STAT-614 FINAL PROJECT REPORT:

PROJECT TITLE:

Analysing Factors Affecting House Prices in Ames, Iowa

INDEX:

Introduction:

The Ames Housing dataset is a rich collection of data encompassing 1460 observations with 81 attributes, providing an extensive overview of the housing market in Ames, Iowa. This report aims to dissect the impact of various factors on house sale prices, offering valuable insights for homeowners, investors, and policymakers.

The following research questions guide our analysis:

1.Is there a significant difference in the average sale price of houses with a central air conditioning system compared to the overall average sale price?

2.Is there a significant difference in the mean sale price between one-family detached houses and townhouses?

3.What is the impact of overall house quality on its sale price, controlling for other factors like house size and age?

4.Is there a relationship between building types and neighborhoods?

Data:

Attribute	**Type**	**Description**
MSSubClass	Nominal Attribute	Type of dwelling
MSZoning	Categorical	Type of dwelling
LotFrontage	Categorical	Linear feet of street connected to property
LotArea	Categorical	Lot size in square feet
Street	Categorical	Type of road access
Alley	Categorical	Type of alley access
LotShape	Continuous	General shape of property
Utilities	Continuous	Flatness of the property
LotConfig	Continuous	Type of utilities available
LandSlope	Categorical	Lot configuration
Neighborhood	Categorical	Slope of property
Condition1	Categorical	Physical locations within Ames city limits
Condition2	Categorical	Proximity to various conditions

Statistical Analysis:

Question 1:

Is there a significant difference in the average sale price of houses with a central air conditioning system compared to the overall average sale price?

Null Hypothesis H0: The average sale price of homes equipped with central air conditioning is not significantly different from the total average sale price (1 = 2).

Alternative Hypothesis H1: The average sale price of homes with central air conditioning differs significantly from the general average sale price (1 2).

$\alpha = 0.05$ $(level\ \ of\ significance)$

Statistical Test: One sample T-test

T-Statistic	2.07
P-value	0.038

Decision:

Since the p-value (0.038) is less than the common alpha level of 0.05, we reject the null hypothesis.

Conclusion:

As a result, we have enough data to rule out the null hypothesis and draw the conclusion that the average sale price of homes equipped with central air conditioning differs significantly from the total average sale price.

Question-2:

Is there a significant difference in the average sale price between houses with different overall conditions?

levels (treatments) and OverallCond is the factor.

Null Hypothesis H0: Homes with various general conditions do not significantly differ in terms of average sale price.Level of Significance: 0.05

Statistical Test: One-way ANOVA test

$F - statistic$: 19.03

$P - value$: $8.90x10^{-27}$

Decision: Because the P-value is below the level of significance, we reject the null hypothesis.

Conclusion: We have enough information to reject the null hypothesis and conclude that average sale prices for properties with different general conditions vary considerably from one another.

Question-3: Is there a significant relationship between the total living area (GrLivArea) and the sale price?

Null Hypothesis H0: GrLivArea and SalePrice don't significantly affect each other($\beta 1 = 0$).

Alternate Hypothesis H1: The link between GrLivArea and SalePrice is considerable.

Level of Significance: 0.05

Statistical Test: Simple linear regression

Results:

Estimate	115.26
Standard Error	3.36
T-value	34.26
P-value	<0.00
R-Squared	0.541

Decision:

We disprove the null hypothesis since the P-value is smaller than the level of significance.

Conclusion: As a result, we are able to rule out the null hypothesis and draw the conclusion that there is a meaningful connection between total living space and sale price.

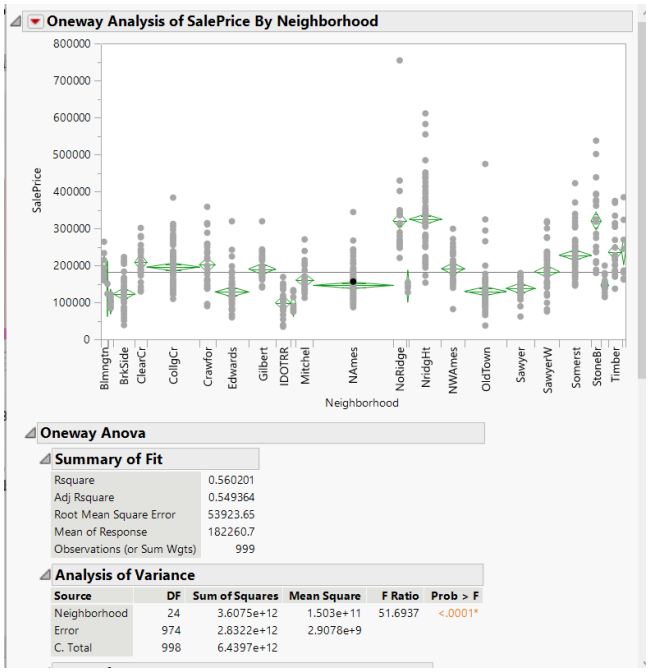
Question-4: Is there a significant association between neighborhood and sale price?

Null Hypothesis H0: There is a strong connection between neighborhood and sale price.

Alternative Hypothesis H1: There is a strong connection between neighborhood and sale price.

Level of Significance: 0.05

Statistical Test: One-way ANOVA test



F-Statistic:51.69

P-Value:$3.34x10^{-155}$

Decision: We disprove the null hypothesis since the P-value is smaller than the level of significance.

Conclusion: As a result, we are able to rule out the null hypothesis and draw the conclusion that there is a meaningful connection between neighborhood and sale price.

Recommendations:

The statistical analysis of the Ames, Iowa housing data has successfully highlighted the crucial factors influencing house prices. This understanding is vital for various stakeholders, including governments, homeowners, and real estate investors. By leveraging this data, homeowners and investors can make more informed decisions about buying or selling properties. Similarly, policymakers can use these insights to create housing policies and interventions aimed at fostering a more equitable housing market.

However, it's essential to acknowledge that the real estate market is subject to a wide array of influences beyond what is captured in this dataset. It's a dynamic and multifaceted sector. Therefore, while these findings are valuable, they should be considered alongside other data sources and market analyses to guide decision-making. This holistic approach is crucial for a comprehensive understanding of the real estate market's complexities.