15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Utilization of Government-Based and Non-Conventional Indicators for Property Value Prediction in the Philippines

**Gabriel Isaac L. Ramolete, Bryan Bramaskara,
Dustin A. Reyes, Adrienne Heinrich**

Presented by:
Gabriel Isaac L. Ramolete
Aboitiz Data Innovation

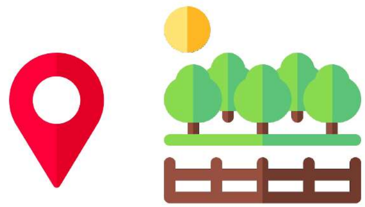
Data Revolution: Big Data, Non-Official Data, Citizen-Generated Data, Big Earth Data, Data Analytics
Crowne Plaza Manila Galleria
05 October 2022, 8:30-10:00AM

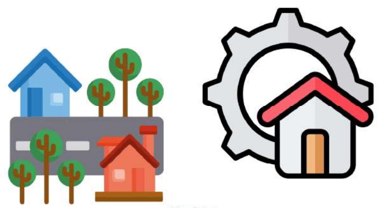
Outline

- Introduction
  - Background
  - Objective and Research Questions
  - Scope and Limitations
- Materials and Methods
  - Data Sources
  - Methodology
- Results and Discussion
  - EDA
  - Non-Segmented Approach
  - Segmented Approach
- Conclusions and Recommendations

Traditional indicators for real estate valuation

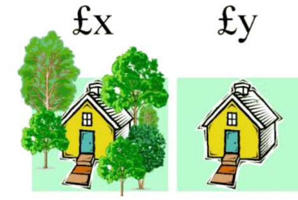
Location
Home Size
Usable Space
Neighborhood comparisons

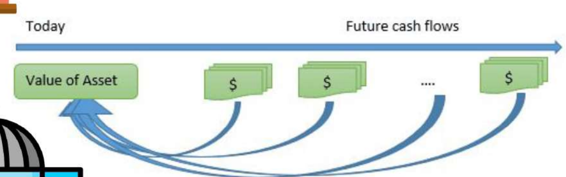




Traditional real estate valuation methods

Hedonic Pricing Model
Sales Comparison Approach
Cost Approach
Income Capitalization Approach
Discounted Cash Flow (DCF) Method





Alternative indicators for real estate valuation

Accessibility
Public service facilities
Commercial places of interest
Safety
Livability

Not commonly utilized or quantified by appraisers and real estate developers, especially in the Philippines

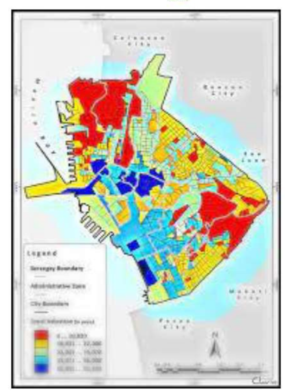
Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

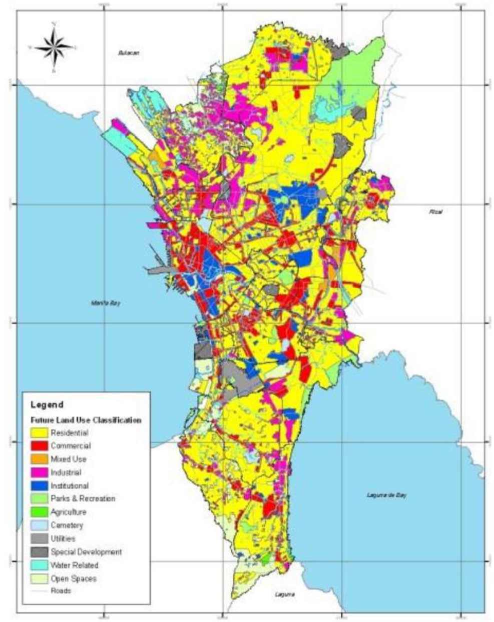
Difficulties and shortcomings of traditional methods

- Outputs change drastically
- Unable to capture macroeconomic factors
- Human error and unconscious bias

Difficulties in Philippine real estate valuation

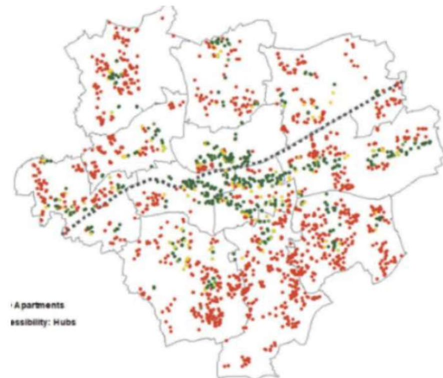
- Multiple valuation systems imposed by different government entities
- Lack of updated zonal-based market values – only 60% of LGUs updated zonal values and only 37% submitted updated market values from 2017-2020
- In order to update zonal values and land use plans, LGUs need to cooperate with multiple agencies with overlapping mandates



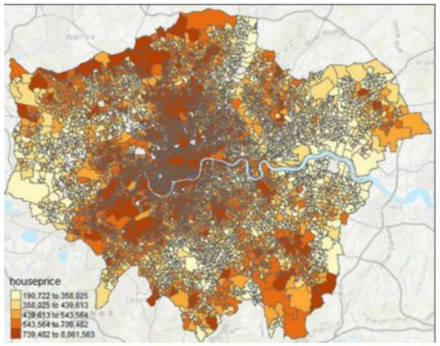


How have other countries adapted?

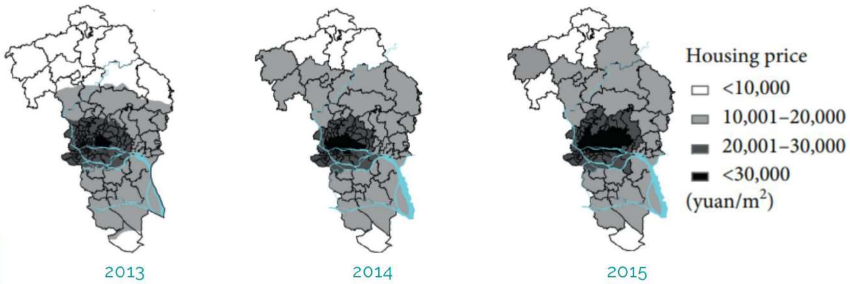
- **Spatial models** found in Dortmund, Kuala Lumpur, Guangzhou, London, and Shanghai studies for price predictions with reliable accuracy
- **Linear** and **tree-based machine learning (ML) models** have been utilized in identifying real estate opportunities around the world
- **Neural networks** and **fuzzy logic** used for apartments and other real estate properties
- **Unsupervised ML techniques** for pre-grouping properties with similar characteristics



Creating adapted walk scores for households in Dortmund



Identifying property hotspots and coldspots with spatial machine learning in London



Spatial and temporal evolution of housing prices in Guangzhou

Could the Philippines adapt as well?

Objective

- To evaluate the **effectiveness of utilizing commonly used ML techniques** of properties in the Philippines
- To verify the usefulness of **alternative data** not commonly used by Philippine real estate agents

Scope and Limitations

- Limited to real estate properties (i.e. **houses** only) found in **Metro Manila** and the province of **Cavite**, with 14,138 and 2,854 houses each
- Data utilized varies from **2021 to 2022**, dependent on latest availability
- Total list of amenities/buildings scraped may not complete or as representative

Research Questions

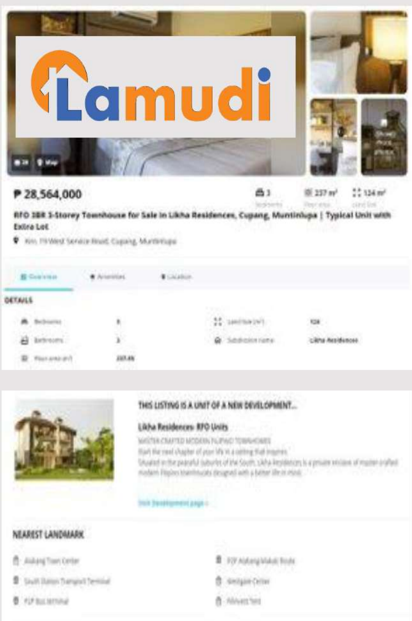
1. Are **commonly used ML techniques** found in similar property prediction publications **also effective** under a Philippine context?
2. Does incorporating **socio-economic indicators** and **geolocation data** provide predictive power in the estimation of property prices in areas from the Philippines?
3. Will the use of **indicators measured by government entities** have a substantial effect in increasing model performance related to machine learning-based property valuations?
4. **How comparable is the effect of geolocation data** to the inherent characteristics of the properties such as floor area and land size?
5. Can **characteristics with LGU granularity** still positively affect the accuracy of property price prediction?

15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

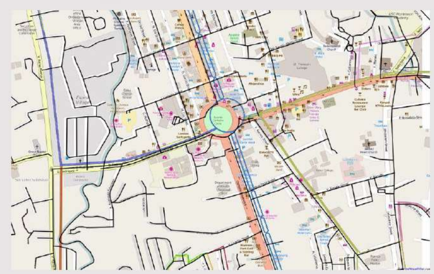
Data Sources

Lamudi



Cities and Municipalities Competitiveness Index (CMCI)

OpenStreetMap (OSM)



Socio-economic datasets from PSA

Utilization of Government-Based and Non-Conventional Indicators for Property Value Prediction in the Philippines

Data Sources – Variable Groups

Land Size (m^2)
House Price (Php)
Target Variable: Price/sqm.

Lamudi

Location
(ex. Longitude, Latitude, Postcode, LGU, Region, Subdivision)

Amenity
(ex. # of AC units, balconies, decks, fences, pools, fitness centers, garages, furnishments, parks, airports, etc.)

Property Specification
(ex. # of bedrooms, # of bathrooms, floor area (m^2), land size (m^2), total rooms, property classification)

Price
In Php

Cities and Municipalities Competitiveness Index (CMCI)

Pillar Indicators
(ex. Economic Dynamism, Government Efficiency, Infrastructure, Resiliency)

Economic Dynamism
(ex. Local Economy Size, Local Economy Growth, Productivity, Cost of Living, Capacity to Generate Employment)

Government Efficiency
(ex. Capacity of Health Services, Security, Presence of Investment Promotions Unit, Social Protection, Capacity of Schools)

Infrastructure
(ex. Existing Road Network, Availability of Basic Utilities, Connection of ICT, Health Infrastructure, Education Infrastructure)

Resiliency
(ex. Land Use Plan, Disaster Risk Reduction Plan, Early Warning System, Local Risk Assessments, Employed Population)

OpenStreetMap (OSM)

Neighborhood Amenities 1, 3, 5km away walking distance
(ex. # of Cafes, Fast Food, Pubs, Restaurants, Colleges, Universities, Gas Stations, ATMs, Banks, Clinics, Hospitals, Pharmacies, Police Stations, Marketplaces)

Neighborhood Buildings 1, 3, 5km away walking distance
(ex. # Residentials, Commercials, Industrials, Retail Stores, Supermarkets, Fire Stations, Government Buildings)

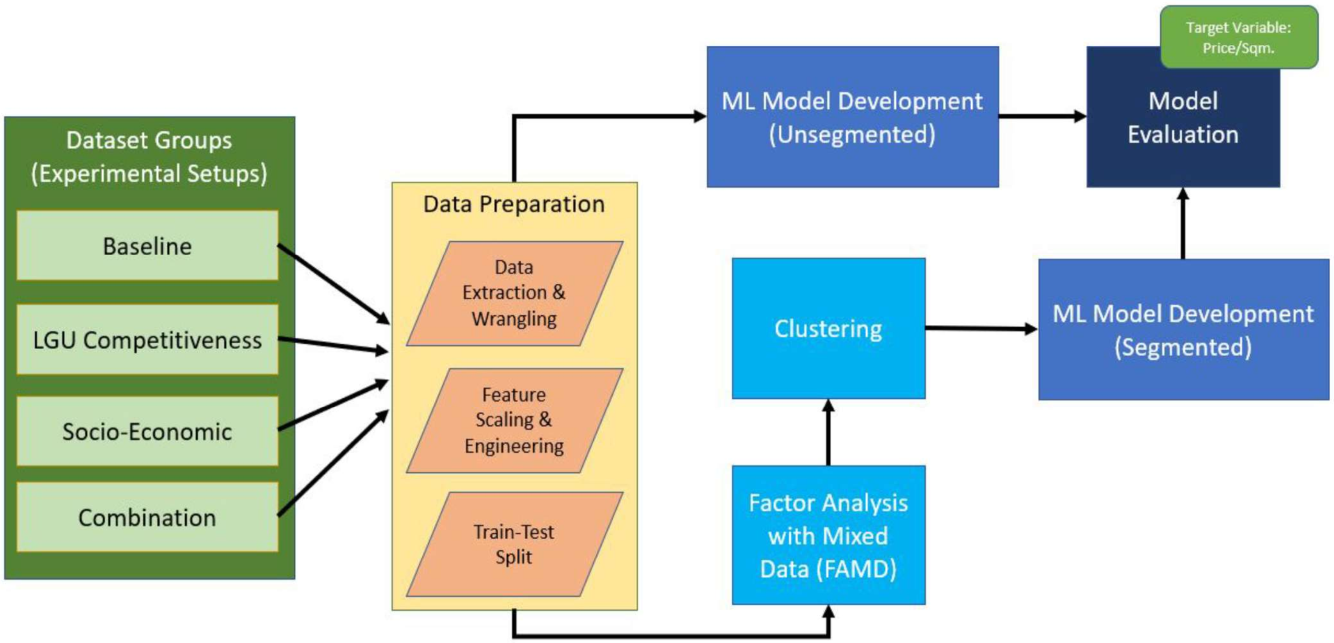
Socio-economic datasets from PSA

LGU Expenditures and Income
(ex. Total Capital Expenditures (2021), Total social services expenditures (2021), Annual Regular Income (2021))

Population and Population Growth
(ex. LGU population (2022), 5- and 10-year population growth rate)

Poverty
(ex. LGU Poverty Incidence Rate (2021), LGU Subsistence Rate (2021))

Methodology and Experimental Setup



Experiment	Datasets Used
True Baseline	Lamudi + OSM
LGU Competitiveness	Lamudi + OSM + CMCI
Socio-Economic	Lamudi + OSM + Government
Combination	Lamudi + OSM + CMCI + Government

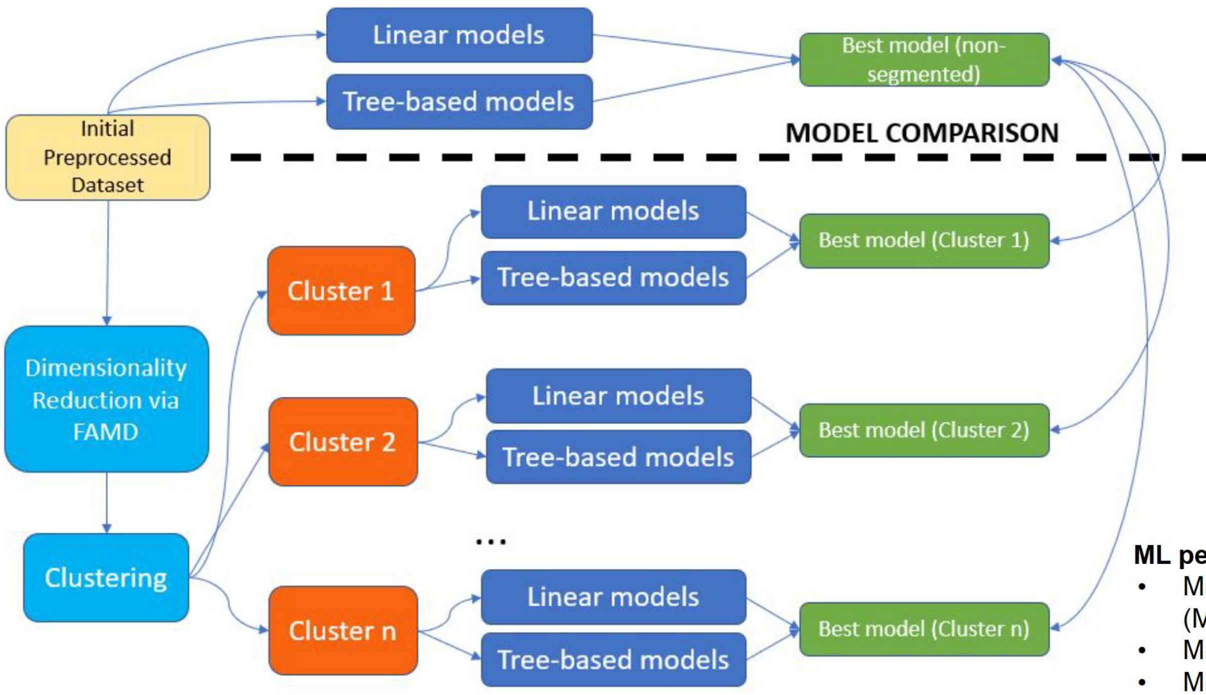
Feature Design and Selection

- Correlation analysis
- One-hot encoding
- Variance threshold
- Select K-Best via mutual information regression
- For segmentation: Factor Analysis of Mixed Data to create principal components for clustering

Validation procedures

- 80/20 train-test split data
- Hyperparameter tuning through five-fold cross validation

Machine Learning Setup and Evaluation



Model Group	Model Name
Linear Models	Ordinary Least Squares (OLS)
	Ridge Regression
	Lasso Regression
Tree-Based Models	Decision Tree Regressor
	AdaBoost estimator on Decision Tree Regressor
	Gradient Boosting Machine Regressor
	Random Forest Regressor
	Extremely Randomized Trees Regressor
	Bagging estimator on Support Vector Regression
	Stacking Regressor
	XGBoost Regressor
	LightGBM Regressor
Clustering Algorithms	K-Means
	Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)

ML performance metrics
- Mean Absolute Percentage Error (MAPE)
- Mean Absolute Error (Mean AE)
- Mean Absolute Error (Median AE)
- R^2 score

Clustering metrics
- Inertia / Sum-of-squares error
- Calinski-Harabasz index
- Silhouette score

15TH NATIONAL CONVENTION ON STATISTICS

03-05 OCTOBER 2022

Discussion and Results – Non-Segmented Approach

Best Models per Data Setup

Location	Data Setup	Best Model	MAPE (%)	Mean AE	Med AE	R^2 Score
Cavite	Baseline	AdaBoost	22.10%	10,774	5,000	0.58
	LGU Comp.	AdaBoost	20.41%	9,630	5,380	0.68
	Socio-Econ.	AdaBoost	21.59%	10,241	5,633	0.70
	Combi.	AdaBoost	20.70%	9,846	5,977	0.74
Metro Manila	Baseline	ERT	58.86%	144,697	41,266	0.76
	LGU Comp.	Stacking	57.15%	140,932	41,471	0.74
	Socio-Econ.	GBM	58.22%	149,307	37,692	0.71
	Combi.	GBM	54.66%	145,320	38,218	0.72

- Cavite: AdaBoost algorithms achieved best performances, with MAPE values in **20-21%**
- Metro Manila: considerably worse, best MAPE values reaching **54-58%**
- **Tree-based models** outperformed linear models consistently

Non-Segmented Approach			Cavite				Metro Manila			
Data Setup	Model Type	Algorithm	MAPE (%)	Mean AE	Median AE	R^2 Score	MAPE (%)	Mean AE	Median AE	R^2 Score
Baseline	Linear	OLS	46.62%	20,864	14,531	0.26	376.90%	349,579	182,378	0.19
		Ridge	46.30%	20,605	14,438	0.27	376.80%	349,552	182,351	0.19
		Lasso	46.30%	20,605	14,445	0.27	376.90%	349,578	182,390	0.19
	Tree-Based	Decision Tree	33.05%	15,744	6,344	0.33	68.31%	195,851	43,892	0.54
		AdaBoost	22.10%	10,774	5,000	0.58	64.88%	147,023	41,800	0.75
		GBM	27.75%	11,610	5,164	0.55	63.25%	165,954	40,344	0.69
		RF	26.33%	11,545	5,705	0.60	59.90%	155,228	41,382	0.73
		ERT	27.43%	11,771	5,921	0.56	58.86%	144,697	41,266	0.76
		Bagging	44.58%	19,944	11,161	0.21	130.85%	241,310	71,970	0.06
		Stacking	26.84%	12,281	5,619	0.54	59.78%	150,028	41,927	0.75
		XGBoost	26.59%	11,892	5,426	0.54	65.57%	162,875	42,361	0.68
		LightGBM	29.21%	13,253	7,438	0.26	67.98%	144,912	46,253	0.78
LGU Comp.	Linear	OLS	39.28%	17,450	13,510	0.38	60.71%	187,381	52,050	0.52
		Ridge	32.88%	15,283	11,114	0.50	74.29%	216,570	58,593	0.37
		Lasso	37.01%	17,010	11,510	0.37	70.52%	195,810	55,690	0.42
	Tree-Based	Decision Tree	33.08%	16,493	8,333	0.02	67.52%	176,700	45,045	0.55
		AdaBoost	20.41%	9,630	5,380	0.68	65.51%	133,924	41,666	0.76
		GBM	23.72%	10,766	5,470	0.61	61.39%	144,719	39,471	0.73
		RF	29.12%	13,078	5,953	0.28	68.32%	167,941	45,264	0.65
		ERT	23.19%	10,695	5,606	0.64	60.65%	135,536	38,033	0.75
		Bagging	34.97%	16,293	11,493	0.41	128.00%	237,270	68,443	0.19
		Stacking	26.05%	12,032	7,937	0.66	57.15%	140,932	41,471	0.74
		XGBoost	23.80%	10,896	5,570	0.63	76.47%	171,120	52,263	0.66
		LightGBM	23.48%	10,954	6,675	0.69	62.76%	134,298	42,407	0.77
Socio-Economic	Linear	OLS	35.26%	16,822	12,703	0.45	62.30%	188,818	51,657	0.56
		Ridge	33.27%	15,856	10,917	0.49	64.54%	189,990	52,234	0.53
		Lasso	35.33%	16,671	11,588	0.44	66.23%	192,390	53,541	0.47
	Tree-Based	Decision Tree	29.04%	14,635	8,648	0.46	66.55%	184,665	43,489	0.55
		AdaBoost	21.59%	10,241	5,633	0.70	65.81%	144,497	41,779	0.74
		GBM	23.52%	10,851	6,343	0.71	58.22%	149,307	37,692	0.71
		RF	28.25%	14,149	7,868	0.49	64.36%	179,674	41,659	0.59
		ERT	25.81%	11,884	7,898	0.68	61.37%	145,617	38,603	0.74
		Bagging	32.84%	14,836	10,034	0.51	136%	245,210	68,398	0.16
		Stacking	27.15%	12,577	8,658	0.66	62.70%	164,017	44,992	0.67
		XGBoost	22.79%	10,499	6,790	0.73	59.13%	145,880	39,499	0.74
		LightGBM	25.14%	11,778	7,804	0.66	60.58%	143,012	41,630	0.77
Combi.	Linear	OLS	29.34%	15,094	10,590	0.51	60.41%	177,371	52,050	0.52
		Ridge	31.11%	14,559	9,936	0.55	73.29%	216,570	58,593	0.37
		Lasso	36.39%	16,918	11,676	0.42	71.52%	195,810	55,690	0.42
	Tree-Based	Decision Tree	31.07%	14,913	8,830	0.43	60.06%	169,665	41,666	0.65
		AdaBoost	20.70%	9,846	5,977	0.74	64.89%	145,554	41,585	0.73
		GBM	23.75%	11,022	6,632	0.68	54.66%	145,320	38,218	0.72
		RF	28.24%	14,130	8,507	0.49	62.35%	174,510	42,182	0.58
		ERT	26.14%	12,006	7,991	0.66	59.15%	141,927	37,492	0.74
		Bagging	32.32%	14,978	10,296	0.50	N/A	N/A	N/A	N/A
		Stacking	26.89%	12,090	7,466	0.66	N/A	N/A	N/A	N/A
		XGBoost	23.69%	10,637	6,764	0.72	61.48%	147,729	39,658	0.74
		LightGBM	25.57%	11,808	7,933	0.66	63.16%	137,118	41,086	0.79

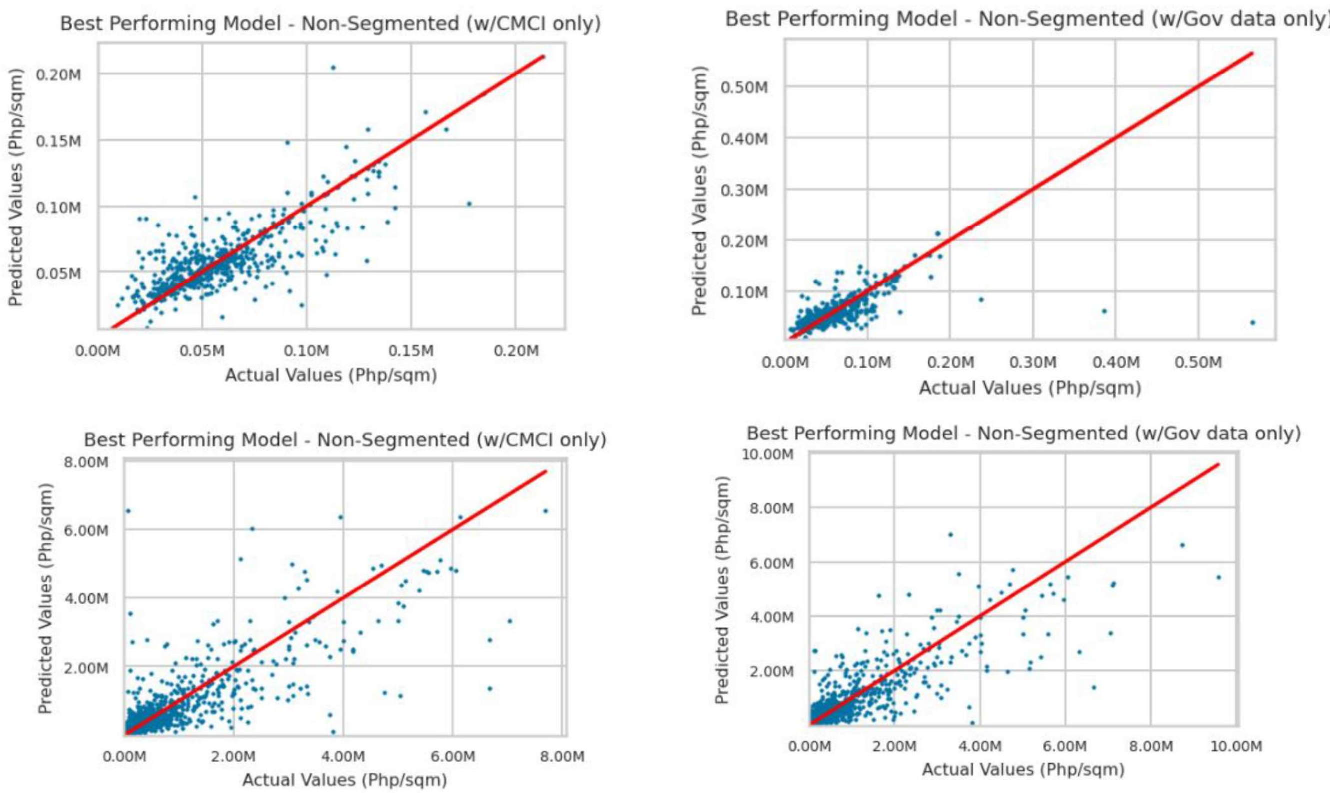
*green = best-performing

Predicted vs. Actual

- Skewed by a few significantly misclassified houses: some off by Php 20,000-90,000/sqm.
- Overfitting on other models may have occurred

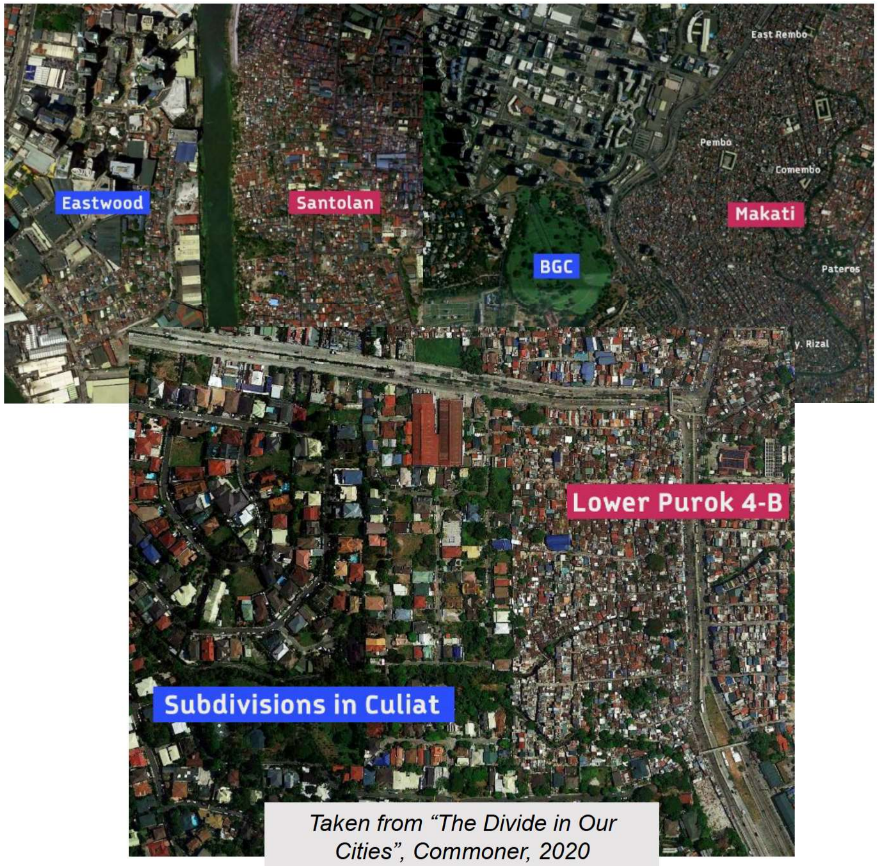
Cavite

Metro Manila



Why is Metro Manila performing subpar?

- Lack of granularity: zonal values, average income, and other factors **may highly vary across barangays and populated areas** within a single LGU
- Recent and trustworthy socio-economic data are **only readily available** at an LGU level



Taken from "The Divide in Our Cities", Commoner, 2020

Feature Importances

- **Property specification** characteristic feature dominantly: "Floor Area", "# of Bedrooms", "# of Car Spaces"
- **Location-based attributes** sourced from OSM are also apparent
- Lack of CMCI and socio-economic indicators in both locations
- Despite lack of appearance, alternative data improves model performance

Feature	Category	Feature Importance
Floor Area	Property Specification	0.30592
# of bathrooms	Property Specification	0.03991
# of residential buildings w/in 5kms	Location	0.03690
# of bedrooms	Property Specification	0.03494
# of pubs w/in 3kms	Location	0.03275
# of schools w/in 5kms	Property Specification	0.02407
# of car spaces	Property Specification	0.02268
# of fences	Location	0.02211
# of fuel stations w/in 1kms	Location	0.02206
presence of local airport	Location	0.01725

Cavite Overall Best-Performing Model
(AdaBoost – LGU Competitiveness)

Feature	Category	Feature Importance
Floor Area	Property Specification	0.24623
LGU Cost of Doing Business	CMCI	0.16773
# of basketball courts	Property Specification	0.10473
# of government buildings w/in 5 kms	Location	0.03438
# of bedrooms	Property Specification	0.03091
# of pools	Property Specification	0.02920
# of Schools w/in 1km	Location	0.02462
Having Postcode of 1231	Location	0.02459
# of Car Spaces	Property Specification	0.02402
# of Grass patches	Property Specification	0.02289

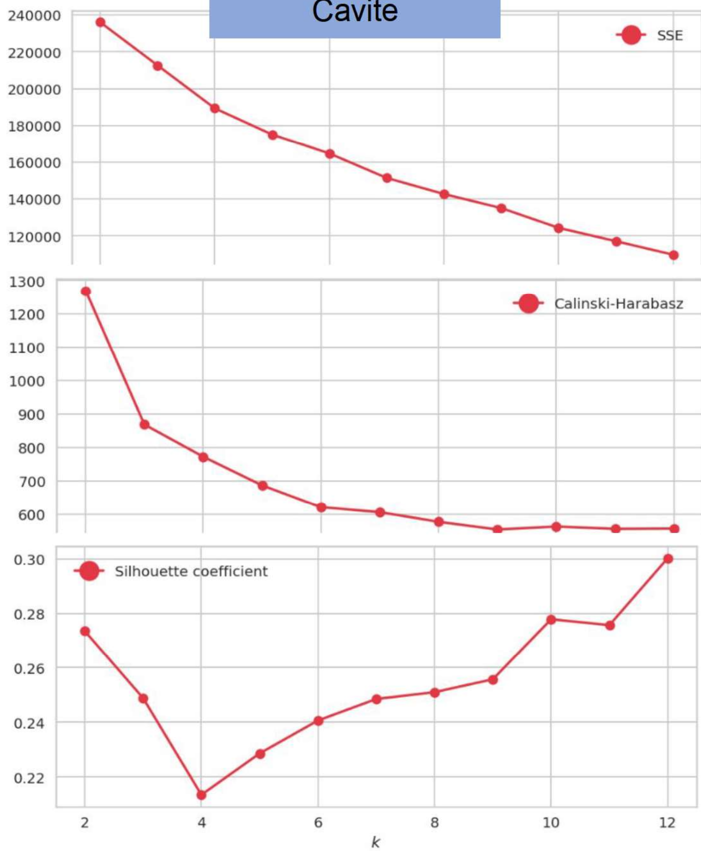
Metro Manila Overall Best-Performing Model
(GBM – Combination)

*Postcode 1231 = Pasong Tamo

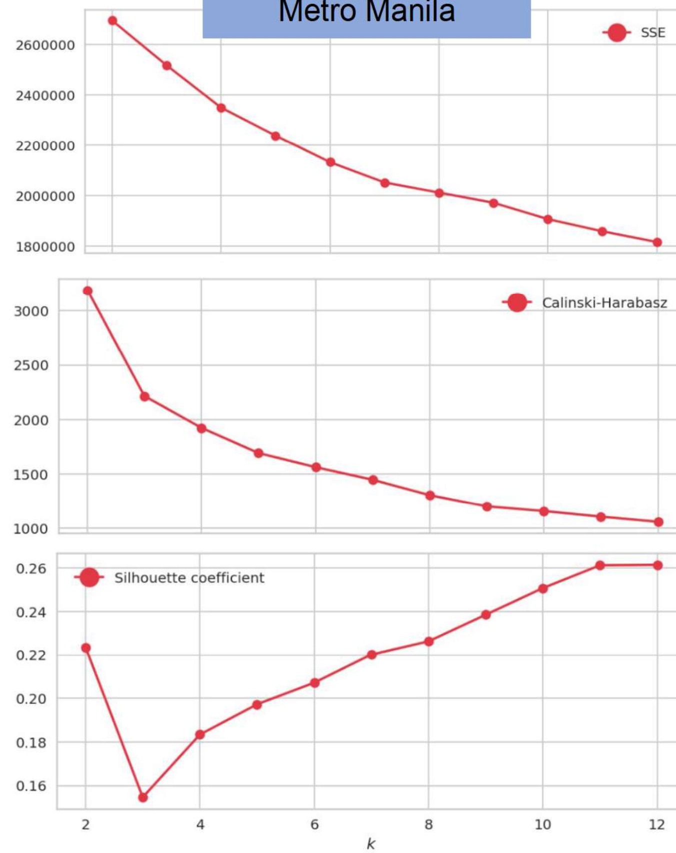
15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

Model Evaluation – Segmented Approach

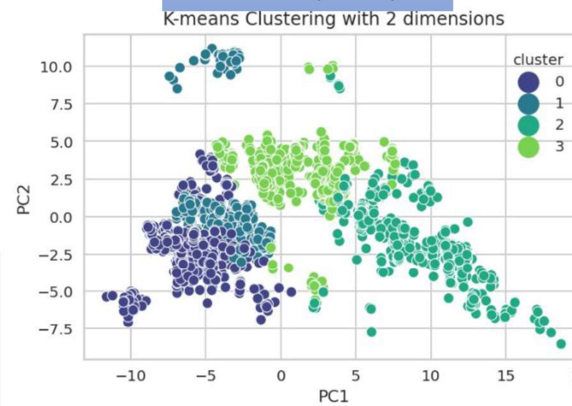
Cavite



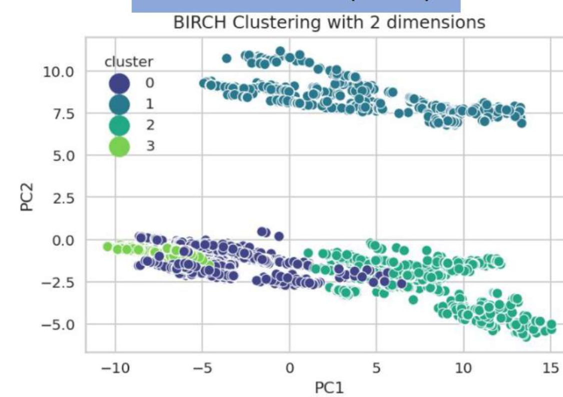
Metro Manila



Cavite (k = 4)



Metro Manila (k = 4)



Results and Discussion – Segmented Approach

Location	Data Setup	Best Model	MAPE (%)	Mean AE	Med AE	R^2 Score
Cavite	Baseline	AdaBoost	22.10%	10,774	5,000	0.58
	LGU Comp.	AdaBoost	20.41%	9,630	5,380	0.68
	Socio-Econ.	AdaBoost	21.59%	10,241	5,633	0.70
	Combi.	AdaBoost	20.70%	9,846	5,977	0.74
Metro Manila	Baseline	ERT	58.86%	144,697	41,266	0.76
	LGU Comp.	Stacking	57.15%	140,932	41,471	0.74
	Socio-Econ.	GBM	58.22%	149,307	37,692	0.71
	Combi.	GBM	54.66%	145,320	38,218	0.72

Segmented		Cavite					Metro Manila				
Data Setup	Cluster #	Best Model	MAPE (%)	Mean AE	Median AE	R2 Score	Best Model	MAPE (%)	Mean AE	Median AE	R2 Score
Baseline	1	Stacking	10.45%	5,140	3,057	0.85	ERT	55.75%	203,501	56,719	0.75
	2	AdaBoost	22.69%	11,083	4,737	0.70	AdaBoost	51.86%	77,070	32,311	0.61
	3	AdaBoost	18.68%	8,784	5,086	0.57	AdaBoost	37.76%	77,070	16,412	0.78
	4	ERT	20.52%	11,129	6,694	0.84	AdaBoost	69.90%	116,232	36,626	0.44
	Ave.	--	18.09%	9,034	4,893	--	--	53.82%	118,468	35,267	--
LGU Comp.	1	GBM	31.49%	12,101	6,843	0.39	AdaBoost	53.16%	95,408	31,146	0.59
	2	AdaBoost	20.69%	9,365	4,175	0.69	AdaBoost	63.21%	125,525	42,140	0.54
	3	AdaBoost	20.84%	8,606	5,696	0.77	RF	57.82%	208,530	49,661	0.84
	4	XGBoost	20.42%	14,684	6,906	0.43	RF	37.49%	81,209	15,914	0.76
	Ave.	--	23.36%	11,189	5,905	--	--	52.92%	127,918	34,715	--
Socio-Economic	1	GBM	21.94%	10,846	4,338	0.72	AdaBoost	42.05%	109,740	29,444	0.44
	2	AdaBoost	24.29%	16,686	8,999	0.36	AdaBoost	55.01%	100,278	41,065	0.54
	3	AdaBoost	19.99%	7,013	4,887	0.56	LGBM	69.67%	266,615	83,741	0.81
	4	AdaBoost	16.69%	9,288	5,363	0.67	AdaBoost	32.41%	86,025	15,622	0.75
	Ave.	--	20.73%	10,958	5,897	--	--	49.79%	140,665	42,468	--
Combi.	1	AdaBoost	20.90%	8,680	4,834	0.85	AdaBoost	55.29%	89,924	31,868	0.60
	2	XGBoost	28.45%	18,717	7,244	0.24	ERT	69.70%	126,574	52,290	0.64
	3	LGBM	33.56%	12,368	7,381	0.35	ERT	47.74%	191,309	46,016	0.87
	4	AdaBoost	18.15%	8,996	5,045	0.66	Stacking	38.43%	79,230	15,288	0.76
	Ave.	--	25.27%	12,190	6,126	--	--	52.79%	121,759	36,366	--

Light green = best-performing

Light blue = cluster is better than best non-performing counterpart

Green = On average, clusters of the data setup are better than its best non-performing counterpart

- AdaBoost still showing its suitability, especially for Metro Manila
- Segmented approach **performed better in data setups for Metro Manila** in all cases (5-15% MAPE, all Mean AE, most Median AE)
- Segmented approach **performed better only partially for Cavite** (1-2% MAPE, total improvement only in Baseline model)

How does the Philippines compare to other countries?

- Regression models for Kuala Lumpur had MAPEs of 11.3-20.9% and R2 scores of 0.74-0.91
- Hong Kong ML models had 32-54% MAPEs and R2 scores of 0.83-0.90
- Other studies utilized different performance metrics (e.g. coefficients, p-value, Moran's I); most comparable was R2 score
  - Shanghai and Xi'an had R2 scores of 0.70 and 0.89
  - Chile models using RF, SVM, and NN models had R2 scores ranging from 0.74-0.96
  - Dortmund utilized OLS and Spatial Lag models which had adjusted scores of 0.35-0.60

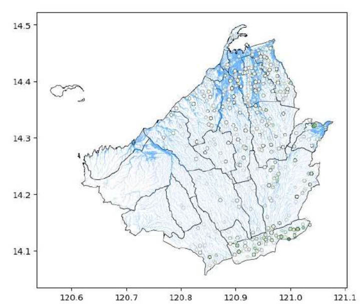
Cavite has comparable models (MAPE of 18-22%) while Metro Manila is worse (MAPE of 50-59%)

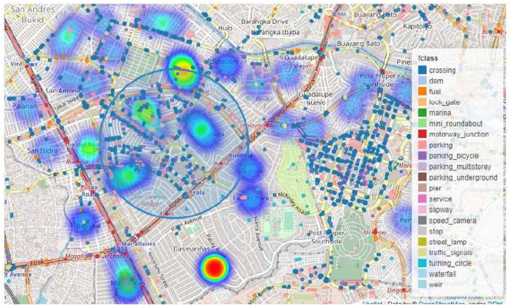
Conclusion

- Non-segmented:
    - Cavite: AdaBoost with **20.41% MAPE,** Php 9,630/sqm Mean AE, Php 5,380/sqm Median AE, and 68% R^2 score
    - Metro Manila: GBM with **54.66% MAPE,** Php 149,307/sqm Mean AE, Php 38,218 Median AE, and 72% R^2 score
- **Lack of finer granularity** may explain Metro Manila's performance, due to its varying contexts within LGUs (barangay levels)
- **Alternative data** (location of nearby amenities, CMCI, socio-economic data) **can improve property valuation prediction**
- Performing property segmentation via K-Means and BIRCH Clustering **improves Metro Manila significantly** but **Cavite partially**

Recommendations

- Comparison with **spatial econometric** models
- **Finer granularity** of government-based indicators (i.e. barangay level or smaller)
- Comparison and possible integration with procedures found in automated valuation machines (AVMs)
- Further hyperparameter tuning and usage of Deep Learning techniques
- Usage of **satellite imagery** for capturing in detail the granular features of a location
- Increase and variety of data points (i.e. houses) found in specified locations
- Obtaining a **prediction interval** of forecasted prices





Thank you!

15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Utilization of Government-Based and Non-Conventional Indicators for Property Value Prediction in the Philippines

Gabriel Isaac L. Ramolete, Bryan Bramaskara,
Dustin A. Reyes, Adrienne Heinrich

Presented by:

Gabriel Isaac L. Ramolete
Data Scientist
Aboitiz Data Innovation

Data Revolution: Big Data, Non-Official Data, Citizen-Generated Data, Big Earth Data, Data Analytics
Crowne Plaza Manila Galleria
05 October 2022, 8:30-10:00AM

Thank you!

http://www.psa.gov.ph/ncs

http://openstat.psa.gov.ph

https://twitter.com/PSAgovph

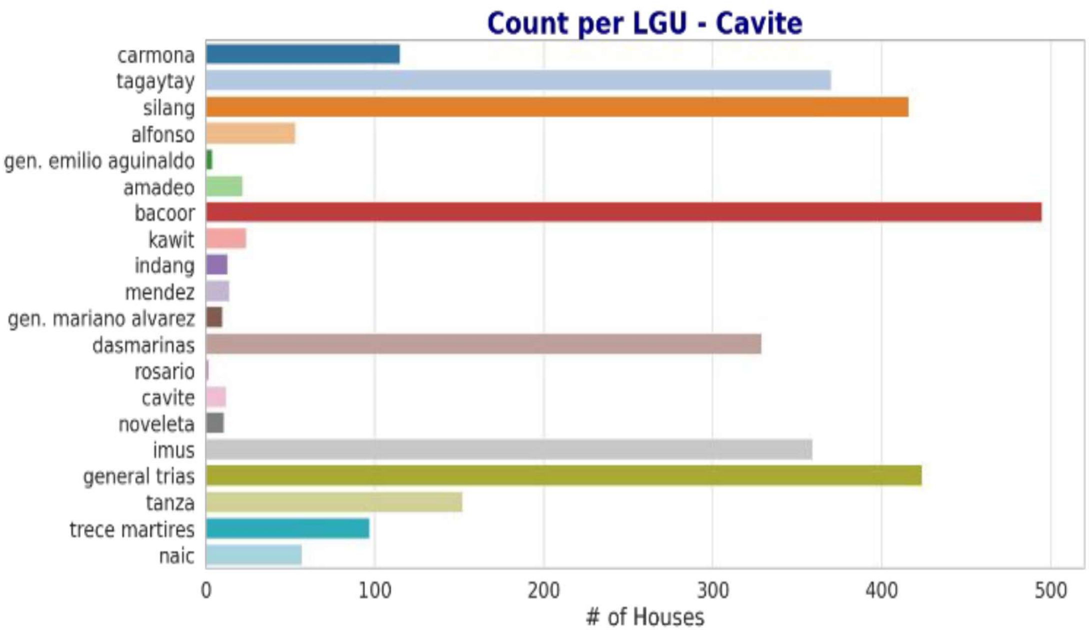
https://www.facebook.com/PSAgovph

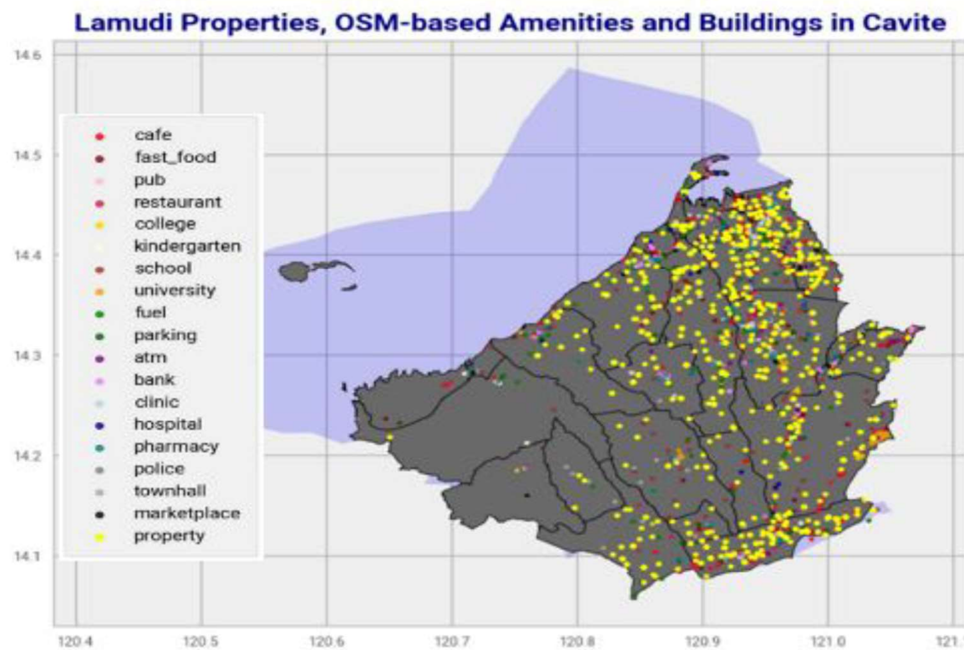
Appendix – Exploratory Data Analysis

15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Exploratory Data Analysis

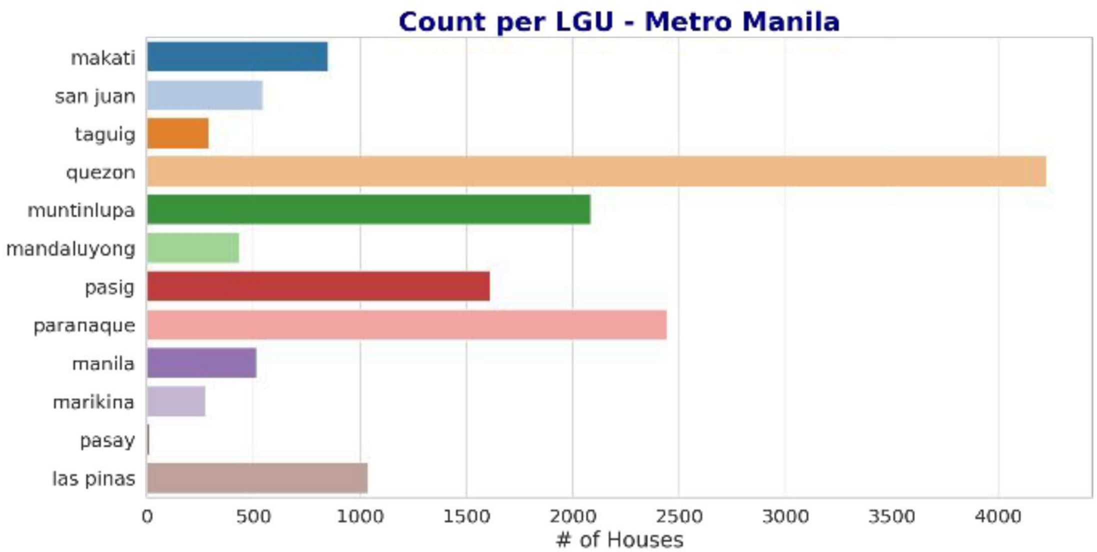


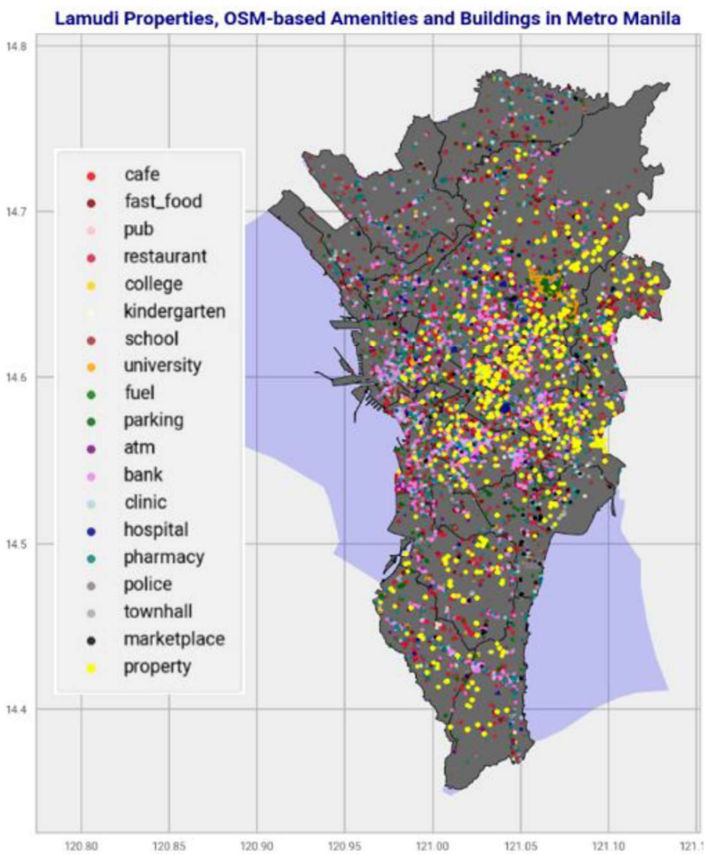


15TH NATIONAL CONVENTION ON STATISTICS
03-05 OCTOBER 2022

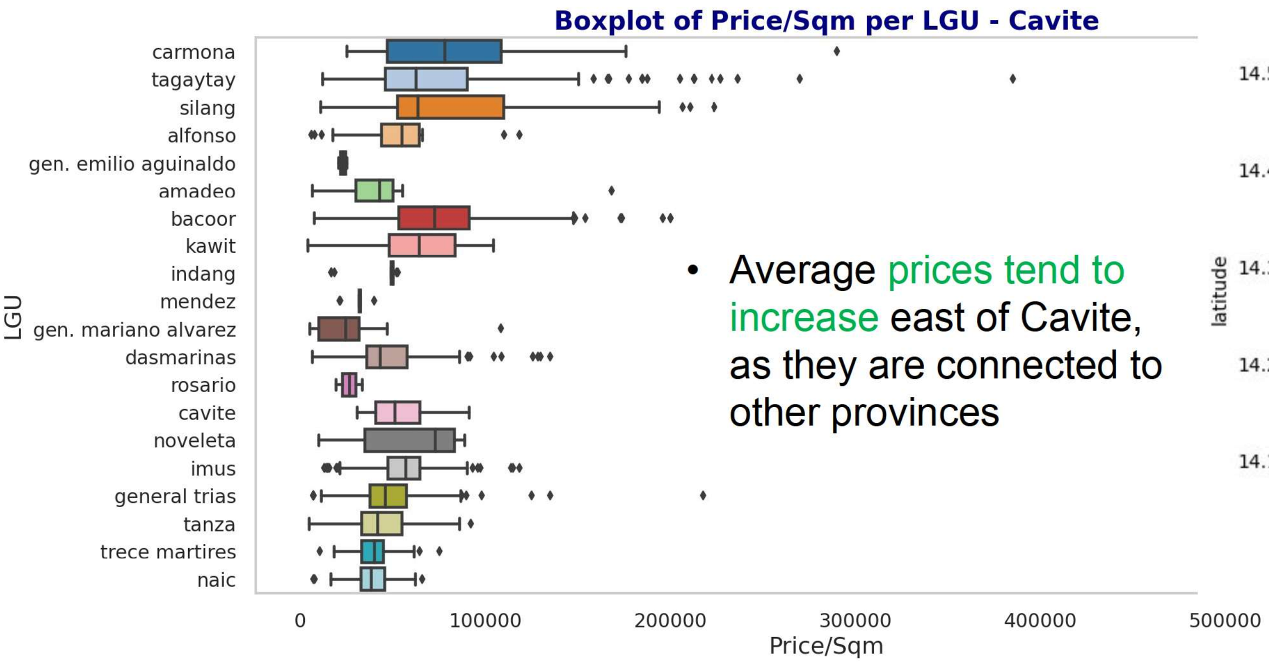
Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Exploratory Data Analysis

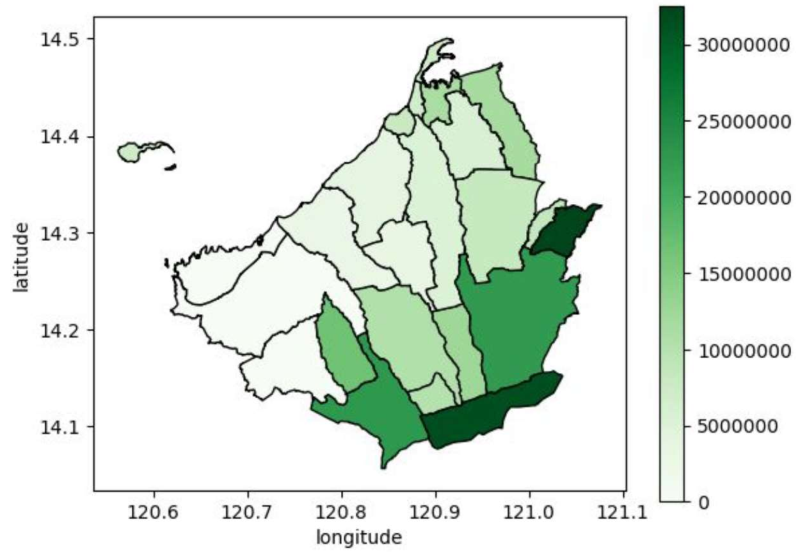




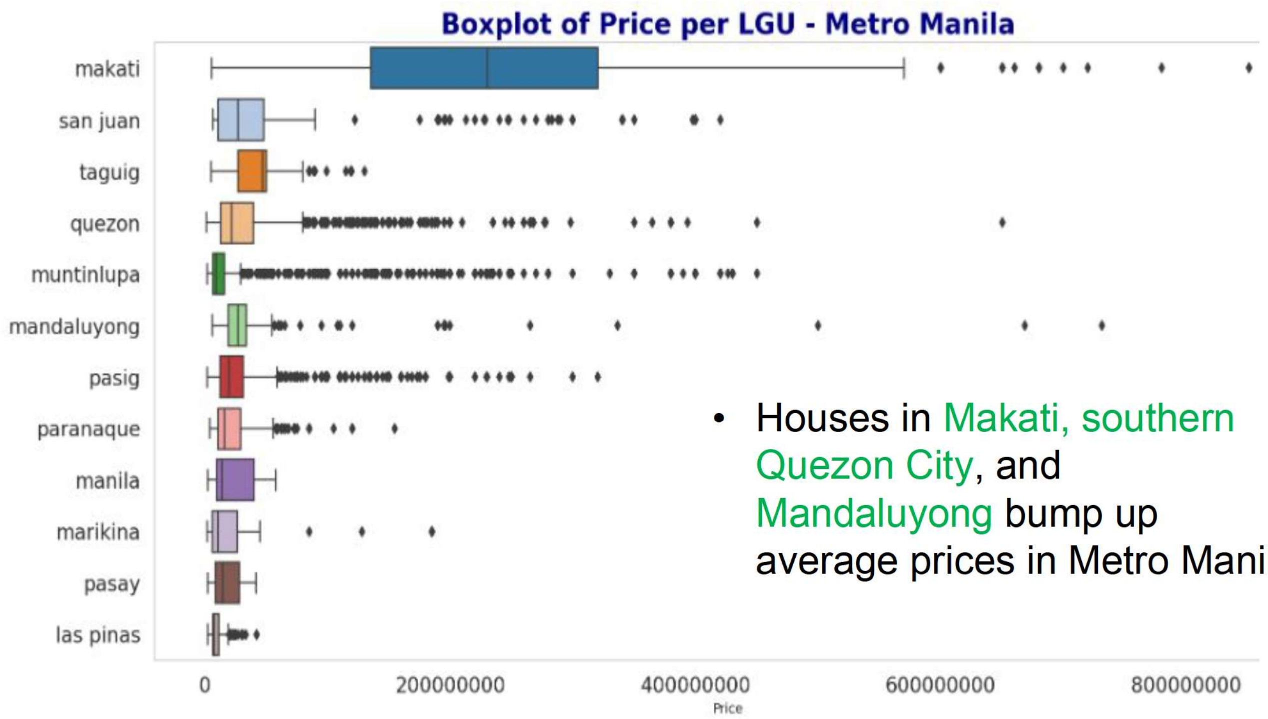
Exploratory Data Analysis



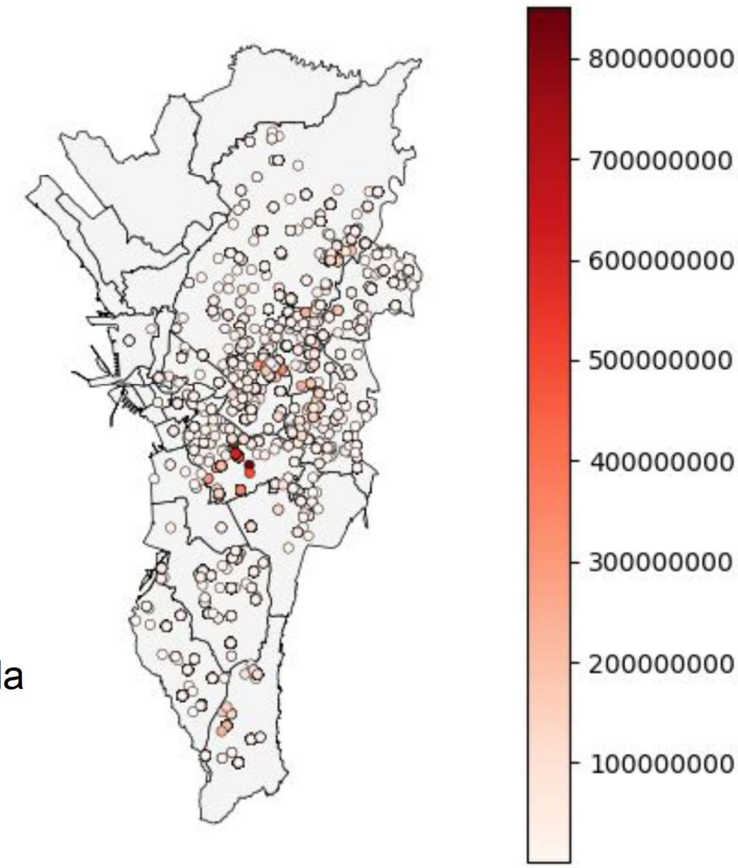
- Average prices tend to increase east of Cavite, as they are connected to other provinces



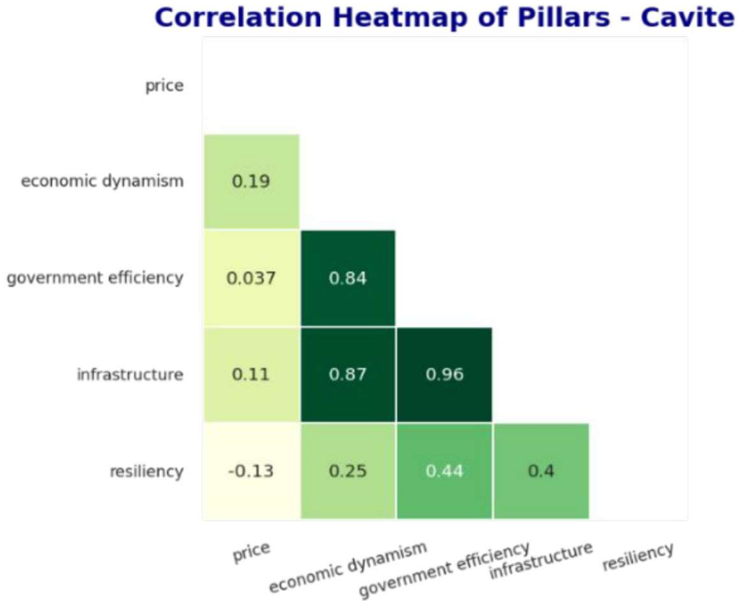
Exploratory Data Analysis

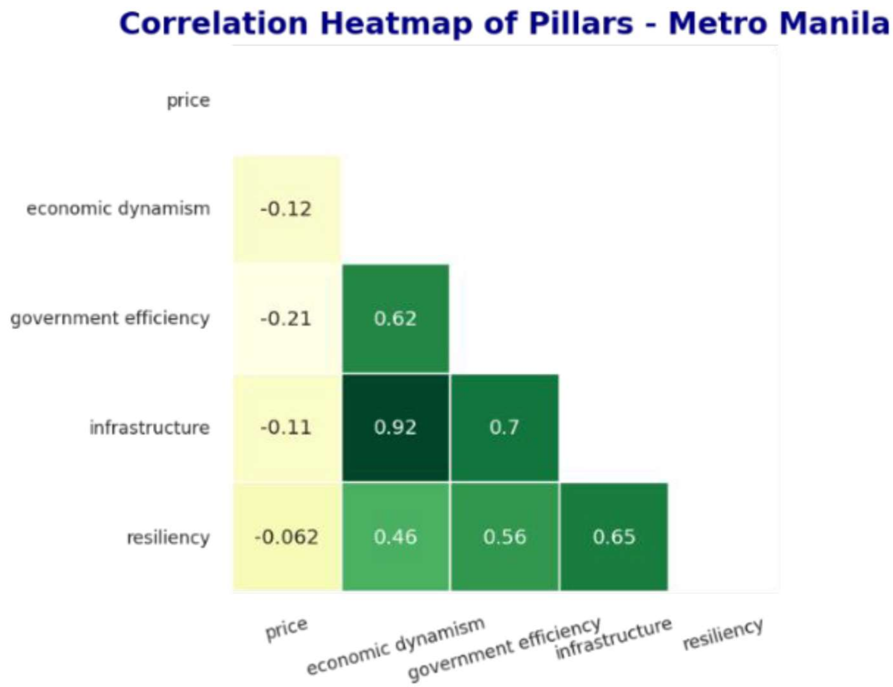


- Houses in Makati, southern Quezon City, and Mandaluyong bump up average prices in Metro Manila



Exploratory Data Analysis





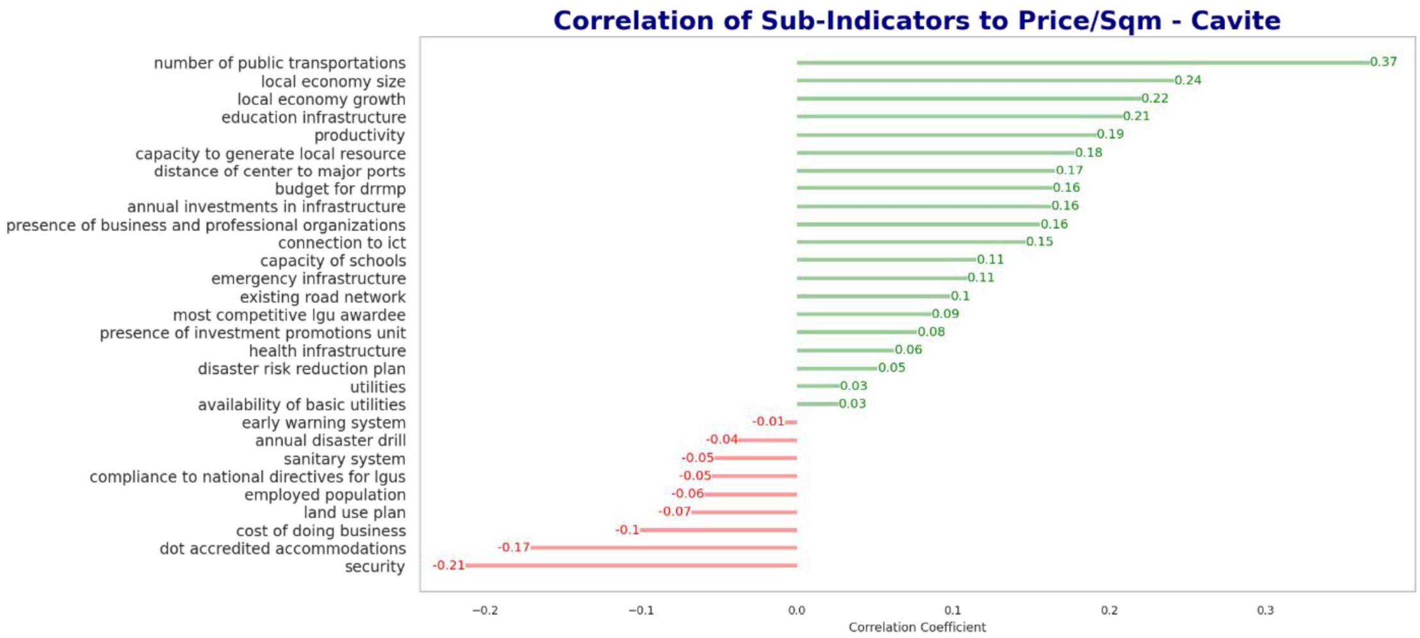
15TH NATIONAL CONVENTION ON STATISTICS

03-05 OCTOBER 2022

Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Exploratory Data Analysis

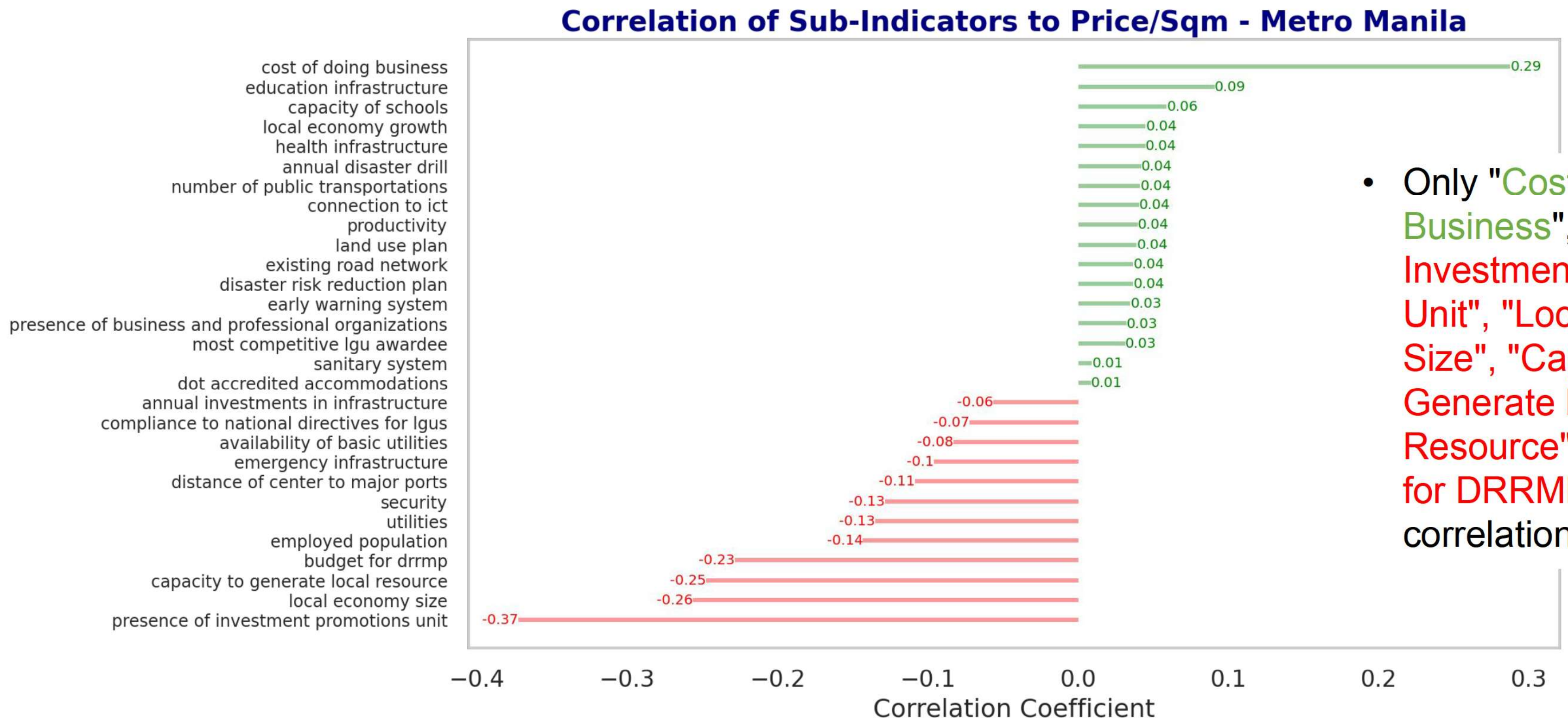
NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU



- Most indicators do not have relatively significant influence on Cavite's price/sqm

Exploratory Data Analysis

NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU

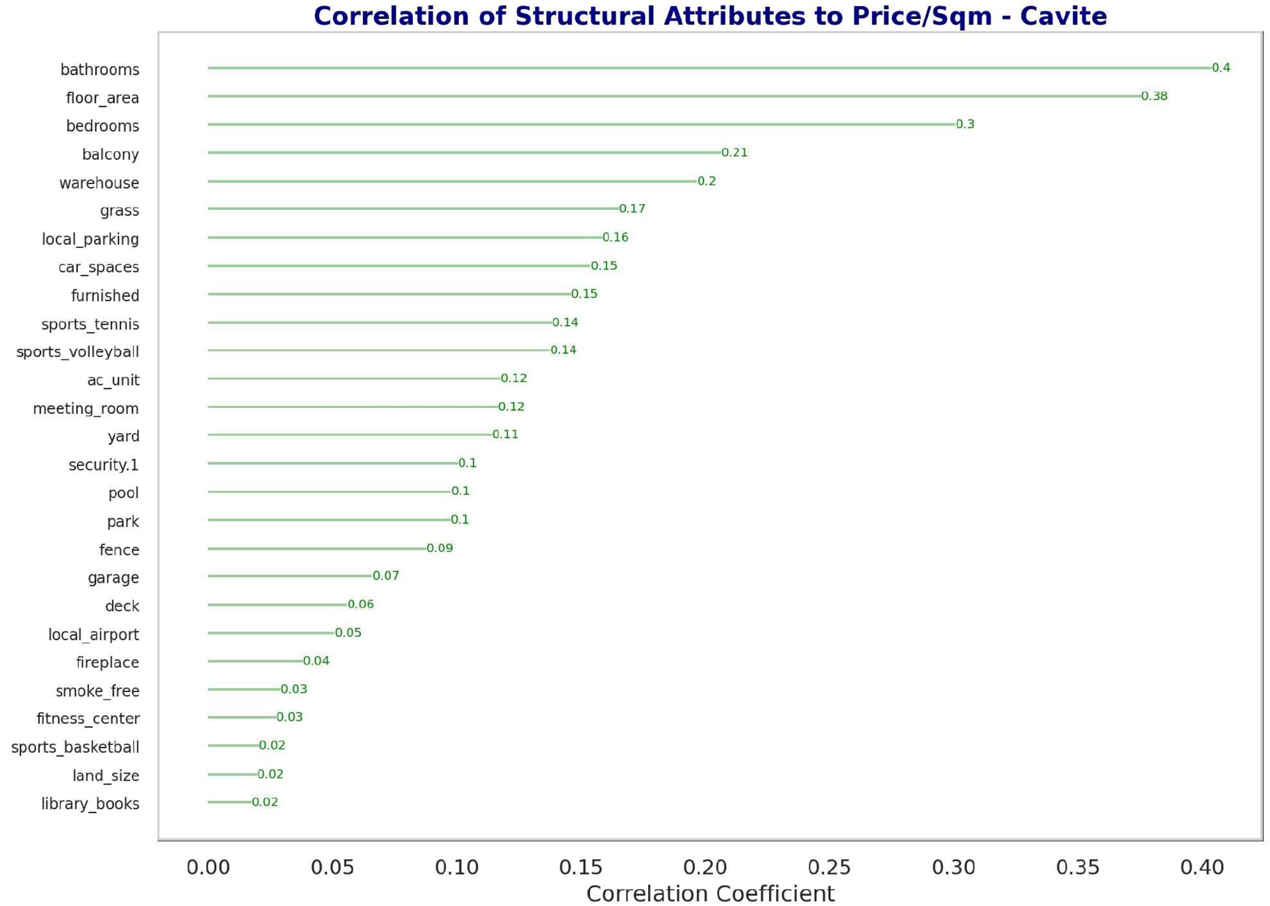


- Only "Cost of Doing Business", "Presence of Investment Promotions Unit", "Local Economy Size", "Capacity to Generate Local Resource", and "Budget for DRRMP" have notable correlations

NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU

Exploratory Data Analysis

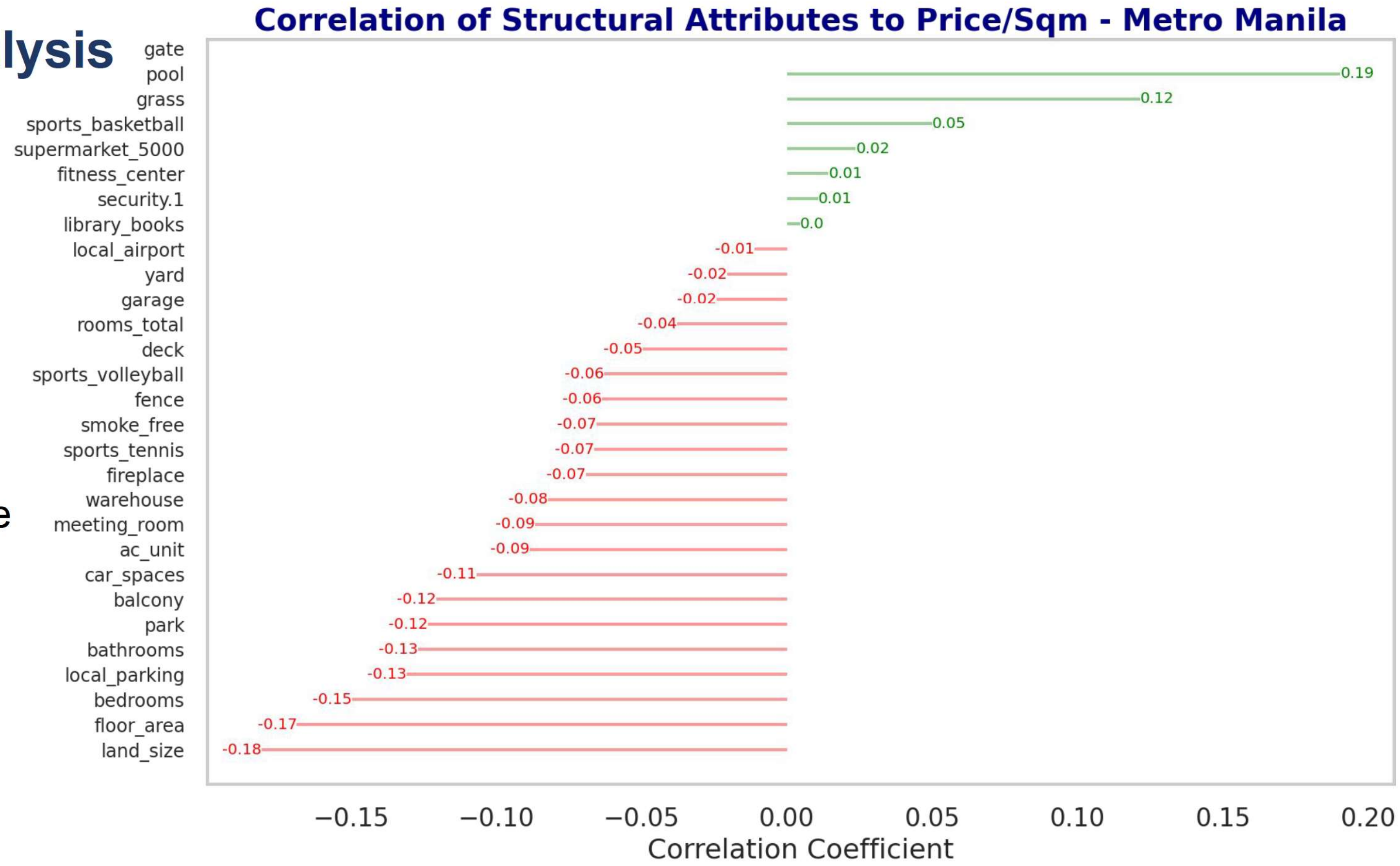
- Bathrooms, floor area, and bedrooms positively correlate to price/sqm
- Other property specifications have relatively minimal correlation



NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU

Exploratory Data Analysis

- A relative lack of correlation to price/sqm for most structural attributes
- Contrary to intuition, bedrooms, bathrooms, and floor area negatively correlate to price/sqm

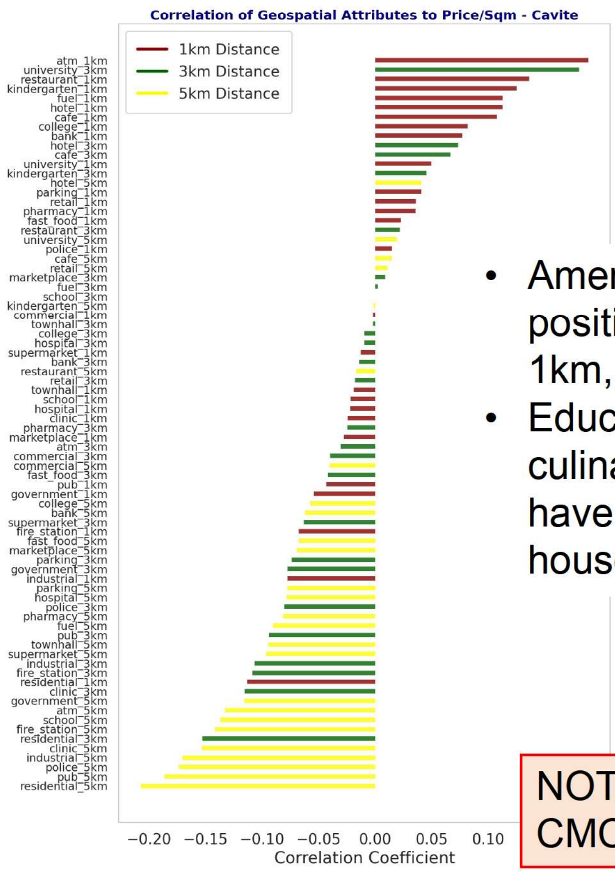


15TH NATIONAL CONVENTION ON STATISTICS

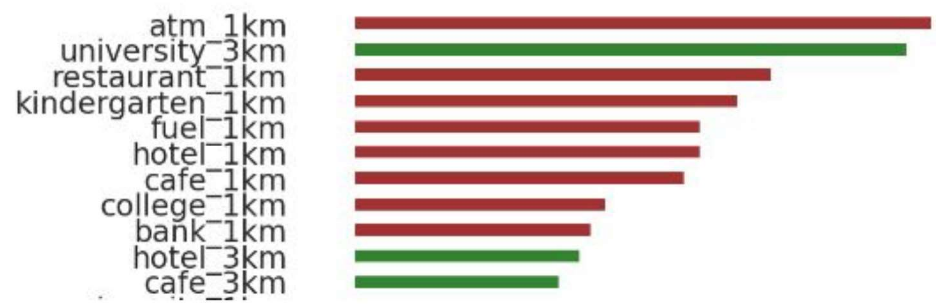
03-05 OCTOBER 2022

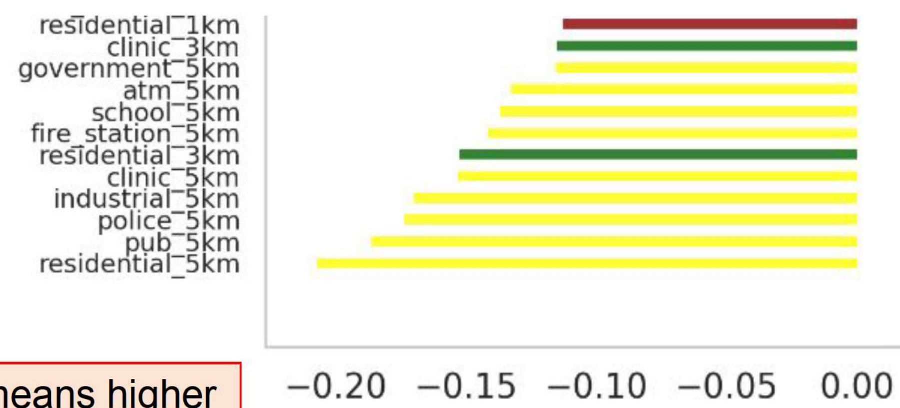
Organized by the Philippine Statistical System Spearheaded by the Philippine Statistics Authority

Exploratory Data Analysis



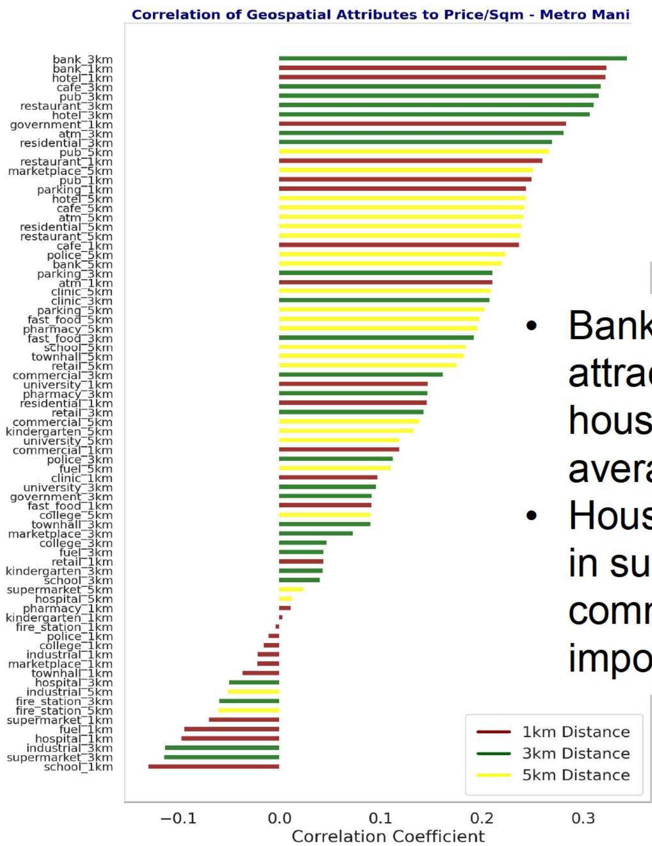
- Amenities nearer houses positively correlate (i.e. 1km, 3km)
- Education, financial, and culinary areas nearby have more expensive houses

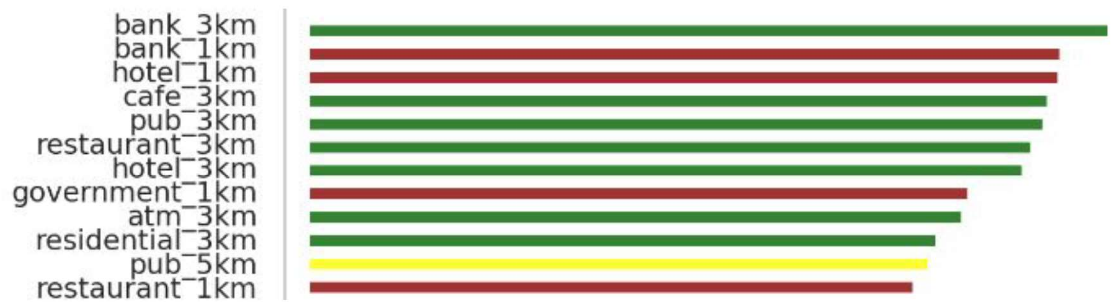




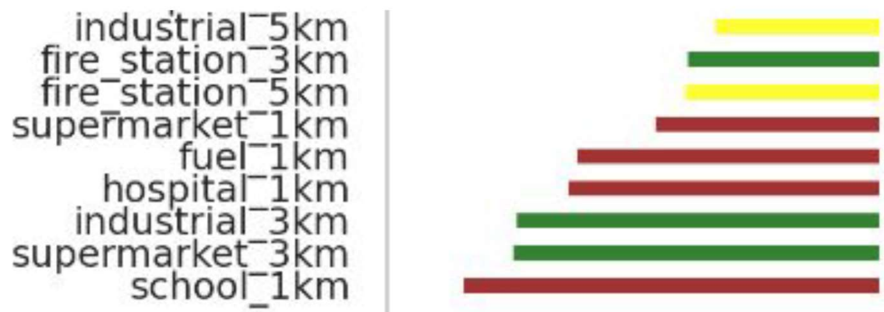
NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU

Exploratory Data Analysis





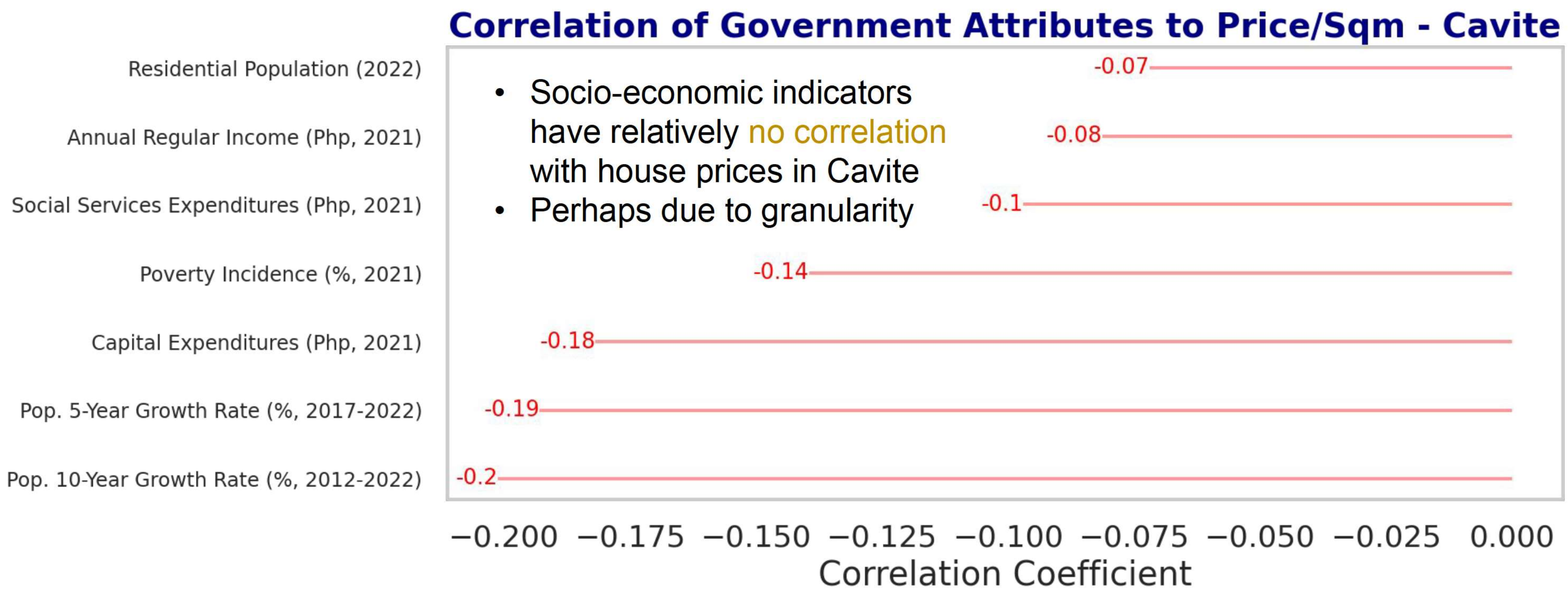
- Banks and culinary attractions relatively near houses seem to drive up average prices
- Houses may usually be in subdivisions or gated communities, hence the importance of 3km



NOTE: Positive correlation means higher CMCI rank, or a lower quality LGU

NOTE: Negative correlation means lower CMCI rank, or a higher quality LGU

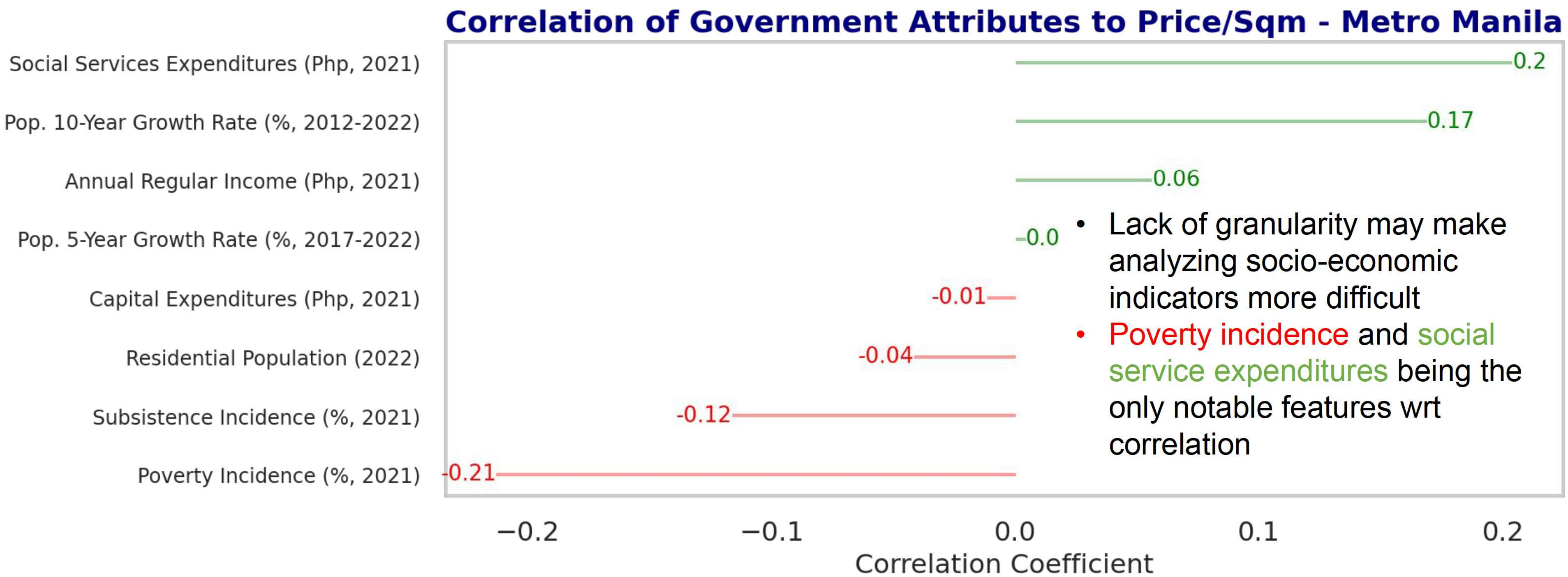
Exploratory Data Analysis



- Socio-economic indicators have relatively no correlation with house prices in Cavite
- Perhaps due to granularity

NOTE: Negative correlation means lower CMCI rank, or a higher quality LGU

Exploratory Data Analysis



- Lack of granularity may make analyzing socio-economic indicators more difficult
- Poverty incidence and social service expenditures being the only notable features wrt correlation