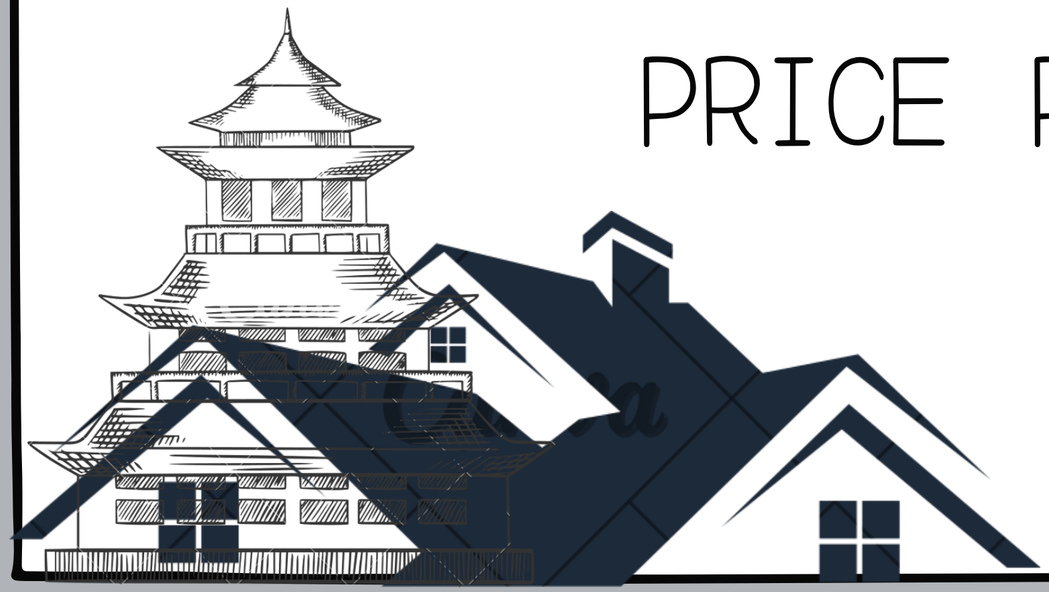
JAPAN

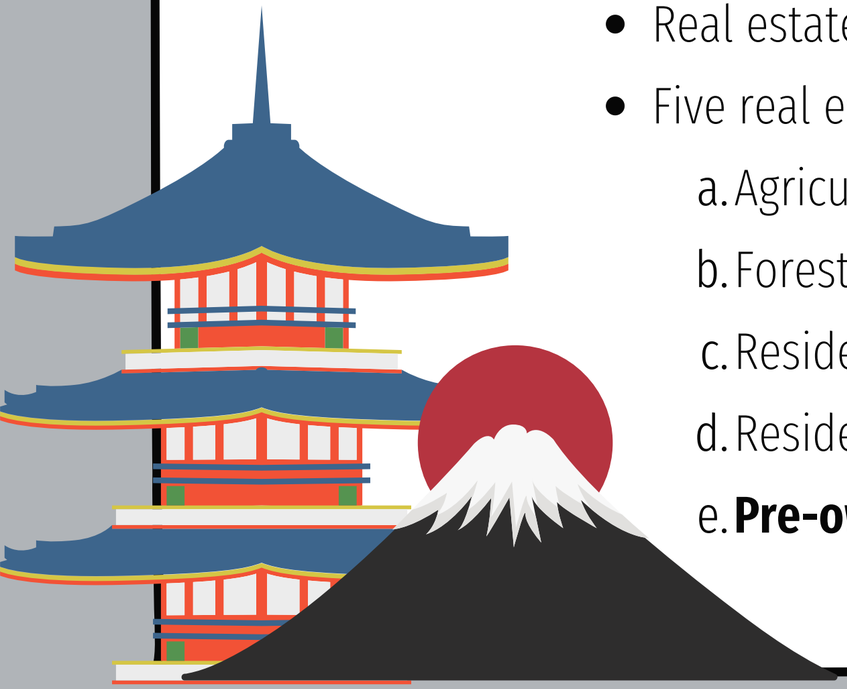
REAL ESTATE

PRICE PREDICTION



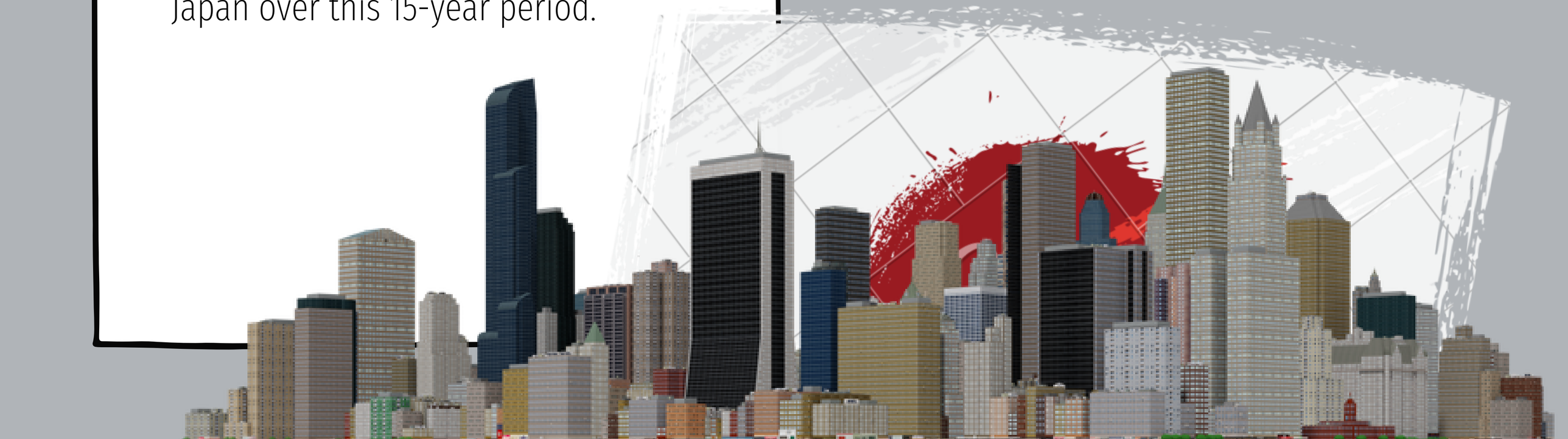
Introduction:

- Dataset contains real estate transaction prices.
- Spans from 2005 to 2019.
- Surveyed by the Ministry of Land, Infrastructure, Transport, and Tourism of Japan (MLIT).
- Real estate transaction prices for 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.



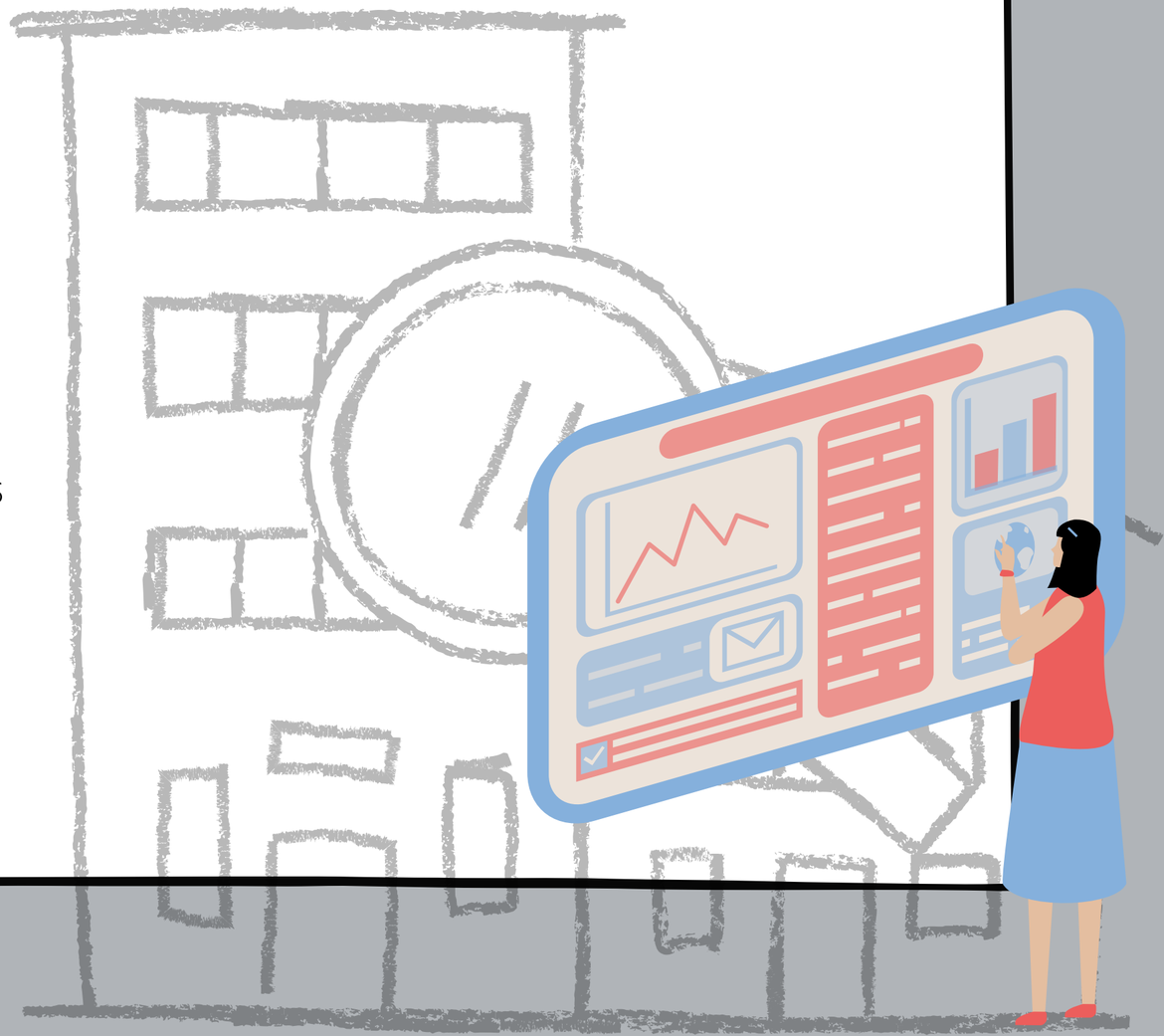
Data Science solutions:

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:

Further for predictions different models can be used:

1. **Regression Models:** Linear regression, Decision Trees to predict property prices based on selected features.
2. **Gradient Boosting:** Improving predictive accuracy.
3. **Neural Networks:** Deep learning models such as neural networks for modeling complex relationships.
4. **Evaluation Metrics:** To assess model performance.



Impact

The analysis of this real estate dataset can have several potential impacts:

- **Market Insights:** It can provide valuable insights into how the Japanese real estate market has evolved, helping stakeholders make informed investment decisions.
- **Investment Decisions:** Investors can use the predictive models to estimate future property prices and make data-driven investment decisions.
- **Regional Comparisons:** Understanding price variations between Tokyo and other prefectures areas can guide property buyers and sellers.

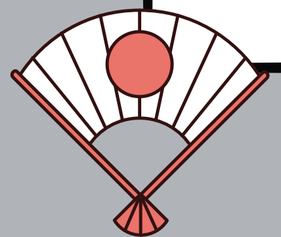
Dataset:

- **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.

Data Quality and other concerns:

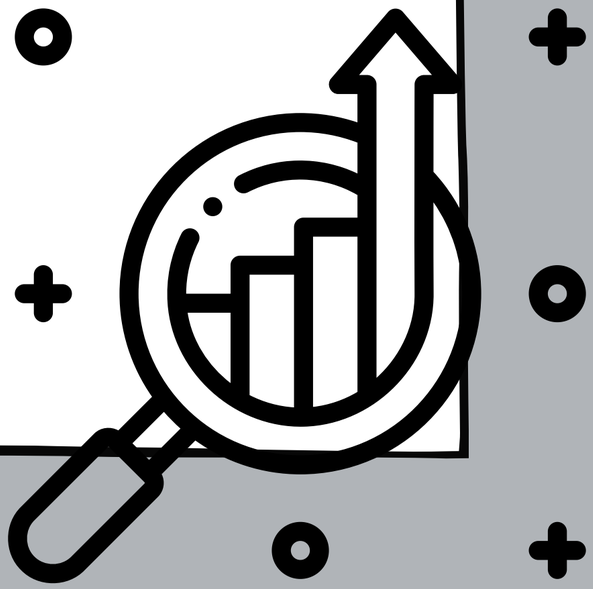
Some of the concerns include:

- Real Estate Type
- No Latitude and Longitude.(granularity of region)
- Null values(617859 rows and 38 columns)
- Nearest Station and time taken.
- Floor Plan

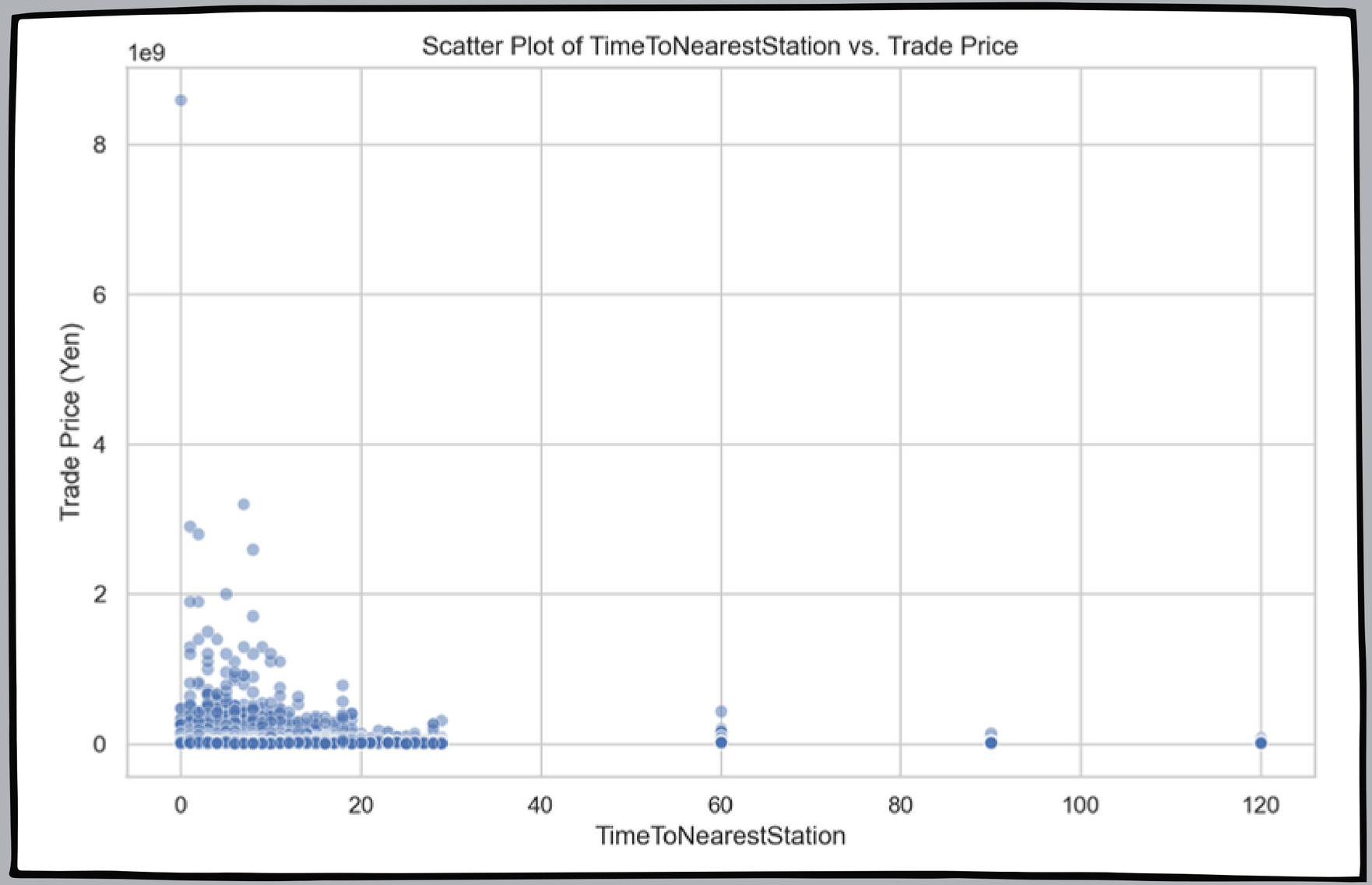


Exploratory Data Analysis

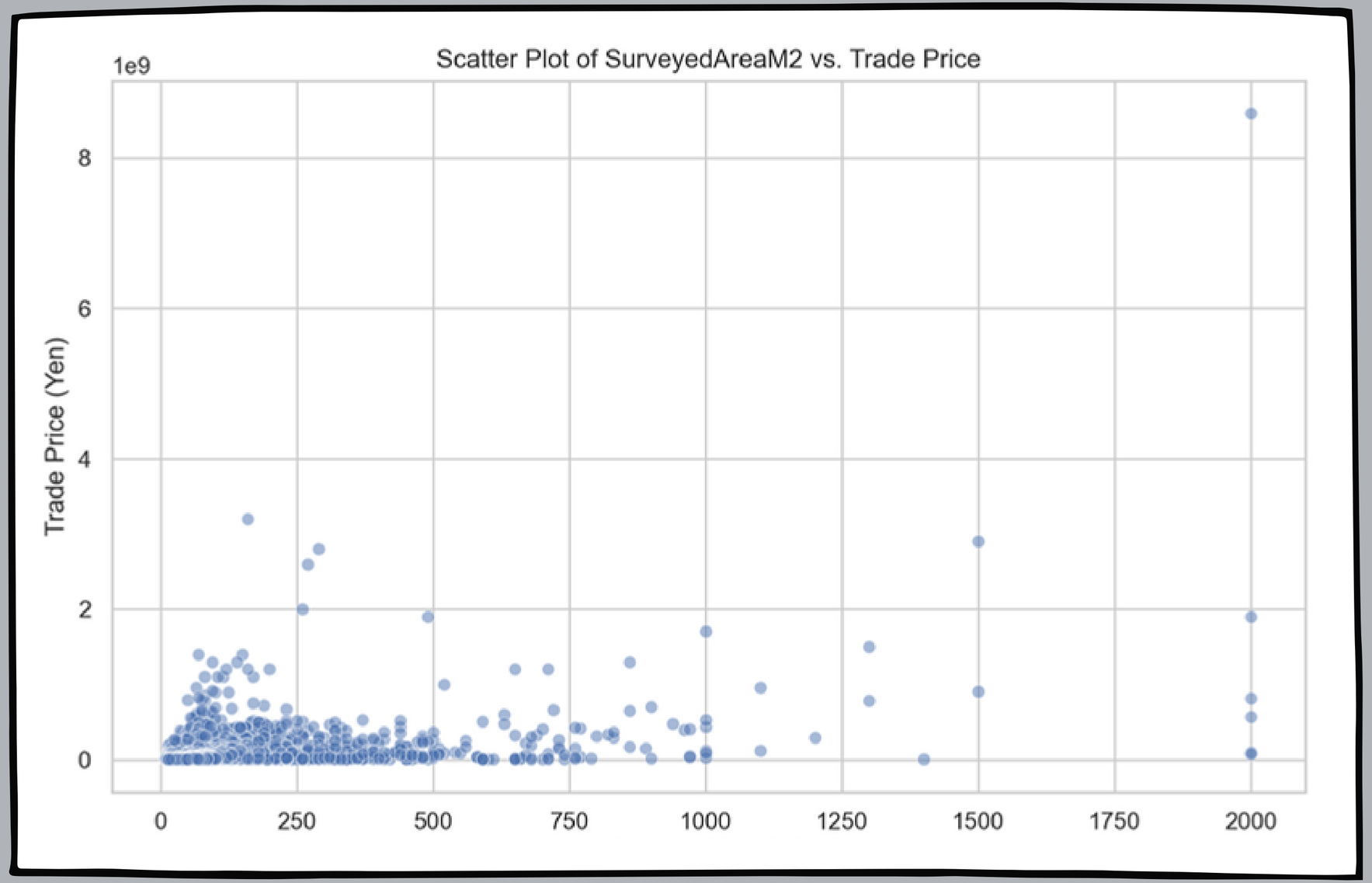
The plots shown in the subsequent slides are the relationship of different variables with **Trade Price.**



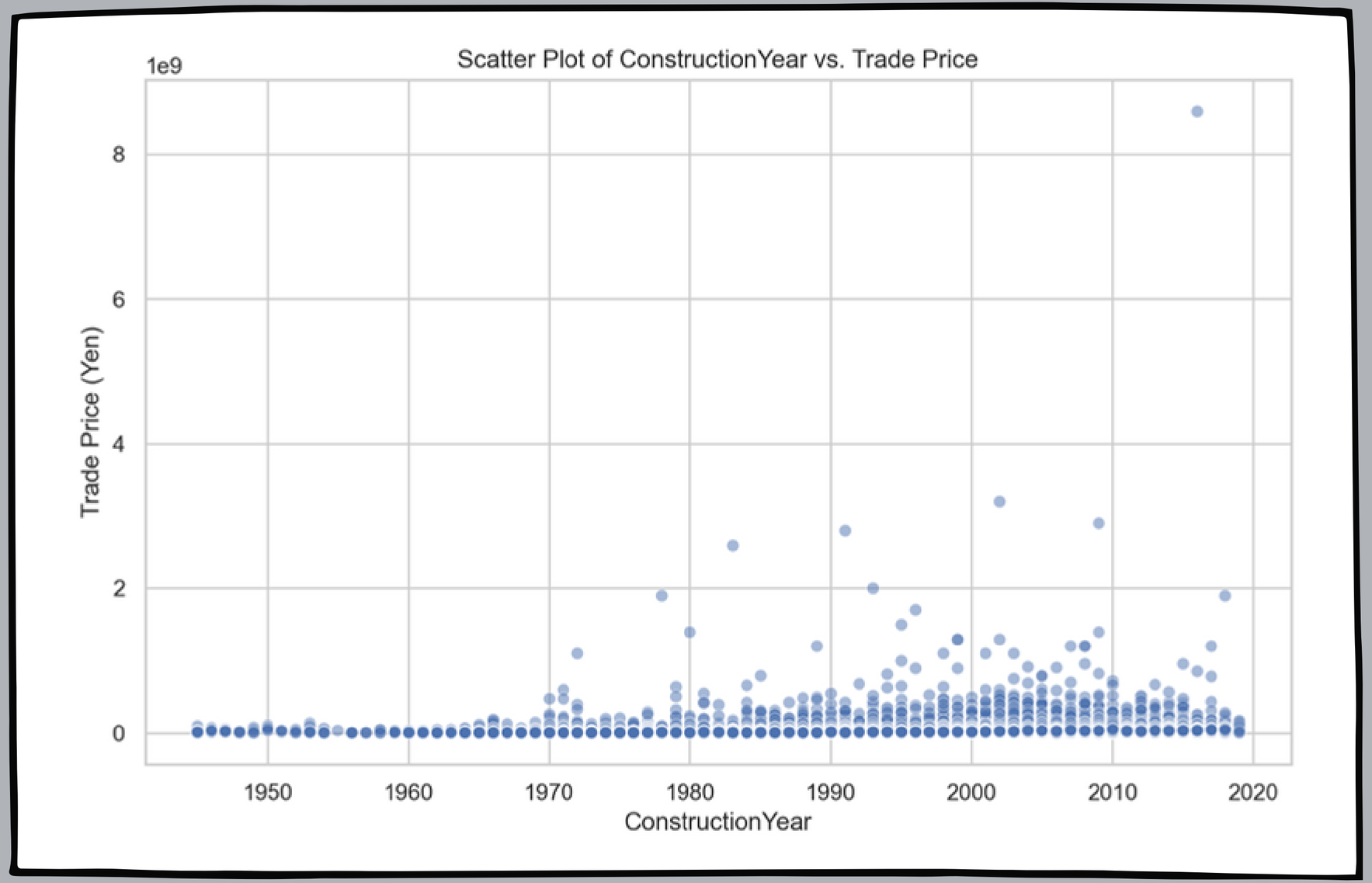
Time to the Nearest Station:



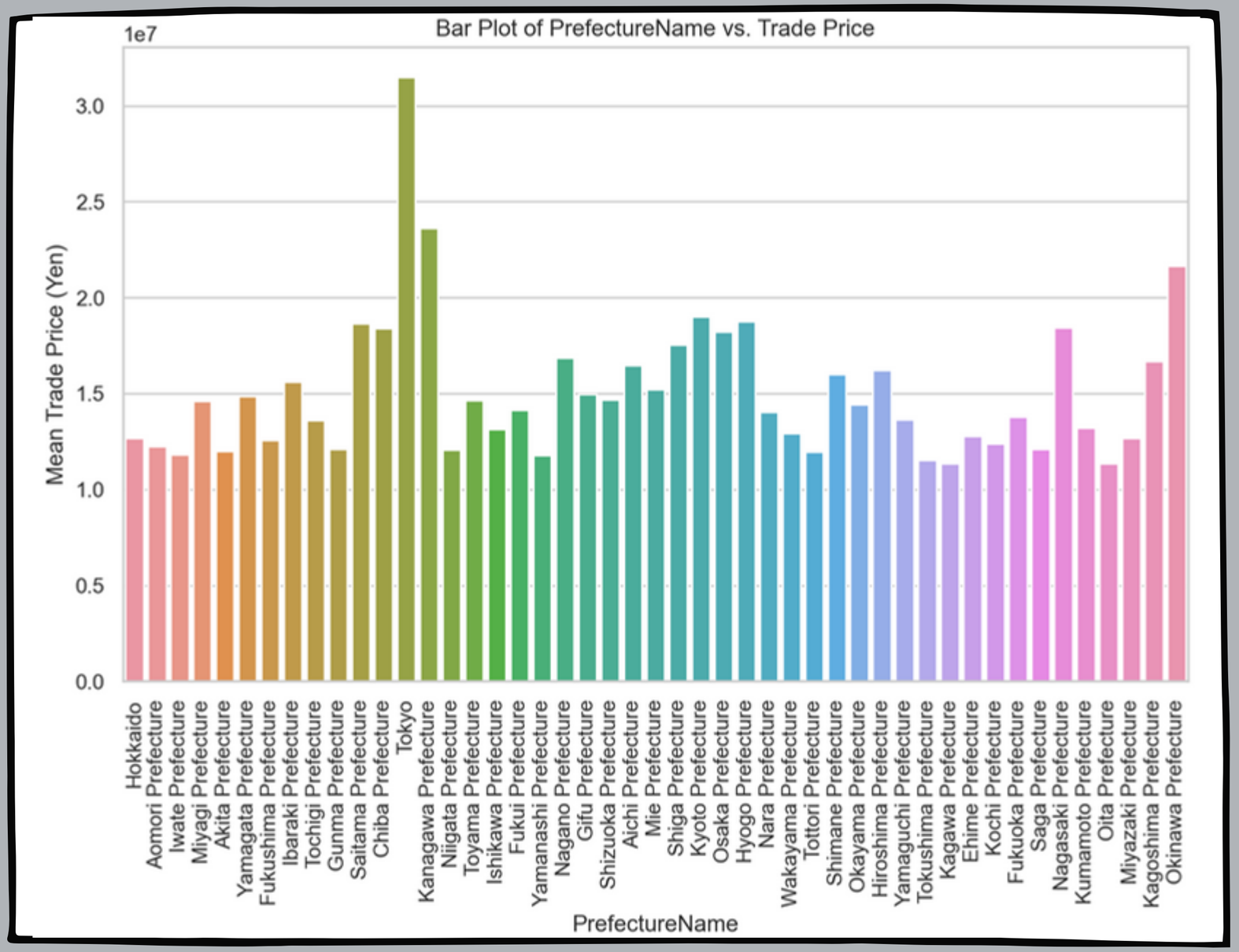
Surveyed Area:



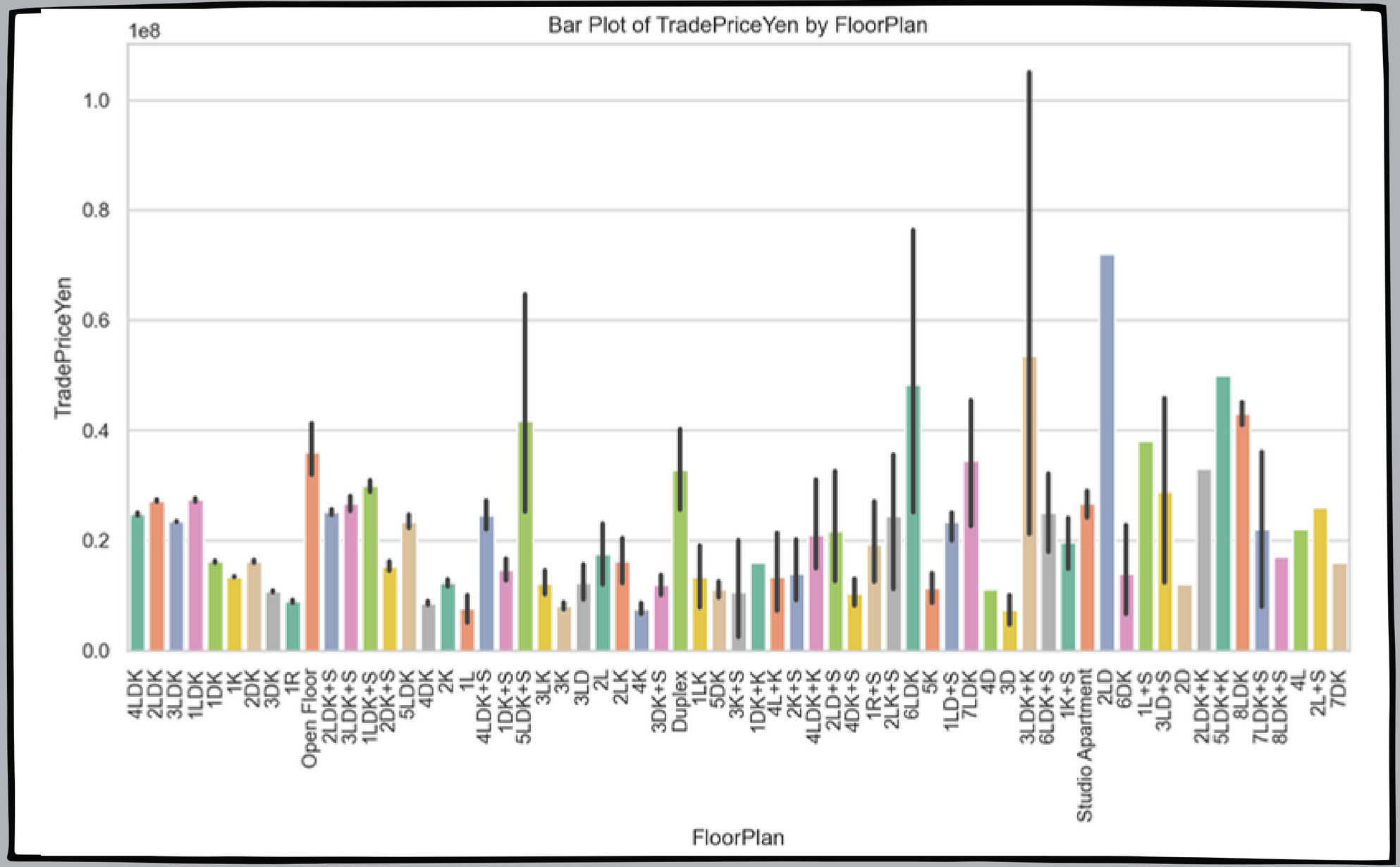
Construction Year:



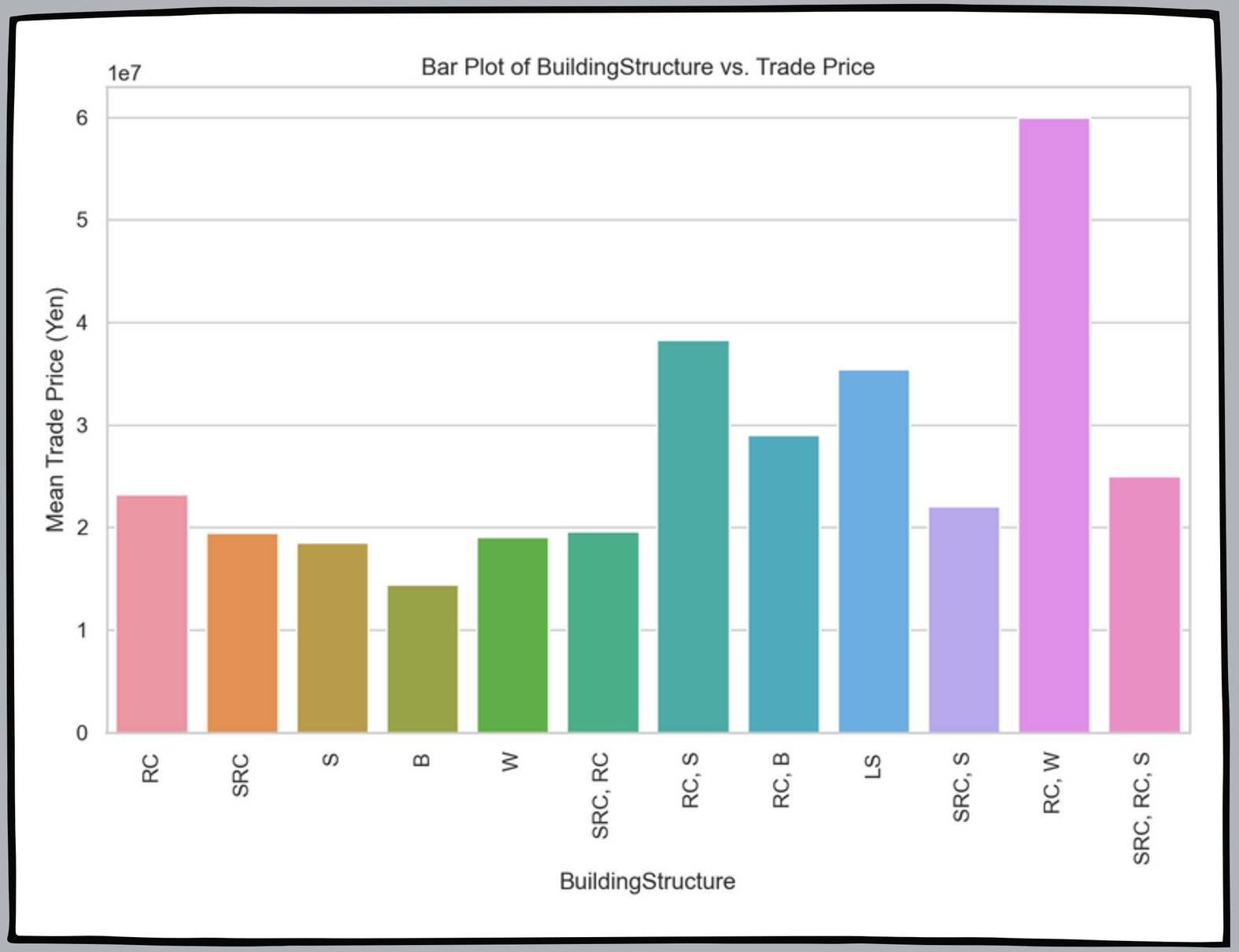
Prefecture Name:



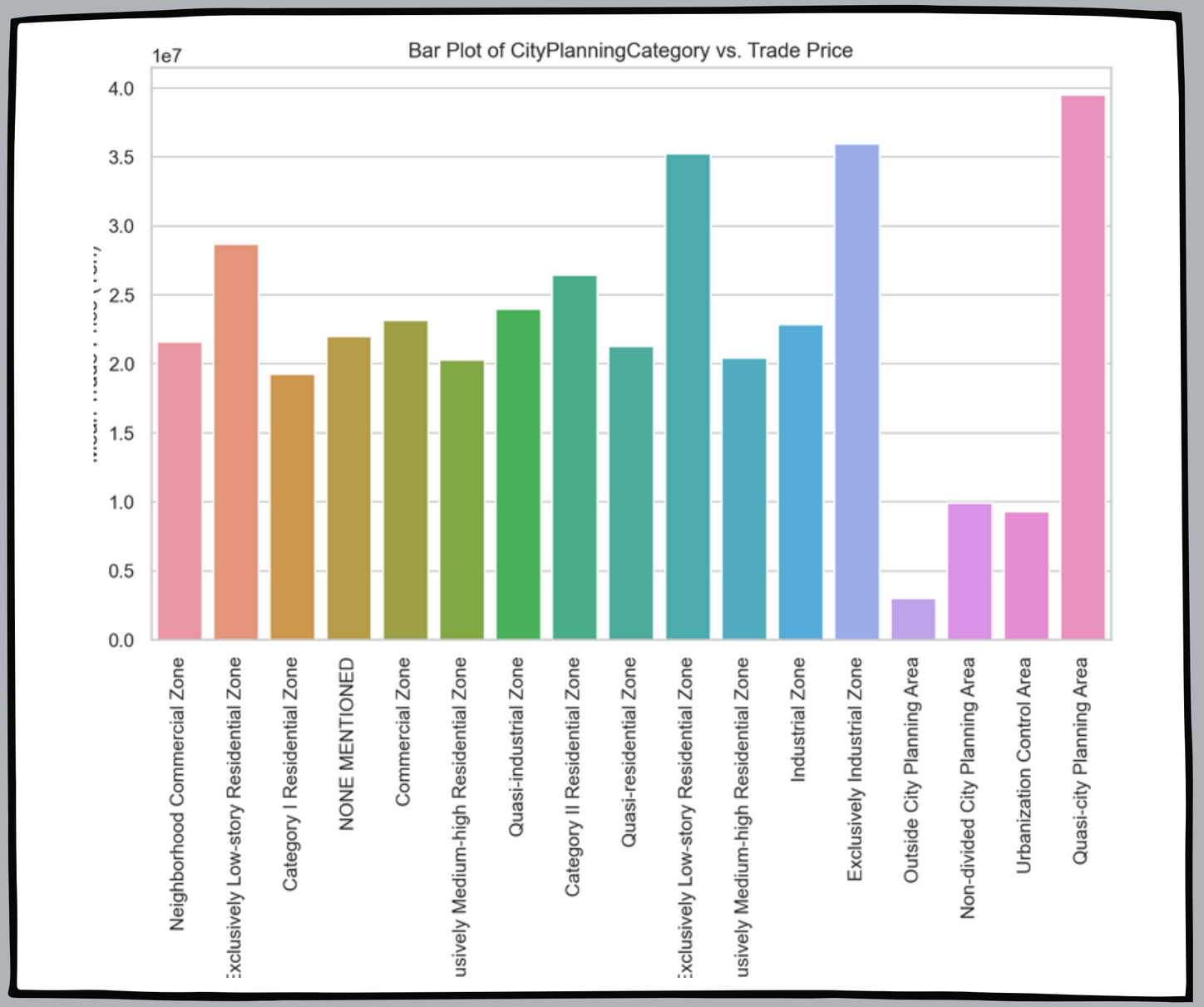
Floor Plan:


Bar Plot of TradePriceYen by FloorPlan
1e8
1.0
0.8
0.6
0.4
0.2
0.0
TradePriceYen
4LDK
2LDK
3LDK
1LDK
1DK
1K
2DK
3DK
1R
Open Floor
2LDK+S
3LDK+S
1LDK+S
2DK+S
5LDK
4DK
2K
1L
4LDK+S
1DK+S
5LDK+S
3LK
3K
3LD
2L
2LK
4K
3DK+S
Duplex
1LK
5DK
3K+S
1DK+K
4L+K
2K+S
4LDK+K
2LD+S
4DK+S
1R+S
2LK+S
6LDK
5K
1LD+S
7LDK
4D
3D
3LDK+K
6LDK+S
1K+S
Studio Apartment
2LD
6DK
1L+S
3LD+S
2D
2LDK+K
5LDK+K
8LDK
7LDK+S
8LDK+S
4L
2L+S
7DK
FloorPlan

Building Structure:



City Planning category:

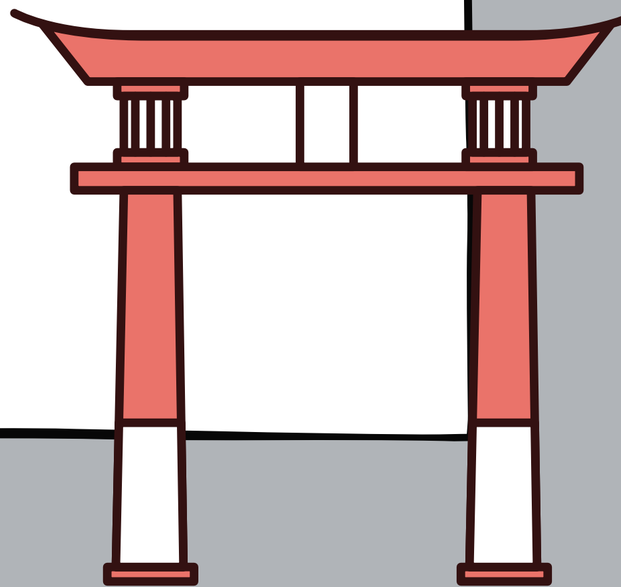


Next steps:

- How to deal with City Names, City code, District Names.
- Multivariate analysis
- Find other ways to deal with null values.

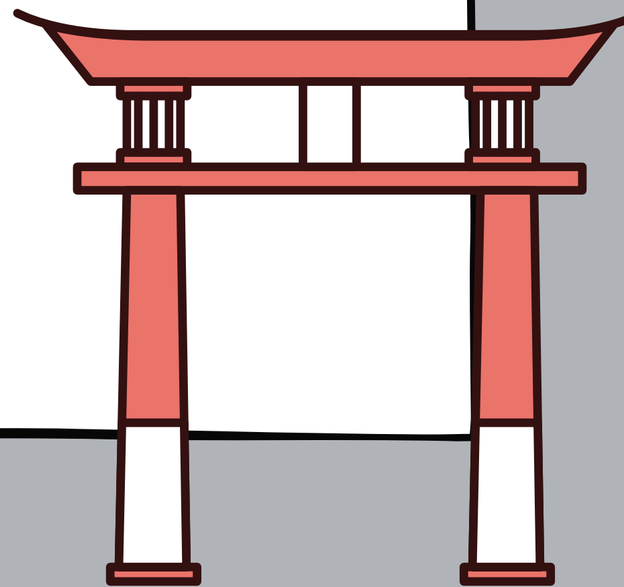
Baseline models:

1. **Linear Regression:** an predict numerical values like property prices based on features such as area, unit price, and more.
2. **Decision Trees:** Decision trees can handle both numerical and categorical features and are useful for regression tasks. They can capture complex relationships in the data.
3. **Random Forest:** An ensemble model that combines multiple decision trees for improved predictive power and generalization.
4. **Lasso Regression:** Lasso is a linear model with L1 regularization, which helps in feature selection. It can be useful for identifying key factors affecting property prices.
5. **Ridge Regression:** Similar to Lasso, Ridge is a linear model with L2 regularization, which can help in reducing overfitting.



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 !