JAPAN

REAL ESTATE PRICE PREDICTION

SPRINT 3

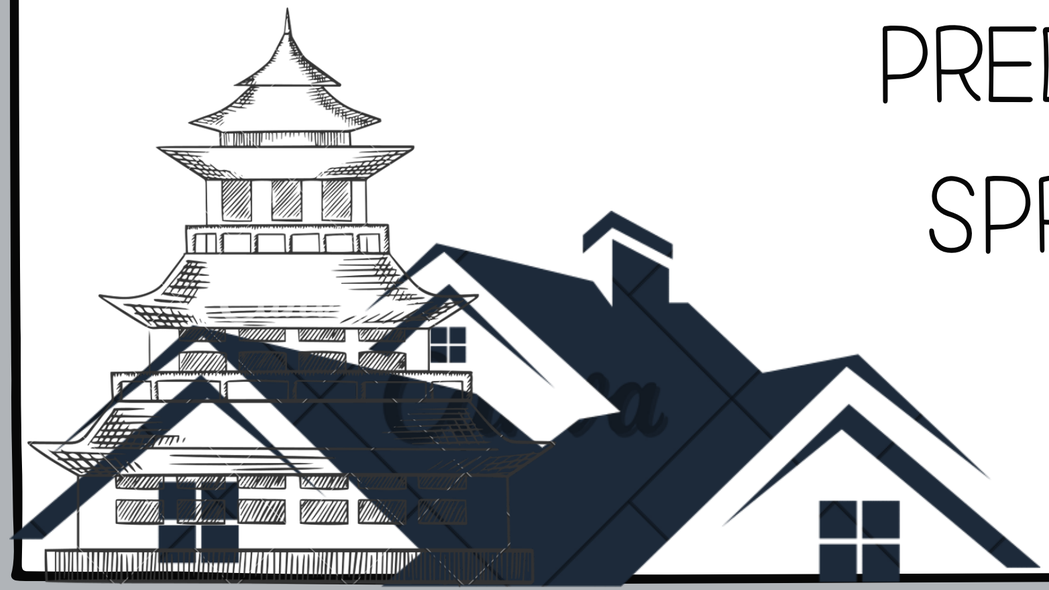
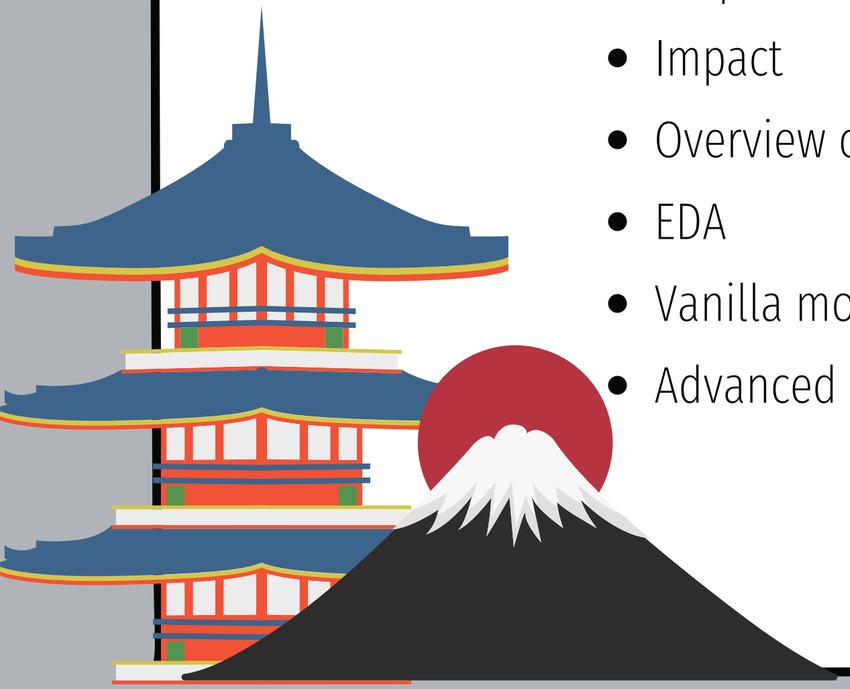
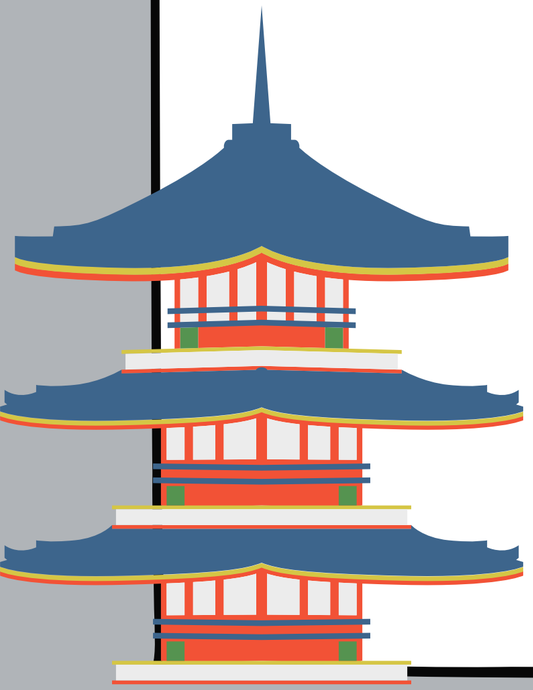


Table of contents:



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.**

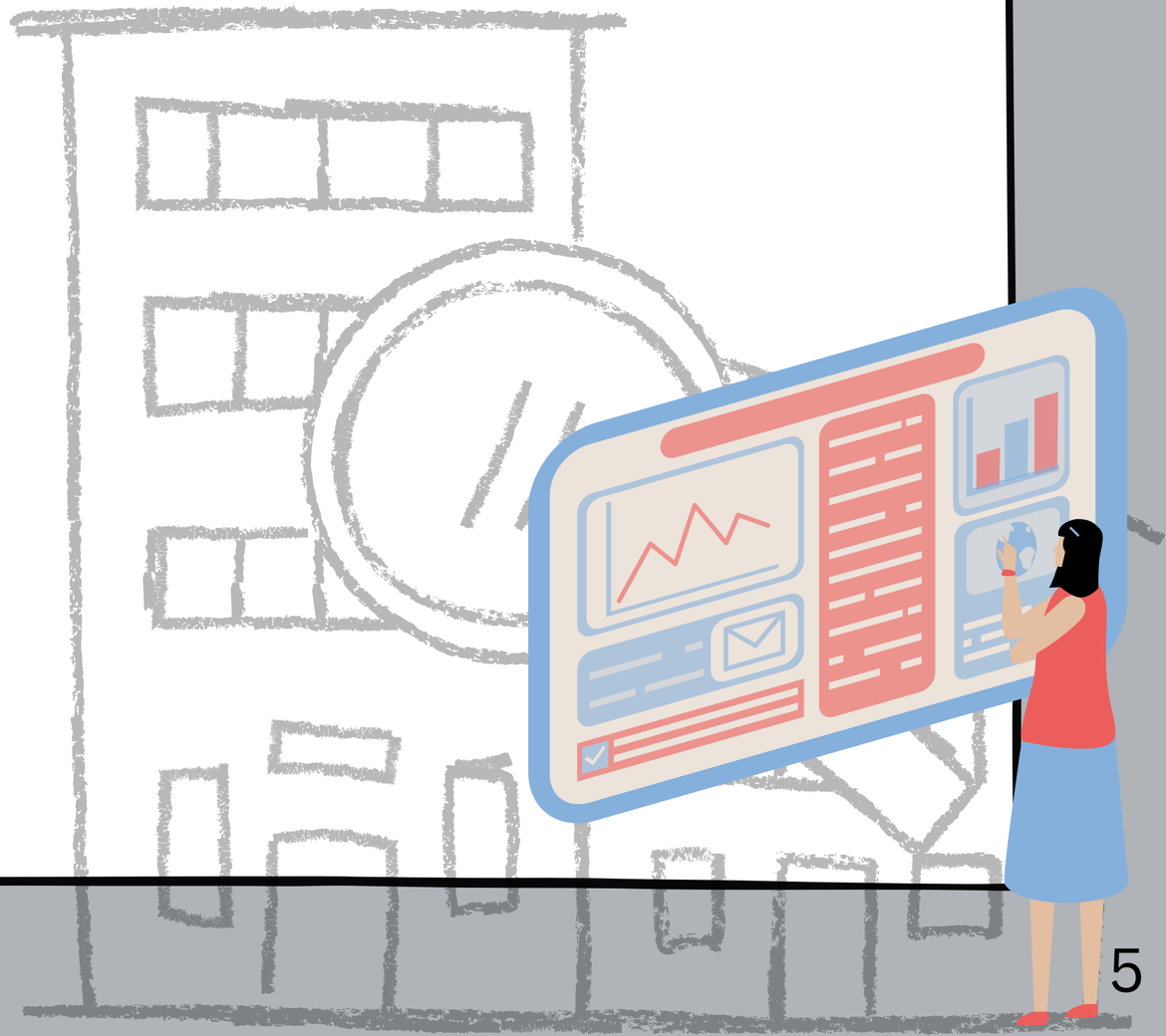
Problem Area:

- The primary goal is to gain insights into the trends and factors influencing real estate prices in Japan over this 15-year period.
- How can we help with National vacancy rate

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:

- **PrefectureName:** Prefecture name of Japan.
- **TimeToNearestStation:** Time to the nearest station (in minutes).
- **TradePriceYen:** Transaction prices in Japanese Yen.
- **TradePriceCAD:** Transaction prices in Canadian Dollars.
- **FloorPlan:** Property floor plan (e.g., 3LDK, 2DK).
- **SurveyedAreaM2:** Surveyed area in square meters.
- **BuildingStructure:** Building structure (e.g., Steel frame reinforced concrete, Wooden).
- **CurrentUsage:** Current property usage (e.g., House, Office, Factory).
- **FutureUsePurpose:** Purpose of future use (e.g., House, Shop, Office).
- **CityPlanningCategory:** Use districts designated by the City Planning Act.
- **ConstructionYear:** Construction year of the building.
- **TransactionYear:** Time of transaction year.
- **TransactionYearQuarter:** Time of transaction year-quarter.
- **RenovationStatus:** Renovation status.

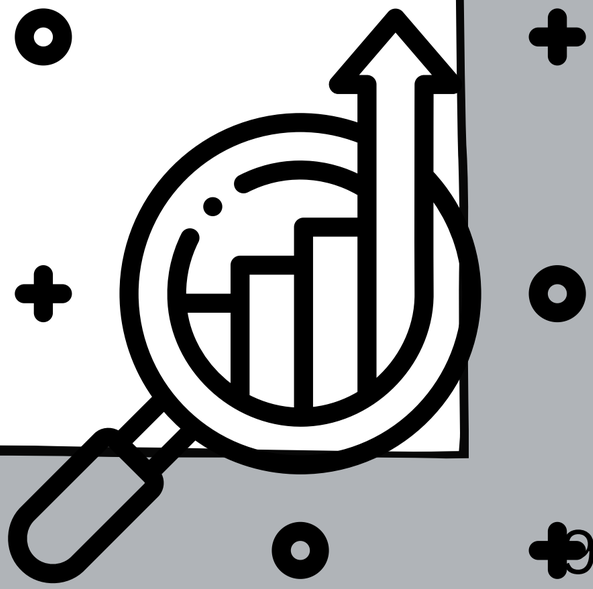
Processing procedures:

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’

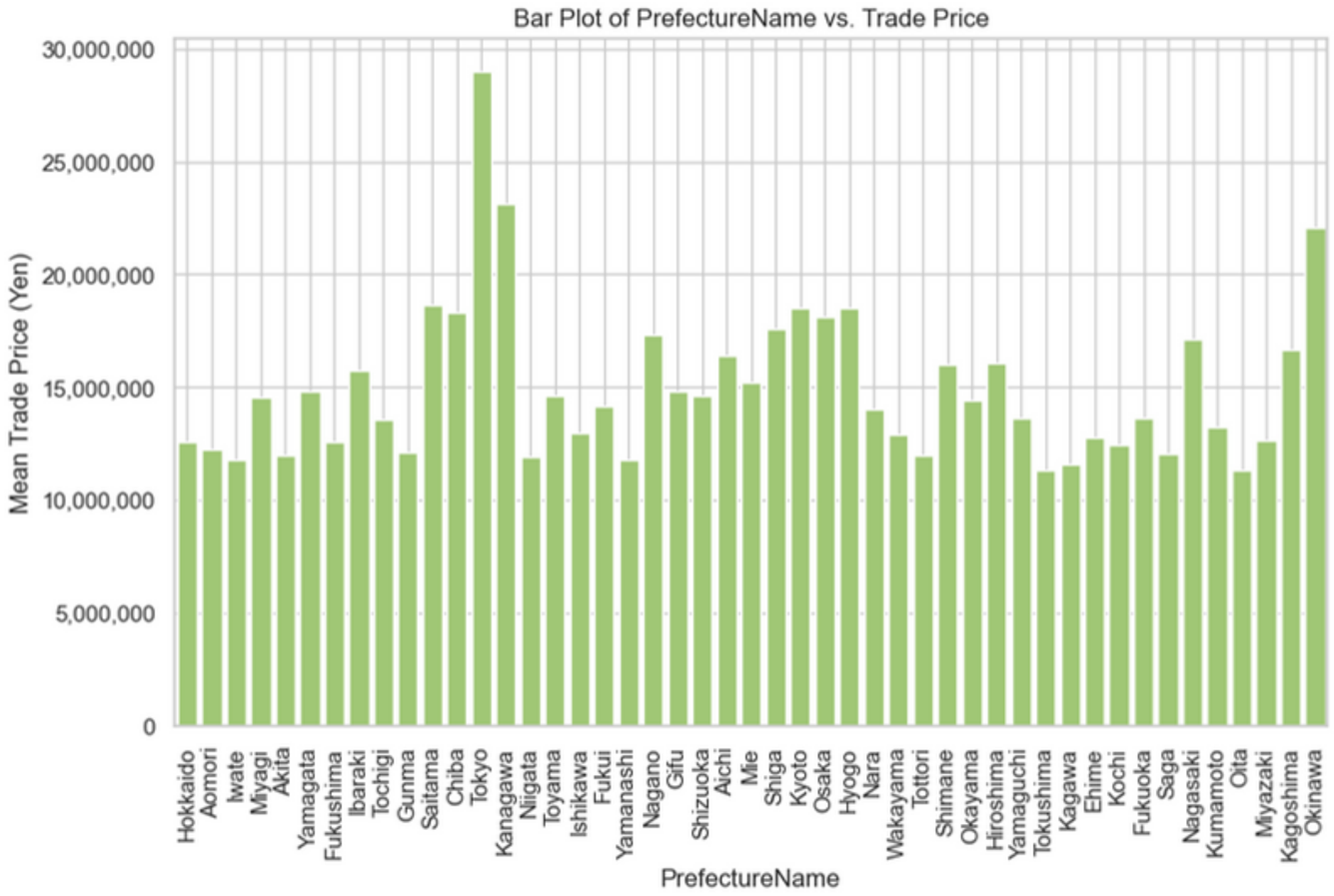
Exploratory Data Analysis

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



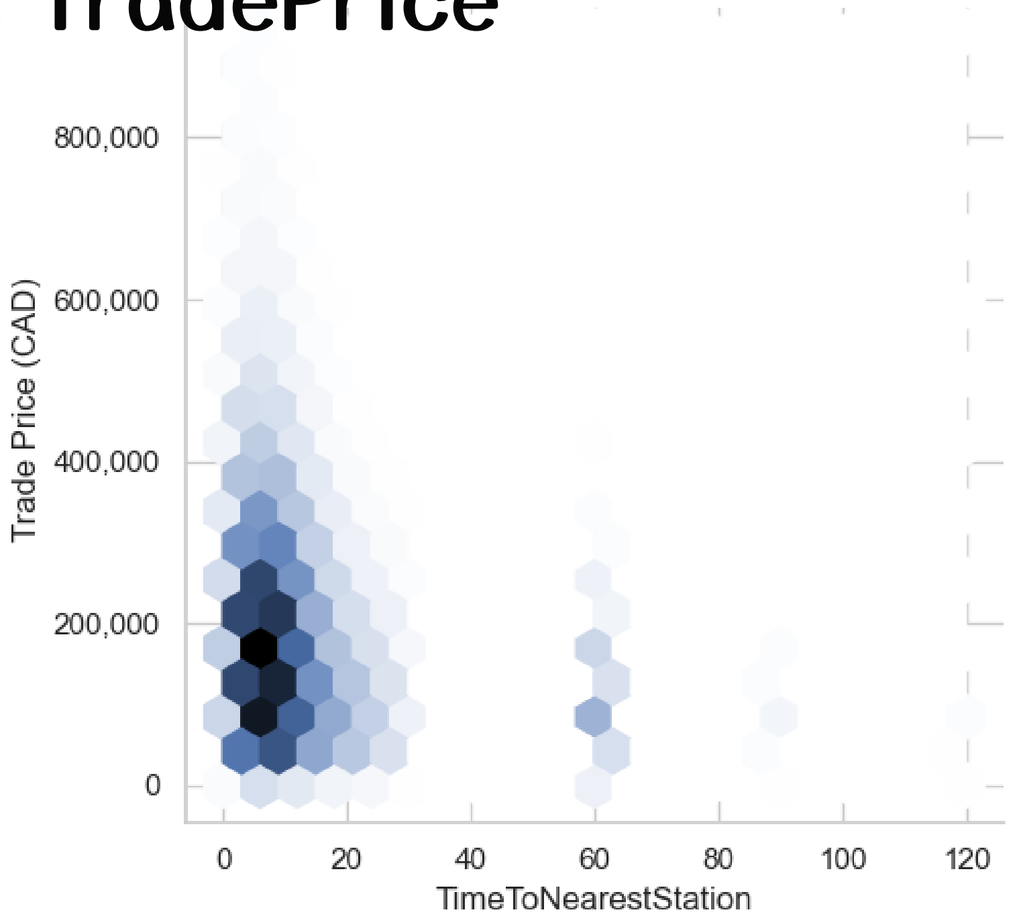
Prefectures Vs TradePrice

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



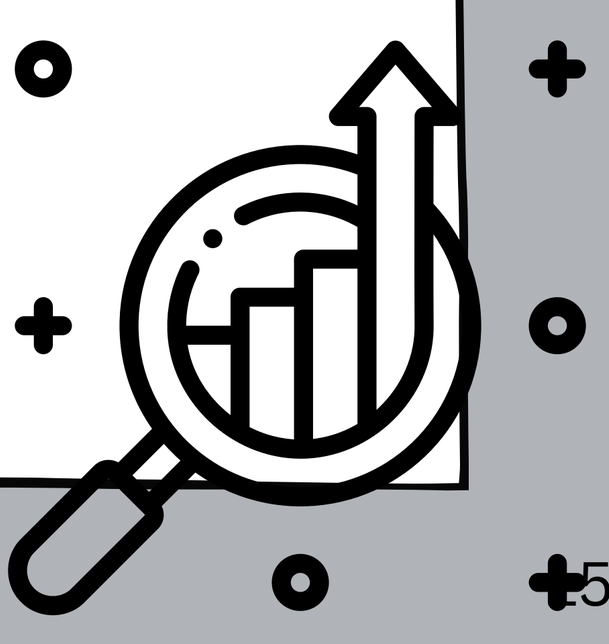
Time to nearest Station Vs TradePrice

- Most of the houses are **lesser than 20 mins** to the Nearest Station
- The **most expensive** houses are the ones **nearest to a station**



Vanilla Models

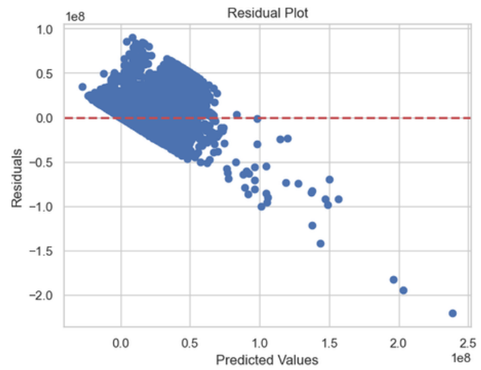
- Linear Regression
- Random Forest
- XG Boost

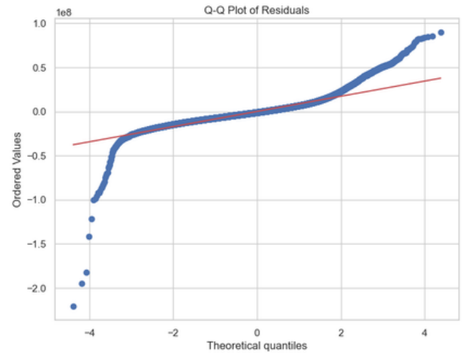


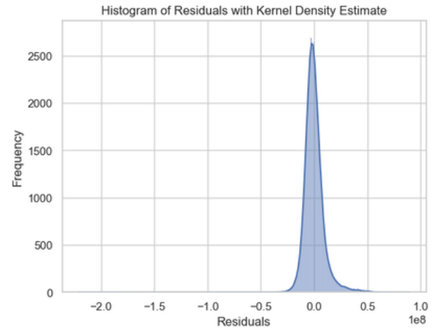
Linear Regression:

- **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** .

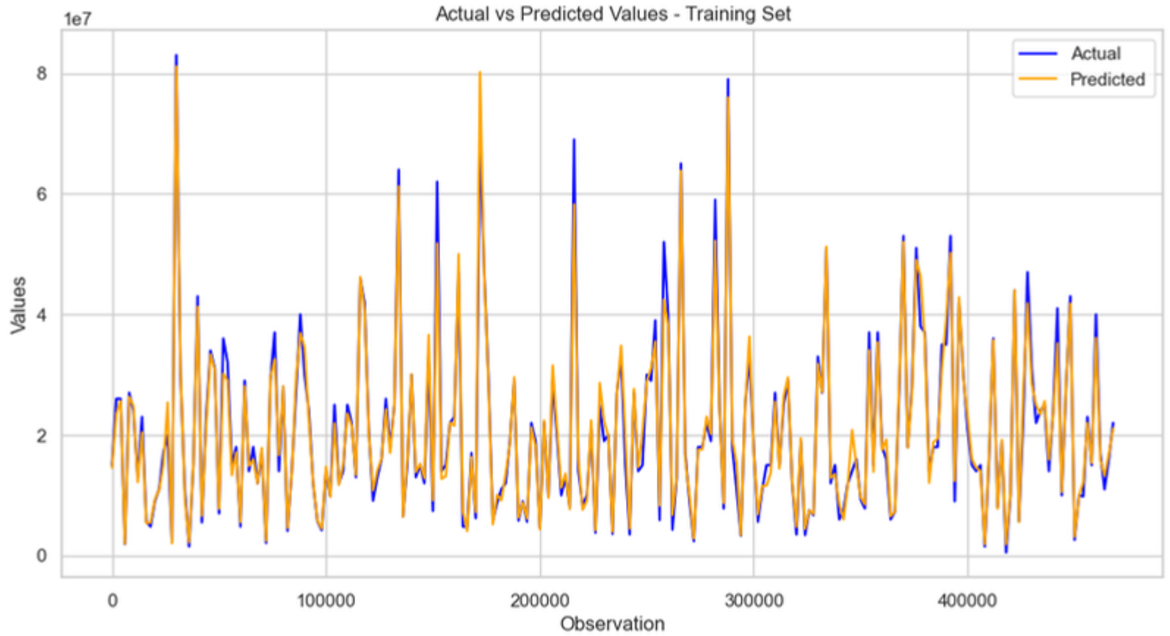
Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
Linear Regression	0.89213128	1.03377171

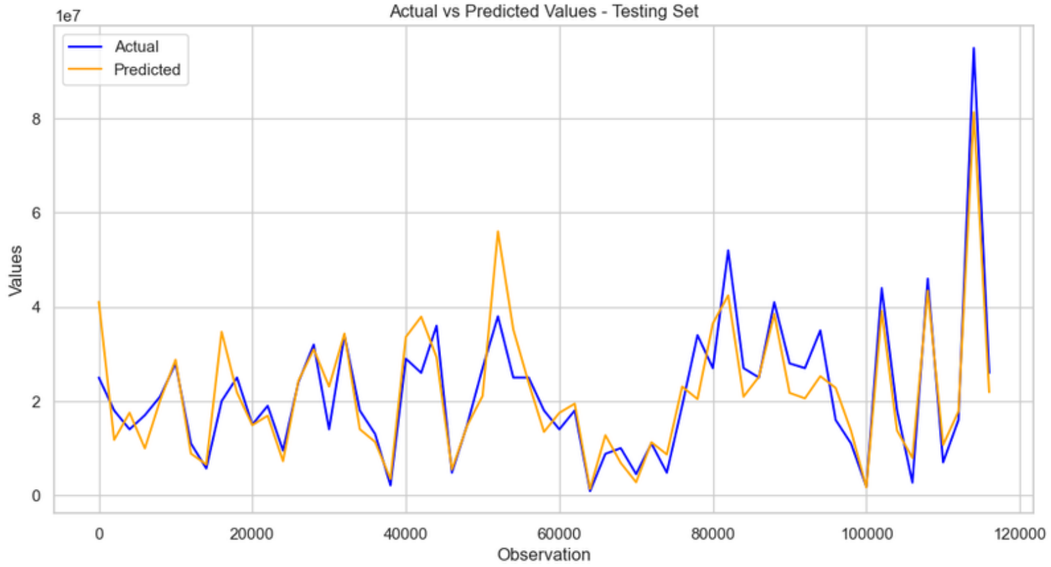






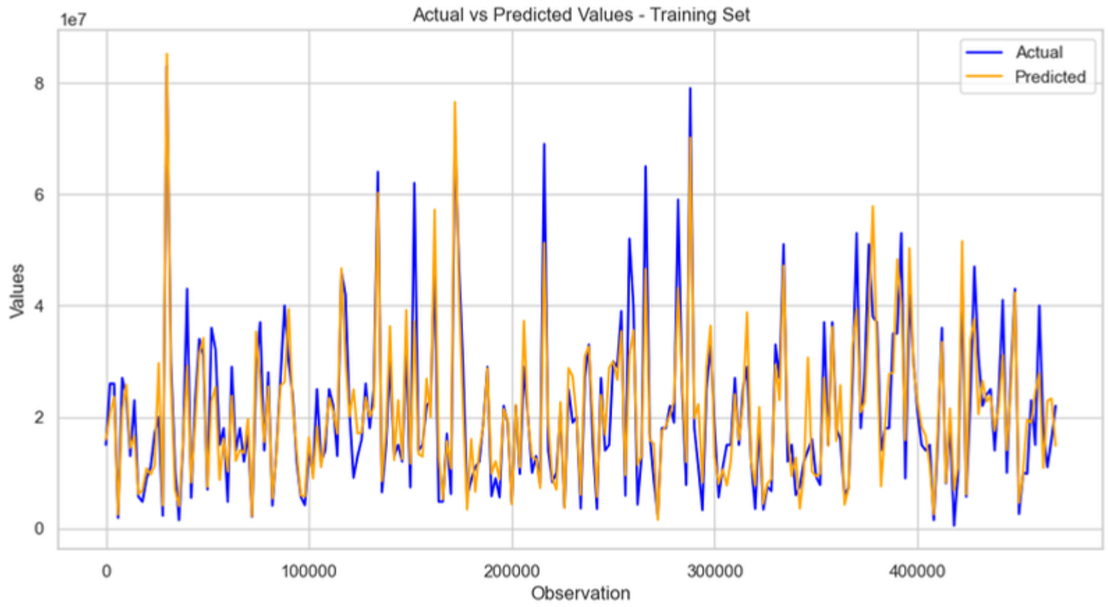
Random Forest

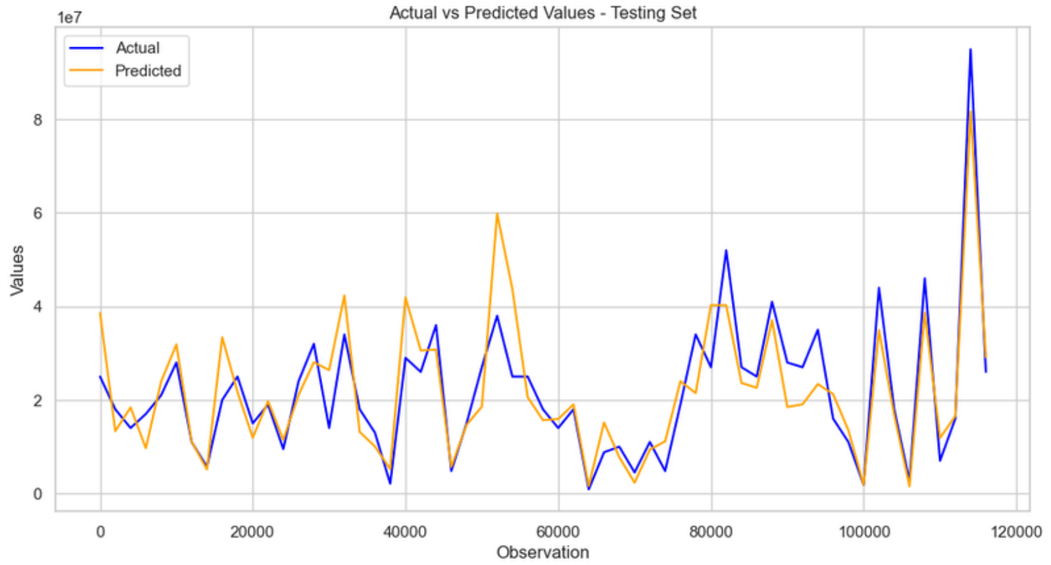




Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
Random Forest Regressor	0.21869275	0.7446585

XGBoost

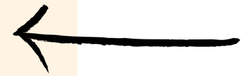




Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
XGBoost	0.66353553	0.743996215

Evaluation Metric of all the models

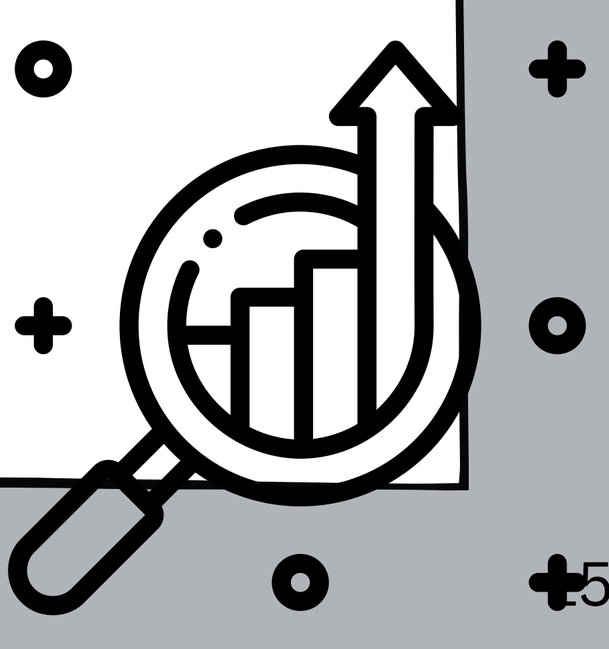
Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
Linear Regression	0.89213128	1.03377171
Random Forest Regressor	0.21869275	0.7446585
XGBoost	0.66353553	0.743996215



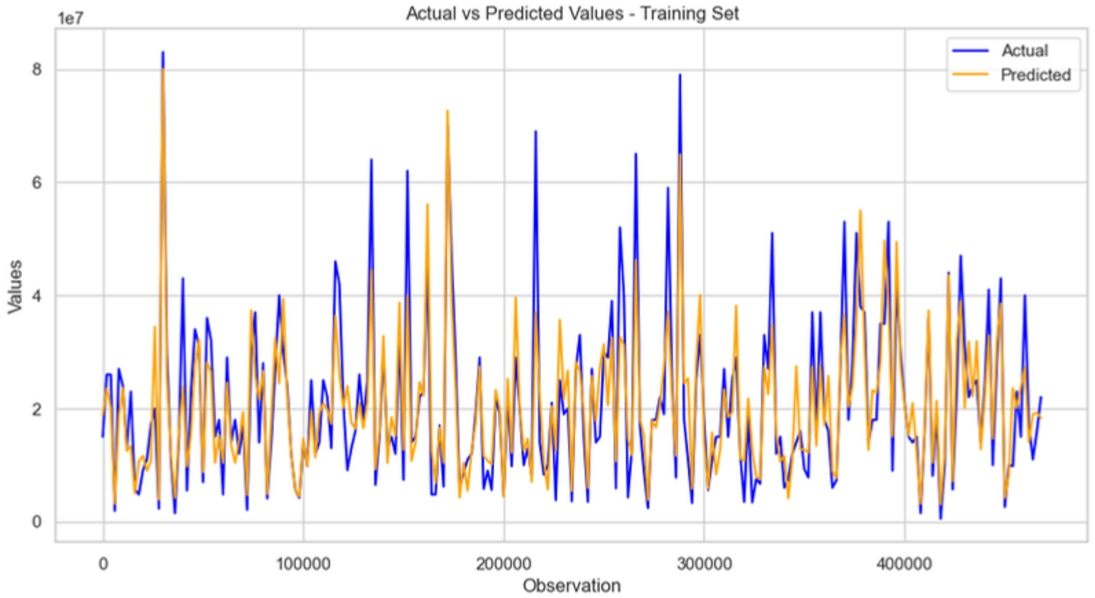
Overfitted

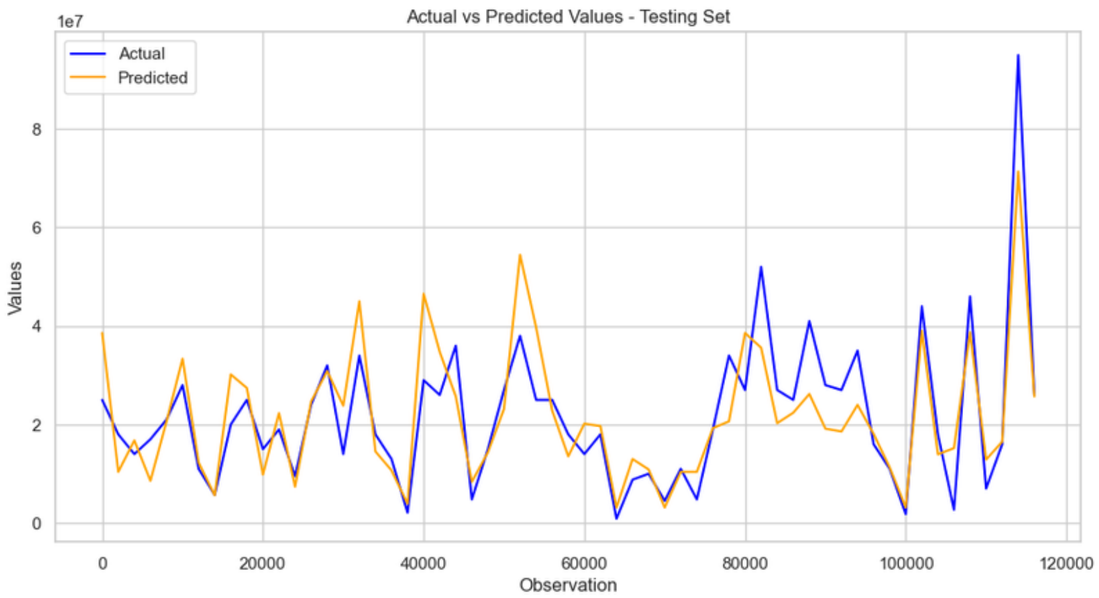
Advanced Modelling

- Random Forest
- XG Boost



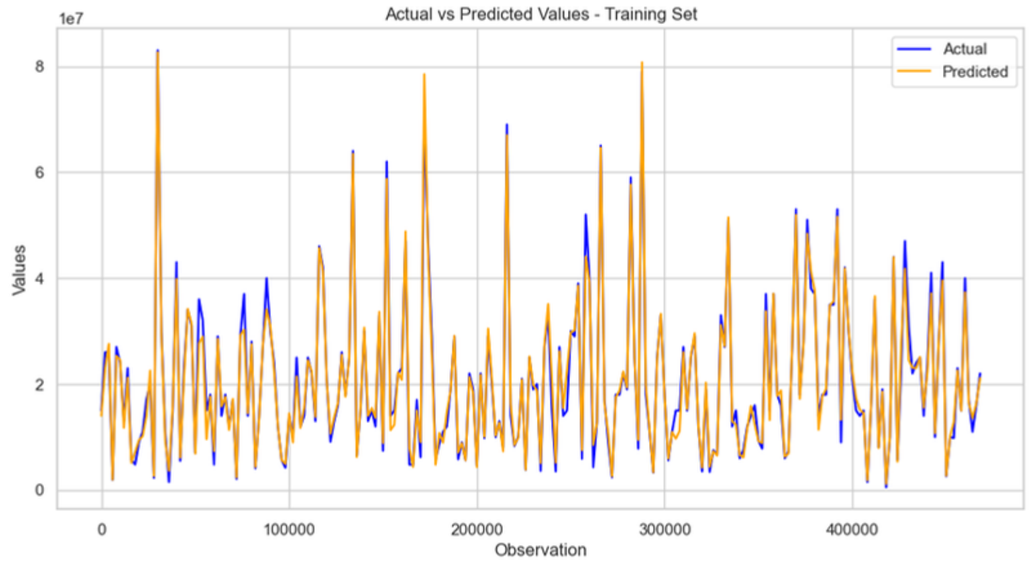
Random Forest with hyperparameter tuning

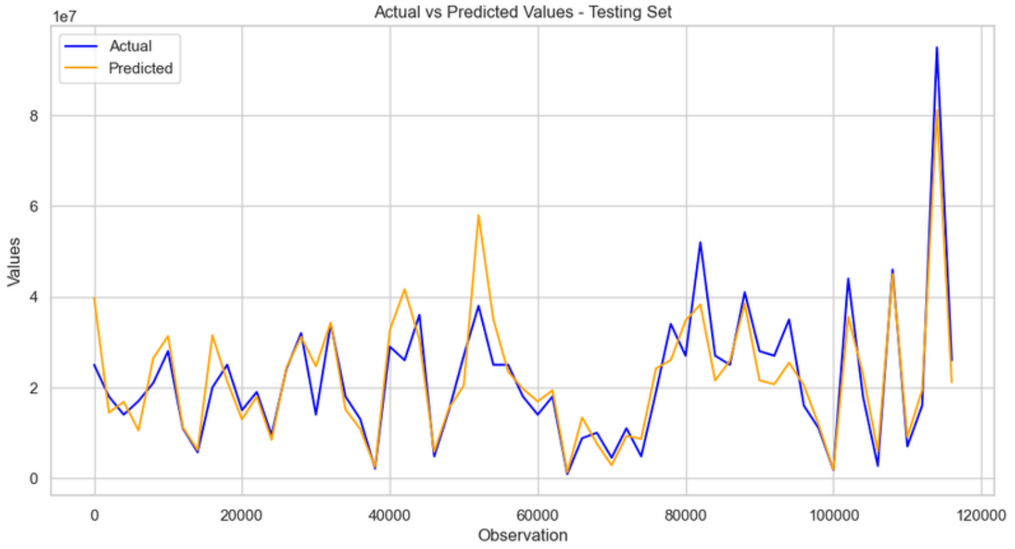




Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
Random Forest Regressor	0.70675262	0.81756413

XGBoost





Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
XGBoost	0.2854715	0.7399581

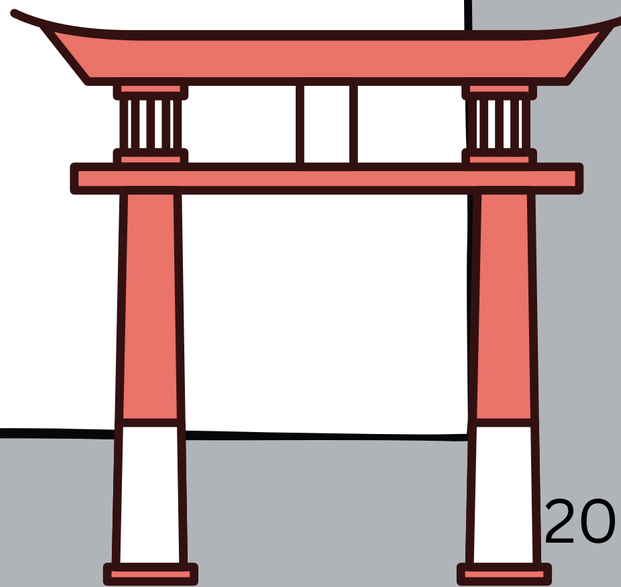
Evaluation Metric of the advanced models

Model	Mean Absolute Percentage Error in TRAIN	Mean Absolute Percentage Error in TEST
Random Forest Regressor	0.607484	0.796113
XGBoost	0.639098	0.7922850

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.



Thank You !