José Rizal University

A MULTIPLE REGRESSION ANALYSIS IN PREDICTION OF DOT PROPERTY PHILIPPINES HOUSE PRICES

A Project Study Submitted to the
Faculty of the College of Computer Studies and Engineering

In Partial Fulfillment of the Requirements for the
Subject Foundation of Business Analytics

by
Jennifer C. Marquez
Rikki Mae M. Martinez
Angelica B. Orido
Marianne C. Rario

ITC C307 – 402I

Submitted to
DR. MENGVI P. GATPANDAN

First Semester, SY 2022 – 2023

ABSTRACT

Predicting house prices is still a hot topic in one of the most rapidly urbanizing nations in Asia -- Philippines. The property prices in this city depend on several interdependent factors. The location, land area, and built-in features were key factors that were taken into consideration. In this research, an analytical study has been conducted by utilizing the dataset from Dot Properties Philippines. An attempt has been made to construct a predictive model for evaluating property prices based on the prominent factors influencing pricing ranges. Modeling explorations apply Multiple Linear Regression to build a predictive model and obtain an optimal prediction result. The findings show a strong correlation between house prices and the features listed by the proponents, such as property location, land area, and built-in features. Given these results, more research is needed to understand better the factors that impact house prices.

Table of Contents

List of Figures

Chapter 1

INTRODUCTION

Background of the Study

People usually know houses are considered the best assets as their prices have increased rapidly over the years (Weeken, n.d.). Buying and selling a home come with a lot of consideration, such as the price will vary depending on the year a house was built, the number of bedrooms and bathrooms it has, the land area on how spacious the place is, and more (Kaushal, 2021). Buyers look for a quality home with a reasonable and worthy price, and sellers highlight all the features of their house for them to be able to sell it at a higher price (Kaushal, 2021).

In the modern economy, predicting house prices is currently a hot topic. A recent study shows by Philippine Statistics Authority (2019) that the Philippines belongs to the most rapidly urbanizing nations in the Pacific and East Asia. The nation's population growth in urban areas was approximately 51.2%, which is 51.73 million Filipinos; this implies a further increase in both on-demand and property selling. Likewise, every year house prices increase. Thus, there is a dire necessity for house prices in the future to be predicted by the model. The model will provide help to those people who are planning to buy a house. Therefore, customers can arrange the right time to purchase a property.

The economics principle states that a property's market price is determined by the point at which the supply and demand curves cross, which might depend on various subjective and objective criteria. Since the real estate market has been too erratic and volatile to be regarded as a perfect market, it is unlikely that the market price of a property would match the market value in practice. Therefore, real estate appraisers must identify the objective variables that account for most of a property's price because many subjective ones heavily influence this component.

Machine learning is one of the most advanced research methods applied to research property today. The approach is used to evaluate the value and price and determine future uses and potential challenges (Trawiński et al., 2017). Property has been transformed from an experience-driven industry to an enterprise industry with a high degree of data and intelligence driven by machine learning (Barr et al., 2017). One of the machine learning algorithms that provides a promising ability for real estate price prediction is multiple regression. This statistical technique

works well in price prediction, given reliable input data with high quality.

Several variables influence house prices. According to Zhang (2021), various factors and features influence house prices, including transportation, environmental, and house features. As part of the research process, the researchers will establish a Multiple Linear Regression model to predict housing prices and test the method using Dot Properties Philippines real estate prices data. The provided data was gathered from Dot Property Philippines's housing market. Dot Property Philippines, launched in April 2015, has quickly become one of the leading property websites. It is one of the nine-nation property portals owned by Dot Property Group, which debuted in May 2013 and is one of Asia's largest property portal networks (Dot Property Moving Asia Online, n.d.). They provide connections for property buyers, renters, and investors with private house sellers, renters, real estate brokers, and the most excellent property developers in the country.

In addition, this organization gives great importance to its partners by providing them with updated data and information about the housing market industry. This data contains information that helps real estate buyers, renters, investors, brokers, and property developers make intelligent decisions about making an economical purchase and leveraging real estate products.

Statement of Objectives

The research aims to create a model for house price prediction using multiple regression and explore factors that affect house prices to obtain optimal prediction results.

The objectives of the study:

1. To identify the significant attributes that contribute to predicting house prices.
2. To develop a multiple regression model to predict house prices.
3. To have a recommendation plan for potential buyers based on the house prices per land area and location.

This study aims to help individuals by providing predicted information about the house prices in the Philippines listed by Dot Property Philippines with the help of multiple regression analysis that manages to give benefit the following:

Significance of the Study

As presented in this research, “A Multiple Regression Analysis in Prediction of Dot Property Philippines House Prices,” its significance is as follows:

Buyers. These findings will benefit the customers or potential buyers looking for a quality property worth their money and help them look for specific house features with their budget in mind.

Sellers. This study will help the sellers have data and information on the house prices based on their characteristics and land locations, enabling them to know what to improve and highlight on the house they are selling to make a large profit.

Proponents. This research will let the researcher devise a predictive model for house price prediction. The researcher can also gain insights into how the data may play a role in determining the house price in the Philippines.

Future Researcher. The results of this study will assist future researchers in developing an innovative way to forecast house prices.

This academic study will give reports about the variables that influence house prices based only on the data provided by Dot Property Philippines. It does not represent the overall factors that affect house prices in the Philippines, as the factors are only limited to the website where the data is gathered.

Moreover, this research aims to create a model for house price prediction using multiple regression and explore factors that affect house prices. Potential buyers’ recommendations were also considered based on the house prices per land area and location. This study will significantly benefit sellers, buyers, proponents, and researchers to ascertain the most important attributes to predict house prices by utilizing machine learning to conduct a study in this field.

Chapter 2

REVIEW OF RELATED LITERATURE AND STUDIES

This chapter presents the review of related literature and studies that the proponents researched thoroughly to strengthen the importance of the current study. This will also give the paper's synthesis for the reader to be informed and have an in-depth knowledge of the study.

Real Estate

The estimated worldwide real estate was $217 trillion in 2015, which is approximately 2.7 times the global GDP. It was roughly 60% of total global resources, thus making it one of the key factors behind the nation's economic stability and growth. In a study by Zulkifley et al. (2020), various global human rights and organizations focus on house importance. Each country house is significantly rooted in the financial, political, and economics. Although, fluctuating house prices were always a concern for real estate, buildings, and house owners. Likewise, as stated, it becomes unaffordable for houses due to substantial price growth in the housing sector for various countries. Since homeowners' Quality of Life (QoL) in the national economy depends on the potential house price increase.

Machine Learning

In recent years, Machine learning has demonstrated significant results in solving real-world problems using various algorithms, according to a study by Pathak et al. (2021). By analyzing house prices utilizing machine learning algorithms, the study provided an outline to better understand the real estate market conditions. Data is the crucial component to exploring any problem. It provides detailed information that can be automated and analyzed using the computer. Several factors influence the frequent variations in real estate values. According to the study by Rawool et al. 2021, these several variables might affect the price of a home, including its physical characteristics, location, and economic conditions. Data is an essential source for analysis and forecasting in the real estate sector. Future home price projections have always piqued the interest of many researchers. This is because the standard approach to calculating property prices based on cost and selling price comparisons does not satisfy the criteria for certification (Ozdemir, 2022).

Much research has been dedicated to determining which algorithm is most appropriate for performing price prediction within this context since many housing property ads are available online. In addition to tabulating and managing the data, developers can forecast trends and make good decisions using it (Manjula et al., 2017). In the paper of Rawool et al. 2021, Automation is becoming increasingly prevalent across all industries. Nevertheless, it is impossible to train a model without data. In essence, machine learning involves building these models from historical data and using them to predict future data. Many machine learning methods, including linear regression, decision trees, and random forests, are employed in creating a prediction model. Step-by-step procedures were used for data collection, preprocessing, analysis, and model construction.

Using machine learning approaches to predict real estate values has several benefits. First off, according to Ho et al. (2021), machine learning algorithms do not require properly distributed training data, in contrast to traditional econometrics models. The assumption of normality is the foundation of many statistical tests. Those statistical tests will be invalid if the data are not regularly distributed. Second, with the iterative processes used by machine learning models, such as grid search in RF or GBM algorithms and stochastic gradient descent in SVM algorithms. The usage of this technology now is less expensive and time-consuming. The operations, which used to take a long time in the past, be swiftly finished with the high-speed processing capacity of current computers. Third, it is possible to reduce estimation errors to the absolute minimum for training and test data sets.

House Price Prediction

According to the study by Gomez et al. (2021), It is common for people, especially those without the proper knowledge to set a price for a home without having a proper evaluation of the house before. In addition, the study states that this is the reason why house price predictions became an interest in the industry, and people started to work as real estate agents. Some even started to build real estate companies. Research from Vishwakarma & Singhal (2020) states that it is significant to manage the process of production and pricing by first analyzing the industry's demand and utilizing it. This pertains to using historical records to have a proper and reliable basis for predicting house prices. The scholarly works of Mohd et al. (2020) expressed that it is indeed a challenge to the industry of real estate to perform a house price prediction due to the difficulty

in finding the best model or algorithm to use, but still, researchers tried and tested the application of different models available such as Linear Regression, Artificial Neural Network, Multiple Regression Analysis, and more. In the study of Baldominos et al. (2018), states that were identifying the opportunities in the real estate industry using machine learning would benefit both sides of the industry, as this will help the sellers to avoid setting a price below the market value and help the customers know the true worth of a house based on its features. According to this study, building a model that can predict a house price based on its features is a huge step up for the industry.

Over the past few years, various research had been conducted concerning the prediction and analysis of house prices. Kushal, A., & Shankar, A. (2021) developed an artificial neural network in which, with the help of machine learning, the buyers can input the type of house they desire to buy. The model can predict the house price and display the estimated price of their desired house. Mubarak et al. (2022) made use of a framework for recommending real estate to buyers. The machine monitors user interactions through an online real estate portal. The framework can make personalized real estate recommendations based on location, collaboration, and content. In addition, the buyer's feedback mechanism tested the recommendation effectiveness with a mean absolute precision of 79% precise generated suggestions. Buyers were interested in at least 3 out of 5 recommendations produced. In a separate study by Wu, Z. (2020), the model solved the problem of forecasting local houses' annual average sales price using multilinear regression. The housing prices' main distribution was attained through data. It analyzed lasso and linear regression, which includes nearby areas, educational resources, floor area, number of rooms, and the proportion of low-income people.

Synthesis

The studies' similarities are that most focus on the same goal. To achieve excellent performance and accurate results when it comes to house price prediction and to identify and discover the relevant features or attributes that affect the overall process of the prediction and its results. This Multiple Regression Analysis in the Prediction of Dot Property Philippines House Prices will enable not just the people of the company but the people in the real estate industry, such as the buyers and sellers, to gain knowledge when it comes to pricing and buying their

properties. This will help the sellers to know their properties' worth and help the buyers to know their money's worth which is also in line with the previous studies' intentions. The studies have similar methodologies and processes, using a machine learning algorithm to perform house price prediction.

However, this also comes with a gap in the study, as the previous studies explored and tried several machine learning algorithms to know the most accurate and practical models to apply. Other studies focused only on reviewing the existing applications they gathered and were only focused on analyzing the models' results and accuracies. These gaps and similarities serve as the proponents' guide in their study, A Multiple Regression Analysis in Prediction of Dot Property Philippines House Prices.

Chapter 3

METHODOLOGY

This study is adjusted to the phase of the CRISP-DM method. CRISP-DM, or Cross-Industry Standard Process for Data Mining methodology, is proven to have a well-structured planning approach as it contains specific phases, tasks, and output in data mining. According to the study of Martinez-Plumed et al. (2021), Schröer et al. (2021), and Wiemer et al. (2019), CRISP-DM is the independent-process model, and the de-facto standard in the industry of data mining as the model is composed of six stages or phases that are connected with an arrow that acts as an indicator for admissible dependencies among stages. The crisp-DM model allows the proponents to go back and forth between the stages if necessary.

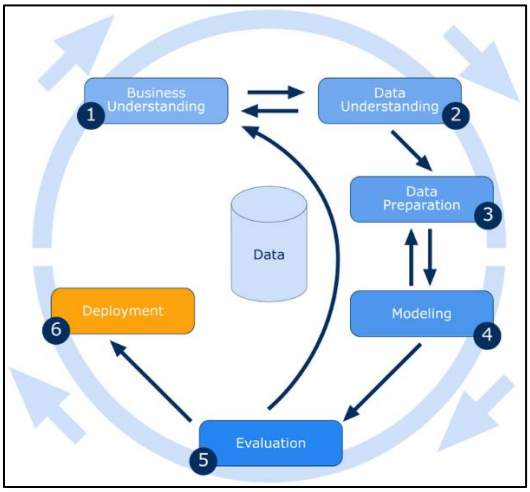


Figure 1. Crisp-DM Methodology Process (Roy, A., 2018)

Business Understanding

Business Understanding is the initial methodology phase for analyzing the business requirement or objectives and understanding the context of the goals. The information is then converted into a problem statement to design and create a project plan. This dataset was gathered from the website of Dot Property Philippines, a real estate company that offers homeowners, buyers, sellers, real estate agents, and property searchers a top-quality platform to communicate

with each other regarding buying, selling, or renting property. Dataset was collected for the study to analyze the data, create a model that can predict house prices, and help individuals gain knowledge of house price prediction and the affecting factors through multiple regression analysis.

The business objectives are to know the factors that can affect house prices for the company to address and improve the quality of the house they sell and maintain a high number of sales. The study's objectives are to identify the significate attributes that contribute to predicting house prices, develop a multiple regression model to predict house prices, and have a recommendation plan for potential buyers based on the house prices per land area and location.

Data Understanding

Data Understanding is understanding the data gathered from the initial collection, familiarization of data, discovering initial insights from the data, and finding the significant data to be used. The dataset contains the following attributes that will be used in analyzing and creating a multiple regression analysis to perform house price prediction successfully.

- Title – Head of the house page posted on the Dot Property Philippines website.
- Lot Area (Land Area) – Total number of property land area, including the yard up to the property of the boundary line for sale in square meters (sqm).
- Floor Area (Usable Area) – Total number of property land areas usable or available to occupy and build a house for sale in square meters (sqm).
- No. of Floors – The total number of floors of the house for sale.
- Bedrooms – Total number of bedrooms of the house for sale.
- Bathrooms – Total number of bathrooms of the house for sale.
- Price – Price of the house for sale in peso (₱).

Data Preparation

Data Preparation is about collecting and setting all the final datasets, which will be fed into the modeling tools needed in the next phase. The data preparation includes selecting, cleaning, constructing, integrating, and formatting. Preparing the dataset is necessary for this phase by

transforming unwieldy raw data into valuable and actionable information. This phase includes the following process:

1. Selecting data
   - Related attributes will be chosen to simplify the process of selecting data. This related attribute has Title, Lot area, Floor area, No. of Floors, Bedrooms, Bathrooms, and Price.
2. Cleansing data
   - The dataset contains missing values, and it is appropriate to replace all the missing values with 0s. Those attributes with missing values are Lot Area, Bathrooms, Floor Area, No. of Floors, Bedrooms, and Price.
3. Transforming data
   - To ensure linearity and normality, transforming data is necessary for multiple regression.
4. Integrating data
   - This stage combines all the final tables into one table into a new dataset.

Modelling

Modeling is when after the data has been preprocessed, particular modeling techniques are chosen, and data can be modeled. The four major modeling tasks include selecting, testing, creating, and assessing the model. The modelling phase involves applying a predictive model to the data and identifying hidden patterns, structures, and relationships among variables. The research aims to create a model for house price prediction by using multiple regression and exploring factors that affect house prices to obtain optimal prediction results.

Multiple regression analysis allows particular price predictions by getting the dependent and independent variables. The model's capability can be seen when the value of the relationship between independent variables (x), such as; Lot Area (Land Area), Floor Area (Usable Area), No. of Floors, Bedrooms, and Bathrooms, and dependent variables (y) such as Price, is measured. Moreover, multiple regression's common characteristics is the relationship between multiple x and result y. The expression of multiple regression can be written as follows:

$$y= w0x0 + w1x1 + w2x2 + ... + wnxn$$

The purpose of multiple regression is to summarize the probability of all occurring events simultaneously, and it can summarize the laws of some unrelated elements. This model can be achieved by using the Price attribute as the target or dependent variable. In contrast, the other remaining attributes are set to be the independent variable and identify the correlation coefficient of each attribute.

Evaluation

Evaluation is where the achieved model is properly examined throughout the evaluation step. In this phase, working with business objectives is necessary to come up with evaluation sheets and process reviewing to see if there is anything that needs to determine for the next steps. The primary evaluation includes evaluating the results, reviewing the process, and determining the next steps.

The stage where the assessment happened whether or not the result successfully met the project's criteria. In this stage, the need for a clear understanding of the project objective is a must. There are three main tasks at this stage. These are evaluating results, reviewing processes, and determining the next step. Evaluating results can be done using the criteria proposed for the project. Do these criteria satisfy the intent of the project? Which of these results should be used in the project? Reviewing the process is the next stage of evaluation. Here, we need to trace back if we forgot to execute a particular procedure, review successful testing of the projects, summarize the result, and revise if things are not in line with the project. Lastly, Determine the next stage; this procedure can determine our next move based on the previous task, whether to create new projects for further deliberation or proceed with deployment. Overall evaluation is an essential task we must look for in creating various projects.

Deployment

Deployment is where the knowledge is acquired from data analysis and translated into actionable recommendations. It will help the business constituents by effectively communicating the analysis results. In addition, it plays a role in a successful analytical project. The primary deployment process includes plan deployment, plan to monitor, plan final report, and review

project. The deployment stage is where the proponents determine a deployment strategy for the data evaluation results of the model. At this step, the process is documented, particularly if it has been identified to create relevance in the model creation. Given that the deployment phase is essential to the project's success, it becomes appropriate to take deployment into account throughout the business understanding phase. This is where predictive analytics may significantly enhance the operational aspect of the target organization.

The research's final report is also written during this phase as well. According to the deployment methodology, this report may only be included if it still needs to be documented. It might also be delivered as a thorough summary of the outcomes of the data modeling procedure.

Chapter 4

RESULTS AND DISCUSSION

This chapter presents the findings and results generated by the proponents after the data analysis, cleansing, and processing using the rapid miner tool. The generated outcome is in line with the listed statement of the study's objectives: A Multiple Regression Analysis in Prediction of Dot Property Philippines House Prices. Understanding the dataset begins with compiling preliminary data and the results to identify data problems. The proponents were able to determine essential insights and detected interesting subsets to form knowledge for hidden information.

In relation to the study's objective no. 1, to identify the significate attributes that contribute to predicting house prices.

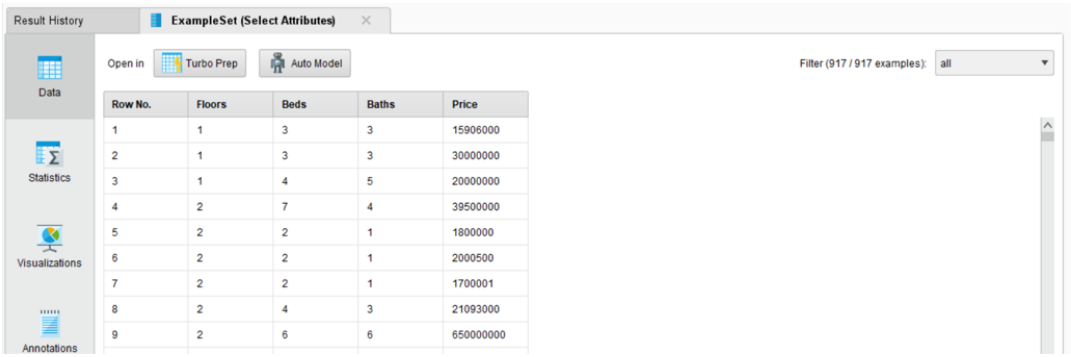
Attributes	Lot Area (Land Area)	Floor Area (Usable Area)	No. of Floors	Bedrooms	Bathrooms	Price
Lot Area (Land...	1	0.998	-0.014	-0.038	-0.060	-0.012
Floor Area (Us...	0.998	1	-0.003	-0.039	-0.059	-0.018
No. of Floors	-0.014	-0.003	1	0.100	0.145	-0.368
Bedrooms	-0.038	-0.039	0.100	1	0.778	0.264
Bathrooms	-0.060	-0.059	0.145	0.778	1	0.401
Price	-0.012	-0.018	-0.368	0.264	0.401	1

Figure 2 Correlation Matrix

Figure 2 shows the generated correlation matrix. The proponents used a correlation matrix to determine the significant attributes or features that affect the dependent variable 'Price'. There is No Correlation between 'Price' and 'Title', 'Price' and 'Lot area' that has -0.012, and 'Price' and 'Floor area' that has -0.018.

However, the correlation between 'Price' and 'No. of Floors' is -0.368, which indicates a weak negative relationship between the two variables. It means that if the 'No. of Floor' increases, the 'Price' decreases and vice versa. While the correlation between 'Price' and 'Bedrooms' is 0.264, indicating a weak positive relationship between the two variables. It means that if the 'Bedrooms' increases, then the 'Price' also increases and vice versa. In addition, the correlation between 'Price' and 'Bathrooms' is 0.401, indicating a moderate positive relationship between the

two variables. It means that if the ‘Bathrooms’ increases, the ‘Price’ also increases and vice versa.



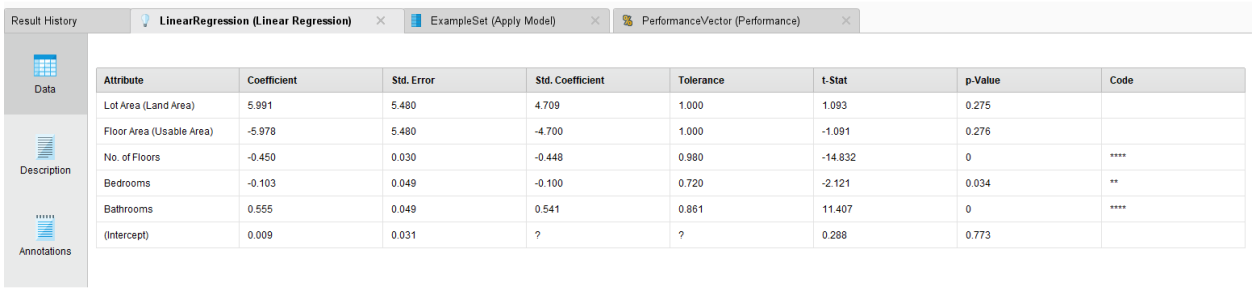
Row No.	Floors	Beds	Baths	Price
1	1	3	3	15906000
2	1	3	3	30000000
3	1	4	5	20000000
4	2	7	4	39500000
5	2	2	1	1800000
6	2	2	1	2000500
7	2	2	1	1700001
8	2	4	3	21093000
9	2	6	6	650000000

Figure 3 Significant Attributes

Figure 3 shows all the significant attributes in the dataset. This was achieved by selecting the significant features using the ‘Select Attributes’ operator. Those significant attributes are ‘No. of Floors’, ‘Bedrooms’, ‘Bathrooms’, and ‘Price’.

Knowing the significant attributes is essential as it will show the factors that can affect the following house prices. According to the research of Baldominos et al. (2018), it is proven that the following features, such as the No. of floors, No. of Bedrooms, and No. of Bathrooms does affect the selling price of the house. Also, it may not correlate with the data provided by Dot Property Philippines. Still, it is also proven in the same study that the floor area and lot area affect the selling price or the house's overall value.

In relation to the study’s objective no. 2, to develop a multiple regression model to predict house prices.

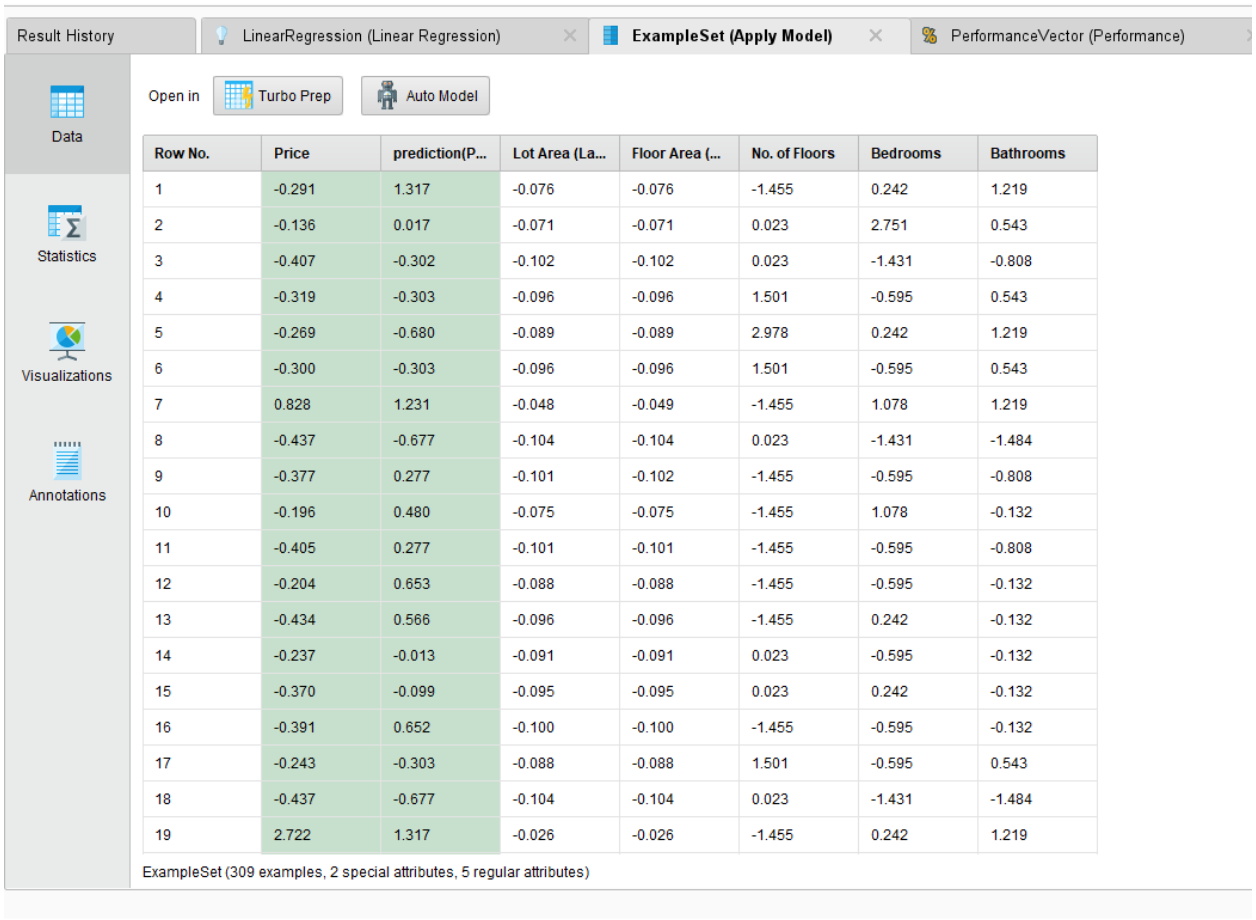


Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Lot Area (Land Area)	5.991	5.480	4.709	1.000	1.093	0.275	
Floor Area (Usable Area)	-5.978	5.480	-4.700	1.000	-1.091	0.276	
No. of Floors	-0.450	0.030	-0.448	0.980	-14.832	0	***
Bedrooms	-0.103	0.049	-0.100	0.720	-2.121	0.034	**
Bathrooms	0.555	0.049	0.541	0.861	11.407	0	***
(Intercept)	0.009	0.031	?	?	0.288	0.773	

Figure 4 Linear Regression operator results

Figure 4 shows the result after applying the 'Linear Regression' operator. The column includes a variety of important results such as intercept value, p-Value, and the Code. These results

show whether a variable has a relationship with the target variable. Moreover, the results of the p-Value column from all of the independent variables, such as Floors, Baths, and Beds, is > 0.05, meaning that these variables have a strong relationship with the Price, which is the target variable.



Row No.	Price	prediction(P...	Lot Area (La...	Floor Area (...	No. of Floors	Bedrooms	Bathrooms
1	-0.291	1.317	-0.076	-0.076	-1.455	0.242	1.219
2	-0.136	0.017	-0.071	-0.071	0.023	2.751	0.543
3	-0.407	-0.302	-0.102	-0.102	0.023	-1.431	-0.808
4	-0.319	-0.303	-0.096	-0.096	1.501	-0.595	0.543
5	-0.269	-0.680	-0.089	-0.089	2.978	0.242	1.219
6	-0.300	-0.303	-0.096	-0.096	1.501	-0.595	0.543
7	0.828	1.231	-0.048	-0.049	-1.455	1.078	1.219
8	-0.437	-0.677	-0.104	-0.104	0.023	-1.431	-1.484
9	-0.377	0.277	-0.101	-0.102	-1.455	-0.595	-0.808
10	-0.196	0.480	-0.075	-0.075	-1.455	1.078	-0.132
11	-0.405	0.277	-0.101	-0.101	-1.455	-0.595	-0.808
12	-0.204	0.653	-0.088	-0.088	-1.455	-0.595	-0.132
13	-0.434	0.566	-0.096	-0.096	-1.455	0.242	-0.132
14	-0.237	-0.013	-0.091	-0.091	0.023	-0.595	-0.132
15	-0.370	-0.099	-0.095	-0.095	0.023	0.242	-0.132
16	-0.391	0.652	-0.100	-0.100	-1.455	-0.595	-0.132
17	-0.243	-0.303	-0.088	-0.088	1.501	-0.595	0.543
18	-0.437	-0.677	-0.104	-0.104	0.023	-1.431	-1.484
19	2.722	1.317	-0.026	-0.026	-1.455	0.242	1.219

ExampleSet (309 examples, 2 special attributes, 5 regular attributes)

Figure 5 Apply Model Operator results

Figure 5 shows the result after applying the 'Apply Model' operator to the test data. This operator tests the accuracy of the created model to the remaining 30% of the data partition. It has 275 rows (examples), 2 special, and 3 regular attributes. The 2 special attributes are the ones highlighted in green. Those were the Actual Price and the Predicted Price of the property and were generated after the operator runs.

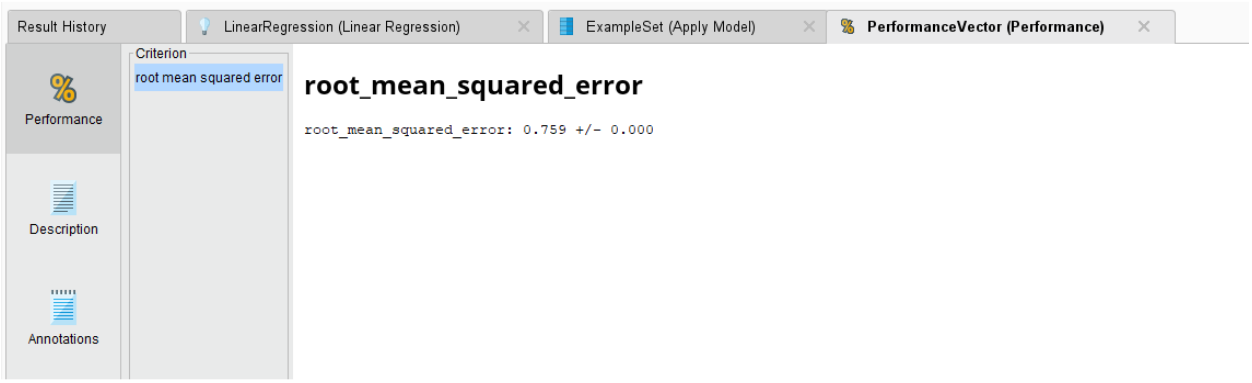


Figure 6 Performance operator results

Figure 6 shows the result of the Root Mean Square Error (RMSE) using the 'Performance' operator. The RMSE is the square root of the mean squared error between the predicted and actual values. The weighted average error between the predictions and actuals in this dataset is 0.759, which indicates that the model can relatively predict the data accurately.

In research conducted by Wu, Z (2020), utilizing multiple linear regression in house pricing prediction is more effective by getting the optimal combination of multiple independent variables to estimate or predict dependent variables. In addition, the RMSE is necessary to measure the error rate of the regression model. It can help to determine how well the algorithm can predict future prices.

In relation to the study's objective no. 3, to have a recommendation plan for potential buyers based on the house prices per land area and location.

When buying a home, one crucial factor that must be in mind is the location of the potential residential property since it is the only thing that cannot change. Changing the size of a house can be very easy, but changing the location is beyond your control. Here are some site recommendations that can be considered based on different factors according to the consulted real estate agent by the proponents. The following locations listed below are some of the locations inside and outside of the metro.

Metro Manila:

Pasig

The house price for Pasig City increases when the location is near Makati and BGC, which has both business district areas. According to the data, the prices for houses in Pasig City range from **7,500,000.00 PHP to 32,000,000.00**. Pasig City, one of Metro Manila's larger cities, is among the most excellent locations. Many families, students, and even foreigners are interested in the city.

Here are some advantages of why Pasig City is an ideal location to move in.

1. Accessibility

Pasig City is well situated close to many establishments, making it a great place to live. It is a very commercialized neighborhood. Therefore, there are an excessive number of local supermarkets, gas stations, shopping centers, and restaurants. Put an end to all concerns about medical emergencies and consultations. The Medical City, a world-class hospital, is one of the area's leading healthcare facilities.

2. Lifestyle

It is possible to pursue various hobbies because there are so many options. Maintain a healthy lifestyle while getting away from the city's bustle. Visit one of the city's well-equipped gyms and swim a few laps in the community pool.

3. The local government unit promotes eco-friendly programs.

It is no accident that Pasig City is renowned as the "green city." The local administration has been putting up initiatives to promote eco-friendly living. Even the city hall's rooftop had been transformed into a gorgeous garden. Along with the arrival of electric vehicles, some days see numerous routes devoid of cars to make room for cyclists and runners. Additionally, people can enjoy the area's well-kept small parks. These are all examples of the city's green initiatives in response to the pressing problem of climate change.

Makati

According to the data, the prices for houses in Makati City range from **95,000,000.00 PHP to 613,500,032.00 PHP**. Makati is considered the most expensive compared to other areas since this location is regarded as one of the top business districts in the country. It is a busy city full of life and individuals from many walks of life. Being a forward-thinking city, Makati has developed into the Philippines' biggest business center, with numerous respected organizations. Due to its rapid financial and infrastructure growth in recent decades, the city is regarded as the best place to be in the metro.

Living in Makati, however, is a unique experience from working and conducting business. Makati is known for enforcing laws more strictly, particularly the financial and traffic rules. As a result, the city is safer and more welcoming to businesses. The following are some benefits of residing in Makati:

1. Centered

Makati serves as a cornerstone of the Philippine National Capital Region. The city has everything, including the hottest clubs, restaurants, and shopping complexes. Real estate corporations also maintain Makati's development by consistently developing condos in the city. Makati is the location of everything a person could need; the city is full of supermarkets, shopping centers, and restaurants.

2. Security

The Makati authorities follow stricter laws when enacting and implementing laws. In addition to a reputable police force and Traffic Enforcement Agency, Makati stands out when enforcing traffic laws.

3. Financial Stability

Makati has grown into the country's financial hub in the past few decades. Many businesses and financial institutions moved to Makati, making Makati its own Wall Street.

Outside Metro Manila:

Rizal Province

Living in Metro Manila with a lot of noise, pollution, and always too busy to live. It is then considered to move to a more serene location to move in. According to the data, the prices for houses in Rizal Province range from **9,500,000.00 PHP to 16,400,000.00**. The province of Rizal, an hour or two away from the city, is a good area for a new home. It is regarded as a preferable place to move in and put down roots. However, moving away from the city is a big decision because Metro Manila has a vast prospect of satisfying the needs of an individual.

Here are some advantages of why Rizal province is an excellent location to move in.

1. Close to Nature

The nature meets city lifestyle that Rizal province provides is one of the primary benefits of living there. The area is surrounded by lush forests, stunning waterfalls, and cold, fresh air at the base of the Sierra Madre. These are just a few things Rizal province has to offer.

2. Accessibility

Due to its tranquil environment and accessibility, Rizal has become a nice place to call home. All these locations are easily reachable by personal automobile or public transportation: North of Metro Manila is Bulacan, east is Quezon, west is Metro Manila, and southeast is Laguna.

3. Work-Life Balance

Work-life balance is facilitated by living in a convenient area, which Rizal province provides. There are many different leisure activities in Rizal, including upscale resorts. The metro is only a half-hour or two away, and there are numerous alternative transit options for clients to have more time for what matters—finishing their work and spending time with their families.

4. Cheap Cost of Living

A home in Rizal can be bought for less money than in the wealthier parts of Metro Manila. It is, without a doubt, a residential haven for growing families and retirees looking for a more peaceful environment.

Because goods are also more affordable in the province than in the city, investing there could lead to more significant savings.

Laguna

According to the data, the prices for houses in Laguna Province range from **1,400,000.00 PHP to 65,000,032.00 PHP**, which makes Laguna one of the less expensive locations. As a home or education district in the city, Laguna serves as an ideal sanctuary for those who work or study in the metropolis. The South Luzon Expressway and Cavite-Laguna Expressway make it easier for residents to commute, but the province has a lot to offer. Property investment in provinces near Metro Manila has become more popular due to new infrastructure.

The following are some benefits of buying a home in Laguna Province:

1. Close to Metro Manila

Despite Laguna being a province, it is relatively easy to get there and back from Metro Manila. The trip will take roughly an hour using the recently constructed Skyway. As a result, from the province of Laguna, it will be simple for you to return to the chaos of Metro Manila.

2. Natural Surroundings

Nowadays, being surrounded by lush, sunny naturalistic environments is more challenging, especially in locations like Metro Manila. Therefore, if people are tired of living in cities, they should consider moving to Laguna.

3. Less Traffic

A major downside to living in a metropolitan area is the incessant traffic; it is also hard to keep up with the ever-changing traffic laws. There is no traffic in Laguna, which makes for an easy drive and commute, so if one is tired of the traffic and busy roads, Laguna is a place to go.

Knowing the property's appraised value is essential for setting a home-buying budget, and one crucial factor influencing the property price is its land area. However, the price of the property is still determined by whether it is in an urban or rural region. Those with modest land sizes in metropolitan locations are nonetheless more costly than properties with larger land areas in rural regions. Here are the comparisons of the property prices based on land areas in both urban and rural regions:

Property Prices Based on Land Areas in Urban Regions:

75 SQM – 150 SQM

5,000,000.00 – 65,000,000.00

155 SQM – 400 SQM

9,000,000.00 – 135,000,000.00

500 SQM – 2000+SQM

130,000,000.00 – 1,000,000,000.00

Property Prices Based on Land Areas in Rural Regions:

75 SQM – 300 SQM

50,000.00 – 350,000.00

400 SQM – 800 SQM

500,000.00 – 750,000.00

1000+ SQM

950,000.00 – 5,000,000.00

The research conducted by Zhang P. et al (2015) explains the reasons behind the dynamics shown. There is mutual causation between land and property prices based on the cost-driven perspective that states land prices contribute to housing prices. The price fluctuation also depends on the context of external shocks for price changes. Land prices influence house prices in highly developed regions, but in rural areas, land price changes are more likely to affect housing prices. As a result, even if the land size of the property in urban areas is smaller, the property becomes more costly since the land price in the area is high.

Living in a city has a higher cost than in most provinces, mainly because most businesses and big establishments are in the city. These businesses can give better jobs and opportunities than provinces. That is why people tend to relocate to the city (Dot Property Moving Asia Online, 2016). According to the research made by Mcdonald (2018), Urban regions are preferred by businesses because they offer one key advantage: proximity. Cities provide proximity to shared infrastructure and integrated supply chains for many people and businesses. Additionally, companies have access to knowledge because of the density of metropolitan areas. Near other businesses' locations encourages innovation, strengthens networks, and facilitates information.

Chapter 5

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

This chapter presents the study's overall summary, conclusions obtained throughout the research, and the proponents' recommendations for future researchers.

Summary

The study aims to perform a multiple regression analysis to predict house prices in the Philippines with the use of the dataset obtained from Dot Property Philippines, a real estate company based in the Philippines with a list of objectives such as; identifying the significant attributes that contribute to predicting house prices, developing a multiple regression model to predict house prices, and having a recommendation plan for potential buyers based on the house prices per land area and location. The proponents then analyzed and studied the data, then performed the multiple regression analysis using the rapid miner tool. After completing the multiple regression analysis, the proponents identified the significant attributes that affect house prices: The number of Floors, Bedrooms, and Bathrooms. Even though the attributes Lot Area and Floor area do not show as significant attributes, the proponents decided to include the two attributes when developing a multiple regression model since, according to other studies, it does have an effect when setting a price for real estate or houses. The proponents successfully developed a multiple regression model that generated a good root-mean error score of 0.759. This allows the proponents to analyze the generated outcome more and be able to recommend a plan to potential buyers regarding where and what property is best to purchase.

Conclusion

Based on the results of the study's objective no. 1, to identify the significant attributes that contribute to predicting house prices. The correlation matrix shows that the No. of Floors and price has a negative correlation, while Bedrooms and Bathrooms are positively correlated with the price, which concludes that variables No. of Floors, Bedrooms, and Bathrooms are proven to affect house price prediction.

Based on the results of the study's objective no. 2, to develop a multiple regression model to predict house prices. It has shown a good root mean squared error score of 0.759, concluding that the attributes used in creating a multiple regression model are a good fit for the study. It also shows the predicted and actual values of the house prices, and it shows that some predicted values are a bit far from the actual values, but there are still a lot of predicted values that are close to the actual values. Overall, this interprets that the proponents' multiple regression model is good, but there is still room for the study's improvement.

Based on the results of the study's objective no. 3, to have a recommendation plan for potential buyers based on the house prices per land area and location. The study's generated output allowed the proponents to successfully create a recommendation plan with the help of the consulted real estate agent, which successfully shows the advantages of buying a house in both the province and the metro. The mentioned advantages can help the buyers decide where to buy a house depending on their needs.

Overall, the proponents successfully understood, analyzed, and performed multiple regression analysis in the Prediction of Dot Property Philippines House Prices.

Recommendation

The following recommendations are listed to help future researchers further improve the study.

1. To find or discover more attributes or independent variables that correlate or affect house price prediction.
2. To analyze more or use a different approach or model to improve or lessen the root mean squared error score and to improve the gap between the actual and predicted house price values.
3. To add a more detailed recommendation plan by considering listing all the possible advantages and disadvantages of buying a house or property in every city and province, and every square meter possible.

References

Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. Applied sciences, 8(11), 2321.

Barr, J., Ellis, E., Kassab, A., & Redfearn, C. (2017, March). *Home Price Index: A Machine Learning Methodology*. International Journal of Semantic Computing. Retrieved September 20,2022,fromhttps://www.researchgate.net/publication/316869611_Home_Price_Index_A_Machine_Learning_Methodology

Centre for Cities. (2018b, April 3). *The geography of demand for residential and commercial space*.https://www.centreforcities.org/reader/city-space-race-balancing-need-homes-offices-cities/geography-demand-residential-commercial-space/

Cost of Living: City Versus Province. (2016, April 14). Dot Property Moving Asia Online. https://www.dotproperty.com.ph/blog/cost-of-living-city-versus-province

Dot Property Moving Asia Online. (n.d.). www.dotproperty.com.ph. Retrieved September 19, 2022, from https://www.dotproperty.com.ph/

Estoque, L. N. C., Fuente, L. M. D. D., Mabborang, R. C., & Molina, M. G. (2022). Forecasting Urban Population Growth in the Philippines using Autogressive Integrated Moving Average (ARIMA) Model *EPRA International Journal of Multidisciplinary Research (IJMR), 8*(7), 132-153.

Gómez, D. O., Manrique, M. A. C., Sierra, O. B., Laniado, H., Carrión, R. M., & Millan, D. A. R. (2020). Housing-Price Prediction in Colombia using Machine Learning.

Ho, W. K., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research, 38*(1), 48-70.

Kaushal, A. (2021, April 25). *House Price Prediction Using Multiple Linear Regression*. Retrieved September 19, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3833734

Kaushal, A., & Shankar, A. (2021, April). House Price Prediction Using Multiple Linear Regression. In *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*.

Manasa, J., Gupta, R., & Narahari, N. S. (2020, March). Machine learning based predicting house prices using regression techniques. In *2020 2nd International conference on innovative*

mechanisms for industry applications (ICIMIA) (pp. 624-630). IEEE.

Manjula, R., Jain, S., Srivastava, S., & Rajiv Kher, P. (2017). Real estate value prediction using multivariate regression models. IOP Conference Series: Materials Science and Engineering, 263(4). https://doi.org/10.1088/1757-899X/263/4/042098

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/tkde.2019.2962680

Mohd, T., Jamil, N. S., Johari, N., Abdullah, L., & Masrom, S. (2020). An Overview of Real Estate Modelling Techniques for House Price Prediction. *Charting a Sustainable Future of ASEAN in Business and Social Sciences*, 321–338. https://doi.org/10.1007/978-981-15-3859-9_28

Mubarak, M., Tahir, A., Waqar, F., Haneef, I., McArdle, G., Bertolotto, M., & Saeed, M. T. (2022). A Map-Based Recommendation System and House Price Prediction Model for Real Estate. *ISPRS International Journal of Geo-Information*, *11*(3), 178.

Özdemir, O. (2022). House Price Prediction Using Machine Learning: A Case in Iowa. https://doi.org/10.13140/RG.2.2.19846.86086

Pathak, S. M., & Chaudhari, A. K (2021), Comparison of Machine Learning Algorithms for House Price Prediction using Real-Time Data, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 12 (December 2021)

Rawool, A. G., Rogye, D. v, Rane, S. G., Vinayk, D. R., & Bharadi, A. (2021). House Price Prediction Using Machine Learning. In IRE Journals | (Vol. 4).

Roy, A. (2018, June 22). Chapter 1 - Introduction to CRISP DM Framework for Data Science and Machine Learning. https://www.linkedin.com/pulse/chapter-1-introduction-crisp-dm framework-data-science-anshul-roy?trk=public_profile_article_view

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Thamarai, M., & Malarvizhi, S. P. (2020). House Price Prediction Modeling Using Machine Learning. *International Journal of Information Engineering & Electronic Business*, *12*(2).

Tranwinski, B., Telec, Z., Krasnoborski, J., & Piwowarczyk, M. (2017, July). *Comparison of expert algorithms with machine learning models for real appraisal.* Retrieved September 20,2022,fromhttps://www.researchgate.net/publication/318980482_Comparison_of_expert_algorithms_with_machine_learning_models_for_real_estate_appraisal

Varma, A., Sarma, A., Doshi, S., & Nair, R. (2018, April). House price prediction using machine learning and neural networks. In *2018 second international conference on inventive communication and computational technologies (ICICCT)* (pp. 1936-1939). IEEE.

Vineeth, N., Ayyappa, M., & Bharathi, B. (2018, April). House price prediction using machine learning algorithms. In *International Conference on Soft Computing Systems* (pp. 425 433). Springer, Singapore.

Vishwakarma, S., & Singhal, S. (2020b). House Price Forecasting Based on Hybrid Multi-regression Model. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3601507

Wang, P. Y., Chen, C. T., Su, J. W., Wang, T. Y., & Huang, S. H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, 55244-55259

Weeken, O. (n.d.). *Asset Pricing and the Housing Market*. Retrieved September 19, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=701272

Wiemer, H., Drowatzky, L., & Ihlenfeldt, S. (2019). Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. Applied Sciences, 9(12), 2407. https://doi.org/10.3390/app9122407

Wu, Z. (2020). Prediction of California house price based on multiple linear regression. *Academic Journal of Engineering and Technology Science*, *3*(7).

Zhang, Q. (2021, October 29). *Housing price prediction based on multiple linear regression*. Scientific Programming. Retrieved September 20, 2022, from https://www.hindawi.com/journals/sp/2021/7678931/

Zulkifley, N. H., Rahman, S. A., Ubaidullah, N. H., & Ibrahim, I. (2020). House Price Prediction using a Machine Learning Model: A Survey of Literature. *International Journal of Modern Education & Computer Science*, *12*(6).

ACKNOWLEDGEMENT

Our persistence in acquiring variables and analyzing data has resulted in the results. The purpose of this would be to represent the knowledge, skills, and abilities that we developed at José Rizal University.

There were several individuals who contributed their insight and assistance in preparing and completing this study in some way or another. This study would not have been possible without the strength and optimism that God blessed us with during its conduct. The research was completed with the help of different individuals who contributed significantly.

It is with gratitude that we acknowledge and thank our professor on this subject, **Dr. Mengvi P. Gatpandan**, for her patience, kindness, and understanding. The way she made comments made us feel respected and valued, and the generosity with which she guided us in our research endeavor enabled us to achieve our study's objective.

Additionally, we appreciate **DOT Property Philippines** for their support and the gracious manner in which we were treated while acquiring the data.

ABOUT THE AUTHORS

Angelica Orido is a 4th-year student pursuing a Bachelor of Science in Information Technology in the Computer Studies and Engineering Department at Jose Rizal University. She has a passion for coding, technology, music, and modeling. She has a little experience in runway modeling, modeling school, and music lessons. She is a CompTIA ITF+ Certified and an NC II (National Certificate II in Computer System Servicing) Passer.

Jennifer Marquez is a 4th-year student pursuing a Bachelor of Science in Information Technology in the Computer Studies and Engineering Department at Jose Rizal University. She has an Associate degree in Multimedia Technology from the University of Makati. She worked as a leader-CAD Engineer at Frontier Intermediary Technology Inc for seven years. She was assigned to Japan in 2018 to lead a three-month Architrend Software training for Filipinos and Japanese.

Rikki Mae Martinez is a 4th-year student pursuing a Bachelor of Science in Information Technology in the Computer Studies and Engineering Department at Jose Rizal University. She was a football player that competes in different prestigious football competitions from Phil Young husband, Coca-Cola, and Alaska Cup. She is a Web Development Level III Certified (American TESOL Institute Philippines).

Marianne Rario is a 4th-year student pursuing a Bachelor of Science in Information Technology in the Computer Studies and Engineering Department at Jose Rizal University. She is currently working as a Mid-Senior Backend Developer at Australian Firm (Frontline AU). She was recently featured in Manila Bulletin Tech News for creating a Proof of Concept (PoC) on how spammers can automate data scraping to any mobile application such as GCash and Viber resulting for GCash to add a layer of customer protection by masking the recipient's name which anonymizes the names of users in the send money service.