Forecasting the prices of villas in the city of Riyadh using Multiple and Simple Linear Regression

- **Fay Abdulaziz Alsaawi**
  Email: Fai_sawi@hotmail.com
- **Said H. Al-Qarni**
  Email: salkarni@ksu.edu.sa
- **Nadia M. Sarhan**
  Email: nsarhan@ksu.edu.sa
- **Anud Hassan Jeariby**
  Email: ajeribi@ksu.edu.sa
- **Marah Miqad Alosaimi**
  Email: Malosaimi6@ksu.edu.sa

College of Business Administration
King Saud University - KSA

Forecasting the prices of villas in the city of Riyadh using Multiple and Simple Linear Regression

Fay Abdulaziz Alsaawi
Said Al-Qarni
Nadia M. Sarhan
Anud Hassan Jeariby
Marah Miqad Alosaimi

College of Business Administration - King Saud University - KSA

Abstract

The purpose of this study is to try to predict the prices of real estate. specifically, villas in different areas of Riyadh by creating simple equations that help estimate the price based on a number of variables the most important of which is the required space for villa. Data were collected from 2019 to the first quarter of 2022, which were 3469 deals. we used the method of simple and multiple linear regression and the use of the statistical program SPSS to find the required equations and deals.

In Riyadh most of the deals were in the north, and the least was in the south, which accounted for 13% of the total transactions, and the availability of services in the neighborhood makes a significant difference in the number of deals in it and demand, also the proximity of the neighborhood to new projects makes a difference but less.

It was also noted that prices in the north of Riyadh are constantly increasing and that the area factor is insignificant in the west and south, which shows that the trend of buyers of small spaces in the north in particular is caused by the rise in prices that vary in the south and west where the area variable becomes minor, and the correlation weakens

Keyword: Linear regression, Villas prices, Forecasting, Estimate, Prediction.

التنبؤ بأسعار الفيلات في مدينة الرياض باستخدام الانحدار البسيط والمتعدد

في عبد العزيز السعوي
سعيد حوفان القرني
نادية محمد سرحان
عنود حسن جريبي
مرح مقعد العصيمي
كلية إدارة الاعمال – جامعة الملك سعود – المملكة العربية السعودية

ملخص

الهـدف من هذا البحث هو محاولة التنبؤ بالأسعار للعقارات وتحديداً الفلل في مناطق الرياض المختلفة وذلـك بإنشاء معادلات بسيطة تساعد على تقدير السعر بالاعتمـاد على عدد من المتغيرات أهمها المساحة المطلوبـة للفيلا، تم جمـع البيانات للصفقات التي تمت في المناطق المختلفـة ومساحاتها وبلغ عددها 3469 صفقـة من عـام 2019 حتى الربع الأول مـن 2022، استخدمنا أسلوب الانحدار الخطـي البسيط والمتعدد والاستعانة بالبرنامج الإحصائي SPSS لإيجاد المعادلات المطلوبة ومعاملاتها.

مـن أبرز النتائج أن توفر الخدمات في الحي يشكل فارق كبـير بعدد الصفقات فيه والطلب عليه، أيضاً قرب الحي من المشاريع الجديدة يشكل فارقًا ولكن أقل، كما كان من الملاحظ أن الأسعار في شمال الرياض تتزايـد بشكـل مستمر، وأن عامل المساحة تقـل أهميته في منطقة الغرب والجنـوب، وذلك يوضح أن اتجاه المشتريـن للمساحات الصغـيرة في منطقة الشمال خاصة سببه الارتفـاع بالأسعار الذي يختلف في منطقة الجنوب والغرب حيث تقل أهمية متغير المساحة ويضعف الارتباط.

الكلمات المفتاحية: الانحدار الخطي، أسعار الفيلات، التقدير، التنبؤ.

Introduction

Real estate especially in large cities in the Kingdom of Saudi Arabia, has made huge leaps in recent years due to many factors, including the rise in value-added tax on goods, which led to a rise in prices of building materials, as well as the Saudi building code and real estate insurance.

All of these factors above caused the citizen to give up many of his requirements and conditions and started to look for small areas that correspond to his income and the surrounding conditions.

In recent days there was an increase in the demand for apartments due to the rise in villa prices, as a comparison between purchases of apartments and residential villas indicates that apartments in Riyadh amounted to 443 during the third quarter of 2021, and villas amounted to 20 during the same period (Ali, Al-Watan newspaper, 2021).

Mohammed bin Salman the Crown Prince of the kingdom announced a strategy for the city of Riyadh, which would make the city's economy among the largest in the world, and said that Riyadh has very special advantages, as it represents about 50% of the non-oil economy in the Kingdom, which makes it "A very big opportunity to create huge economic growth in the Kingdom, and to create industry, tourism, and great progress, furthermore, Riyadh must be given great attention because it is one of the pillars of economic growth in the Kingdom».

The crown prince stated as well that Riyadh is today among the 40 largest economies in the world as a city, adding that the country aims for Riyadh to be among the 10 largest city economies in the world, and we aim to reach a population of 15-20 million people in 2030 and launch initiatives to protect the environment as well as making Riyadh a sustainable city by improving its quality of life (CNN, 2021).

All these factors causes an increase in demand in the city of Riyadh, which influenced the prices there. Moreover, the location of the property and the availability of services in the neighborhood have a strong impact on the prices.

In this research, we will first cover and discuss previous studies related to the subject of our research, which involve six studies discussed real estate prices and the real estate market in the United Arab Emirates, especially Dubai. Then we presented and discussed the results by making statistics for many real estate deals to know the purchase trends in terms of location and areas that have a lot of demand, by dividing the city of Riyadh into several regions and dealing with each region separately.

Since Logistic regression is predict the categorical dependent variable. It›s used when the prediction is categorical, for example, yes or no, true or false, 0 or 1(Menard, 1995) and (Hilbe, 2016). So, we will measure the relationship between some factors which are the space, availability of services, and closeness of the location to the new projects also their impact on the prices of houses and to predict the prices using multiple and simple liner regression. In the end of this research, we will present the important results that we worked on and produce many helpful recommendations.

Importance of the research

Considering the changes that occur in the real estate and economic sector and the observed increase in real estate prices, it has become important to focus on this sector, especially in the capital, Riyadh, due to the increase in the population and accelerated growth. In this research, we are creating a reference equation for those interested in real estate prices to determine the appropriate property for them according to their desires and capabilities.

Objectives of the research

The aim of this research is to measure the relationship between real estate prices and their influencing factors, such as the area of the house, and availability of services in the neighborhood, and its closeness to the new projects, then try to find an equation to predict prices for each region.

Limitations of the research:

- Time limitations: from 2019 to the first quarter of 2022.
- Spatial limitations: the city of Riyadh.

Research Methodology:

The applied analytical method was adopted in this study and the data for deals

were collected through the ministry of justice starting from 2019 to the first quarter of 2022 for various neighborhoods.

The stratified sample was adopted to select neighborhoods and because Riyadh contains 120 neighborhoods after excluding neighborhoods that do not contain purchasing movement, 10% of which are approximately 13 neighborhoods have been taken.

Four neighborhoods were from north Riyadh, four from the east, two from the south and three from the west.

The analysis was done using SPSS program.

Literature review:

There is a multiple, diverse set of studies that aimed at studying and analyzing the real estate market in many countries and in different periods of time in order to identify the most influential factors in the real estate market in multiple variables, the most important of which is the real estate price variable, and some of these studies can be shown as follows:

A study (Zhang, et al, 2023) aimed to predict the price levels of homes through the use of a machine learning and temporal regression model in order to analyze price trends in light of the interaction between land use and transportation, and in order to achieve the objective of the study, the collection of longitudinal data on the prices of house transactions was relied on by applying the regression model to analyze house price trends in the period from 2001 to 2016 in the Toronto and Hamilton area, and the study reached a set of results, the most important of which are is the ability of the machine learning regression model to achieve higher levels of accuracy and emphasize the moral impact of lag on house prices, the results of the study also confirmed the existence of a significant impact of economic, social factors and material conditions on house prices, and finally the study reached results confirming the significant differences in accessibility on house prices in the coming periods.

The main objective of the study (Zhang, 2021) was to analyze the fundamental factors affecting house prices by conducting a multiple linear regression model to predict house price levels by using the Spearman correlation coefficient in order to determine the significance of the factors affecting house prices for a set of data collected in Boston and by conducting multiple regression analysis, the study reached a set of results, the most important of which is the significance of the impact of factors associated with the proportion of low-income groups in the region. The property tax rate, the average crime rate in the city, the weighted average distance from home to employment centers in Boston, and the average number of rooms per home, it stressed that the results are the most significant factors in influencing house prices in the region and thus can be used to formulate various policies related to house prices and home demand more efficiently and effectively.

The study (Shao, 2022) aimed to study and analyze the process of forecasting house prices in Beijing by relying on the descriptive analytical approach to analyze

the reasons for the increase in house prices in Beijing and expectations of the rises of these prices in the coming periods, through a linear regression by collecting data and information on house prices in Beijing in the twentieth century and analyzing the impact of various factors on house price levels through a regression model. The most important of them was the significance of the variable of proximity to the city center in influencing price levels positively, as the closer the houses are to the city center, the more this leads to higher price levels, and also the results of the study confirmed the significant impact of transportation on house price levels in Beijing and through the results it was found that proximity to the city center and the availability of efficient transportation are the most influential factors in house price rises in Beijing.

The main objective of the study (Magid, 2018) is to analyze the basic factors explaining house prices in Iraq, specifically in the Mansour in Baghdad during the time period from 2000 to 2018 in order to predict the significance of each of these factors and its impact on the price levels of homes in the current period and in future price forecasts, and in order to achieve the goal of the study relied on the use of the case space method as one of the most appropriate statistical methods to achieve this goal of the study and that Compared to time series analyses, the study was adopted in analyzing the data reached through program E-views, the ninth version, and by analyzing the data, the study reached a set of results, the most important of which are the significant impact of the variables of the basic area of the house, the age of the house, the neighborhood, and the basic services close to the house in the positive impact on house prices in the current period, and the results reached related to forecasting house prices until 2025 indicated that price levels will witness continuous rises.

Many other studies have tried to focus on the most important factors that affect the price levels of homes in a particular geographical area and in a certain period of time, for example a study (Nahdi, 2015), which tried to show the impact of various factors on the decision to demand real estate in Jeddah by conducting a survey on a sample of 322 participants and based on statistical analyzes of the quantitative data obtained, the results were that the most influential factors in house prices in Jeddah in Saudi Arabia are factors related to the public services provided in the region, the customs and traditions prevailing in the region, the geographical location of the house and other factors; also (Al-Falih and Al-Faraidhi, 2017) in determining the most influential factors in price rises in the Kingdom of Saudi Arabia based on the descriptive analytical approach by collecting the required data the study found that the most influential factors in the rise in house prices are the factors associated with the imposition of taxes on real estate and the low level

of support provided to the real estate sector in the Kingdom of Saudi Arabia, and the results also confirmed that the population density and demand for real estate in Riyadh led to increase house price levels Finally, the results of the study indicated that there is a sensitivity of house prices in Riyadh to the variable availability of new and renewable energy sources in the region.

The objective of the study (Dahan, 2018) was to analyze the real estate market in Dubai in addition to predicting future price levels for real estate and future demand, and the study tried to achieve this objective by building a multiple regression model using the aggregate demand function, and the study reached a set of results, the most important of which is the rapid growth of demand for real estate in Dubai during the past seven years, as confirmed by the results of the study also notes that there is an expected rise in house prices in the coming years in Dubai although the results indicate that there is a non-significant demand for real estate in Dubai.

A study (Louati et al., 2021) also confirmed the case of a rise in the price levels of real estate in Riyadh, Saudi Arabia, based on the analysis of price levels in 5946 plots of land in northern Riyadh and using machine learning, the study tried to achieve an accurate prediction of the price levels of real estate in the coming periods, and the results of the study found an increase in the price levels of real estate in coming years.

AlZain, 2022 also aimed using the artificial neural network approach in estimating the accuracy of forecasting real estate prices in the Kingdom of Saudi Arabia that the level of accuracy of forecasting is low as a result of not taking into account all the factors that affect price levels and development in areas witnessing the establishment of new projects, and the results of the study also confirmed the significant importance of the level of impact of developments on house prices.

The study (Al Obaid, 2020) used the regression model to estimate the factors affecting the demand for real estate and its prices in the Kingdom of Saudi Arabia from 1987 to 2016, relying on variables such as the consumer price level index, population growth and housing loan applications in order to estimate the demand for real estate and its prices in the coming periods. The significant population growth in the demand for homes as well as the impact of villa locations and new developments in the region on the price levels of homes.

The current study is similar to previous studies in some points and differs in others, as there is a similarity in focusing on studying and analyzing house prices and predicting them in the future and analyzing the most factors affecting house prices in the coming period, and also there is similarity with many studies in the

use of the multiple regression model, but the current study differs from previous studies in more than one aspect, including the study's objective focuses on villa prices in particular with the analysis of house prices in general; In addition, the current study uses both multiple regression and simple regression and not only multiple regression, in addition to that the current study relies on recent data that explain the recent developments in the real estate sector in the Kingdom of Saudi Arabia, which contributes to improving the decision-making process associated with these estimates and forecasts for villa prices at the level of individuals and companies.

Results and discussion:

After cleaning up the data and excluding extreme values, the analysis of the 3,469 transactions was conducted using SPSS. Since our data volume is large, the requirement for normal distribution is not necessary in this case "with large enough sample size (> 30 or 40),the violation of the normality assumption should not cause major problems"(Pallant J, SPSS survival manual, 2007).

The variables in the study were as follows:

Dependent variable:

- Home price.

Independent variables:

- Area for the villa: The distances for the villas were taken from 200m2 to 1250m2 because the majority of transactions were within these spaces, and then

placed in categories in the program and reached 21 categories to facilitate descriptive analysis.

-availability of services: Available, unavailable.

-proximity of the neighborhood to new and large projects: close, far

Initially, descriptive analyses of variables were performed:

Table 1:Frequency of deals by regions.

		Frequency	Percent
Valid	East	921	26.5
	North	1235	35.6
	South	447	12.9
	West	866	25.0
	Total	3469	100.0

We note that in table (1) deals in northern Riyadh have had the largest share since 2019, and south of Riyadh was the least in demand, accounting for 13% of the transactions.

Table 2: Frequency of deals by services availability

		Frequency	Percent
Valid	Close	1567	45.2
	Far	1902	54.8
	Total	3469	100.0

Table 3: Frequency of deals by closeness

		Frequency	Percent
Valid	No	456	13.1
	Yes	3013	86.9
	Total	3469	100.0

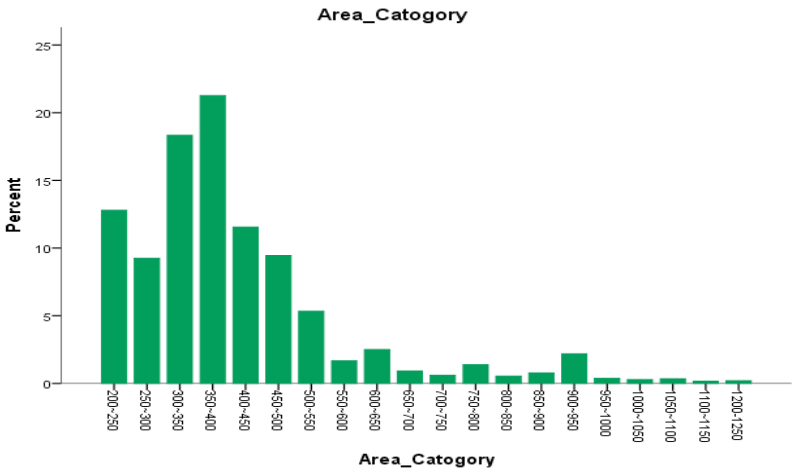

Figure 1: Frequency of Areas

It is clear to us from Figure (1) that the most purchased categories were from 200m2 to 450m2, and then began to decrease, which shows that the demand for large spaces has become very poor and consumer behavior has changed, due to higher prices compared to previous years when large villas were more in demand.

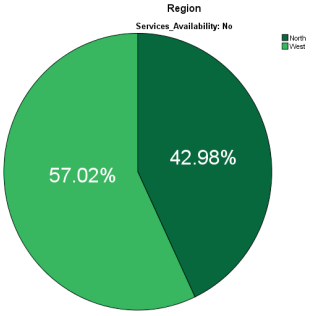

Figure 2:Percentage based on availability of services.

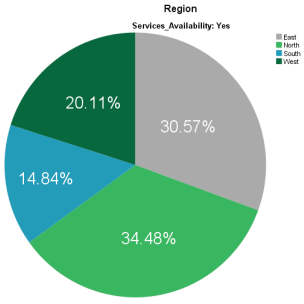

Figure 3:Percentage based on availability of services.

Services as in Figure (2) and (3) are available in most areas, but the lack of availability was in the north and west regions, especially the new neighborhoods of the north, due to the extension, renewal, and expansion of the neighborhoods in these areas.

Because most purchased houses in northern Riyadh were distributed between full-service neighborhoods and no-service ones, so we observed that the percentage of frequencies were approximate.

Table 4: Frequency of deals by and services availability

services Availability	Closeness			Frequency
No	Close	Valid	North	196
	Far	Valid	West	260
Yes	Close	Valid	East	332
			North	1039
			Total	1371
	Far	Valid	East	589
			South	447
			West	606
			Total	1642

Neighborhoods with services have the highest number of deals, for example, as in the table (4), There were 196 deals in the north of Riyadh and in not fully serviced neighborhoods, and in contrast, there were 1,039 deals in the north in full-service neighborhoods, as well as in western Riyadh there was a difference in the number of transactions in service-containing neighborhoods compared to the ones that doesn't have services by a difference of 346 deals.

This shows that the service factor is very necessary when choosing a home and clearly precedes the proximity factor to projects.

Table 5: Average prices for villas in all regions

Price	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
				Lower Bound	Upper Bound		
East	1794510.80	919546.328	30300.082	1735045.50	1853976.10	465000	7195339
North	2540314.68	1382486.168	39339.362	2463135.25	2617494.11	250000	10000000
South	1021778.94	468798.717	22173.418	978201.58	1065356.30	330000	5250000
West	1161864.13	549615.489	18676.695	1125207.19	1198521.07	155000	4300000

The average prices in north Riyadh for deals was as shown in table (5) 2,540,315 SAR is the highest price and in contrast, the lowest price was south of

Riyadh with an average price of 1,021,779 SAR. As we note in the table (6) the availability of services in neighborhoods increases average prices.

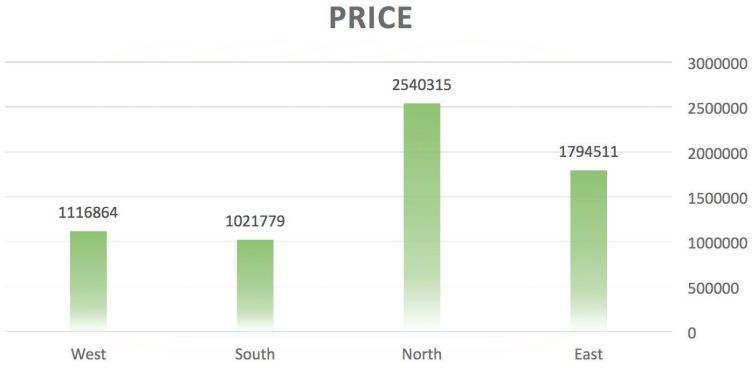

Figure 4: Average prices for villas in all regions.

Table 6: Average prices for villas according to availability of services

services Availability		Minimum	Maximum	Mean	Std. Deviation
No	Price	155000	8350000	1536512.94	975687.869
Yes	Price	301800	10000000	1842778.86	1198394.006

The Compare means test for price averages between the four regions was conducted to ensure before handling the data if they needed to be considered separate groups when analyzing:

Table 7:Test of Homogeneity of variances

		Levene Statistic	df1	df2	Sig.
Price	Based on Mean	154.171	3	3465	.000
	Based on Median	110.011	3	3465	.000
	Based on Median and with adjusted df	110.011	3	2337.030	.000
	Based on trimmed mean	128.342	3	3465	.000

Table 8: ANOVA:

Price	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1300232991881694.000	3	433410997293898.000	429.600	.000
Within Groups	3495740114828843.500	3465	1008871606011.210		
Total	4795973106710538.000	3468			

In table (7) and (8), since Sig < α (0.05) then it means rejecting the zero hypothesis and accepting the alternative hypothesis of differences between the averages of the four regions.

In this case, we need to conduct alternative tests that show the details of the groups, differences in them and the regions, from Post-Hoc multiple comparisons "Dunnett T3 test" as shown in the table (9).

Table 9:Multiple comparisons" Dunnett T3 test"

Dependent Variable: Price						
Dunnett T3						
(I) Region	**(J) Region**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval**	
					Lower Bound	**Upper Bound**
East						
	South	772731.855*	37546.710	.000	673807.39	871656.32
	West	632646.666*	35593.734	.000	538881.19	726412.14
North	East	745803.882*	49655.618	.000	615040.66	876567.11
	South	1518535.738*	45158.011	.000	1399588.76	1637482.71
	West	1378450.549*	43547.725	.000	1263749.06	1493152.04
West						
	South	140085.189*	28991.023	.000	63668.64	216501.73
*. The mean difference is significant at the 0.05 level.						

The table (9) shows significance differences, p-value < α (0.05) ,we note that the most differences were in favour of the north of Riyadh.

thus we're going to make a prices forecast with different equations for each region.

1. North of Riyadh:

Since all data for the proximity variable to projects are "close", we exclude this variable, and the relationship is as follows:

Dependent variable: Home price.

Independent variables: - Space for the villa. - availability of all services in the neighborhood.

Table 10: Model summary for North of Riyadh..

Model Summary[b]				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
1	861[a].	741.	741.	498007.282
a. Predictors: (Constant), services Availability, Area				
b. Dependent Variable: Price				

Table (10) shows that there is a semi-strong relationship between variables hence

$R^2 = 0.741$, which means that 74% of price changes in northern Riyadh depend on the space for the villa and provide services around it, and 26% of price changes are affected by other variables such as villa age, finishing type, internal additions from the availability of a swimming pool or elevator and other necessary variables that the buyer is interested in.

Table 11: Cofficients of North

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	98376.646	50960.956		1.930	.054	-1612.662-	198365.954
	Area	5611.735	99.355	.860	56.482	.000	5416.794	5806.677
	Services Availability	25053.135	39938.566	.010	.627	.531	-53309.397-	103415.668
a. Dependent Variable: Price								

From table (11), The area variable and because its quantitative variable of independent variables, Sig < α (0.05), and indicating its impact and association with the dependent variable.

The variable availability of services as it is of a nominal type that contained «yes» and «no» and was encoded using 0= No, 1= Yes.

From the previous table the difference between the villa located in a neighborhood containing services and which in a neighborhood that does not contain any is 25,053.135 SAR.

So, the linear regression for north of Riyadh is as follows:

Price = 98376.646 + 5611.735(Area) + 25053.135 (Services Availability: Yes=1 or No=0)

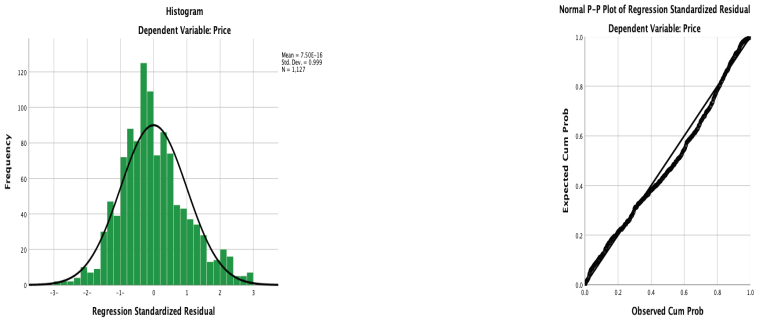

Figure 5:Histogram and Normal P-P plot of regression standardized residual "North"

The figure (5) shows the distribution of residuals for prices as it follows the normal distribution in the north of Riyadh.

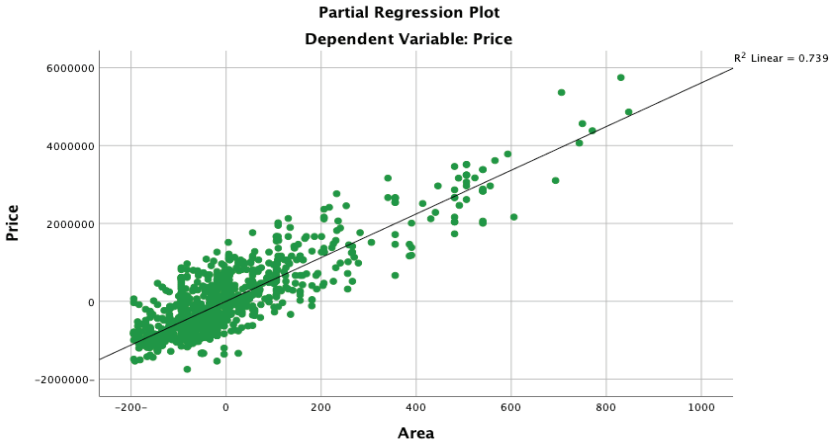

Figure 6: The relationship between the variables. "North"

Figure (6) revealed the linear relationship between areas and prices

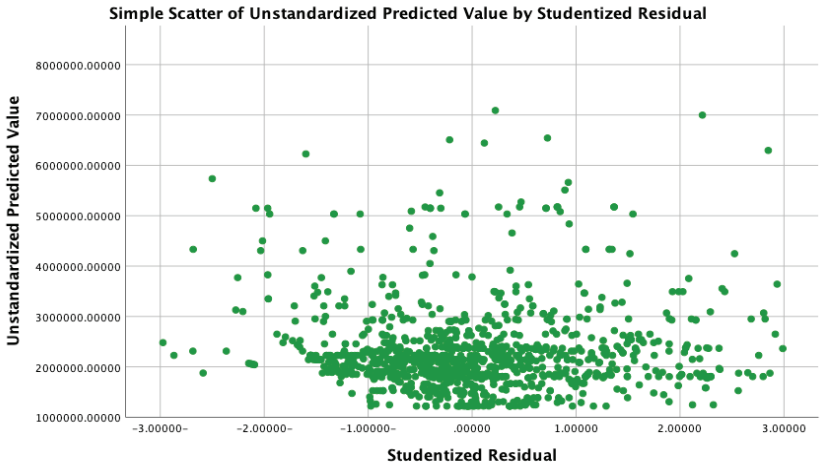

Figure 7: Homoscedasticity "North".

The condition of the stability of the variance as in figure (7), data is widespread, which results in having homoscedasticity.

Prices forecast for north Riyadh:

When compensated manually with the aforementioned equation, assuming that the desired area is 400m2, and in a neighborhood with full services, the expected price will be 2,368,124 SAR, and when full services are not available in the neighborhood the expected price will be 2,343,070 SAR, it is approximately 25,054, SAR for the neighborhood where services are available.

Price = 98376.646 + 5611.735(400) + 25053.135 (1)

Price = 98376.646 + 5611.735(400) + 25053.135 (0)

When using the program to predict, the results will appear as in the following table:

Assuming that the desired area was 400m2, and in a fully serviced neighborhood:

Table 12: Prediction result "North, in a fully serviced neighborhood"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		2368123.973
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		2368123.973
	Std. Error		16261.721
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	2336217.228
		Upper Bound	2400030.717
a. Based on the user-specified contrast coefficients (L›) matrix number 1			

The confidence interval in table (12), has shown that the villa will be priced at a minimum of 2,336,217 SAR, and a maximum of 2,400,030 SAR.

The expected value as indicates by the equivalent compensation is the same as table (16) 2368123.973 SAR.

Assuming that the desired area was 400m2, and in a neighborhood where full services are not available:

Table 13: Prediction results "North, in a neighborhood where full services are not available"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		2343070.837
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		2343070.837
	Std. Error		36539.998
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	2271376.556
		Upper Bound	2414765.119
a. Based on the user-specified contrast coefficients (L›) matrix number 1			

Table (13) has indicated the confidence interval, the villa will be priced at a minimum of 2,271,377 SAR and a maximum of 2,414,765 SAR.

The expected value as shown by the equivalent compensation is the same as shown in the table 2343070.837 SAR.

2. East of Riyadh:

Since all the deals in the east of Riyadh were in neighborhoods that all services are available, so we exclude this variable, and the relationship is as follows:

Dependent variable: Home price.

Independent variable: - Area for the villa. - Near the neighborhood of new and large projects.

Table 14: Model Summary for East of Riyadh.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	850[a].	723.	722.	344219.776
a. Predictors: (Constant), Closeness, Area				
b. Dependent Variable: Price				

Table (14), the R2 = 0.723 shows that there is a semi-strong relationship between variables, which means that 72% of price changes in eastern Riyadh depend on the area of the villa and its proximity to new projects, and 28% of price changes are affected by other variables such as the villa age, finishing type, internal additions from the availability of a swimming pool or elevator and other necessary variables of interest to the buyer when choosing the villa.

Table 15: Cofficients of East

Coefficients[a]						
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	-3376.755	37004.947		-.091	.927
	Area	3734.601	82.721	.807	45.147	.000
	Closeness	437158.297	24453.433	.320	17.877	.000
a. Dependent Variable: Price						

From table (15), The Area variable and its quantitative variable of independent variables Sig < 0.05, indicating its impact and its association with the dependent variable.

therefore, the variable of proximity to the new projects being of a nominal type contained near and far and encoded by 1 = Close, 0 = Far.

the previous table has shown the difference between the villa located in a neighborhood close to the new projects and the villa away from the projects is 437,158 SAR, so the villa far from the projects is less than the villa close to the projects.

Thus, the linear regression for eastern Riyadh is as follows:

Price = -3367.755 + 3734.601 (Area) + 437158.297 (Closeness: Far =0 or Close =1)

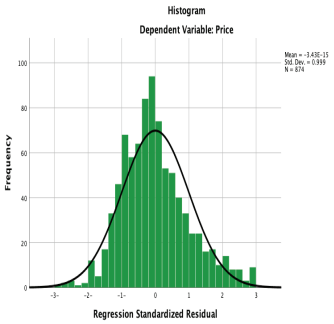

Figure 8: The relationship between variables "East"

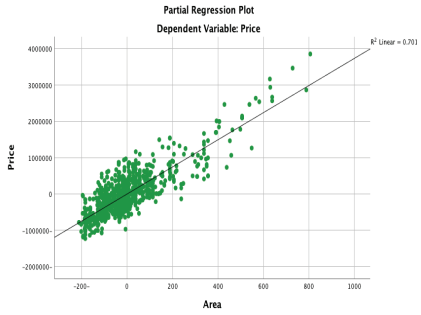

Figure 9: Histogram of regression standardized residual "East"

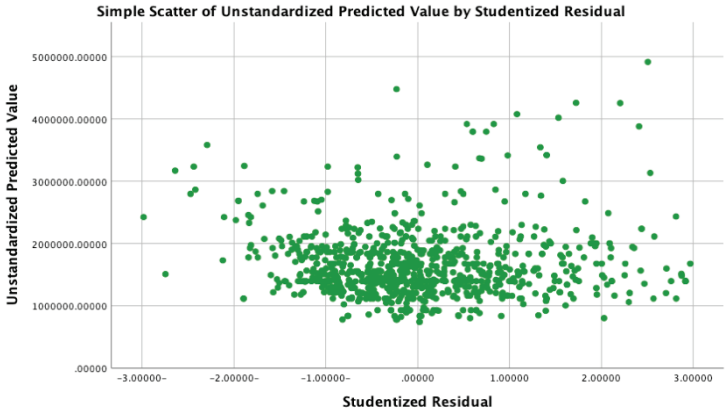

Figure 10: Homoscedasticity "East"

Figure (8) has indicated normal distribution curve for the residuals in the East area and figure (9) as well has shown the linear relationship between quantitative variables.

The condition of the stability of the variance as in figure (10), demonstrates that it is acceptable.

Prices forecast for eastern Riyadh:

When compensated manually with the aforementioned equation, assuming that the desired area is 400m2, and in a neighborhood close to the projects, the expected price will be 1,927,622 SAR, and in the current desire for a villa away from the projects will be the expected price of 1,490,464 SAR.

Price = -3367.755 + 3734.601 (400) + 437158.297 (1)

Price = -3367.755 + 3734.601 (400) + 437158.297 (0)

When you use the program to predict, the results will appear as in the following table:

Assuming that the desired area was 400m2, and in a neighborhood close to the new projects:

Table 16: Prediction results "East, in a neighborhood close to the new projects"

Contrast Results (K Matrix)a			
Contrast			Dependent Variable
			Price
L1	Contrast Estimate		1927622.140
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		1927622.140
	Std. Error		19683.726
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	1888989.062
		Upper Bound	1966255.218
a. Based on the user-specified contrast coefficients (L›) matrix number 1			

Table (16) demonstrated in the confidence interval, the villa will be priced at a minimum of 1,888,989 SAR and a maximum of 1,966,255 SAR.

The expected value as shown by the equivalent compensation is the same as that of 1927622.140 SAR.

Assuming that the desired area was 400m2, and in a neighborhood far from new projects:

Table 17: Prediction results "East, in a neighborhood far from new project"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		1490463.843
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		1490463.843
	Std. Error		14476.462
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	1462051.017
		Upper Bound	1518876.668
a. Based on the user-specified contrast coefficients (L') matrix number 1			

Table (17), in the confidence interval described, the villa will be priced at a minimum of 1,462,051 SAR, and a maximum of 1,518,877 SAR.

The expected value as shown by the equivalent compensation is the same as the one in the schedule 1490463.843 SAR.

3 . West of Riyadh:

Since all the deals in western Riyadh were in neighborhoods relatively far from the new projects, we exclude this variable, and the relationship is as follows:

dependent variable: Home price.

Independent variables: - area for the villa. – If full services are provided in the neighborhood.

Table 18: Model summary for West of Riyagh

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.764[a]	.584	.583	286649.645
a. Predictors: (Constant), services Availability, Area				
b. Dependent Variable: Price				

Table (18), the R2 = 0.584 shows that there is a relationship that tends to decline between variables, which means that about 59% of price changes in western Riyadh depend on the area of the villa and its proximity to new projects, and 41% of price changes are affected by other variables such as the villa age, finishing type, internal additions from the availability of a swimming pool or elevator and other necessary variables of interest to the buyer when choosing the villa.

This also shows that the area factor in particular does not have much impact on prices in the west, unlike the north and east regions, and may be due to the level of prices in this area, most of the transactions that were made were based on the neighborhood of Laban, Namar and Duraihimiyah, and perhaps when prices start to decline, buyers are less interested in the area factor, and attention is paid to many other factors.

Table 19: Cofficients of West

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	18190.030	35724.080		.509	.611	-51932.144-	88312.204
	Area	2153.161	64.304	.824	33.484	.000	2026.940	2279.382
	services Availability	420336.166	23862.909	.434	17.615	.000	373496.078	467176.255
a. Dependent Variable: Price								

From table (19), The Area variable and its quantitative variable of independent variables Sig <0.05, indicating its impact and its association with the dependent variable.

And the variable provides full services in the neighborhood because it is of a nominal type that contained yes and no and was encoded by 0 = No,1= Yes.

From the previous table, the difference between a villa located in a fully serviced neighborhood and a villa in a neighborhood without services is 420,336 SAR, when services are available that increase the price at the said value.

The linear regression for western Riyadh is therefore as follows:

Price = 18190.030+ 2153.161 (Area) + 420336.166 (Services Availability: Yes=1 or No=0)

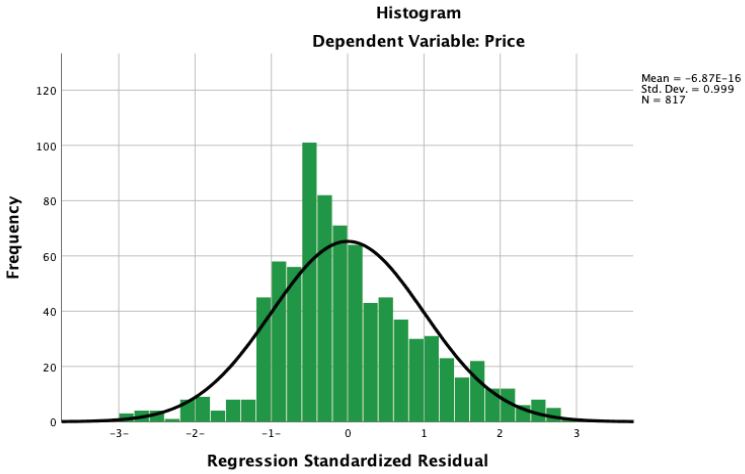

Figure 11: Histogram of regression standardized residuals «West»

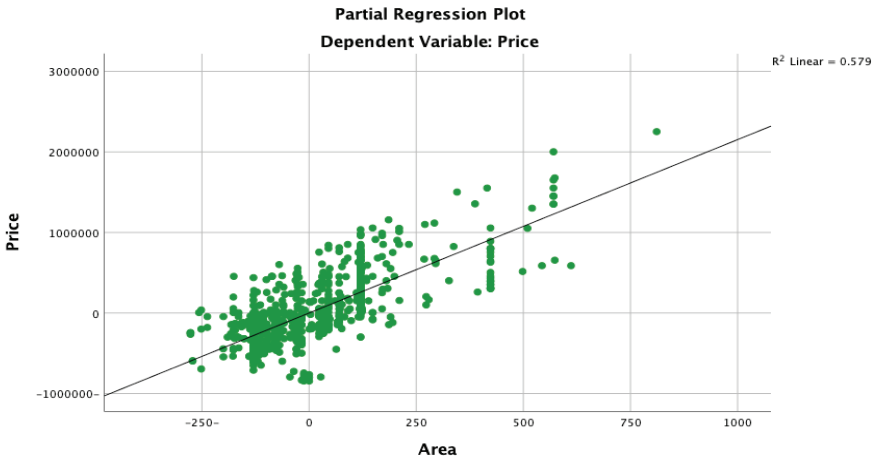

Figure 12: the relationship between variables "West"

Figure (11) Shows normal Distribution of residuals in the west. And figure (12) demonstrates the linear relationship between quantitative variables.

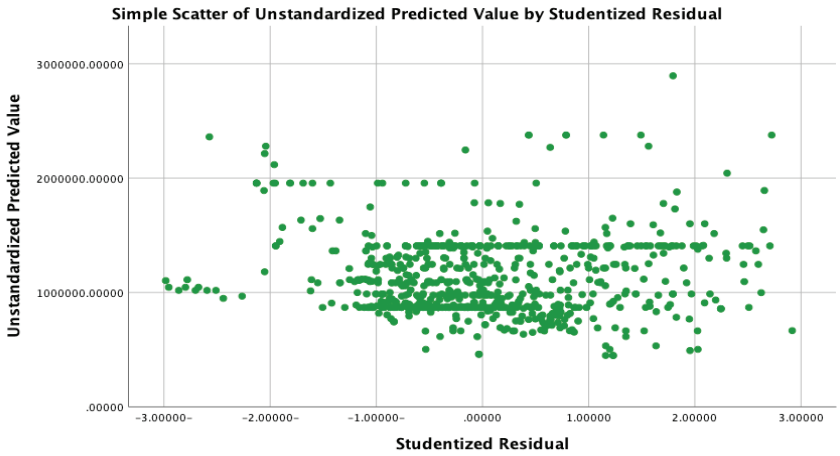

Figure 13: Homoscedasticity «West».

The condition of variance stability in figure (13) indicates that the data is widespread, which results in having Homoscedasticity.

Prices forecast for western Riyadh:

When compensated manually by the aforementioned equation, assuming that the desired area is 400m2, and in a neighborhood with full services, the expected price will be 1,299,791 SAR, and when full services are not available in the neighborhood the expected price will be 879,454 SAR. it is approximately 420,337 SAR for the neighborhood where services are available, and it is a big difference and a strong influence on the price of the villa.

Price = 18190.030+ 2153.161 (400) + 420336.166 (1)

Price = 18190.030+ 2153.161 (400) + 420336.166 (0)

When using the program to predict, the results will appear as in the following table:

Assuming that the desired area was 400m2, and in a fully serviced neighborhood:

Table 20: Prediction results "West, in a fully serviced neighborhood"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		1299790.499
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		1299790.499
	Std. Error		12799.639
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	1274666.311
		Upper Bound	1324914.687
a. Based on the user-specified contrast coefficients (L') matrix number 1			

Table (20), in the confidence interval described, the villa will be priced at a minimum of 1,274,666 SAR, and a maximum of 1,324,915 SAR.

The expected value as shown by the equivalent compensation is the same as the one in the schedule 1299790.499 SAR.

Assuming that the desired area was 400m2, and in a neighborhood where full services are not available:

Table 21: Prediction results "West, in a neighborhood where full services are not available"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		879454.333
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		879454.333
	Std. Error		19001.289
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	842157.033
		Upper Bound	916751.632
a. Based on the user-specified contrast coefficients (L') matrix number 1			

Table (21), in the confidence interval described, the villa will be priced at a minimum of 842,157 SAR and a maximum of 916,752 SAR.

The expected value as shown by the equivalent compensation is the same as the one in the schedule 879,454,333 SAR.

4. South of Riyadh:

Since all the deals in the south of Riyadh were in neighborhoods far from the new projects, and all services are available as they are old neighborhoods, so we exclude these variables, and the relationship is simple as follows:

Dependent variable: Home price.

Independent variable: - area for the villa.

Table 22:Model summary for South of Riyadh

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.734a	.538	.537	185086.490
a. Predictors: (Constant), Area				
b. Dependent Variable: Price				

Table (22), the R2 = 0.538 shows that there is a relationship that tends to decrease between variables, which means that about 54% of price changes in southern Riyadh depend on the area of the villa, and 46% of price changes are affected by other variables such as villa age, finishing type, internal additions from the availability of a swimming pool or elevator and other necessary variables of interest to the buyer when choosing the villa.

This also shows that the area factor does not have much impact on prices in the south as in the west, unlike in the north and east, and may be due to the level of prices in these areas, since when prices begin to decline, the buyer's interest in the area factor is reduced, and attention is paid to many other factors.

Table 23: Coefficients of South

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	367206.439	28722.362		12.785	.000	310748.951	423663.926
	Area	1727.953	78.096	.734	22.126	.000	1574.445	1881.460
a. Dependent Variable: Price								

Table (23) demonstrated the Area variable and its quantitative variable of independent variables Sig < 0.05, indicating its impact and its association with the dependent variable.

The linear regression for South Riyadh is therefore as follows:

Price = 367206.439+ 1727.953 (Area)

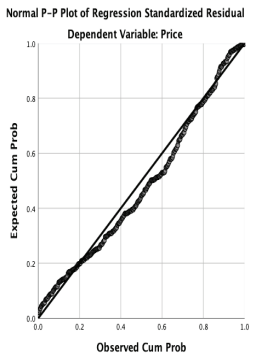

Figure 14: Normal P-P plot

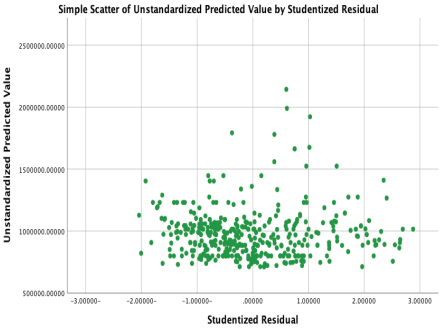

Figure 15: Homoscedasticity «South»

Figure (14) shows the distribution of the residual in the south of Riyadh, and in figure (15) there is a spread of data that is acceptable.

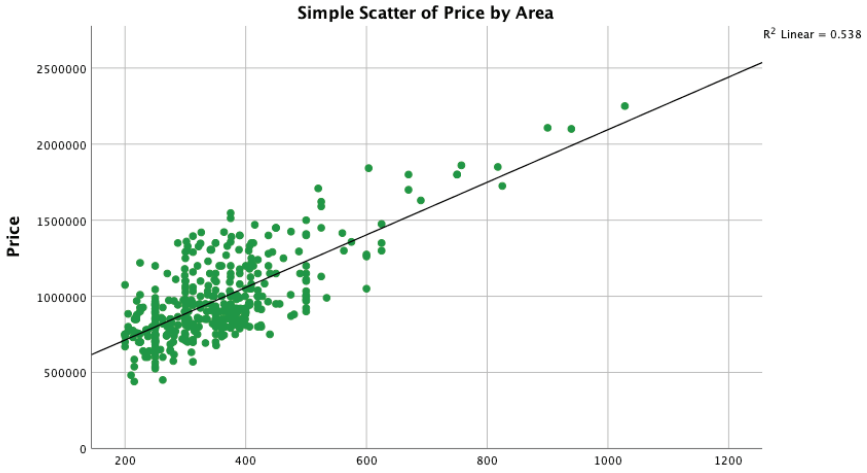

Figure 16: The relationship between variables "South"

In figure (16) the linear relationship between variables appears.

Prices forecast for south of Riyadh:

When compensated manually with the aforementioned equation, assuming that the desired area is 400m2, the expected price will be 1,058,387 SAR.

Price = 367206.439+ 1727.953 (400)

When using the program to predict, the results will appear as in the following table:

Assuming that the desired area was 400m2:

Table 24: Prediction results "South"

Contrast Results (K Matrix)[a]			
Contrast			**Dependent Variable**
			Price
L1	Contrast Estimate		1058387.473
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		1058387.473
	Std. Error		9844.044
	Sig.		.000
	95% Confidence Interval for Difference	Lower Bound	1039037.743
		Upper Bound	1077737.204
a. Based on the user-specified contrast coefficients (L') matrix number 1			

Table (24), in the confidence interval described, the villa will be priced at a minimum of 1,039,038 SAR and a maximum of 1,077,737 SAR.

The expected value as shown by the equivalent compensation is the same as the one in the schedule 1058387.473 SAR.

After viewing each area separately, we will collect them and predict prices for the upper limit in SR using SPSS for multiple areas then compare them as follows:

Table 25: Prediction equations

North of Riyadh: Price = 98376.646 + 5611.735(Area) + 25053.135 (Services Availability: Yes=1 or No=0)
East of Riyadh: Price = -3367.755 + 3734.601 (Area) + 437158.297 (Closeness: Far =0 or Close =1)
West of Riyadh: Price = 18190.030+ 2153.161 (Area) + 420336.166 (Services Availability: Yes=1 or No=0)
South of Riyadh: Price = 367206.439+ 1727.953 (Area)

Table26: Forecasts for all regions with different areas

Areas in square meters	North		East		West		South
	full services are provided in the neighborhood	full services are not available in the neighborhood	Neighborhood far from new projects	Neighborhood Close to new projects	full services are provided in the neighborhood	full services are not available in the neighbor-hood	
200	1,295,289	1,298,498	788,117	1,230,552	897,803	498,990	741,751
250	1,568,885	1,575,719	968,929	1,412,589	1,002,384	602,479	822,553
300	1,843,757	1,854,108	1,150,670	1,595,697	1,108,277	706,501	904,847
350	2,120,573	2,133,774	1,333,808	1,780,172	1,215,777	811,205	989,700
600	3,541,704	3,550,691	2,279,056	2,725,619	1,771,853	1,349,328	1,446,353
800	5,161,665	5,160,651	3,053,402	3,497,842	2,224,900	1,795,154	1,820,997
1250	7,307,931	7,300,648	4,803,872	5,246,256	3,248,492	2,813,688	2,666,554

Note: the prices mentioned in table (26) do not cover the value added tax and the commission of the real estate marketer.

Figure (17) demonstrates prices of villas between regions for an area of 300m2, and figure (18) demonstrates prices of villas between regions for an area of 1250m2.

Figure 17: Average prices of villas for an area of 300m2

Figure 18:Average prices of villas for an area of 300m2

It came to our attention that most of the deals that took place from 2019 till the first quarter of 2022 were in the north of Riyadh which shows that people tend to buy in this region. Additionally north region prices are high compared to other areas due to many reasons, most importantly, most of new projects are in this area, structure and streets are new, additionally the extension that took place in north and new neighborhoods made the demand higher thus it has led to higher prices.

when comparing current prices with former years or even former months we will notice the rapid difference growing, and this increase is also expected to continue to grow in near future for the North region specifically. Therefore, the estimate equation for northern Riyadh may not be effective and realistic for the

coming period due to the reasons mentioned, as it is a sensitive and constantly changing region.

The area factor also changes its importance and relevance from region to region as in the north and east of Riyadh it is associated and affects prices significantly, but the less correlation and impact in the south and west of Riyadh, which means that the higher the price, the less interest in large areas. therefore, the attention is paid to other factors.

Conclusion and recommendation:

Based on the results of this study, Riyadh seems like one of the exceptional cities where it is difficult to predict the prices of real estate, especially in recent times and with the developments in it and the population density and high demand for ready-made villas, especially north of Riyadh area prices are constantly changing towards the rise, so when analysis needs to be kept up with each period by updating the equations according to the prices of the deals that have been made at the time.

It is important to widen the range of the study in the future to include apartments as well, comparing the demand for them and for houses, and predict future trends.

Also, the population must be considered as one of the factors affecting the increase in prices, future studies must be conducted in this regard, especially as we mentioned earlier that one of the goals of the Kingdom's vision is to increase the population of Riyadh almost double, so it is expected that real estate and its prices will be significantly affected.

The results of the study also support many of the results of previous studies, as it is noted that real estate prices in Riyadh are constantly rising like what happens in Dubai, and that the Kingdom's Vision 2030 has a significant impact on the real estate field.

People need more of these studies that help them understand the situation, prices, and future of real estate to make it easier for them to make their decisions appropriately, especially since this field has a lot of uncertainty and a lot of contradictory expectations.

A significant recommendation to be mentioned is that it is important when the competent authorities provide the prices and areas of deals to provide the interested person with details about the property in terms of the age of the villa sold, the type of finishing in it, and the advantages contained in the villa from the presence of a swimming pool or elevator and others, all these information are

considerable for decision-making and the use of data in analysis to make good decisions, and because when information are partially available it is difficult for the buyers to make their decision considering that many influences that make a difference in price are hidden,

through the data, we might encounter many villas in the same region, area, and neighborhood, but prices vary very much, it is important to know these factors to be able to analyze and benefit from the data to make accurate and realistic results. Finally, the information mentioned in this study can be used in the future when conducting research to compare prices for real estate in Riyadh while considering that the neighborhoods in central Riyadh were excluded because there were not enough deals. and see how much the prices has changed over the years.

A comparison between geographic areas within the cities of Saudi Arabia and neighboring countries, along with working to provide the necessary statistical information for planning and analysis, and building a database on property, land prices, and rental rates.

The information presented in this study can be useful for future research into the real estate market in Riyadh. By comparing prices over the years, researchers can gain insights into how the market has changed and what factors have contributed to these changes.

References

Abdelgalil, E., & Bakheet, B. A. (2007). Dubai Real Estate Sector. Data Management and Business Research Department. Retrieved from https://www.researchgate.net/publication/242417789_Dubai_Real_Estate_Sector

Al Obaid, H. M. A. (2020). FACTORS DETERMINING HOUSING DEMAND IN SAUDI ARABIA. International Journal of Economics and Financial Issues, 10(5), 150–157. https://doi.org/10.32479/IJEFI.10262

Al-Falih, Saad bin Muhammad, and Al-Faraidhi, Al-Bashir Misbah. (2017). The Kingdom's vision 2030 and the correction of the situation of the Saudi real estate market. Journal of Economic and Financial Research: University of Larbi Ben Mhidi Oum El Bouaghi - Faculty of Economic, Commercial and Management Sciences, Volume 4, Volume 2, 98 - 124. Retrieved from:

Ali, Zina. (2021, November 27). Buying apartments is 15 times more than villas in 3 years. Al-Watan Newspaper. Retrieved from:

Al-Nahdi, T. S. (2015). Factors Influencing Purchase Intention of Real Estate in Jeddah, Saudi Arabia with the Moderating Effect of Public Service Location and Purchaser's Demographics. Retrieved from http://eprints.usm.my/30589/1/TAWFIK_SALAH_AL-NAHDI.pdf

Alzain, E., Alshebami, A. S., Aldhyani, T. H. H., & Alsubari, S. N. (2022). Application of Artificial Intelligence for Predicting Real Estate Prices: The Case of Saudi Arabia. Electronics (Switzerland), 11(21). https://doi.org/10.3390/ELECTRONICS11213448

CNN Arabic. (2021, January 28) Mohammed bin Salman: Riyadh will be one of the largest city economies in the world. Retrieved from:

Dahan, A. A. (2018). The Future of The Real Estate Industry of Dubai: The Demand for Real Estates. Journal of Global Economic.6(4) Retrieved from https://www.researchgate.net/publication/330500443_The_Future_of_the_Real_Estate_Industry_of_Dubai_The_Demand_for_Real_Estates_2000-2020

Donghao Shao. (2022). Prediction of Housing Price in Beijing Based on Linear Regression Model. BCP Business & Management, Volume 32. Prediction of Housing Price in Beijing Based on Linear Regression Model | BCP Business & Management (bcpublication.org)

Hilbe, Joseph M.(2016). Practical guide to logistic regression. CRC press/Taylor & amp; Francis.

http://search.mandumah.com.sdl.idm.oclc.org/Record/984390

https://arabic.cnn.com/business/article/2021/01/28/mbs-saudi-riyadh-future-investment-initiative

https://www.alwatan.com.sa/article/1093450

https://www.moj.gov.sa/ar/pages/default.aspx Ministry of justice.

Louati, A., Lahyani, R., Aldaej, A., Aldumaykhi, A., & Otai, S. (2022). Price forecasting for real estate using machine learning: A case study on Riyadh city. Concurrency and Computation: Practice and Experience, 34(6), e6748. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6748

Magid, Ragaa Kamel. (2018). Use a state space model and forecast house prices in Baghdad. Journal of Economics and Administrative Sciences 24(108). SSM (emarefa.net)

Menard, Scott. (1995). Applied logistic regression analysis. Sage Publications.

Ministry of Municipal & Rural Affairs & Housing. https://gisapps.momra.gov.sa/locator/EWC_locator.aspx

Pallant, J. (2020). SPSS survival manual: A step by step guide to data analysis using IBM SPSS. Routledge. Retrieved from https://www.taylorfrancis.com/books/mono/10.4324/9781003117452/spss-survival-manual-julie-pallant

Zhang, Qingqi (2021). Housing Price Prediction Based on Multiple Linear Regression. Hindawi Scientific Programming, Volume 2021. Longitudinal modelling of housing prices with machine learning and temporal regression l Emerald Insight

Zhang, Y., Rahman, A. and Miller, E. (2023), “Longitudinal modelling of housing prices with machine learning and temporal regression”, International Journal of Housing Markets and Analysis, Vol. 16 No. 4, pp. 693-715. https://doi.org/10.1108/IJHMA-02-2022-0033

Zhao, Y., Ravi, R., Shi, S., Wang, Z., Lam, E. Y., & Zhao, J. (2022). PATE: Property, Amenities, Traffic and Emotions Coming Together for Real Estate Price Prediction. 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/DSAA54385.2022.10032416