Developing in King County

A Study of Housing Prices from 2014-2015

Dataset Description & Observations

- Began with a dataset of 20k+ observations.
- In addition to the sale price, observations included data about the house's location and neighborhood, layout, when it was built and if it was renovated, and its condition.
- Noted Quirks:
    - There are a handful of houses that appeared multiple times in the dataset, indicating that they had been sold multiple times.
    - Another handful of houses appeared to have been built or renovated after the date of sale.

Notable Data Manipulation

- Dropped 60 observations which were missing values for the number of times the house was viewed before sale.

- Dropped another ~50 observations that were deemed to be extreme outliers (defined as having a z-score of over 10).

- Created two location-based categorical variables:
  - Using zipcodes, we identified those homes in the Medina area--a notably rich and attractive area outside of Seattle.
  - Using longitude & latitude, we identified those houses which were within 7.5 miles of downtown Seattle, the primary business center of the county.

Parameter Selection

- We started with 21 parameters, and then used the backwards stepwise algorithm in order to identify the subset of parameters that resulted in a regression model that best fit our business goal.
  - The backwards stepwise algorithm reduces the chances of overfitting
  - It provides a easily readable result that can be interpreted clearly and concisely.
- An initial subset of models was selected using R-Squared metric.
- With this subset, we ran cross-validation tests (k=10) to determine the optimal model.
- The final subset of parameters that we ended up with after the stepwise algorithm was:
  - Layout-Related (6): Living Space square footage (squared as well), amount of space above ground, and the number of bedrooms, bathrooms and floors
  - Location-Related (4): Binary indicators for a waterfront property, located near Seattle's downtown area, and located in Medina; and a measure of the house's square footage relative to its neighbors.
  - Condition-Related (7): Housing grade (a measure of workmanship and material quality), its condition, its age and its renovation status.
  - Seasonal Adjustment (1) for sales completed during Spring season.
- The model's test scores ranged from 0.77 to 0.69.

Model Results & Interpretation

Our 14-parameter model, once again trained on the entire dataset returned a score of 0.75. However, the more result are the coefficient values our model returned.

Using the regression output coefecients, we can come to three important conclusions:

1. One of the largest factors that affects price is square footage of living space, not necessarily lot space. Living space design is key to capitalize on this trend. More rooms and/or floors does not necessarily lead to a higher price, rather living space ALONE. This would indicate that a house with a more streamlined design that maximizes open space would be more attractive than one which feels more cluttered. It is also worth considering that building these kinds of open designs are usually cheaper to build.

2. According to the regression coefficients, location is undisputedly the most influential factor when it comes to the sale price of a home. Houses built in Seattle and/or Medina have consistently higher sale prices than that of those located elsewhere. The sale price is even more positively affected if built on a location with a waterfront view.

3. Quality matters. After location and living space design, the regression model highlights the two measures of quality--condition of the house and the grade (which according to King County website, is a measure of workmanship and material quality).

Conclusions & Further Considerations

The development company this project was created for should try their best to use high quality materials and skilled workers to build/renovate a home located in Medina, Seattle or other areas that attractive to residents. This 'attractiveness' can be evaluated through other means, such as education, landmarks, etc.

The company should also try their best to secure lots that have a waterfront view in order to maximize the potential sale price of their home.

Additional research can be conducted by looking closely at price per square foot of living space.