

# FACTORS FOR HOUSING PRICES

Angela Deslandes



# Table of contents

01

## Introduction

Business objective &  
Research questions

02

## Hypotheses

Testing research  
questions and  
interesting observations

03

## Data & Methods

Descriptive statistics &  
testing

04

## Results

Proving the hypotheses,  
Analysis, & visualizations

05

## Recommendations

Call to Action / Insights for  
decision-making



# 01

## Introduction



# About the Data



## Ames, Iowa

- This data set was a sample of 1,460 houses that were sold between 2006 and 2010 in Ames, Iowa.
- There are 81 columns with a good distribution between numerical and categorical data.
- The description of each variable can be found in the data\_description.txt file on the website.



## Kaggle

This data was retrieved from [Kaggle](#)



# Business Objective

This presentation analyzes the factors that drive home prices by providing data-driven insights for an investment bank. Variables that provide impact on prices will be identified in order to provide informed decisions on how to allocate dollars earmarked for investment into mortgage-backed securities.



# Research Questions

1. What factors influence the selling price of a home? (characteristics such as quality, number of bedrooms, etc.)
2. Which type of homes are most frequently sold in the market; is there a relationship between the age or condition and their selling prices?
3. Are there any observable patterns or trends indicating which combinations of home types and neighborhoods attract the highest number of buyers/selling price?
4. Do certain neighborhoods have higher home prices compared to others?





---

02



## Hypotheses

# Hypotheses

1. There is a significant difference in the number of bedrooms between the treatment and control groups that would affect housing value.
2. There is a significant difference in the number of car garages between treatment and control groups that would affect housing value.
3. There is a significant difference in the number of fireplaces between treatment and control groups that would affect housing value.
4. There is a significant difference in the overall quality between treatment and control groups that would affect housing value.
5. There is a significant difference if homes have central air that would affect housing value.



# 03

# Data & Methods



# Descriptives

By utilizing descriptive statistics, important features of the dataset were condensed and examined, offering valuable observations on the central tendencies, variabilities, and interconnections among variables.

SalePrice	
Mean	180921.1959
Standard Error	2079.105324
Median	163000
Mode	140000
Standard Deviation	79442.50288
Sample Variance	6311111264
Kurtosis	6.53628186
Skewness	1.88287576
Range	720100
Minimum	34900
Maximum	755000
Sum	264144946
Count	1460

GarageCars	
Mean	1.767123288
Standard Error	0.019558128
Median	2
Mode	2
Standard Deviation	0.74731501
Sample Variance	0.558479724
Kurtosis	0.220997764
Skewness	-0.34254893
Range	4
Minimum	0
Maximum	4
Sum	2580
Count	1460

Fireplaces	
Mean	0.613013699
Standard Error	0.01687169
Median	1
Mode	0
Standard Deviation	0.644666386
Sample Variance	0.41559475
Kurtosis	-0.217237208
Skewness	0.649565183
Range	3
Minimum	0
Maximum	3
Sum	895
Count	1460

OverallQual	
Mean	6.099315068
Standard Error	0.036194674
Median	6
Mode	5
Standard Deviation	1.382996547
Sample Variance	1.912679448
Kurtosis	0.096292778
Skewness	0.216943928
Range	9
Minimum	1
Maximum	10
Sum	8905
Count	1460



# Methods

Find correlations between numerical data and the housing sale price

Used descriptive statistics, A/B testing with t-tests, PivotTables, and PivotCharts to provide insights into central tendencies and relationships between variables.

Assess the quality of the data, handling missing values, and detecting outliers.

Incorporate visualizations to identify patterns and relationships with data.



# Methods explained

This project was done using Excel, splitting data into treatment and control groups, while utilizing pivot tables and the Data Analysis ToolPak to manipulate data.

Based on insights from using Exploratory Data Analysis, new columns of data were created in order to get the proper data needed for t-tests.

Confidence intervals were also calculated for the hypotheses in order to accept or reject them.



---

04

# Results



# Neighborhoods



The top neighborhood to bring in the highest average sales for homes is: North Ames, which accounted for 14.42% of total sales and generated a revenue of \$32,815,593.

College Creek and Northridge Heights are other neighborhoods that could be of interest.



# Bedrooms

3 bedroom homes were sold the most, comprising of 55.07% of total home sales with a revenue of \$145,569,724.

With 95% confidence,  
the difference in  
means is between  
\$32,101 and \$13,617.

Rejected the null that  
there is no significant  
difference in means  
between houses that  
have 2 bedrooms and  
those that have 3.

t-Test: Two-Sample Assuming Unequal Variances		
	2 bed	3 bed
Mean	158197.6592	181056.8706
Variance	6114787424	4069664855
Observations	358	804
Hypothesized Mean Difference	0	
df	577	
t Stat	-4.857927362	
P(T<=t) one-tail	7.65105E-07	
t Critical one-tail	1.647498746	
P(T<=t) two-tail	1.53021E-06	
t Critical two-tail	1.964083864	
total sample size	1162	
mean difference	-22859.21143	
standard error of difference	4705.548215	
margin of error	9242.091322	
c.i. lower limit	\$ (32,101.30)	
c.i. upper limit	\$ (13,617.12)	



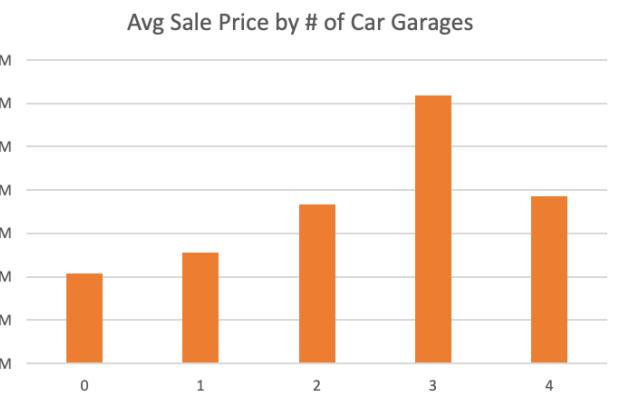
# Car Garages

2 car garages sold the most consisting of 56.44% of homes sold with the revenue at \$151,493,771.

With 95% confidence, the difference in means is between \$73,817 and \$90,507.

Rejected the null that there is no significant difference in means between houses with and without garages.

t-Test: Two-Sample Assuming Unequal Variances		
	Yes	No
Mean	185479.5112	103317.284
Variance	6244775286	1076825760
Observations	1379	81
Hypothesized Mean Difference	0	
df	143	
t Stat	19.4619501	
P(T<=t) one-tail	2.49937E-42	
t Critical one-tail	1.655579143	
P(T<=t) two-tail	4.99874E-42	
t Critical two-tail	1.976692198	
total sample size	1460	
mean difference	82162.22729	
standard error of difference	4221.685229	
margin of error	8344.972255	
c.i. lower limit	\$ 73,817.26	
c.i. upper limit	\$ 90,507.198	



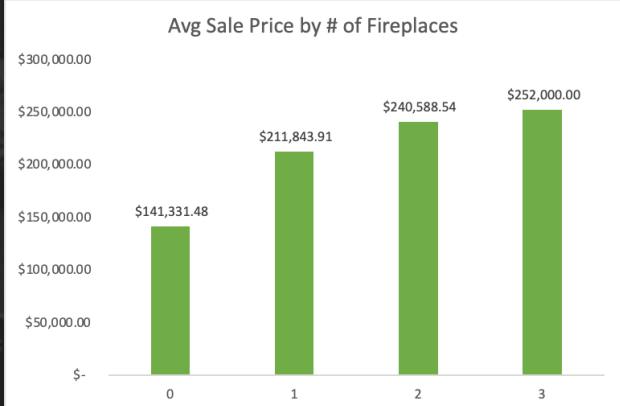
# Fireplaces

The average sale price was higher for homes with 2-3 fireplaces inside, but the most sales came from homes with 0-1 fireplace, with 47.25% (0) and 44.52% (1) of total home sales along with \$97,518,723 (0) and \$137,698,541 (1) in revenue.

With 95% confidence, the difference in means is between \$68,087 and \$82,044.

Rejected the null that there is no significant difference in means between houses that have fireplaces and those that don't.

t-Test: Two-Sample Assuming Unequal Variances		
	Yes	No
Mean	216397.6922	141331.4826
Variance	7541863186	1970453180
Observations	770	690
Hypothesized Mean Difference	0	
df	1172	
t Stat	21.10537632	
P(T<=t) one-tail	2.31292E-84	
t Critical one-tail	1.646154806	
P(T<=t) two-tail	4.62583E-84	
t Critical two-tail	1.961990163	
total sample size	1460	
mean difference	75066.2096	
standard error of difference	3556.734002	
margin of error	6978.277125	
c.i. lower limit	\$ 68,087.93	
c.i. upper limit	\$ 82,044.49	



# Overall Quality

Higher quality homes (6-10 rating) sold more than lower quality homes, accounting for 63.15% of total home sales and brought in \$196,554,295 in revenue.

With 95% confidence, the difference in means is between \$93,420 and \$81,678.

Rejected the null that there is no significant difference in means between houses with a quality rating of 1-5 (low) and those with a quality rating of 6-10 (high)

t-Test: Two-Sample Assuming Unequal Variances		
	<i>low</i>	<i>high</i>
Mean	125633.1803	213182.5325
Variance	960961515	6609906283
Observations	538	922
Hypothesized Mean Difference	0	
df	1299	
t Stat	-29.2559096	
P(T<=t) one-tail	2.9591E-145	
t Critical one-tail	1.646027501	
P(T<=t) two-tail	5.9183E-145	
t Critical two-tail	1.961791887	
total sample size	1460	
mean difference	-87549.35224	
standard error of difference	2992.535643	
margin of error	5870.732146	
c.i. lower limit	\$ (93,420.08)	
c.i. upper limit	\$ (81,678.62)	



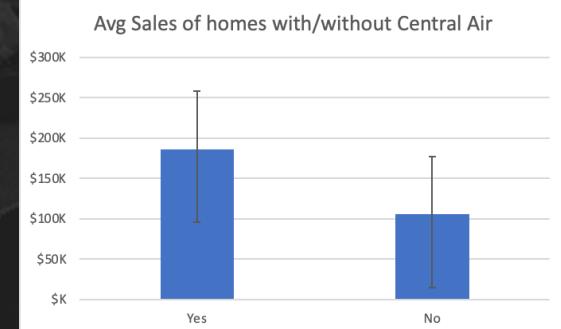
# Central Air

Homes with Central Air sold more than homes without it, accounting for 93.49% of total home sales and brought in \$254,144,859 in revenue.

With 95% confidence,  
the difference in  
means is between  
\$71,662 and \$90,182.

Rejected the null that  
there is no significant  
difference in means  
between houses that  
have central air and  
those that do not.

t-Test: Two-Sample Assuming Unequal Variances		
	Yes	No
Mean	186186.7099	105264.074
Variance	6210260622	1654152526
Observations	1365	95
Hypothesized Mean Difference	0	
df	149	
t Stat	17.26777256	
P(T<=t) one-tail	1.10128E-37	
t Critical one-tail	1.655144534	
P(T<=t) two-tail	2.20256E-37	
t Critical two-tail	1.976013178	
total sample size	1460	
mean difference	80922.63621	
standard error of difference	4686.339011	
margin of error	9260.267642	
c.i. lower limit	\$ 71,662.37	
c.i. upper limit	\$ 90,182.90	





# Top variables that drive home prices

---

1. Overall Quality
2. Garage Cars
3. Fireplaces
4. Year remodel added
5. Year built
6. Neighborhoods

# 05

## Recommendations



# Call to Action

Based on these findings, specific factors exert a significant influence on the prices of homes in the market. When making informed decisions about investing in mortgage-backed securities, it is crucial to take into account these factors and their impact on home prices.





# Recommendations

Give priority to three-bedroom homes, as they have the highest sales volume and generate substantial revenue.

Pay close attention to neighborhoods like North Ames, College Creek, and Northridge Heights, which have shown promising sales performance.

Give preference to homes with Central Air Conditioning, as they are in high demand among buyers.

Consider properties with two-car garages, as they have demonstrated strong sales performance.

Evaluate homes with an overall quality rating of 6 to 10, as they attract a significant number of buyers.