

Project 02.1: Boston House Prices

Jacob Cluff

March 2019

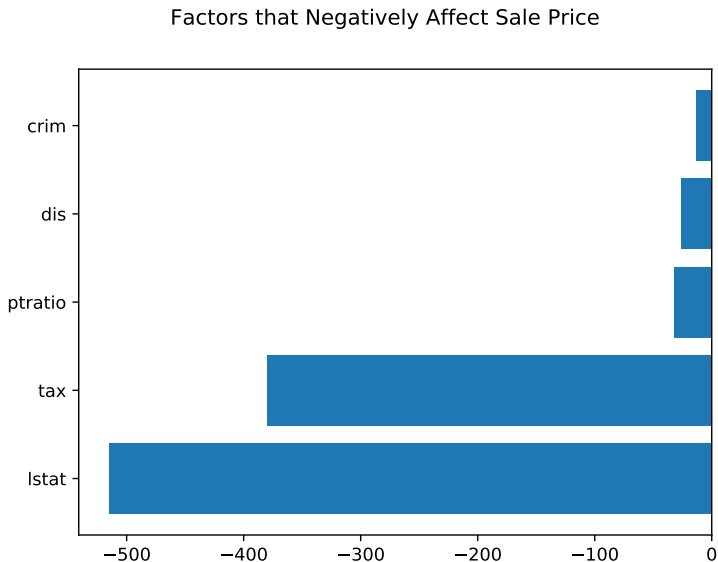
Terms

- ▶ crim: per capita crime rate by town
- ▶ zn: proportion of residential land zoned for lots over 25,000 sq.ft
- ▶ indus: proportion of non-retail business acres per town
- ▶ chas: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
- ▶ nox: nitrogen oxides concentration (parts per 10 million)
- ▶ rm: average number of rooms per dwelling

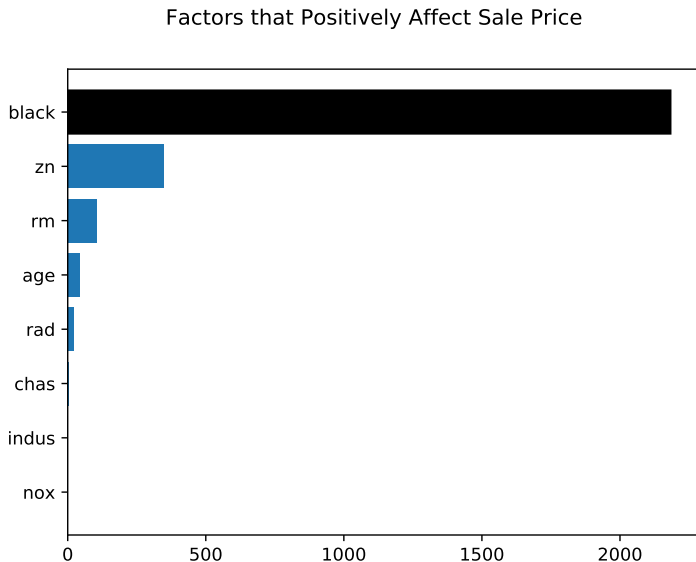
Terms (cont')

- ▶ age: proportion of owner-occupied units built prior to 1940
- ▶ dis: weighted mean of distances to five Boston employment centres
- ▶ rad: index of accessibility to radial highways
- ▶ tax: full-value property-tax rate per \$10,000
- ▶ ptratio: pupil-teacher ratio by town
- ▶ black: $1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town
- ▶ lstat: lower status of the population (percent)
- ▶ medv: median value of owner-occupied homes in \$1000s

Factors that Negatively Affect House Values



Factors that Positively Affect House Values



A Synopsis of the Factors Affecting Sale Price

Factor	
crim	-12.878698
zn	346.495384
indus	0.740451
chas	2.479102
nox	0.385008
rm	106.203218
age	44.423986
dis	-26.358841
rad	21.949803
tax	-380.071080
ptratio	-32.118406
black	2183.254644
lstat	-515.200294

High Income House vs. Low Income Status

On one hand, it seems like the more expensive houses suffer a penalty on taxes. On the other hand, the sale price is also adversely effected by higher percentages of lower income houses. With some additional analysis, we should be able to find a target price range to focus on to maximize profits.

Expanding Suburbs

Since the distance from radial highways does not seem to affect the sale price much, developing the suburbs may be the right Business. To expand on this idea, we would need to determine the price difference between real estate in developed and non-developed areas, then we would have to determine if the development cost and minor hit in sale cost would be worth exploiting.

How Does Race Affect Sale Price?

Do African Americans systematically pay more for houses, or does having a high population of African Americans in the area positively affect sale prices?

Since this seems to be the largest factor increasing the sale price of Houses, it may be worth looking into. To Look into this, we would need data correlating the race between buy/sell transactions of Houses.

Technical Notes

Training R^2	Validation R^2	Testing R^2
0.762	0.776	0.584

- ▶ Used train-validate-test method with 60-20-20 split.
- ▶ Using k-folds validation may improve testing R^2 and result in a more stable model output.
- ▶ The model does a random shuffle of the data every time it loads and it results in a wide range of resulting R^2 values for the test data. That and the fact that there are a few factors that are essentially zero hint that the model might be over-fit and should look into removing some features and/or increasing regularization.

Normalized and Denormalized Weights

	Normalized	DeNormalized
bias	13.684331	1.000000
crim	-0.281779	-12.878698
zn	3.647320	346.495384
indus	0.000017	0.740451
chas	2.479102	2.479102
nox	0.000017	0.385008
rm	21.220223	106.203218
age	0.408766	44.423986
dis	-2.872138	-26.358841
rad	0.910861	21.949803
tax	-1.086178	-380.071080
ptratio	-5.199815	-32.118406
black	5.540810	2183.254644

Comparing Actual Sale Prices to the Modal Sale Prices

