

BAN-6025

Machine Learning for Business Analytics

Real Estate Price Prediction and Investment Analysis

Housing Price Data Overview

Considered Factors

Size:

Total area;
First Floor area...

Maintenance:

Year built;
Year remolded ...

Built-in facilities:

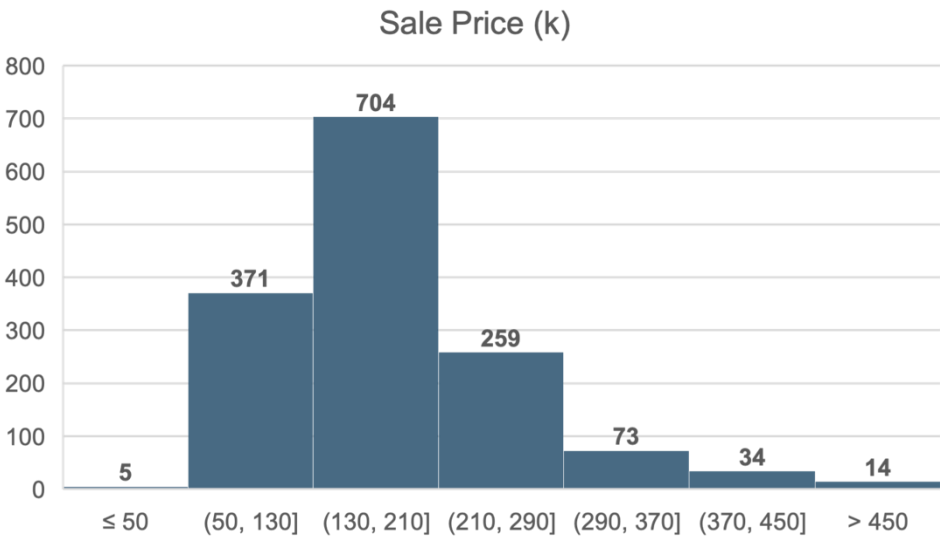
Bathroom;
Bedroom;
Central air...

External Environment:

House type;
Foundation;
Drive paved...



Prediction	Selling Price
Number of Variance	23
Number of Houses	1460
Houses Sold Year	2006 - 2010
Number of Variables	22



Final Model & Model Reliability

Log-Linear Model	Log(Sale Price) = $\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon$		
Adjusted R Square	80.7%		
Variable (x)	Coefficient (β)	p-value	Base Case
Numerical Variables			
Lot Area	0.000001731	0.001	
Total Basement Area	0.0002	0	
1 st Floor Area	0.0002	0	
2 nd Floor Area	0.0002	0	
Full Bathroom No.	0.0853	0	
Half Bathroom No.	0.0569	0	
Fireplaces No.	0.0575	0	
Car No. in Garage	0.1084	0	
Wood Deck Area	0.0002	0	
Screen Porch Area	0.0003	0.001	
Pool Area	-0.0004	0	

(Table continued...)

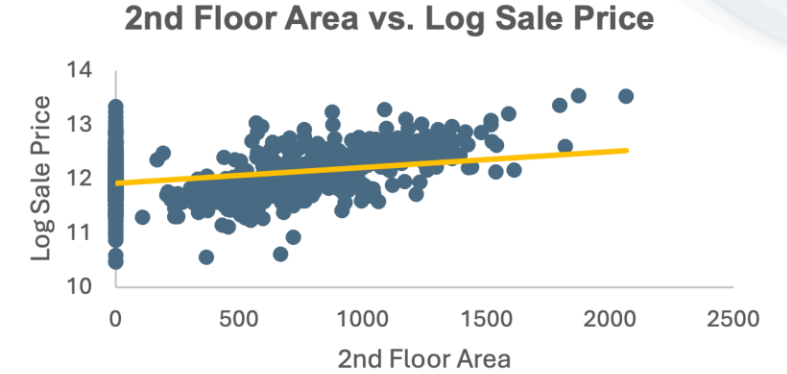
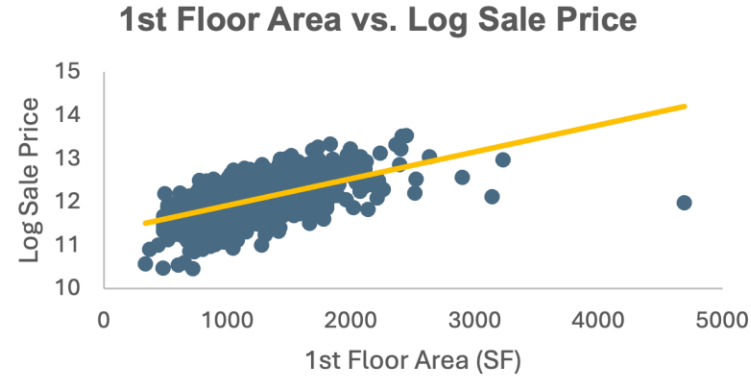
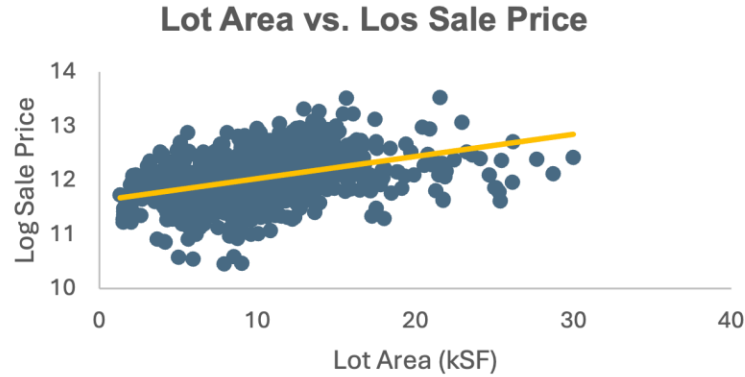
Final Model & Model Reliability

Variable (x)	Coefficient (β)	p-value	Base Case
Categorical Variables			
Central Air (Yes)	0.193	0	Central Air (No)
Kitchen Quality (Fair)	-0.3383	0	Kitchen Quality (Excellent)
Kitchen Quality (Good)	-0.1513	0	
Kitchen Quality (Unknown)	-0.3081	0	
Paved Drive (Partially)	0.0607	0.105	Paved Drive (None)
Paved Drive (Completely)	0.0976	0	

Model Reliability:

- **Categories to Numbers:** Categorical features are turned into numbers (0/1) for better evaluation.
- **Adjusted $R^2 = 80.7\%$:** The model explains 80.7% of what affects house prices.
- **$p < 0.05$:** Most 17 variables are significant at the 5% significance level. Low possibility of random relationships.
- **VIF < 10:** No Major Multicollinearity. Variables aren't too similar.
- **Normal Residuals:** Converting to a log-linear model due to an abnormal QQ plot for multiple linear regression.
- **Log-Linear Model:** Ensuring a consistent comparison across various sizes by accounting for proportional area changes.

Land Space & Living Space:



→ Key Findings:

Lot Area (0.000001731): **Very moderate positive** relationship

1st Floor Area (0.0002): **Moderate positive** relationship

2nd Floor Area (0.0002): **Moderate positive** relationship

→ Key Suggestions:

Living space > land space: Buyers may value living space more.

Outdoor and Auxiliary Features:

→ Key Findings:

Basement Area (0.0002): **Moderate positive** relationship

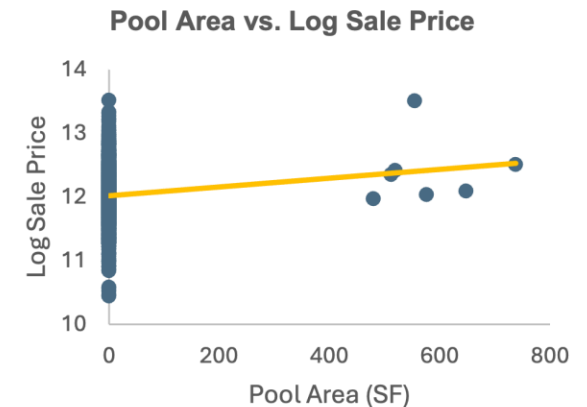
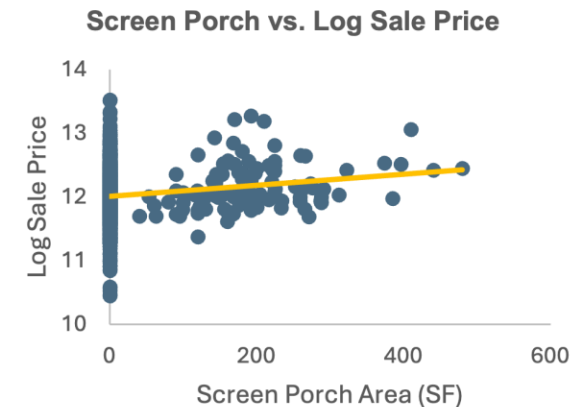
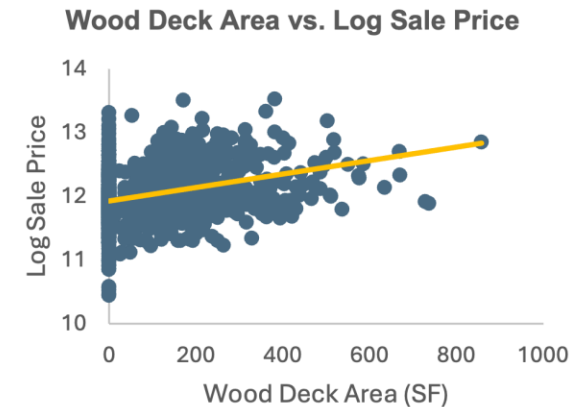
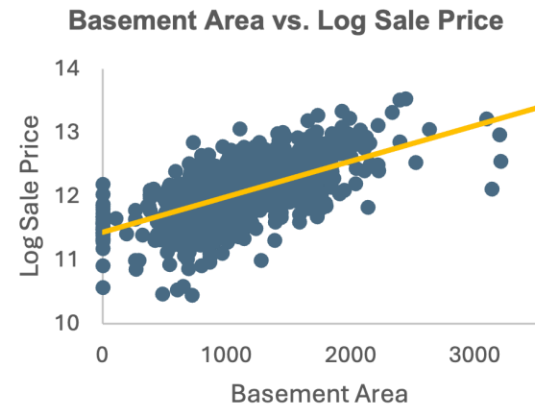
Wood Deck Area (0.0002): **Moderate positive** relationship

Screen Porch Area (0.0003): **Moderate positive** relationship

Pool Area (-0.0004): **Moderate negative** relationship

→ Key Suggestions:

Be cautious about investing in pool areas and prioritize the basement, wood deck, and screen.



Built-in Features:

→ Key Findings:

Full Bathroom Number (0.0853): **Strong positive** relationship

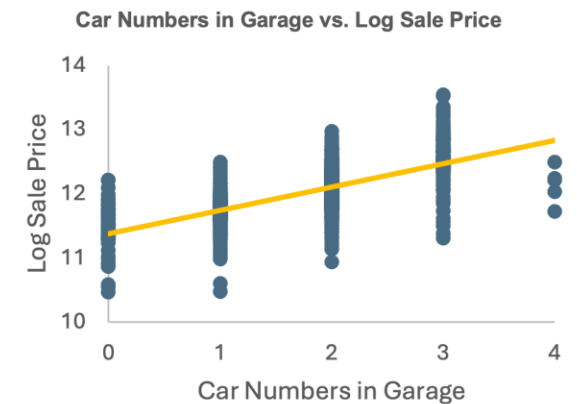
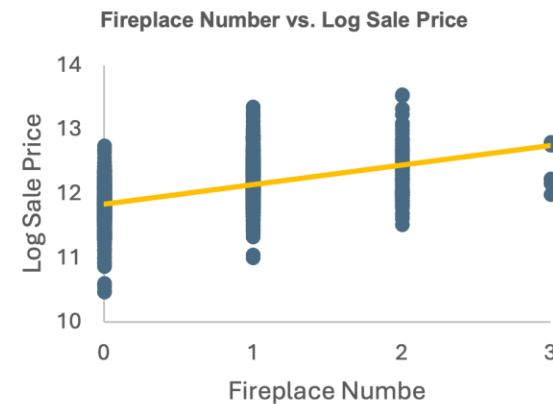
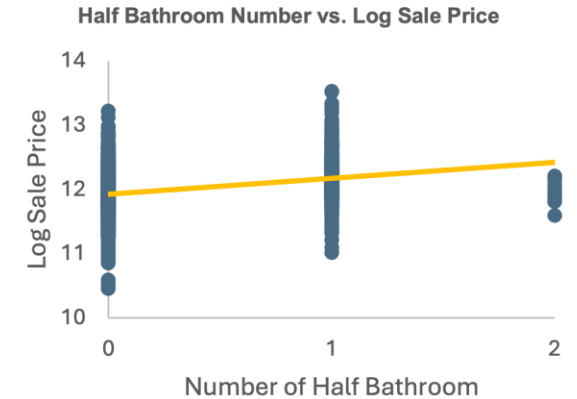
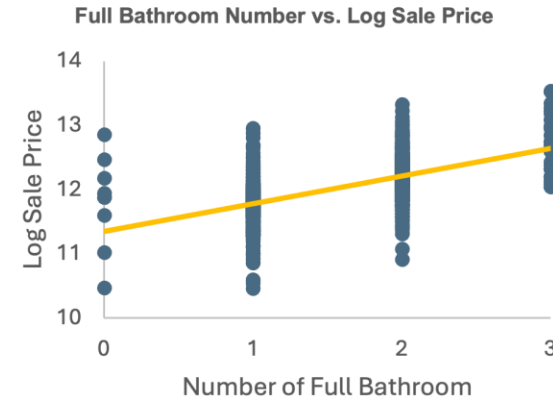
Half Bathroom Number (0.0569): **Strong positive** relationship

Fireplace Number (0.0575): **Strong positive** relationship

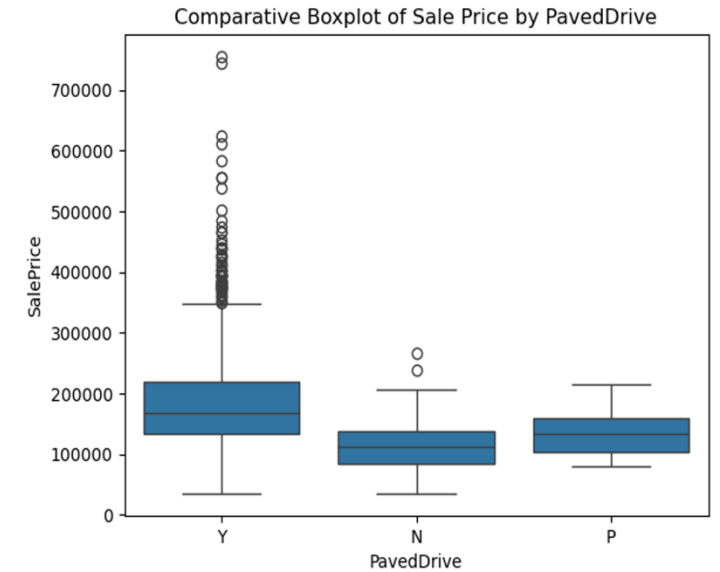
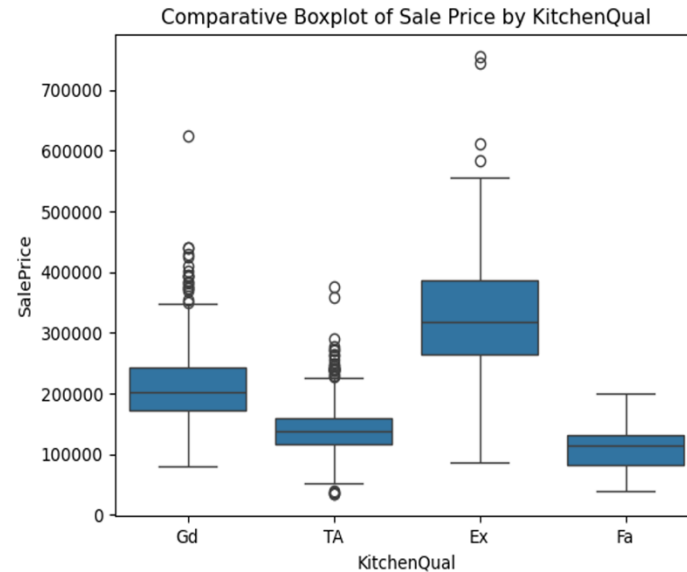
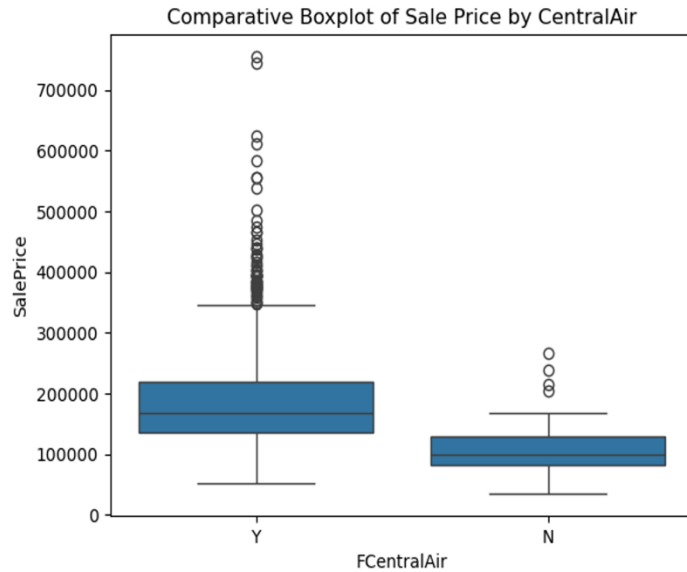
Car Number in Garage (0.1084): **Strong positive** relationship

→ Key Suggestions:

Investing in bathroom, fireplace and garage.



Categorical: Essential Home Comforts



→ Key Findings:

Central Air (Yes: 0.193): **Strong positive** relationship

Kitchen Quality (Fair: -0.3383; Good: -0.1513; Unknown: -0.3081): **Strong positive** relationship

Paved Drive (Partially: 0.0607; Completely: 0.0976): **Strong positive** relationship

→ Key Suggestions:

Investing in central air, kitchen quality and paved drive.

Recommendations:

Relationship	Strength	Variables
Positive	Strong	Full & Half Bathroom Number
		Fireplace Number
		Car Number in Garage
		Central Air
		Kitchen Quality
		Paved Drive
	Moderate	Lot Area
		1 st & 2 nd Floor Area
		Basement Area
		Wood Deck Area
		Screen Porch Area
Negative	Moderate	Pool Area

1

2

!

Sale Price Predictions:

Id	11_1	11_2	11_3	11_4	11_5
Predicted Sale Price (\$)	102,445	127,635	154,923	275,019	93,417
Lot Area (SF)	8450	9600	11250	11899	10791
1 st Floor Area (SF)	856	1262	920	1200	1296
2 nd Floor Area (SF)	854	0	866	1142	0
Total Basement Area (SF)	0	1262	920	1175	0
Wood Deck Area (SF)	0	298	0	147	0
Screen Porch Area (SF)	0	0	0	0	0
Pool Area (SF)	0	0	0	0	0
Full Bathroom No.	2	2	2	3	2
Half Bathroom No.	1	0	1	1	0
Fireplaces No.	0	1	1	2	0
Cars No. in Garage	2	2	2	3	2
Central Air	Y	Y	Y	Y	Y
Kitchen Quality	Fa	TA	Gd	Ex	TA
Paved Drive	Y	Y	Y	Y	Y