A comparison of a graph

Description automatically generated

The results for the Los Angeles home price reduction prediction model tell us several important things about its performance:

### Accuracy Score

- The \*\*accuracy score\*\* of 0.71875 (or approximately 72%) indicates that, overall, the model correctly predicted whether a home's price was cut in about 72% of the cases in the test dataset. This means out of all predictions made, 72% were correct, regardless of the class (price cut or no price cut).

### Confusion Matrix

- The \*\*confusion matrix\*\* provides a more detailed breakdown of the model's predictions:

- \*\*True Negatives (TN): 19\*\* - The model correctly predicted 19 instances where the price was not cut.

- \*\*False Positives (FP): 7\*\* - The model incorrectly predicted 7 instances as price cuts when they were not.

- \*\*False Negatives (FN): 2\*\* - The model missed 2 instances where the price was cut, predicting them as no price cut.

- \*\*True Positives (TP): 4\*\* - The model correctly identified 4 instances where the price was cut.

- This matrix helps us understand not just the accuracy, but how the model is accurate, revealing a stronger performance in predicting no price cuts over price cuts.

### Classification Report

- The \*\*classification report\*\* dives deeper into the model's performance for each class (0 for no price cut, 1 for price cut):

- For \*\*Class 0 (no price cut)\*\*, the model has a \*\*precision of 0.90\*\*, meaning that when it predicts a property's price won't be cut, it is correct 90% of the time. The \*\*recall of 0.73\*\* indicates that it correctly identifies 73% of all actual no price cut instances. The \*\*F1-score of 0.81\*\* is a harmonic mean of precision and recall, suggesting a balanced performance for this class.

- For \*\*Class 1 (price cut)\*\*, the model has a \*\*precision of 0.36\*\*, showing it is correct only 36% of the time when predicting a price cut. The \*\*recall of 0.67\*\* means it is able to identify 67% of the actual price cut instances. The \*\*F1-score of 0.47\*\* indicates a lower performance for predicting price cuts compared to no price cuts.

- The \*\*macro average\*\* provides an unweighted mean of the precision, recall, and F1-score across both classes, which gives a general idea of the model's overall performance without taking class imbalance into account.

- The \*\*weighted average\*\* takes the support (the number of true instances for each class) into account, offering a measure that considers the imbalance between classes. Here, the weighted averages suggest the model performs better than the macro averages might imply, primarily because it does well on the more common class of no price cuts.

### Key Insights

- The model is significantly better at predicting when a property's price will not be cut than when it will be cut. This is evident from the higher metrics across the board for the no price cut class.

- The relatively low precision for price cuts (Class 1) suggests a high rate of false positives, which could be problematic depending on the application. For example, if this model were used to advise buyers on which properties might experience a price cut, a lot of the advice might lead to waiting for price cuts that never happen.

- The high recall for price cuts, however, indicates that the model is relatively good at catching instances when they occur, despite its precision issues. This could be more useful for buyers willing to sift through more potential options to find those few genuine price cuts.

- The accuracy figure alone might seem decent, but the detailed metrics reveal that the model's performance is uneven across classes. Improving the model might involve addressing this imbalance, perhaps by gathering more data, using techniques like SMOTE for oversampling the minority class, or exploring different models or tuning parameters for better balance.

A blue and orange bar chart

Description automatically generated

The results for the New York City home price reduction prediction model provide insights into its performance in predicting whether the price of a property will be cut. Here's a breakdown:

### Accuracy Score

- The \*\*accuracy score\*\* is 0.8 (or 80%), indicating that the model correctly predicted the price cut status (whether a price was cut or not) in 80% of the cases in the test dataset. This suggests a relatively high level of overall predictive accuracy.

### Confusion Matrix

- The \*\*confusion matrix\*\* offers a detailed view of the prediction outcomes:

- \*\*True Negatives (TN): 18\*\* - The model correctly identified 18 instances where no price cut occurred.

- \*\*False Positives (FP): 2\*\* - In 2 cases, the model incorrectly predicted a price cut when there was none.

- \*\*False Negatives (FN): 4\*\* - The model failed to identify 4 actual price cuts, predicting no price cut instead.

- \*\*True Positives (TP): 6\*\* - It correctly predicted 6 instances where a price cut did occur.

- This matrix reveals the model's strengths and weaknesses in predicting each class, showing a better capability in identifying properties without price cuts than those with.

### Classification Report

- The \*\*classification report\*\* delves into metrics for each class (0 for no price cut, 1 for price cut):

- \*\*Class 0 (no price cut)\*\* has a \*\*precision of 0.82\*\*, indicating a high level of accuracy when the model predicts no price cut. The \*\*recall of 0.90\*\* suggests the model is very effective at capturing the majority of actual no price cut instances. The \*\*F1-score of 0.86\*\* highlights a strong balance between precision and recall for this class.

- \*\*Class 1 (price cut)\*\* shows a \*\*precision of 0.75\*\*, meaning when the model predicts a price cut, it is correct 75% of the time. The \*\*recall of 0.60\*\* indicates it correctly identifies 60% of actual price cuts. The \*\*F1-score of 0.67\*\* reflects a reasonable, though not outstanding, balance between precision and recall for predicting price cuts.

- The \*\*macro average\*\* metrics give an unweighted average across both classes, offering a sense of overall performance without considering class imbalance.

- The \*\*weighted average\*\* takes the number of instances in each class into account, showing a more nuanced performance metric that reflects the actual distribution of classes. Here, the weighted averages closely align with the overall accuracy, indicating consistent performance across the classes.

### Key Insights

- The New York City model demonstrates good accuracy and is more balanced in its predictive capability across classes compared to the Los Angeles model. It shows a relatively strong ability both to identify properties where the price will not be cut and to predict price cuts, albeit with slightly less accuracy for the latter.

- The lower false positive rate for New York City (compared to Los Angeles) suggests a more conservative approach in predicting price cuts, which could be beneficial in applications where false alarms are costly.

- The difference in recall rates between the two classes suggests room for improvement in identifying price cuts more reliably. Strategies might include collecting more balanced data, employing advanced modeling techniques, or applying targeted feature engineering to better capture signals associated with price reductions.

- Overall, the New York City model is quite effective but, like any model, can benefit from further tuning and validation to ensure its predictions are as reliable and useful as possible in practical scenarios.

A graph of different colored squares

Description automatically generated

The results for the Miami real estate market prediction model, regarding price reductions, indicate a perfect performance across all metrics. Here's a detailed explanation:

### Accuracy Score

- The \*\*accuracy score is 1.0 (or 100%)\*\*, meaning the model correctly predicted every instance in the test dataset regarding whether or not there would be a price cut. This level of accuracy is exceptionally high and somewhat unusual in real-world applications, suggesting that the model was able to perfectly distinguish between properties that experienced a price cut and those that did not.

### Confusion Matrix

- The \*\*confusion matrix\*\* confirms the perfect predictive performance:

- \*\*True Negatives (TN): 14\*\* - The model correctly identified all 14 instances where no price cut occurred.

- \*\*False Positives (FP): 0\*\* - There were no instances where the model incorrectly predicted a price cut.

- \*\*False Negatives (FN): 0\*\* - The model also did not miss any actual price cuts; it did not predict a no price cut when there was one.

- \*\*True Positives (TP): 3\*\* - It correctly predicted all 3 instances where a price cut did occur.

- This matrix shows that the model achieved perfect classification with no errors in predictions.

### Classification Report

- The \*\*classification report\*\* provides perfect metrics for both classes (0 for no price cut, 1 for price cut):

- \*\*Precision, Recall, and F1-score are all 1.00 for both classes\*\*, indicating flawless performance. Precision measures the accuracy of positive predictions (how many predicted price cuts were correct), recall measures the ability to find all positive instances (how many actual price cuts were correctly identified), and the F1-score provides a balance between precision and recall.

- \*\*Macro and weighted averages\*\* are also 1.00, reflecting the model's overall perfect performance across different measures of accuracy, precision, recall, and F1-score.

### Key Insights

- Achieving a 100% accuracy and perfect precision, recall, and F1-scores across all categories is very rare in machine learning, especially on unseen data. It suggests that the model was exceptionally well-suited to the patterns present in the Miami dataset.

- While impressive, such perfect results may also prompt skepticism, especially in complex real-world tasks like predicting real estate price reductions. It's essential to consider whether the test dataset was representative of the broader data or if there were unique circumstances that made the prediction task particularly straightforward for this dataset.

- Overfitting could be a concern with perfect scores, where the model may have learned the training data too closely, including its noise and outliers, rather than generalizing from the underlying patterns. However, since the results are based on the test set, this scenario suggests that the model genuinely performed well on the given task.

- Before deploying this model in a broader context, it would be prudent to validate its performance on a larger and more diverse dataset to ensure its findings are robust and applicable in varied scenarios within the Miami real estate market.

In summary, the Miami model's results are extraordinary, showcasing a model that perfectly predicts price reductions in the dataset provided. Further validation would be essential to ensure its applicability and reliability in broader real-world applications.

A graph of different colored squares

Description automatically generated with medium confidence

The results for the Chicago real estate market prediction model regarding price reductions indicate a significantly low performance across most metrics. Here's a breakdown of what each component of the result signifies:

### Accuracy Score

- The \*\*accuracy score is approximately 0.22 (or 22%)\*\*, indicating that the model correctly predicted only about 22% of the instances in the test dataset regarding whether or not there would be a price cut. This low accuracy suggests that the model struggled significantly to predict price cuts accurately for properties in the Chicago dataset.

### Confusion Matrix

- The \*\*confusion matrix\*\* provides insight into the type of errors made:

- \*\*True Negatives (TN): 2\*\* - The model correctly identified 2 instances where no price cut occurred.

- \*\*False Positives (FP): 2\*\* - There were 2 instances where the model incorrectly predicted a price cut when there wasn't one.

- \*\*False Negatives (FN): 5\*\* - The model missed 5 instances where there actually was a price cut, predicting no price cut instead.

- \*\*True Positives (TP): 0\*\* - It failed to correctly predict any instances where a price cut did occur.

- This matrix shows that the model was notably more likely to miss actual price cuts (False Negatives) and was not effective in correctly identifying when price cuts would happen.

### Classification Report

- The \*\*classification report\*\* reveals critical weaknesses in the model's performance:

- \*\*Precision for class 1 (price cut) is 0.00\*\*, meaning the model did not correctly predict any price cuts. Precision measures the accuracy of positive predictions.

- \*\*Recall for class 1 is also 0.00\*\*, indicating the model's inability to find any of the actual price cut instances. A recall of 0 means every instance that had a price cut was missed by the model.

- \*\*The F1-score for class 1 is 0.00\*\*, reflecting the poor balance between precision and recall for predictions of price cuts.

- \*\*For class 0 (no price cut), the precision, recall, and F1-score are relatively better but still indicate low performance.\*\*

- \*\*The macro and weighted averages\*\* further highlight the model's overall weak performance across these metrics.

### Key Insights

- The results suggest that the model was largely ineffective in predicting price reductions for the Chicago dataset, with significant issues in both identifying actual price cuts and avoiding false alarms.

- The notably low accuracy and complete failure to predict any true price cuts (True Positives) may indicate several potential issues, such as:

- \*\*Data quality or representativeness\*\*: The training data may not have been representative of the test set or lacked enough examples of price cuts.

- \*\*Model suitability\*\*: The decision tree classifier, as configured, might not be suitable for the patterns in the Chicago real estate market data.

- \*\*Feature selection and preprocessing\*\*: The chosen features and preprocessing steps (including one-hot encoding and handling of missing values) may not have been optimal for capturing the nuances required to predict price cuts accurately.

- Given these results, it would be crucial to revisit the data preprocessing steps, feature selection, and possibly consider alternative models or additional features that might capture the dynamics of the Chicago real estate market more effectively.

In summary, the Chicago model's performance is markedly poor, indicating significant room for improvement in model configuration, feature engineering, or even reevaluation of the data used for training and testing.

Based on the non-linear dataset we got from Zillow for the cities we chose to analyze, we wanted to know if the initial price offering will get reduced. For that we used supervised machine learning models like decision tree and random forest because we understand that these models are better suited for non-linear datasets.