Apartment Daegu Price Analysis

By. Vito Muhammad

Background

- Apartments are one of the answers to the housing needs of modern society due to limited residential land and dense business activities in urban areas especially Daegu.

- Individuals or companies usually offer apartment units. Bidders can sell these units through a platform by determining the prices of their apartments. Therefore, for apartment owners, it is often difficult to adjust the prices of their apartments to market prices. Similarly, it can also be challenging for buyers to determine the appropriate purchase price based on market conditions.

BUSINESS UNDERSTANDING

Background	Apartments in Daegu have become the preferred choice for residential dwellings among its residents. This has resulted in a significant number of real estate agents offering their services to facilitate apartment transactions.
Analytic Approach	Building a model that will assist real estate agents in providing a predictive tool for estimating the selling price of an apartment based on given features.
Objectives	● Make Machine Learning Model to Prediction Sell Price
Business Metrics	*Sell Price (USD)*
Metric Evaluation	Among the commonly used evaluation metrics for prediction such as MSE, RMSE, and MAPE, this time I choose to use MAE (Mean Absolute Error).

02 EXPLORATORY DATA ANALYSIS (EDA)

❑ Data Description
❑ *Business Insight*
❑ Data Visualization

Data Descriptions

HallwayType	Type of hallway entrance to the apartment.
TimeToSubway	Time required to reach the nearest subway station (in minutes).
SubwayStation	Name of the nearest subway station.
N_FacilitiesNearBy(ETC)	Number of facilities near the apartment, such as hotels, special needs schools.
N_FacilitiesNearBy(PublicOffice)	Number of public office services near the apartment.
N_SchoolNearBy(University)	Number of universities near the apartment
N_Parkinglot(Basement)	Number of parking lots in the basement.
YearBuilt	Year the apartment was built
N_FacilitiesInApt	Number of facilities for residents, such as swimming pool, gym, playground
Size(sqf)	Apartment size in square feet
SalePrice	Apartment price (USD)

Data Descriptions

Daegu Apartment Data

About Dataset
The dataset contains information about sell price of apartment in Daegu with any feature.

Shape
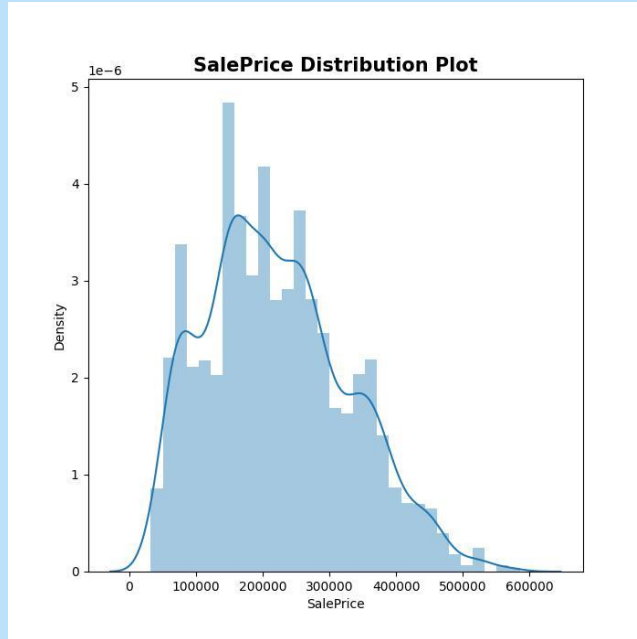*4123 rows x 11 columns (10 Features and 1 Target : Sell Price)*
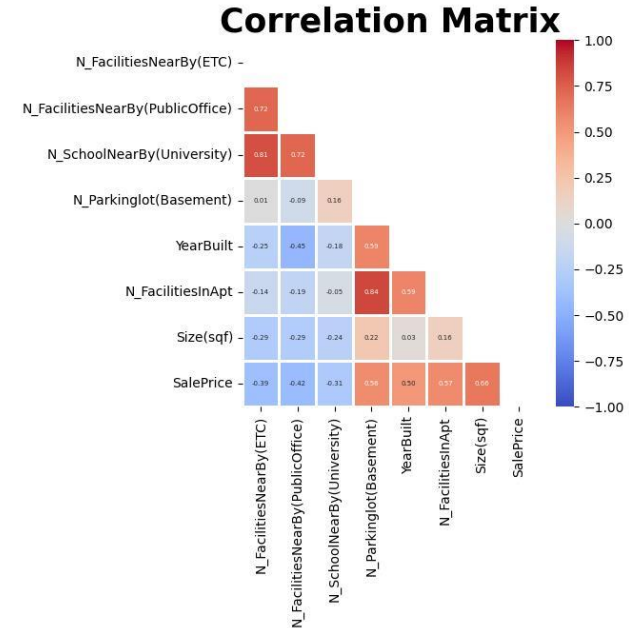
Missing Value
0 Null Values on Income Column

Duplicated Data
1422 duplicated data
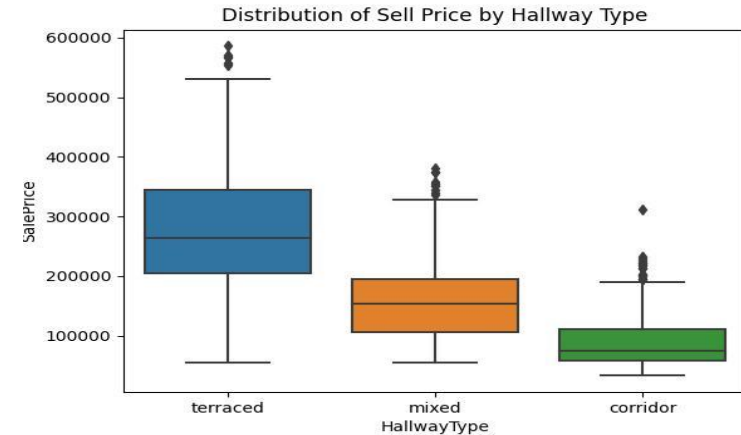
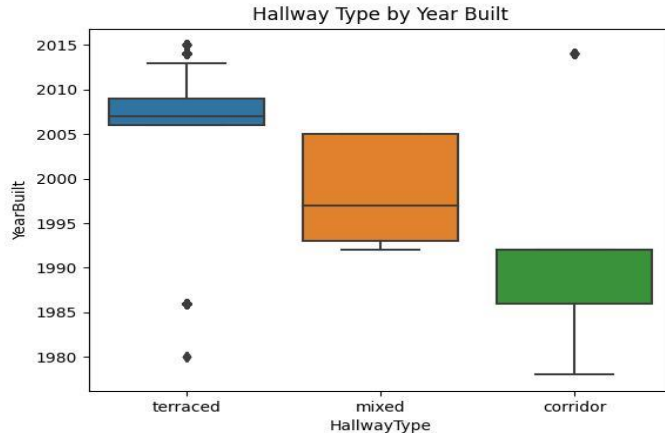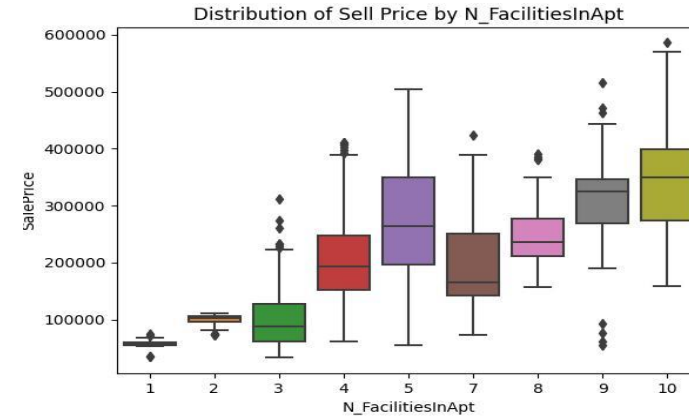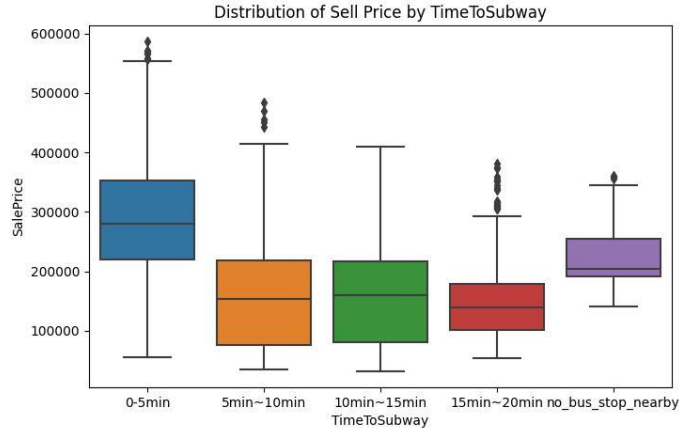# Sell Price Distribusion

Data is not normal and have right skewed

Correlation Between Feature

There are 4 feature have medium correlation to sale price

Feature to Response Category

Distribution of Sell Price by TimeToSubway

Distribution of Sell Price by N_FacilitiesInApt

Hallway Type by Year Built

Distribution of Sell Price by Hallway Type

Purwadhika
Digital Technology School

PEDESTAL PROPERTIES

03
Data
Preprocessing

"Preparing the data before the modeling process"

Data Cleansing

Handling Duplicate

Remove duplicate data to avoid bias result.

Handle Outlier

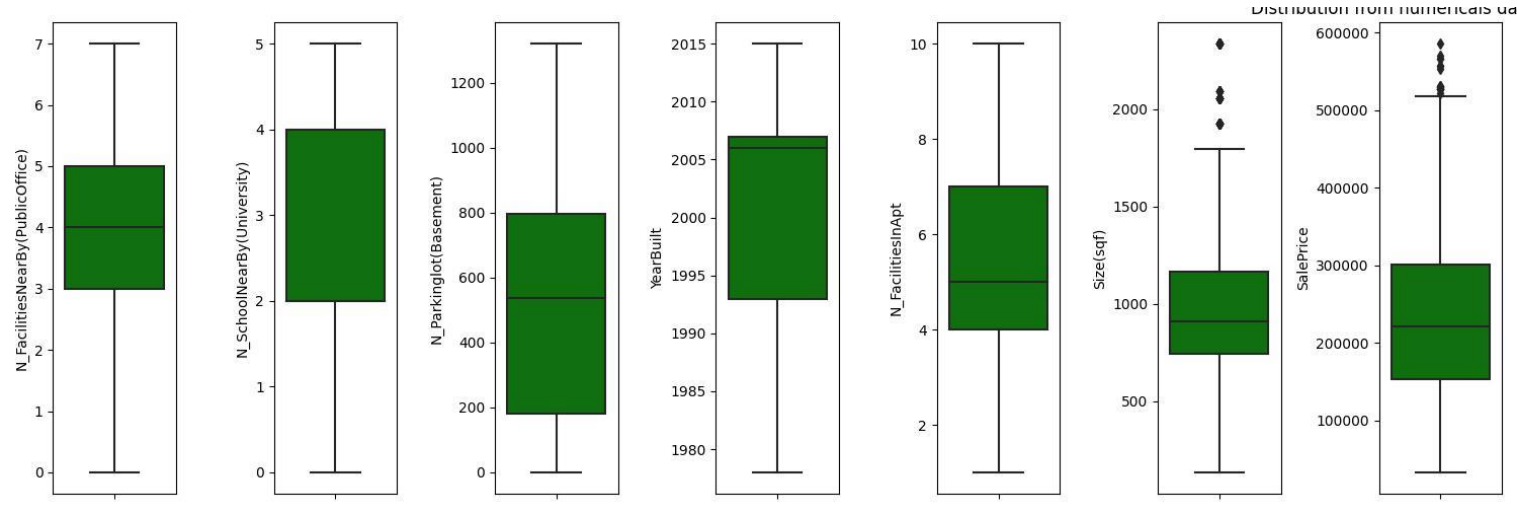Identify and handle outliers in the dataset by making informed decisions based on domain knowledge.

Handle Duplicate

Total Data Before Remove Duplictae	4123
Total Data Duplicate	1422
Total Data After Remove Duplicate	**2701**

Handle Outlier

After we research in one of Real Estate Agent in Korea [My Home Real Estate], it turns out that there are valid outlier data in terms of size and sale price of apartments, and they are a genuine part of the data. Therefore, the outliers on the boxplot do not need to be removed.

Data Treatment
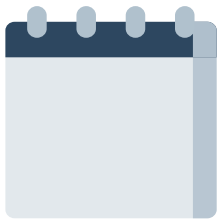
Data Transformation

Used function **Robust Scaller** before Split Data

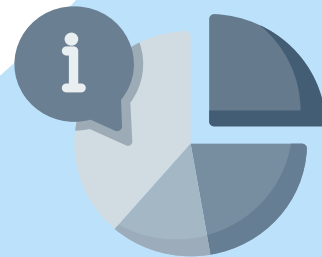**Feature Encoding**

HallwayType
(**OneHotEncoding**)

SubwayStation
(**Binary Encoding**)

TimeToSubway
(**Ordinal Encoding**)

Split Data

Rasio
75 : 25
Train : Test

04
Machine Learning
Model & Evaluation

STRATEGY

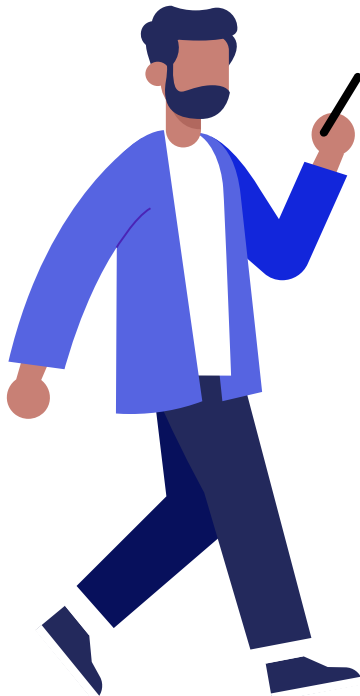| STEP 1	STEP 2	STEP 3
Choosing the model algorithm.	Performing model.	Evaluation stage and Hyperparameter Tuning process.

Model Preparation

Regression Model Candidates :

- Linear Regression
- KNN Regressor
- Decision Tree Regression
- Random Forest Regression
- Extreme Gradient Boosting (XGBoost)
- Light Gradient Boosting Machine (LGBM)

Evaluation Metrics :

- RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and **MAE** (Mean Absolute Error)

Why was MAE chosen? Because MAE (Mean Absolute Error): MAE is suitable for predicting apartment prices. **MAE** (Mean Absolute Error) is an evaluation metric that **measures prediction errors by calculating the average of the absolute differences between predicted values and actual values.**

Benchmark Model

	model	MAE	MAPE	RMSE
2	Decsion Tree	36793.783123	0.189502	45658.591413
4	XGBoost	36816.916154	0.189754	45685.413717
3	Random Forest	36849.929911	0.190183	45717.063493
5	LGBM Regressor	37011.331196	0.191228	45826.812352
1	KNN	39004.363654	0.203405	49446.803655
0	Linear	42375.725805	0.219933	52389.060930

Take top 4 algorithm models that have the best values by MAE, namely DecisionTree,XGBoost, RandomForest, and LGBM Regressor.

Model Before & After Tuning

In concept, LGBM Boost and XGBoost are similar, as they both build an ensemble model from multiple weak learners. working mechanisms,

BEFORE

Working mechanisms, LGBM uses a "leaf-wise" approach to build trees,

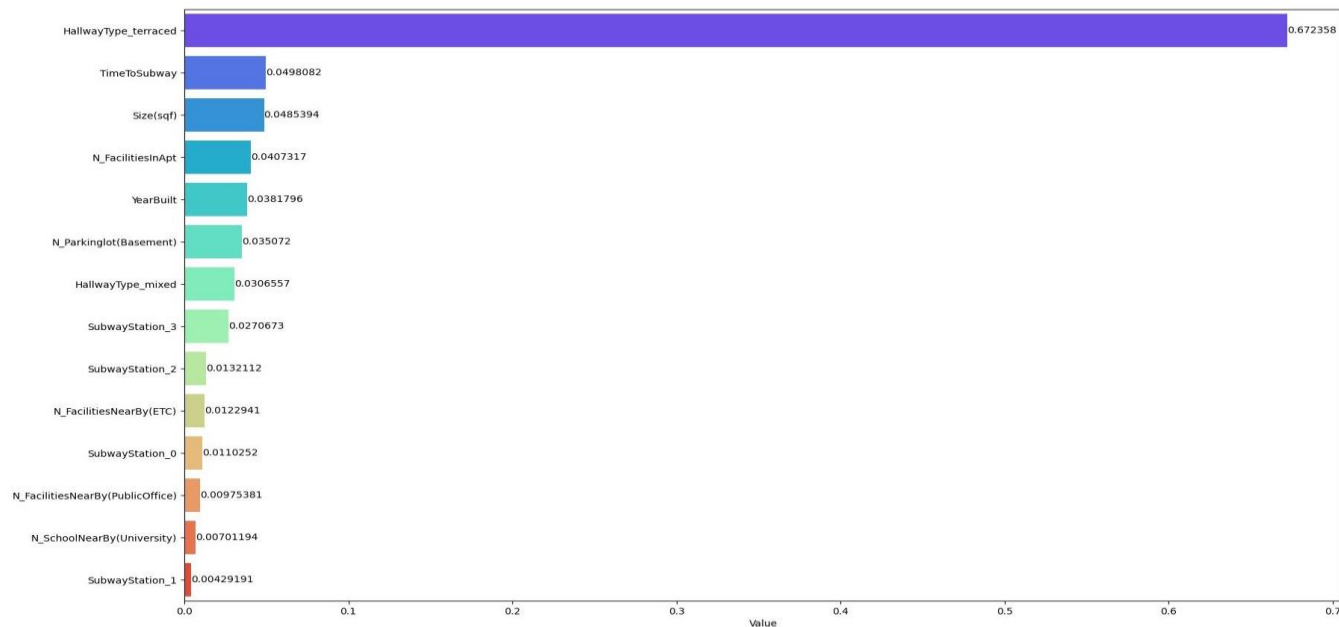|	model	MAE	MAPE	RMSE
3	LGBM Regression Before Tuning	38831.895503	0.195765	47852.900793
1	RandomForest Regression Before Tuning	39052.333906	0.197426	48125.141812
2	XGBoost Regression Before Tuning	39095.706268	0.197346	48299.482363
0	DecisionTree Regression Before Tuning	39188.868617	0.198920	48637.842951

Tuning Metrics :

- XGBoost

AFTER

XGBoost uses a "level-wise" approach, where trees are built level by level by evenly splitting at each level.

	model	MAE	MAPE	RMSE
1	XGBoost Regression After Tuning	38515.607468	0.191499	2.295363e+09
0	LGBM Regressio After Tuning	38990.431886	0.196650	2.308726e+09
3	Random Forest Regression After Tuning	39260.801760	0.199579	2.381850e+09
2	DecisionTree Regression After Tuning	44261.707414	0.232128	2.929687e+09

Metrics Evaluation

Feature Importances :

Based on the Feature Importance with XGBoost plot above, This Feature importance is facing how impact feature to the sale price.

Purwadhika
Digital Technology School

PEDESTAL PROPERTIES

05 Business Recommendation

Business Recomendation

Baseline Price

Obtaining price references for each apartment price.

Direct estimate Predict Price

Model Predict Apartment Daegu

Variable Influence

consider another variable that which is not in the model

Thank You