Goal: Analyze the factors that affect the price of real estate

Target Variable: Sales Price

Predictors:

1. Year Built
2. Lot Size
3. Finished Square Feet
4. Garage Size
5. Quality
Overview of the data

- 522 observations
- 10 variables
- No missing values
- Several Outliers
  - Apply Anomaly node to discard outliers
Basic Statistics of Target Variable

- Average: $277,894
- Max: $920,000
- Min: $84,000
SQRT
Correlation

- Positive Correlations between Predictors and Target Variable
- Finished square feet has the largest correlation with sales price
- Both models are not overfitted
- Model with lower validation data has higher quality
- Finished square feet has the most significant effect on sales price.
- Garage size has the least significant effect on sales price.
- Finished square feet, Quality=Best, Year built, Lot size and Garage size have positive impact on target variable.
Cluster Analysis (Two Step)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>cluster-1</th>
<th>cluster-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>88.8%</td>
<td>11.2%</td>
</tr>
<tr>
<td>Inputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sales Price</td>
<td>241,731.19</td>
<td>577,980.80</td>
</tr>
<tr>
<td>Finished square feet</td>
<td>2,105.69</td>
<td>3,517.89</td>
</tr>
<tr>
<td>Garage size</td>
<td>1.98</td>
<td>2.93</td>
</tr>
<tr>
<td>Year built</td>
<td>1,965.61</td>
<td>1,989.02</td>
</tr>
<tr>
<td>Lot size</td>
<td>22,666.05</td>
<td>27,015.80</td>
</tr>
</tbody>
</table>
Business applications

Decision Tree: 5 rules

1. If Sales price \( \leq 200,375 \), Finished square feet \( \leq 1668.500 \), then the quality of the house is Fair.
2. If Sales price \( \leq 200,375 \), Finished square feet \( > 1668.500 \) and Year built \( \leq 1955.500 \), then the quality is Fair.
3. If Sales price \( \leq 200,375 \), Finished square feet \( > 1668.500 \) and Year built \( > 1955.500 \), then the quality is Moderate.
4. If Sales Price >200,375, and \( \leq 489,750 \), then the quality is Moderate.
5. If Sales Price >489,750, the quality of the house is Best.