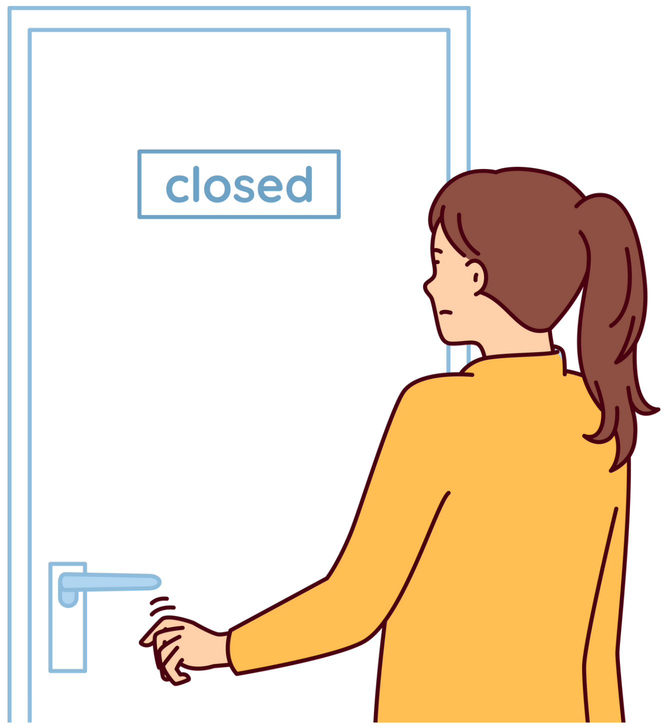
How much does property cost?

The trouble with pricing property

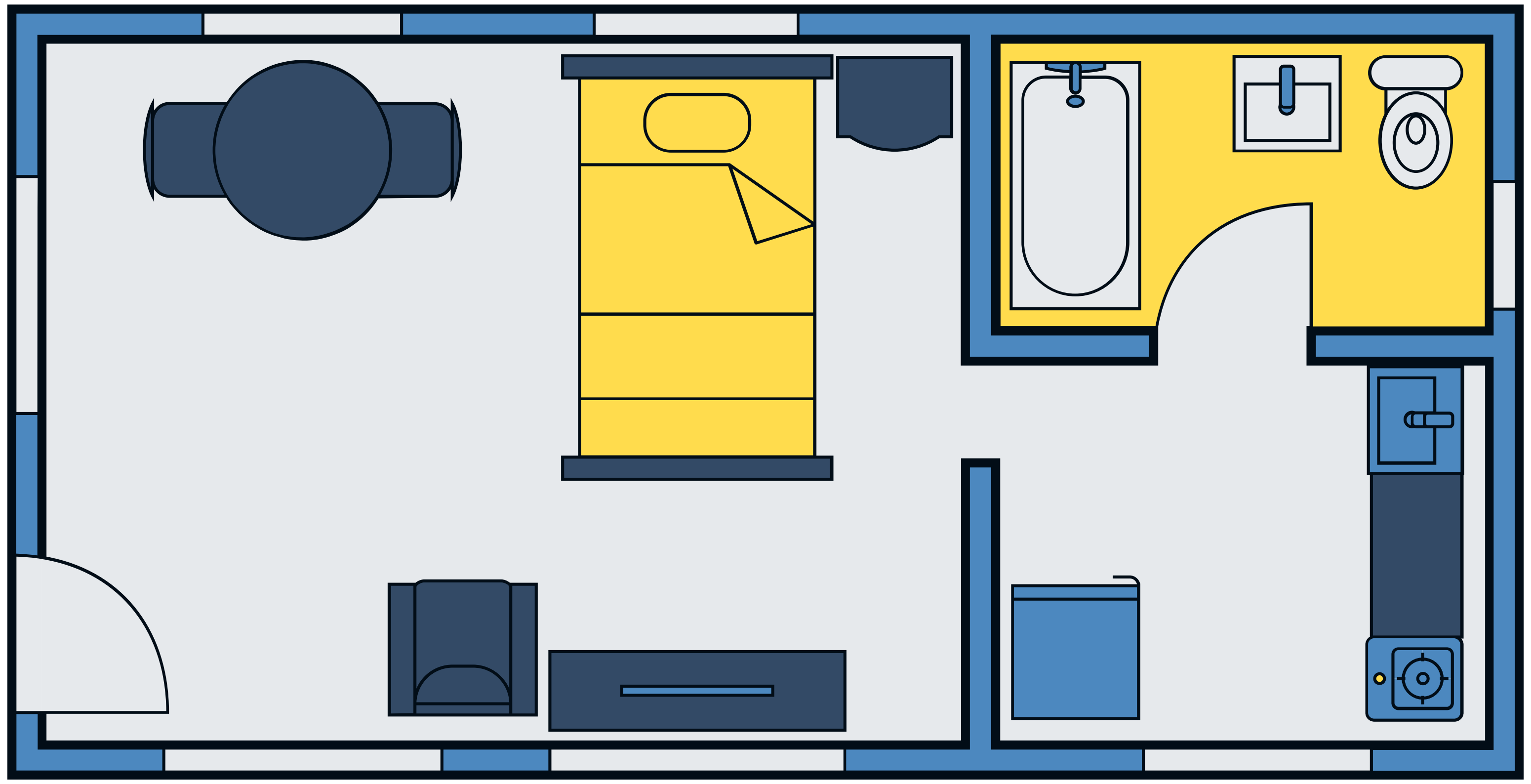
Overpricing



Inaccessible information

Subjective appraisals

The trouble with pricing property



Naive way of pricing property

= Floor Area x (Price / m2)

Property valuation needs to be data-driven.

Business Value

Seller

Buyer

Informed Pricing

Market Trends

Problem Statement:

Can we produce an accurate housing price prediction model and determine the factors that drive property prices?

Machine Learning 2

The House of Us: Metro Manila Housing Price Prediction using Machine Learning

Bulatao, de la Paz, Satiada, Soliman

Previous work on housing price prediction models (PH)

Tree-based models are better

(Jagan Chowhaan et al., 2023)

Linear regression models for Metro Manila

(Abellana et al., 2021)

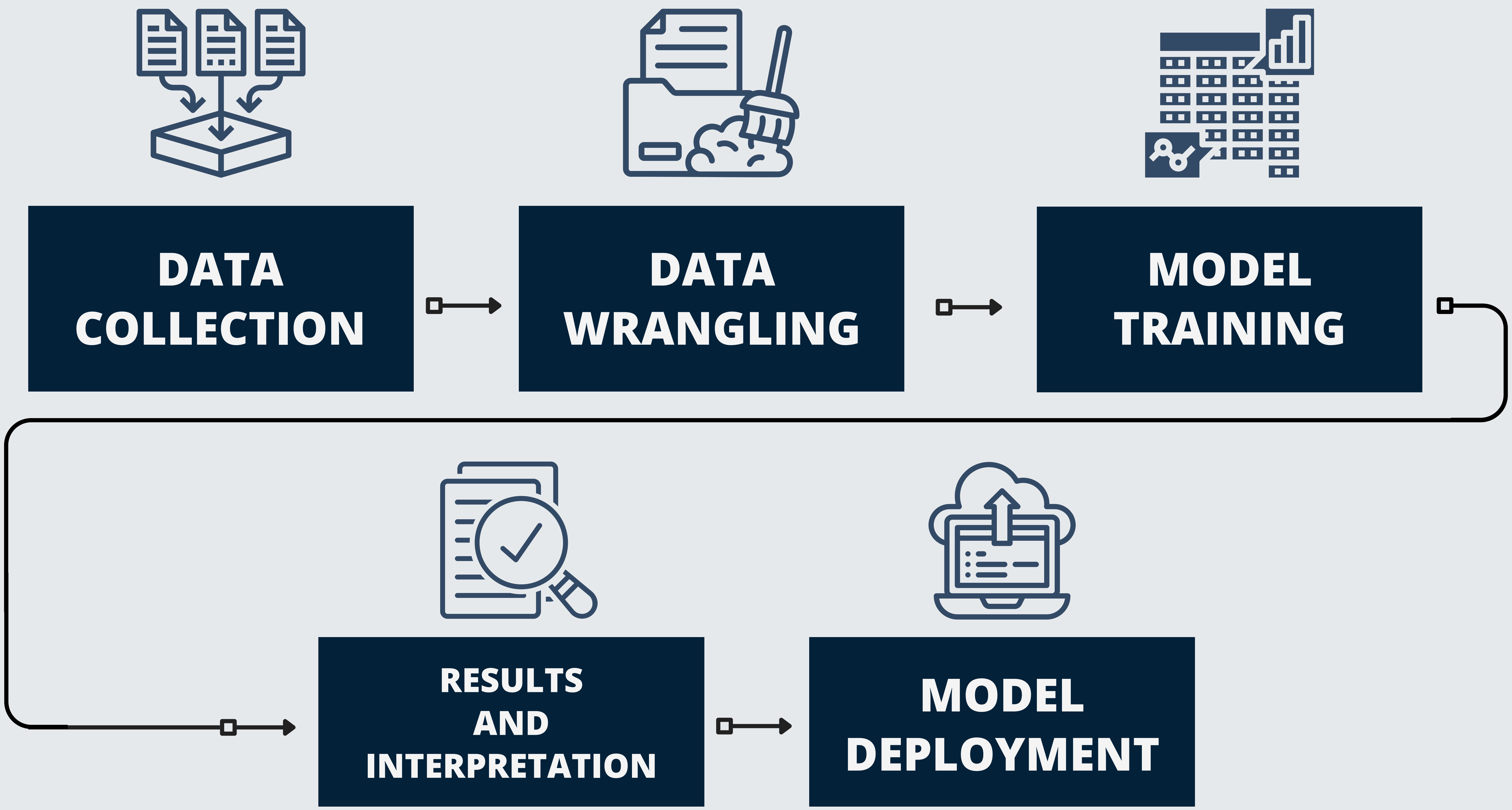
No instance-based valuation models for Metro Manila

Objectives

01 Create an **instance-based machine learning model** that can approximate property prices in Metro Manila using property features and amenity data.

02 Investigate **factors that drive property prices** aside from floor area.

Methodology



Data Description

Metro Manila Lamudi
Listings Dataset (2022)

Price

Bedrooms

Bath

Geographic
Coordinates

Floor Area

Data Description

Amenity Types

Healthcare

Food

Transportation

Finance

Public Service

Education

OpenStreetMap Dataset

Data Description

Metro Manila Lamudi
Listings Dataset (2022)

OpenStreetMap
Dataset

List of Models Trained

GradientBoosting	Linear Regression
Random Forest	Linear Regression (L1)
XGBoost	Linear Regression (L2)
AdaBoost	KNearestNeighbor

R^2 Score

	Base (Random Forest)	with Amenities (Gradient Boosting)
Mean Train-Validation Score	99.16	99.86
Mean Test Score	98.91	99.08

NOTE: One sample T-Test statistic provides 95% confidence that the change is significant!

Mean Absolute Error (MAE)

	Base (Random Forest)	with Amenities (Gradient Boosting)
High (27,431,000)	PHP 710,000	PHP 575,000
Medium (15,492,000)	PHP 279,000	PHP 279,000
Low (7,667,973)	PHP 510,000	PHP 262,000

NOTE: One sample T-Test statistic provides 95% confidence that the change is significant for high and low-priced properties.

Mean Absolute Error (MAE)

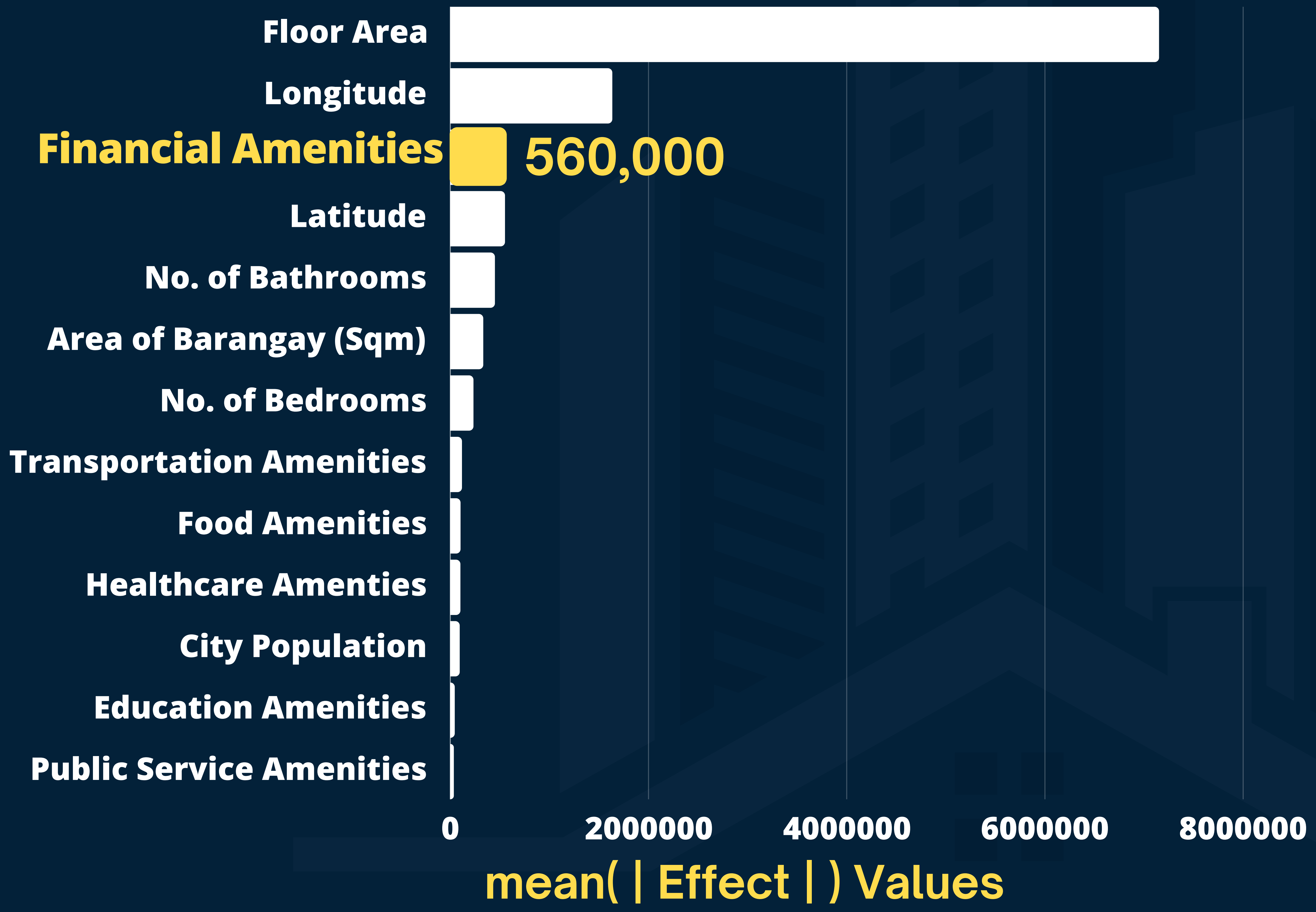
	Base (Random Forest)	with Amenities (Gradient Boosting)
High (27,431,000)	(2.31%)	(2.02%)
Medium (15,492,000)	(1.94%)	(1.94%)
Low (7,667,973)	(7.89%)	(3.39%)

NOTE: One sample T-Test statistic provides 95% confidence that the change is significant for high- and low-priced property.

What now?

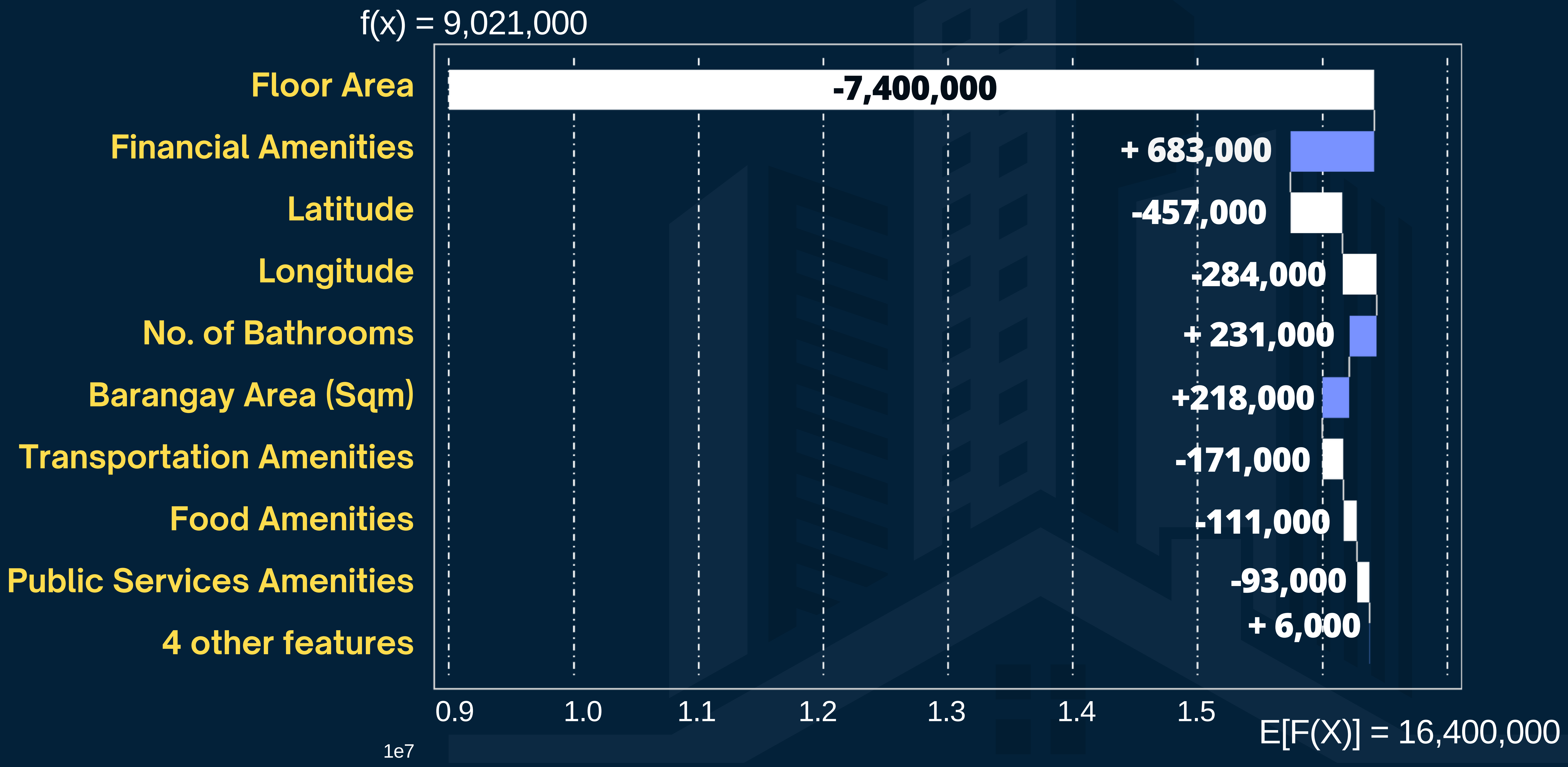
"Where is a good place to invest in?"

Shap Global Explainability



“How much will financial amenities in the area affect my property‘s value?”

The Sapphire Bloc Ortigas I 1 Bedroom for sale



"I want to sell my property at 10 million pesos but it's currently being valued at 9 million.

The Sapphire Bloc Ortigas | 1 Bedroom

currently being listed at PHP 9 million

Existing Amenity Counts

Food	Education	Healthcare	Public	Finance	Transpo
71	4	5	-	63	8

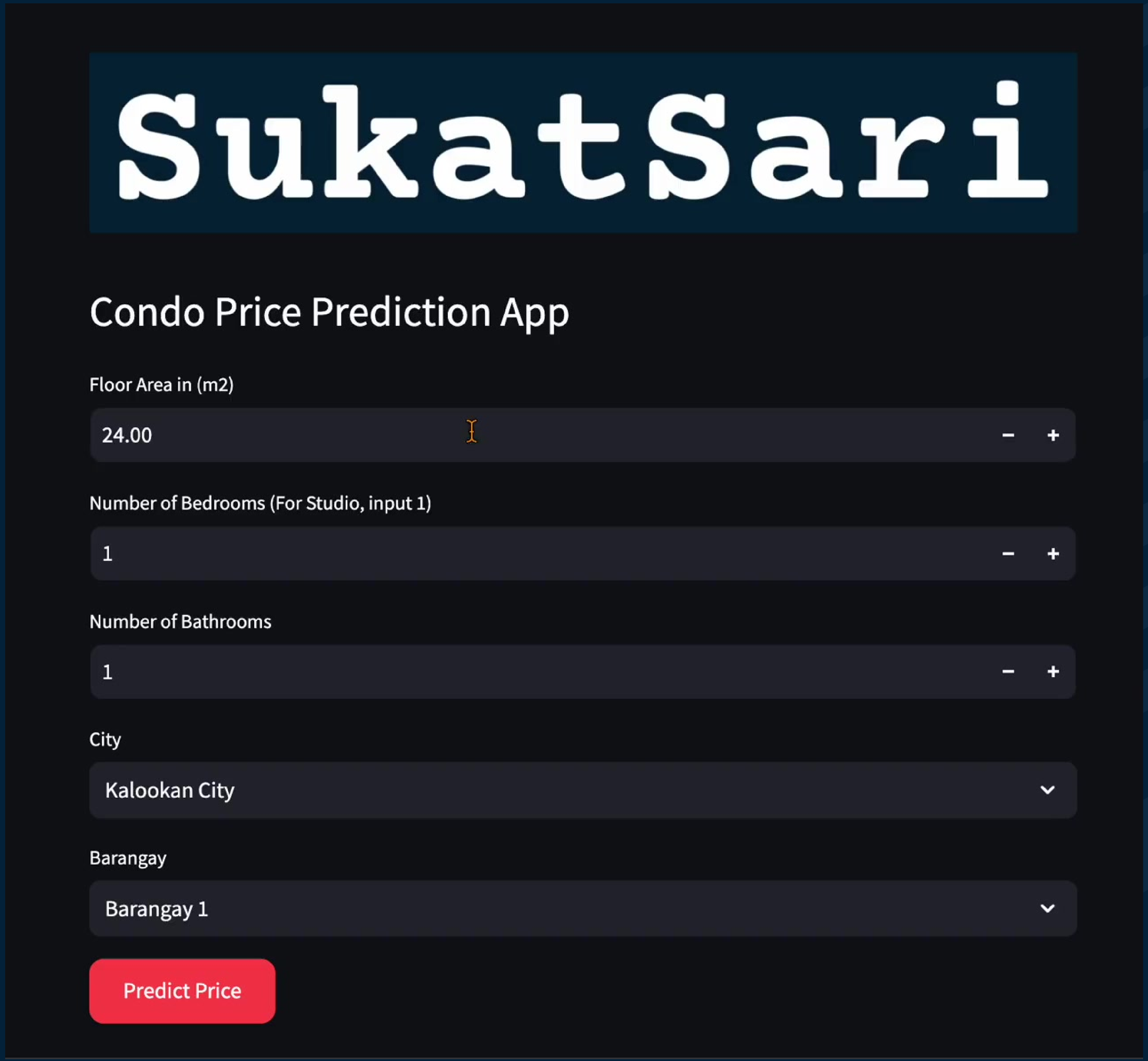
Counterfactuals

Food	Education	Healthcare	Public	Finance	Transpo
71	4	5	-	63	8
-	9	-	-	86	-

Conclusion

Accurate

Interpretability



Try our Prototype App!

SukatSari

Machine Learning 2

Thank you!

Try our Prototype App!

SukatSari

References

Abellana, J. A., & Devaraj, M. (2021). Hedonic Modeling for Predicting House Prices during COVID-19 Pandemic in the Philippines. In *Proceedings of the 2021 3rd International Conference on Management Science and Industrial Engineering* (pp. 21–26). Retrieved from https://doi.org/10.1145/3460824.3460828

Fontinelle, A. (2023, April 30). *Avoid these mistakes when selling your home.* Investopedia. Retrieved February 24, 2024, from https://www.investopedia.com/articles/mortgages-real-estate/08/home-seller-mistakes-selling-house.asp

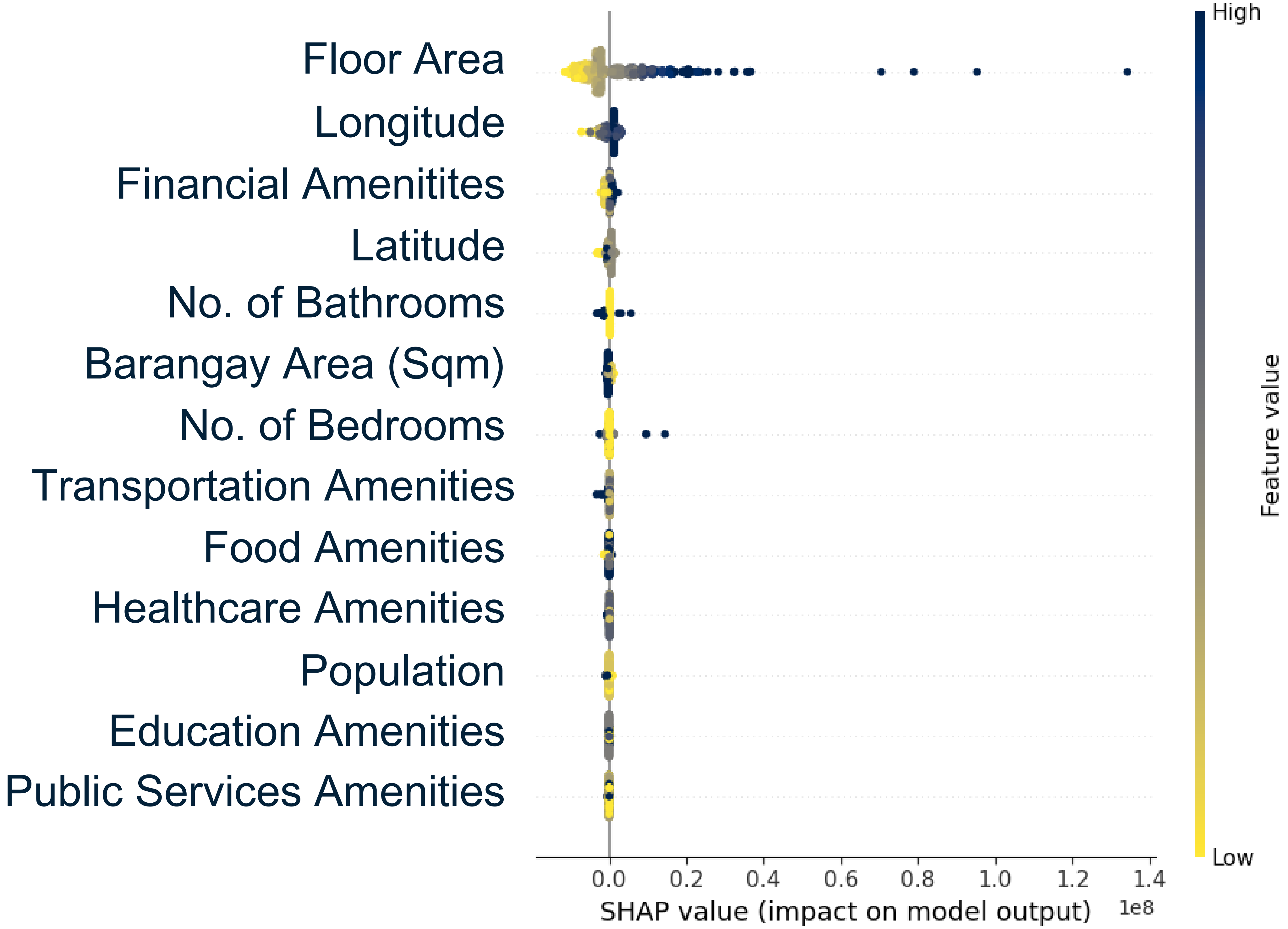
Jagan Chowhaan, M., Nitish, D., Akash, G., Nelli, S., & Shaik, S. (2023). Machine Learning Approach for House Price Prediction. *Asian Journal of Research in Computer Science, 16,* 54-61. Retrieved from https://doi.org/10.9734/ajrcos/2023/v16i2339

Philippines Statistics Authority. (2023, May 10). *Housing Characteristics in the Philippines (2020 Census of Population and Housing).* Retrieved February 26, 2024 from https://psa.gov.ph/content/housing-characteristics-philippines-2020-census-population-and-housing.

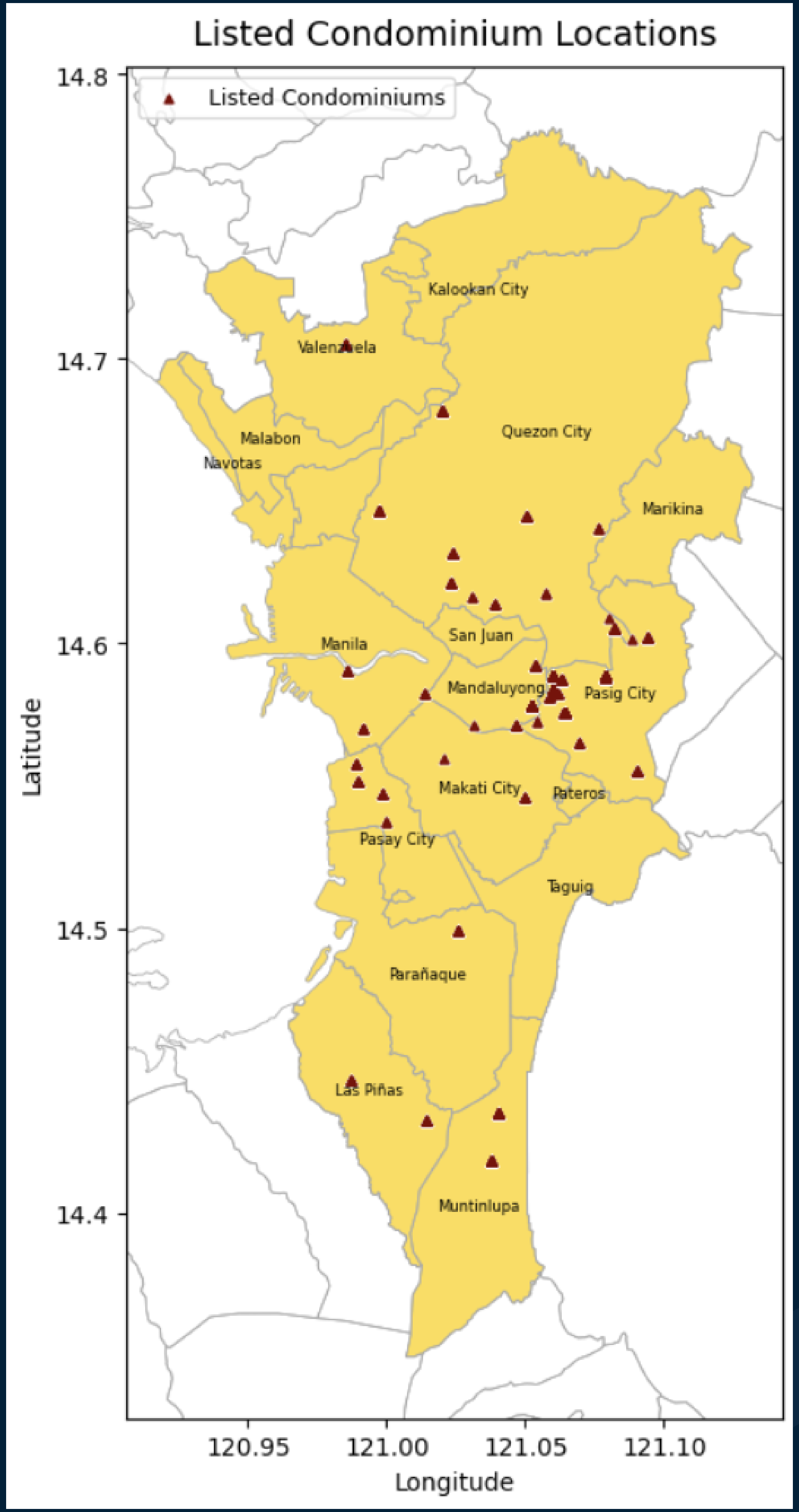
Philippines Statistics Authority & University of the Philippines Population Institute. (2019). *2018 National Migration Survey.* Quezon City, Philippines: PSA and UPPI. Retrieved February 26, 2024 from https://psa.gov.ph/sites/default/files/2018%20NMS%20Final%20Report.pdf

Statista. (2024, January 11). *Philippines: Age structure from 2012 to 2022.* Retrieved February 26, 2024 from https://www.statista.com/statistics/578779/age-structure-in-philippines/.

Umbrasas, K. (2022, August 8). *9 mistakes to avoid when selling your home.* Zillow. Retrieved February 24, 2024, from https://www.zillow.com/learn/mistakes-to-avoid-when-selling-your-home/



Floor Area
Longitude
Financial Amenitites
Latitude
No. of Bathrooms
Barangay Area (Sqm)
No. of Bedrooms
Transportation Amenities
Food Amenities
Healthcare Amenities
Population
Education Amenities
Public Services Amenities
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1e8
SHAP value (impact on model output)
High
Low
Feature value



Listed Condominiums and their Locations

Scope and Limitations

Listings are limited to the 2022 Lamudi dataset

- Only condominiums were considered.
- Datapoints pertaining to houses were limited.

OpenStreetMap Dataset is from 2018

Recommendations

Widen our scope and acquire updated datasets

Reduce scope of surrounding amenities from barangay-level

- Use various radius values (500m, 1km, 2km) around the properties.

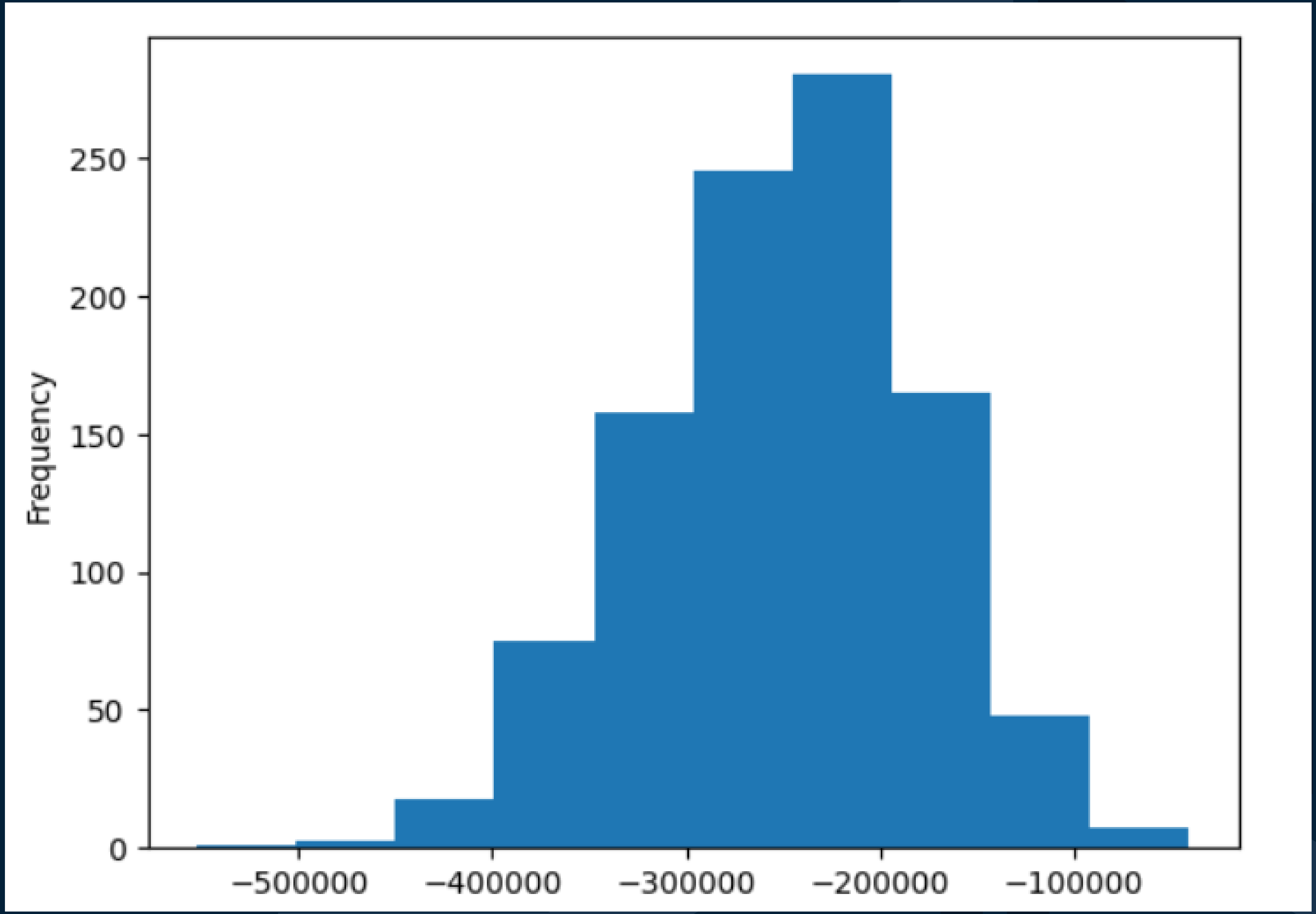
Results - Base (MAE)

Model	Mean Train Score	Mean Validation Score
Random Forest	**498,707**	**864,257**
XGBoost	278,246	919,671
Gradient Boosting	627,730	991,921
KNearestNeighbor	966,041	1,320,325
Linear Regression (L1)	2,523,344	2,561,009
Linear Regression (L1)	2,523,345	2,561,009
Linear Regression	2,523,097	2,561,909
AdaBoost	2,273,031	2,659,213

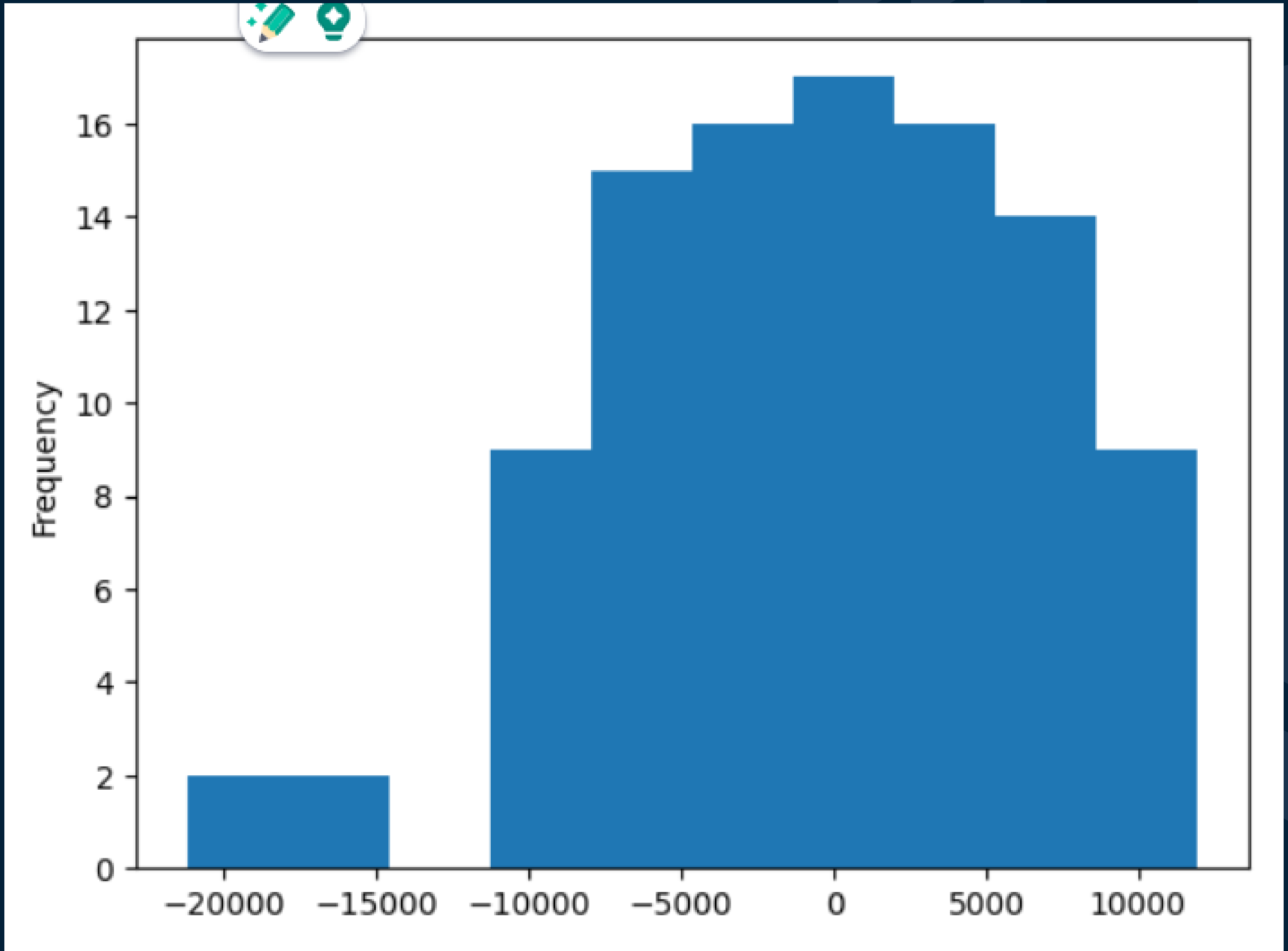
Results - with Amenities (MAE)

Model	Mean Train Score	Mean Validation Score
Gradient Boosting	**272,093**	**743,401**
XGBoost	272,093	869,627
Random Forest	509,386	871,004
KNearestNeighbor	321,299	955,220
Linear Regression (L1)	2,351,576	2,445,311
Linear Regression	2,351,576	2,445,311
Linear Regression (L2)	2,351,576	2,445,320
AdaBoostRegressor	2,189,373	2,587,680

Distribution of MAE difference at Low Price



Distribution of MAE difference at Medium Price



Distribution of MAE difference at High Price