#### **Analysis and Prediction of Singapore HDB Resale Flat Prices (2017-2024)**

#### **Objective**

The objective of this project is to analyze the resale prices of HDB flats in Singapore from January 2017 to June 2024 and build predictive models to estimate future resale prices based on various features.

#### **Data Description**

The dataset contains detailed information on HDB flat resale transactions, including:

* **month**: Month of the sale
* **town**: Designated residential area
* **flat\_type**: Classification by room size
* **block**: Block number of the flat
* **street\_name**: Street name of the flat
* **storey\_range**: Estimated range of floors
* **floor\_area\_sqm**: Total interior space in square meters
* **flat\_model**: Classification by generation
* **lease\_commence\_date**: Start date of the lease
* **remaining\_lease**: Remaining time on the lease
* **resale\_price**: Cost of the flat sold

#### **Data Preprocessing**

1. **Handling Missing Values**:
   * There were no missing values in the dataset.
2. **Encoding Categorical Variables**:
   * flat\_type and storey\_range were ordinal encoded based on their natural order.
   * flat\_model was one-hot encoded to handle categorical data without imposing any ordinal relationship.
3. **Feature Engineering**:
   * Extracted flat\_age from the lease\_commence\_date and current year.
   * Converted remaining\_lease to numeric format for better model compatibility.

#### **Exploratory Data Analysis (EDA)**

1. **Trends Over Time**:
   * Analyzed resale price trends over the months and years.
   * Observed seasonal trends and price fluctuations.
2. **Distribution Analysis**:
   * Analyzed the distribution of flat prices based on different flat types, towns, and storey ranges.
   * Visualized the relationship between floor area and resale price.
3. **Correlation Analysis**:
   * Investigated correlations between different features and the resale price to identify significant predictors.

#### **Model Building and Evaluation**

We experimented with several regression models to predict the resale price:

1. **Polynomial Regression**:
   * Added polynomial features and fitted a linear regression model.
   * Performance:
     + MAE: 69549.3
     + RMSE: 91689.5
     + R²: 0.72
2. **Neural Network**:
   * Built a neural network model with two hidden layers.
   * Performance:
     + MAE: 83575.0
     + RMSE: 106484.6
     + R²: 0.63
3. **Decision Tree Regressor**:
   * Applied a decision tree model with grid search for hyperparameter tuning.
   * Performance:
     + MAE: 38713.0
     + RMSE: 55757.8
     + R²: 0.90
4. **Random Forest Regressor**:
   * Used randomized search for hyperparameter tuning.
   * Performance:
     + MAE: 35900.5
     + RMSE: 51043.3
     + R²: 0.91
5. **Gradient Boosting Regressor**:
   * Applied gradient boosting with grid search for hyperparameter tuning.
   * Performance:
     + MAE: 48223
     + RMSE: 63155.3
     + R²: 0.87
6. **XGBoost Regressor**:
   * Used XGBoost with grid search for hyperparameter tuning.
   * Performance:
     + MAE: 48072.7
     + RMSE: 62809.6
     + R²: 0.87

#### **Comparison of Model Performance**

Based on the evaluation metrics, the model performances were compared as follows:

| Model | MAE | RMSE | R² |
| --- | --- | --- | --- |
| Polynomial Regression | 69549.3 | 91689.5 | 0.72 |
| Neural Network | 83575.0 | 106484.6 | 0.63 |
| Decision Tree | 38713.0 | 55757.8 | 0.90 |
| Random Forest | 35900.5 | 51043.3 | 0.91 |
| Gradient Boosting | 48223 | 63155.3 | 0.87 |
| XGBoost | 48072.7 | 62809.6 | 0.87 |

#### 

#### **Conclusion**

* The models provide a comprehensive view of the factors influencing HDB flat resale prices in Singapore.
* **Random Forest** generally performed the best in terms of accuracy and prediction consistency.
* **Decision Tree** also showed strong performance with a high R² value.
* **Gradient Boosting** and **XGBoost** performed similarly, both providing reliable predictions.
* **Neural Network** and **Polynomial Regression** performed adequately but were less accurate than the tree-based models.

The findings can be used by policymakers, real estate agents, and buyers to understand and predict the resale prices of HDB flats in Singapore, helping them make more informed decisions.