

Ames, Iowa: Alternative to Boston Housing Data Set

ADTA 5130 - Data Analytics - 1

July 26th 2023 - Final Project

Blake Manuel, Kadala Sai Chandana, Manohar Ravulapalli, Yog Chaudhary

Blake Manuel

ANOVA: Sales Price and Housing Condition

H0: There is no significant relationship between the two variables.

HA: There is a significant relationship between the two variables.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
average	1654	340768703	206027.0272	7573604902		
above average	533	80151414	150377.8874	2556089815		
good	390	59670763	153001.9564	2405173738		
very good/excellent	185	30478075	164746.3514	4121426566		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.845391392681E+12	3	6.151304642271E+11	108.9407159	1.16776	2.60812997
Within Groups	1.557296375641E+13	2758	5646469817			
Total	1.741835514909E+13	2761				

The data in the table indicates that there is a difference in the mean sales price of houses in Ames, Iowa for the four housing conditions: average, above average, good, and very good/excellent. The average housing condition has the highest mean sales price, while the above average housing condition has the lowest mean sales price.

P-Value (1.16776): The P-Value tells us that there is 11.68% chance that the observed difference in mean sales prices for the four housing conditions is due to chance. This is above the significance level of 0.05, so we fail to reject the null hypothesis and conclude that the difference in mean sales prices is likely due to the housing condition of the houses.

F-critical (2.60812997) & F-statistic (108.94): Since the F-statistic of 108.94, is much greater than the F-critical of 2.60812997, we can conclude that the housing condition has a significant effect on sales price.

These results suggest that the housing condition of a house is a significant factor in determining its sales price. Houses in better condition tend to sell for more than houses in worse condition.

Sai Chandana K:

ANOVA:Sales Price with Housing Style

(H0): Rejecting the null hypothesis & (Ha): We can conclude that there is significant relationship with the Sales Price and the Housing Styles.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
1.5Fin	314	43184396	137529.9	2.23E+09		
1.5Unf	19	2083600	109663.2	4.23E+08		
1Story	1442	2.57E+08	178092.6	6.59E+09		
2.5Fin	8	1760000	220000	1.4E+10		
2.5Unf	24	4251800	177158.3	5.79E+09		
2Story	856	1.78E+08	207558.2	7.16E+09		
SFoyer	78	11261231	144374.8	8.69E+08		
SLvl	126	20897005	165849.2	1.19E+09		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.45E+12	7	2.07E+11	35.34316	0.00	2.01278
Within Groups	1.68E+13	2859	5.87E+09			
Total	1.82E+13	2866				

The Results of the ANOVA show that the group means are significantly different from one another.This condition is supported by the low p -value(1.27E-47),which is below the accepted significance level of 0.05.

The F-value (35.34315682) is used to compare the variance within the groups to the variance between the groups.A high F- value suggests that the difference between the group means are more significant than the “Sales Prices” &”Housing Style”.

The Significance of the group differences is illustrated by the relatively high F-value in this case.

Overall,the results of the ANOVA demonstrate that the difference between “Sales Price”&”Housing Style” is statistically more significant with the “Sales Price”.

Yog Chaudhary :

ANOVA: Sales Price and Housing Condition: Year Built, 1st Flr Sf, Full Bath, Bedroom AbvGr & Kitchen AbvGR

HO: There is no significant relationship between two the variable (Year Built, 1st Flr Sf, Full Bath, Bedroom AbvGr & Kitchen AbvGR)

HA: There is a Significant relationship between two the variable (Year Built, 1st Flr Sf, Full Bath, Bedroom AbvGr & Kitchen AbvGR)

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
SalePrice	2930	529732456	180796.0601	6381883616		
Year Built	2930	5776074	1971.356314	914.7818396		
1st Flr SF	2930	3397504	1159.557679	153578.4659		
Full Bath	2930	4590	1.566552901	0.30574332		
Bedroom AbvGr	2930	8363	2.854266212	0.685138843		
Kitchen AbvGr	2930	3060	1.044368601	0.045828638		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	7.9268E+13	5	1.58536E+13	14904.57903	0	2.21460799
Within Groups	1.8693E+13	17574	1063673018			

The data in the table indicate that there is a statistical to analyze the differences between the means of two or more groups.

The variance in the sales price across differents groups, including Year Built, 1st flr SF, Full Bath, Bedroom AbvGr, and Kitchen AbvGr

The ANOVA results show that there are significant difference among these groups based on the extremely low P-Value < 0..05 and the large F-statistic 14904.58,

The indicates that the independent variables have a significant impact on sale price. House is better condition tend to sell for more than house in worse conditions.

Manohar ravulapalli

Anova: How does Sales price differs by building type?

H0: There is no significant difference in "SalesPrice" across the different building types (2fmCon, Duplex, TwnhsE, Twnhs, and 1Fam).

HA: There is a significant difference "SalesPrice" across the different building types (2fmCon, Duplex, TwnhsE, Twnhs, and 1Fam).

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
2fmCon	62	7786066	125581.7	9.67E+08		
Duplex	109	15239174	139808.9	1.56E+09		
TwnhsE	233	44808676	192311.9	4.38E+09		
Twnhs	101	13729340	135934.1	1.76E+09		
1Fam	2425	4.48E+08	184812	6.86E+09		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	6.45411E+11	4	1.61E+11	26.15136	2.48E-21	2.37497
Within Groups	1.80471E+13	2925	6.17E+09			
Total	1.86925E+13	2929				

From the analysis we can see that the F-statistic tests whether the between-group variation is significantly larger than the within-group variation. A low p-value (2.48E-21) associated with the F-statistic indicates that the between-group variation is significant, which we can suggest that there is a difference in "SalesPrice" among the building types.

From the ANOVA analysis results we can conclude that the building type is a statistically significant factor in determining the "SalesPrice".

Blake Manuel

Regression: Total SF (1st floor, 2nd floor, & basement) and Sales Price

H0: There is no significant difference in the mean sales price of houses for the four housing conditions.

HA: There is a significant difference in the mean sales price of houses for the four housing conditions.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.79312725							
R Square	0.62905083							
Adjusted R Square	0.62892414							
Standard Error	489.978325							
Observations	2930							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	1192054607	1192054607	4965.26479	0.00E+00			
Residual	2928	702950606	240078.759					
Total	2929	1895005213						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1102.483	22.3999647	49.2180686	0.0E+00	1058.5617	1146.4043	1058.5617	1146.4043
SalePrice	0.00798572	0.00011333	70.464635	0.0E+00	0.0077635	0.0082079	0.0077635	0.0082079

Multiple R value (0.793127249); This indicates there is a strong positive correlation between sales price and square footage. As the square footage of a house increases, so does its sales price.

R-squared (0.629050833): This indicates that 62.90% of the variation in sales price can be explained by the square footage of the house.

The F-statistic (4965.264786): This indicates there is a very low probability that the observed relationship between the square footage of a house and its sales price is due to chance.

Coefficients (0.007985718): This indicates that for every additional square foot, the sales price of a house is expected to increase by \$7.98. Additionally, the standard error for the coefficient is 0.000113329, which means that the coefficient is statistically significant.

95% confidence intervals (0.007763504 and 0.008207931) This indicates that we can be 95% confident that the true coefficient for the square footage variable is between 0.007763504 and 0.008207931.

Sai Chandana K:

Regression: Sales Price with Total Rooms Above Ground

(H0): Rejecting the null hypothesis & (Ha):We can conclude that there is significant relationship with the Sales Price and the Total Rooms Above Ground.

Regression: SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.498975							
R Square	0.248976							
Adjusted R Square	0.248463							
Standard Error	69266.51							
Observations	2929							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2	4.65398E+12	2.33E+12	485.0068	0.00			
Residual	2926	1.40385E+13	4.8E+09					
Total	2928	1.86925E+13						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	43339.71	8763.474944	4.945494	0.00	26156.50543	60522.912	26156.50543	60522.9118
TotRms AbvGrd	24921.17	817.3165343	30.49145	0.00	23318.59548	26523.743	23318.59548	26523.7432
Overall Cond	-4150.67	1156.134594	-3.59012	0.000336	-6417.585449	-1883.746	-6417.585449	-1883.74567

The expected home price is represented by the intercept(0),which equals 43339.70863 when all independent variables are zero.

Total Rms Abv Grd :The Coefficient(1) is 24921.16936, which indicates that,typically,each additional room above grade is associated with an increase in home price of 24921.17 units.

Overall Cnd:The Coefficient (2) is -4150.66556 , which means that, on average,each unit increase in overall condition rating correlates to a 4150.67 unit decrease in housing price.

The results of the regression analysis show that the quantity of above grade rooms and the total condition score have a big impact on home prices.

Yog Chaudhary:**Regression:** Sales Price: Year Built, 1st Flr, Full Bath, Bedroom AbvGr & Kitchen AbvGr**HO:** There is no significant relationship between the dependent variable and the independent variables (Year Built, 1st Flr, Full Bath, Bedroom AbvGr & Kitchen AbvGr). The coefficients of all independent variables are equal to zero.**HA:** There is significant relationship between the dependent variable and the independent variables (Year Built, 1st Flr, Full Bath, Bedroom AbvGr & Kitchen AbvGr)

Regression Statistics								
Multiple R	0.776466317							
R Square	0.602899941							
Adjusted R Square	0.602220906							
Standard Error	50384.32182							
Observations	2930							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5	1.12697E+13	2.25395E+12	887.87669	0			
Residual	2924	7.42281E+12	2538579885					
Total	2929	1.86925E+13						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Upper 95.0%
Intercept	-1383313.924	74024.27657	-18.68730082	1.04647E-73	-1528458.921	-1238168.927	-1528459	-1238168.927
Year Built	736.2584919	37.6041264	19.57919415	2.60404E-80	662.5252374	809.9917463	662.5252	809.9917463
1st Flr SF	91.01401414	2.599186778	35.01634238	1.213E-224	85.91759206	96.11043622	85.91759	96.11043622
Full Bath	37073.04452	2216.042586	16.72939173	4.55469E-60	32727.88223	41418.2068	32727.88	41418.2068
Bedroom AbvGr	5916.233876	1264.191268	4.679856623	3.00132E-06	3437.438453	8395.029299	3437.438	8395.029299
Kitchen AbvGr	-64936.10921	4609.821372	-14.08646973	1.19137E-43	-73974.93459	-55897.28383	-73974.9	-55897.28383

In the table R-Squared values of 0.6029 indicates that approximately 60.3% of the variation in the sales prices.

The Coefficients Year Built 736.2585 and 1st Flr SF 91.0140 indicates that, on average, for each additional square foot of the first floor area, the sale price increases by \$91.01

The Regression analysis helps us different factors contribute to the variation in house sale price, enabling better decision-making and predicting future sale price based on these variables.

Manohar ravulapalli

Regression: How does "Year Built," "Overall Qual," and "Lot Area" collectively influence the "SalePrice" ?

H0: There is no significant relationship between "Year Built," "Overall Qual," and "Lot Area" with the "SalePrice" of properties. The coefficients of all three independent variables are equal to zero.

HA: There is a significant relationship between "Year Built," "Overall Qual," and "Lot Area" with the "SalePrice" of properties.

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.828765926
R Square	0.68685296
Adjusted R Squ	0.686531894
Standard Error	44727.13909
Observations	2930

ANOVA

	df	SS	MS	F	Significance F
Regression	3	1.2839E+13	4.28E+12	2139.284	0
Residual	2926	5.85351E+12	2E+09		
Total	2929	1.86925E+13			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-792697.1272	64690.37989	-12.2537	1.06E-33	-919540.4114	-665853.8431	-919540.4114	-665853.8431
Year Built	361.336315	34.09343367	10.59841	8.84E-26	294.4867602	428.1858697	294.4867602	428.1858697
Overall Qual	39551.33156	734.0705477	53.87947	0	38111.98433	40990.67879	38111.98433	40990.67879
Lot Area	1.98167865	0.105476854	18.78781	1.93E-74	1.774862265	2.188495035	1.774862265	2.188495035

From the analysis we can see that -

Year Built: The coefficient of 361.34 can suggest that for each year increase in the "Year Built," the "SalePrice" is expected to increase by approximately \$361.34.

Overall Qual: The coefficient of 39551.33 indicates that for each unit increase in the "Overall Qual" rating, the "SalePrice" is expected to increase by approximately \$39,551.33.

Lot Area: The coefficient of 1.98 suggests that for each additional unit increase in "Lot Area" (e.g., 1 square foot), the "SalePrice" is expected to increase by approximately \$1.98.

Overall, the regression model shows that "Year Built," "Overall Qual," and "Lot Area" have a significant positive impact on the "SalePrice" of properties.

Recommendation

ANOVA

- **Blake:** The ANOVA results indicate that housing condition is a significant factor in determining its sales price. Houses in better condition are more likely to sell for more than houses in worse condition. However, houses in average condition sell for more than houses in very good/excellent condition. Therefore, our client should focus on buying houses in better condition, as these will be more likely to appreciate in value and generate higher rental income.
- **Yog:** The ANOVA result indicates that housing is significant difference exist among the group . The extremely low p-value ($P < 0.005$) and the large F-statistic (14904.58) provide strong evidence for specific comparisons to understand factors impacting Sales price mostly will be likely to appreciate in values and meaningful decisions.
- **Sai:** The ANOVA results indicates that the housing style is a significant factor in determining its sales price. We can conclude that the housing styles are more likely to sell based on the style of the houses with its Sales price. Finally, based upon the P-value & F-Statistic value we can provide the strong evidence that the factors impacting Sales Price are more likely significant with the housing style.
- **Manohar:** From the ANOVA results, a low p-value (2.48×10^{-21}) associated with F-statistics (26.151) indicate that building type is a significant factor in determining the sales prices. Hence it provide strong evidence for specific comparisons to understand building types impacting Sales price.

Regression

- **Blake:** The regression results indicate that there is a strong positive correlation between sales price and square footage. As the square footage of a house increases, so does its sales price. Houses with more square footage will be more expensive, but they will also be able to rent for more.
- **Yog:** The regression results to consider the significant predictors (Year Built, 1st Flr, sf, Full Bath, Bedroom AbvGr, Kitchen AbvGr) in analyzing Sales price have positive impacts, while Bedroom AbvGr and Kitchen AbvGr have negative impacts on sale price. Therefore results are essential for better understanding the relationships.
- **Sai:** The results of the regression analysis show that the number of rooms above grade and the total condition score have a big impact on home prices. The model accounts for about 24.9% of the variation in home prices, indicating that there may be more factors that are not considered in the analysis that affect home prices.
- **Manohar:** Manohar: From the regression analysis we can conclude that "Year Built," "Overall Qual," and "Lot Area" have a significant positive impact on the "SalePrice" of properties. The R-squared value indicates that the model explains about 68.69% of the variance in "SalePrice," suggesting that it is important portion of the price variation based on the given independent variables. With an understanding of how "Lot Area," "Year Built," and "Overall Qual" can impact property values, the client can make more informed decisions regarding real estate investments.