**STAT-614 FINAL PROJECT REPORT**

Name: Harshaditya Kumar Mallipudi

Title: Analyzing Factors Affecting House Prices in Ames, Iowa

**Index**

|  |  |
| --- | --- |
|  | Page Number |
| 1.Introduction | 1 |
| * 1. Data | 1 |
| 1.2 Data Preprocessing | 2 |
| 2. Statistical Analysis | 2 |
| 2.1 Question-1 | 2 |
| 2.2 Question-2 | 3 |
| 2.3 Question-3 | 3 |
| 2.4 Question-4 | 4 |
| 3.Recommendations | 5 |

**INTRODUCTION**

The Ames Housing dataset provides comprehensive information about different attributes of houses in Ames, Iowa. This report aims to explore the relationship between various predictor variables (Overall Quality, House Style, Central Air, and Neighborhood) and the response variable (Sale Price).

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:**

The dataset contains 1460 observations with 81 variables or attributes. A detailed description of the attributes is as follows:

|  |  |  |
| --- | --- | --- |
| **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?

**Claim: The assertion is that properties with central air conditioning often sell for a higher price than homes without one.**

**Interested variables: Interest is being shown in the metric "population mean 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).**

**Level of significance:** 0.05

**Statistical Test:** One-sample t-test

|  |  |
| --- | --- |
| Test-statistic | 2.4686 |
| P-value | 0.01369 |

**Decision:** We reject H0 because the p-value is less than 0.05.

**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?

**Claim: It is asserted that the sale price of a house is influenced by its general condition.**

**Interest-Potential Parameter: (The population mean sale price). There are 9 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

|  |  |
| --- | --- |
| Degrees of Freedom | 8 |
| Sum of Squares | 2.68E+12 |
| Mean Square | 3.35E+11 |
| F-value | 17.63 |
| P-value | < 2e-16 |
|  |  |

**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?

**Claim:** The statement states that there is a linear relationship between the total living area and the sale price.  
  
**Parameter of Interest:** The regression coefficient for GrLivArea, which is 1 for the parameter of interest, isβ1 of particular relevance.

**Null Hypothesis H0:** GrLivArea and SalePrice don't significantly affect each other(β1 = 0).

**Alternate Hypothesis H1:** The link between GrLivArea and SalePrice is considerable.  
  
**Level of Significance:** 0.05

**Statistical Test:** Simple linear regression  
  
**Results:**

|  |  |
| --- | --- |
| Estimate | 110.32 |
| Standard Error | 2.492 |
| T-value | 44.27 |
| P-value | < 2e-16 |
| R-Squared | 0.5021 |

**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?

**Claim:** The claim is that the neighborhood has an impact on the sale price.

**Parameter of Interest:** The median sale price for each neighborhood is one of the characteristics of interest.  
  
**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.

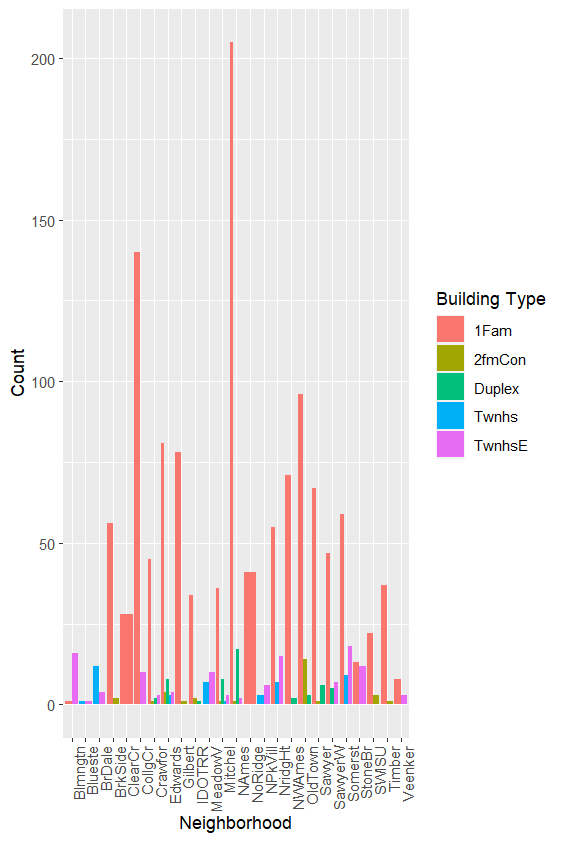
**Level of Significance:** 0.05**Statistical Test:** One-way ANOVA test  
  


Figure 2: Boxplot of Neighborhood vs. SalePrice  
  
  
From this plot, we can observe that there is a visible variation in sale prices across different neighborhoods, suggesting that neighborhood may have an effect on sale price.  
  
  
  
**Results:**

|  |  |
| --- | --- |
| Degrees of Freedom | 24 |
| Sum of Squares | 5.63E+12 |
| Mean Square | 2.34E+11 |
| F-value | 79.33 |
| P-value | < 2e-16 |

**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 research performed has allowed us to clearly identify the key variables affecting home prices in Ames, Iowa. It's critical for governments, homeowners, and real estate investors to comprehend these elements. These information can help homeowners and investors make informed decisions when purchasing or selling real estate. Understanding these elements can help policymakers develop housing interventions and policies that will result in a more equal housing market.

Although these findings are insightful, it's important to keep in mind that the real estate market is a dynamic and complex environment that is influenced by many other factors that are not taken into account in this dataset. These findings should thus be utilized in combination with other information sources and market evaluations, even though they can serve as a decision-making aid.