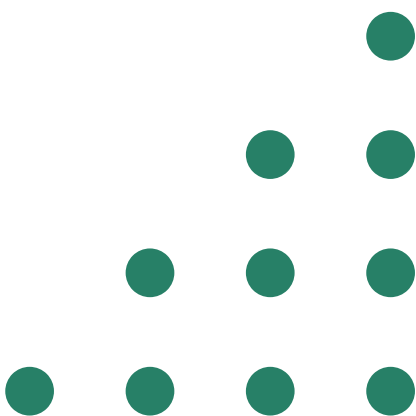


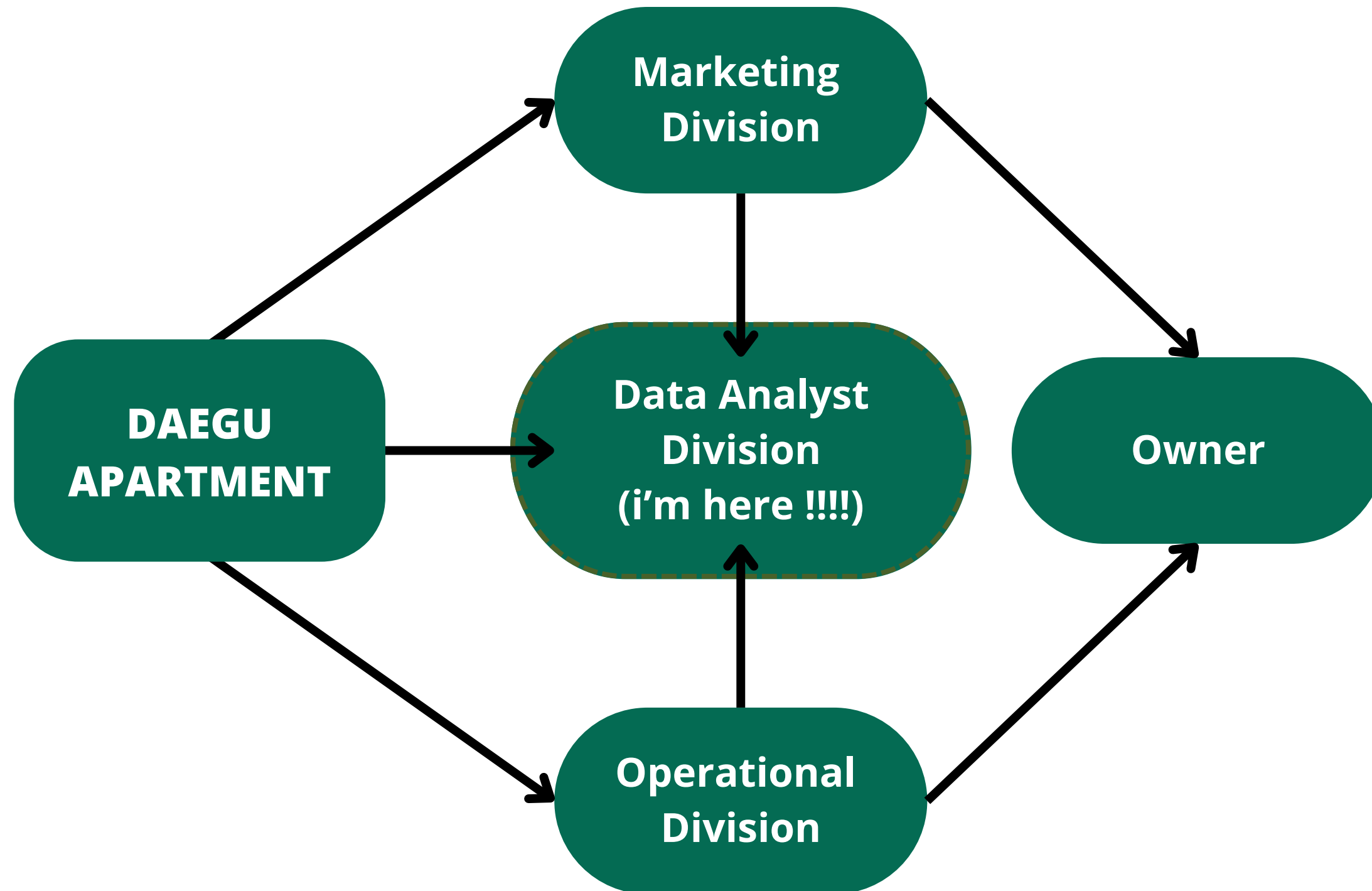
PROFIT MEETS PRECISION:

APARTMENT PRICING POWERED BY MACHINE LEARNING

Project By Dhika Wahyu Pratama



STAKEHOLDER



PRELIMINARY DATA INFORMATION

The data used comes from **Daegu Apartment**, which contains information about Sale Price Apartment in Daegu. In addition to Sale Price information, this data also includes , Hallway type, Number of Facilities nearby, Number of Facilities in Apartment, Size, Parkinglot, etc.

PROBLEM STATEMENT

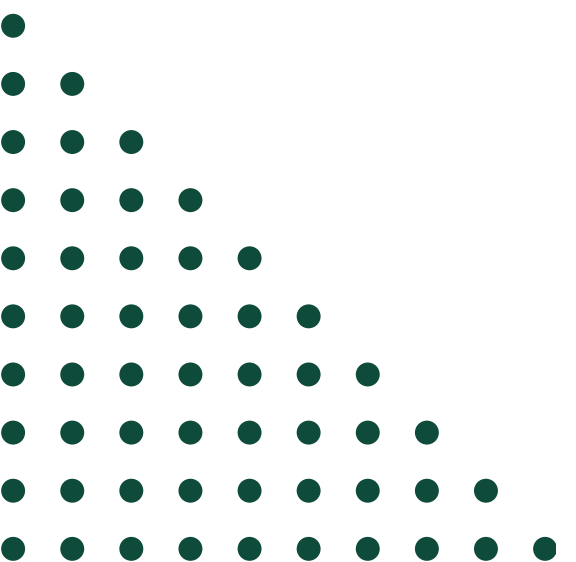
*Dengan bertambahnya jumlah unit yang tersedia di pasar, **menetapkan harga yang tepat** menjadi semakin penting **agar penjualan tetap kompetitif** tanpa mengorbankan potensi keuntungan*

Objective

- Membangun alat prediksi harga jual/sewa apartemen berbasis machine learning.
- Membantu pemilik apartemen menentukan harga yang sesuai berdasarkan fitur properti.
- Menyediakan estimasi harga yang adil, akurat, dan mudah digunakan.
- Mendukung peningkatan pendapatan perusahaan melalui pemanfaatan prediksi harga yang optimal.

Target

- Menentukan harga jual properti yang tepat untuk tiap properti yang baru akan mereka jual
- Mencapai RMSE, MAE, dan MAPE serendah mungkin untuk memastikan akurasi prediksi harga properti.



Data Understanding & Data Cleaning



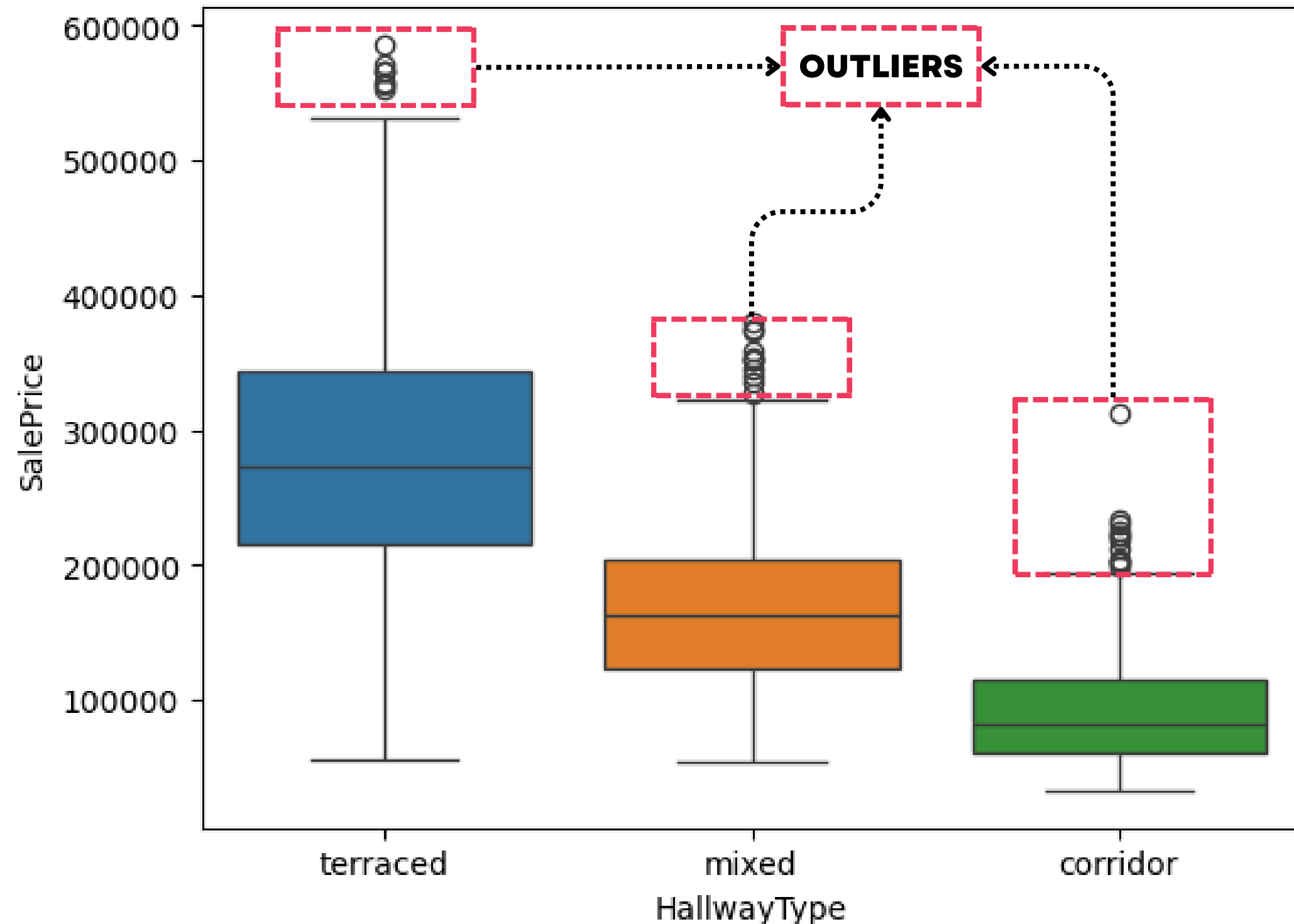
Dataset Variable

Nama Fitur	Deskripsi
Hallway Type	Tipe lorong atau jenis apartemen berdasarkan desain bangunan
TimeToSubway	Waktu tempuh menuju stasiun subway terdekat (dalam menit)
SubwayStation	Nama stasiun subway terdekat
N_FacilitiesNearBy (ETC)	Jumlah fasilitas umum lainnya di sekitar apartemen
N_FacilitiesNearBy (PublicOffice)	Jumlah kantor pemerintahan di sekitar apartemen
N_SchoolNearBy (University)	Jumlah universitas di sekitar apartemen
N_Parkinglot (Basement)	Jumlah tempat parkir bawah tanah (basement) yang tersedia
YearBuilt	Tahun pembangunan apartemen selesai
N_FacilitiesInApt	Jumlah fasilitas internal apartemen
Size (sqft)	Luas apartemen dalam satuan square feet (kaki persegi)
SalePrice	Harga jual apartemen dalam satuan Won (₩)

DATA UNDERSTANDING

- Total Data Row = **4123**
- Missing Values = **0**
- Duplicated Data = **1422**
- Data Non Duplicated = **2701**
- Outliers = **30**

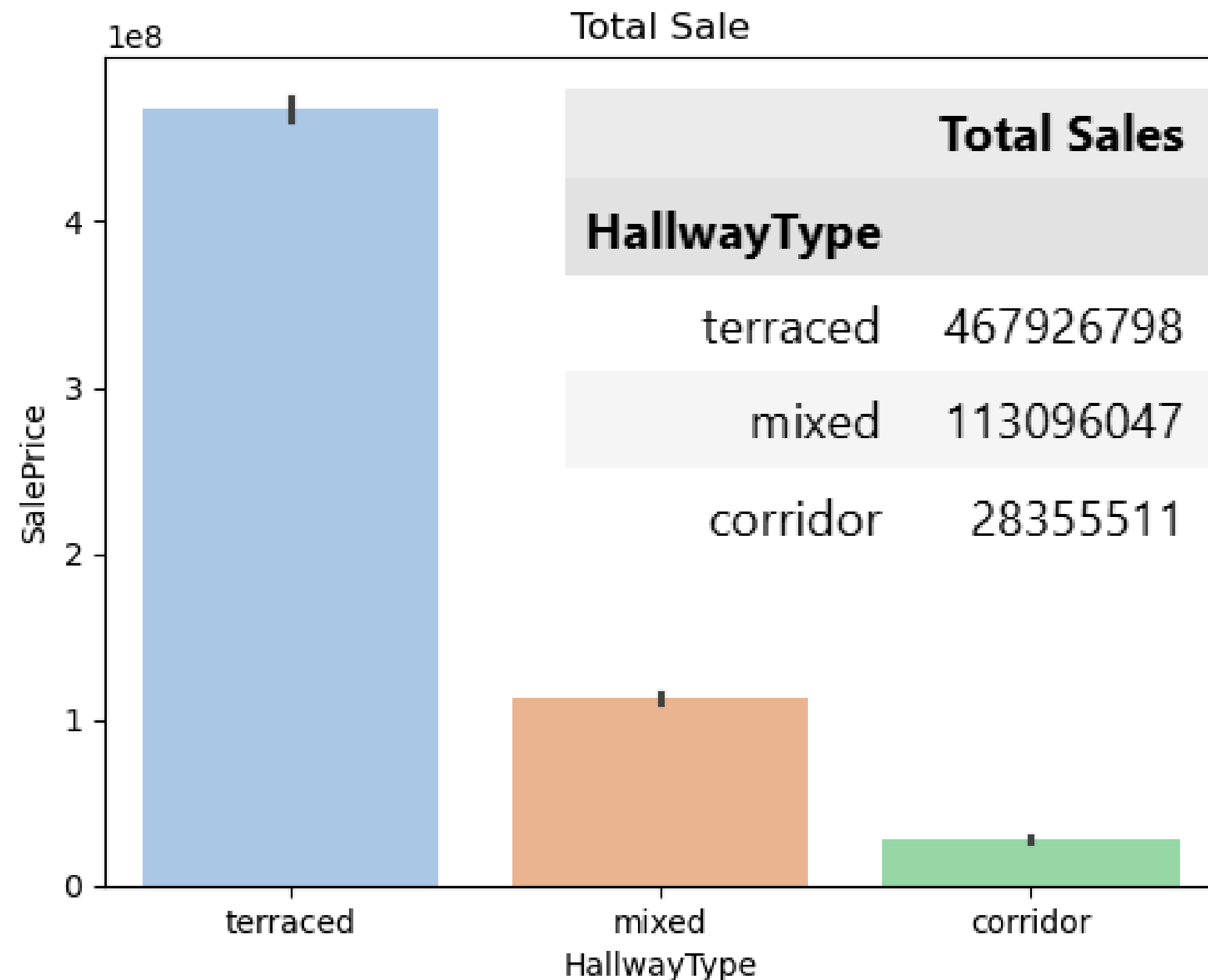
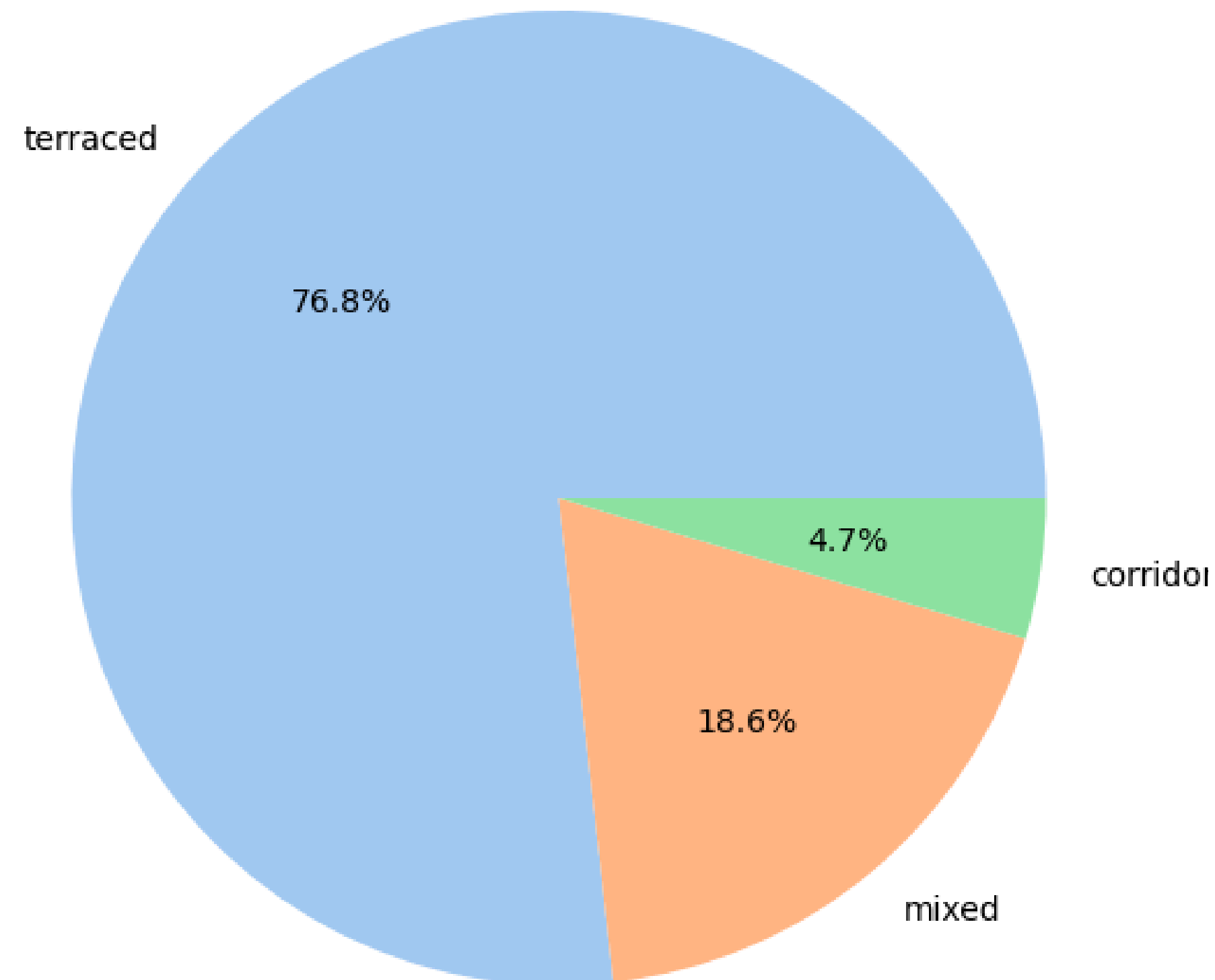
Clean data = 2671



Data Analysis

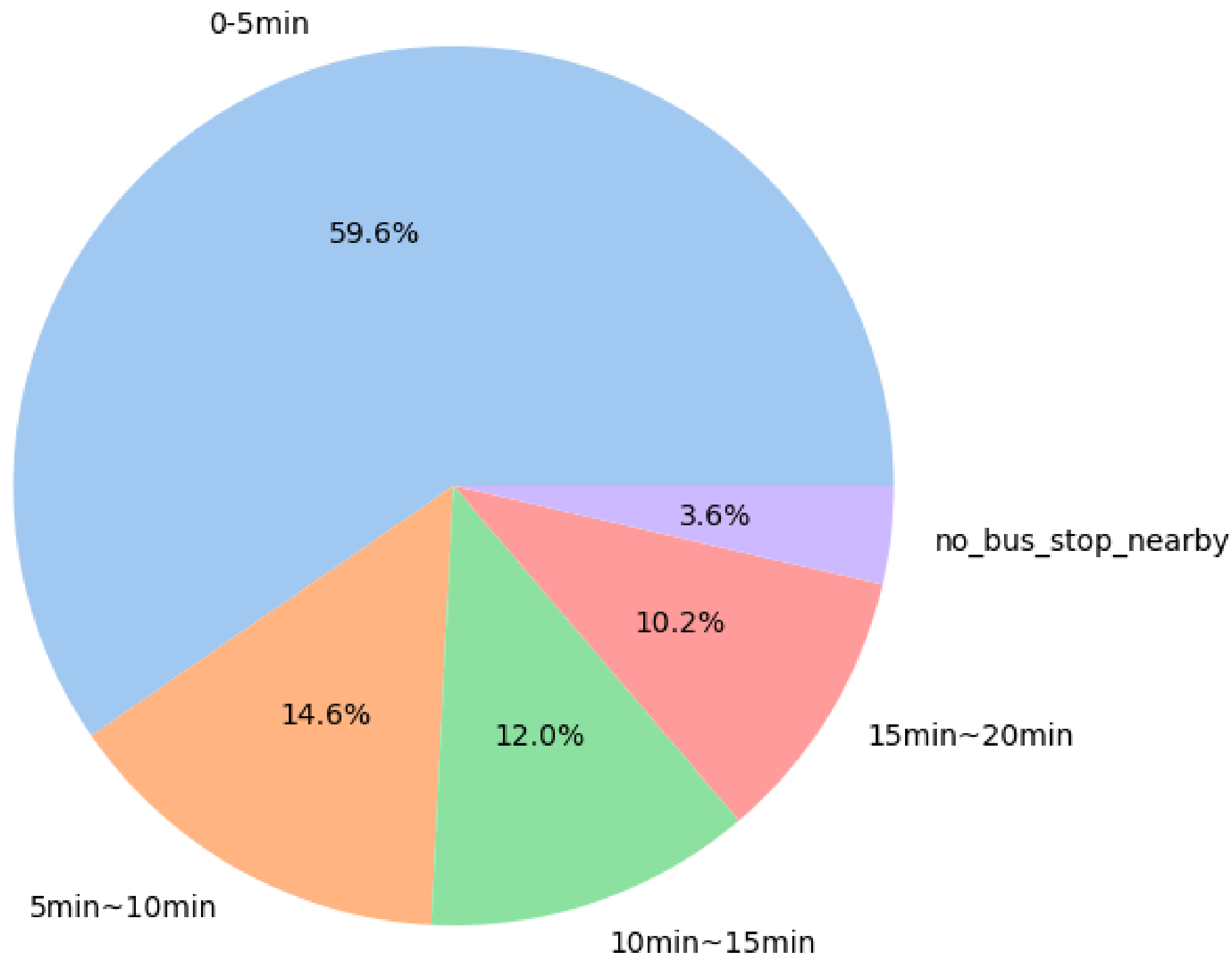
1. TYPES OF PROPERTY RELATED TO PRICE

Proporsi Total Penjualan Berdasarkan Tipe Hallway

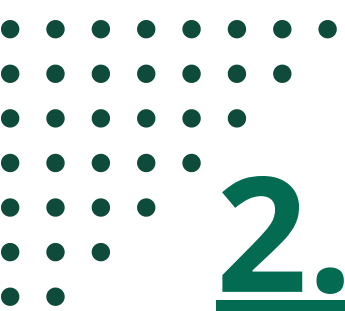


2. ACCESSIBILITY RELATED TO PRICE

Proporsi Total Penjualan Berdasarkan Waktu Ke Subway

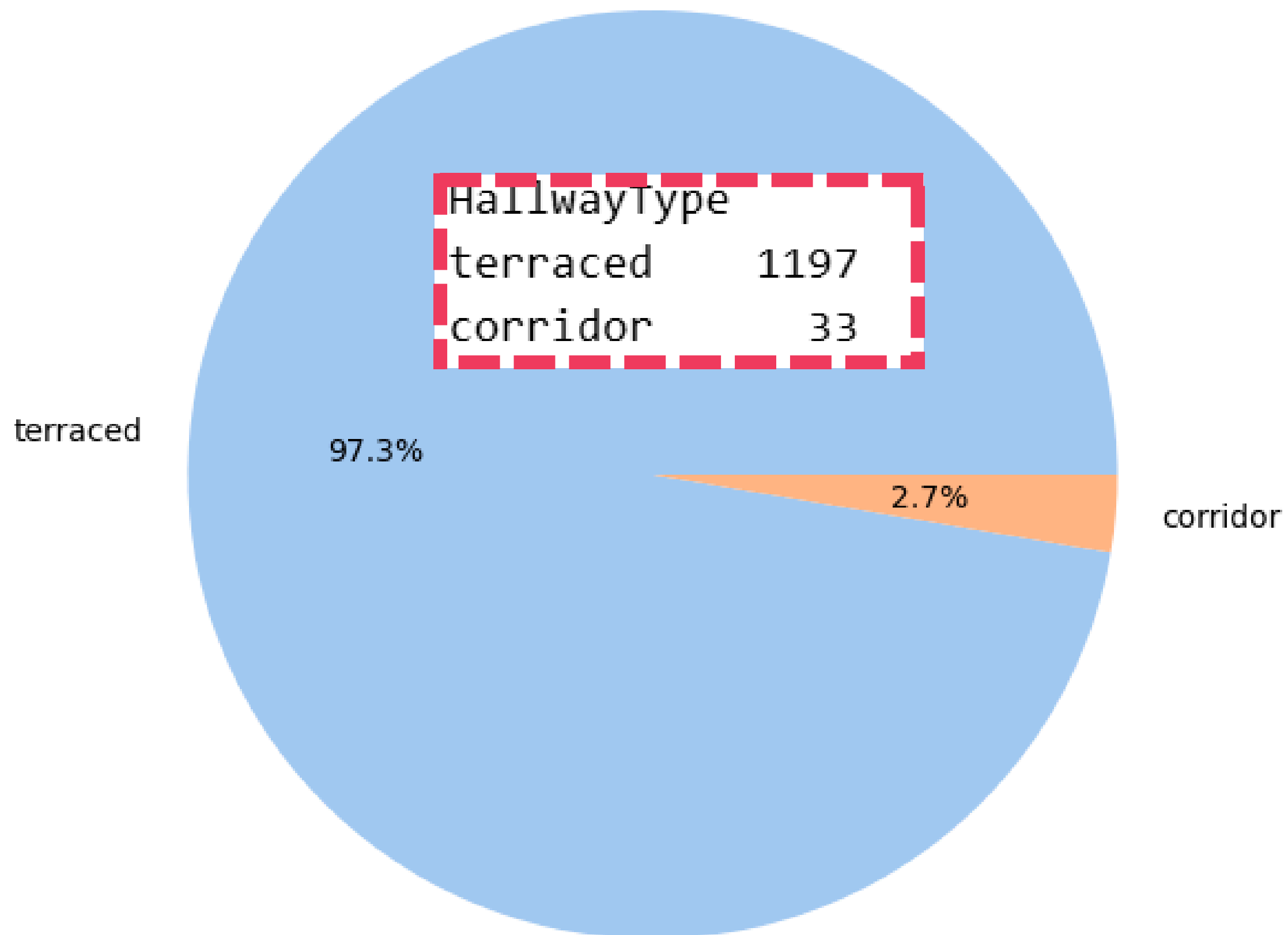


	Total Sales	Total Unit
TimeToSubway		
0-5min	363266627	1230
5min~10min	88830291	499
10min~15min	73349303	434
15min~20min	61950439	408
no_bus_stop_nearby	21981696	100

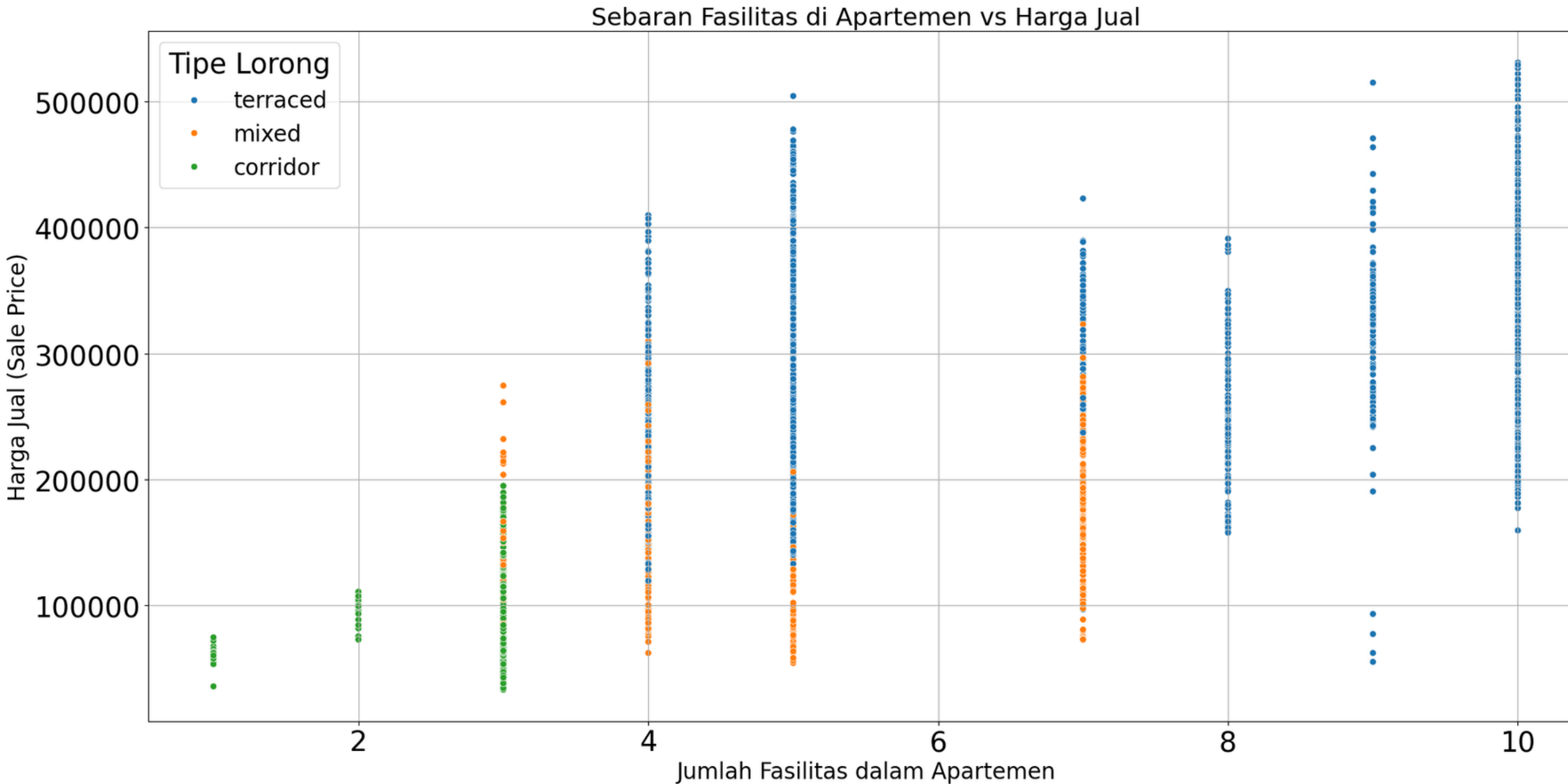


2. ACCESSIBILITY RELATED TO PRICE

Proporsi Jumlah Properti Terjual Berdasarkan Tipe Hallway (0-5 menit ke Subway)

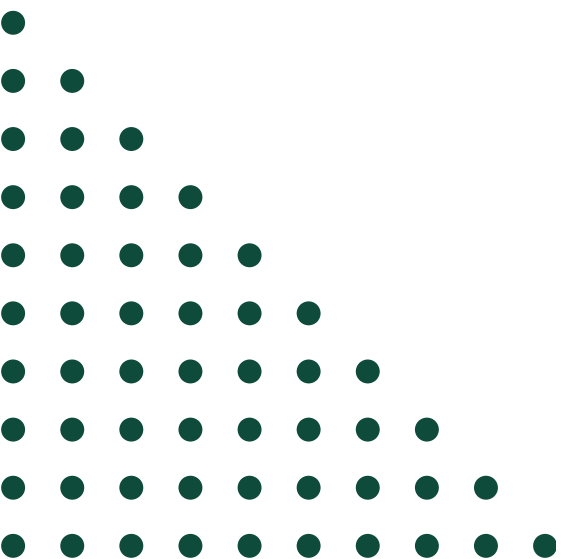


3. FACILITIES RELATED TO PRICE



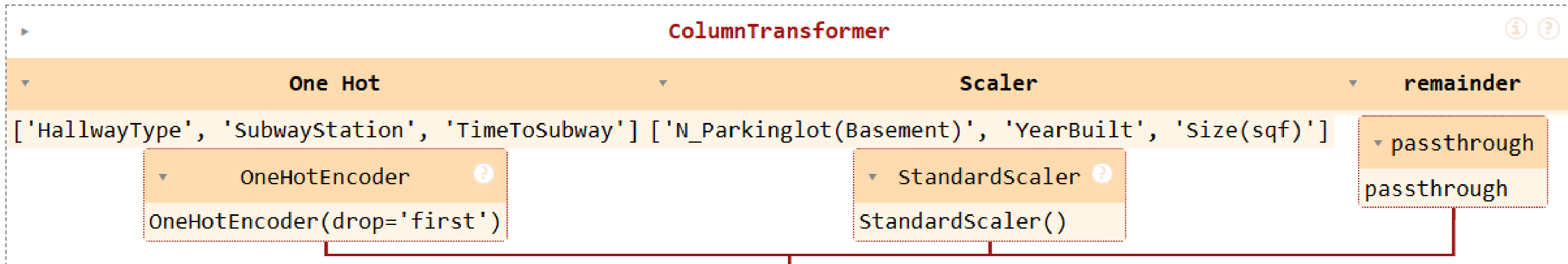
**TIPE TERRACED
MEMANG MEMILIKI
JUMLAH FASILITAS
YANG CENDERUNG
LEBIH BANYAK
DIBANDINGKAN
DENGAN JENIS
LAINNYA,**

Model Machine Learning



PREPROCESSING & MODELING

- **SPLITTING DATA**
 - TRAIN : 80%
 - TEST : 20%
- **TRASNFORM DATA**
 - PIPELINE



MODELING

5 FOLD CROSS VAL

	Model	Mean_RMSE	Std_RMSE	Mean_MAE	Std_MAE	Mean_MAPE	Std_MAPE
0	Linear Regression	-51835.117441	2327.952867	-40989.834230	1749.277428	-0.206764	0.009180
1	KNN Regressor	-47994.833522	1310.550343	-37481.299499	1104.581336	-0.195279	0.009442
2	DecisionTree Regressor	-44437.000857	1417.396295	-35342.231023	935.898694	-0.177789	0.004704
3	RandomForest Regressor	-44430.328193	1402.818764	-35339.101804	925.709891	-0.177870	0.004689
4	XGBoost Regressor	-44426.937979	1391.609476	-35320.089998	920.807719	-0.177647	0.004731
5	Adaboost Regressor	-53242.152424	2052.489493	-42172.569330	1632.716984	-0.202731	0.004559
6	GradientBoosting Regressor	-44281.130447	1258.956581	-35263.863341	1009.801476	-0.177190	0.004989
7	SVR	-45526.877343	1203.490612	-35942.497607	918.043380	-0.181669	0.004938

MODEL TERPILIH

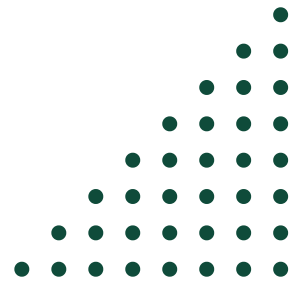
- **BEST RMSE**
- **BEST MAE**
- **BEST MAPE**



MODELING

TUNING HYPERPARAMETER

Hyperparameter	Nilai-nilai yang Dicoba
max_depth	1, 3, 5, 7, 9, 11, 13, 15, 17, 19
learning_rate	0.01, 0.02, ..., 0.10 (increment 0.01)
n_estimators	1000, 1100, ..., 2000 (increment 100)
subsample	0.2, 0.3, ..., 0.9 (increment 0.1)
max_leaf_nodes	1, 2, ..., 10
min_impurity_decrease	0.0, 0.1, ..., 1.0 (increment 0.1)
max_features	0.1, 0.2, ..., 0.9, 'sqrt', 'log2', 'auto'





MODELING

TUNING HYPERPARAMETER

TRAIN

BEST SCORE RMSE TRAINING
==> 44172.29

TEST

BEFORE TH

	RMSE	MAE	MAPE
Grad	48566.012311	39333.243739	0.195183

AFTER TH

	RMSE	MAE	MAPE
Grad	48839.993243	39424.042247	0.195126

RandomizedSearchCV

best_estimator_: Pipeline

preprocessing: ColumnTransformer

One Hot

Scaler

remainder

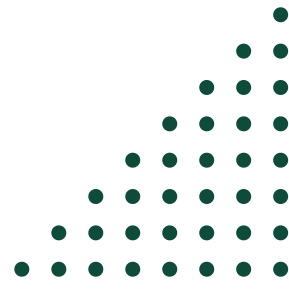
OneHotEncoder

StandardScaler

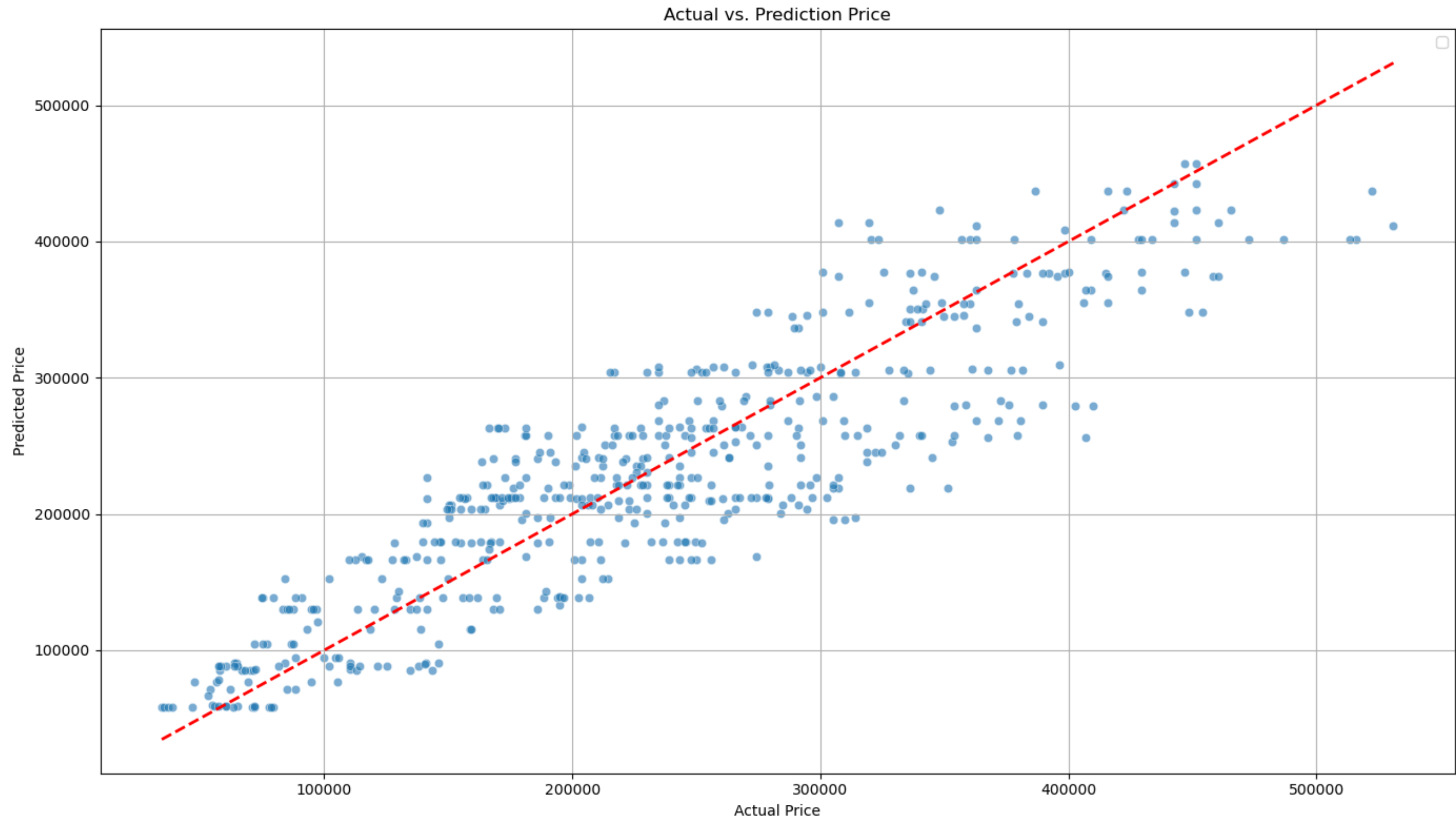
passthrough

GradientBoostingRegressor

GradientBoostingRegressor(learning_rate=0.01, max_depth=7,
max_features=0.30000000000000004, max_leaf_nodes=6,
min_impurity_decrease=0.30000000000000004,
n_estimators=1800, random_state=42, subsample=0.8)



MODEL INTERPRETATION

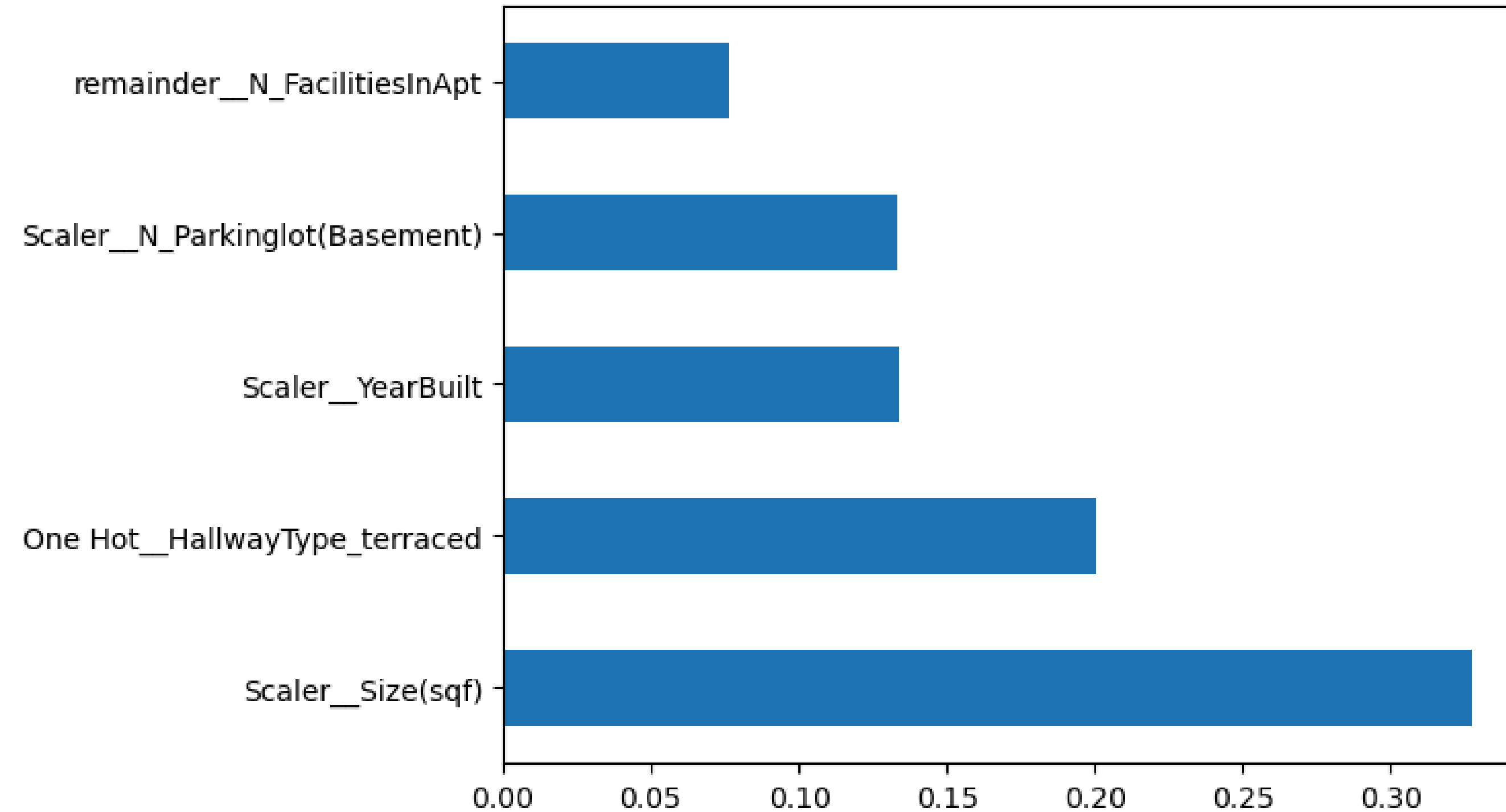




MODEL INTERPRETATION

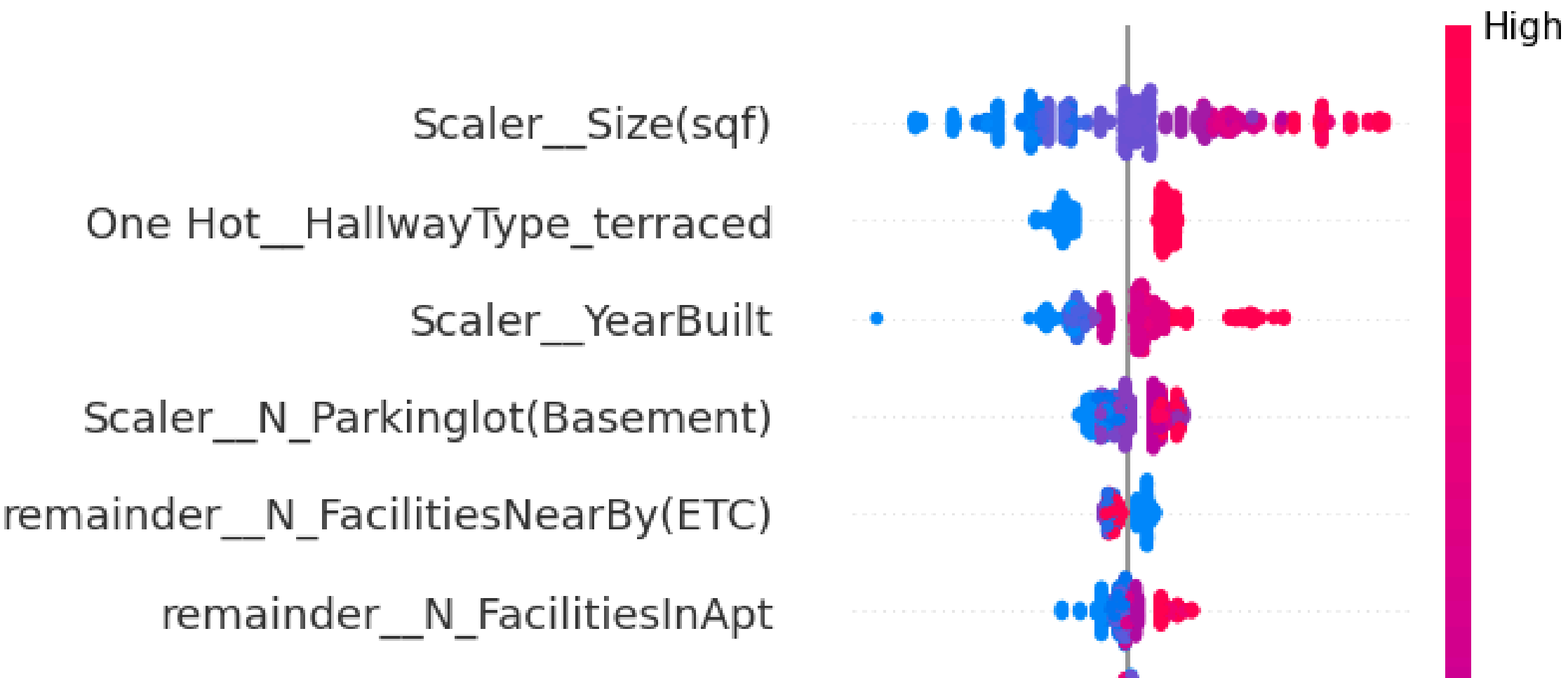
FEATURE IMPORTANCES

Feature Importances



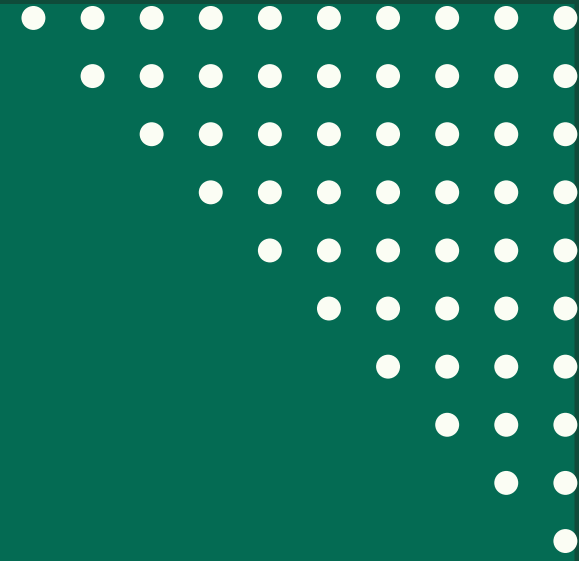
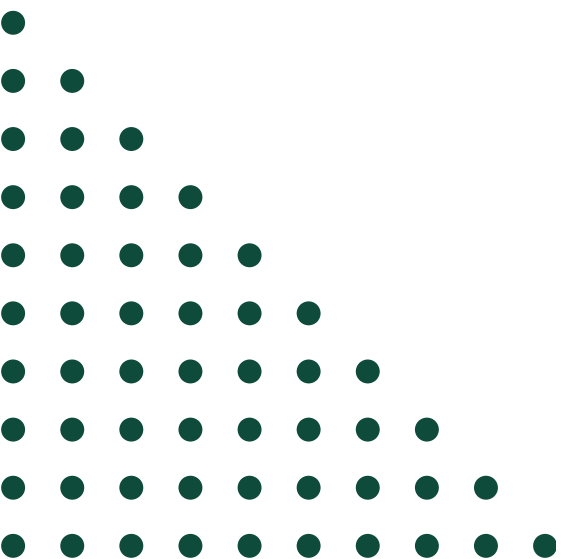
MODEL INTERPRETATION

SHAP



- **SCALER__SIZE(SQF) → LARGER SIZE = HIGHER PRICE.**
- **ONE HOT__HALLWAYTYPE_TERRACED → "TERRACED" HALLWAY TYPE INCREASES VALUE.**
- **SCALER__YEARBUILT → NEWER CONSTRUCTION YEAR = HIGHER PRICE.**
- **SCALER__N_PARKINGLOT(BASEMENT) → MORE BASEMENT PARKING = HIGHER VALUE.**
- **REMAINDER__N_FACILITIESNEARBY(ETC) → NEARBY FACILITIES BOOST PRICE.**
- **REMAINDER__N_FACILITIESINAPT → MORE FACILITIES, HIGHER VALUE**

Model Machine Learning



CONCLUSION

- **Fitur yang Paling Berpengaruh adalah:**
 - a. Luas apartemen (Size(sqf))
 - b. Tipe lorong (HallwayType)
 - c. Tahun dibangun (YearBuilt)
 - d. Jumlah fasilitas dalam apartemen (N_facilitiesinApt)
 - e. Ketersediaan parkir (ParkingLot)
- **Performa Model**
 - a. Evaluasi model menggunakan metrik RMSE, MAE, dan MAPE.
 - b. MAPE setelah hyperparameter tuning sebesar ~19%, yang berarti prediksi harga rata-rata meleset sekitar 19% dari nilai sebenarnya.
- **Keterbatasan Model**
 - a. Model masih menunjukkan bias yang cukup tinggi, terlihat dari visualisasi antara harga aktual dan prediksi.
 - b. Hal ini kemungkinan besar disebabkan oleh keterbatasan fitur yang tidak merepresentasikan faktor eksternal seperti keamanan lingkungan, akses makanan, dan fasilitas umum sekitar, Pemandangan, Tingkat Kebisingan dan lainnya.

RECCOMENDATION

- **Lakukan A/B Testing**

- Uji efektivitas model terhadap kemungkinan lakunya listing, dibandingkan dengan pengguna yang menentukan harga jual (Sale Price) secara manual.

- **Perkaya Fitur yang Lebih Relevan**

Tambahkan fitur yang berkaitan dengan kenyamanan tinggal, seperti:

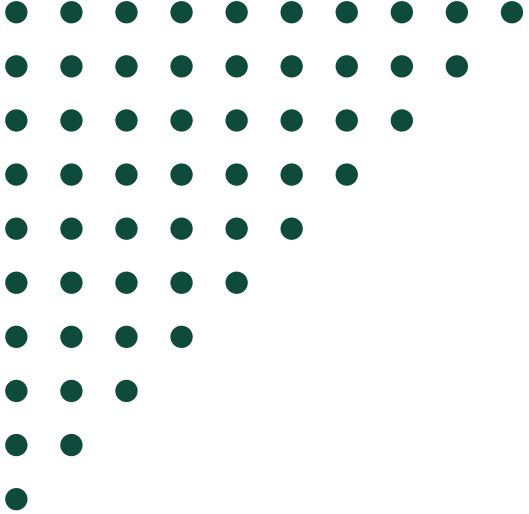
- Jarak ke pusat kota
- Tingkat kebisingan
- Akses makanan dan fasilitas umum Serta perluas cakupan data apartemen di Daegu untuk meningkatkan kapasitas generalisasi model.

- **Perbaiki Kualitas dan Kuantitas Data**

- Hapus data duplikat untuk menghindari bias
- Jika jumlah data memadai, pertimbangkan penggunaan model yang lebih kompleks seperti Neural Networks (NN)

- **Kembangkan Prediksi Jangka Panjang**

- Model saat ini dapat dikembangkan lebih lanjut untuk memprediksi tren harga apartemen di masa depan, membantu pengguna memperkirakan potensi kenaikan nilai properti.



THANK YOU

