Price Prediction for Dubai’s Residential Real-Estate Market

1. Dubai Real Estate Market Overview

Market Introduction
Market transactions worth $9.8Bn within the first 5 months of 2021.

Transaction value growth of 44% in the last year.

Key Challenges
- 70% of home sales are off-plan investments, generating no income in early years
- Inflated supply as developers construct new buildings when demand isn’t there
- Opaque market prices controlled by developers and brokers, difficult for investors to assess deals
- High price volatility in the early-life of properties

2. Problem Statement

Goal
Develop a price prediction model for the residential real-estate market that is data-rich, method-driven, and scalable.

System Comparison

<table>
<thead>
<tr>
<th>Inclusion of Factors</th>
<th>Standard System</th>
<th>Analytics-driven System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Based on standard factors and Underutilize the existing and available data</td>
<td>Explores comprehensive, novel, dynamic and granular engineered features</td>
</tr>
<tr>
<td></td>
<td>1 week worth of data input with monthly updates and discontinued due to efficiency</td>
<td>13 years of historical data, covering entire economic cycle with easy and fast to update approach</td>
</tr>
<tr>
<td></td>
<td>Discounted Cash Flow with simple statistics and management inputs</td>
<td>Big data analytics-driven machine learning models for fair market price prediction</td>
</tr>
<tr>
<td></td>
<td>Low Scalability and restricted to limited number of property candidates</td>
<td>High Scalability that enables scaling to select from all deals across Dubai</td>
</tr>
</tbody>
</table>

3. Overview of Analytical Approach

Data Strategy
Sourced 38,000 historical records between 2009-21 across 4 price driver categories with 211 independent features

Modeling
Final price prediction model is an ensemble of 3 decision-tree based regression algorithms

<table>
<thead>
<tr>
<th>Building-Level</th>
<th>Area-Level</th>
<th>Macro-Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bedrooms</td>
<td>Number of buildings nearby</td>
<td>S&amp;P500 Index</td>
</tr>
<tr>
<td>Apartment size</td>
<td>Number of high-end buildings nearby</td>
<td>OPEC oil prices</td>
</tr>
<tr>
<td>Building age</td>
<td>Total apartment units</td>
<td>Case Shiller index</td>
</tr>
<tr>
<td>Building amenities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R² = 65.7% | 72.0% | 71.6%

4. Management Impact

Existing System
- 3 Building-level features
- Unknown accuracy

Proposed System
- 44 Features across 4 levels
- 13 Years of data
- 95% Acquisition automation with weekly updates
- 72.4% accuracy, compared to 12.4% in literature using similar ensemble of decision-trees
- Combine quantitative & qualitative