JAPAN

REAL ESTATE PRICE PREDICTION

SPRINT 2

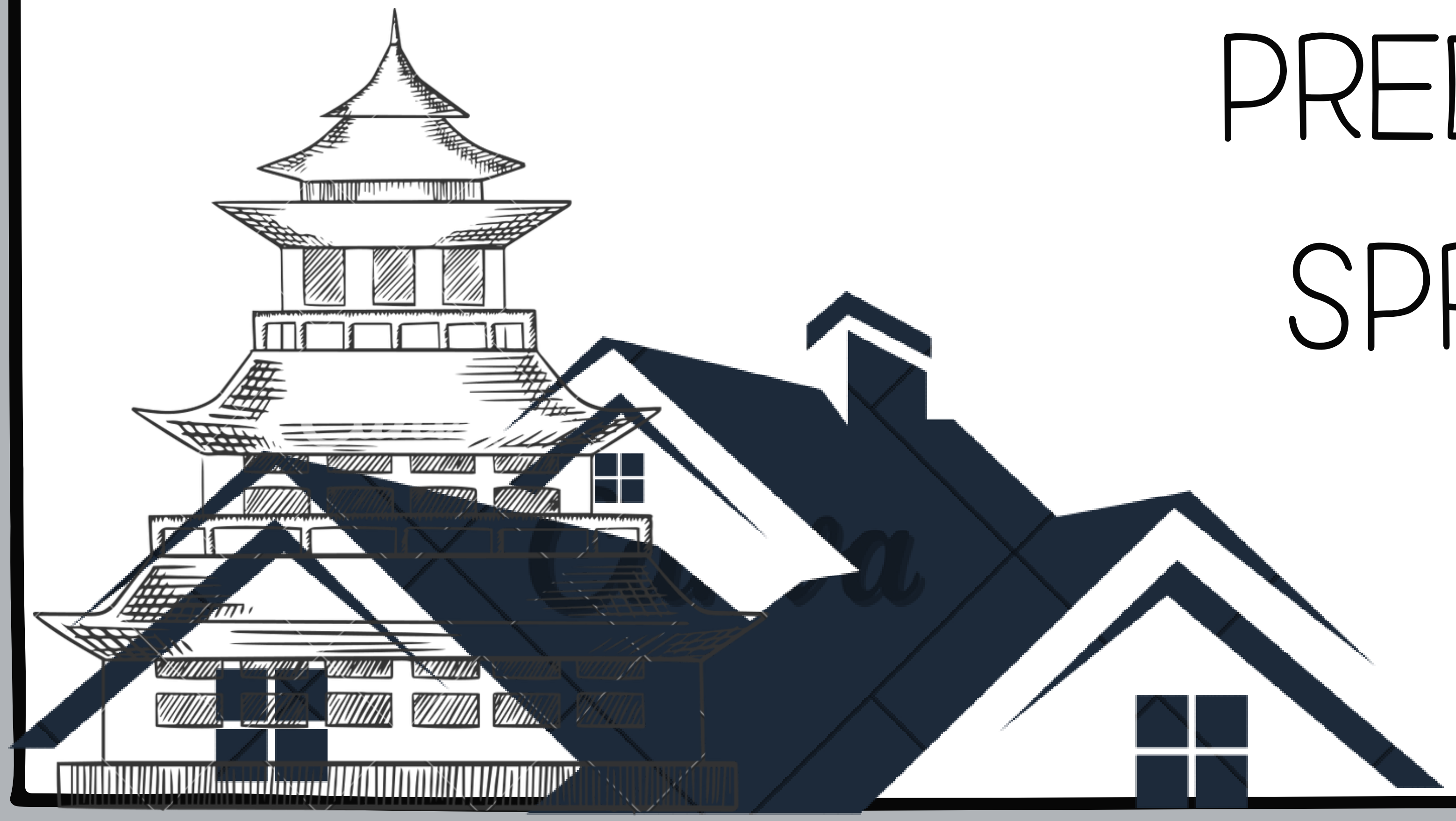
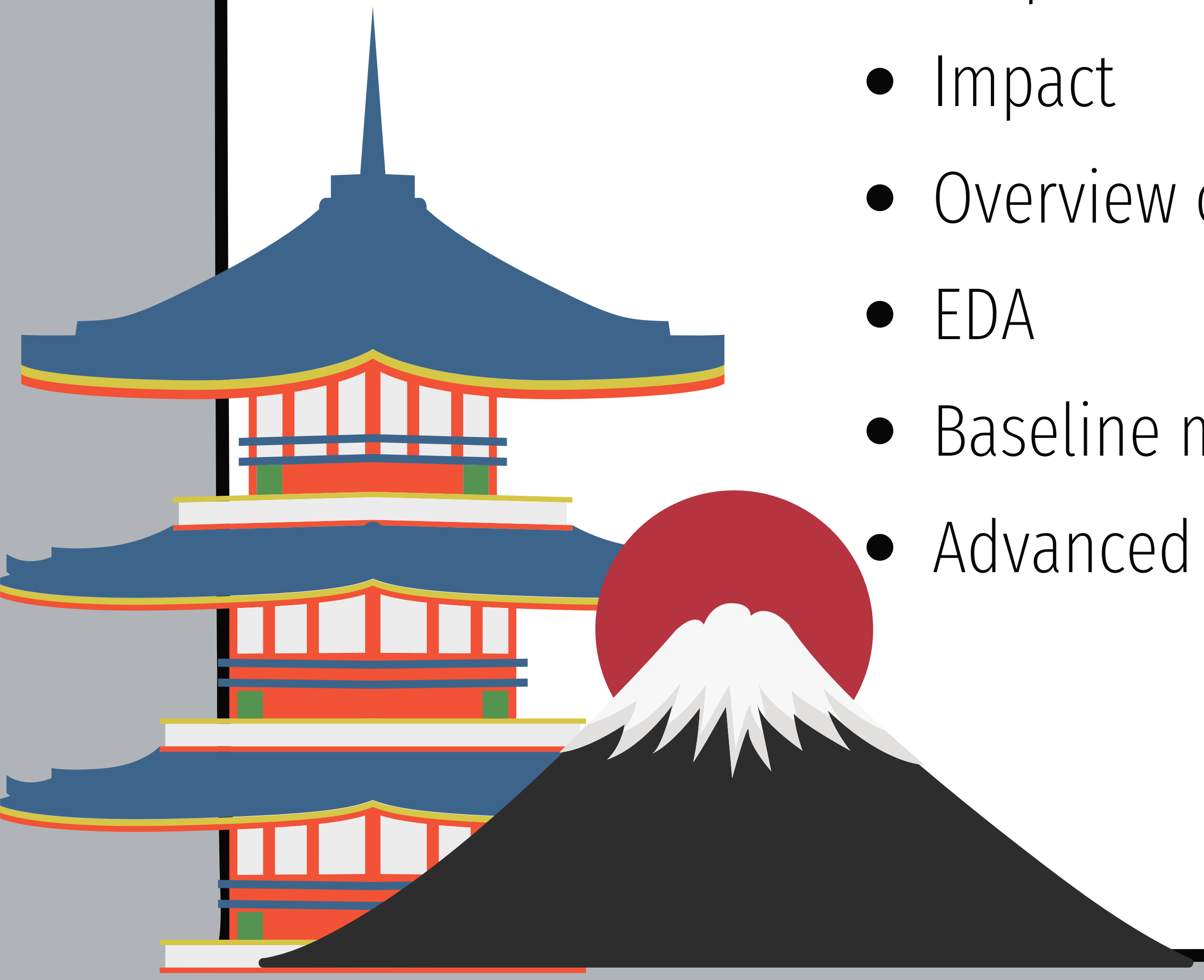
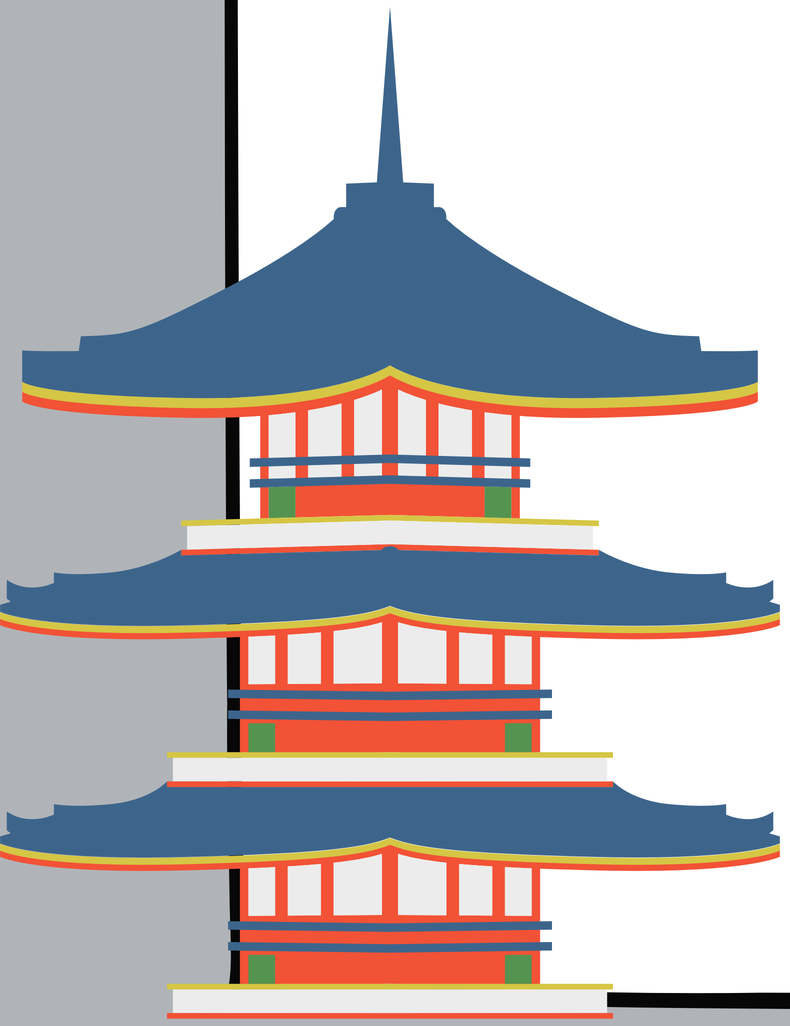


Table of contents:



Introduction:

- From 2005 to 2019.
- Surveyed by the Ministry of Land, Infrastructure, Transport, and Tourism of Japan (MLIT).
- 47 prefectures in Japan.
- Five real estate types namely:
  a. Agricultural land
  b. Forest Land
  c. Residential Land(Land Only)
  d. Residential Land(Land and Building)
  e. **Pre-owned Condominiums, etc.**



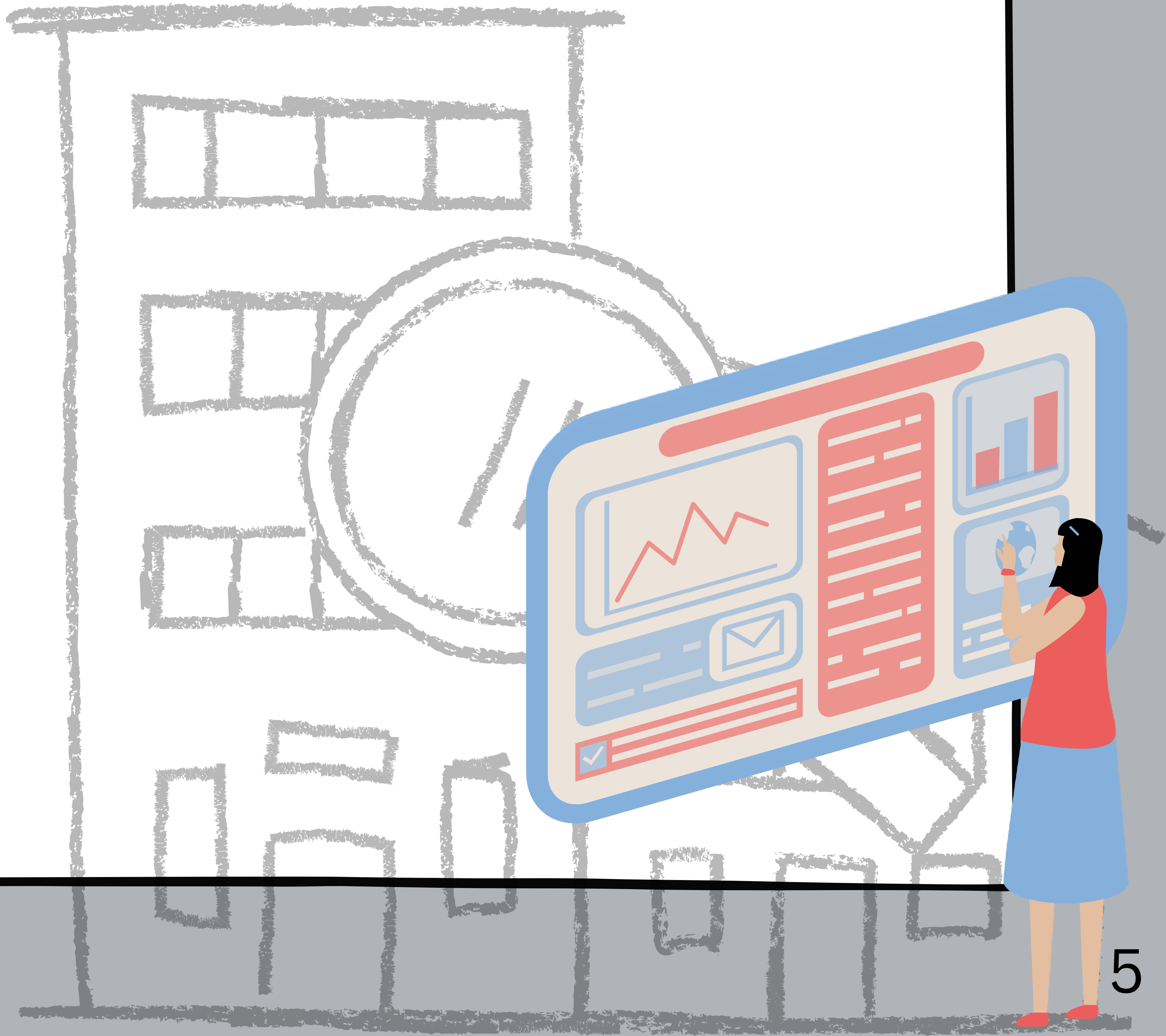
Problem Area:

The primary goal is to gain insights into the trends and factors influencing real estate prices in Japan over this 15-year period.

Data Science solutions:

The proposed data science solution involves the following key components:

1. Data Exploration
2. Descriptive Analysis
3. Time period Analysis
4. Spatial Analysis
5. Factors Affecting Prices
6. Predictive Modeling:



Impact

- Japan's population is aging quickly. Housing supply exceeds housing demand, the national vacancy rate reached **13.6%** for all housing types.
- **5% improvement** in predicting trade prices can contribute to more affordable housing options for buyers.
- Market efficiency, it might lead to a **10% increase** in the overall effectiveness of property transactions.

Dataset:

- **Type**: Real Estate Type (e.g., Residential Land, Agricultural Land, Condominiums).
- **Region**: Characteristics of surrounding areas (e.g., Residential Area, Commercial Area).
- **MunicipalityCode**: City code of Japan.
- **Prefecture**: Prefecture name of Japan.
- **Municipality**: City name.
- **DistrictName**: District name.
- **NearestStation**: Nearest station name.
- **TimeToNearestStation**: Time to the nearest station (in minutes).
- **TradePrice**: Transaction prices in Japanese Yen.
- **FloorPlan**: Property floor plan (e.g., 3LDK, 2DK).
- **Area**: Surveyed area in square meters.
- **UnitPrice**: Unit Land Price (Yen) per square meter.
- **PricePerTsubo**: Unit Land Price (Yen) per Tsubo.
- **Frontage**: Frontage in meters.
- **BuildingYear**: Construction year of the building.
- **PrewarBuilding**: Buildings built before 1945.
- **Structure**: Building structure (e.g., Steel frame reinforced concrete, Wooden).
- **Use**: Current property usage (e.g., House, Office, Factory).
- **Purpose**: Purpose of future use (e.g., House, Shop, Office).
- **Direction**: Frontage road direction.
- **Classification**: Frontage road classification (e.g., City Road, National Highway).
- **Breadth**: Frontage road breadth in meters.
- **CityPlanning**: Use districts designated by the City Planning Act.
- **CoverageRatio**: Maximum Building Coverage Ratio (%).
- **FloorAreaRatio**: Maximum Floor-area Ratio (%).
- **Period**: Time of transaction.
- **Year**: Time of transaction year.
- **Quarter**: Time of transaction year-quarter.
- **Renovation**: Renovation status.
- **Remarks**: Additional notes and remarks.

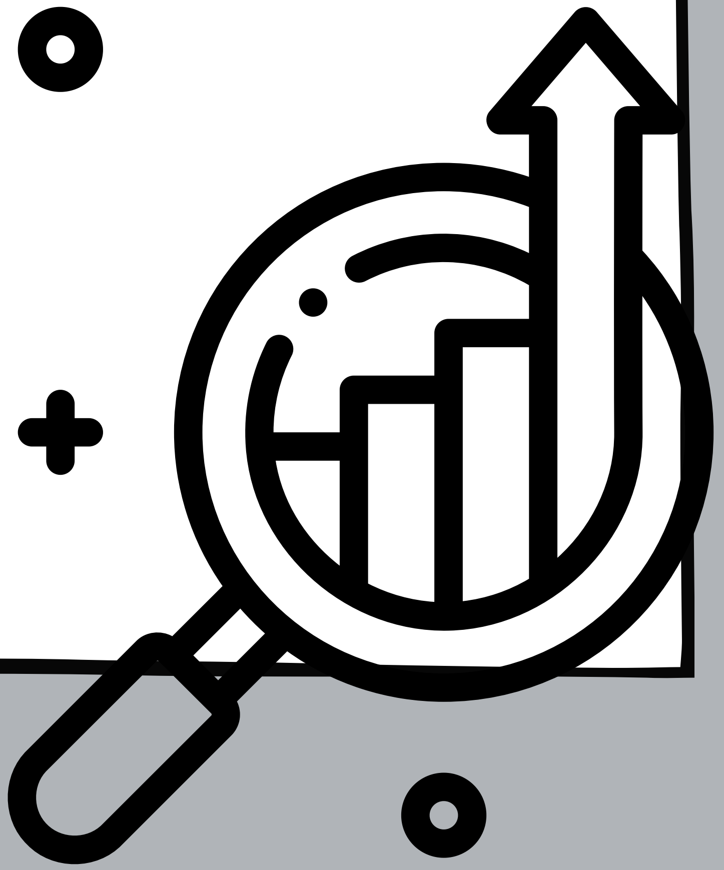
Processing procedures:

Some of the procedures include:

- Imputed null values
- Removal of outliers
- Removed unwanted columns.
- Prefecture level prediction
- Single time to Nearest Station excluding the names of the stations
- Categorical columns with fewer data points converted to 'Others'

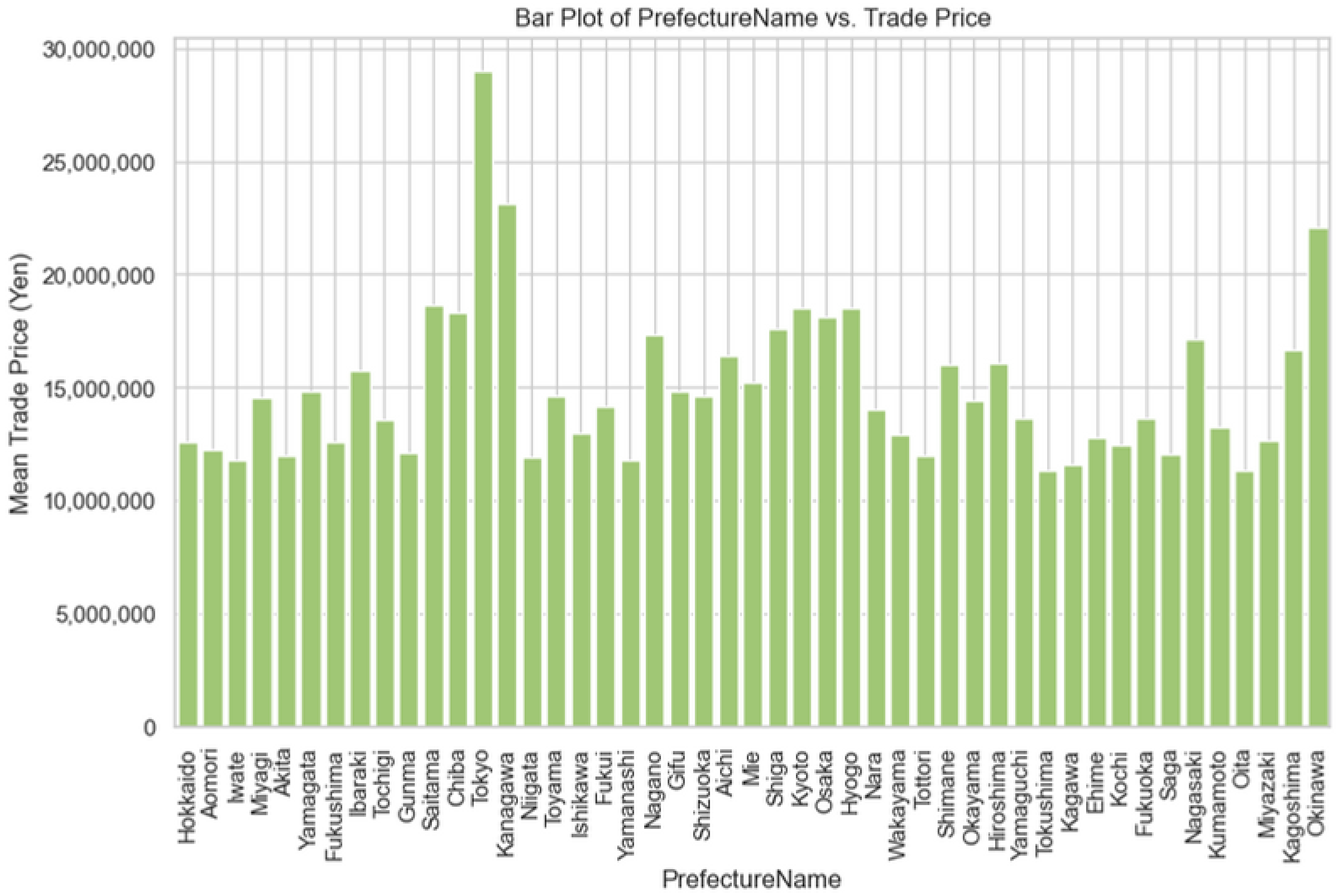
Exploratory Data Analysis

The plots shown in the subsequent slides are the relationship of different variables with **Trade Price.**



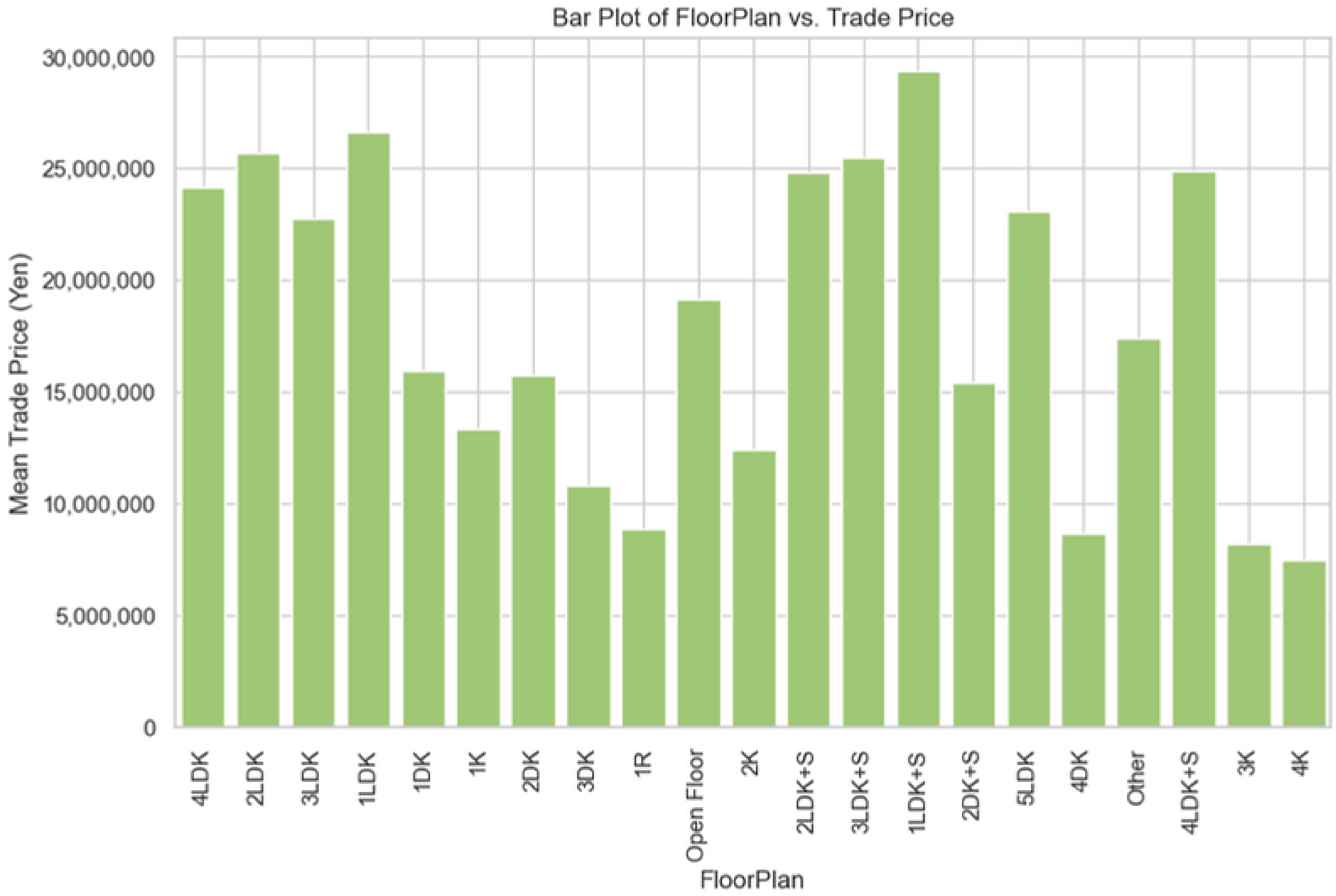
Prefectures Vs TradePrice

As expected Tokyo has the highest Prices out of all Prefectures.



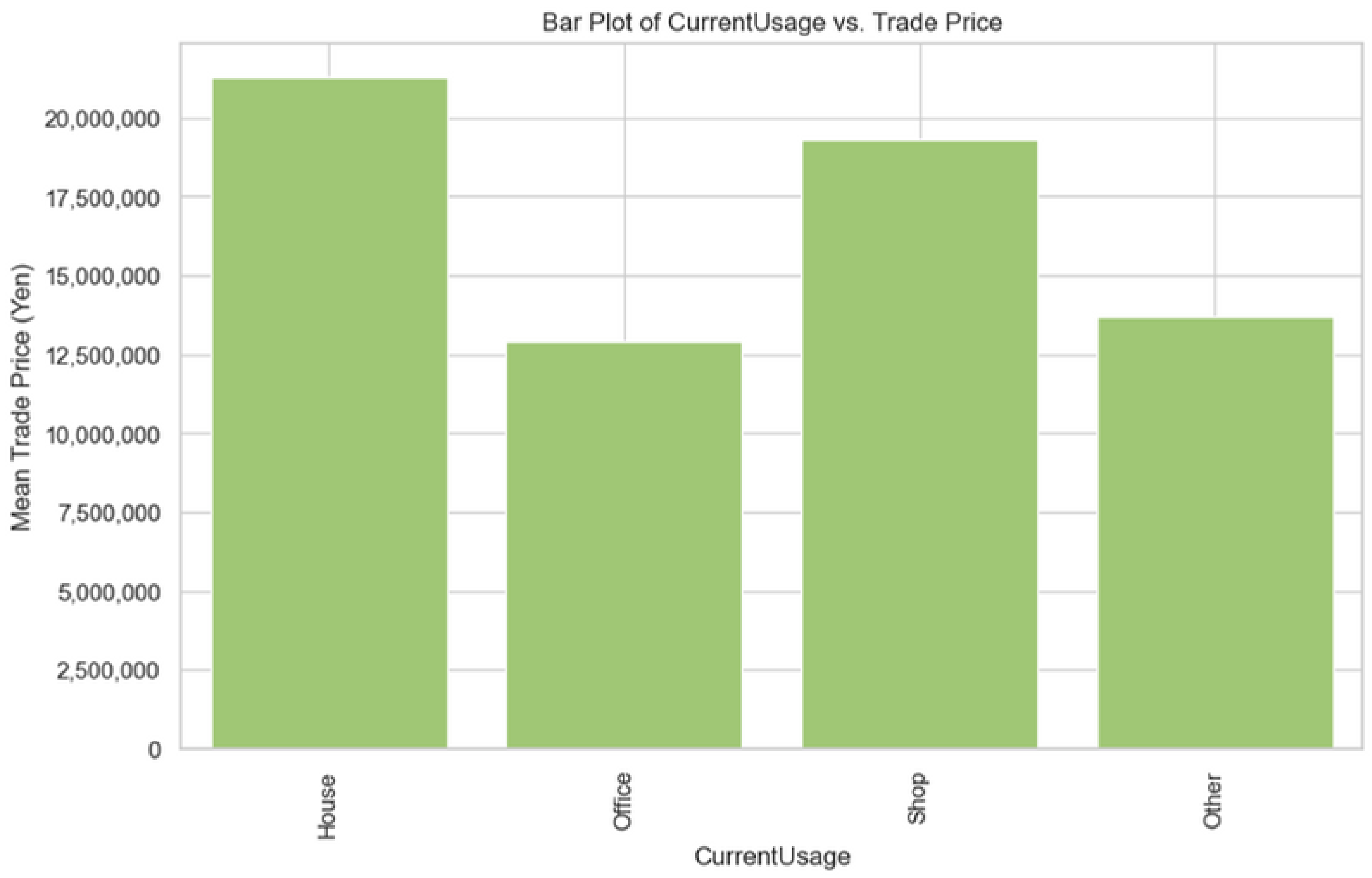
Floor Plan Vs TradePrice

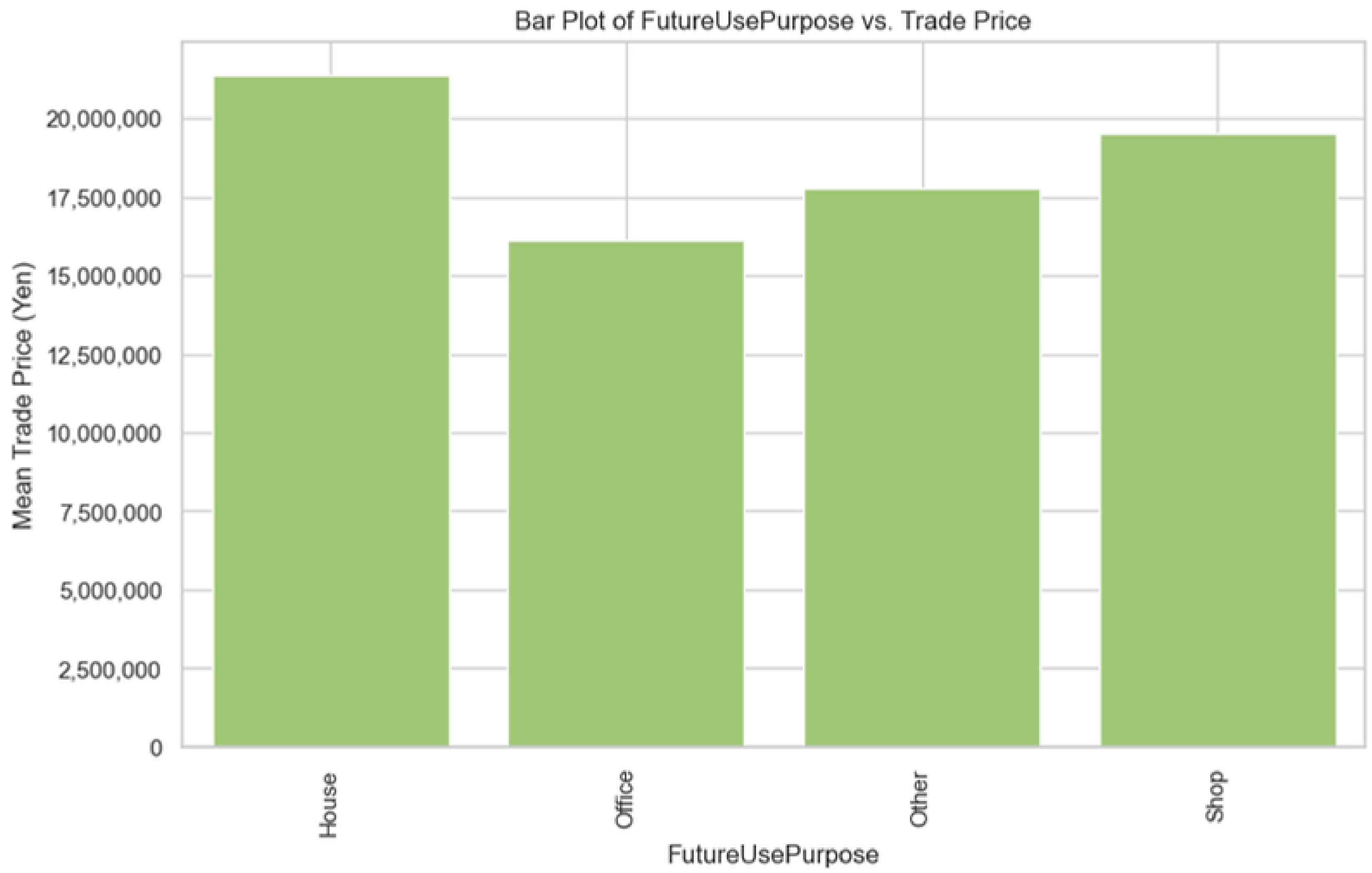
1LDK+S floor plan has the highest mean trade price. So most sold are 1 bedroom houses



Use Vs TradePrice

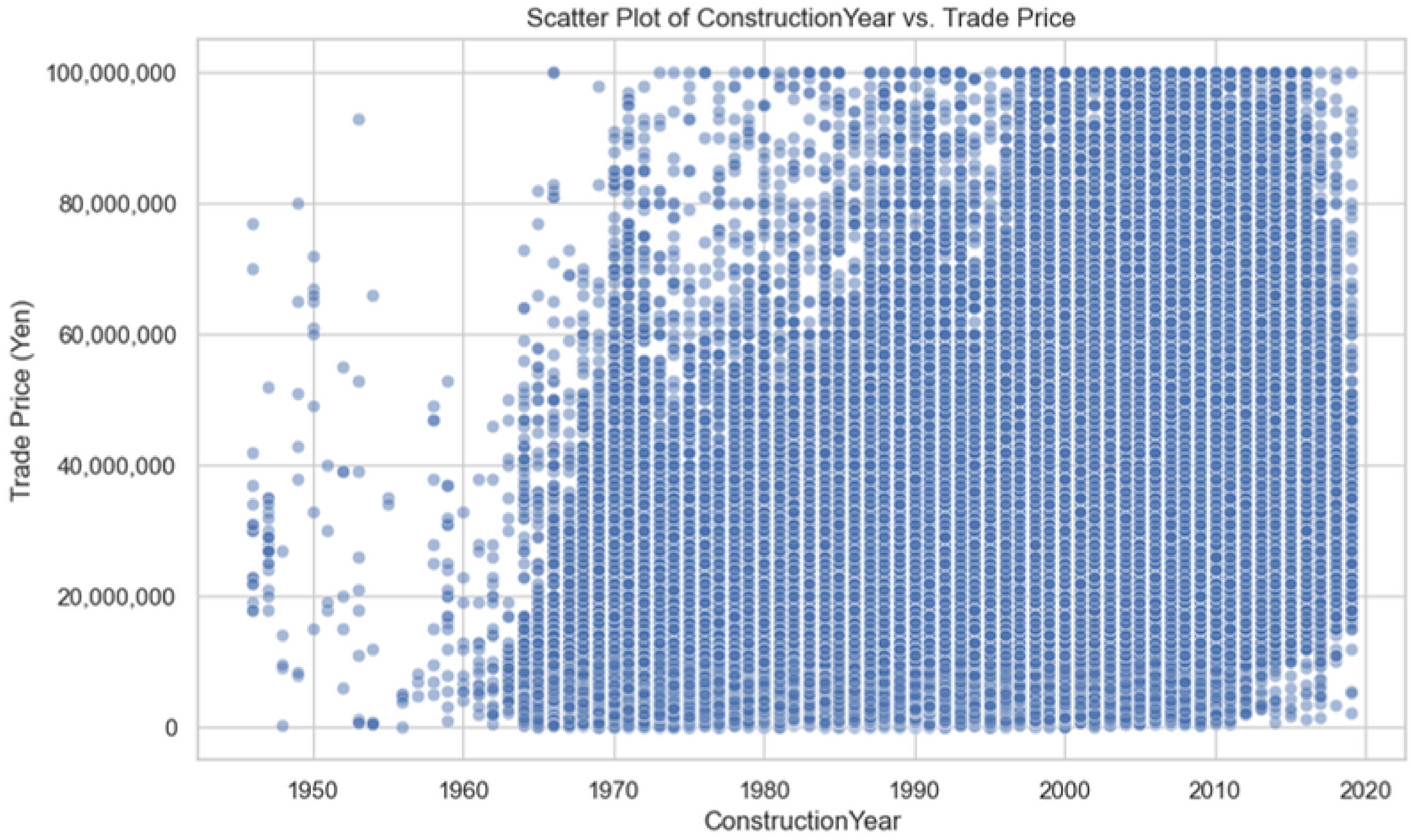
Houses and shops are the ones that are most sold.





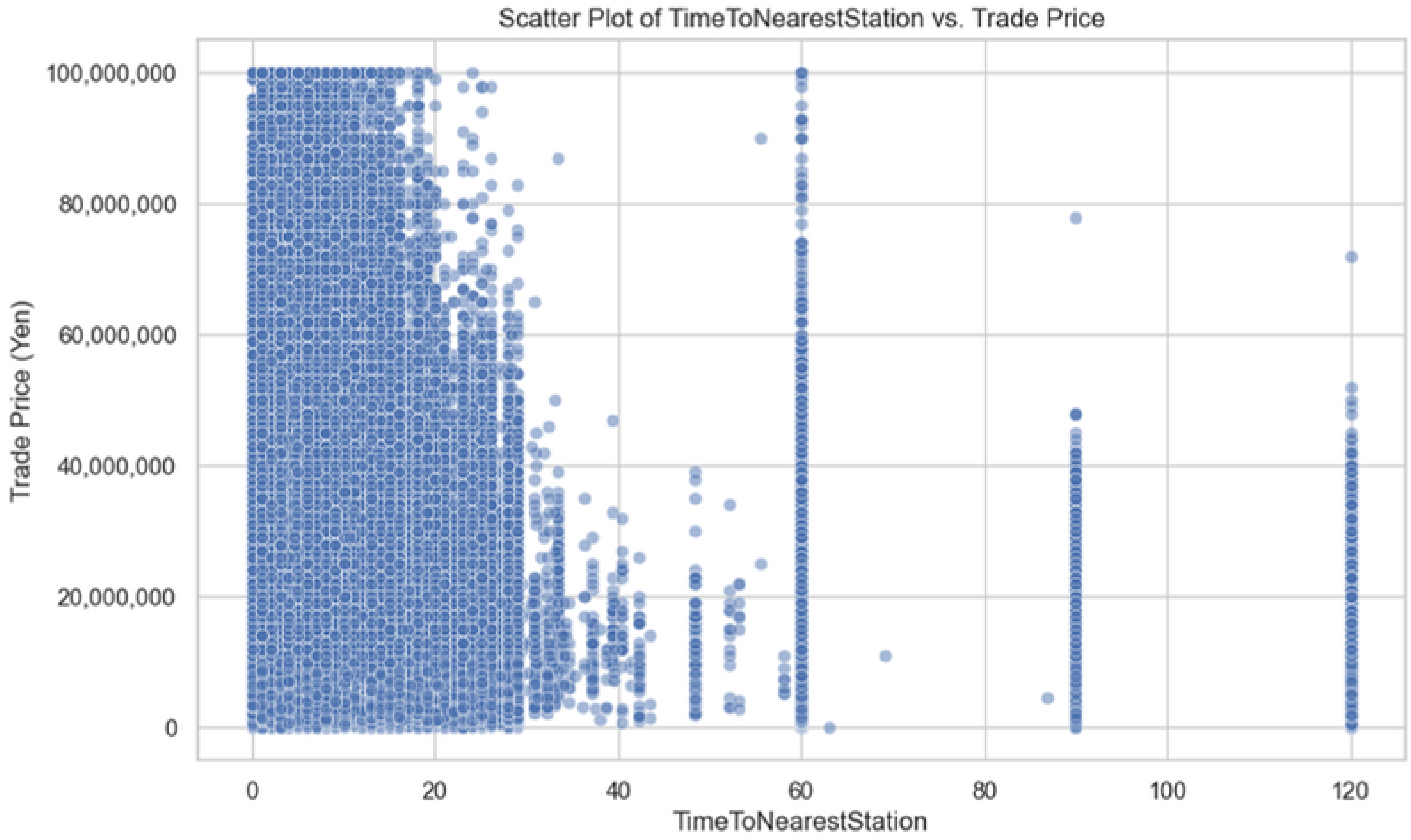
Construction Year Vs TradePrice

The most expensive units are that were constructed recently as expected.



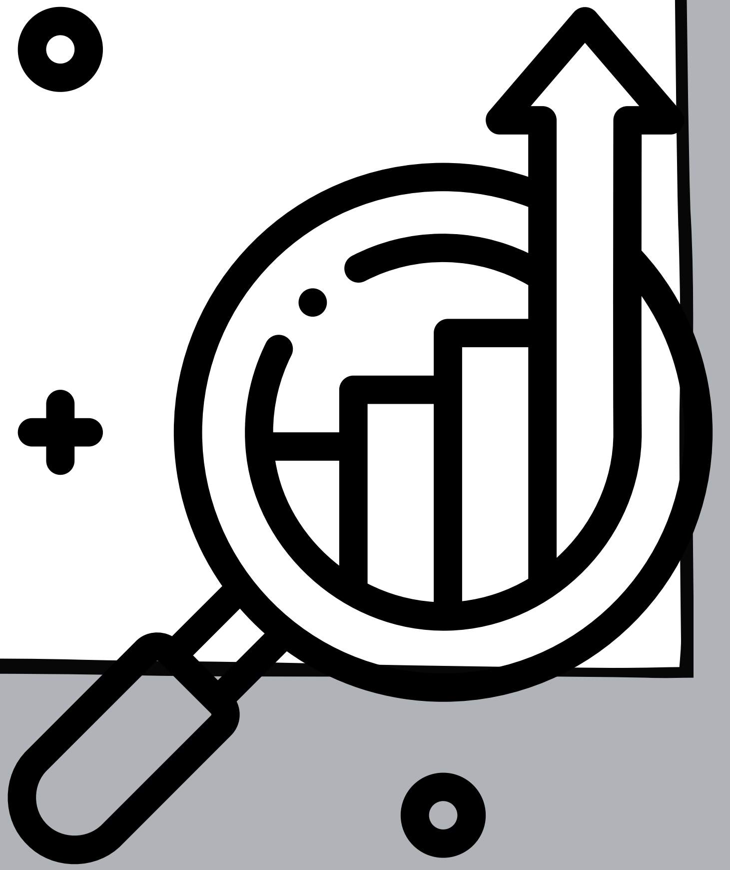
Time to nearest Station Vs TradePrice

Clearly more units are sold near stations and are the most expensive too.



Baseline modelling

- Linear Regression
- Random Forest



Range of trade price to compare RMSE

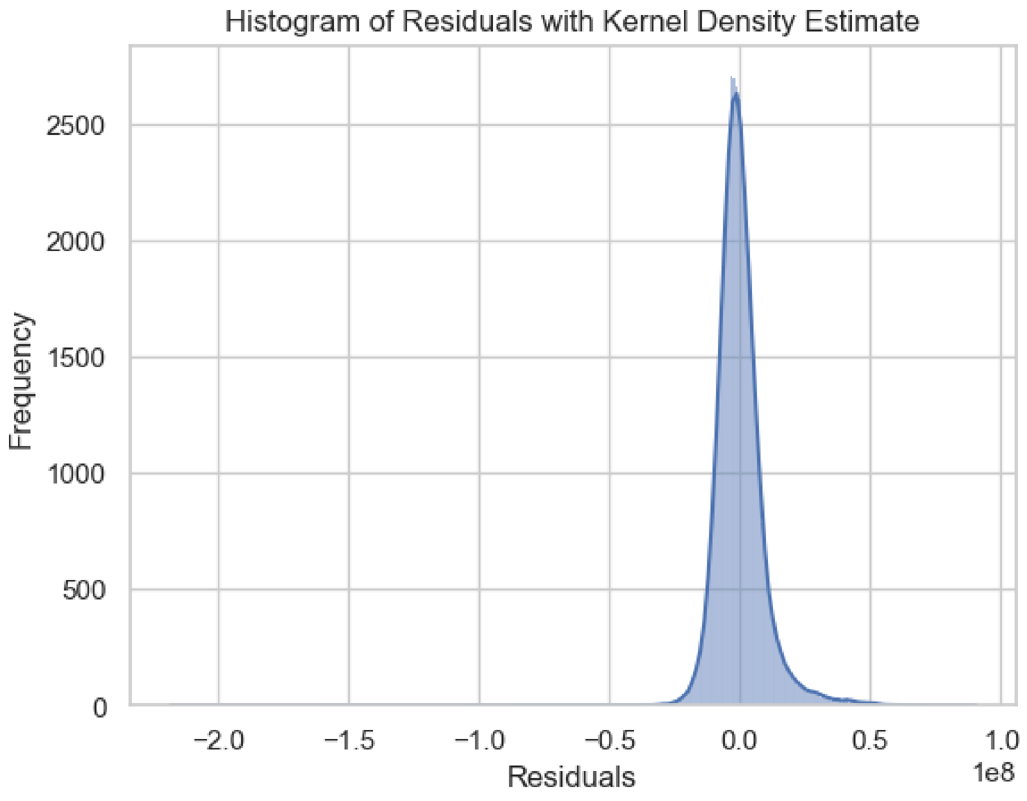
- Minimum TradePriceYen: 450
- Maximum TradePriceYen: 100,000,000
- Range of TradePriceYen: 99,999,550

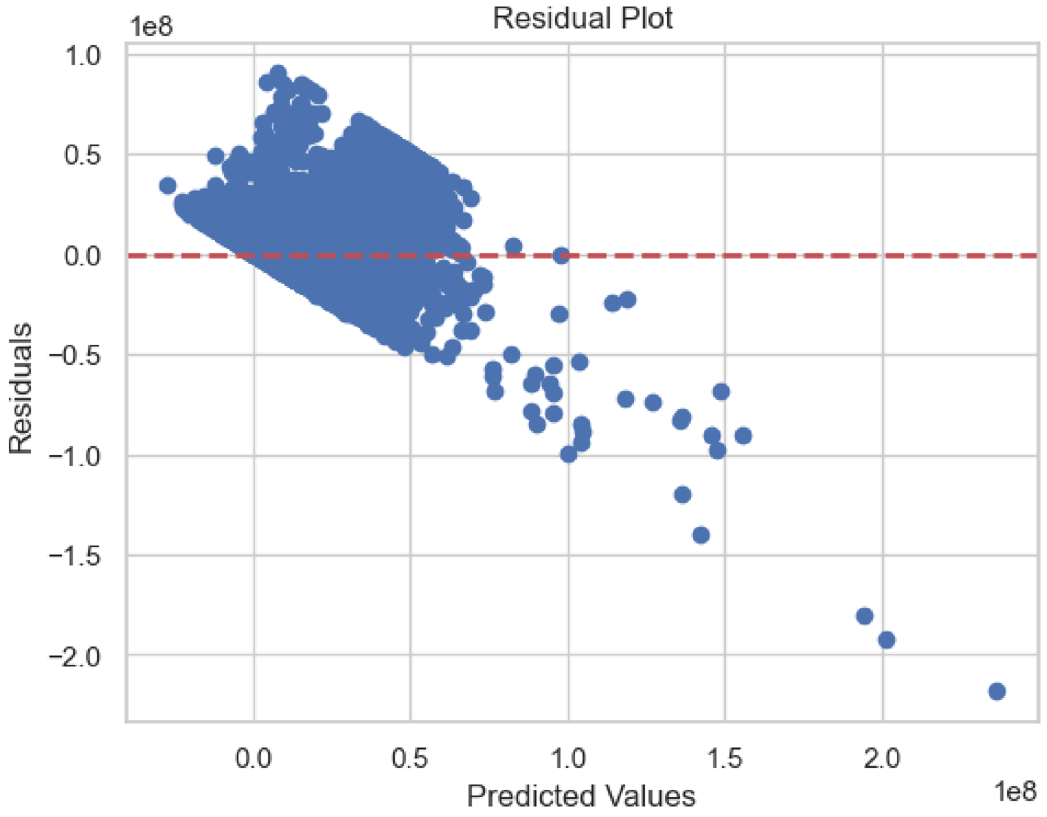
Evaluation Metric of the baseline models

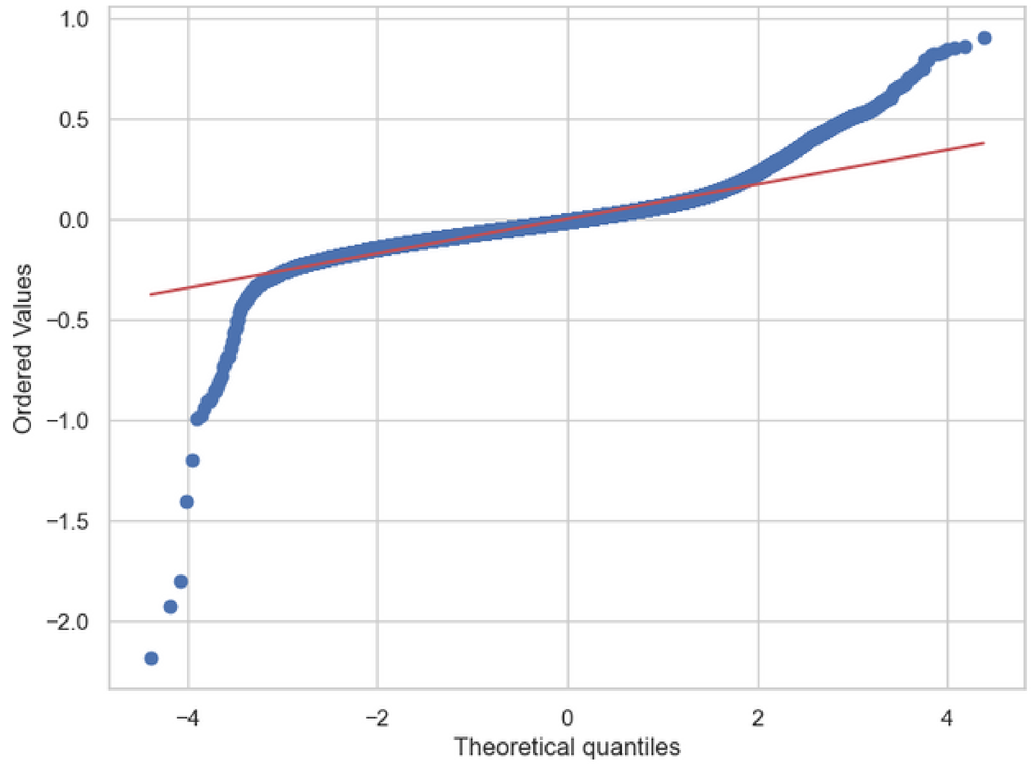
Model	RMSE (Test)	RMSE (Train)	R-squared (R^2)
Linear Regression	9,077,746.713	9,106,036.330	0.6332
Random Forest Regressor	6,789,084.464	2, 575,431.610	0.7948

More on Linear Regression:

- **Heteroscedasticity**
- **QQ-plot deviates from the straight line**, it's an indication that the residuals do not follow a perfect normal distribution
- It shows what is called **excess kurtosis** (heaviness of tails) compared to a normal distribution.

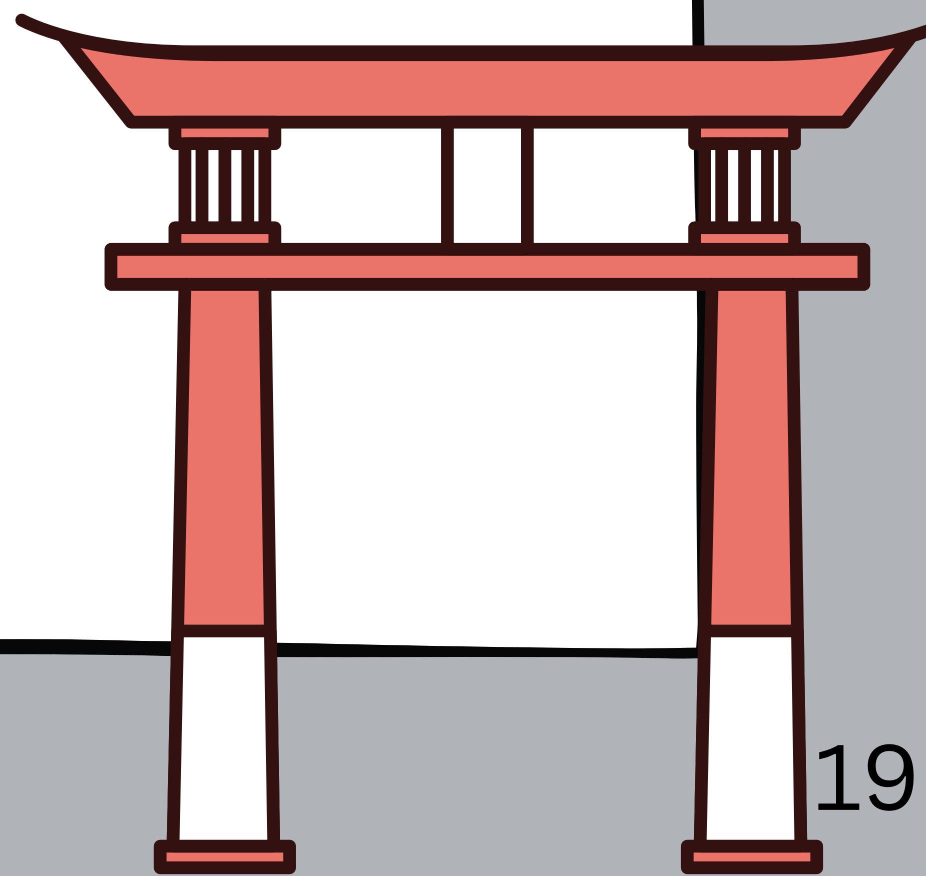






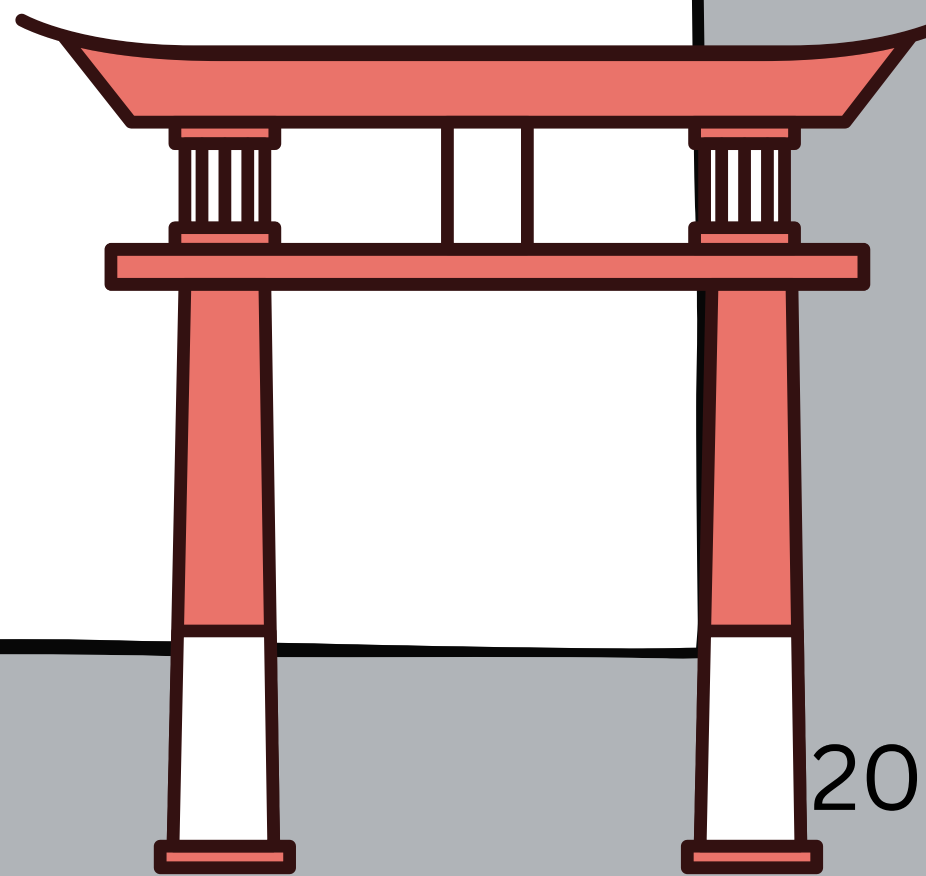
Next steps:

- Advance Modelling
  - **Cross validation**
  - **Fine-tune hyperparameters.**
  - **Ridge and Lasso Regression -** handles collinearity
  - **Gradient Boosting Regressor -** captures complex relationships in the data.
  - **Support Vector Machines (SVM) -** capture non-linear patterns.
  - **Neural Networks -** Using TensorFlow or PyTorch.



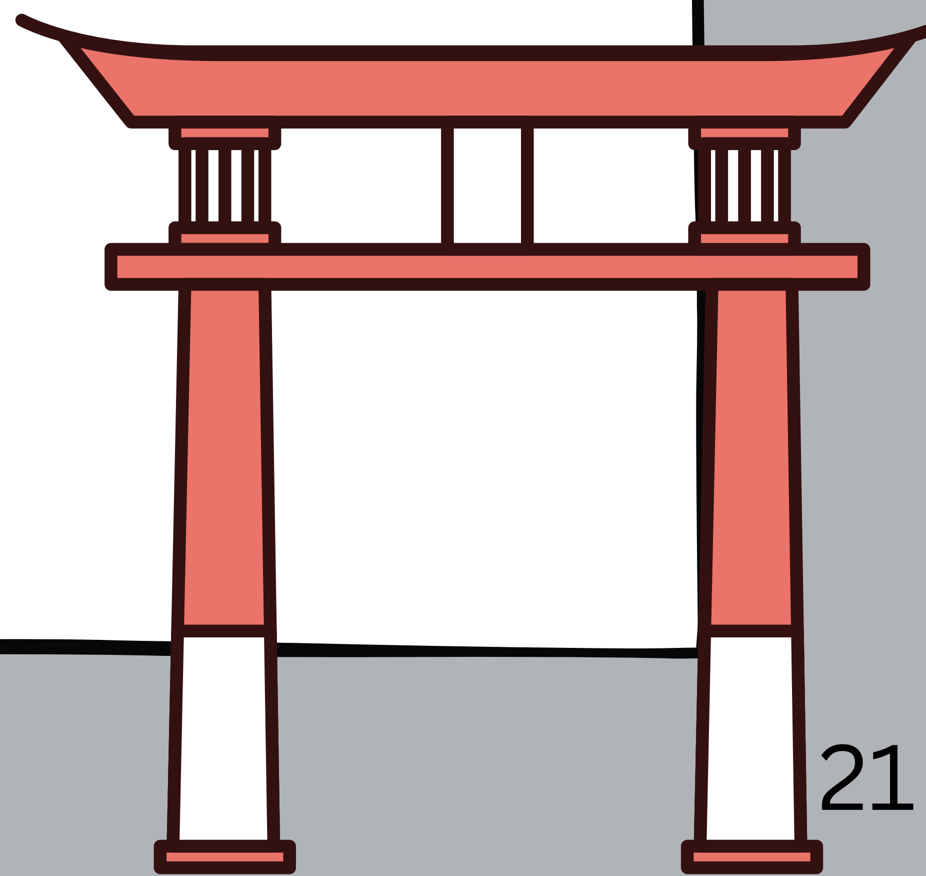
Next steps:

- Productizing work
  - **Create a clean and intuitive UI** that is easy for users to navigate and enter their preferences.
  - Will produce the **price prediction** based on their choices.
  - Bonus: Other budget friendly options from the dataset we can suggest.



Conclusion:

- The impact of predicting real estate prices is significant, as it aids in making informed investment decisions and understanding the dynamics of the real estate market. It addresses questions about the direction of property prices, factors influencing price changes, regional variations, and the potential for future price movements.
- Combining data preprocessing, feature engineering, data analysis, and predictive modeling we can provide valuable insights for stakeholders in the real estate industry.



Thank You !