

Questions/Answers for Predicting Property Value

Ten Audience Questions

- Can you clearly declare what features drive property value?
- If you omitted so many rows from your dataset, what makes you believe that the remaining rows are reliable?
- What are other possible features which drive value?
- Can you generalize from the one year of data that you gathered for this year? And for upcoming years?
- Who else aside from real estate buyers and sellers can use this model for profit?
- How do you plan to keep stakeholders informed and educated about your black-box model?
- What did PCA produce and why did you not use it?
- How did you prepare your data for your models?
- How did you deal with categorical variables and why?

- How does accurate real estate property value predictions make the market more robust?

Ten Answers

- GROSS SQUARE FEET is the greatest driver of property value, as is evident in its correlation with SALE PRICE (0.52). No other feature comes close. A mixture of other features like units and borough contribute together to drive property value.
- I deliberated whether to include or exclude each row. Therefore, the rows which I included were verified to be good data. Moreover, the plethora of rows which I had to omit from the models were truthful data. It was just not useful to this study. Sales without considerations are a reality but are irrelevant because I am modeling for sales with considerations which are the barometer of property value.
- Other features to explore are proximity to amenities and schools. Crime rates and neighborhood trends presumably contribute to price.
- Real estate market conditions are dynamic and require one to carefully consider how to apply the present project to another year which brings its own drivers of price. For starters interest rates are expected to fluctuate. One could generalize the patterns and trends from the modeled year to other years.
- Lenders can leverage accurate property values when determining how much they can loan an owner of such a property and what is the level of risk of default manifest in the interest rate.
- The data team must maintain open lines of communication to consistently update and reinforce in laymen's terms the mechanics and the limitations of the model. We must educate the stakeholder about how to leverage the insight given by the model.
- I created PCA analyses and built models with the principal components, but the models were not good. Using regular features led to more robust models.
- I extracted only the rows which included sales with consideration because those rows are the only pertinent ones for predicting values. I carefully assessed the outlier of the \$2 Billion sale and confirmed its veracity.
- I chose one-hot encoding, and not label encoding, for instance, because the categories are nominal. Encoding enables one to use ML models on categorical variables.
- Clarity builds trust. Accurate predictions lead to better-informed decisions,

which create a more stable real estate market.