

---

# PREDICTING AMES HOUSING SALES PRICE

*Rhoeun Park*

*dsi-flex-222*



---

# Problem Statement

**iBuyer companies** rely their profit on **selling** real estate properties **at higher prices** than the prices that the **properties were bought**. In such cases, **accurate predictions** of housing property market prices are key in securing profit from the real estate transactions. This project seeks to explore the various features of housing properties and **find out which features can most accurately predict the housing sales prices and make the predictions, specifically for Ames, IA.**

---

# Background

iBuyer companies and  
Automated valuation  
models (AVMs)

Ex) Zillow Offers  
Program

**Zillow, unable to predict housing prices, to 'wind down' Offers program everywhere**

ZILLOW

IBUYERS

TRIANGLE HOUSING MARKET

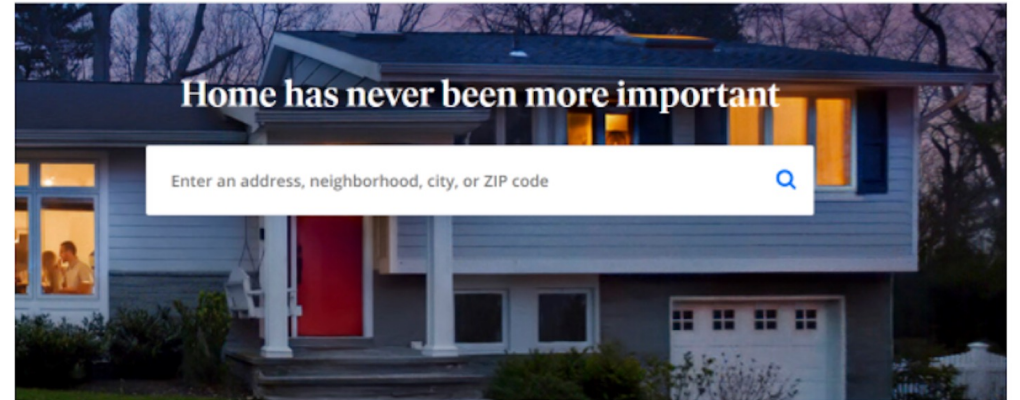
Home Loans Agent finder

 **Zillow**

Manage Rentals Advertise Sign in or Join

**Home has never been more important**

Enter an address, neighborhood, city, or ZIP code



---

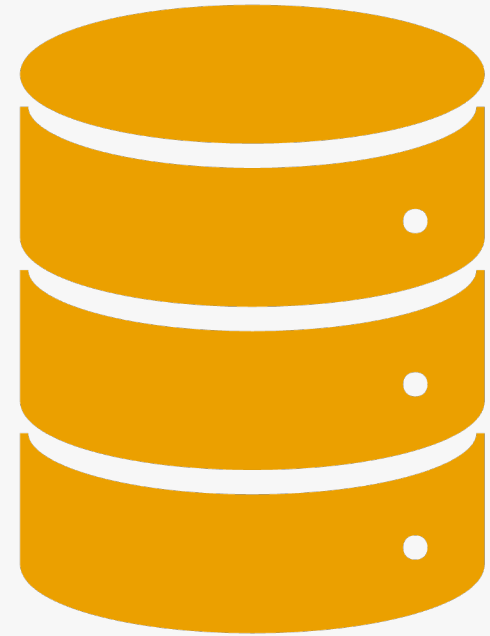
# Datasets

---

## Ames Housing Data

---

- source: Ames Assessor's Office
- 
- Assessed values and sale price in Ames, IA from 2006 to 2010.



---

# Limitations

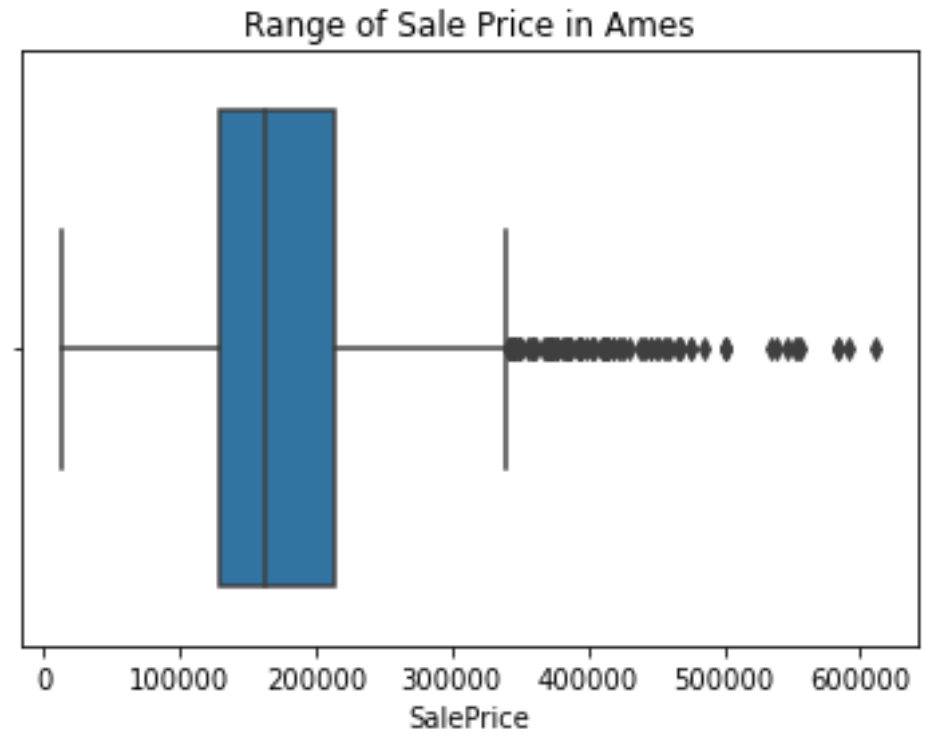
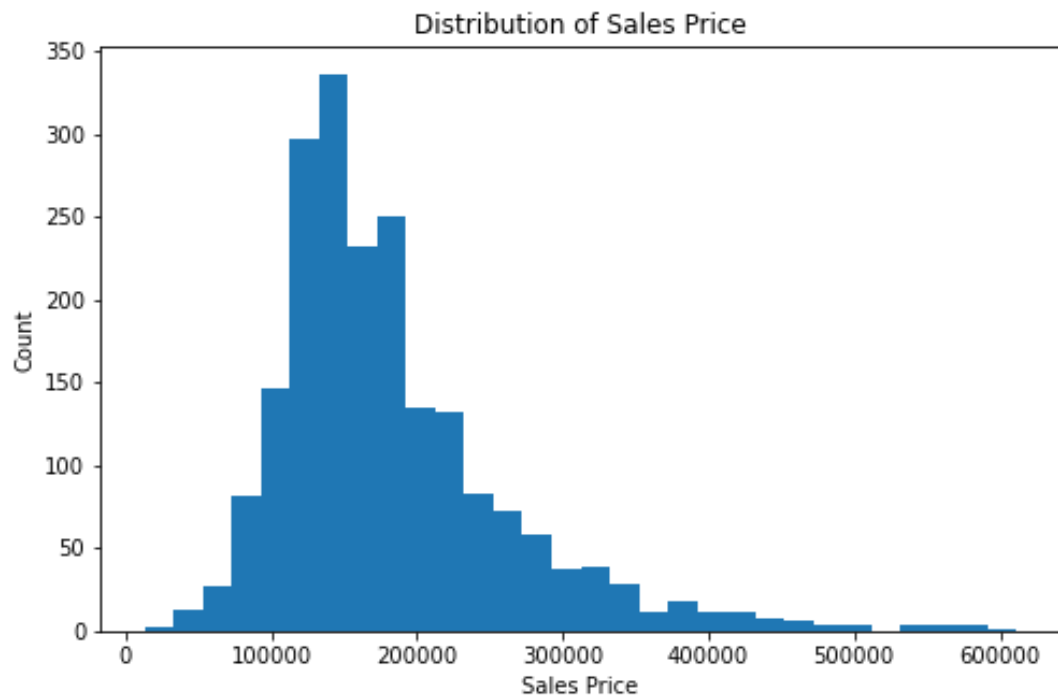
DATA LIMITED TO ONE  
CITY; AMES, IA

OUTDATED

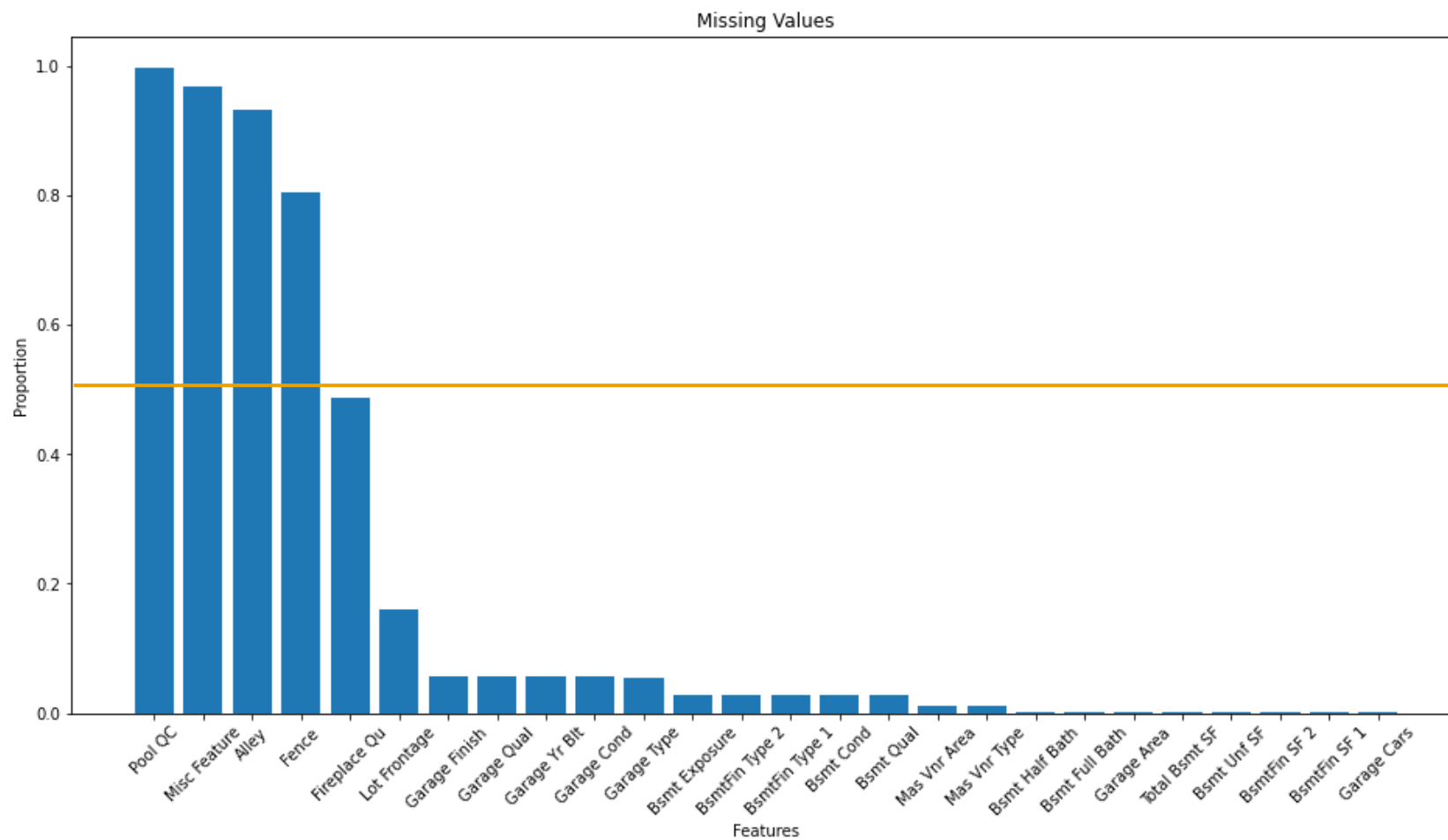
PREDICTION ONLY BASED  
ON INTRINSIC FACTORS

---

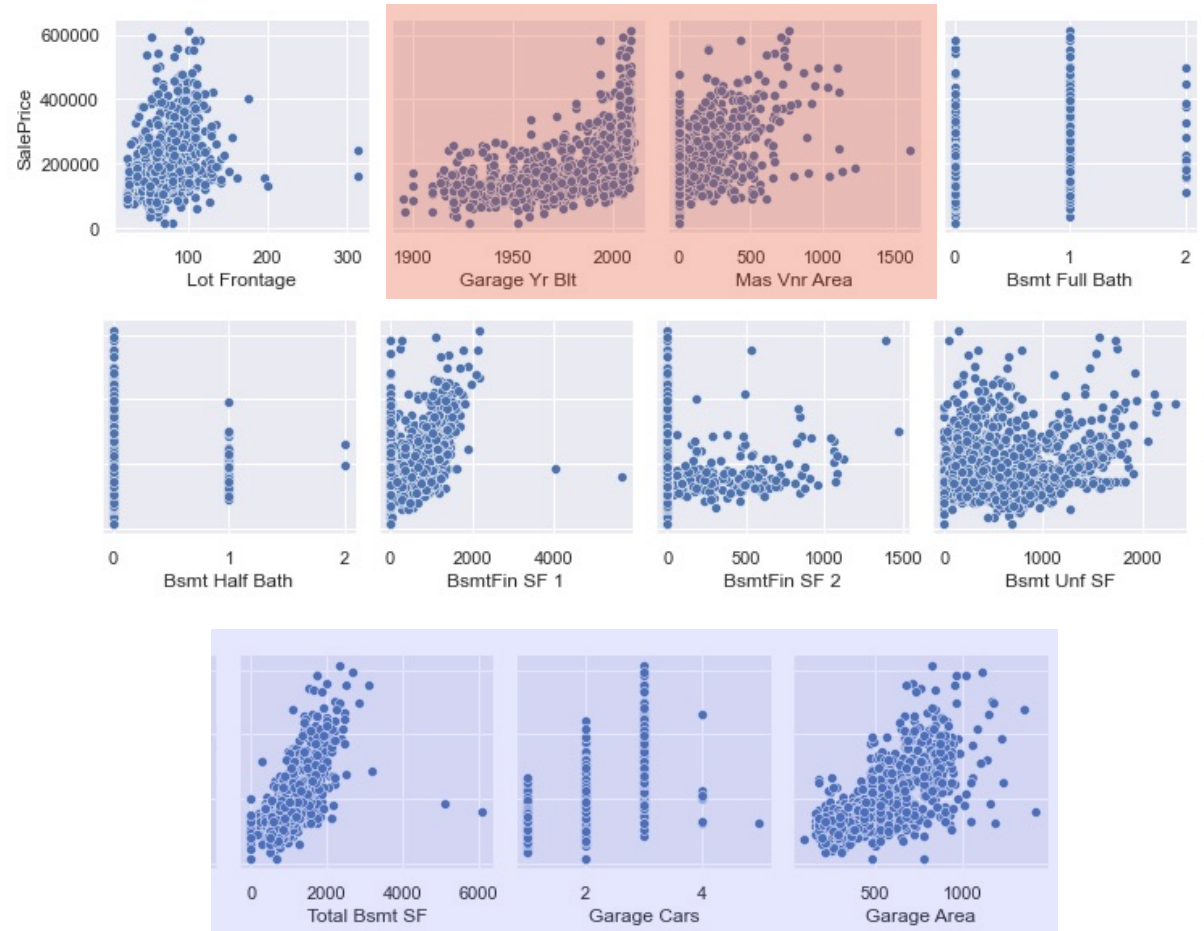
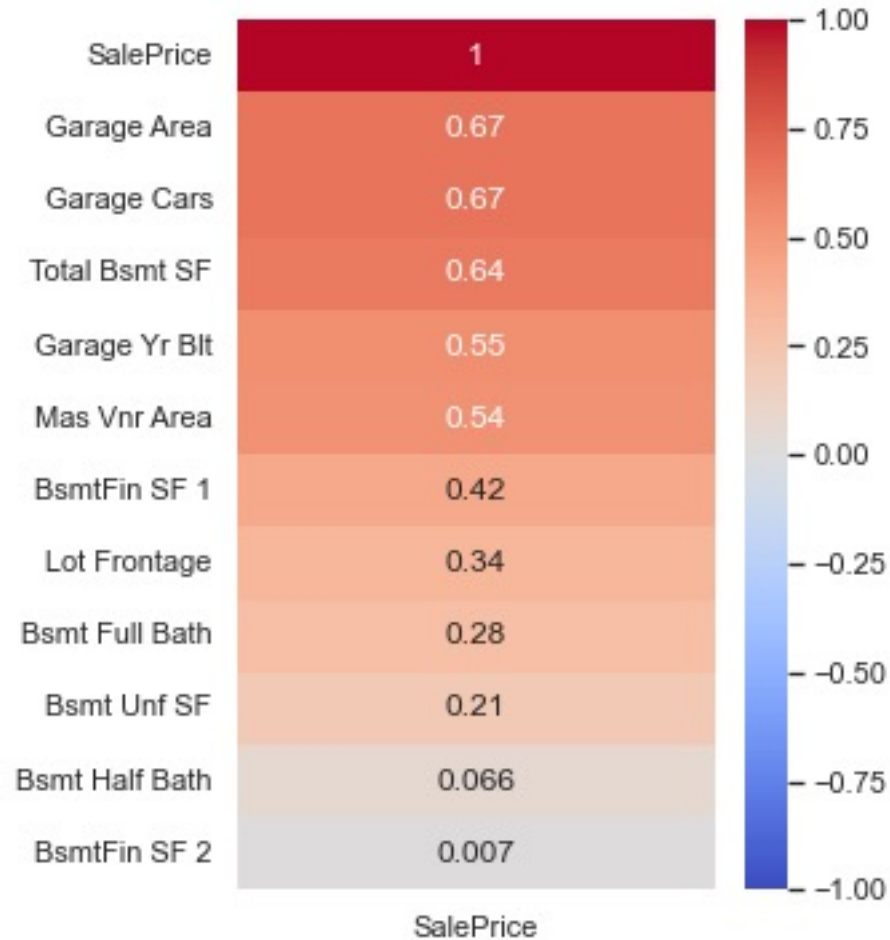
# Exploratory Data Analysis



# Null Values



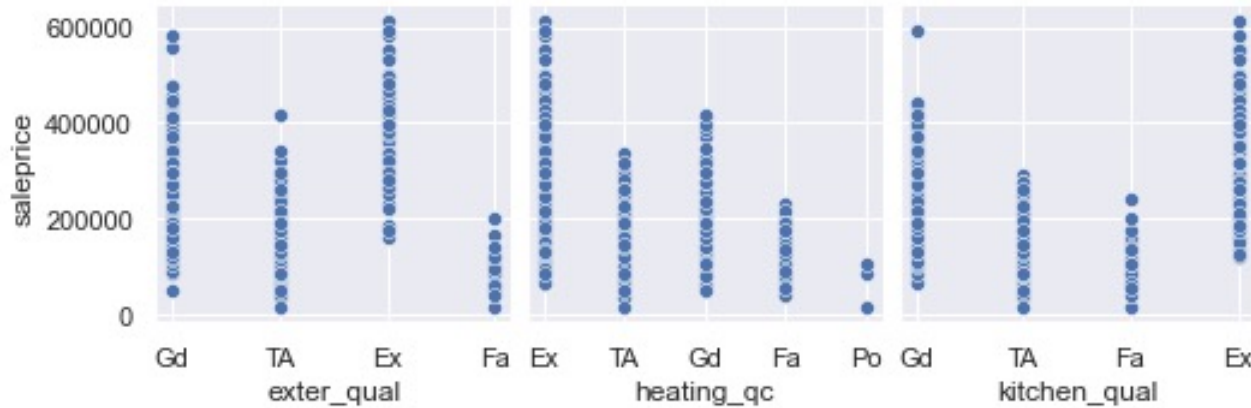
# Null Values: Numerical



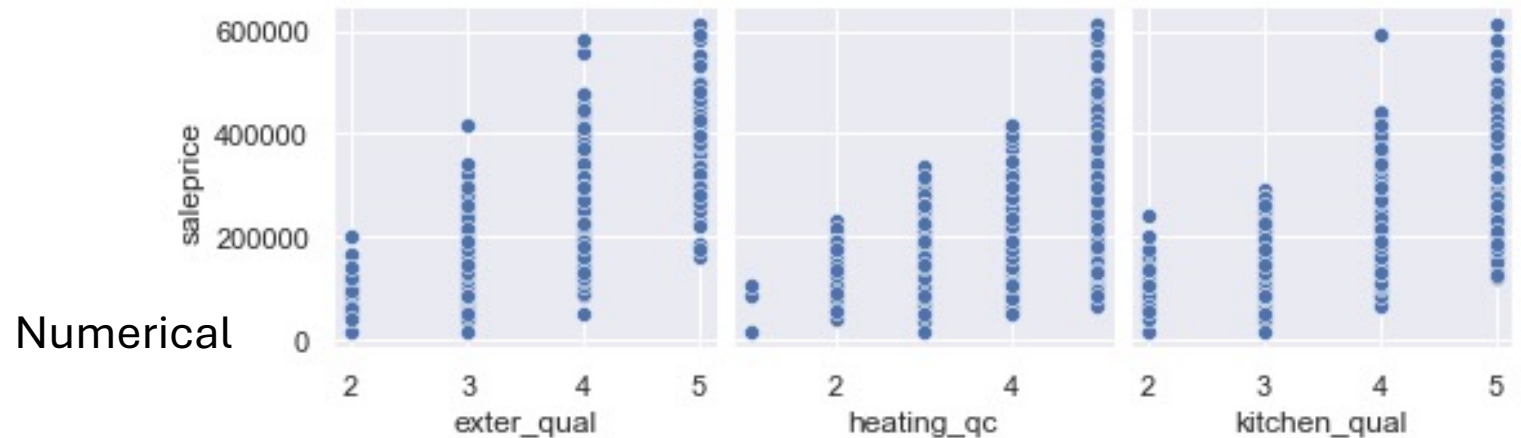
Keep



# Feature Engineering – Ranking Ordinals



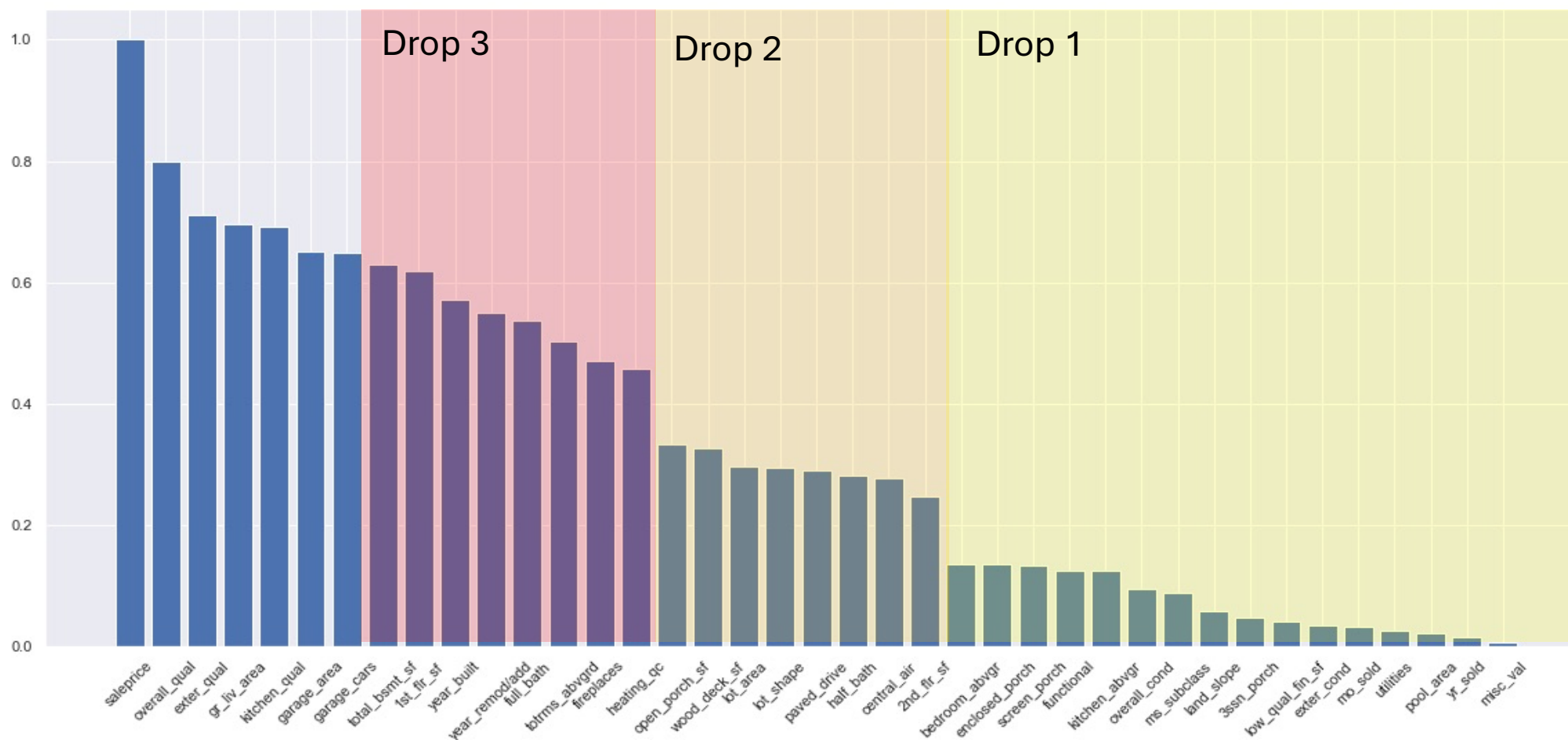
Categorical



Numerical

# Feature Selection – Numerical

Correlation Coefficients of Features vs Sale price



## CV Score

0.8087



-0.0015

0.8072



-0.0011

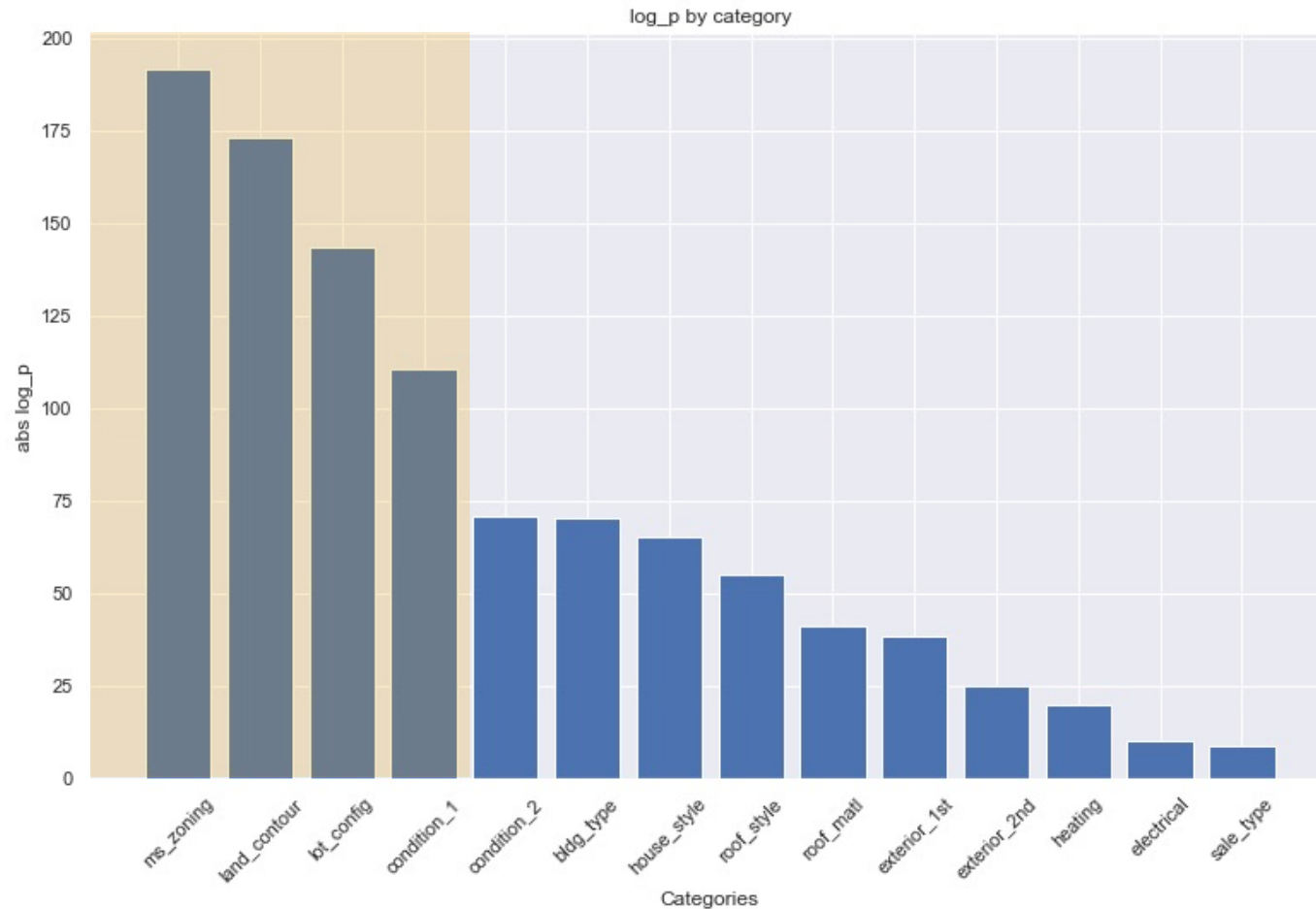
0.8061



-0.0076

0.7985

# Feature Selection - Categorical

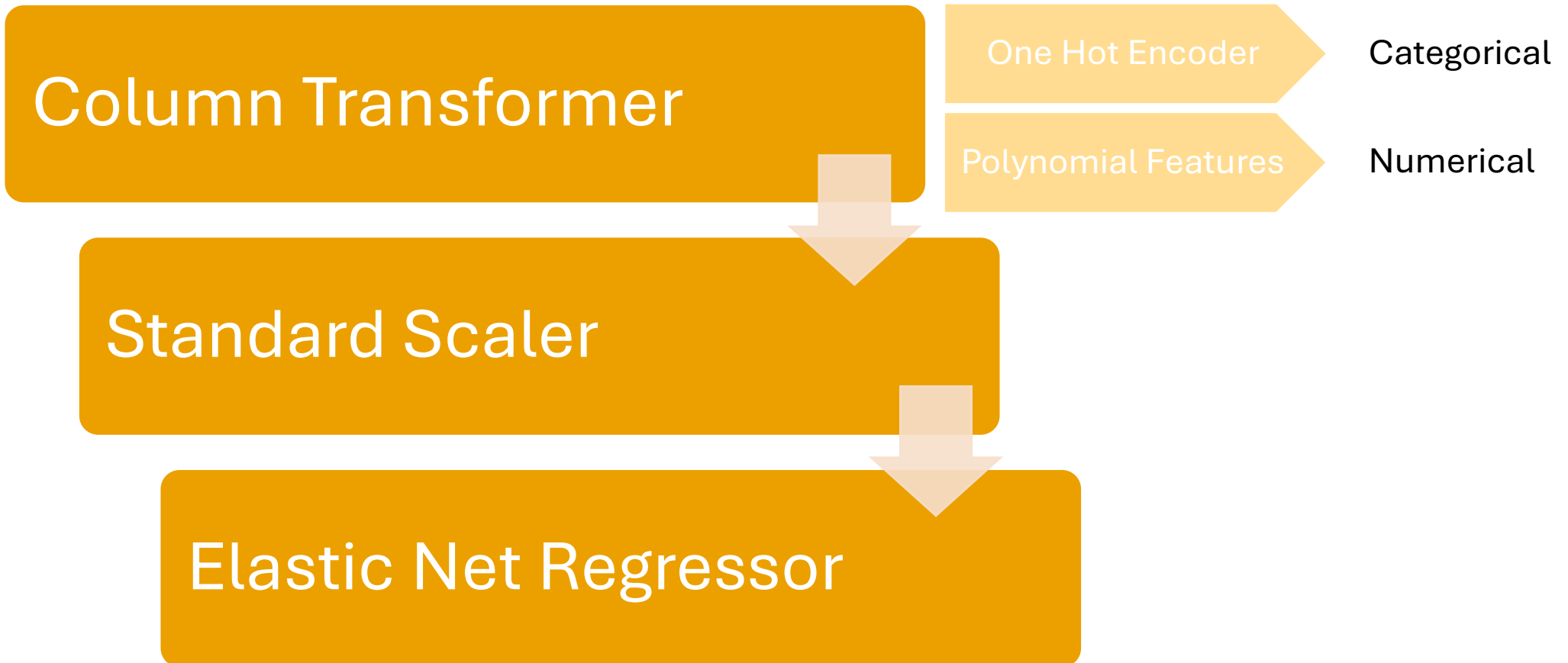


## ANOVA Stats

- Are the mean 'sale price' different for each group in the categorical feature?
- p-value

---

# Pipeline



---

# Analysis

	RMSE	R-squared
Train Set	22796	0.92
Validation Set	24950	0.90

- 21 total features
- 14 numeric, 7 categorical



---

# Feature Groups



**Location:** zoning classification, neighborhood



**Quality:** overall quality, exterior quality, heating quality, kitchen quality, fireplace quality,



**Age:** year built, year remodeled



**Material:** exteriors, foundation



**Number of features:** bathrooms, total rooms, fireplaces



**Usable space:** basement area, 1<sup>st</sup> floor area, living area, garage cars, garage area

---

# Conclusion



90% of Instances fit  
the model



Reduce Variance



Consider  
External factors