NYC Property Sales Analysis

Matthew Scott

March 20, 2025

Contents

Executive Summary

This report presents a comprehensive analysis of New York City property sales data, focusing on residential properties across different boroughs. We employ various machine learning models to predict property prices and analyze market trends.

Data Overview

The analysis utilizes two primary datasets: 1. NYC Property Sales data (`nyc-rolling-sales.csv`) 2. NYC Building Codes (`NYC_Codes.csv`)

Data Summary
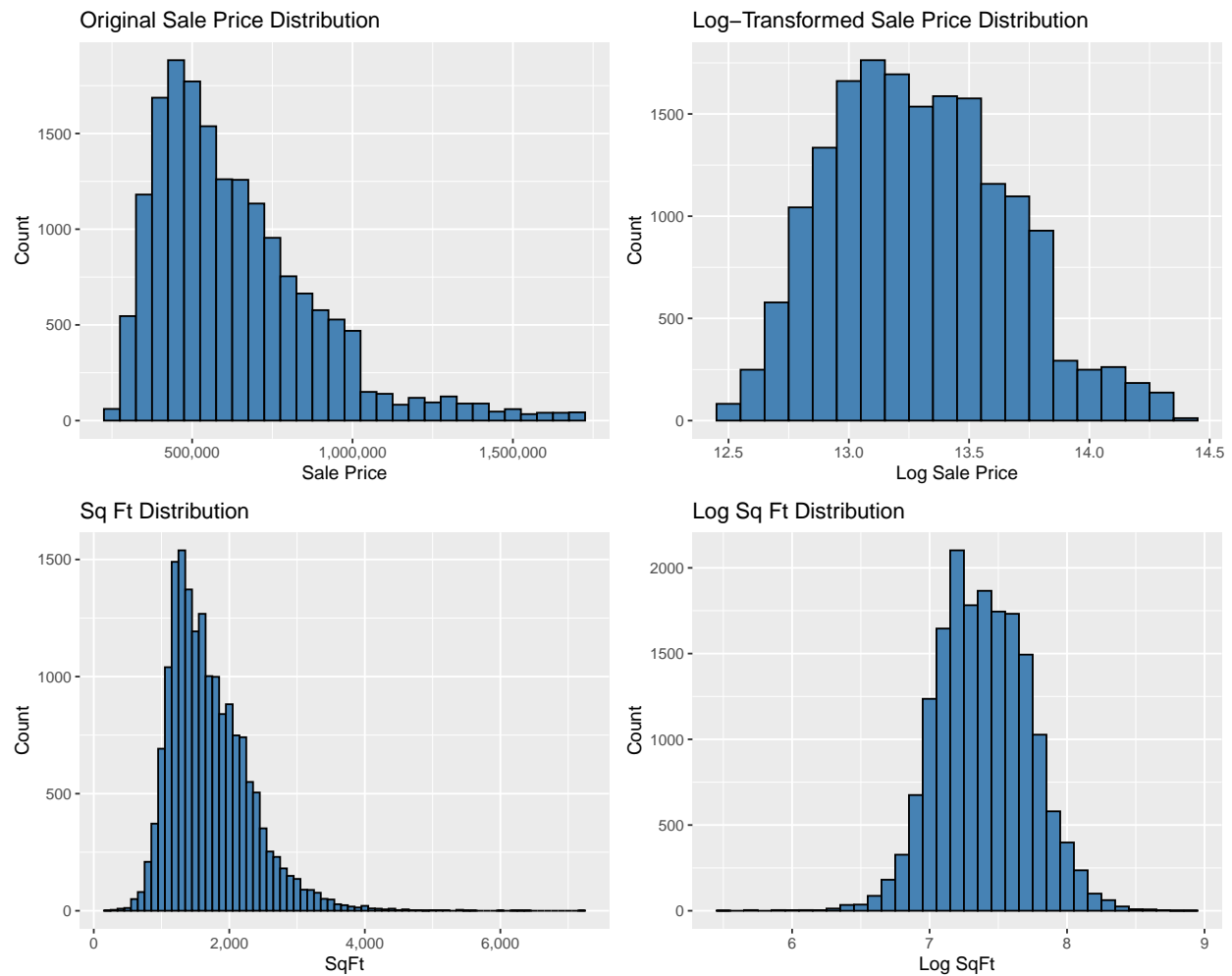
Table 1: Summary Statistics

Total_Properties	Avg_Price	Median_Price	Avg_SqFt	Median_SqFt
17,425	637,567.6	579,000	1,715.304	1,600

Table 2: Borough-wise Summary

borough_name	Properties	Avg_Price	Median_Price
BRONX	1,904	463,545.6	454,475
BROOKLYN	4,379	825,154.5	750,000
QUEENS	7,333	638,022.6	625,000
STATEN ISLAND	3,809	508,021.2	495,000

Exploratory Data Analysis
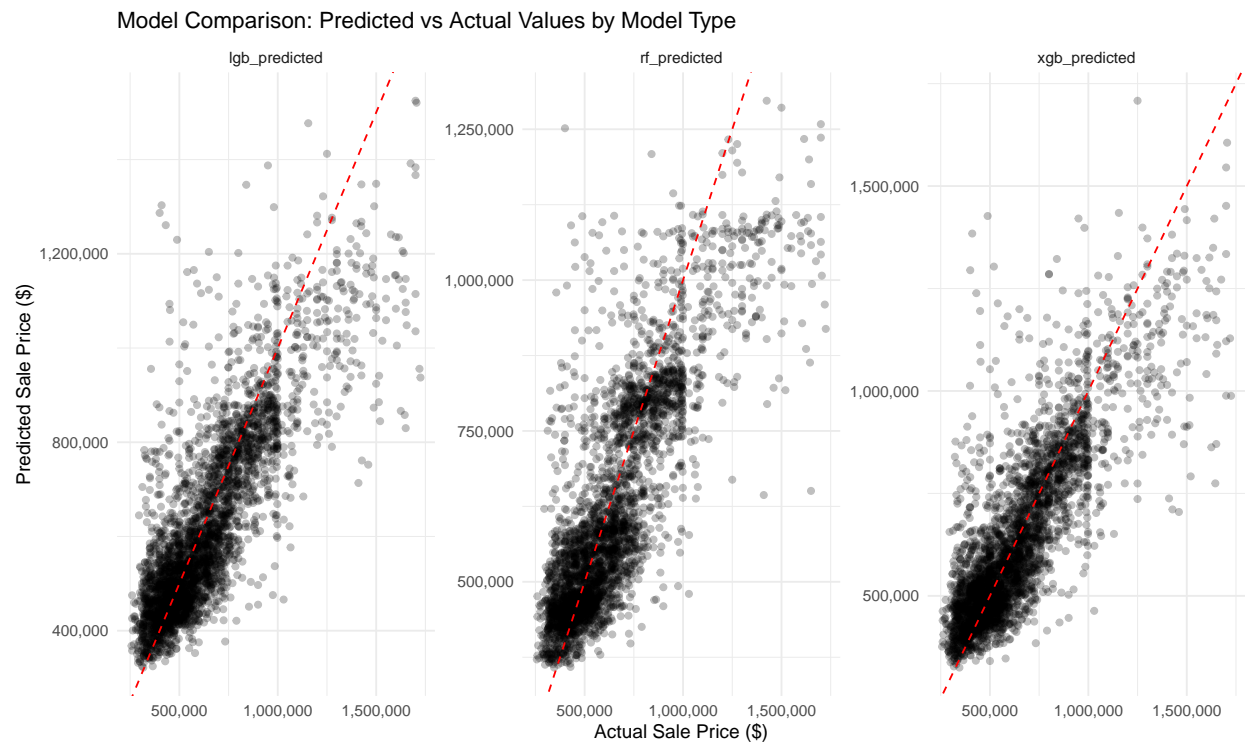
Price Distributions

Geographic Analysis

Price Distribution by Borough

Sale Price Distribution by Neighborhood and Borough

3
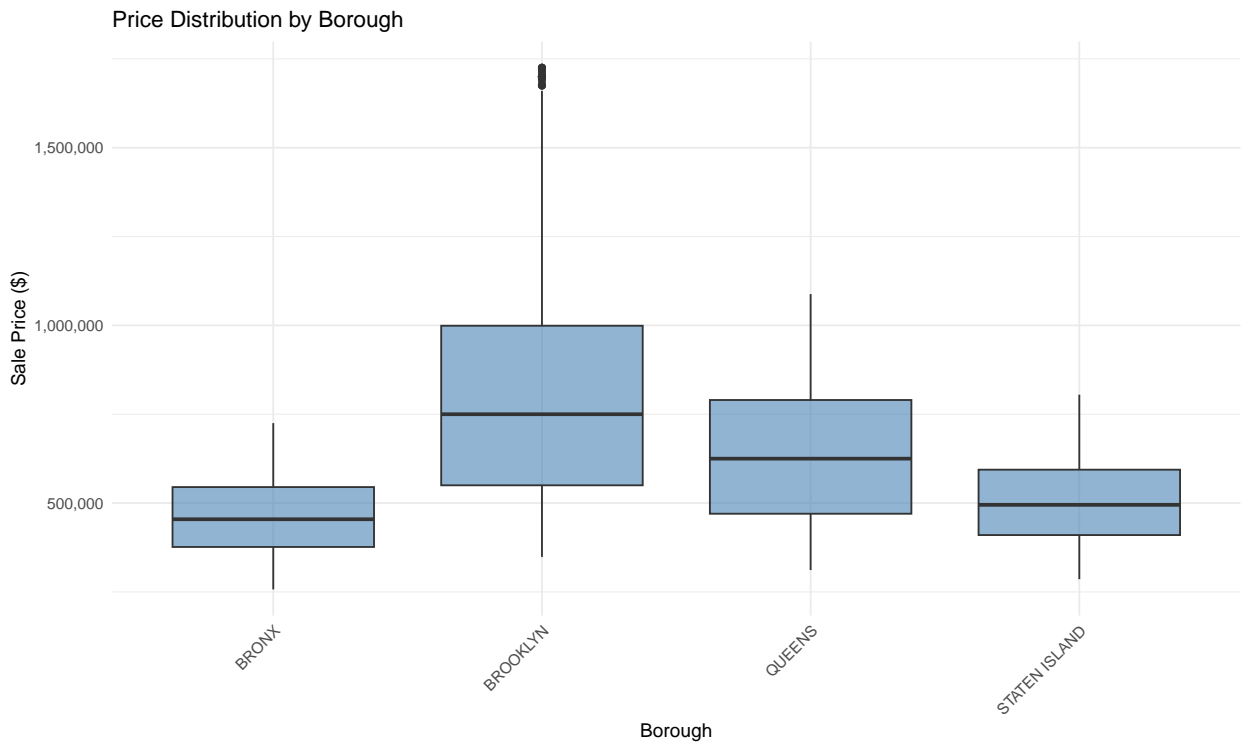
Model Performance Analysis

Model Comparison

Model Comparison: Predicted vs Actual Values by Borough

Model Comparison: Predicted vs Actual Values by Model Type

Error Analysis
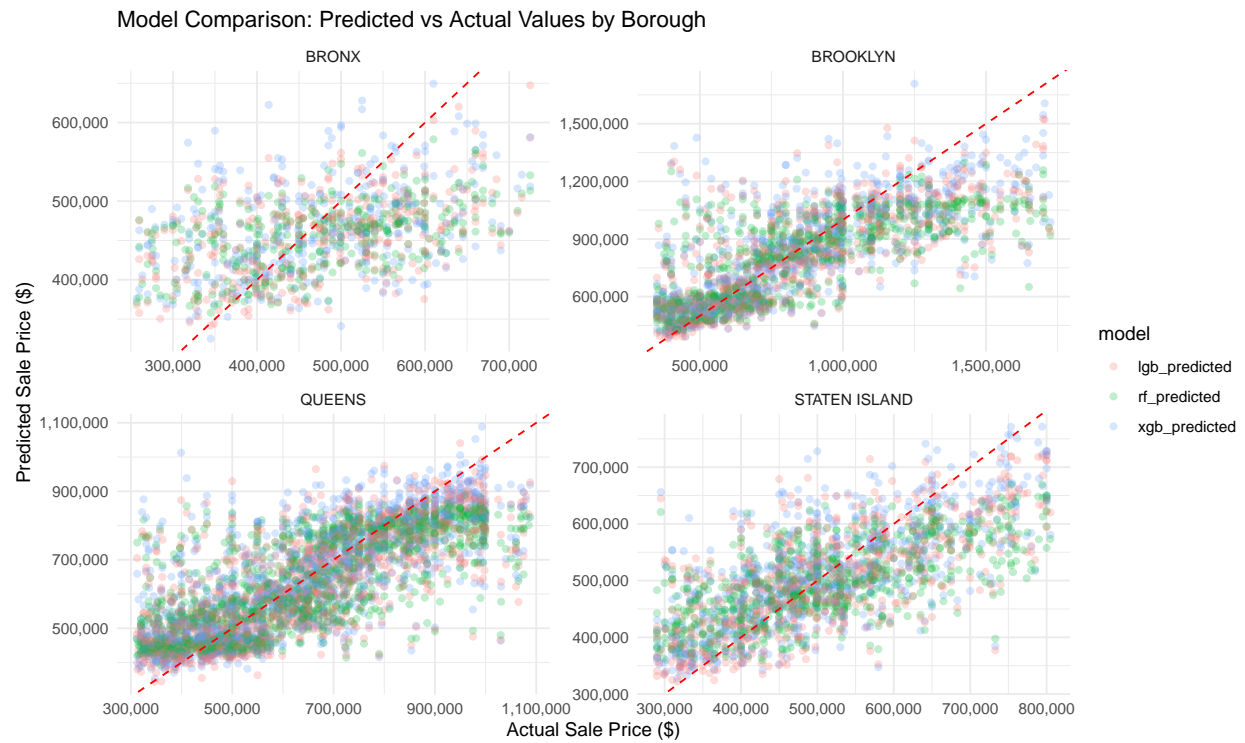
Error Distribution by Model
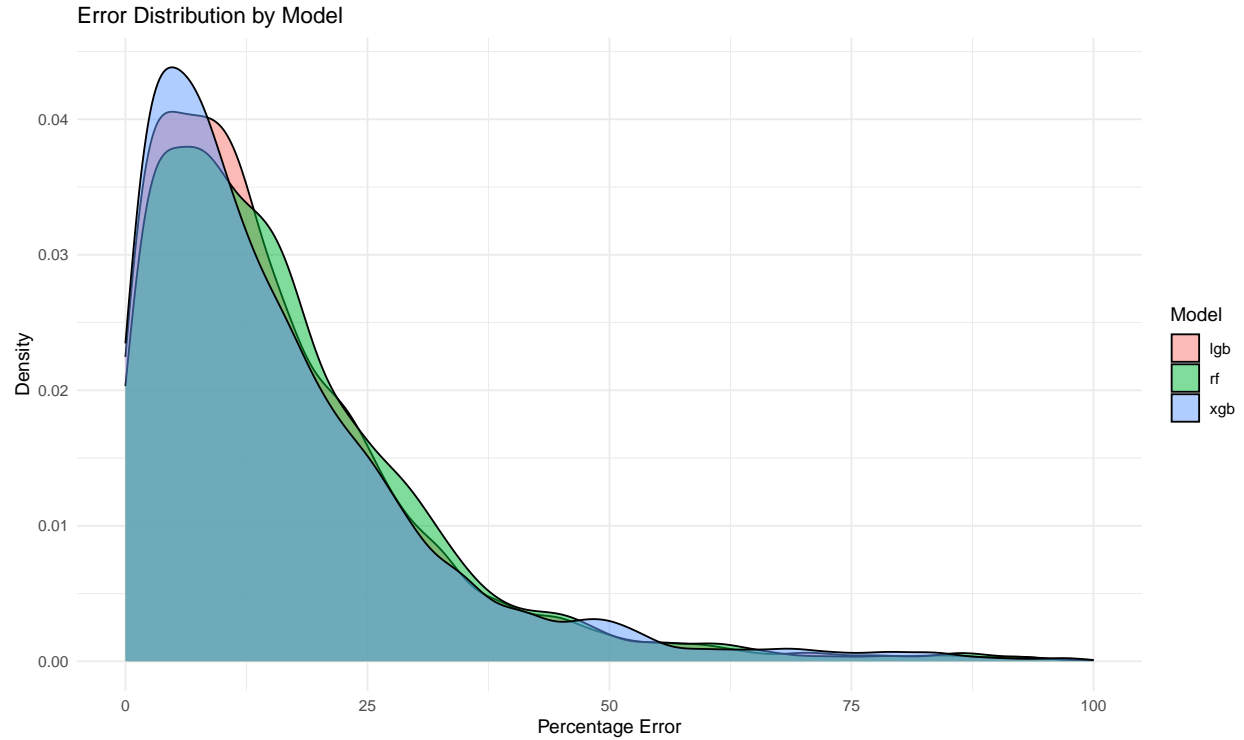
Table 3: Mean Absolute Percentage Error by Model

XGBoost_MAPE	RF_MAPE	LightGBM_MAPE
17.14	17.44	16.77

Conclusions

1. **Price Distribution**: Property prices in NYC show significant variation across boroughs and neighborhoods, with Manhattan consistently showing higher median prices.

2. **Size-Price Relationship**: There is a strong correlation between property size and price, though this relationship varies by borough.

3. **Model Performance**:

   - XGBoost shows robust performance across different price ranges
   - Random Forest provides consistent predictions
   - LightGBM demonstrates competitive results, particularly in mid-range properties

Recommendations

1. **For Buyers**: Consider properties in emerging neighborhoods that show stable price trends but haven't yet reached peak valuations.

2. **For Sellers**: Use the model predictions as a baseline for pricing strategy, considering the specific characteristics of the property and neighborhood.

3. **For Future Analysis**:

   - Incorporate additional features such as proximity to amenities
   - Develop borough-specific models for more accurate predictions
   - Include temporal analysis to capture market trends