**Team Alpha Review**

Page 2: Please provide a little more detail about why you excluded some variables. It makes sense to exclude a variable if it is essentially a copy of another variable. However, we don’t want to remove a variable simply because we think it won’t be important.

Data Section: I think your discussion regarding the distribution of some of your explanatory variables should come before you fit your first model. This will give your writing a proactive approach, as if you are fixing issues before you model, rather than reacting to problems after the fact. In reality, we ALL are reacting to problems after the fact, but when writing we have the luxury of making the story flow seamlessly.

Interaction terms: Your reason for the crime/age interaction needs to be adjusted. You make it sound as though you were hoping that building age could help predict crime, but the interaction should have everything to do with how each variable (age or crime) affects the other variables relationship with price. This needs to be clarified in the text.

Figure 7 says it includes the VIF, but it does not. Either the figure or the caption need to be updated.

Page 9: In your final model, please organize your variables in a more intuitive way. All the borough variables should be grouped together. Same would go for seasons.

Page 9: While the raw leverage values are small, they are still way above the threshold. Your assurance that you don’t have influential points needs to be removed from the text. You also have several extreme outlier residual values. I doubt you will be able to fix them, but the outlier values need to be better addressed in the text.

Page 11: Try to expound a little bit on your regression tree results. Which variables were most important for prediction in the tree? Walk the reader the first few tree splits so that they know how predictions work.

Page 12: Please do look more into the poor performance of the tree and see what you can uncover. It seems shocking to me that the tree would do worse than no model. It could be an overfitting problem like you mentioned, but it could also be an issue with how the prediction error is being calculated, or how the data were transformed.