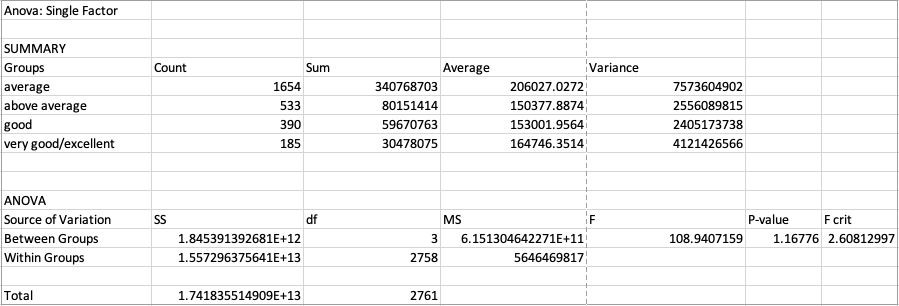
Ames, Iowa: Alternative to Boston Housing Data Set

**ADTA 5130 - Data Analytics - 1**

July 26th, 2023 - Final Project

Azhan Saleem

# ANOVA: Sales Price and Housing Condition



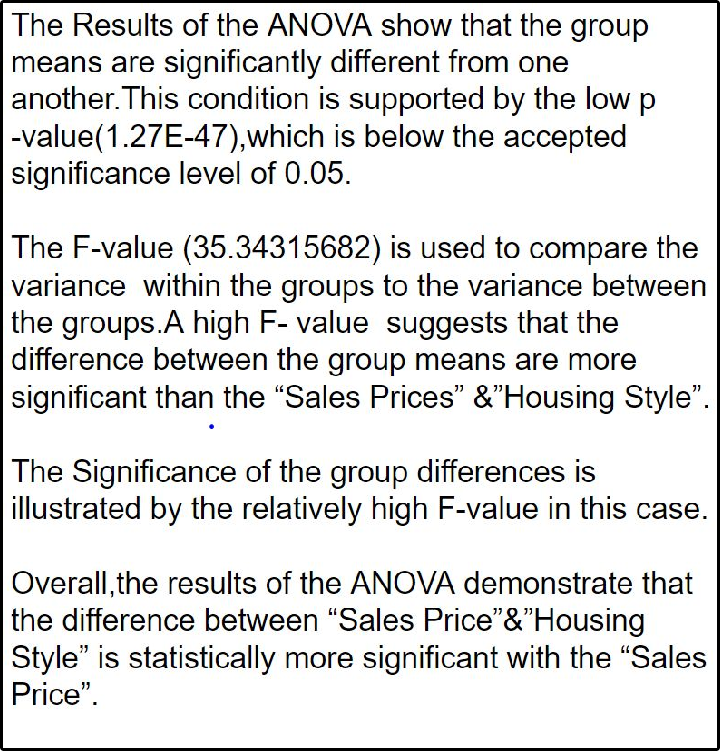
H0: There is no significant relationship between the two variables. HA: There is a significant relationship between the two variables.

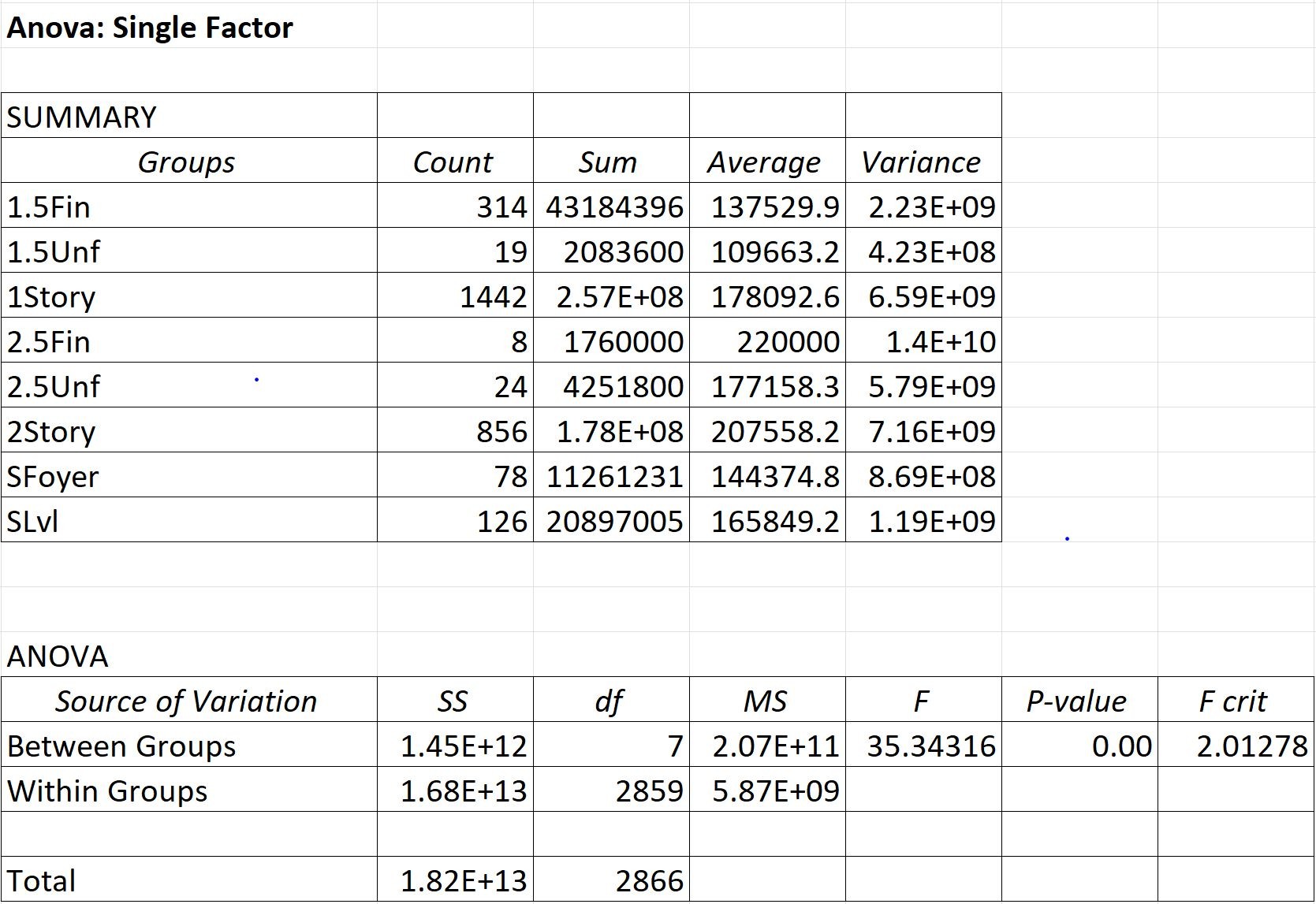
The data in the table indicates that there is a difference in the mean sales price of houses in Ames, Iowa for the four housing conditions: average, above average, good, and very good/excellent. The average housing condition has the highest mean sales price, while the above average housing condition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

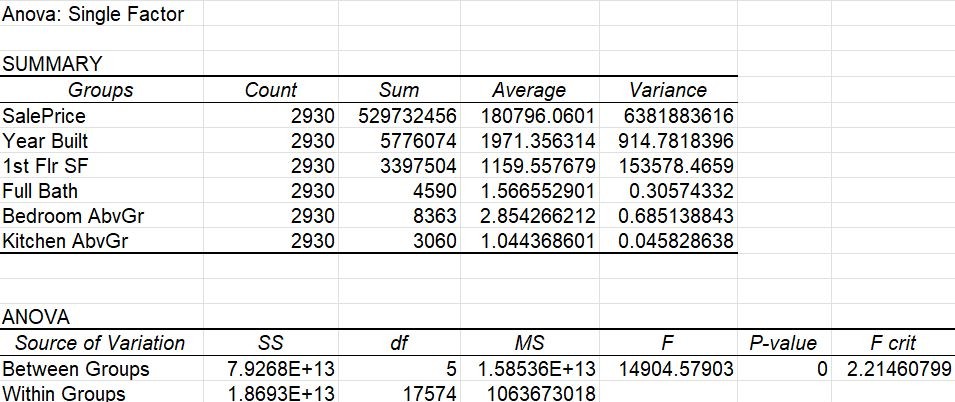




**ANOVA:** Sales Price and Housing Condition: Year Built, 1st Flr Sf, Full Bath, Bedroom AbvGr & Kitchen AbvGR

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 )



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

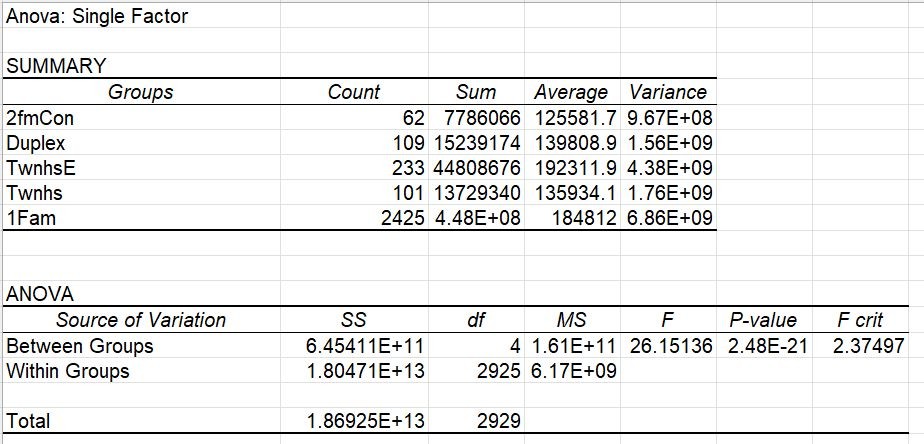
The ANOVA results show that there are significant difference among these groups based on the extremely low P-Value < 0..05 and the large F-statistic 14904.58,

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.

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).



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

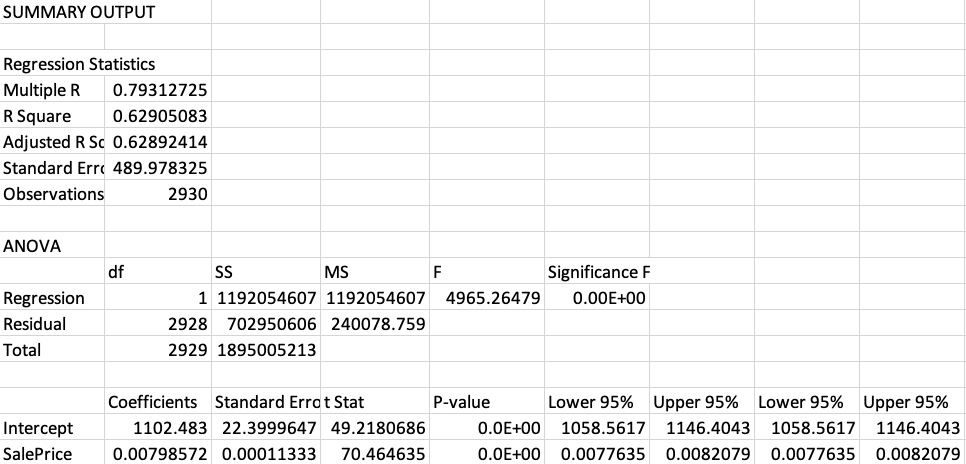
F-statistic indicates that the

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".

# 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.



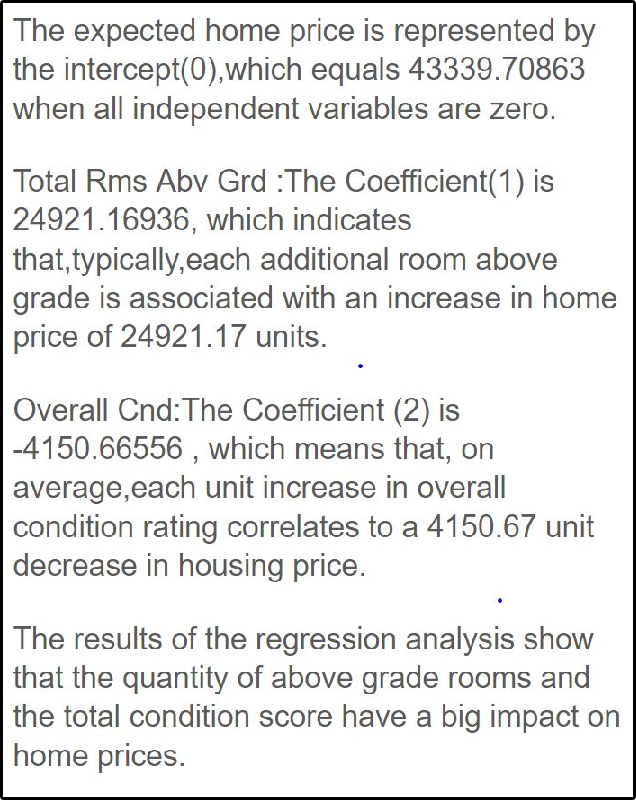
Multiple R value (0.793127249); This indicates there there is a strong positive correlation between sales price and square footage. As the square footage of a house increases, so does its sales price.

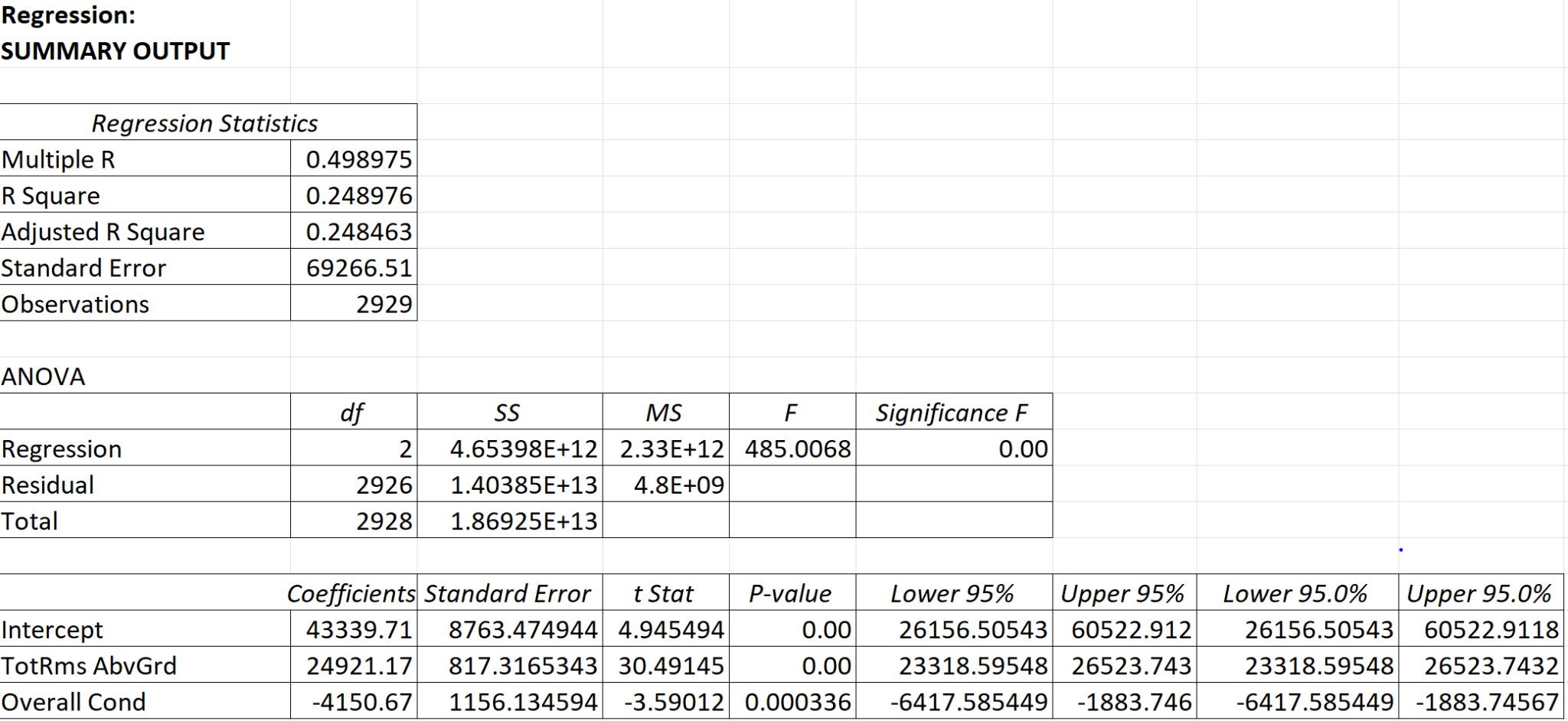
R-squared (0.629050833): This indicates that 62.90% of the variation in sales price can be explained by the square footage of the house.

The F-statistic (4965.264786): This indicted 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.

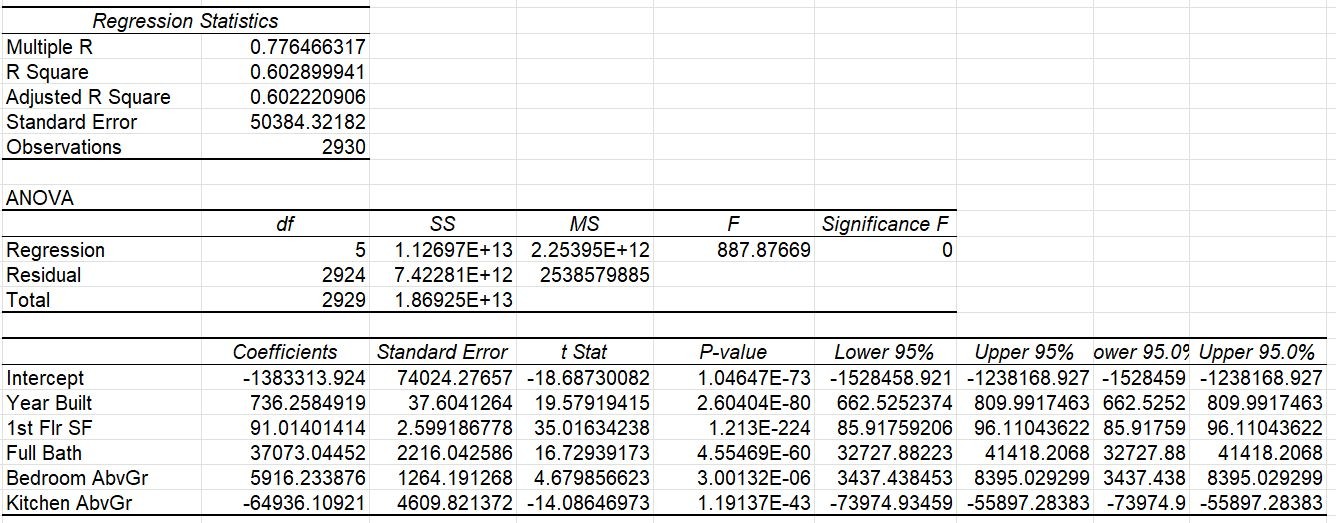




**Regression:** Sales Price: Year Built, 1st Flr, Full Bath, Bedroom AbvGr & Kitchen AbvGr

**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)



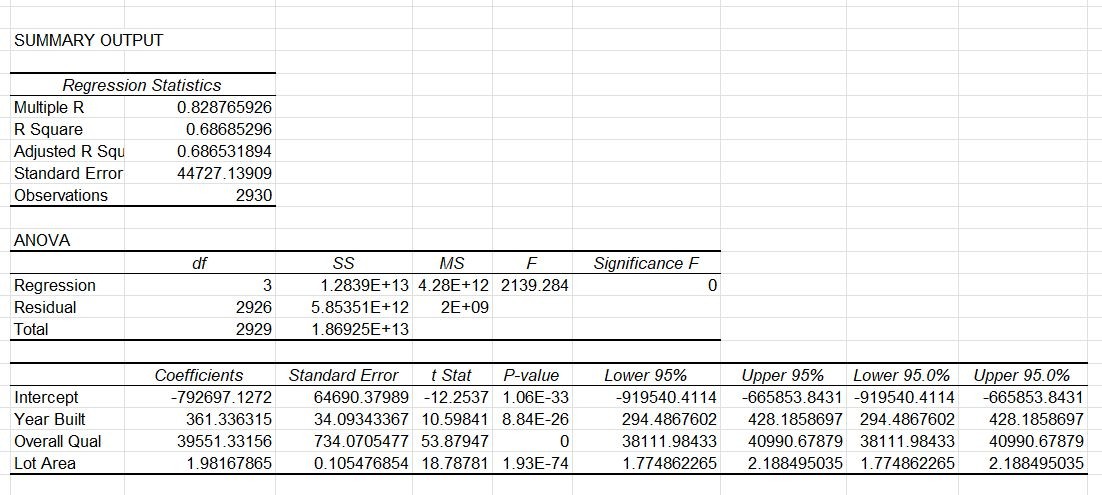
In the table R-Squared values of 0.6029 indicates that approximately 60.3% of the variation in the sales prices.

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.

**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.



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

* 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.
* 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.

: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.

* :From the ANOVA results, a low p-value(2.48E-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

* 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.

The regression results to consider the significant predictors ( Year Built, 1st Flr, sf, Full Bath,Bedroom AbvGr, Kitchen AbrGr) in analyzing Sales price have positive impacts, while Bedroom AbvGr and Kitchen AbGr have negative impacts on sale price.Therefor results are essential for better understanding the relationships.

* 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: 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.