# XOrithm

Presentation by
Peter Romany

Data Science THA

Ames Housing Overview

Overview

- ✓ **Upper and Lower Bound :** The spread or dispersion of data points around the mean (real range of mean) – (115.18K , 261.95K)
- ✓ **Mean :** The value which most values are centered about it (balance point)- (188.57K)
- ✓ **Coefficient of Variation (CV) Interpretation for Sale Price:**

- • **Moderate Coefficient of Variation** (CV): A CV of 39% indicates a moderate level of dispersion in sale prices. This means the sale prices are somewhat spread out from the mean, but not excessively so.
- • **Interpretation:** The sale prices exhibit a moderate degree of variability across transactions. While there is some diversity in the sale prices, with both lower-priced and higher-priced properties, the spread is not extremely wide. This suggests a balanced real estate market with a reasonable range of property values.

- ✓ **CV Rate:** This represents how much of the data falls within this range. For the Sale Price, around 72.00% of the transactions are within boundaries of the mean.
- ✓ **Approximately 61.29%** of the data points fall within a low range. Specifically, out of 899 data points, 551 fall within this range.
- ✓ **The distribution** is described as "most in low," indicating that the majority of data points are concentrated towards the lower end of the distribution where mode(most frequently occurring value) locate.
- ✓ **This is further** elaborated by specifying that the mode is less than both the median and mean(mode < median= 174K < mean= 188.57K), suggesting a concentration of values towards the lower end of the distribution.


Count of SalePrice by SalePrice (bins)
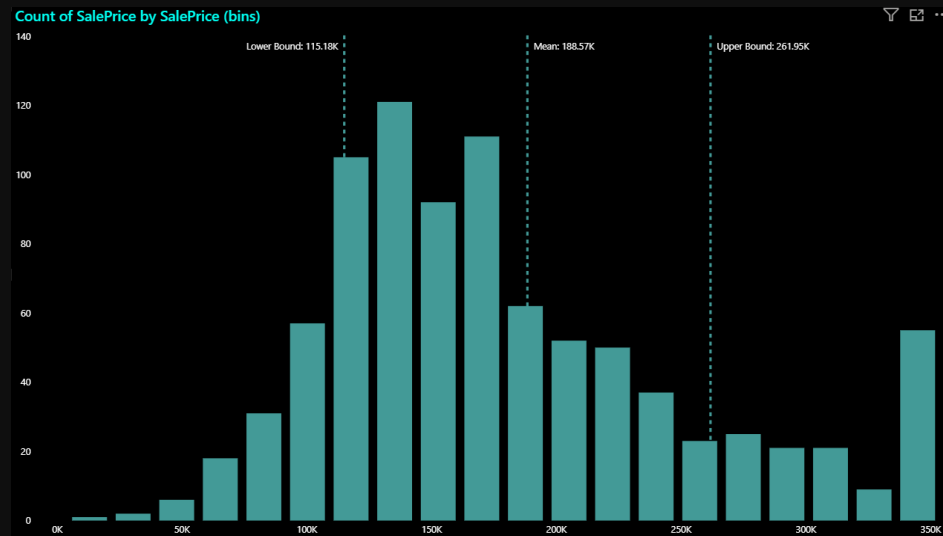
Insights

- ✓ **Maximum Sale Price:** The highest sale price recorded: $755,000
- ✓ **Minimum Sale Price:** The lowest sale price recorded: $13,100
- ✓ **Total Count of Sale Prices:** The total number of sale prices recorded: 899
- ✓ **Top 5 Distinct Sale Prices:** The five most frequently occurring sale prices and their counts: ($145,000, $155,000, $140,000, $185,000, $147,000)
- ✓ **Mode for Sale Price:** The most common sale prices: ($145,000, $155,000)
- ✓ **Total Sum of Sale Prices:** The sum of all sale prices recorded: $172,815,786
- ✓ **Top 5 Largest Sale Prices:** The five highest sale prices recorded: ($755,000, $745,000, $615,000, $591,587, $500,067)
- ✓ **Top 5 Smallest Sale Prices:** The five lowest sale prices recorded: ($13,100, $40,000, $46,500, $52,000, $55,000)

Market Overview : House Sale Price
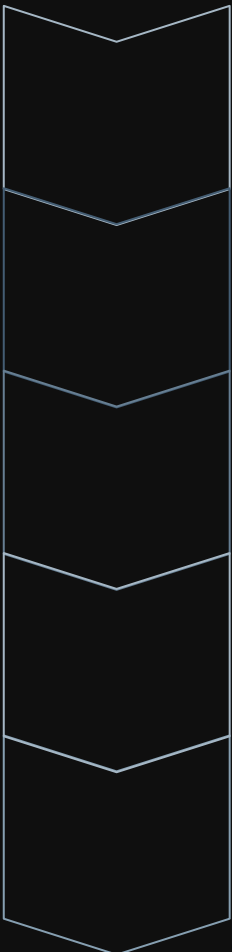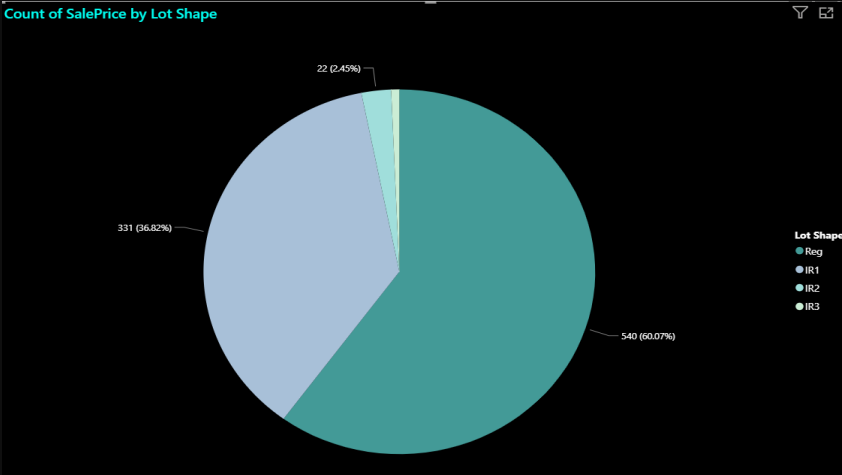
Overview

- ✓ This Pie Chart clarifies the importance of the shape of the home's area. The chart shows that more than 60% of sales go to homes with a regular shape, followed by 36% for homes with a slightly irregular.
- ✓ This also indicates that there is a lot of land that must be used for another project, unlike homes.

Volume wise lot shape Market

Count of SalePrice by Lot Shape

22 (2.45%)

331 (36.82%)

540 (60.07%)

Lot Shape
- ● Reg
- ● IR1
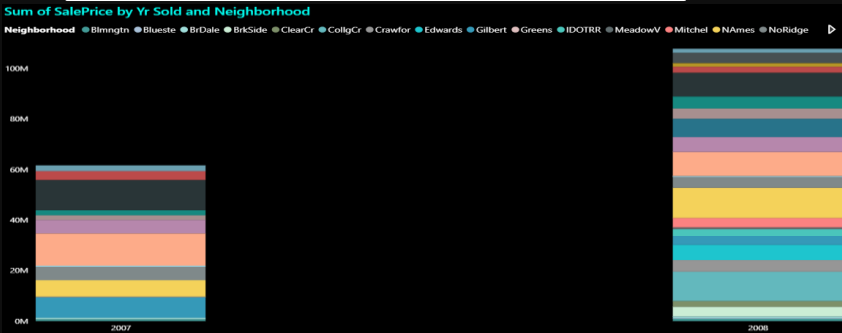- ● IR2
- ● IR3

House Sale Industry

- ✓ The year 2008 indicates more sales than 2007, as its sales reached more than 100 million, while sales in 2007 were only 60 million.
- ✓ We find that regions such as (Nridght and somerst) accounted for the most sales in 2007.
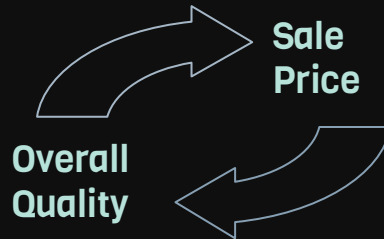- ✓ While the best selling areas in 2008 called NAmes and CollgCr.

Sale Total Volume by Region 2007and2008

Sum of SalePrice by Yr Sold and Neighborhood

Neighborhood ● Blmngtn ● Blueste ● BrDale ● BrkSide ● ClearCr ● CollgCr ● Crawfor ● Edwards ● Gilbert ● Greens ● IDOTRR ● MeadowV ● Mitchel ● NAmes ● NoRidge ▶

100M

80M

60M

40M

20M

0M

2007   2008

Source: Euromonitor International

Spearmanr Correlation

Sale Price

Overall Quality

r Interpretation

"r = 0.813 is Very strong positive correlation. "

R-squared

"0.66% of the variance in [Sale Price] can be explained by [Overall Quality]."

p-value

" 1.11 (statistically significant with 99% confidence)."

Focus on Property Quality: Invest in higher-quality properties to potentially maximize sale prices.

Pricing Strategy Adjustments: Sellers could adjust their pricing strategies to reflect the quality rating of their properties, potentially setting higher prices for those with better overall quality ratings and vice versa.

Marketing Emphasis on Quality: Real estate agents and developers may emphasize the quality aspects of properties in their marketing efforts, highlighting features that contribute to higher overall quality ratings to attract potential buyers.

Ridge Regression

Sale Price

Independent variables
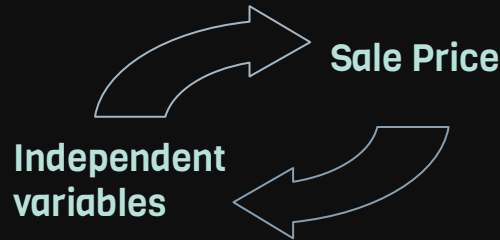
r Interpretation

"By 0.93 %, my points follow a constant rate of change. "

R-squared

"By 0.87 %, my X(IV) explains or determines the change observed in y(DV)."

Intercept(bias)

"An intercept of 188578 might represent an estimated baseline."
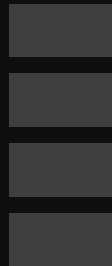
Slope (Coefficient)

"Slope: 19479, indicating the change in [Independent Variables] for each unit change in [Sale Price]. "

Model Evaluation

"Evaluating the model prediction with Mean Squared Error: 557967."

Model Prediction

"R-squared of Prediction: 0.88, indicating that the model will make correct predictions by: 0.88 %."

Robust ANOVA Table:

	df	sum_sq	mean_sq	F	PR(>F)
Q("Lot Shape")	3.0	3.111446e+11	1.037149e+11	20.513599	7.333581e-13
Residual	895.0	4.525038e+12	5.055908e+09	NaN	NaN

Dunn's test

ANOVA Test

Finding: Dunn's test results indicate significant differences in pairwise comparisons between some categories (IR1, IR2, IR3, Reg) regarding the variable of interest.

Benefit: Identifying these differences helps stakeholders understand the unique characteristics of each category and their potential impact on the variable being studied.

Practical Applications: Stakeholders can tailor their approaches more effectively to address specific characteristics or needs within each category, ultimately enhancing outcomes.
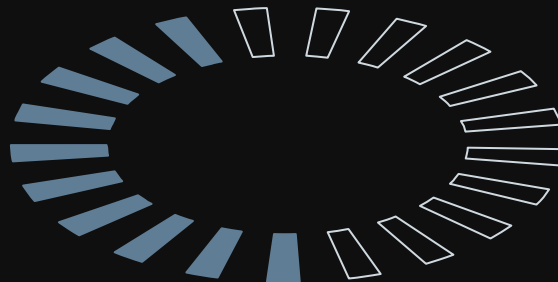
Dunn's test Results:

	IR1	IR2	IR3	Reg
IR1	False	False	False	True
IR2	False	False	False	True
IR3	False	False	False	False
Reg	True	True	False	False

Finding: ANOVA indicates that Statistically significant differences exist in the mean sale prices across different categories of lot shape.

Benefit: This information allows stakeholders to recognize the influence of lot shape on sale prices, aiding in property valuation and investment decisions.

Practical Applications: Stakeholders can strategically assess properties based on lot shape when buying or selling real estate.

Sale Price Analysis

We conducted an analysis comparing prices with and without air conditioning. The results revealed a substantial difference, emphasizing the significant impact of AC presence on sale prices.

On average, properties with AC commanded a price of $192,137, while those without averaged $98,088. This considerable variance underscores the importance of AC in determining property values.

independent sample t-test

This data enhances decision-making concerning property investments and renovations, providing stakeholders with vital market insights.

Stakeholders need to be informed about these results, underlining AC's critical role in price determination.

Invest in Properties with AC: Stakeholders may prioritize investing in properties equipped with air conditioning, recognizing their higher sale prices and potentially greater returns on investment.

Upgrade Existing Properties with AC Installation: Stakeholders owning properties without AC may decide to invest in installing air conditioning systems to increase their resale value based on the observed price differentials.

Adjust Pricing Strategies: Sellers may adjust their pricing strategies for properties with and without AC, reflecting the observed price disparities to optimize their sales and returns.
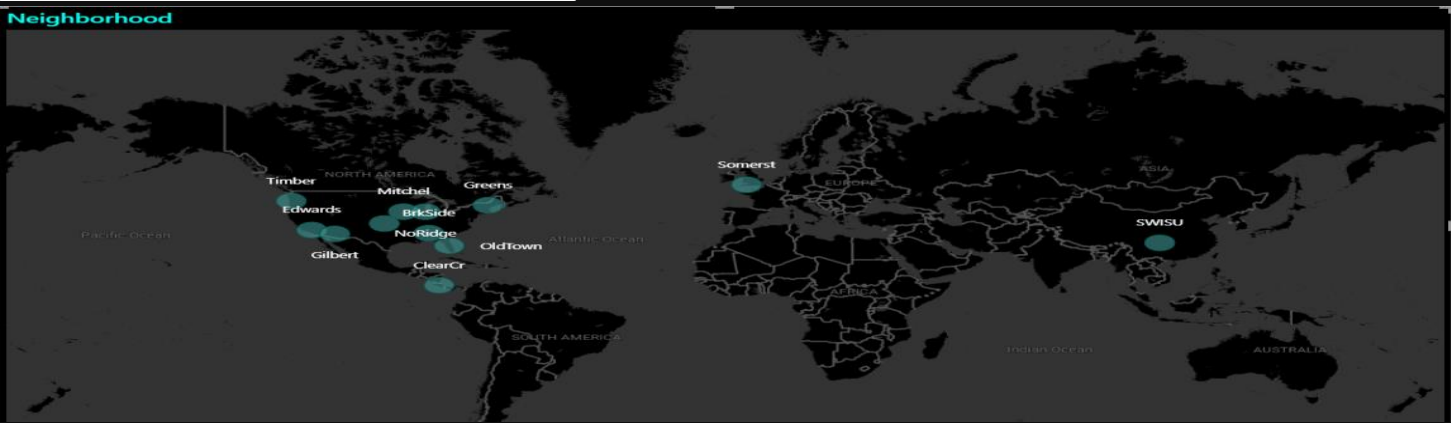
Market Positioning and Marketing Efforts: Real estate agents and developers could focus their marketing efforts on highlighting the presence of AC in properties as a key selling point, emphasizing the added value it brings to potential buyers.

Here are some potential decisions that stakeholders could make based on the findings of the sale price analysis:

Reasons to acquire : Geographical Presence & Segment Coverage

➢ When looking at the reasons for the acquisition, you will find that the geographic Presence and segment coverage are an important reasons for the increasing population growth, the huge sales rate.

➢ It takes up a large area geographically and also the technological development. Due to the strong economy, it is one of the luxurious cities.
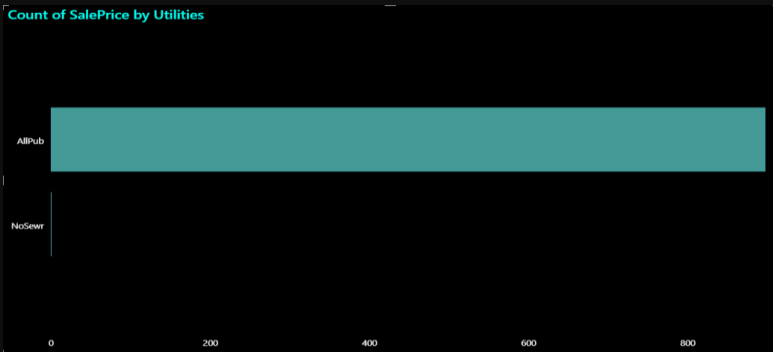
Geographical Presence

Neighborhood

Timber
Mitchel
Edwards
BrkSide
NoRidge
Gilbert
ClearCr
OldTown
Greens
Somerst
SWISU

NORTH AMERICA
Pacific Ocean
Atlantic Ocean
EUROPE
ASIA
AFRICA
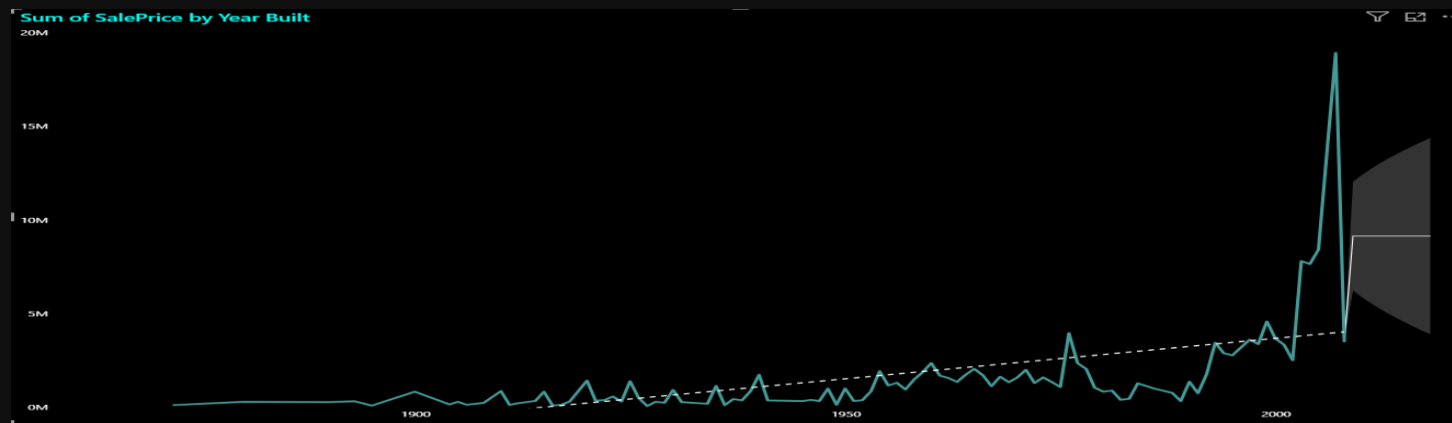SOUTH AMERICA
Indian Ocean
AUSTRALIA

Market Share Overview

✓ Can you imagine that percentage?
✓ Is it possible that one cannot buy a housing unit that does not have all public utilities?
✓ This is a modern and important city and is considered one of the high housing categories and wealthy groups of people.

Market share of top entities

Count of SalePrice by Utilities

AllPub

NoSewr

0   200   400   600   800

Trend and Forecasting

Sum of SalePrice by Year Built

Forecasting

From historical data, we see that the trend is always for there to be many sales over time, and with the current advancement of technology, housing sales in the city of Ames will increase significantly in the coming years and then will stabilize or decrease slightly.

Trend Sale Price Over Time

✓ It is normal for there to be a lot of zigzags over time.
✓ From 1872 until 1950, the city of Ames did not attract people
✓ But from 1993 until 2007, the city of Ames flourished due to its economy and beautiful climate, so sales became very high, but with the presence of other more beautiful places, the percentage of sales decreased.

Source: Euromonitor International

Thanks!