Analyzing Housing Prices in Floating Village Residential Zone

Noah Anderson

2023-02-21

Abstract

This study examines the relationship between zoning class, above ground living area, and sale price in the city of Ames. The research hypothesis posits that when controlling for year built, Floating Village-Residential (FV-R) homes are not more expensive on average than the dominant Residential Low Density (RL) zones. The analysis begins by adding a new column to the dataset to clearly distinguish between the two zoning values. A series of t-tests are conducted to compare the mean sale prices between RL and FV-R zones, accounting for year built and filtering out homes built before 1997. The results reveal a significant difference in sale prices between the two zoning classes ($p < .001$), with FV-R homes (M = \$21.9, SD = 5.3) being less expensive on average than RL homes (M = \$25.4, SD = 8.7). Moreover, the study explores the impact of above ground living area on sale price and identifies a strong positive correlation between the two variables (cor = .74). An ANOVA test uncovers a significant interaction between zoning class and above ground area ($F = 9.048$, $p < .003$), indicating that RL houses have a steeper rate of increase in price per square footage compared to FV-R houses. The findings highlight the importance of considering year built and above ground living area when evaluating the differences in sale prices between RL and FV-R zones. Furthermore, the paper discusses the implications of these findings for the city of Ames and challenges the common perception of FV-R as solely a retirement community. The study suggests that FV-R offers more affordable and diverse housing options while promoting environmental sustainability through increased green space and walkability. The authors recommend recognizing FV-R as a legitimate zoning class rather than a niche retirement community, considering its potential benefits for residents and the environment.

```r
library(AmesHousing)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
##
## The following object is masked from 'package:stats':
##
##     filter
```

About

This project was a class assignment for the intro to Data Science for Truman States Data Science and Analytic story masters program. This project showcases my skills in data tidying and performing complex statistical analyses. Methods such as multiple linear regression, proper handling of outliers, and analysis of variance are used throughout this project.

Introduction

Zoning codes play a pivotal role in regulating construction activities and spatial allocation. However, it is important to recognize that zoning codes can have unintended consequences on housing prices. For instance, suburban homes characterized by expansive square footage and widely spaced lots, designed to accommodate automobile usage, exhibit higher carbon intensity compared to more densely populated and pedestrian-friendly areas (Fredrick, 2022). The prevailing belief since the 1950s has been that the construction of low-density suburban housing trades off environmental conservation for economic stability, primarily by increasing the supply of affordable housing. Although this may have held true during the economic boom of the 1950s, it is no longer the case in the contemporary United States. In fact, a study conducted in 2022, which examined the impact of housing stock diversity on unemployment rates across 146 mid-sized cities, found that cities characterized by higher density and walkability experienced greater socioeconomic stability (Fredrick, 2022).

Since 1950, low-density housing has dominated new development projects, leaving homeowners and renters with a choice between two extremes: high-density housing in major cities or low-density housing in small to mid-sized towns and cities. This phenomenon, often referred to as "The Missing Middle" by renowned author and urban planning professor Daniel Parolek, has recently attracted the attention of urban planners who consider single-family and low-density zoning regulations as one of the fundamental causes of the housing crisis in the United States. Restricting density and imposing minimum lot sizes artificially limits the housing supply, consequently driving up home prices (Monkkonen, 2019).

The city of Ames, Iowa, has undergone significant changes to its zoning code over the years. In 1930, the Ames City Council introduced a zoning code that clearly defined classifications for residential, commercial, and heavy industrial districts (Prather, 1939). Concerns regarding overcrowding prompted the City Council to expand upon the code in 1939 (Prather, 1939). This amendment aimed to proactively limit density in Ames by establishing two residential districts: District A, which restricted new housing developments to single-family or two-family dwellings, and District B, which restricted housing developments to multi-family zoning. The 1939 code reflected the imminent national trend of suburban expansion. By 1983, a considerably complex zoning code was implemented in Ames, establishing seven distinct residential zones (Gibbons, 1983). This revised code included two low-density districts with minimum lot requirements of 20,000 square feet for the "Suburban Zone" and 10,000 square feet for the "Low-Density District." Such lot minimums impose a prerequisite for clearing a minimum lot size before constructing a home and serve as another mechanism to control density.

Presently, the City of Ames has comprehensively overhauled its zoning code and adopted four main residential districts, each with specific definitions. These include the Low-Density district, allowing no more than 7.26 dwellings per acre, the Urban Core Residential Medium-Density district (UCRM), also allowing no more than 7.26 dwellings per acre, the Residential Medium-Density district (RM), accommodating between 7.26 and 22.31 dwellings per acre, and the Residential High-Density district (RH), accommodating between 11.2 and 38.56 dwellings per acre (Ames Article 7, 2022). Each district has established a standard minimum lot size of 6,000 square feet for single-family dwellings and 7,000 square feet for two-family dwellings, both of which fall well below the state average. The significant reduction in lot size requirements suggests a trend toward greater flexibility regarding lot sizes.

Additionally, the city has introduced a new zoning concept called Floating Village Residential (F-VR) Districts, designed to resemble traditional walkable neighborhoods reminiscent of pre-1940 urban design in the

United States. The earliest recorded house built in an F-VR district in Ames dates back to 1997. Floating zones are used to describe designated areas that must meet certain conditions before receiving approval, as opposed to being predetermined spots on a zoning map (American Planning Association). In other words, floating zones serve as experimental areas where cities can test specific codes before implementing them on a larger scale. In the context of Ames, FV-R zones are areas where development adheres to planning standards prevalent in the United States before 1940 (Ames Article 12, 2022). Consequently, these zones prioritize walkable and densely populated neighborhoods that can be traversed within a 10 to 15-minute walk, with close proximity to green spaces, residential units, workplaces, and commercial amenities. This marks a decisive departure from the anti-density paradigm established in the 1940s.

Before officially designating FV-R as a zoning class with land allocation that could rival RL zones, the city must conduct a thorough study to evaluate the outcomes of the current experimental zones. Is this new housing model a viable planning structure accessible to all? A project conducted in 2016, utilizing the Ames data set, characterized FV-R as a high-priced retirement community with the highest median price among all zoning codes (Liu, 2016), while another project reported FV-R as having the highest price per square foot (Zheng, 2019). In this study, we challenge this notion and assert that, after controlling for the year of construction, FV-R houses actually exhibit statistically significant lower prices than RL houses.

Methods

In our analysis, we implemented the following methods to investigate the relationship between zoning classifications and housing prices in Ames:

1. Zoning Classification Column: We started by adding a new column to the Ames dataset specifically for zoning values. When the zoning classification was designated as “Residential_Low_Density,” we assigned the value “Low” to this new column. Similarly, when the zoning classification was identified as “Floating_Village_Residential,” we assigned the value “Floating_Village” to the new column. This step aimed to provide a clear distinction and facilitate further analysis based on zoning.

2. Creation of Filtered Table: Once the zoning classification column and corresponding values were established, we generated a new table by applying filters. Specifically, we filtered out homes with a living area greater than 4000 square feet and those built before 1997. This filtering allowed us to focus on a subset of homes for subsequent analyses.

3. T-Tests for Homes Built Before 1997: We conducted separate t-tests for homes built before 1997 and homes built after 1997. To begin, we examined the mean sale price differences between the “Low” zoning (Residential Low Density) and “Floating_Village” zoning (Floating Village Residential) groups. In this test, we divided the Sale_Price of each home by 10,000 using the Ames data. The data was then grouped based on the new zoning column, and summary statistics such as mean, standard deviation, and the total number of homes were calculated. These results allowed us to evaluate the significance of the mean differences between the two zoning groups.

4. T-Tests for Homes Built After 1997: Following the initial t-test analysis, we focused on excluding homes built before 1997. To do this, we first examined the correlation between the “Year_Built” and “Sale_Price” variables in the Ames dataset. By identifying the correlation, we were able to perform another t-test, this time excluding homes built before 1997. This test aimed to determine if there were significant mean and standard deviation differences between homes built before 1997 and those built after 1997, categorized by the zoning groups.

5. T-Test for Ground Level: Lastly, we conducted a t-test to investigate if there was a significant difference in ground level and whether it influenced housing prices. Initially, we examined the correlation between living area and sale price. Subsequently, we performed a t-test on the living area variable in relation to the new zoning column we created in the Ames dataset. This analysis provided mean and standard deviation data to evaluate any significant differences in the total square footage within specific zoning categories.

By employing these methods, we aimed to assess the relationship between zoning classifications and housing prices, specifically focusing on the differences in mean prices, standard deviations, and other relevant factors within different zoning groups.

```
ames <- make_ames()

ames <- ames %>%
  mutate(RL_or_FV = case_when(MS_Zoning == "Residential_Low_Density" ~ "Low", MS_Zoning == "Floating_Vi


# creating a data set that filters out pre-1997 houses
ames_1997 <- ames %>%
  filter(Gr_Liv_Area <4000) %>%
  filter(Year_Built >= 1997)

# This data frame contains only RL and FV-R which will simpilfy visualizations later in the paper.
FV_and_Low.df <- ames_1997 %>%
  filter(MS_Zoning == "Residential_Low_Density"| MS_Zoning == "Floating_Village_Residential")
```

Results

Research Hypothesis: When you control for year built, FV-R homes are not more expensive on average than RL zones.

First, we conducted an independent t-test to compare two zoning class groups: RL (Residential Low Density) and FV-R (Floating Village Residential), without accounting for the year built. In this t-test, the null hypothesis (H_0) states that there is no significant difference in the mean sale price of houses between the low density and FV-R residential zones.

```
t.test(Sale_Price/10000 ~ RL_or_FV, data = ames)
```

```
##
##  Welch Two Sample t-test
##
## data:  Sale_Price/10000 by RL_or_FV
## t = 5.7922, df = 180.88, p-value = 3.023e-08
## alternative hypothesis: true difference in means between group Floating_Village and group Low is not
## 95 percent confidence interval:
##  1.826624 3.714116
## sample estimates:
## mean in group Floating_Village              mean in group Low
##                        21.89869                        19.12833
```

```
descriptives <- ames %>%
  group_by(RL_or_FV) %>%
  summarize(mean = mean(Sale_Price)/10000, sd = sd(Sale_Price)/10000, n = n())
descriptives
```

```
## # A tibble: 3 x 4
##   RL_or_FV          mean    sd     n
##   <chr>            <dbl> <dbl> <int>
```

```
## 1 Floating_Village  21.9  5.27   139
## 2 Low               19.1  8.13  2273
## 3 <NA>              12.5  4.82   518
```

With $p < .001$, we rejected the null hypothesis and concluded that there was a significant difference in the scores for sale price (10k USD) between the FV-R group (M=21.9, SD=5.3) and the RL group (M=19.1, SD=4.8). This t-test indicated that the sample means differed by approximately $19,000, with the FV-R group exhibiting higher prices. However, the previous t-test did not account for the fact that the first recorded FV-R home was in 1997, which raises the possibility that our results may have been influenced by the sale prices of older homes in the RL group. Therefore, we proceeded to conduct the same analysis after excluding all houses built before 1997.

Excluding Houses Built Before 1997

Since the establishment of FV-R zones occurred in 1997, and considering the reasonably high correlation between sale price and year built (R=.558), we deduced that the higher sale prices observed for RL houses compared to FV-R houses could be attributed to the presence of older homes in the RL group. To address this disparity, we applied a filter to exclude all houses built before 1997 and focused solely on RL houses constructed during the era of FV-R. The null hypothesis for this analysis posits that there is no significant difference in sale prices between FV-R and RL houses built in 1997 or later.

```
cor(ames$Year_Built,ames$Sale_Price)
```

```
## [1] 0.5584261
```

```
t.test(Sale_Price/10000 ~ RL_or_FV, ames_1997)
```

```
##
##  Welch Two Sample t-test
##
## data:  Sale_Price/10000 by RL_or_FV
## t = -6.3207, df = 304.61, p-value = 9.24e-10
## alternative hypothesis: true difference in means between group Floating_Village and group Low is not
## 95 percent confidence interval:
##  -4.573158 -2.401717
## sample estimates:
## mean in group Floating_Village              mean in group Low
##                       21.89869                       25.38613
```

```
descriptives <- ames_1997 %>%
  group_by(RL_or_FV) %>%
  summarize(mean = mean(Sale_Price)/10000, sd = sd(Sale_Price)/10000, n = n())
descriptives
```

```
## # A tibble: 3 x 4
##   RL_or_FV          mean    sd     n
##   <chr>            <dbl> <dbl> <int>
## 1 Floating_Village  21.9  5.27   139
## 2 Low               25.4  8.67   718
## 3 <NA>              18.7  6.63    57
```

With $p < .001$, we rejected the null hypothesis and concluded that there was a significant difference in the scores for sale price (10k USD) between the FV-R group (M=21.9, SD=5.3) and the RL group (M=25.4, SD=8.7). This t-test revealed a sample mean difference of approximately $36,000, with the RL group exhibiting higher prices. By accounting for the year a house was built, we have demonstrated a reversal in the relationship, with the RL group being significantly more expensive than the FV-R group..

Total Square Footage and Average Sale Price

The aforementioned t-test did not provide insights into other variables that could potentially account for the difference in sale price. However, it is worth noting that the above ground level area of houses exhibits a relatively strong positive correlation with sale price (cor = .74), as indicated below. This implies that larger living areas of houses tend to correspond to higher mean prices.

```
cor(FV_and_Low.df$Gr_Liv_Area, FV_and_Low.df$Sale_Price)
```

```
## [1] 0.7403462
```

We conducted a t-test to examine whether there was a significant difference in above ground level area between the two groups.

```
t.test(Gr_Liv_Area~ RL_or_FV, ames_1997)
```

```
##
##  Welch Two Sample t-test
##
## data:  Gr_Liv_Area by RL_or_FV
## t = -5.4301, df = 223.77, p-value = 1.463e-07
## alternative hypothesis: true difference in means between group Floating_Village and group Low is not
## 95 percent confidence interval:
##  -242.4318 -113.3257
## sample estimates:
## mean in group Floating_Village              mean in group Low
##                        1581.568                       1759.447
```

```
descriptives <- ames_1997 %>%
  group_by(RL_or_FV) %>%
  summarize(mean = mean(Gr_Liv_Area), sd = sd(Gr_Liv_Area), n = n())
descriptives
```

```
## # A tibble: 3 x 4
##   RL_or_FV          mean    sd     n
##   <chr>            <dbl> <dbl> <int>
## 1 Floating_Village 1582.  341.   139
## 2 Low              1759.  412.   718
## 3 <NA>             1296.  343.    57
```

With $p < .001$, we rejected the null hypothesis and concluded that there was a significant difference in the scores for total square footage between the FV-R group (M=1580, SD=340) and the RL group (M=1760, SD=340). This t-test revealed a sample mean difference of approximately 170 SF, with the RL group having a larger average. Considering the positive correlation between sale price and total square footage, we inferred that the greater square footage of RL houses contributes to some extent to the price difference.

While the disparities in average total square footage shed light on the potential reasons for the price variation, they do not account for any potential differences in the relationship between square footage and sale price within the two groups. In the subsequent subsection, we compared linear regressions for sale price and total square footage separately for the FV-R and RL groups.

Testing Interaction

To examine the potential impact of zoning class on the relationship between above ground level area and sale price, we conducted an ANOVA test to determine if there is a statistically significant interaction.

```
FV_and_Low.df %>% anova_test(Sale_Price ~ MS_Zoning*Gr_Liv_Area)
```

```
## ANOVA Table (type II tests)
##
##                  Effect DFn DFd        F         p p<.05   ges
## 1             MS_Zoning   1 853    2.448  1.18e-01       0.003
## 2           Gr_Liv_Area   1 853 1005.235 2.23e-146     * 0.541
## 3 MS_Zoning:Gr_Liv_Area   1 853    9.048  3.00e-03     * 0.010
```

We observed a statistically significant interaction between zoning class and total square footage as it relates to sale price ($F = 9.048$, $p < .003$). The effects of total square footage on sale price differ depending on whether we are examining FV-R houses or RL houses, with RL exhibiting a steeper slope. The scatter plot below depicts the relationship between sale price and above ground level area, including trend lines for RL and FV-R. Notably, the trend line for RL demonstrates a steeper slope compared to FV-R, indicating that the positive impact of total square footage on sale price is more pronounced for RL houses.

```
ggplot(FV_and_Low.df, mapping = aes(x = Gr_Liv_Area, y = Sale_Price/10000, col = MS_Zoning, shape = MS_Z
  geom_point() +
    ggtitle("Regressions for RL and FV-R") + xlab("Above Ground Living Area (SF)") + ylab("Sale Pricek
  guides(col = guide_legend("Zoning Class"), shape = guide_legend("Zoning Class")) +
  geom_smooth(method = lm, fill = NA, fullrange = TRUE) + xlim(0,6000) + scale_color_discrete(labels = c
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Discussion

Our findings provide robust statistical evidence that FV-R homes are generally less expensive than the dominant RL homes, even after accounting for the year built. A significant portion of this price difference can be attributed to variations in the average above ground living area, with RL homes exhibiting significantly larger living areas.

Furthermore, when considering the relationship between price-per-square-foot and zoning classifications (FV-R and RL), our analysis revealed a substantial impact of year built. We discovered a significant interaction between square footage and these two zoning classifications, with RL houses exhibiting a steeper slope. This implies that constructing larger homes in RL areas is more advantageous, as it corresponds to a higher price per square foot compared to FV-R. Consequently, there is a market incentive to build larger homes in RL areas, while FV-R houses do not experience the same dramatic effects on sales price. This finding, coupled with the limited space in denser neighborhoods, helps explain why the average square footage is significantly higher in RL homes compared to FV-R.

These findings have significant implications for the city of Ames when evaluating the overall success of FV-R areas. Our analysis challenges the common perception that FV-R homes are the most expensive and primarily suitable as retirement communities. By comparing FV-R prices to houses in RL built in the same year or after the establishment of the first FV-R, we demonstrate that RL homes are likely to be much more expensive. It is crucial for the city to consider these findings and avoid drawing incorrect conclusions by failing to account for important variables such as year built.

We propose that FV-R areas actually offer more affordable and diverse housing options for Ames residents, in addition to their lower environmental impact due to increased green spaces and walkability, which contributes to reduced car emissions. Moreover, the differences in price-per-square-foot rates provide market incentives to utilize fewer resources in FV-R zones, leading to positive environmental outcomes. The city should view FV-R as a legitimate and valuable zoning classification rather than a niche retirement community, thereby providing residents with more housing choices and contributing to sustainable urban development.

Limitations

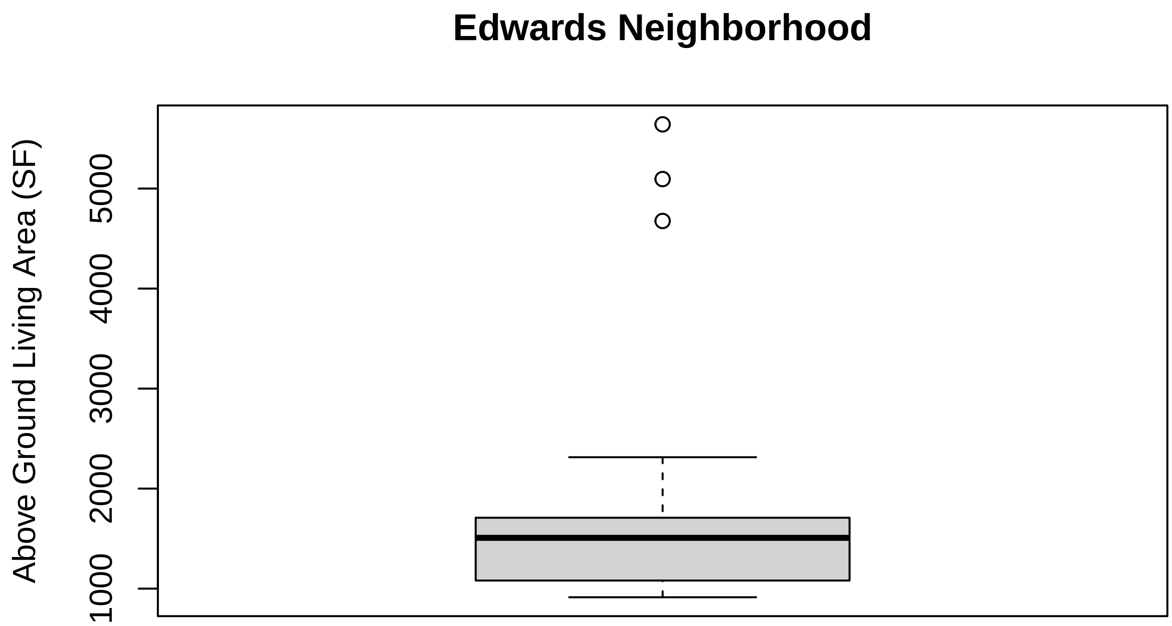
One crucial aspect to consider in our analysis of the interaction is the exclusion of outliers in the overall above ground square footage. Specifically, we encountered three data points originating from the Edwards neighborhood, which were constructed and sold between 2007 and 2008. These data points exhibited remarkably high above ground living areas but surprisingly low sale prices. The inclusion of these three outliers significantly disrupted the interaction test between zoning class and above ground area, rendering it inconclusive. Despite our diligent efforts to investigate the reasons behind these unusually low sale prices, we were unable to identify a probable explanation. It is worth noting that these data points shared a median year built and year sold of 2007, effectively ruling out the impacts of the Great Recession, as they were sold during a period of substantial price escalation before the economic downturn. Furthermore, these properties exhibited average overall condition and displayed excellent overall quality. By meticulously excluding these outliers from our analysis, we aimed to ensure the robustness and reliability of our findings regarding the interaction between zoning class, above ground area, and sale price. Although the anomalous data points remain unexplained, their exclusion enables us to derive more accurate and dependable conclusions concerning the general trends and patterns observed in the dataset. Consequently, it is vital to exercise caution and prudence when handling outliers in statistical analyses to ensure accurate interpretations of the underlying data patterns.

```
ames_outlier_1997 <- ames %>%
  filter(Gr_Liv_Area  > 4000) %>%
  filter(Year_Built >= 1997)


edwards.df <- ames %>%
  filter(Year_Built >= 1997) %>%
  filter(Neighborhood == "Edwards")
```

We were intrigued by the peculiar nature of these three data points, all originating from the same time period and neighborhood. Consequently, we conducted a more in-depth examination of the Edwards neighborhood. To visualize the distribution of above ground living area in this neighborhood, we present a box plot below. It is noteworthy that the three outlier data points stand out as highly atypical, even within the context of the Edwards neighborhood. While we acknowledge that our inability to provide a satisfactory explanation for these outliers presents a limitation to our interaction test results, we still maintain that the insights gained from examining the interaction between zoning class, total square footage, and sale price are meaningful. Importantly, we observed that the inclusion or exclusion of these outliers had no discernible impact on the results of the average sale price and square footage t-tests, thus further affirming the validity of our findings in those aspects.

```
boxplot(edwards.df$Gr_Liv_Area, ylab = "Above Ground Living Area (SF)", main = "Edwards Neighborhood")
```



References

Prather, J.W. (1939, March 24). Ames Daily Tribune, p. 4.

Gibbons, N. (1983, March 19). Ames Daily Tribune, p. 59.

Monkkonen, Paavo (2019) The Elephant in the Zoning Code: Single Family Zoning in the Housing Supply Discussion, Housing Policy Debate, 29:1, 41-43, DOI: 10.1080/10511482.2018.1506392

Frederick, C. (2022). Economic Sustainability and 'Missing Middle Housing': Associations between Housing Stock Diversity and Unemployment in Mid-Size US Cities. Sustainability, 14(11), 6817.

Liu, Ziqiao (2016). NYC Data Science Blog.

Ames, IA. City Code. Chapter 29 Article 7 (2022).

Ames, IA. City Code. Chapter 29 Article 12 (2022).

Property topics and concepts. American Planning Association. (n.d.). Retrieved March 5, 2023, from https://www.planning.org/divisions/planningandlaw/propertytopics.htm#Floating

Zheng, Jingwen . (2019, August 6). House prices: Regression Techniques. Jingwen Zheng. Retrieved March 5, 2023, from https://jingwen-z.github.io/house-prices-regression-techniques/

Boylan, J. (2023, February 2). How much value does a finished basement add? Springs Homes. Retrieved March 5, 2023, from https://springshomes.com/blog/how-much-value-does-a-finished-basement-add/#:~:text=The%20Appraiser%20does%20give%20credit,the%20main%20level%20square%20footage.