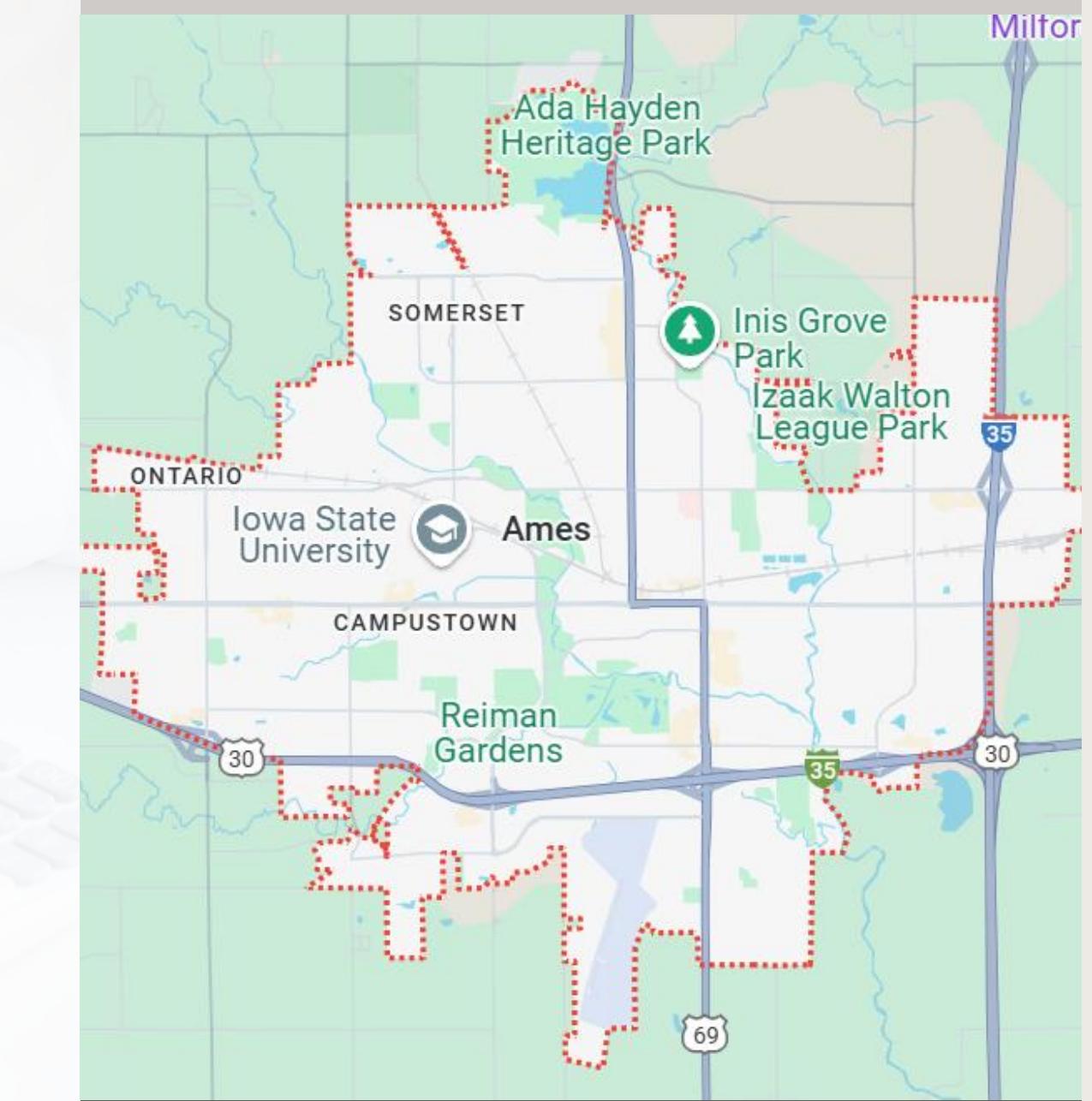


# BU ABA Coding Challenge

Prabu Jeyabalan

AMES HOUSING - IOWA



REAL ESTATE



# Abstract

## 01 Data Overview:

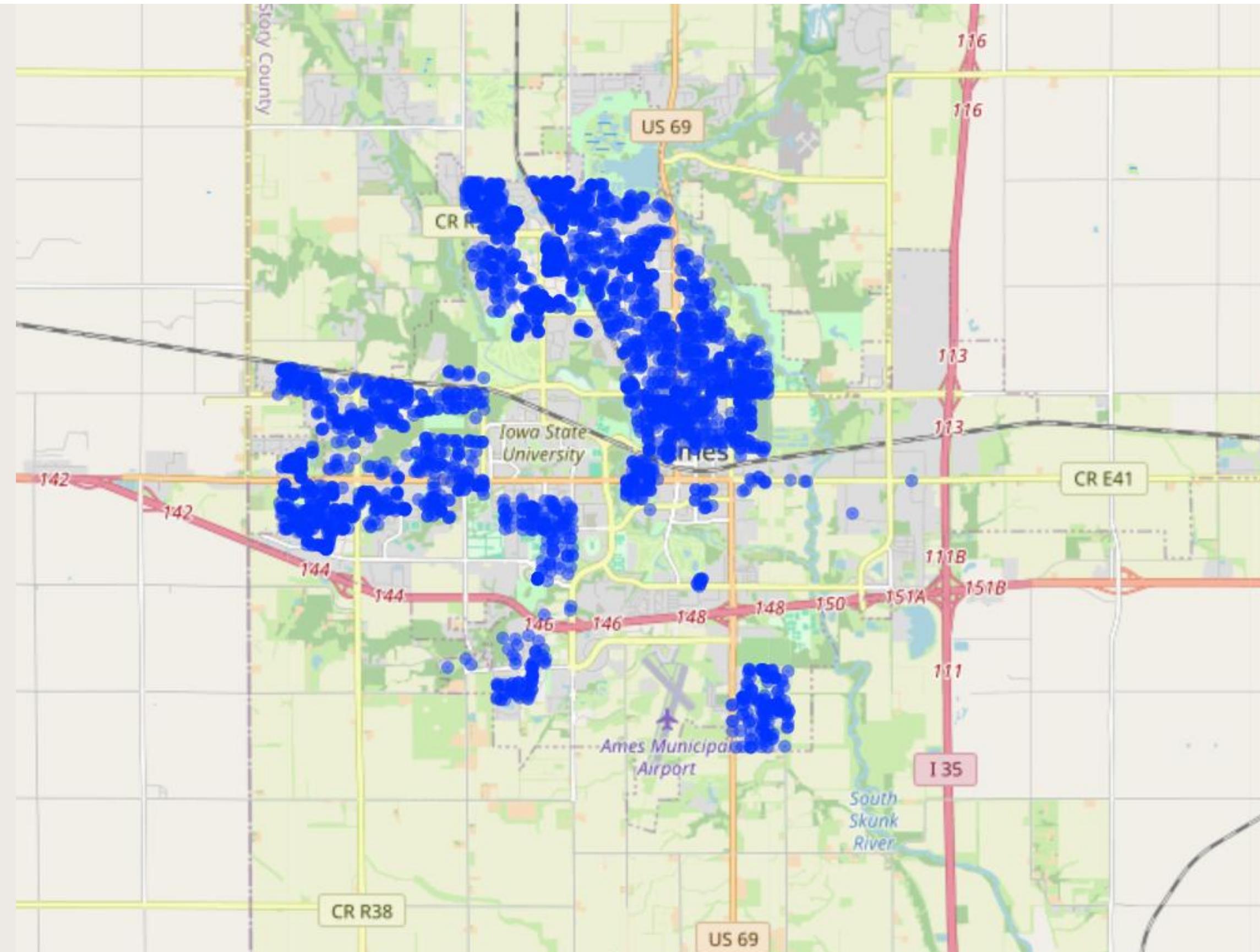
This dataset covers the sale of 2,930 residential properties in Ames, Iowa, from 2006 to 2010. It includes a variety of variables—23 nominal, 23 ordinal, 14 discrete, and 20 continuous—all of which are used to predict the sale price of the properties.

## 02 Assumption:

The majority of buyers intend to purchase a home to live in with minimal renovations. Thus, factors such as the property's condition and features, along with its location, are important. If a complete reconstruction were the goal, only the location would matter, making the current condition irrelevant.

Based on this assumption, a customer looking to move to Ames City is interested to understand the market and what mostly influences the price.

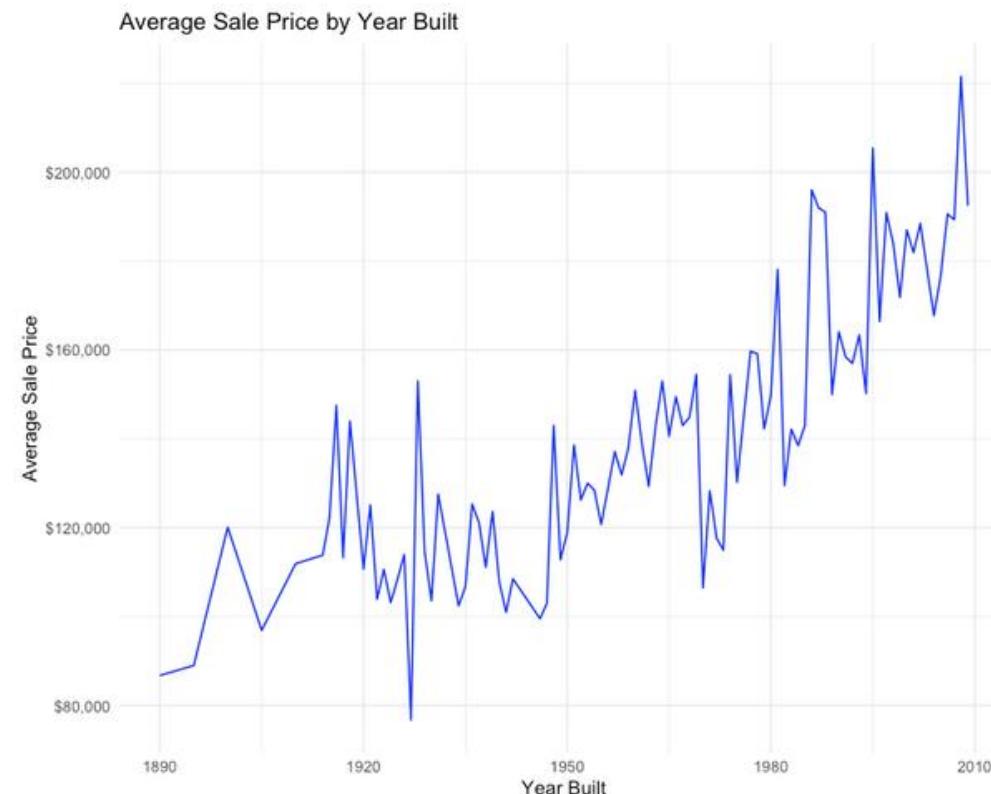
# Property Sale Types & Key Factors Influencing Sale Price



# Dashboard

## Average Sale Price

\$180,796.10



## Dominant Sale Method

Warranty Deed – Conventional



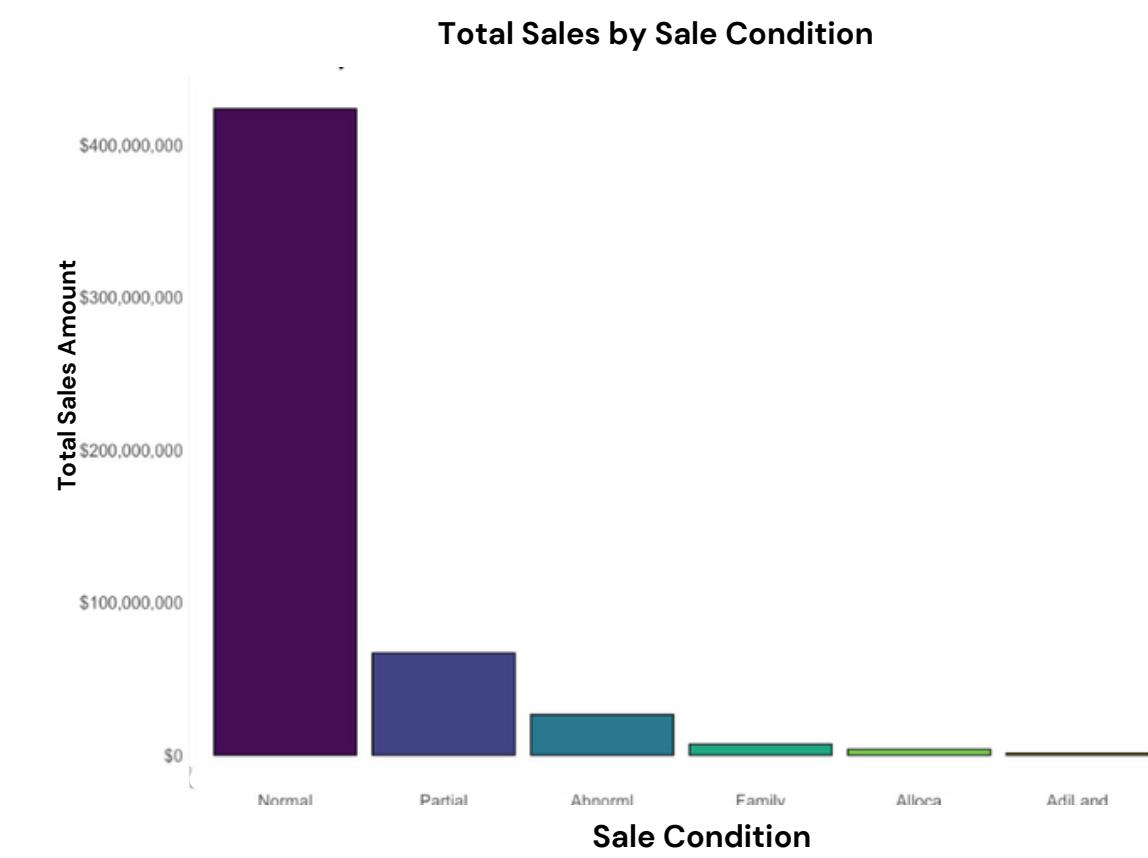
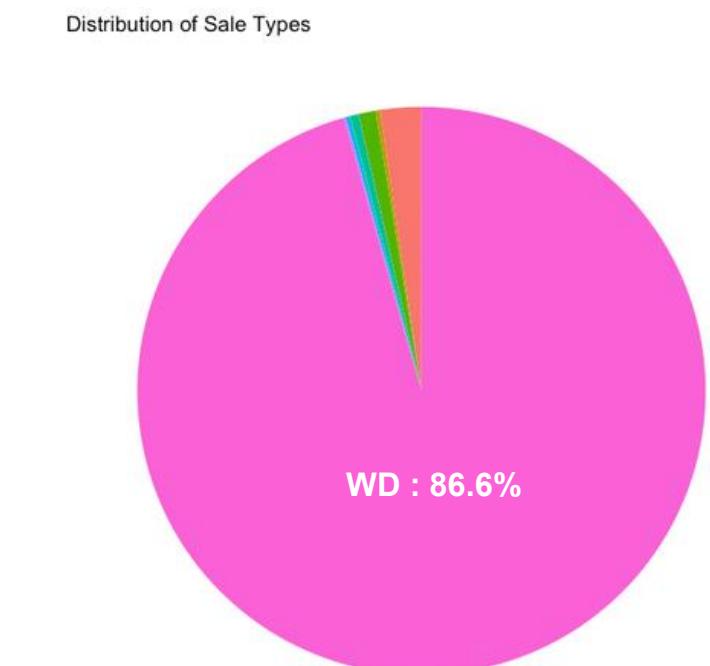
Price increases with time. Almost every increase in price is followed by a decline in price.

Warranty Deed - Conventional is the dominant sale method with the mid-range price. New construction sales being the second most common with highest average price.

Most sales have a condition of Normal type.

## Dominant Sale Condition

Normal (80%)

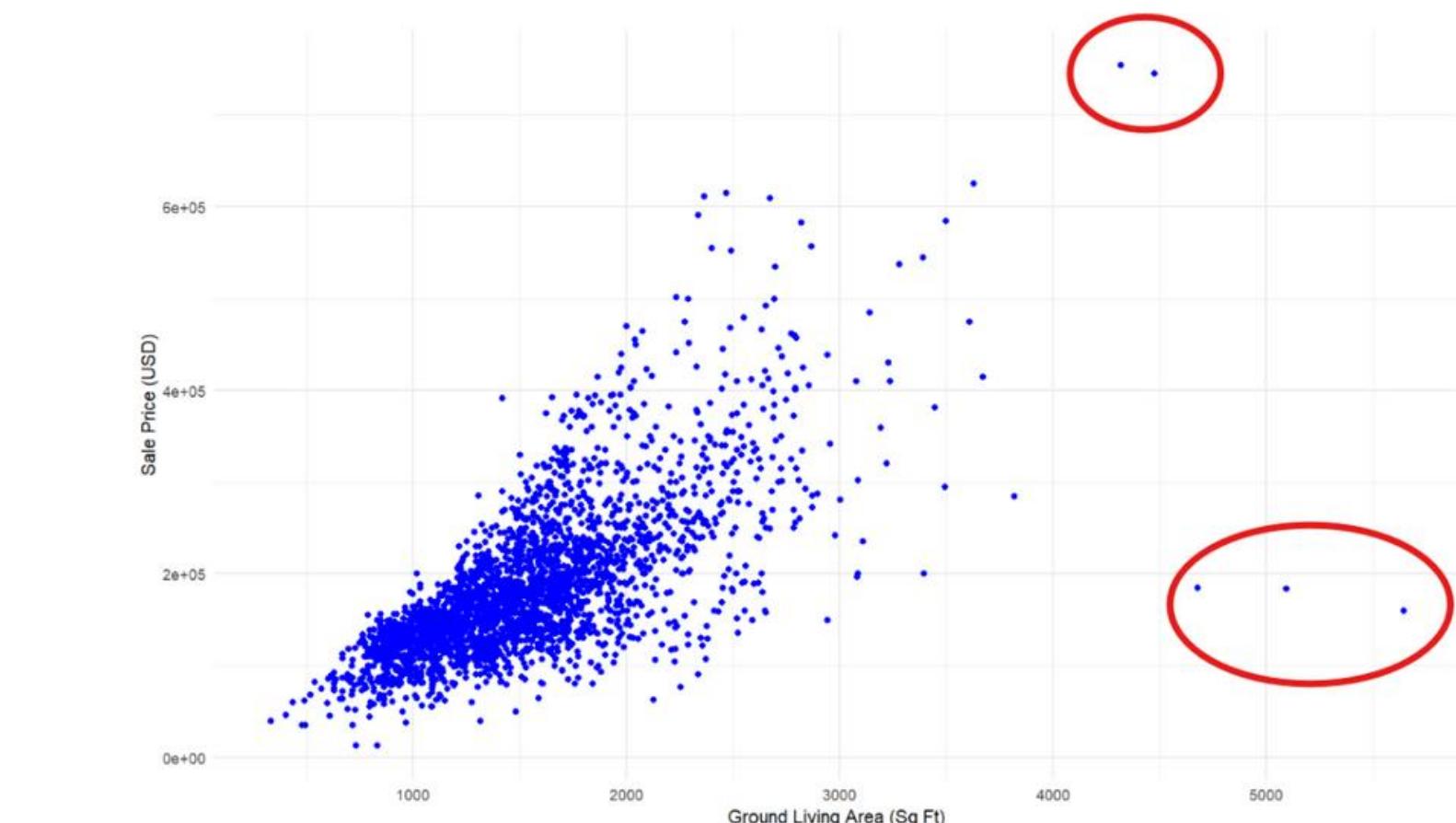


# Special Notes on Data Filterings:

- Above grade (ground) living area square feet less than 1500
- Sale Method as “Normal” only

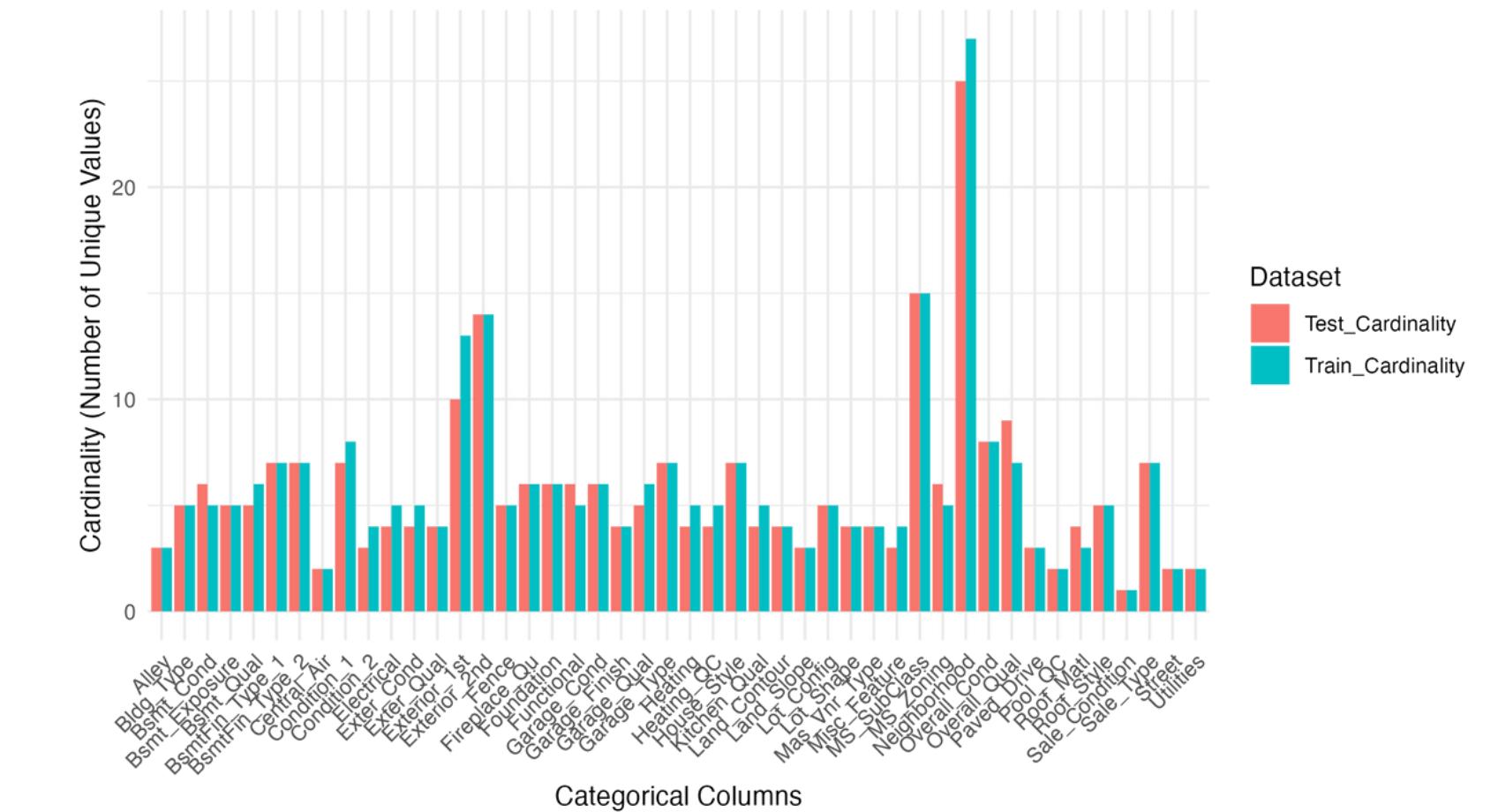


Relationship Between General Living Area (Sq Ft) and Sale Price)



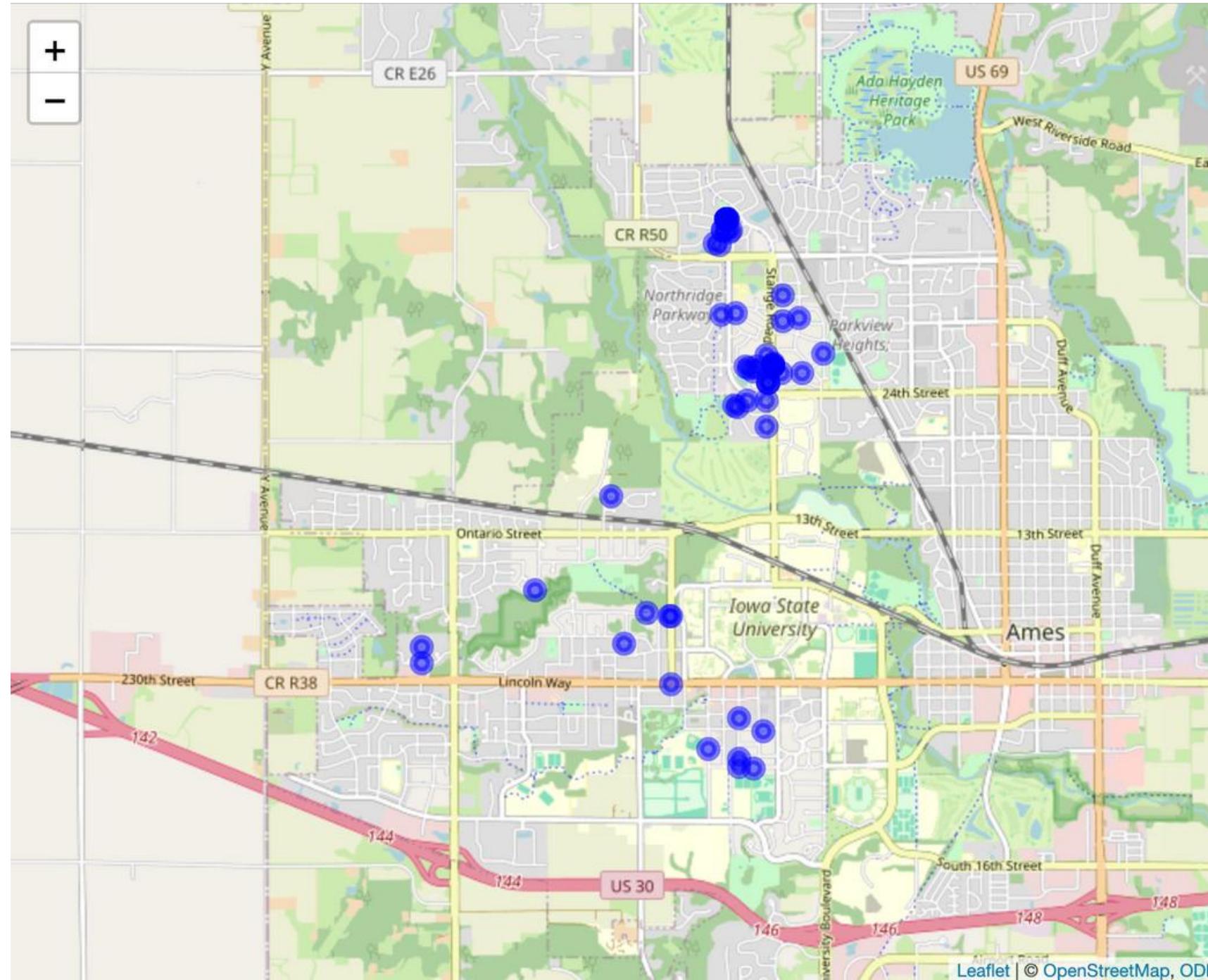
**Cardinality of Train vs. Test columns**

Cardinality of Categorical Columns: Train vs. Test



# Location

- Top 5 Neighborhoods: Veenker, Clear Creek, Somerset, Northridge Heights and S-W of Iowa State University



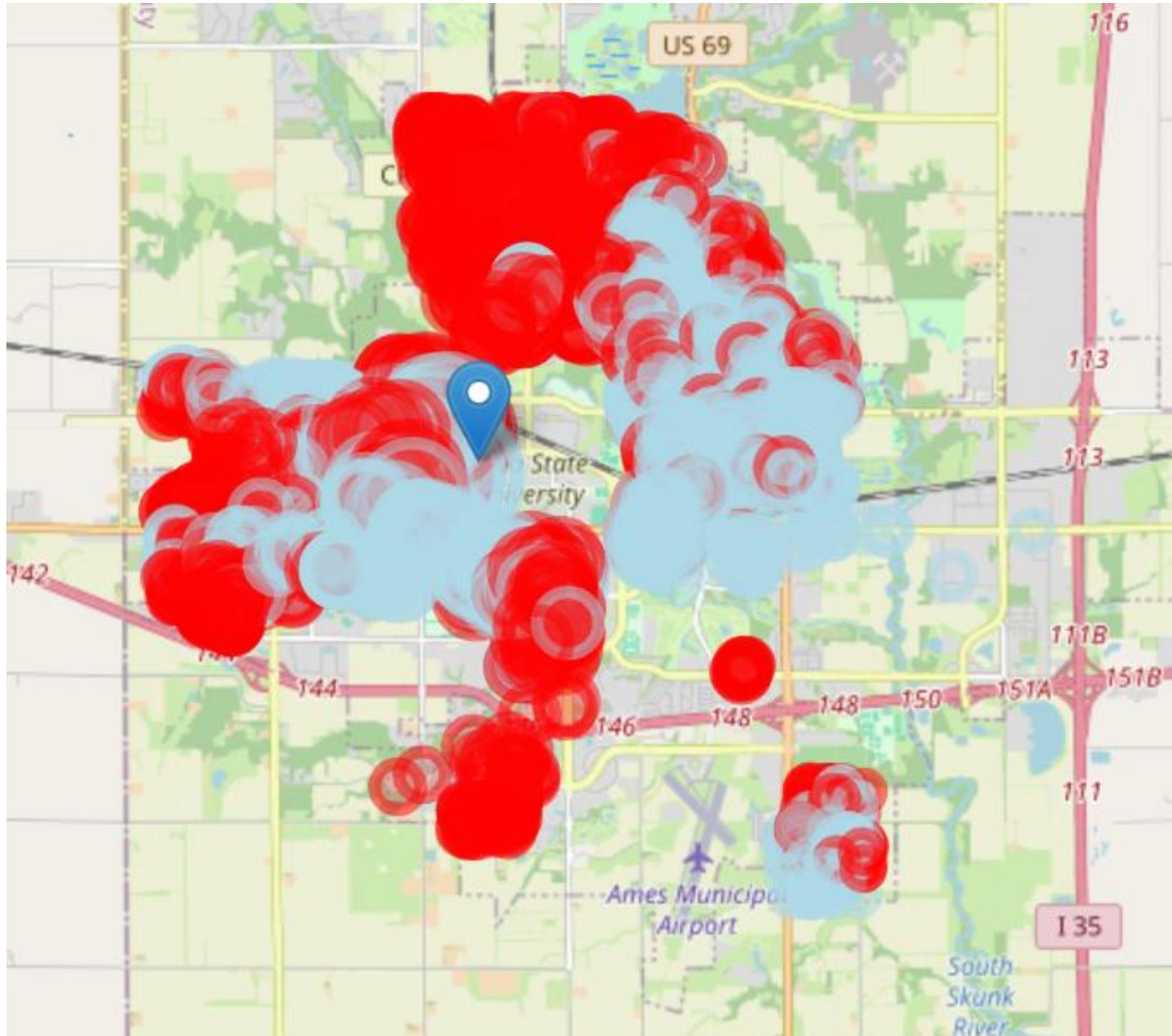
All the expensive neighborhoods are grouped around two points, either to the north side or towards the Iowa State University.

Somerset and Northridge Heights are present in the northern area.

Iowa South-West of State University and the Clear Creek Neighborhoods are present in the southwest

# Location

- Iowa State University: latitude and longitude as 42.02931225906169, and -93.65156268149819
- Airport



● Red is applied to properties with a sale price greater than the median sale price

● Blue is applied to properties with a sale price less than or equal to the median sale price

Properties near the northeast of Iowa State University and Highway 30 (red circles) tend to be more expensive, likely due to proximity to major transportation routes and the university.

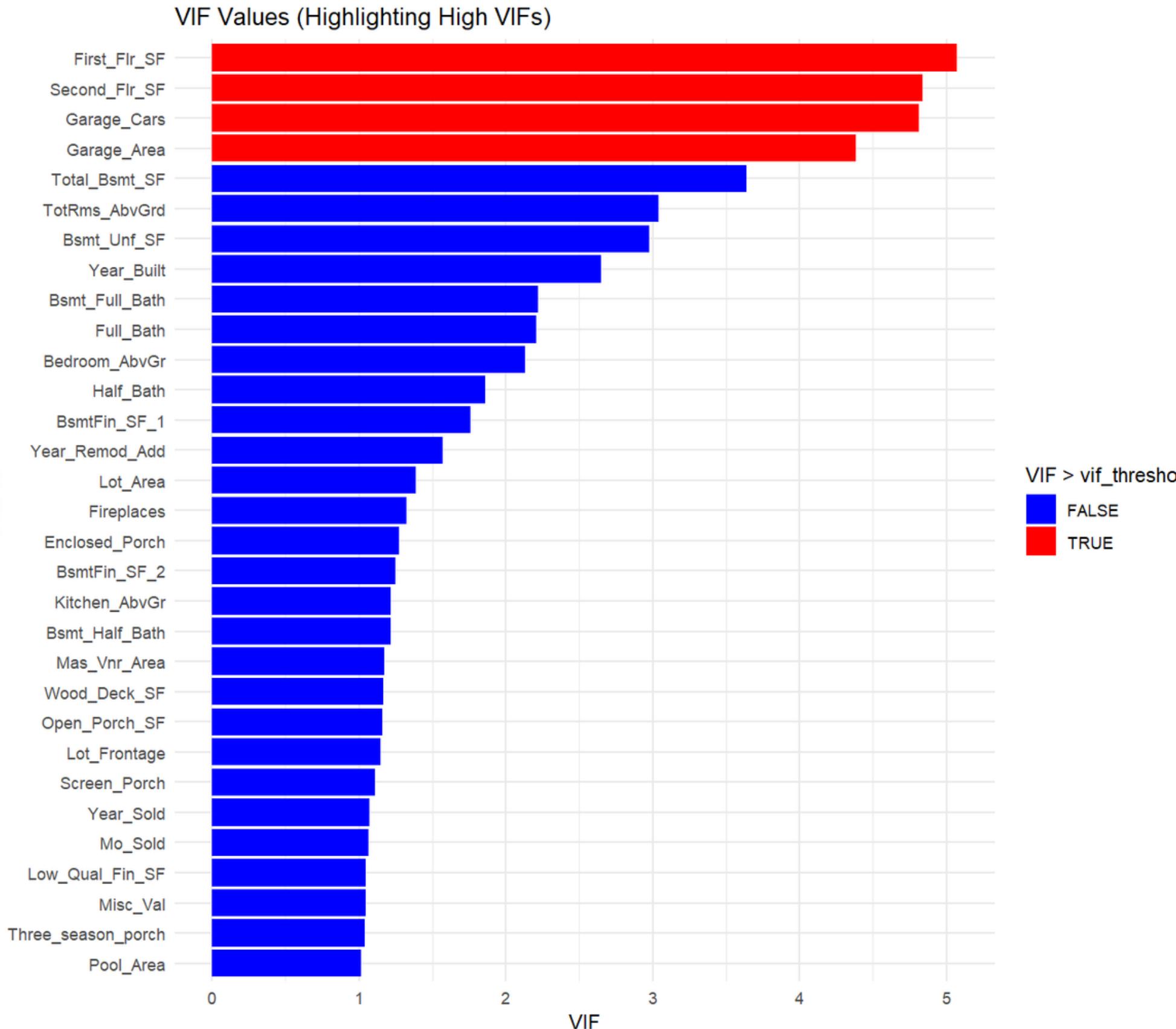
Properties farther from the university, closer to Ames Airport (blue circles), tend to have lower sale prices, possibly due to noise pollution from airplanes.

# Predictive Analysis



# Multi-Linear

Select number of features that are not highly correlated with each other.



# Multi-Linear Regression

```
Call:
lm(formula = Sale_Price ~ ., data = train.df_reduced)
```

Residuals:

Min	1Q	Median	3Q	Max
-60765	-8822	426	8363	95200

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.019e+06	8.558e+05	-1.191	0.233962
<b>Lot_Frontage</b>	4.255e+01	1.916e+01	2.221	0.026654 *
<b>Lot_Area</b>	6.300e-01	1.659e-01	3.799	0.000157 ***
<b>Year_Built</b>	3.808e+02	3.287e+01	11.587	< 2e-16 ***
<b>Year_Remod_Add</b>	3.142e+02	3.296e+01	9.533	< 2e-16 ***
Mas_Vnr_Area	-7.468e+00	5.145e+00	-1.451	0.147069
BsmtFin_SF_1	-2.713e+02	3.176e+02	-0.854	0.393152
BsmtFin_SF_2	-5.191e+00	3.514e+00	-1.477	0.140037
<b>Bsmt_Unf_SF</b>	-9.657e+00	2.648e+00	-3.647	0.000283 ***
<b>Total_Bsmt_SF</b>	2.681e+01	3.247e+00	8.259	6.25e-16 ***
Second_Flr_SF	-3.545e+00	3.771e+00	-0.940	0.347530
Low_Qual_Fin_SF	-1.821e+01	2.126e+01	-0.856	0.392001
<b>Gr_Liv_Area</b>	5.122e+01	4.703e+00	10.891	< 2e-16 ***
Bsmt_Full_Bath	2.063e+03	1.626e+03	1.269	0.204923
<b>Bsmt_Half_Bath</b>	4.594e+03	2.172e+03	2.115	0.034737 *
<b>Full_Bath</b>	6.017e+03	1.759e+03	3.420	0.000659 ***
Half_Bath	-1.946e+03	1.769e+03	-1.100	0.271717
<b>Bedroom_AbvGr</b>	-4.178e+03	1.181e+03	-3.538	0.000426 ***
Kitchen_AbvGr	-1.644e+04	5.206e+03	-3.157	0.001653 **
TotRms_AbvGrd	1.876e+03	9.338e+02	2.009	0.044893 *
<b>Fireplaces</b>	6.440e+03	1.102e+03	5.843	7.50e-09 ***
<b>Garage_Area</b>	2.607e+01	3.660e+00	7.121	2.42e-12 ***
<b>Wood_Deck_SF</b>	1.745e+01	5.309e+00	3.287	0.001058 **
<b>Open_Porch_SF</b>	3.474e+01	1.098e+01	3.164	0.001615 **
<b>Enclosed_Porch</b>	2.171e+01	1.062e+01	2.044	0.041315 *
Three_season_porch	9.103e+00	1.764e+01	0.516	0.605983
<b>Screen_Porch</b>	3.770e+01	1.148e+01	3.284	0.001071 **
Pool_Area	3.245e+01	2.720e+01	1.193	0.233260
Misc_Val	1.264e+00	7.235e-01	1.747	0.080963 .
Mo_Sold	2.476e+02	2.243e+02	1.104	0.269996
Year_Sold	-1.514e+02	4.271e+02	-0.355	0.722988

Residual standard error: 15550 on 784 degrees of freedom  
Multiple R-squared: 0.811, Adjusted R-squared: 0.8038  
F-statistic: 112.1 on 30 and 784 DF, p-value: < 2.2e-16

```
> cat("Training Set Accuracy (Reduced Model):\n")
Training Set Accuracy (Reduced Model):
> print(train_accuracy_mlr_reduced)
      ME    RMSE     MAE     MPE     MAPE
Test set -4.608072e-10 15251.93 11351.67 -1.319113 8.765986
> cat("Validation Set Accuracy (Reduced Model):\n")
Validation Set Accuracy (Reduced Model):
> print(valid_accuracy_mlr_reduced)
      ME    RMSE     MAE     MPE     MAPE
Test set 1114.693 17644.9 12702.29 -1.308013 9.843143
```

The highlighted variables are the ones that are highly correlated with the target variable (**Sale\_Price**).

```
Call:
lm(formula = Sale_Price ~ ., data = train.df_reduced)

Residuals:
    Min      1Q  Median      3Q      Max 
-62355  -8882     119   8461  101020 

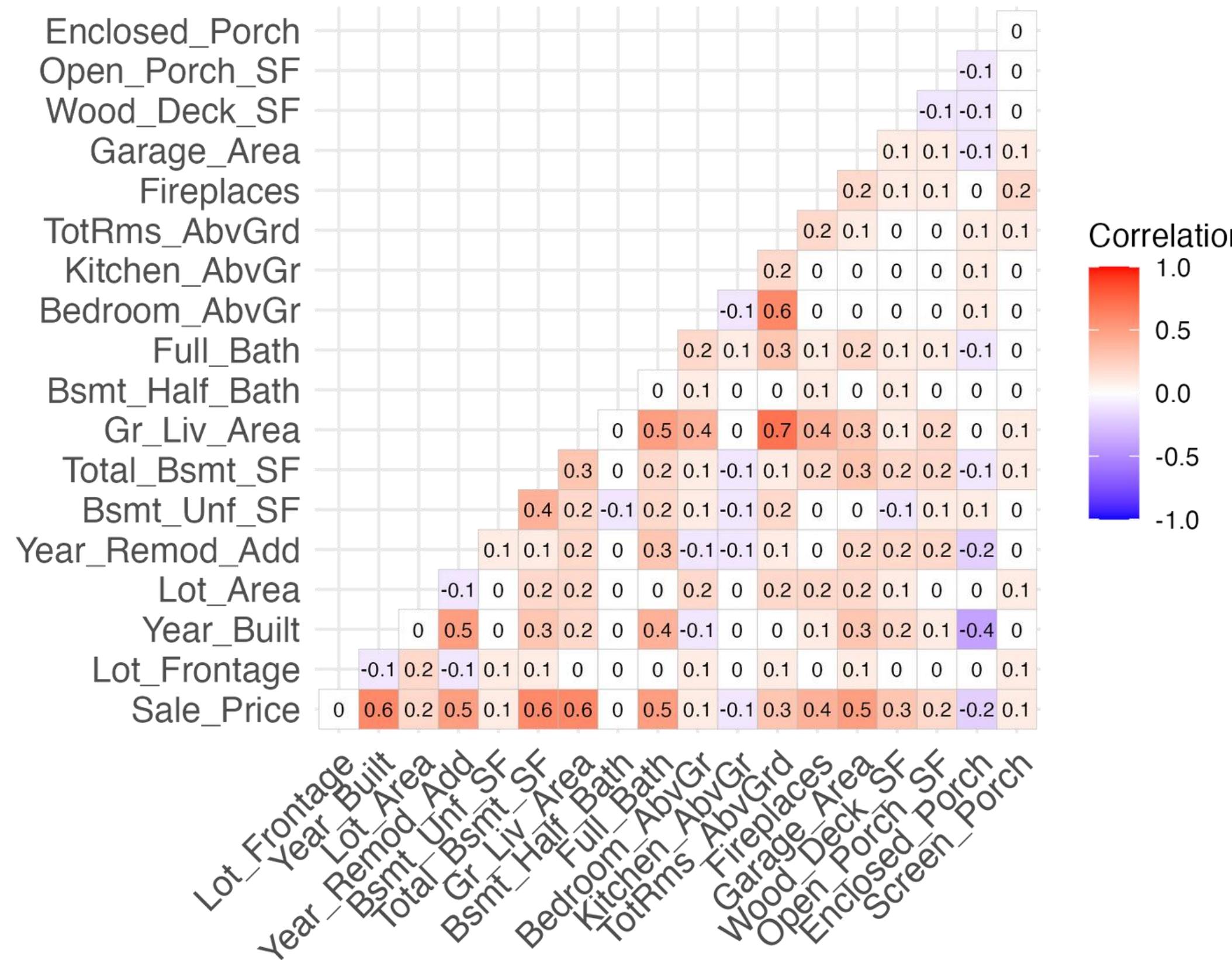
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.321e+06  7.072e+04 -18.678 < 2e-16 ***
Lot_Frontage  4.430e+01  1.910e+01   2.319 0.020663 *  
Lot_Area      7.505e-01  1.586e-01   4.733 2.62e-06 ***
Year_Built    3.750e+02  2.952e+01   12.704 < 2e-16 ***
Year_Remod_Add 3.184e+02  3.243e+01   9.817 < 2e-16 ***
Bsmt_Unf_SF  -1.119e+01  1.852e+00  -6.040 2.37e-09 *** 
Total_Bsmt_SF 3.099e+01  2.245e+00   13.805 < 2e-16 ***
Gr_Liv_Area   4.621e+01  4.032e+00   11.461 < 2e-16 ***
Full_Bath      6.167e+03  1.613e+03   3.823 0.000142 *** 
Bsmt_Half_Bath 3.987e+03  2.020e+03   1.974 0.048698 *  
Bedroom_AbvGr -4.443e+03  1.158e+03  -3.838 0.000134 *** 
Kitchen_AbvGr -1.438e+04  5.152e+03  -2.790 0.005391 ** 
TotRms_AbvGrd  1.861e+03  9.284e+02   2.005 0.045339 *  
Fireplaces     6.466e+03  1.094e+03   5.913 4.98e-09 *** 
Garage_Area    2.672e+01  3.652e+00   7.318 6.17e-13 *** 
Wood_Deck_SF   1.795e+01  5.280e+00   3.399 0.000710 *** 
Open_Porch_SF  3.565e+01  1.089e+01   3.273 0.001108 ** 
Enclosed_Porch 2.210e+01  1.061e+01   2.082 0.037660 *  
Screen_Porch   3.660e+01  1.148e+01   3.190 0.001480 ** 

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 15620 on 796 degrees of freedom  
Multiple R-squared: 0.8063, Adjusted R-squared: 0.8019  
F-statistic: 184.1 on 18 and 796 DF, p-value: < 2.2e-16

# Correlation of Predictor Variables

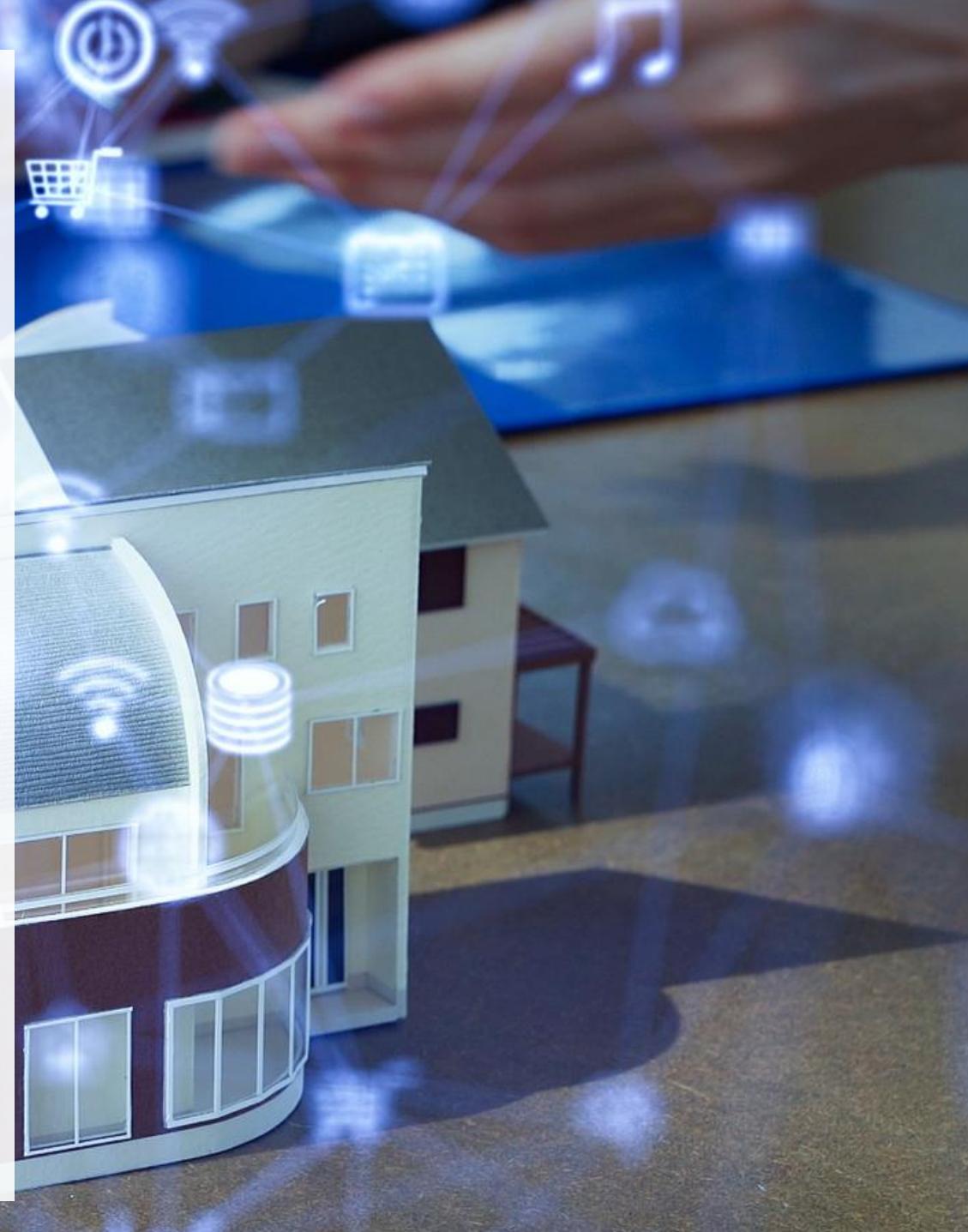
Correlation Plot of Predictor Variables



Year Built, total basement square footage, and general living area are important features contributing to the sale price

# clustering

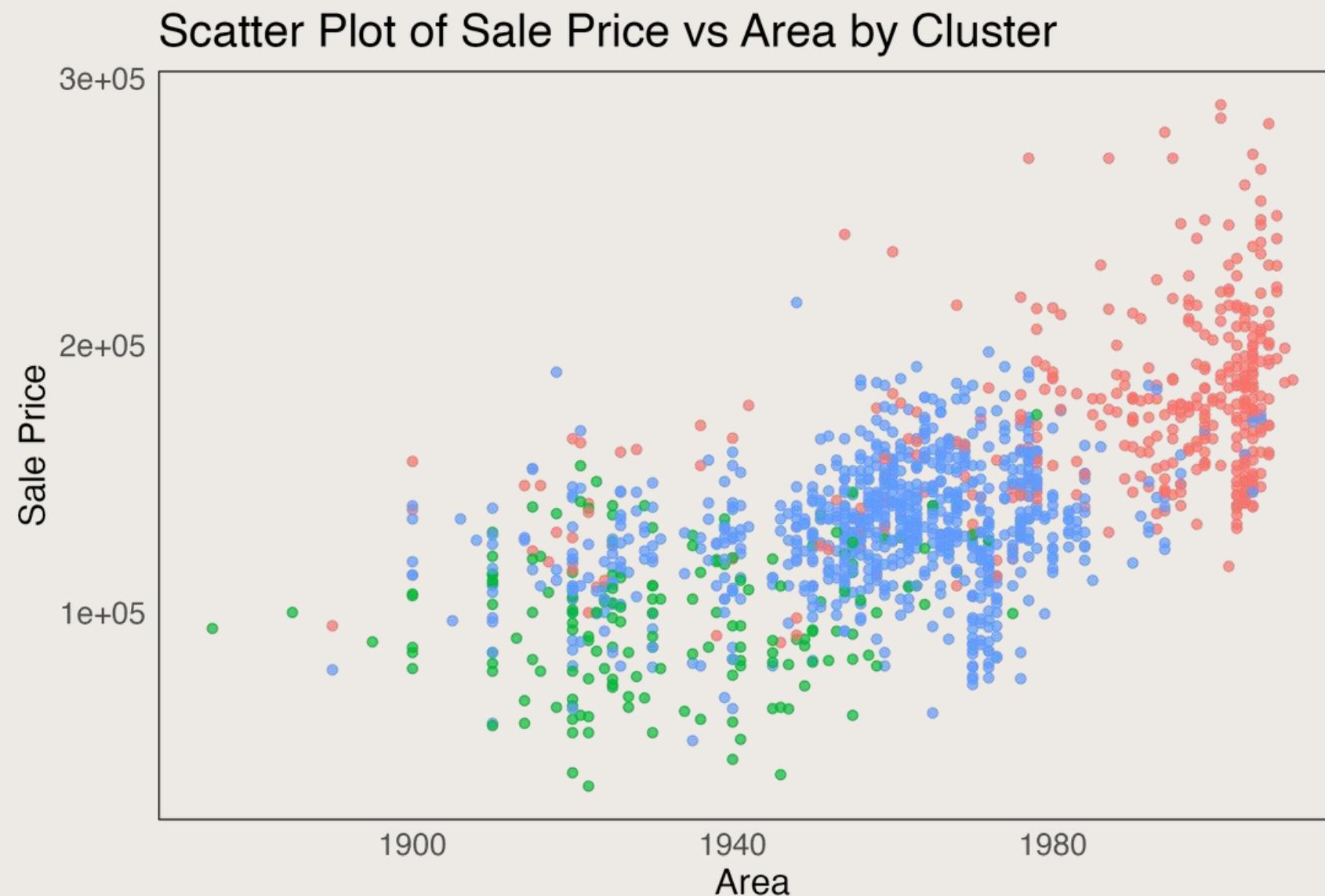
Ordinal Categorical Variables &  
Numerical Variables



# Findings

## Ordinal Categorical Variables

The model's price is influenced by Garage Quality and Overall Quality; as these values rise, the total price increases. Additionally, having a fence is a requirement, and Central Air conditioning is always included. However, factors such as Basement Quality, Heating Quality, and Kitchen Quality are not of concern.

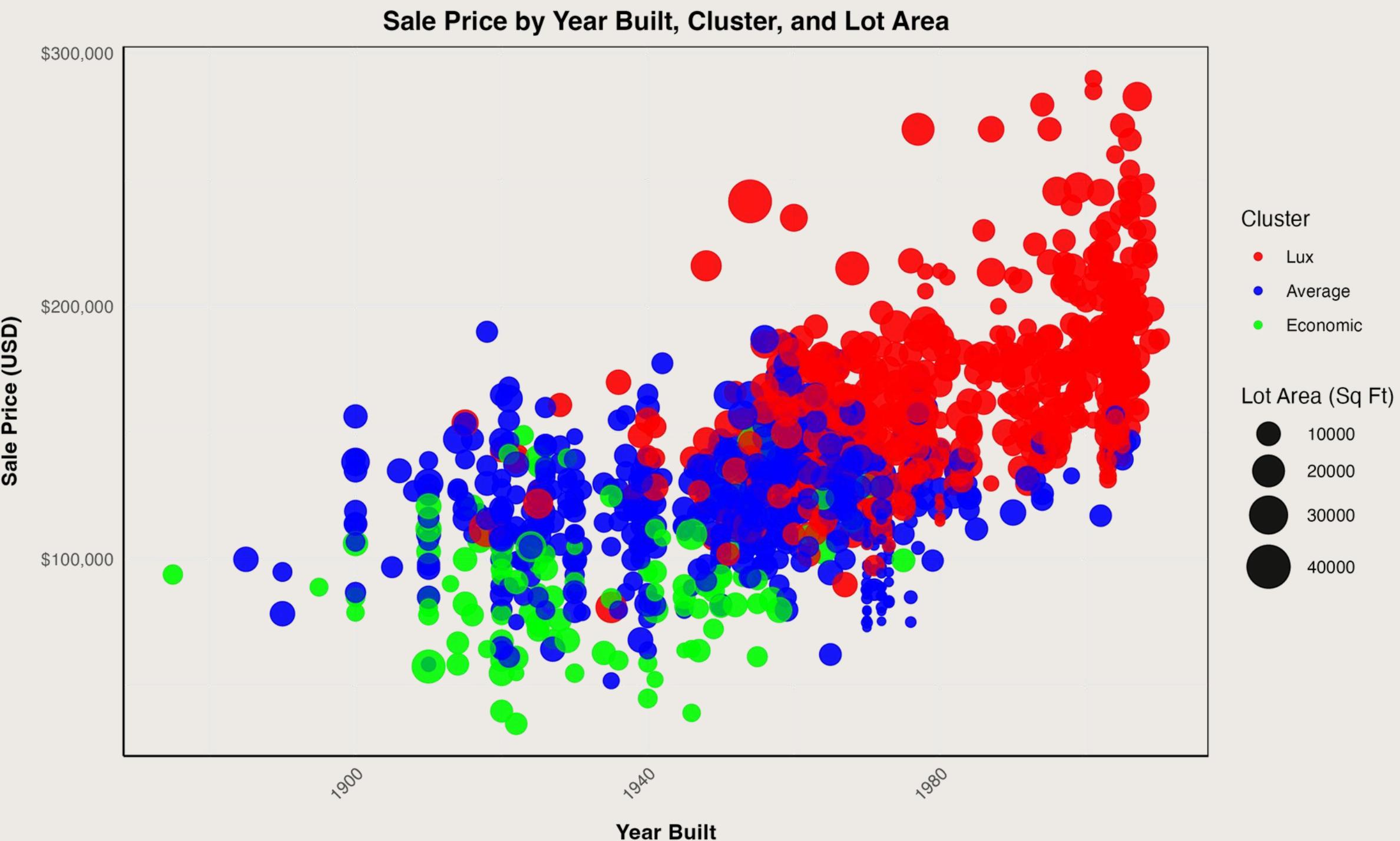


Cluster	Overall_Qual	Bsmt_Qual	Heating_QC	Central_Air	Kitchen_Qual	Garage_Qual	Fence
Lux	6.53	2.99	1.67	1	3.48	5.94	4.86
Average	5.2	5.41	3.53	1	4.7	5.88	4.1
Economic	4.43	4.99	3.35	0.34	4.24	3.74	4.48

# Findings

## Numerical Variables

- The Lux cluster represents more modern homes with higher prices, likely due to the incorporation of newer features like central air conditioning.
- The Average cluster falls between Lux and Economic in terms of price and construction dates.
- The Economic cluster includes homes with the lowest prices, and smaller lot areas.
  - World War II and great depression (1939–1945).



# Key Insights

## **Location:**

Focus on developing or marketing properties in higher-value locations like properties near the northeast of Iowa State University and Highway 30.

## **Certain Home Features that Affect Sale Prices Are:**

Highlighting key features like basements, living areas, bathrooms, bedrooms, kitchens, lot size, year built, fireplaces, garages, and wood decks can positively influence the sale price. Emphasizing these features in listings can attract higher bids.

## **Clusterings:**

The cluster can give us insight into the type of listings to be shown thinking on different customers needs. Tailoring sales pitches and pricing strategies can then be implemented according to clustering the insights; such as for high-end buyers, target the Lux cluster, which has properties with the highest quality and garage space.

A photograph of a modern two-story house with light-colored walls and dark wood trim around the windows and doors. Large glass windows provide a view of a well-maintained garden with various plants and a clear blue sky.

Thank You For Your  
Attention!

# Resources:



History of air conditioning: A brief HVAC history - when & how {timeline}. Trade School Programs in Chicago. (2022a, May 23).  
<https://www.coynecollege.edu/a-brief-history-of-hvac-air-conditioning/#:~:text=What's%20the%20story%20of%20the,built%20with%20central%20air%20conditioning>

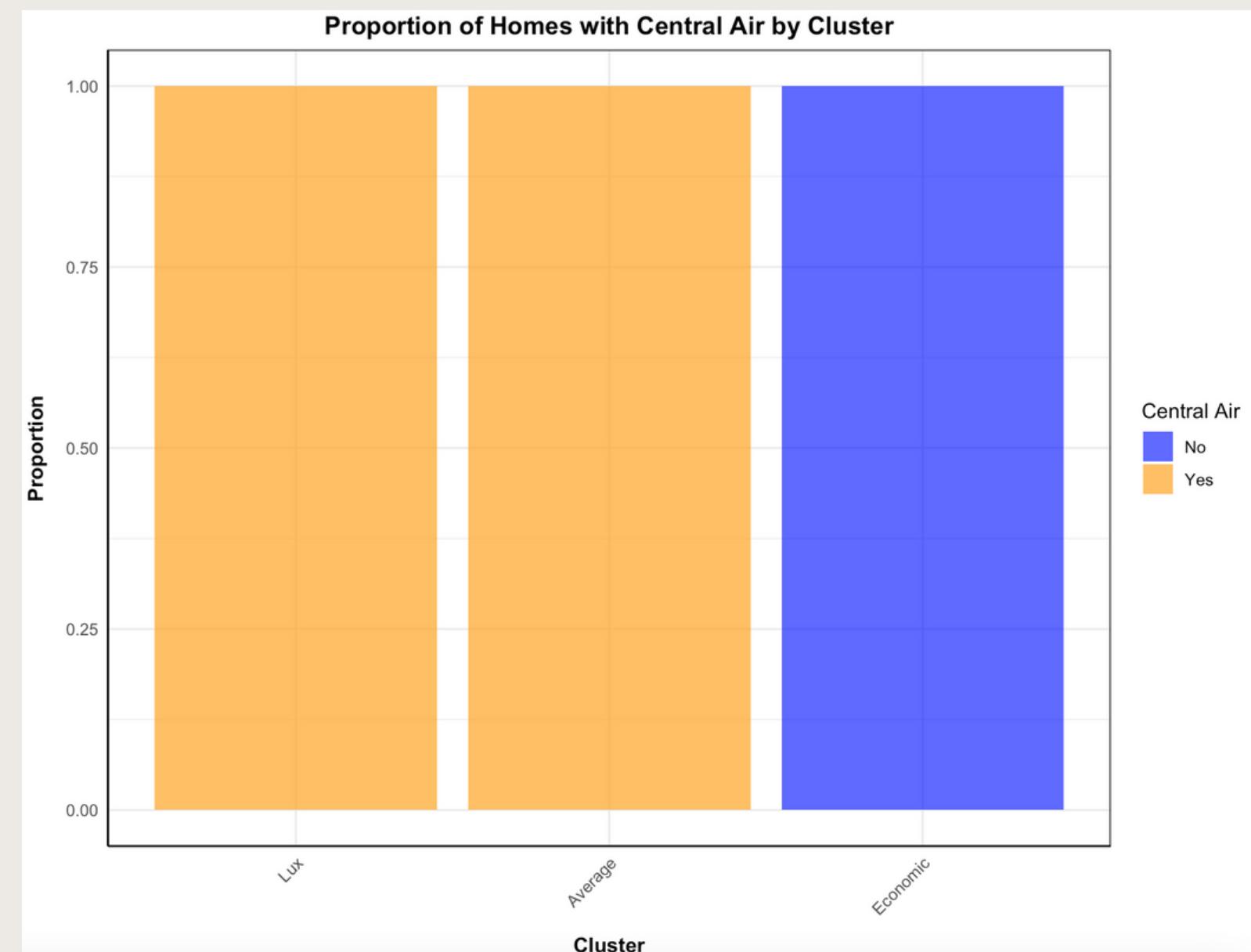
Wolfeco. (2023, February 14). A look back at the housing market: From the 1920s to Today. Winter Homes.  
<https://www.builtbywinterhomes.com/news/a-look-back-at-the-housing-market/#:~:text=The%201940s:%20A%20Decade%20of,of%20the%20previous%20two%20decades.>

World War II. Iowa PBS. (n.d.). <https://www.iowapbs.org/iowapathways/mypath/2555/world-war-ii#:~:text=The%20second%20world%20war%20in,in%20a%20variety%20of%20ways>

# Appendix

## Clustering with Numerical Variables

```
selected_features <- data_1 %>%  
  dplyr::select(Lot_Area, Year_Built, Central_Air) %>%  
  mutate(Central_Air = ifelse(Central_Air == "Y", 1, 0))
```



# RFM: What are the most important ones?

```
> train_1 <- train %>% select(-c("Total_Bsmt_SF", "First_Flr_SF", "Garage_Cars", "BsmtFin_Type
_1", "BsmtFin_SF_1", "Low_Qual_Fin_SF",
+                               "Latitude", "Longitude" ))
>
> rf_model <- randomForest(Sale_Price ~ ., data=train_1, importance=TRUE, ntree=100)
>
> # Get feature importance
> importance <- as.data.frame(importance(rf_model))
> importance <- importance[order(importance`%IncMSE`, importance$IncNodePurity, decreasing =
TRUE), ]
>
> # Display top features
> print(importance)
```

	%IncMSE	IncNodePurity
Gr_Liv_Area	16.5267457	104138908170
Neighborhood	13.0720099	163588333642
Overall_Qual	11.2457016	101511392704
Year_Built	9.9960278	153493064487
Year_Remod_Add	9.6039134	17994688370
MS_SubClass	9.2829020	62273248457
Garage_Area	8.2511914	68955404204
Garage_Type	6.1226894	21682173558
Lot_Area	6.1181062	16772381673
Full_Bath	6.0122747	20629157885
Second_Flr_SF	5.7353172	5931486840
...	...	...

Year_Built	Total_Sales	YoY_Increase	Color
1872	122000	NA	No Change
1875	94000	-28000	Decrease
1879	185000	91000	Increase
1880	933479	748479	Increase
1882	168000	-765479	Decrease
1885	222500	54500	Increase
1890	1019900	797400	Increase
1892	582500	-437400	Decrease
1893	325000	-257500	Decrease
1895	356600	31600	Increase
1896	50138	-306462	Decrease
1898	106000	55862	Increase
1900	3419026	3313026	Increase
1901	230000	-3189026	Decrease
1902	97500	-132500	Decrease
1904	157500	60000	Increase
1905	389300	231800	Increase
1906	135000	-254300	Decrease
1907	93369	-41631	Decrease
1908	367000	273631	Increase
1910	4805729	4438729	Increase
1911	130000	-4675729	Decrease
1912	603300	473300	Increase

Sale_Type	Description
WD	Warranty Deed - Conventional
CWD	Warranty Deed - Cash
VWD	Warranty Deed - VA Loan
New	Home just constructed and sold
COD	Court Officer Deed/Estate
Con	Contract 15% Down payment regular terms
ConLw	Contract Low Down payment and low interest
ConLI	Contract Low Interest
ConLD	Contract Low Down
Oth	Other