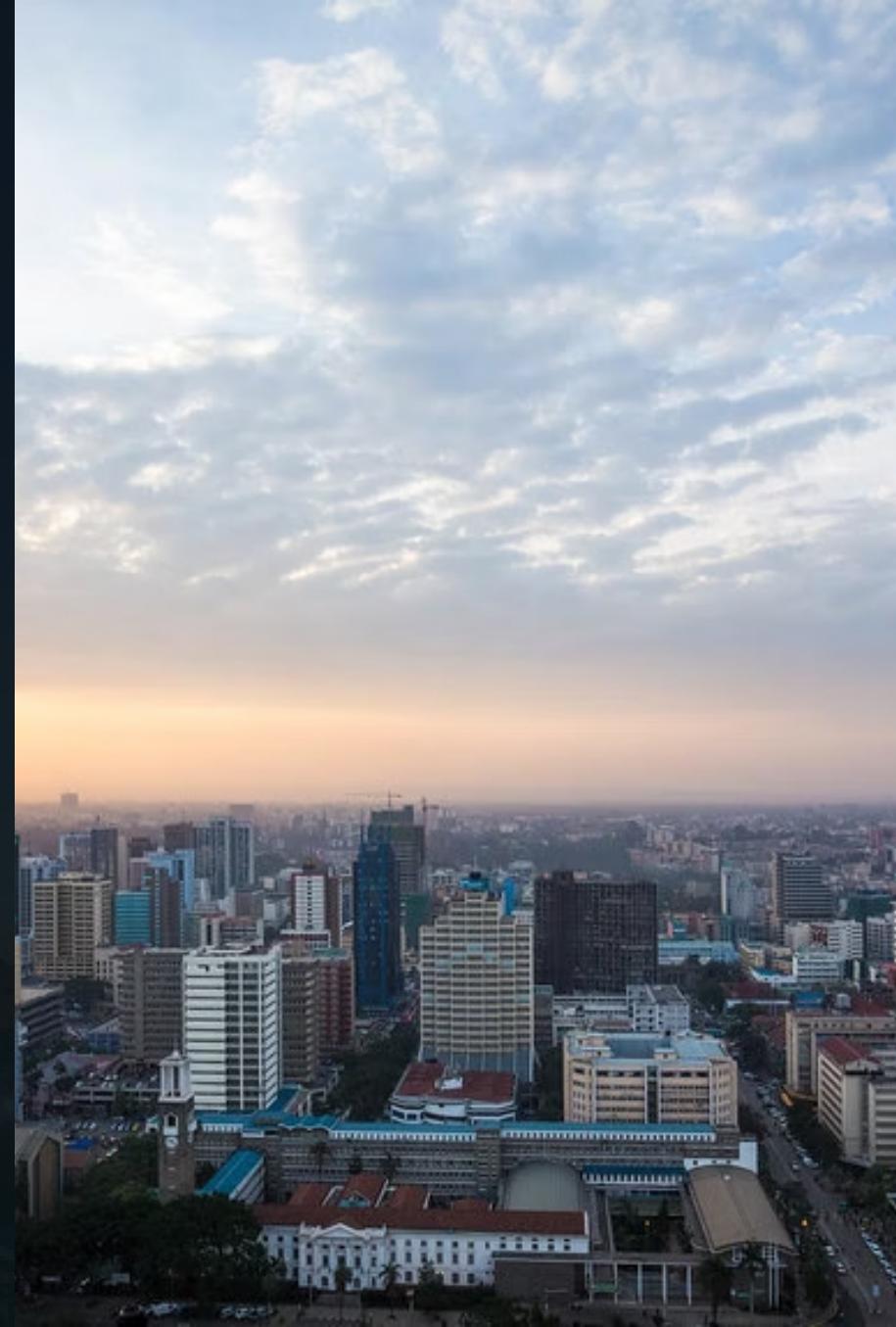


# Nairobi House Price Prediction

Building a data-driven tool to bring transparency and reliable price benchmarks to Nairobi's dynamic real estate market.



# The Challenge: Market Opacity

## The Problem

Nairobi's property market operates in the shadows. Buyers, sellers, and investors navigate without reliable pricing data or transparent benchmarks.

## Real-World Impact

- Buyers overpay by millions of shillings
- Sellers under-price valuable properties
- Investors miss lucrative opportunities
- Market inefficiency drives uncertainty

**Our Goal:** Build a machine learning model that delivers accurate, data-driven house price estimates.

# Day 1: Data Collection Sprint

## Source & Scale

Scraped **800+ active listings** from BuyRentKenya, the leading online real estate platform in Nairobi.

## Key Fields Extracted

- Location & neighborhood
- Price (asking)
- Size (sqm/acres)
- Bedrooms & bathrooms
- Amenities & listing date

## Major Challenge

Inconsistent size formats plagued the dataset: ranges, mixed units (acres vs sqft), and missing values required a custom parser.

A comprehensive data dictionary was created to standardize fields and ensure consistency across the entire pipeline.

# Day 2: Cleaning & Feature Engineering



## From Raw to Ready

Reduced **800 listings to 359 usable records** after aggressive cleaning: removed duplicates, handled missing values, and eliminated statistical outliers.

## Transformations Applied

- Standardized location names using basic string matching
- Created **price\_per\_sqm** metric for normalization
- Built **amenity\_count** feature from listing descriptions
- Extracted temporal features: month and year

**Critical Insight:** Location and size emerged as dominant predictors, while amenity count showed minimal predictive power.

# Days 3-4: Model Development

01

## Baseline Model

Linear Regression with log-transformed target achieved  $R^2$  of 0.70 – a solid foundation.

02

## Advanced Algorithms

Tested Random Forest and XGBoost regressors with hyperparameter tuning via grid search.

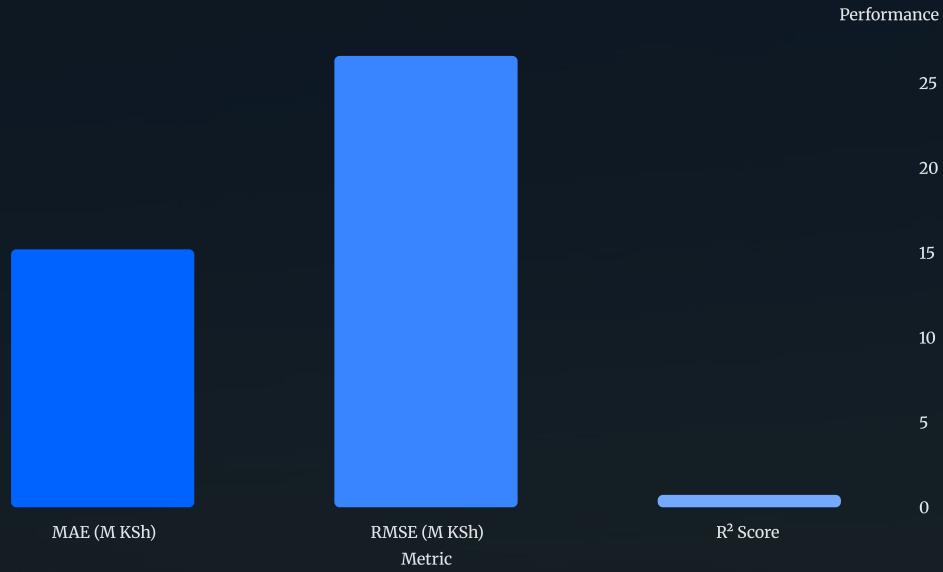
03

## Winner Selected

XGBoost with log transformation delivered the best balance of accuracy and interpretability.



# Model Performance Metrics



## XGBoost Results

**Mean Absolute Error:** 15.2M KSh – predictions typically within 15 million shillings of actual price.

**R<sup>2</sup> Score of 0.73** indicates the model explains 73% of price variance across the dataset.

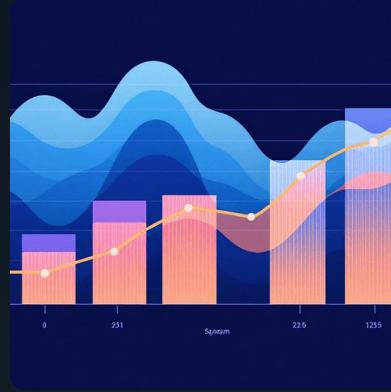
- ❑ **Key Learning:** Simpler models with proper transformations can outperform complex architectures when working with limited data. Feature engineering trumps algorithm complexity.

# Business Intelligence Dashboard



## Top Locations by Price

Karen, Runda, and Lavington dominate the premium segment with median prices 70M+ KSh above average areas.



## Price per SQM Analysis

Wide variation across neighborhoods reveals arbitrage opportunities and investment hotspots.

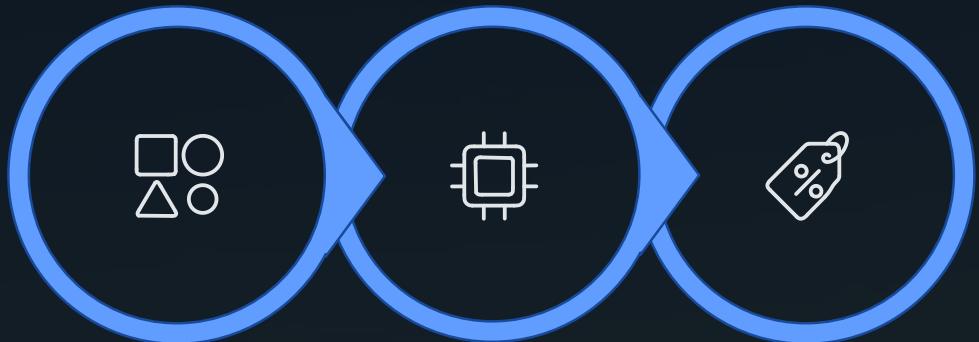
**Critical Takeaway:** Location is king. A property in Karen commands a 70M KSh premium compared to average neighborhoods, independent of size or amenities.



## Amenity Impact Study

Surprisingly minimal effect on final price – location and size remain the decisive factors.

# Live Application Demo



Enter  
Property

Model  
Processes

Get Estimate

## Predictor App Features

- Interactive form for property specifications
- Real-time price estimation engine
- Confidence intervals displayed
- Comparable properties shown

## Dashboard Capabilities

- Interactive neighborhood price maps
- Historical trend analysis
- Market segment breakdowns
- Export functionality for reports

Both tools are deployed and accessible for testing, with local and cloud deployment options available.

# What Was Achieved



**Working MVP Delivered**  
Fully functional price prediction model with dual interfaces: web app and analytical dashboard.



**73% Accuracy Rate**  
Model explains nearly three-quarters of price variance, providing reliable estimates for stakeholders.



**Market Insights Uncovered**  
Quantified location premiums and identified key price drivers in Nairobi's real estate market.

**Current Limitation:** Dataset size (359 listings) constrains model generalization. More data would significantly improve accuracy and market coverage.

# Future Roadmap



## Expand Data Sources

Scrape additional platforms and historical listings to reach 5,000+ properties for robust training.



## Add Geocoding

Calculate distance to CBD, schools, and transport hubs as location-based features.



## Property Type Segmentation

Distinguish apartments, townhouses, and standalone homes with separate prediction models.



## Public Cloud Deployment

Launch on AWS or Azure for broader access and real-time market updates.

## Take Action

Explore the [GitHub repository](#), test the [prediction app](#), and share your feedback to help us refine this tool for Nairobi's real estate community.