HDB Resale Price Prediction Report

Author: Wong Wen Hui

Submission Date: 5/8/2025
GitHub Repository: wongwenhui262/HTX

---

1. Problem Statement

The goal of this project is to build a predictive model to estimate the resale prices of HDB flats in Singapore. This model helps identify the key factors influencing prices and enables stakeholders to strategize on mitigating housing inflation.

---

2. Methodology Overview

a. Data Handling

- Multiple CSVs from different years were extracted from a ZIP archive and concatenated.
- Inconsistent column formatting (e.g., spacing) was cleaned.
- Columns like `storey_range` and `remaining_lease` were parsed into numerical representations.

b. Preprocessing

- Categorical variables (`town`, `flat_type`, `flat_model`) were one-hot encoded.
- Date column `month` was converted to datetime, with new `year` and `month_num` columns extracted.
- Rows with missing values were dropped after parsing key fields.

---

3. Exploratory Data Analysis

a. Resale Price Trends:

- Average resale prices increased steadily from 2017 to 2023.
- A spike was observed in 2021–2022, likely due to macroeconomic shifts and increased demand.

b. Town-Level Pricing:

- Central regions such as Queenstown, Bukit Merah, and Bishan showed significantly higher median prices.
- Outlying towns like Sembawang and Woodlands had lower average prices.

---

4. Model Development

a. Model Chosen: Random Forest Regressor

- Handles both numerical and categorical data well
- Provides built-in feature importance
- Robust against overfitting for tabular data

b. Train-Test Split: 80% training, 20% testing

c. Evaluation Metrics:

Metric	Value
MAE	~23,428 SGD
RMSE	~33,594 SGD
R^2	0.9528

These indicate a strong fit, with the model explaining ~95% of the variance in resale prices.

---

5. Feature Importance & Insights

Top Features Impacting Price:

1. `floor_area_sqm`
2. `remaining_lease_years`
3. `lease_commence_date`
4. `storey_avg`
5. Flat type & model
6. Town (e.g., Queenstown, Bukit Timah)

Key Insights:

- Larger flats predictably cost more.

- Flats with more remaining lease have higher resale values.
- Higher floor units tend to have a price premium.
- Town location plays a major role — proximity to CBD inflates value.

---

6. Policy Recommendations

To curb housing price inflation, consider:

- Encouraging decentralization by developing amenities in lower-cost towns.
- Extending lease for older flats to retain value and reduce volatility.
- Implementing flat size caps or subsidies for lower-income buyers in high-cost areas.

---

7. Future Work

- **Model Tuning:** Use `GridSearchCV` for hyperparameter optimization.
- **Explainability:** Implement SHAP values for transparency.
- **Temporal Modeling:** Explore time-series forecasting techniques.
- **Deployment:** Wrap model in an API for stakeholder use.

---

8. Conclusion

This model provides accurate and interpretable predictions for resale HDB flat prices. With further tuning and deployment, it can serve as a valuable tool for policy-making and public transparency in Singapore's housing market.