

The background of the slide is a dense field of three-dimensional, light blue numbers (0-9) of various sizes and orientations, creating a sense of depth and data. A solid black rectangular box is positioned on the right side of the slide, containing the title, author, and date.

House Sale Price Prediction

Olorunfemi Adeyemo

July 18, 2025

Problem Statement



This project aims to predict house sale prices using Machine learning. etc.

Why is this important

- ❖ Critical for Mortgage lenders and Banks
- ❖ Aids in Urban Planning & Development
- ❖ Guides seller in selling their homes.

Stakeholders

- ❖ Policy Markers
- ❖ Real estate agency & Banks
- ❖ Mitchel and Francis

Data Overview

- ◆ **Source: Ames Housing Dataset (Kaggle)**
Total Records: 2042 rows (train), 878 row(test)
- ◆ **Features: 80+ variables, including:**
- ◆ **Lot Area, Overall Quality, Year Built etc.**

23 Nominal

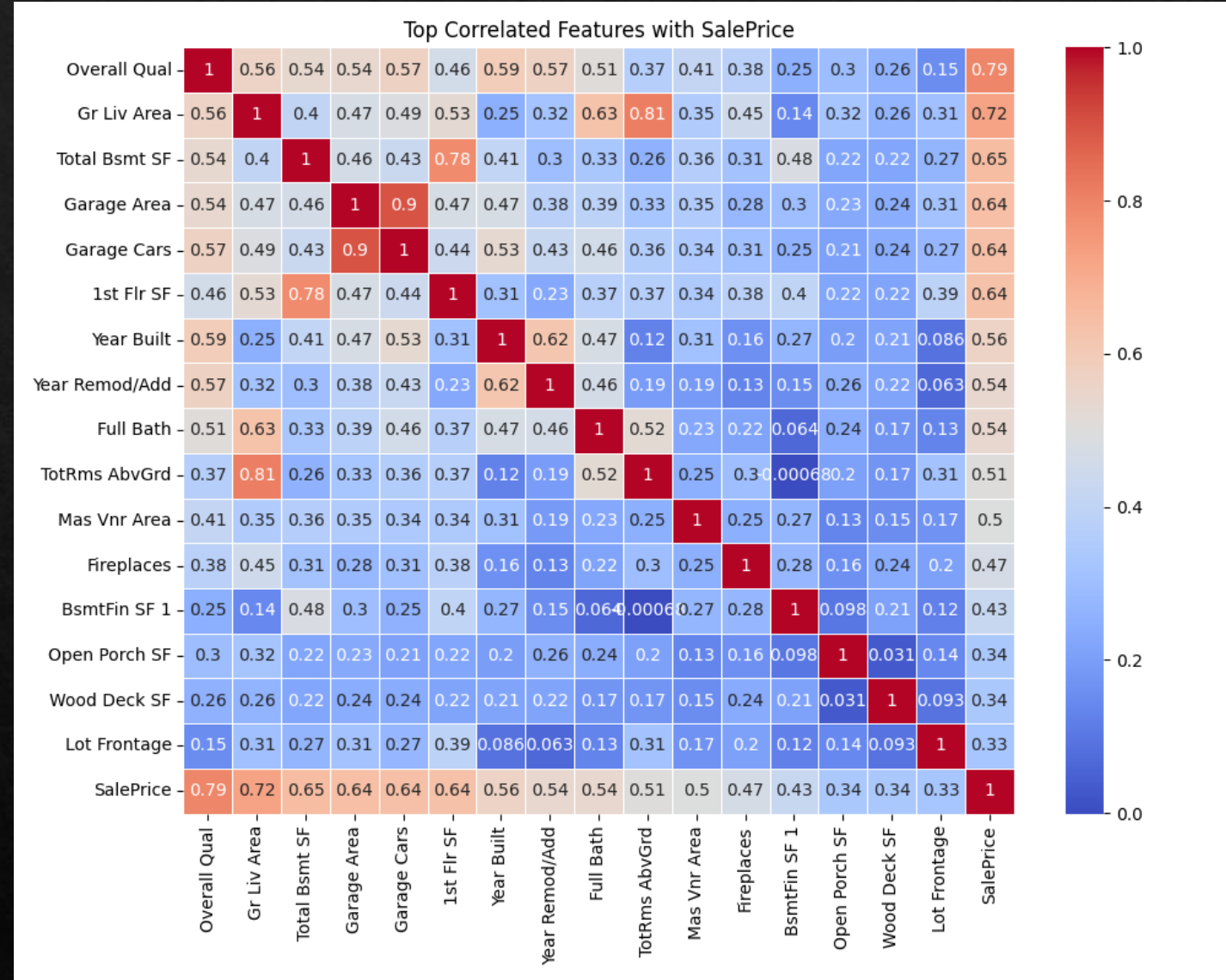
14 Discreet

20 Continuous



Data Cleaning & Preprocessing

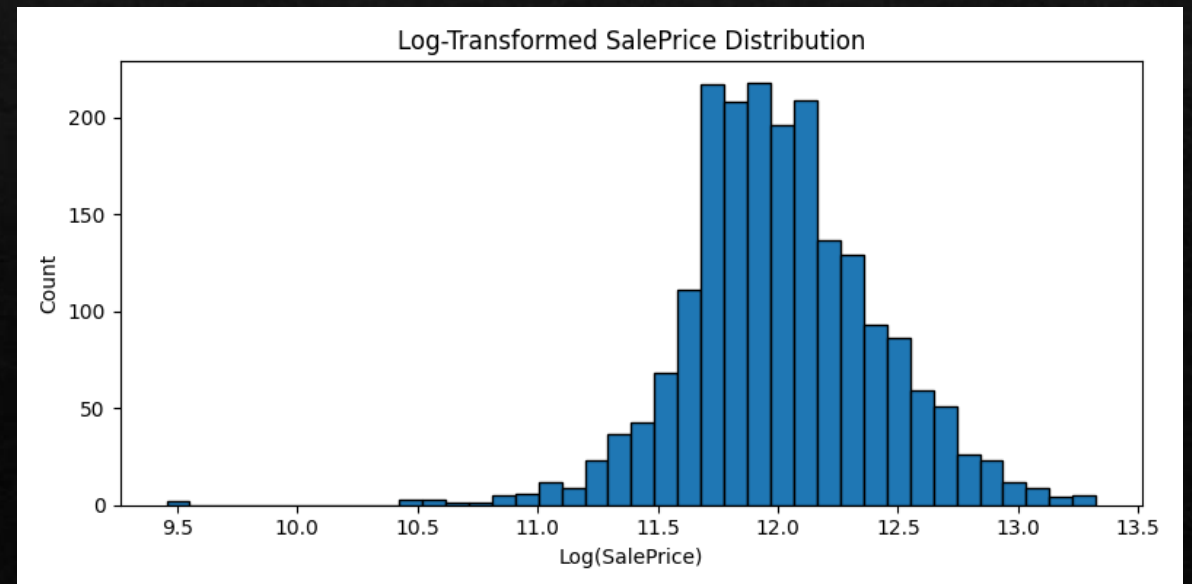
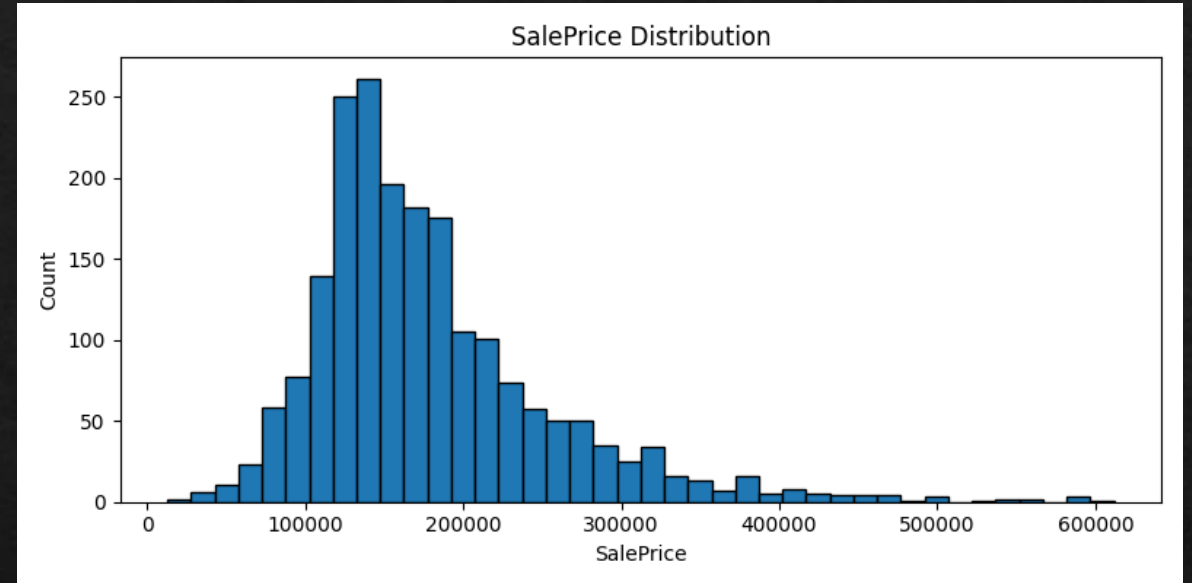
- Handled missing values (None, 0, mode, median)
- Combined rare labels (e.g., alley, pool QC)
- Converted categorical to numerical
- Created log of target variable
- Feature Scaling**



Feature Engineering

Data was grouped into

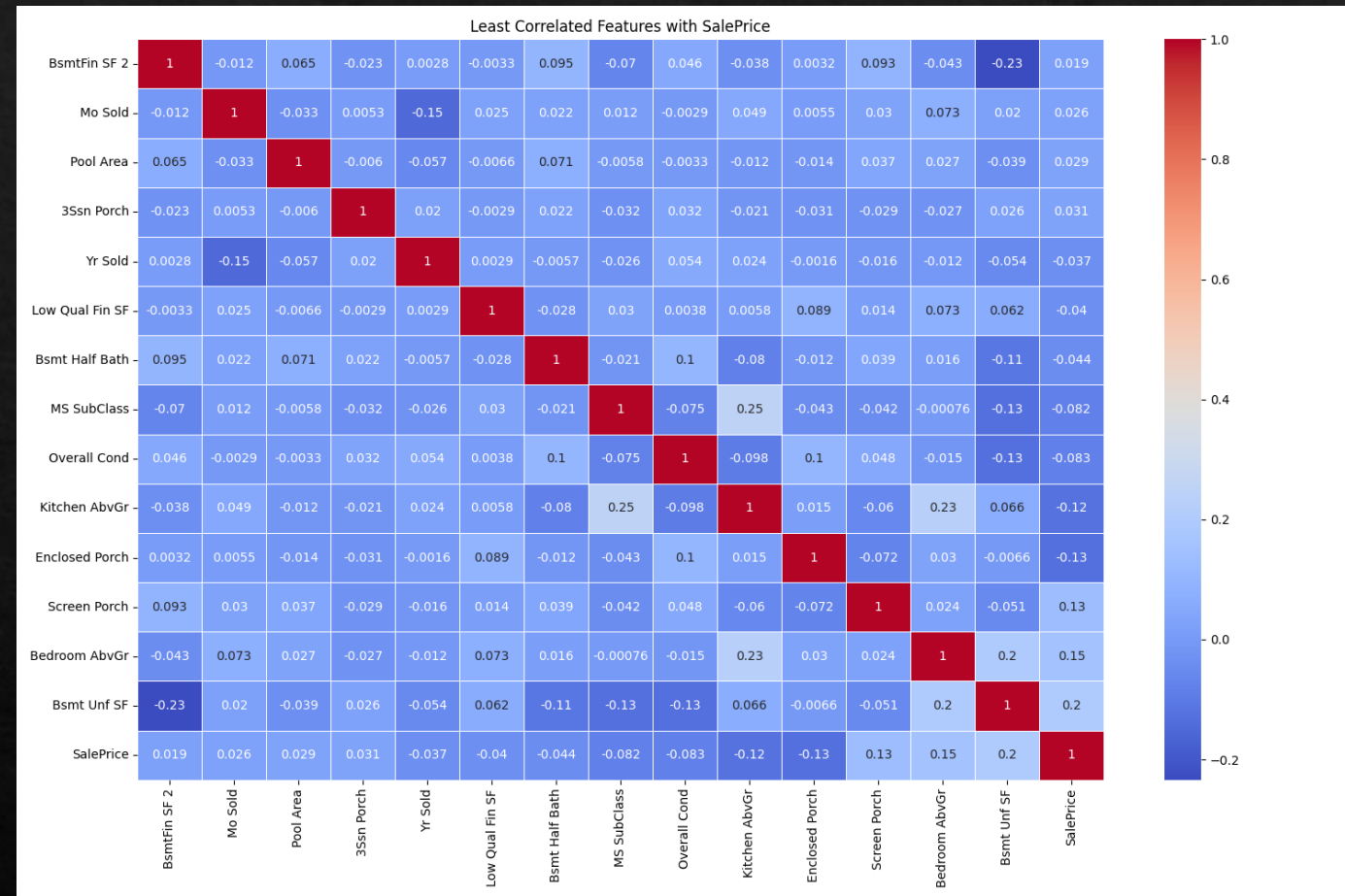
- ◆ **Skewed numerical features: log-transformed**
- ◆ **Ordinal categorical features: ordinal encoded**
- ◆ **Nominal categorical Features: One-hot encoded**



Feature selection

Data was grouped into

- ◆ **Poor correlated features e.g. Mo Sold were dropped.**
- ◆ **Certain outlier features like Pool Area were selected**
- ◆ **Research article and domain knowledge also played a role in the choice of features to retain.**



Model Selection

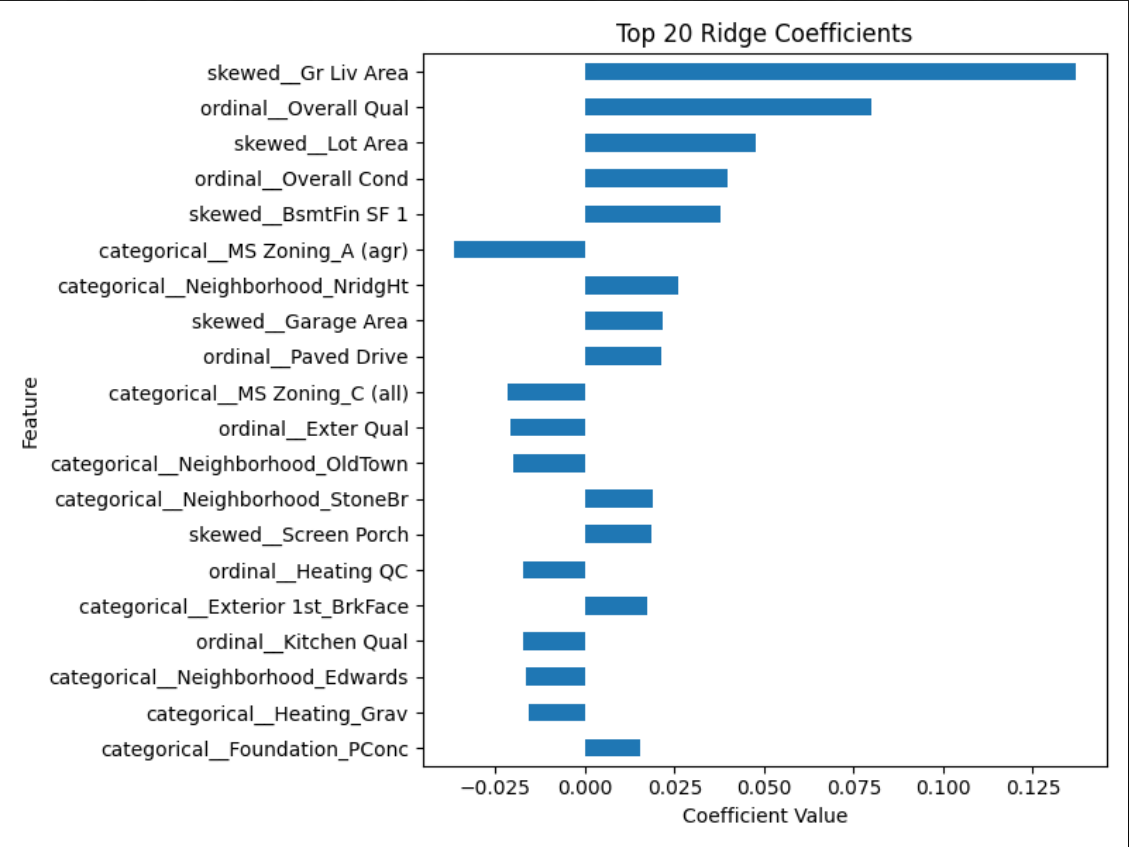
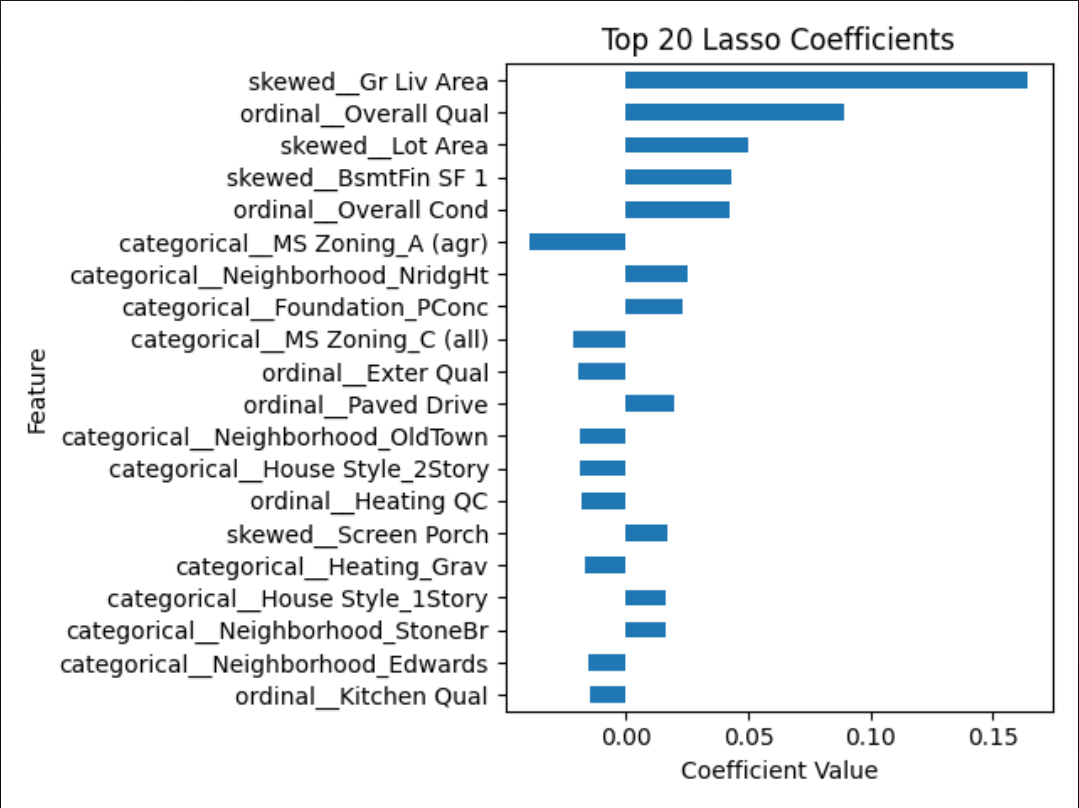
Tested and tuned:

- ◇ **Linear Regression**
- ◇ **Ridge Regression (L2)**
- ◇ **Lasso Regression (L1)**
- ◇ **Random Forest Regressor**

Model Performance Comparison

Model	RMSE	R ²
Linear Regression	0.141	0.873
Ridge Regression	0.140	0.874
Lasso Regression	0.141	0.872
Random Forest	0.135	0.884

Feature Importance



Recommendation

- ◆ **Enhance overall and Exterior quality: Overall Quall and Exter Qual has a coefficient of +0.07985**
- ◆ **Target High- Value Neighborhoods: Neighborhood like NrdgHt significantly boosted prices.**
- ◆ **Prioritize Larger Living Areas in Listings : Features like Gr Liv Area had the strongest positive relationship. Coefficient of +0.13698**
- ◆ **Beware of Detrimental Features: Ms Zonning_A(Agric) has a coefficient of -0.037.**

Thank you!