Real Estate Los Angeles Data Analysis

Author:
Mayah Bosworth

Table of Contents

Background

Objective

To build statistical models that help real estate companies maximize profit along with visual analysis to help interpret a data set. Specifically, which types of property factors affect property price and which property factors affect property performance.

Project Summary

The motivation behind my project is to understand real-estate market trends and how real-estate agencies can gain a competitive advantage. Some of the important insights companies focus on are location-based demand, customer service advantages, and tailored property segments. Location based-demand helps agencies focus on understanding the specific needs and desires of buyers in particular neighborhoods or areas. This helps real-estate agencies better meet their clients' needs.[1] Following up on customer inquiries, providing personalized attention, and ensuring customer satisfaction helps agencies stand out amongst their competitors.[2] Overall, real estate agencies that can stay on top of evolving market trends and adapt to evolving buyer preferences are better positioned for success in the competitive real estate industry.

---

[1] Koutroumanis, Dean A. "Developing Competitive Advantage for Real Estate Agents: Links to Organizational Culture, Service Quality and Behavioral Intentions." Journal of Applied Business & Economics, vol. 23, no. 2, Feb. 2021, pp. 45–53. EBSCOhost, https://doi.org/10.33423/jabe.v23i2.4085.

[2] Zhang, Xiuzhi, et al. "Online Advertising and Real Estate Sales: Evidence from the Housing Market." Electronic Commerce Research, vol. 23, no. 1, Mar. 2023, pp. 605–22. EBSCOhost, https://doi.org/10.1007/s10660-022-09584-2.

Data Processing

Data Collection

The data set provides information regarding properties in Los Angeles and their market performance. There is also specific information on the real estate agencies of each property. The data set was provided by Professor Ghasemi at Loyola Marymount University.

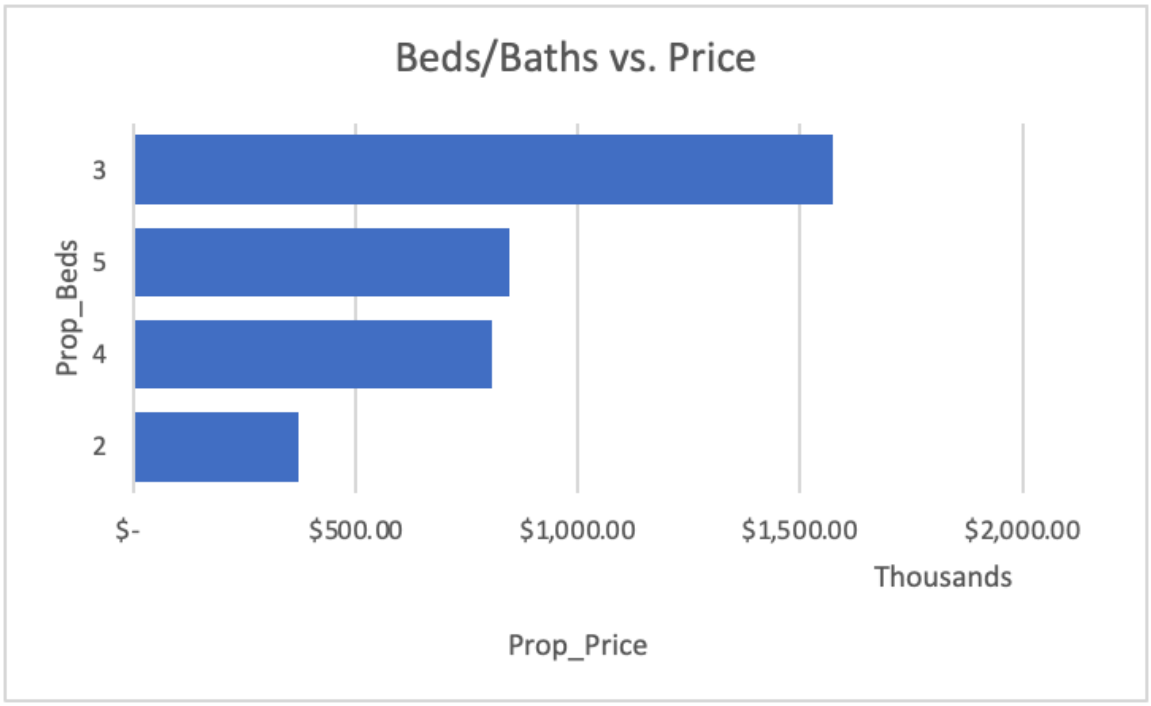
Data Cleaning

- Found no duplicate rows
- Found no blank rows
- Created copy of data set to disregard unnecessary rows that were not used in our regression analysis
    - Observation number, Property State, Listing Agent Phone Number, LC Address, LC State, LC Zip
- Assigned values to performance
    - Excellent = 4 (made 71 replacements)
    - Very good = 3 (68 replacements)
    - Good = 2 (62 replacements)
    - Fair = 1 (59 replacements)
- Changed some property details to numeric values as opposed to general values
    - Number of beds, number of baths, house size, and lot size

Data Analysis

Specifics

Multi-regression model determining which number of combined beds and baths has the highest average property price:



Prop_Baths	3	
Prop_Beds	**Average of Prop_Price**	
3	$	1,571,475.76
5	$	845,857.14
4	$	804,625.03
2	$	369,180.00
Grand Total	**$**	**1,112,749.68**

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.521523795							
R Square	0.271987069							
Adjusted R Square	0.266321599							
Standard Error	847310.6016							
Observations	260							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	6.89331E+13	3.44665E+13	48.00785385	1.92696E-18			
Residual	257	1.84509E+14	7.17935E+11					
Total	259	2.53442E+14						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	315354.9675	166549.1372	1.893465033	0.059419254	-12619.83841	643329.7733	-12619.83841	643329.7733
X Variable 1	-118143.0506	69610.39709	-1.69720409	0.090868691	-255222.4547	18936.35343	-255222.4547	18936.35343
X Variable 2	429385.4713	54530.76035	7.874188229	9.64617E-14	322001.4533	536769.4892	322001.4533	536769.4892

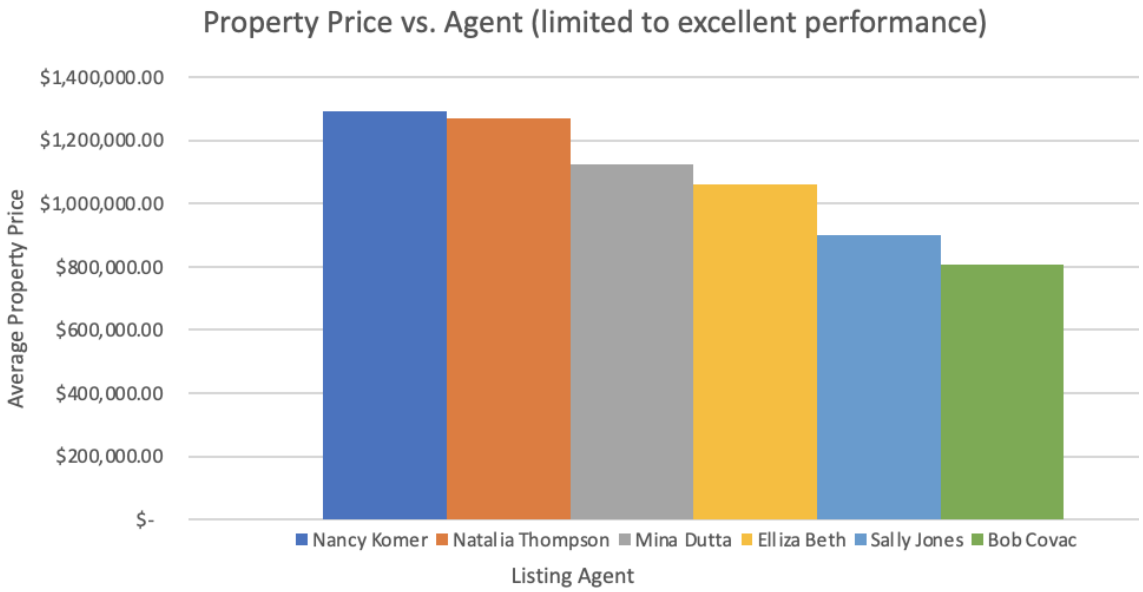
The most profitable property is found with 3 bedrooms and 3 bathrooms, with an average property price of $1,571,475.76. The Multiple R is 0.52 which indicates there is a moderate positive relationship between the number of beds, number of baths, and property price. In the regression, 27% of the variability found in the property price explains the input variables, number of property beds and baths. In addition, the Significance F of $1.927E^{-18}$ shows the relationship between the variables is statistically significant. However, with the adjusted R Squared being 0.266, running a multiple regression will not provide any additional value to our model.

Pivot table determining which cities have the highest average property prices:

Property_City	Average of Prop_Price
Venice	$ 1,923,847.83
West Los Angeles	$ 1,056,499.95
Westchester	$ 993,302.03
Diamond Bar	$ 694,795.46
Grand Total	**$ 1,052,580.62**

The order of highest to lowest prices in terms of location include Venice, West Los Angeles, Westchester, and Diamond Bar. Venice has an average property price of $1,923,847.83.

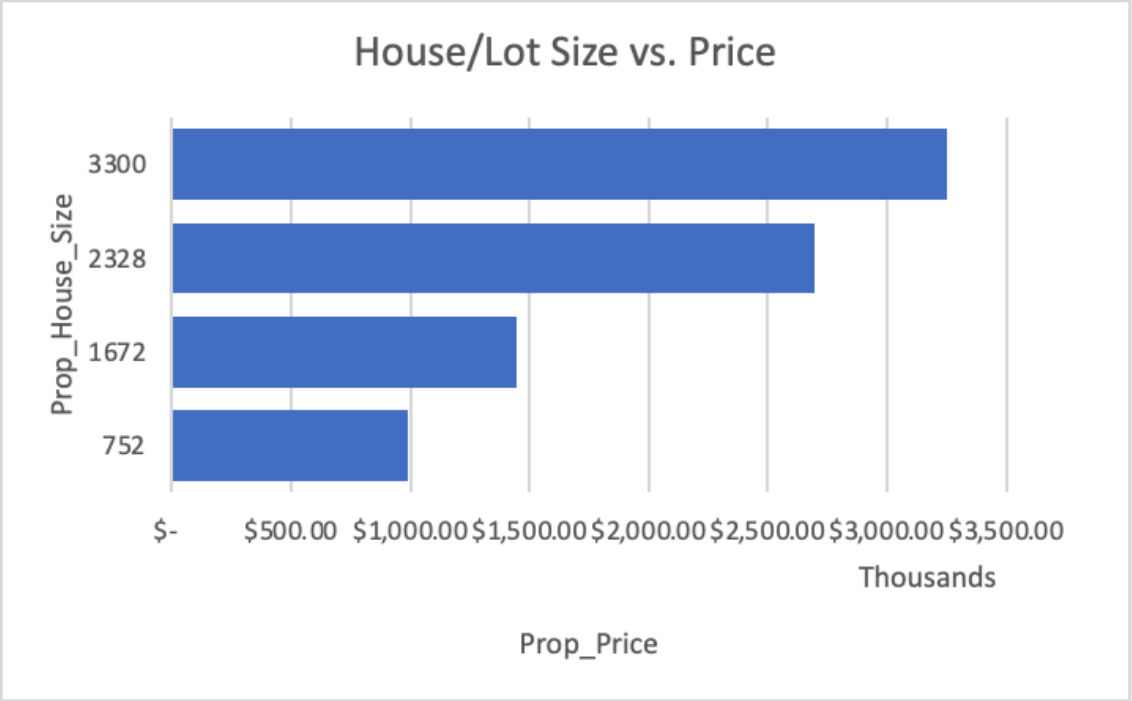
Pivot table determining which agent makes the most revenue and have the highest average performance rates (for homes in the excellent performance category):



Average of Prop_Price	Listing_Agent						
LA_Performance	Nancy Komer	Natalia Thompson	Mina Dutta	Elliza Beth	Sally Jones	Bob Covac	Grand Total
Excellent	$ 1,294,888.80	$ 1,269,163.55	$ 1,126,354.55	$ 1,059,970.64	$ 900,273.53	$ 806,327.18	$ 1,058,218.49
Grand Total	**$ 1,294,888.80**	**$ 1,269,163.55**	**$ 1,126,354.55**	**$ 1,059,970.64**	**$ 900,273.53**	**$ 806,327.18**	**$ 1,058,218.49**

Here the amount of revenue each agent brought in when selling a home with an excellent performance rating is being analyzed. This helped identify that Nancy brought in the most revenue at the highest performance.

Multi-regression determining which lot sizes and house sizes have the highest average property prices:



Prop_Lot_Size	2700

Prop_House_Size	Average of Prop_Price
3300	$ 3,250,000.00
2328	$ 2,697,000.00
1672	$ 1,449,000.00
752	$ 989,000.00
Grand Total	**$ 2,096,250.00**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.683244419
R Square	0.466822936
Adjusted R Square	0.462673698
Standard Error	725118.2293
Observations	260

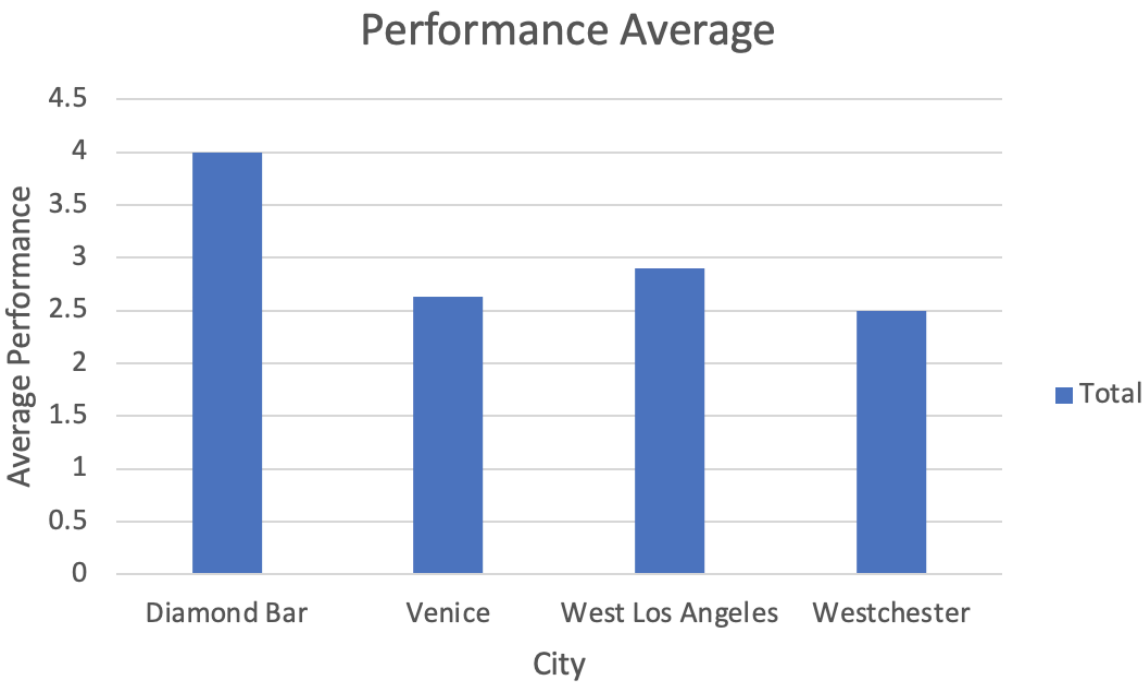
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.18313E+14	5.91564E+13	112.508116	7.998E-36
Residual	257	1.3513E+14	5.25796E+11		
Total	259	2.53442E+14			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	421.9352129	84404.81049	0.004998947	0.99601531	-165791.182	166635.053	-165791.182	166635.053
Prop_House_Size	601.2831188	41.75522309	14.40018935	7.1527E-35	519.057168	683.50907	519.057168	683.50907
Prop_Lot_Size	-19.73356823	4.468119689	-4.416526326	1.4789E-05	-28.5323571	-10.9347794	-28.5323571	-10.934779

The highest property prices are found with a property lot size of 2700 and a property house sizes of 3300 and 2328. The average property price for a house size of 3300 is $3,250,000. The average property price for a house size of 2328 is $2,697,000. The Multiple R is 0.68 which indicates there is a moderate positive relationship between the number of beds and baths and property price. In the regression, 46.6% of the variability found in the property price explains the input variables, property beds and baths. In addition, the Significance F of $7.998E^{-36}$ shows the relationship between the variables is statistically significant. However, with the adjusted R Squared being 0.462, running a multiple regression will not provide any additional value to our model.

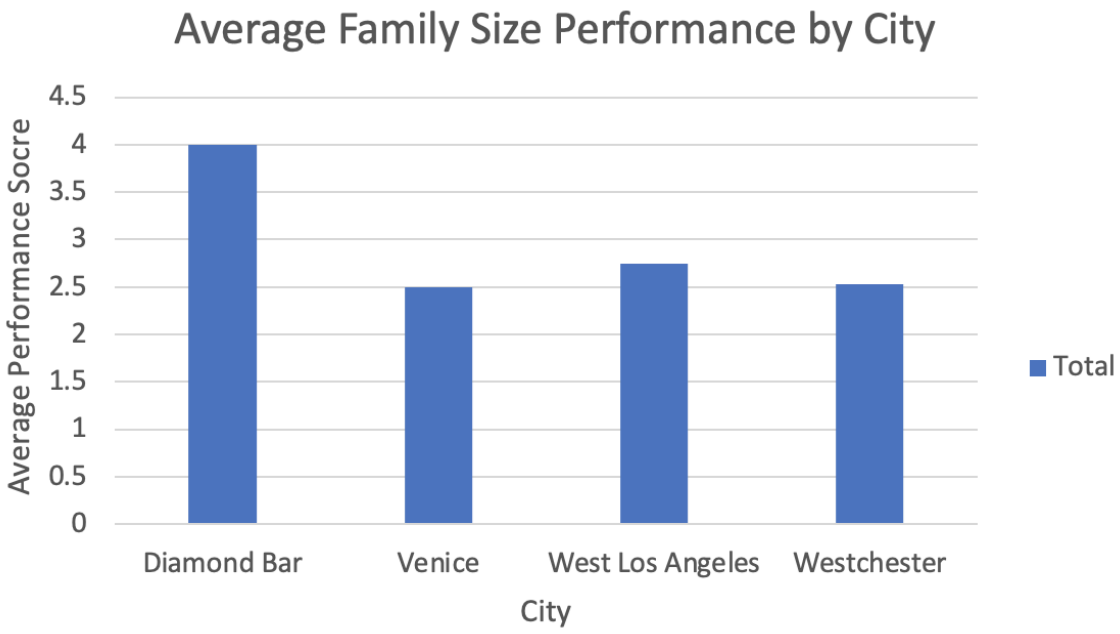
Pivot table determining cities that have collected the highest average performance score:



City	Performance Average
Diamond Bar	4
Venice	2.63
West Los Angeles	2.9
Westchester	2.5
Grand Total	**12.03**

The cities listed in order from highest to lowest ratings for property performance include Diamond Bar, West Los Angeles, Venice, and Westchester. This means Diamond Bar has the properties that sell best because it has the highest average property rating.

Pivot table determining highest average family-sized home performance by city:



City	Avg Family Size Property Performance by City
Diamond Bar	4
Venice	2.5
West Los Angeles	2.75
Westchester	2.531914894
Grand Total	**11.78191489**

For this analysis, I will assume that houses with 4+ bedrooms are considered family sized homes. Diamond Bar has the highest city performance for these homes. Following Diamond Bar is West Los Angeles, Venice, and Westchester respectively.

General Patterns[3]

Multi-regression analysis between all factors to determine price:

When comparing property price to the number of beds, number of baths, house size, lot size, and performance score, the Multiple R is 0.757 which indicates there is a strong positive relationship between our independent and dependent (property price) variables. In the regression, 57.3% of the variability found in the property price explains the input variables, property beds and baths. In addition, the Significance F of $5.459E^{-45}$ shows the relationship between the variables is statistically significant. However, with the adjusted R Squared being 0.564, running a multiple regression as opposed to a single linear regression will not provide any additional value to our model.

Multi-regression analysis between all factors to determine performance score:

When comparing property performance to number of beds, number of baths, house size, lot size, and price, the Multiple R is 0.335 which indicates there is a moderately positive relationship between the independent and dependent (property performance) variables. In the regression, 11.2% of the variability found in the property price explains the input variables, property beds and baths. In addition, the Significance F of $1.267E^{-05}$ shows the relationship between the variables is statistically significant. However, with the adjusted R Squared being 0.094, running a multiple regression as opposed to a single linear regression will not provide any additional value to our model.

[3] Regression statistics and ANOVA tables with calculated values to support these numbers are provided in the attached excel file.

Trends:

Diamond Bar maintains the strongest performance level even when limited to family size.

There is a stationary trend for performance rates of the real-estate agents.

Business Recommendations

For Managers

- Assuming higher property values obtain higher profit for real estate company:
    - Represent homes with 3 beds and 3 baths
    - Represent homes located in Venice
    - Assign Nancy Komer to properties with a performance level of excellent
    - Represent homes with a house size of 3300 and lot size of 2328
    - Represent homes in Diamond Bar to increase property performance rates
    - Market Diamond Bar homes to families (4+ beds)

Calculations

Excel File Link

Click to open working excel file.