

An Investigation of the Impact of Amenities on Condominium Apartment Prices in Colombo, Sri Lanka with Machine Learning

Rifat Mahamood*, Gayani Ranasinghe

Department of Town and Country Planning, Faculty of Architecture, University of Moratuwa, Sri Lanka.

Abstract

Condominium prices of a location are heavily impacted by the availability of neighbourhood amenities, which attract people to that location. Studies that employ machine learning to evaluate the impacts of amenities on condominium prices are a few, despite indications of increased accuracy. Hedonic models are used in the majority of current studies on housing pricing. Yet, the application of machine learning will improve accuracy and be more effective for the identification of multi-collinearity and non-linear relationships between property prices and the availability of amenities for a wide range of parameters. The aim of this study is to investigate the possible relationship between urban amenities surrounding condominium apartments and their market prices. The study uses data from the Google Maps Platform to examine the link between neighbourhood amenities and the prices of 236 condominiums in Colombo, Sri Lanka. 56 significant amenity characteristics were found using the eXtreme gradient boosting (XGB) algorithm, and the variety of correlations between amenities and condominium prices as restricted positive, accelerated positive, crooked, humped, and negative was explained. Results demonstrated that while a beautiful urban environment requires a variety of facilities, the popularity and other attributes of amenities affect condominium prices in numerous non-linear ways. Therefore, public and private organizations must work together to create integrative strategies that enhance and maintain the variety and accessibility of urban amenities.

Keywords: Condominium; Google Maps; Housing Price; Machine Learning; Urban Amenities


1. Introduction

Condominiums can be identified as a type of housing that Western societies used commonly as means of living. It has been introduced to urban areas of the rest of the world to overcome the issues related to the scarcity of developable lands and increasing population density (Wong, 2004; Senaratne et al., 2010; Ariyawansa & Udayanthika, 2011). A condominium is known as a form of tenure or a type of ownership that the owner has the full title to the individual property unit and shares a set of common amenities and services with other individuals (Rajapakshe & Jayaweera, 2019). In Sri Lanka, the growth of these condominium escalated last

decade, after the end of civil war in 2009 (Abesinghe & Selvarajan, 2018).

The vertical living changed the skyline of many urban areas in Sri Lanka including Kandy, Galle, and Nuwara Eliya, but a vast majority of them has been emerged Colombo, the economic hub of Sri Lanka. There are handful of residential and commercial projects evolving in the suburbs also in the suburbs of Colombo, where nowadays, there is a popular trend among upcoming middle class populations to live in comfortable surrounding with modern lifestyles (Karsten, 2007; Salama et al., 2017; Madhushani & Piyadasa, 2019).

Corresponding Author:

 <https://orcid.org/0009-0005-8435-9874>

E-mail address: mahamoodmr.23@uom.lk; gayanip@uom.lk

DOI: <https://doi.org/10.4038/bhumi.v10i2.103>

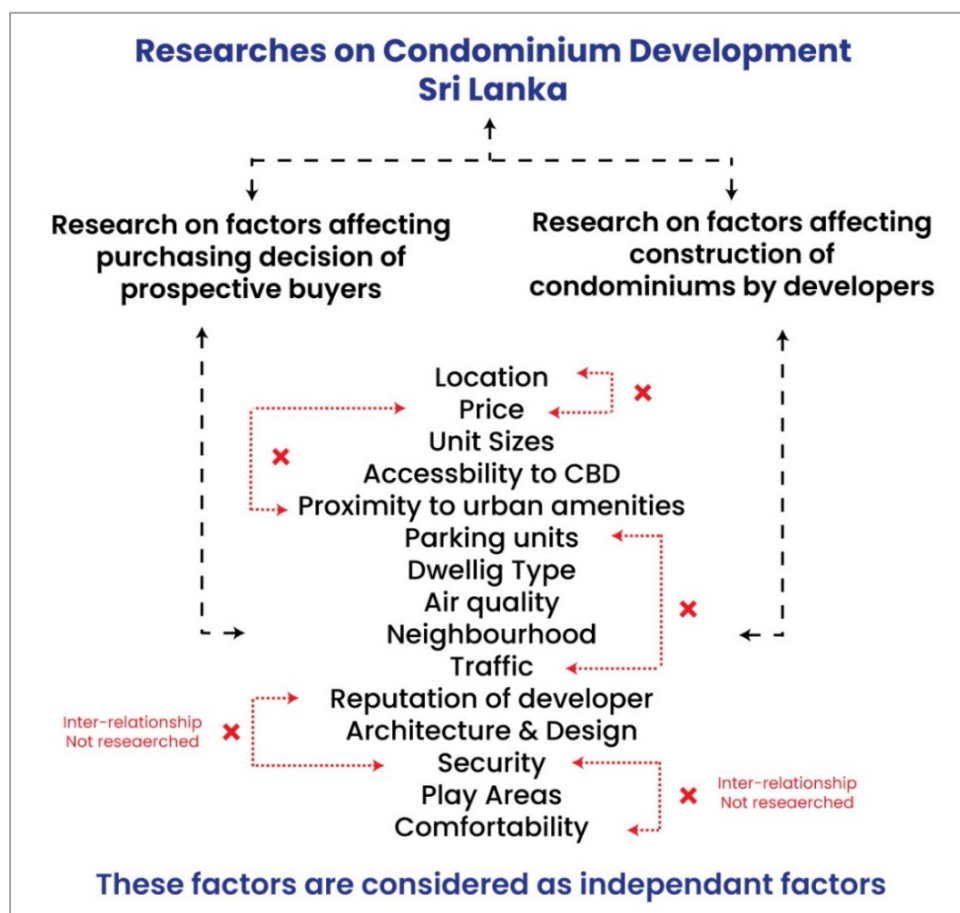


Figure 1: Previous research areas on Condominium Developments
Source: Author's development, based on literature

Buying such an apartment involves a complex decision-making process with many influencing factors for prospective buyer. Out of those factors, location, price, surrounding neighbourhood, scenic beauty and characteristics of the building are most commonly affluent (Ariyawansa & Udayanthika, 2011). - A widely used rhetoric is that location, location and location are the three most important determinants for a choice of a house (Abesinghe & Selvarajan, 2018). In that sense, highly demanded condominiums are generally located with good access to schools, hospitals, workplaces and transportation facilities (Senaratne et al., 2010). Thereby, neighbourhood amenities can be taken as the most important factor out of location category that influence condominium price and demand (Taecharungroj, 2021). But, due to attraction of people into these centrally located places, agglomeration of amenities can also be seen, and it leads to the growth of metropolitan areas. Therefore, the decision-making process of purchasing a

condominium apartment is significantly influenced by two key determinants: location and price (Madhushani & Piyadasa, 2019; Ariyawansa & Udayanthika, 2011).

However, it is worth noting that the aforementioned studies do not explicitly examine the interrelationship between these two factors. Therefore, further studies to develop thorough knowledge on the impact of urban amenities and surrounding environmental features on condominium apartment prices are necessary, because those two factors are the most concerning for the demand and eventually, to improve sustainable city planning and policy making (Taecharungroj, 2021).

The hedonic technique has been employed in several research, including the aforementioned ones, to identify the primary factors influencing a given phenomenon through the application of linear regression analysis. However, this approach has several limitations, including

its inability to account for non-linear interactions and potential concerns with multicollinearity. Hence, the utilization of machine learning techniques is proposed as a more effective way to address these constraints by investigating complex interdependencies among several factors and capturing non-linear connections. The proposed method has fewer statistical assumptions and does not necessitate data dispersion, hence resulting in enhanced accuracy in the final output. The current body of research also exhibits constraints pertaining to the acquisition of data from authoritative sources. In conventional methodologies, the process of manual data collecting necessitates a substantial investment of time and exertion. Yet, using open data sources such as Google Maps, a plethora of large-scale data pertaining to urban amenities may be accessed through the Places Application Programming Interface (API) (Taecharungroj, 2021). The examination of the correlation between geographical location and pricing is of significant use in enhancing comprehension.

This paper discusses two research questions which derive the key neighbourhood characteristics that influence condominium prices and the relationship between urban amenities and condominium apartment prices in the context of urban areas in Sri Lanka. There are two objectives in this research to fulfil the gaps in the existing literature. The first objective is to Investigate the important neighbourhood features that impact luxury condominium unit pricing in Colombo and its suburbs. The second objective is to examine the possible relationship between urban amenities and condominium prices. For this study, 236 residential condominium units in Colombo were selected. Colombo is the economic hub of Sri Lanka and recorded a significant growth in condominium volume in 2021 with a growth of 48.4% compared to 2020 (Central Bank of Sri Lanka, 2022).

The results will support housing developers and urban amenity developers in their decisions to invest on condominiums and

amenities in appropriate locations as well as to guide city planners to evolve sustainable development policies and planning interventions. Since the study attempts to clarify the process of identifying the deficiencies in neighbourhood amenities and pinpointing the areas that require community-led efforts. The existing ties will serve as indicators for determining the appropriate amenity to initiate in order to rejuvenate or attract individuals to the area. The business community gets the opportunity to conduct temporary evaluations of suggested facilities, such as pop-up bars, fashionable restaurants, temporary market stalls, parklets, and similar offerings, prior to their permanent integration.

Consequently, the results of this study will shed light on the significance of diversity and variety in neighbourhood amenities, as well as the potential impact of multifaceted amenity development on the enhancement and sustainability of housing demand. In due course, it will facilitate the establishment of a connection between public and private entities accountable for amenity development, with the aim of attaining sustainable urban planning.

2. Literature review

2.1 Condominium developments and types

According to Condominium Management Authority, a condominium plan is a means of dividing property in such a way that one individual has title to a piece of a building, or "unit," as well as a share of the remaining property that is shared by all individual unit owners. Condominiums can be residential, commercial, or industrial establishments. Buildings might be either new or old. According to the target group of condominium buyers, condominiums can be classified into three categories: luxury condominiums located within the city limits of Colombo, middle-income condominiums situated in the outskirts of Colombo and its suburbs, and low-budget apartments located outside the suburbs of

Colombo (Prathapasinghe et al., 2018). Pongprasert (2022) has classified condominium prices into six divisions based on the price per square meter. Classification is listed in the Table 1. However, this study encompasses all condominium types without being selective on a particular type. Condominiums can be categorized according to various themes such as with respect of ownerships, management structure, physical structure etc. (Swaminathan, 2023; Madsen et al., 2022).

Table 1: Classification of Condominium Prices

Type	Price per square meter
Entry Level	Below 2,018.45 USD
Mid-range	2,018.45 – 2,595.13 USD
Upper Mid-range	2,595.13 - 3,460.21 USD
High-End	3,460.21-5,767.01 USD
Luxury	5,767.01 - 8,650.49 USD
Super Luxury	Above 8,650.49 USD

Source: Pongprasert (2022)

In Singapore, there are two distinct categories of condos: Executive condominiums and Private condominiums (Jayasekara & Chandrathilake, 2022). This classification pertains to the aspect of ownership, wherein executive condominiums are established through a public-private collaboration, with the latter being fully owned by a private entity. According to Senaratne, et al. (2010) and Wijayamali, et al. (2014) condominiums may be classified into three categories based on the amount of facilities they possess: Super Luxury, Luxury, and Semi Luxury. Madsen, et al. (2022) have classified condominiums according to management structure. Thereby following three categories being identified, (1) Single management structure, (2) Two-tier management structure (3) Linked management structure. The owners' association is in charge of managing, maintaining, and operating the single building unit, which includes common properties. Therefore, in such a type of a mixed-use condominium building, just one management organization is formed. When a property has two buildings, a two-tier management system is required. The two-tier management structure will thereafter

include a master association that oversees the broad common property and two subsidiary associations that oversee each building and the restricted common property. In large lateral condominium development with different buildings, a linked management system is employed (Madsen et al., 2022).

In this paper, a novel approach is used to classify condominiums. Three categories have been identified considering the ownership, management structure, facilities, and physical structure to provide a full understanding of how the type of condominiums impacts the pricing along with the influencing factors. is the three categories are as follows; (1) Stand-Alone condominium apartments - condominium property in which all of the apartment units are housed in a single building unit and share a set of facilities that are also housed in the same building (2) Gated community condominium apartments, housed in several building units and have a controlled entrances for pedestrians, bicycles, and automobiles (3) Mix of commercial and gated community condominium apartments - apartments possess a set of single-family housing units with commercial units which has separate access to roadway and parking.

2.2 Factors influencing condominium development and purchases.

A comprehensive study is being done by Madhushani & Piyadasa (2019) to reveal the aspects that substantially affect consumer purchase decisions in relation to luxury apartments in Colombo and its surroundings. According to the study, the most influential aspect is the consumer's financial capability, although physical characteristics and locational relevance are equally important. The following categories have been ranked: Basic Amenities, Environment, Goodwill, Emotional, and Recreation. Abesinghe & Selvarajan (2018) attempted to lessen and address the difficulties that intentional developers get while building luxury apartments since they are unable to satisfy

client demands. According to the study, the following factors influence the preference for luxury condominium apartments in the Colombo Metropolitan Area: financial factor, environmental and condominium general element, security factor, and project team reputation. Furthermore, Ariyawansa & Udayanthika (2011) also examined the contributing factors for living in high-rise that has created a market for condominiums in the city of Colombo. The study considers price, rental value, income, demographic features, occupation and education background, migration, and consumer choice as demand variables.

In his study, Pongprasert (2022) examined three distinct categories of variables, specifically locational, structural, and neighbourhood characteristics, in order to identify the factors influencing luxury condominium prices. The researcher considered two locational variables, eight structural variables, and six neighbourhood variables. Soon et al. (2016) identified locational and neighbourhood characteristics, structural factors, and government regulations as possible determinants of price fluctuations in high-rise buildings. The parameters encompassed in this study are tenure, ownership of strata titles, built-up area, number of storeys, number of rooms, age of the building, GDP growth rate, and distances to neighbourhood elements.

When we summarize the above variables, we may identify 3 key groups. Specifically, structural or physical attributes, the developer's marketing approach and reputation, and the location of the condominium property. The location element may then be divided into three categories: city accessibility, environmental factors, and neighbourhood amenities (Taecharungroj, 2021). The accessibility factor relates to the distance between the condominium and the nearest central business districts (CBDs). This was identified to be quite essential and has a substantial impact on prices. However, even if locations are remote from CBDs, they may be easily accessible via the

expressway network. As a result, the distance to the nearest expressway interchange influences both the demand and the price of the condominium.

2.3 Amenities affecting condominium prices.

Numerous studies have been conducted to determine the influence of amenities on housing prices. In Shenzhen, China, educational institutions, health care facilities, natural amenities, commercial facilities, and job prospects are regarded possible factors for neighbourhood characteristics that impact housing values (Hu et al., 2019).

Hong, et al. (2020) discovered that the distance to the following neighbourhood facilities is crucial for South Korean house prices: national parks, high schools, redevelopment areas, universities, general hospitals, museums, and subway stations. All of these amenities are not directly tied to or focused on condominium pricing, but we may utilize them to establish links between them. In addition to the aforementioned factors, Tajima (2019) also takes into account the proximity to amenities such as rail stations and core business districts. In a study conducted by Tajima (2019), the author examined the influence of shared amenities on the resale value of condominiums. However, this study takes into account both the brand-new prices and the resale prices.

The distance to one's workplace, educational institutions, commercial establishments, healthcare facilities, and other significant variables contribute to the overall concept of connection and accessibility. This encompasses several aspects such as utilities, current and forthcoming projects, as well as social infrastructure. Furthermore, an essential aspect of enhancing quality of life is the infrastructure that supports it. This includes many elements such as transportation systems, neighboring communities, recreational facilities, water supply, sewage systems, electricity grids, phone

connectivity, waste management, and amenities. Within the realm of social infrastructure, features like playgrounds, sufficient parking spaces, and security measures play a crucial role in ensuring a satisfactory home life (Melesse, 2020). Hence, both the frequencies of amenities and the distance from the condominium to each specific amenity are crucial factors to consider. In their study assessing the drivers of condominium pricing, Soon et al. (2016) examined many neighbourhood factors, including the proximity to Setia City Mall, the blue mosque, Taman Botani Negara, MAZ International School, the nearest KTM station, and the Kesas highway. In his study, Pongprasert (2022) examined six distinct neighbourhood amenities, including the presence of a department store, school, hospital, recreational park, expressway, and main road. Scholar has exclusively focused on the proximity of amenities to condominiums as a determining factor for assessing the affordability of condominiums in Thailand.

The majority of the facilities discussed in a foreign setting are also applicable to Sri Lanka. The primary factors that have been taken into consideration are the presence of expressways, supermarkets, hospitals, and schools, which are likewise available within the Sri Lankan environment. However, the majority of the research has imposed a constraint on the number of dependent variables, often ranging from 15 to 20, which therefore limits the ability to account for potential multi-collinearity. In the study on the importance of neighbourhood amenities on condominium prices in Bangkok, Thailand, by Taecharungroj (2021), has shown that accessibility considerations and structural elements had a greater impact on condominium pricing compared to amenity factors. However, among the amenity variables, facilities like bars, motels, and restaurants were found to hold more significance. Further, this study revealed that 95 Amenities of google maps explained 52% of condominium price variance. There five types of amenities

have been studied: commercial, cultural, healthcare, natural, and services. In addition, the study stated that popularity of amenities could also drive condominium prices.

2.4 Estimating Housing prices using machine learning.

To estimate house prices, most research combined linear regression with a hedonic model (Paniagua-Molina et al., 2021; Pongprasert, 2022; Soon et al., 2016; Nilsson, 2015). It is assumed that real estate is a heterogeneous good whose price is determined by structure, neighbourhood, and locational qualities. The hedonic model, on the other hand, is simply put into linear regression form, which is restricted in its capacity to identify intricate and nonlinear correlations between housing prices and variables (Hu et al., 2019). As a result, alternative robust strategies for projecting housing prices are needed. Machine-learning algorithms fall under this category because they provide prospective advantages in three areas: (1) Capability to investigate multilevel interactions and nonlinear relationships (2) There are no strict assumptions requirements (3) Capability to process data in a variety of forms. In recent housing research, academics have proved the resilience of machine learning techniques. (Taecharungroj, 2021; Hu et al., 2019; Hong et al., 2020). Unfortunately, only a few researchers have used machine learning techniques in conjunction with hedonic modelling to forecast fine-scale housing or condominium prices in Sri Lanka

In Sri Lanka, there is very little literature on estimating housing prices using machine learning. Liyanaarachchi, et.al. (2021) developed a website for housing price forecast system, which included nine factors with property prices. The impact that nearby local amenities have on property prices is also taken into consideration very marginally. The machine learning methods employed in the system include linear regression, forest

regression, and boosted regression, and the results are fed into a neural network to compare the forecast and its accuracy with ground level findings. This system was designed to predict whether a customer would purchase a house or not. Accordingly, this study was conducted from the perspective of a buyer in order to aid in their decision-making process.

Hu, et al (2019) have used six different machine learning algorithms to determine the best algorithm for dynamics of housing prices and as a conclusion. In order to ascertain the optimal algorithm, the researchers have taken into account the root mean square error (RMSE), mean absolute error (MAE), and R-squared values associated with the various model algorithms. The findings suggest that the models achieved the following R-squared values: 0.74 for Random Forest regression, 0.73 for Extra trees regression, 0.69 for Gradient boosting regression, 0.40 for Support vector regression, 0.71 for multi-layer perceptron neural network, and 0.70 for k-nearest-neighbor algorithm. The results showed that except Support Vector Machine algorithm, all other algorithms undergone good performance. Out of them Random Forest Regression and Extra-Trees Regression are the best algorithm for dynamics of housing prices.

Hong, et al. (2020) used Random Forest to compare house pricing prediction with hedonic pricing model. In order to assess the precision of the author's findings, the author has employed three statistical measures: Mean Absolute Percentage Error (MAPE), Coefficient of Dispersion (COD), and R-squared. The R-squared value for the random forest technique is 0.97, whereas the R-squared value for the hedonic approach is 0.72. Results indicate that because the Random Forest technique better depicts the complexity or nonlinearity of actual housing markets, it may be a beneficial addition to hedonic models in the practice of bulk evaluation. To assess its use in the Malaysian real estate mass assessment context, J. McCluskey, et al. (2014) deployed Boosted

Regression Trees (BRT) to a heterogeneous data set of residential property taken from Kulai Municipality, in the State of Johor, Malaysia. The findings demonstrate that the BRT model outperforms the conventional models in terms of MAPE, COD and R-squared. The model achieved an R-squared value of 0.91, while other conventional models, such as Linear MRA and Semi-Log MRA, attained R-squared values of 0.74 and 0.78, respectively. Therefore, employing machine learning to predict house prices produces more accurate results.

In their study, Abidoye and Chan (2018) conducted a comparison between conventional and current approaches in order to enhance the accuracy of property assessment. Consequently, the researchers employed the hedonic pricing model (HPM) and artificial neural network (ANN) as methodologies for their investigation. The evaluation of the models under consideration involved the assessment of many metrics, including R-squared, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE).

The findings indicate that the artificial neural network (ANN) exhibits higher accuracy compared to the high-performance model (HPM), as evidenced by the respective R-squared values of 0.81 and 0.77 obtained for ANN and HPM. To determine the association between 95 neighbourhood amenities and condominium prices, Taecharungroj (2021) employed random forest and extreme gradient boosting regression. In his investigation, the researcher used R-squared, RMSE, and MAE as evaluation metrics for the models. The Extreme Gradient Boosting model achieved an R-squared value of 0.73 across all models, whereas the Random Forest model achieved an R-squared value of 0.66 across all models. This study is most pertinent to this paper since it used a comparable

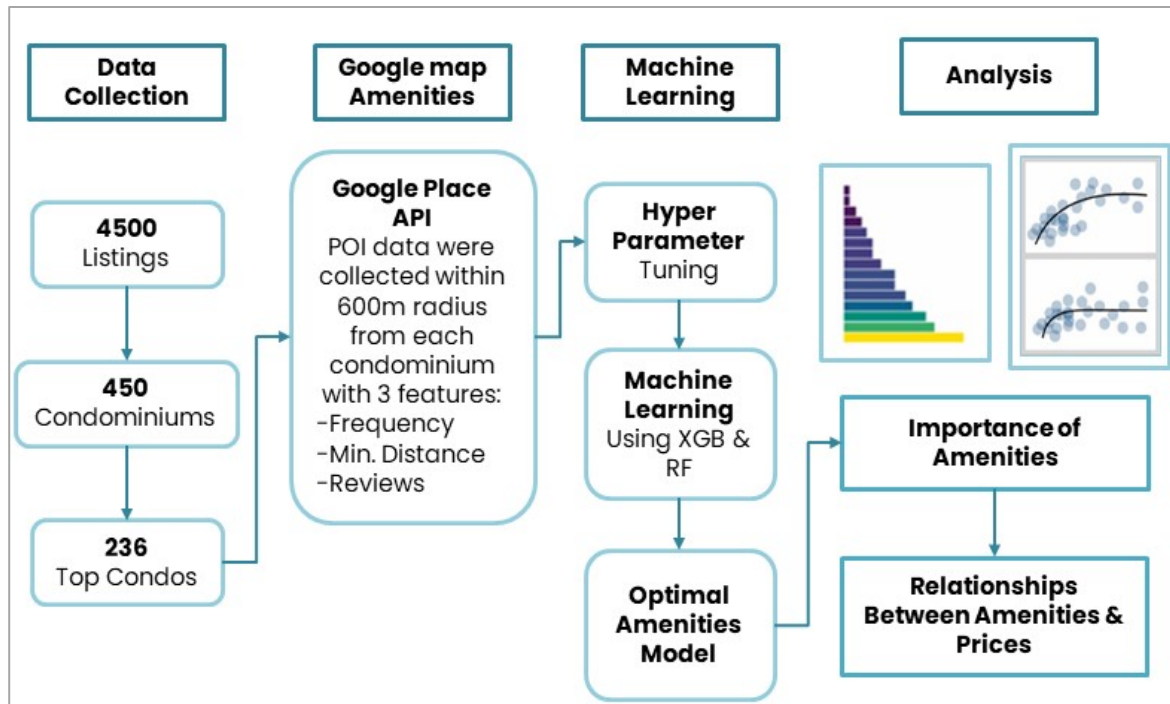


Figure 2: Overall Framework
Source: Author's development

methodology to determine how amenities affect condominium pricing.

In past scholarly works, it has been observed that Random Forest and Boosting models exhibited superior performance. Additionally, the evaluation of models in literature often relies on metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared values. Utilizing the aforementioned methods and assessment criteria would be most advantageous for the present investigation.

3. Methodology

3.1. Overall methodological framework

The overall methodological framework followed four basic processes. First, pricing for condominium apartments were gathered and filtered from electronic platforms. Second, data on amenities was acquired utilizing the Google Maps platform. As the third step, trained two machine learning algorithms to determine the association between apartment pricing and amenities. Finally, the relative importance of the various categories was compared, and

various relationships among them have been classified.

3.2. Data collection and process

As the initial step of the research data related to condominium apartment prices were collected from websites of the apartment developers, advertisements, social media posts, as well as by making direct calls. Prices of around 4500 condominium apartments were collected in January 2023, covering the area within the tentative boundary given in figure 2. Web scraping technique was widely used to scrape the listed condominium prices. Beautiful Soup Library in Python was used to extract data. A unique python code for each website was created to extract the data since each electronic platform was different and unique. The price, location, number of bedrooms, area of the apartment, coordinates data was extracted related to the condominium developments,. The data was extracted from Lanka Property Web (<https://www.lankapropertyweb.com>), Ikman.lk (<https://ikman.lk>), Selling.lk (<http://www.selling.lk>) etc. The collection of data through internet platforms may not provide a comprehensive representation of the whole condominium market. However,

a significant amount of data could be gathered to ensure that a wide range of research locations would be included.

The process of data gathering mainly relied on the technique of web scraping. This approach has several advantages, such as the ability to gather a large volume of condominium pricing from various platforms. However, it is essential to carefully analyze the constraints associated with this method. Due to the presence of adverts on paid websites, there is a possibility of seeing duplicate listings for a certain property inside a given locality. Certain advertising exhibited a deficiency in providing crucial information such as pricing details, geographical location, and the specific name of the condominium. The data has undergone manual filtration in order to enhance data accuracy. The coordinates of the apartments are not freely available in the electronic platform. Therefore, the coordinates are manually gathered when filtering the data.

Afterward, 4500 listings were condensed into separate condominium units (450 condominiums). Then the study was done on those 236 top condominiums. Each condo's price was calculated as the "10th percentile" of all listings' prices per square foot. This method used to remove the sellers' posted outliers and mistakes (eg: Taecharungroj, 2021). Then by using Google Maps Platform, data about amenities were collected in the form of Point Of Interest (POI). The utilization of this particular approach was subject to several restrictions, including the possibility that the services being assessed might not be currently available or obsolete in nature.

Hence, it is important to take into account this factor throughout the discussion of outcomes. The Places Application Programming Interface (API) was very useful in this regard. For this "Place Search Nearby" from places API was obtained. This function was not open source and fee levied according to data retrieval. The data

was collected within a particular radius from each condominium.

According to the literature, a suitable walking distance from a place should be between 300 and 600 meters, and it should also be sufficient to capture built environment characteristics (Taecharungroj, 2021). Thereby, POIs within 600 meters from a luxury condominium apartment collected with 3 important features. They were frequency of each amenity in the neighbourhood, distance between the amenity and the condominium, and number of public reviews received for the particular amenity. 54 POI data types were taken by nearby search and distances to each amenity were calculated from 'geopy package' in python.

3.3. Machine learning algorithms

Machine learning tools have been taken to derive the relationship between condominium prices and amenities. The machine learning tools used were Random Forest (RF) and extreme gradient boosting regression (XGB).

The literature typically identified Random Forest (RF) and Boosting Regressions (XGB) as highly effective and often employed machine learning algorithms (Taecharungroj, 2021; Hu et al., 2019; Hong et al., 2020; J. McCluskey et al., 2014). RF is a reliable and efficient tree-based bagging technique. It combines several decision trees into a forest and makes a prediction based on a vote by the majority. Without the need for precise specification, RF can effectively analyze a big dataset with several input variables, handle missing and categorical variables, handle unequal distribution, and find interactions and non-linear connections (Taecharungroj, 2021).

XGB was the other algorithm employed. This is a member of the boosted tree model family. Contrary to RF, which independently and arbitrarily generates trees, boosting approaches generate trees based on the residuals of the prior tree. Trees are constructed until the residuals are

low or the maximum number of trees is achieved, at which point a forecast is issued. It was discovered that boosted tree models were quicker and more accurate than RF (Taecharungroj, 2021).

Before running the models, the hyperparameters were tuned using the sci-kit learn package's "RandomizedSearchCV" function in Python. Each round examined hundreds of random hyperparameter combinations. To avoid model overfitting, a five-fold cross validation was done, and the process was repeated ten times.

The optimal hyperparameters of the XGB model were 0.7 colsample_bytree, 0.1 learning_rate, 0.1 gamma, 8 max_depth, 7 min_child_weight, 100 n_estimators and 1 subsample. These hyperparameters were used in all subsequent models for consistency. The accuracy level of models was tested using root mean square error (RMSE), R-squared (R²) value and mean absolute error (MAE).

3.4. Data Analysis

Table 2: Variables

Dependent Variables	Independent Variables	Control Variables
Condominium Price	Frequency of Amenities	Structural Factors <ul style="list-style-type: none"> Median Unit Size Modal Number of Bedrooms Built Year Number of Floors Number of Units Type
	Minimum Distance of Amenities	Accessibility Factors <ul style="list-style-type: none"> Distance to nearest Expressway interchange Distance to two main CBDs
	Reviews of Amenities	

Source: Author's development

For the analysis of the data, seven models including the following dependent, independent and control variables were formed.

The seven models were as follows: (1) amenities, (2) structural, (3) accessibility, (4) structural and accessibility, (5) structural and amenity, (6) accessibility and amenity and (7) all models. Thereby, the

Table 3: Amenities

Category	Amenities
Commercial	Supermarkets, Convenient store, Department store, Book Store, Home good store, Hardware store, Electronic Store, Meal delivery, Meal takeaway, Shopping Malls, Clothing Store, Restaurants, Bakery, Lodge, Movie Theatres
Cultural	Museum, Art Gallery, Bar, Liquor store, Gym, Library, Amusement Park, Aquarium, Casino, Zoo, Stadium, Night Club, Mosque, Church, Hindu Temple, Cemetery
Healthcare	Hospitals, Pharmacy, Drugstore, Physiotherapy, Haircare, Beauty Salon, Dentist, Doctor, Spa
Natural	Park, Natural feature
Educational	Universities, Schools
Services	Train Station, Bus Station, Transit station, ATM, Bank, Laundry, Police, Post office, Gas station
Job opportunities	Government offices

Source: Author's development

correlation of different factors was considered to see which factors are highly inter-related.

But the focus was on the first model which is Amenities vs Condominium Price. If there was a considerable variance with amenities, then each amenity factor was considered, and individual relationships to be developed. 54 amenities were considered under 7 categories for the development of relationships.

To determine the appropriate number of factors, the relative importance of the model's 162 amenity features was first computed. The least significant components were then eliminated to produce a new model. To identify the best amenity model, models were constructed and compared using RMSE, R², and MAE. The same procedure was followed for all models, including structural and accessibility variables. Following that, the relative importance of each feature on the condominium price was disclosed and ranked. Following that, the link between each important amenity and the expected pricing was displayed in order to see the similarities and differences. The primary benefit of Machine Learning (ML) is the discovery of non-linear correlations; hence, the partial dependence (PD) of each amenity element was determined in Python

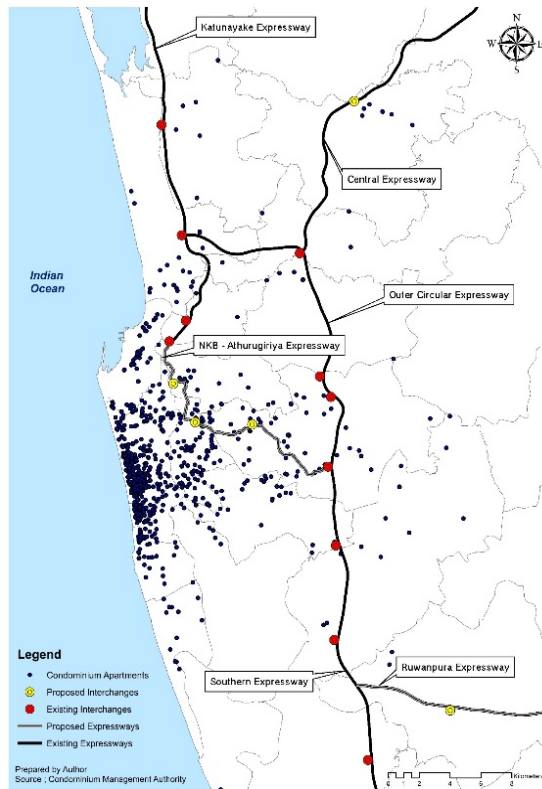


Figure 3: Condominium spread.

Source: Author

using the 'pdpbox' package. After accounting for the influence of other factors, PD describes the effect of a single factor on condominium prices.

3.5. The area under study

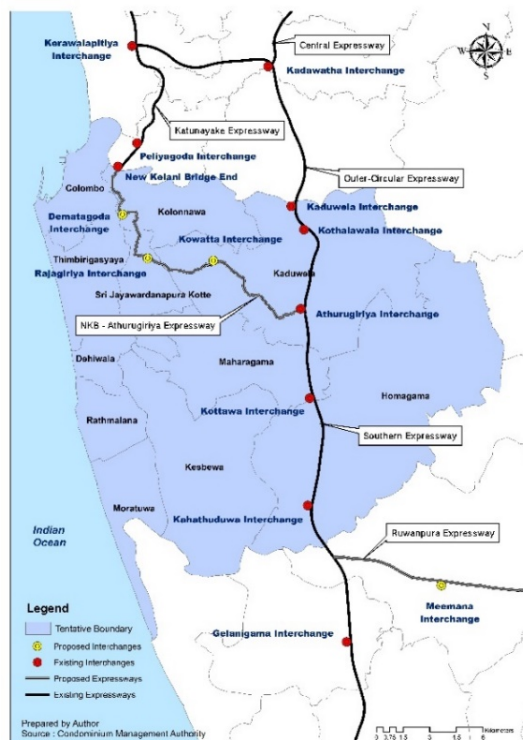


Figure 4: Study boundary

Source: Author's development

Over 90% of condominiums and existing expressways were currently located around the Colombo and suburban areas which includes following Divisional Secretariat Divisions; Colombo, Ratmalana, Maharagama, Homagama, Thimbirigasyaya, Moratuwa, Kaduwela, Kotte, Dehiwala, Kesbewa and Kolonnawa

4. Results

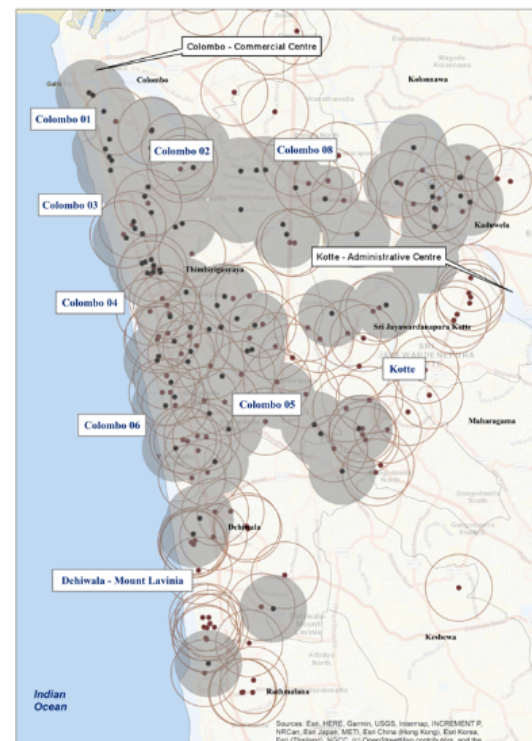
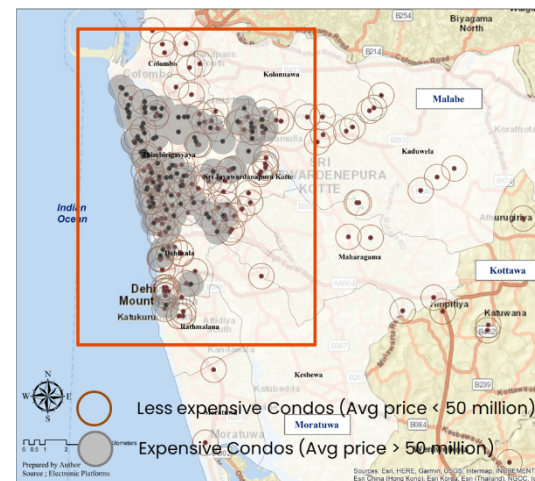


Figure 5: 236 condominium locations and neighbourhood

Source: Author's development

Figure 5 depicts the locations of the 236 condominiums included in the study as well as their adjacent areas within a 600-meter buffer. Most condominiums were concentrated in the two core business areas, Colombo, and Kotte.

Some condominiums were located in the eastern portion of Colombo, near Kottawa and Malabe, away from the main cluster. A few condominiums could be found in the southern section of Colombo, closer to Moratuwa. More costly condominiums (average price > Rs. 50 million) were marked in grey colored area, whereas almost all condominiums in Colombo's outskirts were less expensive (average price = Rs. 50 million).

4.1. Extreme Gradient Boost (XGB) and Random Forest (RF) results

Seven models were constructed utilizing the XGB and RF algorithms, with all seven models acceptable from the XGB and RF algorithms, with R² values ranging from 0.48 - 0.83. The lowest performing model for both XGB and RF was structural model, with a root mean square error (RMSE) of 9,185 and 11,597 respectively. When training the model, the lower performing models exhibit overfitting or underfitting of the data. As a result, additional data about condominiums was required to build the model to a highly acceptable level.

The amenities model (Model 1) surpassed the structural model (Model 2) and accessibility model (Model 3). When all

factors were considered, the Model 7 performed better than all other models except for the structural and accessibility model (Model 4). This was because more variables produced good correlation and increased model performance and predictability.

Model 1 is the focus of the research, and its performance is superior to that of other models, producing useful results. Model 1 was developed to comprehend the importance of neighbourhood amenities in the absence of other factors. The R² value for both XGB and RF was 0.76, meaning that Model 1 explained 76% of the variation in condominium prices. The results also demonstrated that the XGB and RF models perform similarly in all seven models. XGB is regarded as a strong regression model that produces more accurate findings; hence, XGB models and outcomes were used in subsequent investigations.

Models 1 and 7, which had 163 and 172 factors, respectively, were fine-tuned further to identify the optimal factor number. To construct a new model, the least significant factors were deleted at a time. To tune the factors, the XGB algorithm was run ten times for each attempt at deleting factors and obtaining the best R². The best models with the lowest R² were selected. As an outcome, the optimal amenities model had 56 factors (RMSE = 9,023, R² = 0.78, and MAE = 6,787), whereas the optimal all models had 20 factors (RMSE = 7,573, R² = 0.83, MAE =

Table 4: Model Details

No	Model	Extreme Gradient Boosting			Random Forest		
		RMSE	R ²	MAE	RMSE	R ²	MAE
1	54 (162 factors) Amenities (AME)	8,715	0.76	6,102	8,966	0.76	6,384
2	Structural (STR)	9,185	0.58	7,657	11,597	0.48	8,710
3	Accessibility (ACC)	8,669	0.76	6,451	10,108	0.71	6,509
4	STR + ACC	7,505	0.83	5,434	8,327	0.68	6,621
5	STR + AME	10,001	0.74	6,963	9,195	0.75	6,639
6	ACC + AME	10,636	0.67	6,943	9,881	0.70	6,766
7	All	9,510	0.76	7,465	8,766	0.77	5,396

6,160). Both models also had the highest R^2 and outperformed the original models.

4.2. Relative importance of factors

Figure 6 depicts the normalized XGB significance value for each of the 20 model factors. Number of Floors of the condominium had the highest relative importance (22.75%), followed by frequency of art gallery (12.43%). The factors were divided under the study's three main categories: structural, amenities, and accessibility. However, only one accessibility element was filtered throughout the tuning phase and make it into the final 20.

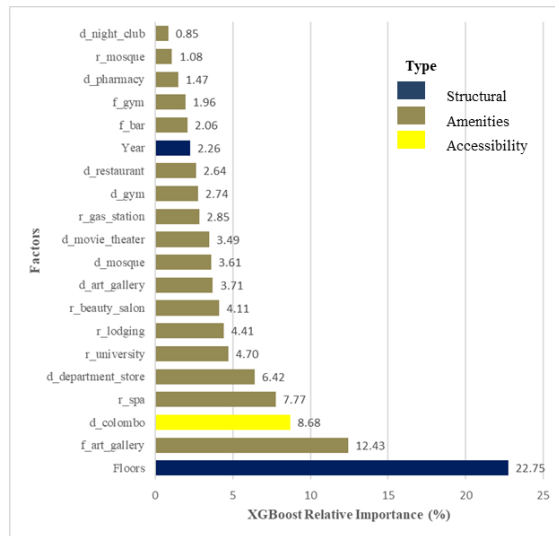


Figure 6: Normalized relative importance of optimum all models.

Source: Author's development

The 56 amenities factors in the optimal amenities model were used to determine the relevance of amenities in relation to condominium apartment prices. Although the optimal all model de-scribed up to 74% of the variance in condominium pricing, the optimal amenities model explained up to 78%, shedding insight on the previously unknown effects of amenities.

Figure 7 illustrates the normalized importance level rankings of the 56 amenity factors. The most significant amenity factor was user reviews of lodgings (9.95%), followed by the minimum distance to a movie theatre (7.2%). There are 28 factors with normalized importance greater than 1%.

Some amenities were included with several features, such as hotels (user reviews and minimum distance), art galleries (frequency and minimum distance), hospitals (user reviews and minimum distance), and so on. (Figure 6) also categorized amenities based on the literature review as commercial, cultural, healthcare, educational, natural and services amenities.

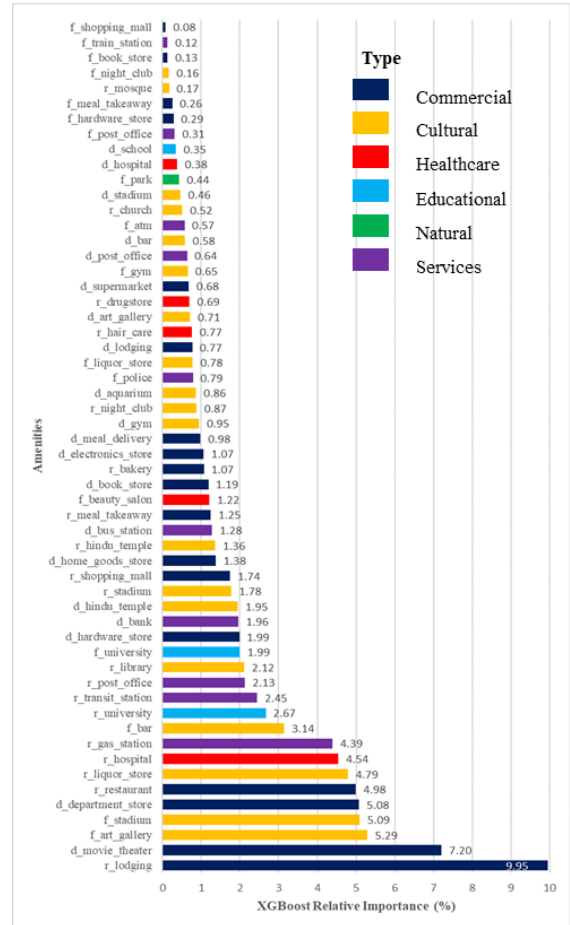


Figure 7: Normalized relative importance of optimum amenities model.

Source: Author's development

4.3. Relationship among structural factors, accessibility factors and condominium prices

2D Partial dependence (PD) was used to fit and show the relationships between structural and accessibility factors with predicted condominium prices (in LKR per sq.ft).

Figure 8 depicts 2D partial dependence curves between median apartment size with distance to Colombo. The size - distance to

Colombo graph demonstrates that the effect of size is insignificant when compared to the distance component, since expected price margins are parallel with distance.

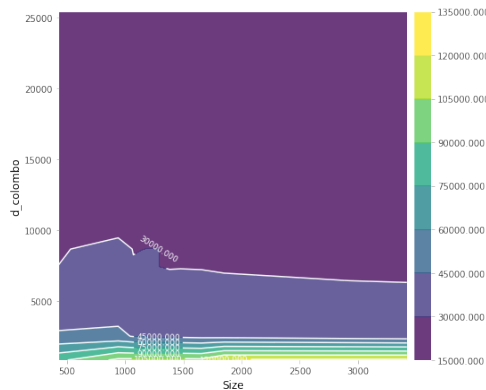


Figure 8: PD curve – Size (sq.ft) vs Distance to Colombo (m)
Source: Author's development

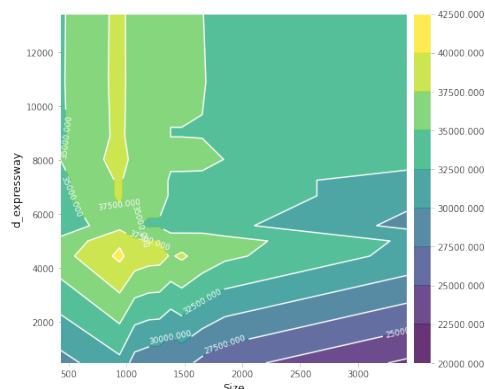


Figure 9: PD curve – Size (sq.ft.) vs Distance to Expressway (m)
Source: Author's development

Figure 9 examines median apartment size vs distance to expressway; both had a significant effect on cost. The highest price per square foot, or the sweet spot, was found around 4.5 kilometers from expressways, with a condominium size of approximately 900 sq.ft

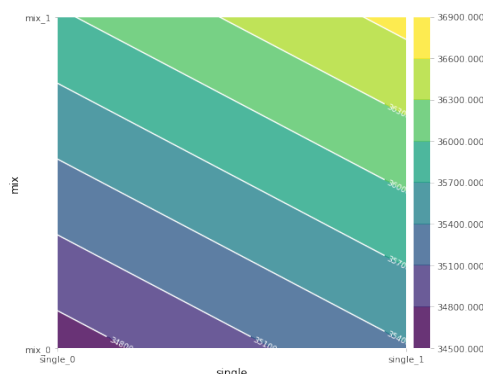


Figure 10: PD curve – stand-alone apartment vs mixed-use apartment
Source. Author's development

Three different types of condos were the subject of pricing and type analyses. If Apartments that were neither single (stand-alone) nor mixed fall under the gated category. Thus, prices have been low when an apartment is part of a gated community and are estimated to be roughly 34,800.00 LKR per square foot. Nevertheless, the price rises to 37,000 LKR when it displays mixed characteristics. As a result, developers should take that fact more seriously as a condominium with commercial activities like restaurants, cafés, pubs, banks, etc. would increase the value per square foot.

4.4. Relationship between amenity factors and condominium prices.

PD curves are used to visually show the relationships between the 56 amenity parameters. There are five sorts of relationships: restricted positive, accelerated positive, crooked, humped, and negative.

Out of 56 amenity parameters nine amenity aspects are categorized as restricted positive. It displays amenities on the X-axis and projected condominium values on the Y-axis. All of these curves are positive and have the same form in that beyond a certain degree of rise, subsequent increments have no effect on pricing (Figure 11). However, up until that point, prices steadily rise ($r_{\text{restaurant}}$ & d_{gym}) or sharply rise ($d_{\text{bus_station}}$).

According to the categorization, when there is just one liquor shop nearby, the prices of those condominiums are comparatively higher than those without. However, if the neighbourhood has more than one liquor store, the cost would not alter. As a result, at least one liquor store needed to be located within 600m boundary to enhance the value of condominiums. But liquor stores which are in supermarkets are not considered in the analysis. However, in the case of restaurants, prices may alter up to one level of user reviews (about 90 reviews), but they will not raise again until another level (approximately 150 reviews) is reached. However, they then rise and

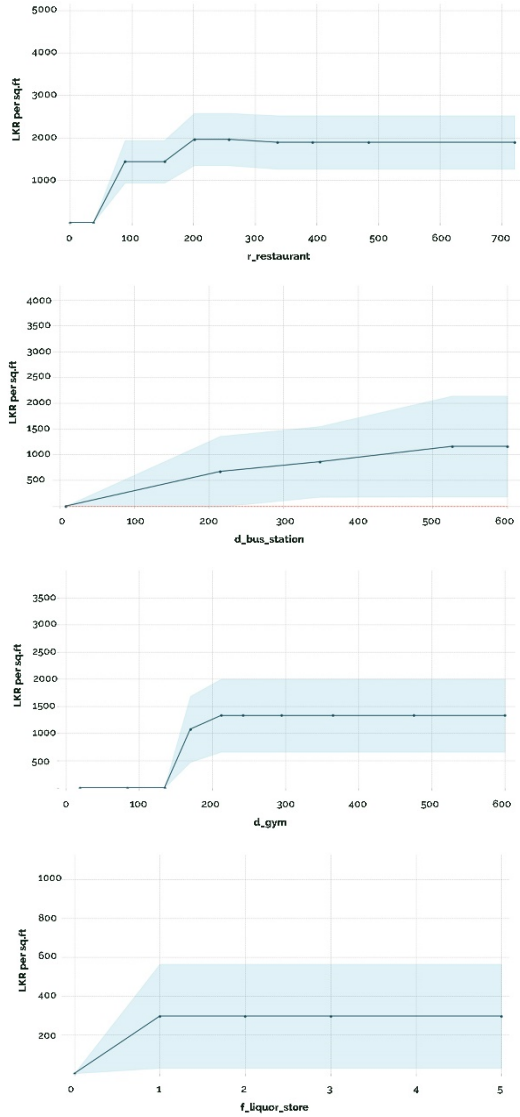


Figure 11: PD curves of restricted positive relationships

Source: Author's development

stops at the peak (about 200 reviews), after which prices remain unchanged.

As a result, to raise the price of a condominium, there should be a restaurant with user reviews ranging from 0 to 90 or 150 to 200. Otherwise, the prices will remain unchanged. Other amenities' results demonstrate a similar variation in the restricted positive category.

Out of 56 amenity factors, ten amenity factors have been classified as accelerated positive, however the distance to the nearest amenity is not one of them. A common feature of this type was a modest relationship between amenity and predicted pricing at first, but it escalated after it

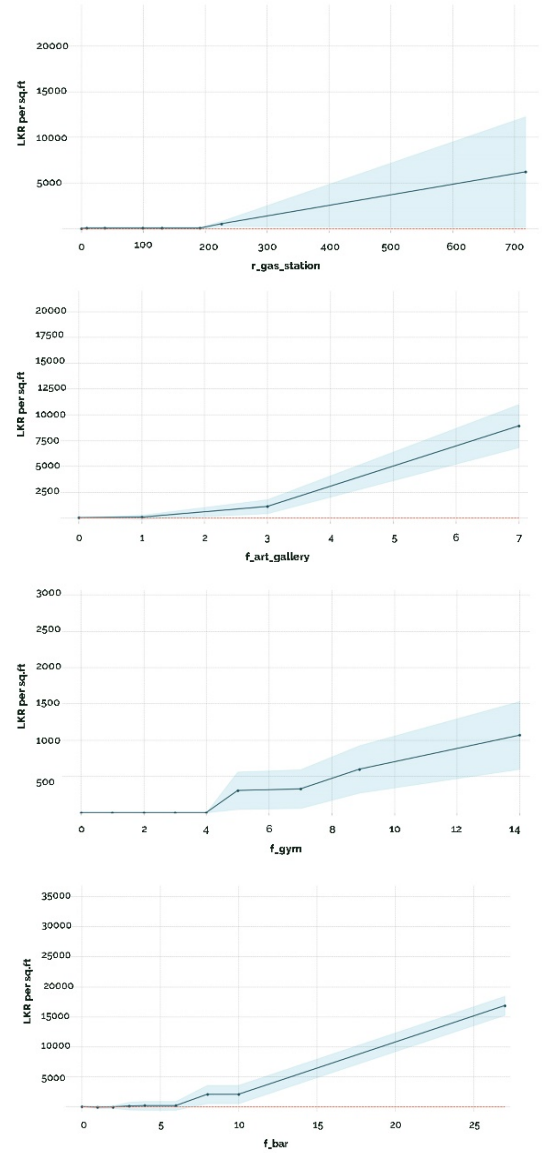


Figure 12: PD curves of accelerated positive relationships

Source: Author's development

reached a certain level. Figure 12 demonstrates the shapes of four amenity factors of accelerated positive relationship.

As the term implies, the 10 amenity factors out of 56 classified as humped relationships had a positive impact on predicted prices until they peaked (Figure 13).

The relationship declined after a certain point of value. For example, when there is a bar located approximately 75m, the predicted price peak is Rs. 500 per sq.ft, and when there is a Hindu temple located around 425m, the predicted price peak is

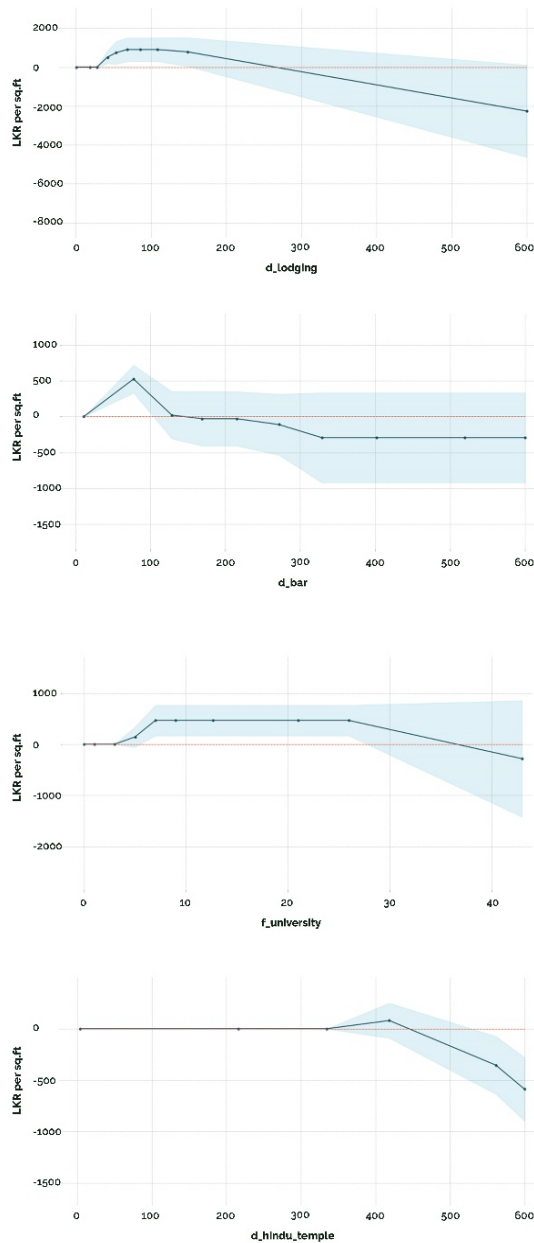


Figure 13: PD curves of humped relationships
Source: Author

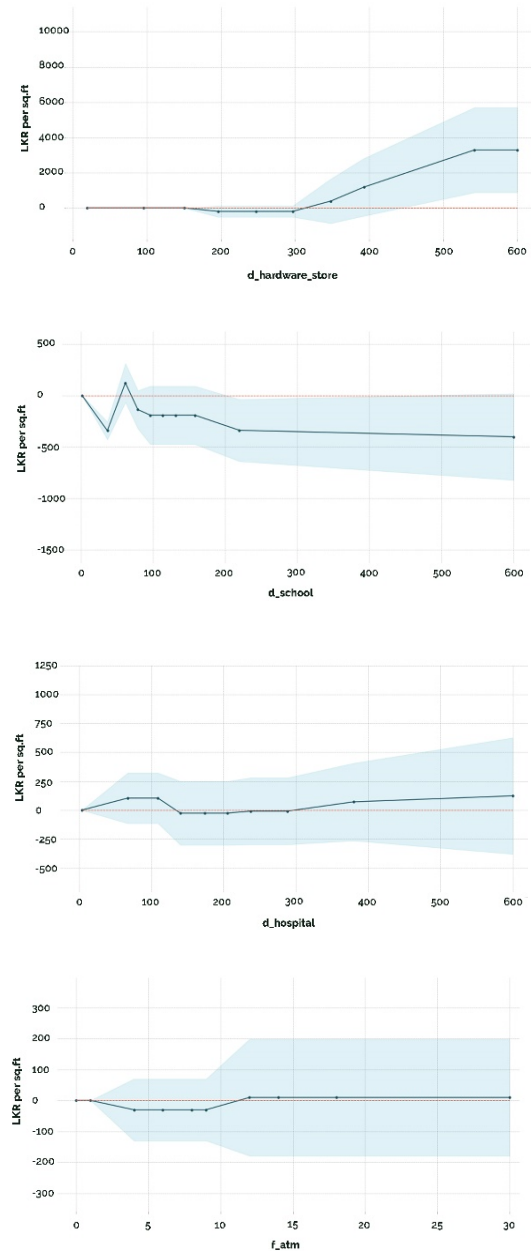


Figure 14: PD curves of crooked relationships
Source: Author

Rs. 100 per sq.ft. The predicted prices then fell below these levels. The crooked category features a distinct 'S' curve that goes from negative to positive or positive to negative. If we consider the minimum distance to the hospital, the cost increases when the hospital is within 100 meters, but it decreases when the hospital is between 150 and 300 meters. Finally, prices rise and stabilize at a steady level of about 375 m. Figure 14 depicts further crooked type variations.

Finally, out of 56 amenity factors, eight amenity factors correlated negatively with predicted condominium prices (Figure 15). These kind of amenity factors viewed as “disamenities” by Taecharungroj (2021) which had adverse implications on condominium prices.

The partial dependence curve demonstrates that anticipated prices fall as the number of beauty salons or popular bakeries in the neighbourhood grows.

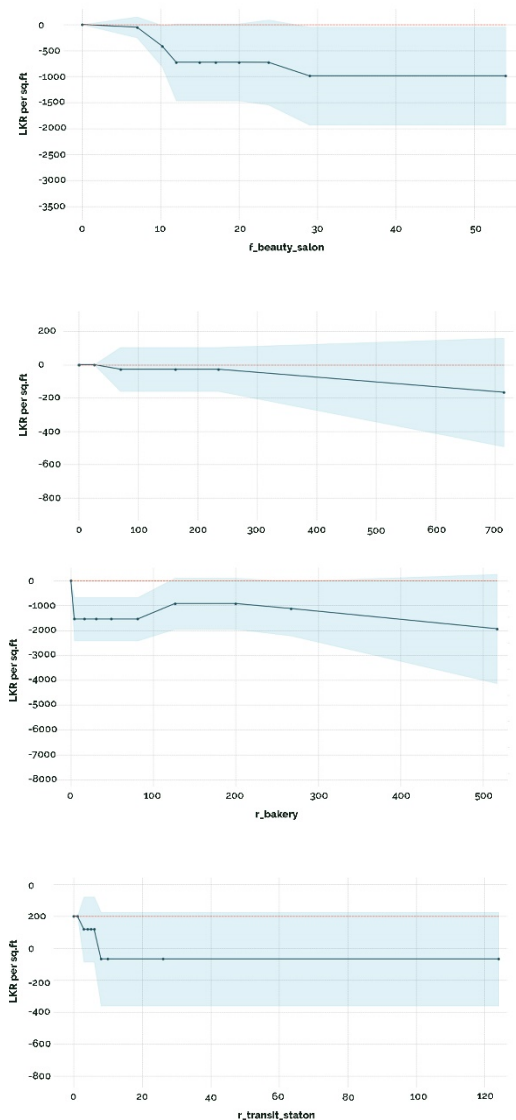


Figure 15: PD curves of negative relationships
Source: Author's development

5. Discussion

As the first instance to the use data from Google Maps to analyze and anticipate condominium prices in the Sri Lankan context, this study brought in new insights on the relevance of neighbourhood amenities on condominium prices in Colombo, Sri Lanka. A total of 162 elements from 56 different categories of facilities were examined. The models also included two structural and three accessibility characteristics as control variables. Instead of typical hedonic modelling, two ML algorithms were utilized to evaluate the high number of parameters. In most simulations, the results revealed that XGB beat RF by a little

margin. The findings revealed that 56 Google Maps amenities explained 76% of the price variance, supporting the current research indicating links between amenities and housing price/demand (Taecharungroj, 2021; Hu et al., 2019).

The model was optimized using 56 amenity parameters that demonstrated the highest performance to further study the amenities. Each amenity factor's significance was determined. Commercial (lodging and movie theatres) and cultural amenities (art galleries and stadiums) were voted first and second, respectively. According to Hidalgo et al. (2020), cultural and commercial amenities are central nodes — that attracted people and other facilities — inside networks of urban amenities or the so-called "amenity space". According to this study, they also have an impact on house prices and demand.

These findings also supported the findings of a recent study by Liu et al. (2018), who suggested that entertainment and sports facilities such as nightclubs, stadiums, and art galleries were highly key appealing elements for properties in city centers. The present study indicated hospitals to be the top ten most significant amenities, which also supports the preceding study by Hu et al (2019). PD was utilized to depict non-linear correlations between amenity parameters and forecasted condo prices as restricted positive, accelerated positive, crooked, humped, and negative relationships (Taecharungroj, 2021).

6. Conclusion

Previous research on neighborhood amenities relied on factors such as frequency, density, and distance. However, the innovative study incorporates reviews as a component to indicate the popularity of POIs and investigates the popularity of amenities on housing prices. Taecharungroj (2021) employs user reviews for the first time in his housing study, and this research continues and confirms his work. The results revealed that reviews of numerous amenities, such as lodgings, restaurants,

hospitals, universities, shopping malls, and so on, were significant drivers of condominium costs.

This study found that individual amenity criteria have an influence on anticipated pricing, it is crucial to increase various amenities in a neighborhood. Individual amenity elements might possibly boost anticipated costs between 1000 and 20000 rupees at their ideal values, according to the PD curves. As a result, in order to raise prices further, many neighborhood features must be improved.

A link between public and private organizations is essential for more sustainable urban development. This study examined the impact of public and private urban amenities on condominium pricing and suggest that the development of some facilities are important for the planning of residential condominium development. However, in Sri Lanka, these private and public organizations responsible for urban amenity development are not integrated and their actions are also not synchronized.

A common goal of an integrated an urban plan is to improve urban amenities in desperately needed areas of the city and catalyze residential demand, which, in turn, will foster neighborhood regeneration and deliver a more sustainable city over the long term. For that, constant communication and collaboration between urban amenity providers and housing providers are crucial.

By using innovative urban interventions such as tactical urbanism, Colombo may also temporarily improve its neighborhoods. So, as a practical implication of this study, identification of important amenities that Colombo lack and introduce community-led initiatives such as pop-up bars and trendy restaurants, temporary market stalls or parklets. Then, the community may then assess the results

and decide whether to make these proposals permanent features.

This study had various limitations despite its significant contributions. The condominium price listings on the website may not be actual prices in market due to advertised values may differ from selling prices which has been utilized in the models. The prices are collected from online websites and prices may change over time in the condominium market. Therefore, the analysis is highly depending to the data collection time period. Comparing e-platforms of condominium market in Sri Lanka to foreign e-platforms, data collection through web scraping is not as simple as it is. Local websites lack accurate unit names and precise location information in the advertisements but foreign e-platform displayed those advertisements in detail. Thereby web scraping technique for local websites is time consuming than usual since some data needed to collect and clean manually.

Massive amounts of data could be rapidly collected with the support of Google API, but only up to 60 POIs of each category could be collected. Places nearby search is no longer an open source due to a change in Google Maps' policies. Additionally, the 54 categories of POIs chosen did not include non-customer-facing POIs like factories or path-like infrastructures like rivers and roadways. The classified relationships of amenity factors are not validated with ground level testing and merely depending on the data collected through POI and analysis done by machine learning algorithms. This study suggests new directions for future research, such as establishing a model to predict the pricing of neighborhood amenities with the highest relative importance. The ability of XGB to regularly outperform other ML algorithms in price prediction should be investigated further, as should alternative practical ways of defining non-linear connections.

7. References

- Abesinghe, A., & Selvarajan, P. (2018). A study on factors affecting the preference for luxury condominium apartments in Colombo metropolitan Area, Sri Lanka. *International Journal of Advanced Research in Management, Architecture, Technology and Engineering*, 4(6), 1–7.
- Abidoye, R. B., & Chan, A. P. C. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, 24(1), 71–83. <https://doi.org/10.1080/14445921.2018.1436306>
- Ariyawansa, R., & Udayanthika, A. (2012). Living in high-rise: An analysis of demand for condominium properties in Colombo. *International Journal of Sociology and Anthropology*, 4(1), 31–37. <https://doi.org/10.5897/IJSA11.104>
- Condominium Management Authority. (2023). Condominium management authority. Retrieved from Condominium Plan. <http://www.condominium.lk/en/condominium-plan.html>
- Hong, J., Choi, H., & Kim, W. (2020). A house price valuation based on the random forest approach: The mass appraisal of residential property in South Korea. *International Journal of Strategic Property Management*, 24(3), 140–152. <https://doi.org/10.3846/ijspm.2020.11544>
- Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: an integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, 82, 657–673. <https://doi.org/10.1016/j.landusepol.2018.12.030>
- Jayasekara, S., & Chandrathilake, M. A. N. (2022). Protecting the rights of Purchasers in Condominium Properties in Sri Lanka: A Comparative Analysis with United States of America. *Sri Lankan Journal of Real Estate*, 19(1), 30–60. <https://doi.org/10.31357/sljr.v19i1.6023.g4456>
- Karsten, L. (2007) Housing as a Way of Life: Towards an Understanding of Middle-Class Families' Preference for an Urban Residential Location, *Housing Studies*, 22:1, 83-98, DOI: 10.1080/02673030601024630
- Larwiang, A. (2014). *Factors affecting the purchasing decision of high-end condominium in Sathorn Area, Bangkok* ([Master's Thesis]. Stamford International University).
- Liyanaarachchi, L., Wijethunga, I., & Madushanka, M. (2021). Housing price prediction using machine learning 14th International Research Conference (pp. 526–532). Kyungdong University. General Sir John Kotelawala Defence University.
- Madsen, M. D., Paasch, J. M., & Sørensen, E. M. (2022). The many faces of condominiums and various management structures – The Danish case. *Land Use Policy*, 120, 106273. <https://doi.org/10.1016/j.landusepol.2022.106273>
- Madushani, K., & Piyadasa, R. (2019). An empirical investigation on factors influencing to consumers' purchasing decision towards the luxury apartments in Colombo and suburbs, Sri Lanka. In *Proceedings of 12th International Conference of Faculty of Architecture Research Unit* (pp. 125–133). Colombo: University of Moratuwa. Sri Lanka.
- McCluskey, J., Zulkarnain Daud, D., & Kamarudin, N. (2014). Boosted regression trees: An application for the mass appraisal of residential property in Malaysia. *Journal of Financial Management of Property and Construction*, 19(2), 152–167. <https://doi.org/10.1108/JFMPC-06-2013-0022>
- Melesse, H. (2020). Factors determining the continuous price appreciation of condominium units in Addis Ababa. *World Journal of Civil Engineering and Construction Technology*, 4(1), 64–78.
- Nilsson, P. (2015). The influence of urban and natural amenities on second home prices. *Journal of Housing and the Built Environment*, 30(3), 427–450. <https://doi.org/10.1007/s10901-014-9421-6>
- Paniagua-Molina, J., Solórzano-Thompson, J., González-Blanco, C., & Barboza-Navarro, D. (2021). Hedonic price for amenities in rural and urban residential condominiums in Costa Rica. *Real Estate Management and Valuation*, 29(3), 52–64. <https://doi.org/10.2478/remav-2021-0021>
- Pongprasert, P. (2022). Determinants of luxurious condominium prices in Bangkok CBD: A case study of encouraging housing affordability in Bangkok, Thailand. *International Review for Spatial Planning and Sustainable Development*, 10(1), 167–182. https://doi.org/10.14246/irspsd.10.1_167
- Prathapasinghe, D., Perera, M. P. R. I., & Ariyawansa, R. G. (2018). Role of technology in providing better basic facilities for condominium properties. In *Proceedings of the 2nd International Conference on Real Estate Management and Valuation* (pp. 418–426). Colombo, Sri Lanka.
- Rajapakshe, R., & Jayaweera, N. (2019). *Condominium development in Sri Lanka: Current status and recommendations* (2nd ed). Department of Architecture, University of Moratuwa.
- Salama, A.M., Wiedmann, F., & Ibrahim, H. G., (2017) Lifestyle trends and housing typologies in emerging multicultural cities, *Journal of Architecture and Urbanism*, 41:4, 316-327, <https://doi.org/10.3846/20297955.2017.1415773>
- Senaratne, S., Zainudeen, N., & Weddikkara, C. (2006). Factors affecting condominium development in Sri Lanka. *Built-Environment Sri Lanka*, 7(1), 23–28. <https://doi.org/10.4038/besl.v7i1.1948>
- Soon, L. T., Tang, S. J., Kam, K. J., & Tik, B. (2016). Determinants of condominium prices in Shah Alam. *Journal of Valuation and Property Services*, 16.

- Swaminathan, J. M. (2023). Julius and creasy. Retrieved from <http://www.juliusandcreasy.com/publications>
- Taecharungroj, V. (2021). Google Maps amenities and condominium prices: Investigating the effects and relationships using machine learning. *Habitat International*, 118, 102463.
<https://doi.org/10.1016/j.habitatint.2021.102463>
- Tajima, K. (2019). Shared amenities' impacts on condominium resale values. *International Journal of Housing Markets and Analysis*, 13(2), 281–297. <https://doi.org/10.1108/IJHMA-03-2019-0038>
- Wijayamali, D. M. D., Waidyasekara, K. G. A. S., & Dahanayake, K. W. D. K. C. (2014). Living in low-income condominiums: End users' perspectives. In *Proceedings of the 3rd World Construction Symposium 2014: Sustainability and Development in Built Environment*. 356–364. Colombo, Sri Lanka.
- Wong, K. M. G. (2004). Vertical cities as a solution for land scarcity: The tallest public housing development in Singapore. *Urban Design International*, 9(1), 17-30.
<https://doi.org/10.1057/palgrave.udi.9000108>