### **Executive Summary Report**

**Team**: Foresight

#### **1. Objective**

The primary objective of this project was to predict the assessed value of properties in the dataset predict\_property\_data.csv as accurately as possible. Using historic\_property\_data.csv, we aimed to develop a robust machine learning model that minimizes the Mean Squared Error (MSE) of predictions compared to actual sales prices when these homes sell.

#### **2. Approach**

To achieve the objective, the following steps were undertaken:

**Data Preprocessing**:

* **Outlier Handling**: Sale prices above the 99th percentile were capped to reduce skewness and the impact of extreme values.
* **Log Transformation**: The target variable (sale\_price) was log-transformed to stabilize variance and improve model training.

**Feature Engineering**:

* New features were engineered based on domain knowledge:
  + **Building Coverage Ratio**: Ratio of building size to lot size.
  + **Room Density**: Ratio of the number of rooms to building size.
  + **Bath-to-Bedroom Ratio**: Number of full bathrooms per bedroom.
* Feature importance analysis was used to focus on high-impact predictors, improving model interpretability and performance.

**Model Training**:

* The **xgboost** algorithm was employed for its efficiency and suitability for regression tasks.
* Hyperparameter tuning was conducted using a grid search to optimize key parameters:
  + **Learning rate** (eta), **maximum tree depth** (max\_depth), **sampling fractions** (subsample and colsample\_bytree), and **minimum child weight** (min\_child\_weight).

**Evaluation**:

* Performance was assessed using multiple metrics:
  + **Mean Squared Error (MSE)**: Measures the average squared error of predictions.
  + **Root Mean Squared Error (RMSE)**: Represents the average prediction error in dollar terms.
  + **Normalized MSE**: Indicates error relative to the mean target value.
  + **Mean Absolute Percentage Error (MAPE)**: Reflects average percentage error relative to actual values.

#### **3. Findings**

The final tuned model achieved the following performance metrics on the validation set:

* **Normalized MSE**: 0.1126
* **MSE in dollars**: $9,763,295,913
* **RMSE**: $98,809
* **MAPE**: 32.64%
* **Accuracy**: 95% of predictions were within $200,000 of the actual value.

**Key Observations**:

* The most impactful features in the model were:
  1. **meta\_certified\_est\_bldg**: Certified building assessment value (highest importance).
  2. **meta\_certified\_est\_land**: Certified land assessment value.
  3. **econ\_tax\_rate**: Economic tax rate for the property.
  4. **char\_bldg\_sf**: Building square footage.
* **meta\_certified\_est\_bldg** contributed most to predictions, highlighting its importance in property valuation.

#### **4. Recommendations for CCAO**

Based on the model's findings, we recommend the following actions for the Chief County Assessment Office (CCAO):

1. **Focus on Building Value Estimates**:
   * Given the high importance of **meta\_certified\_est\_bldg**, ensure that building value assessments are accurate and reflect market trends.
2. **Neighborhood-Level Adjustments**:
   * Incorporate neighborhood-specific features (e.g., median income, amenities) to refine property value predictions further.
3. **Monitor Age-Related Value Depreciation**:
   * The impact of property age on value should be closely monitored, with adjustments for depreciation incorporated into assessments.
4. **Automate Validation Checks**:
   * Implement automated systems to flag discrepancies in property assessments, ensuring consistency and accuracy.

#### **5. Appendix Figures**

The following visualizations support the findings and methodology:

1. **Feature Importance Plot**: Highlights the most influential predictors in the model.
2. **Distribution of Prediction Errors**: Illustrates the spread of errors across properties.
3. **Actual vs Predicted Values Scatter Plot**: Compares model predictions to actual property values.
4. **Training Convergence Plot**: Shows how the model's error metrics evolved during training.

### **Conclusion**

This project successfully developed a predictive model for property assessments, achieving high accuracy and interpretability. The recommendations outlined above provide actionable insights for improving the valuation process, with a focus on leveraging data-driven techniques to enhance assessment accuracy and fairness.

