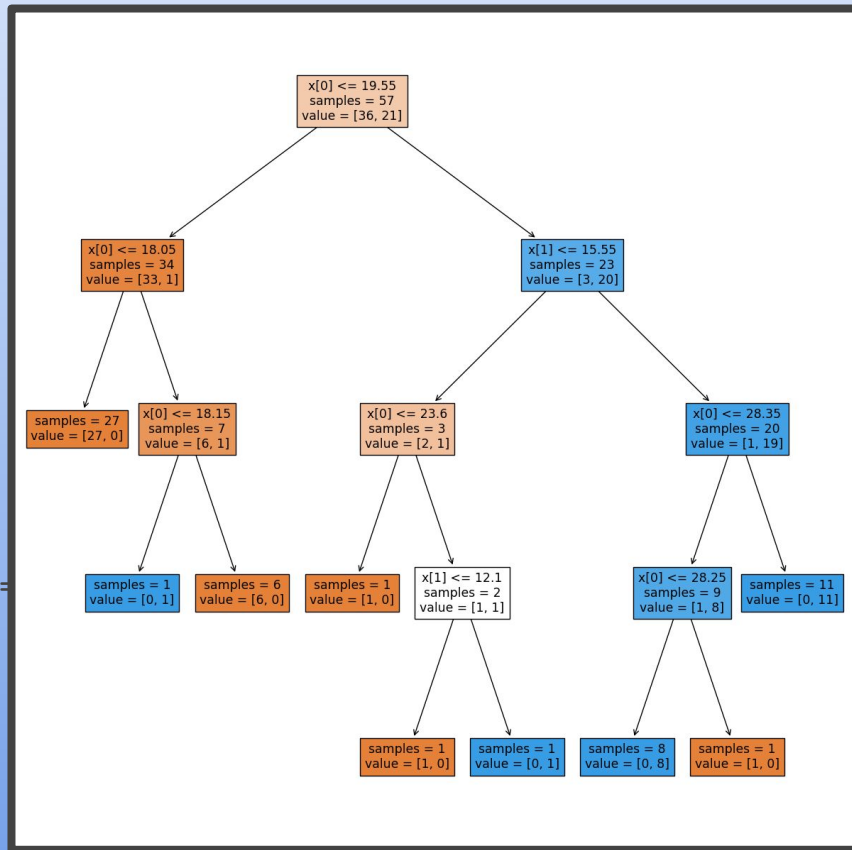
What predictors did you choose?

- No High School Diploma & Birth Rate

How did you choose the predictors?

- Came from our Data Visualization
 - The most related to our assigned variable
- No High School Diploma = Less ability to afford a bigger house = Crowded Housing
- Birth Rate = More people per less space = Crowded Housing

Plot the decision tree





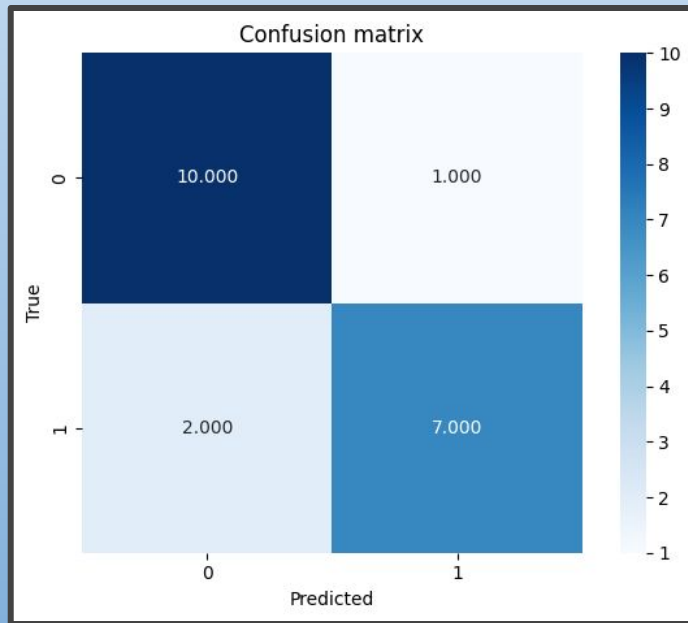
Classification

Plot the confusion matrix

What is the accuracy and error of the decision tree on the test set?

Accuracy = .85

Error = .15





Conclusions

What can you conclude about your assigned variable after analyzing the dataset?

Our assigned variable was Crowded Housing, and we can conclude that Crowded housing is a half decent predictor of No high school diploma and birth rate

Looking at scatterplot and correlation matrix helps find what the variable is correlated with