AMES HOUSING PRICE PREDICTION

Presented by,
-Abirami Rajamanickam

Project Overview

Objective:

- Predict the sale price of residential homes in Ames, Iowa using features like the number of rooms, neighborhood, and house condition.

- Goal: Build an accurate predictive model to estimate sale prices for real estate decisions.

Goal of the Model:

- Develop a model that generalizes well to unseen data and can provide accurate price predictions for homes of various price ranges.

Data Science Approach:

- **Preprocessing**: Handle missing values, scale numerical features, and encode categorical variables.

- **Modeling**: Apply **Ridge regression** to capture relationships between features and target price while preventing overfitting through regularization.

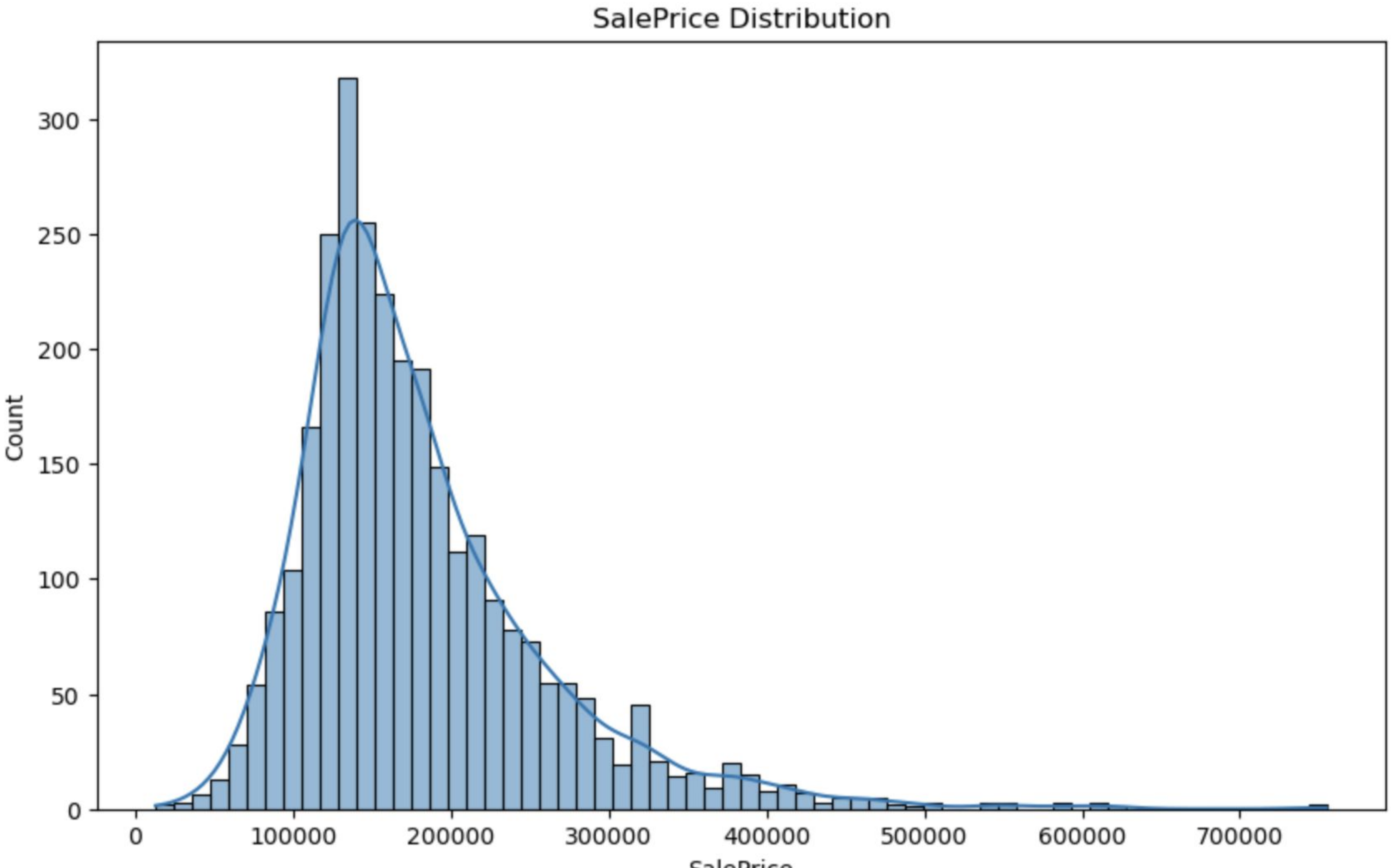
- **Evaluation**: Use metrics like **RMSE** and **R^2** to assess model performance and make improvements.

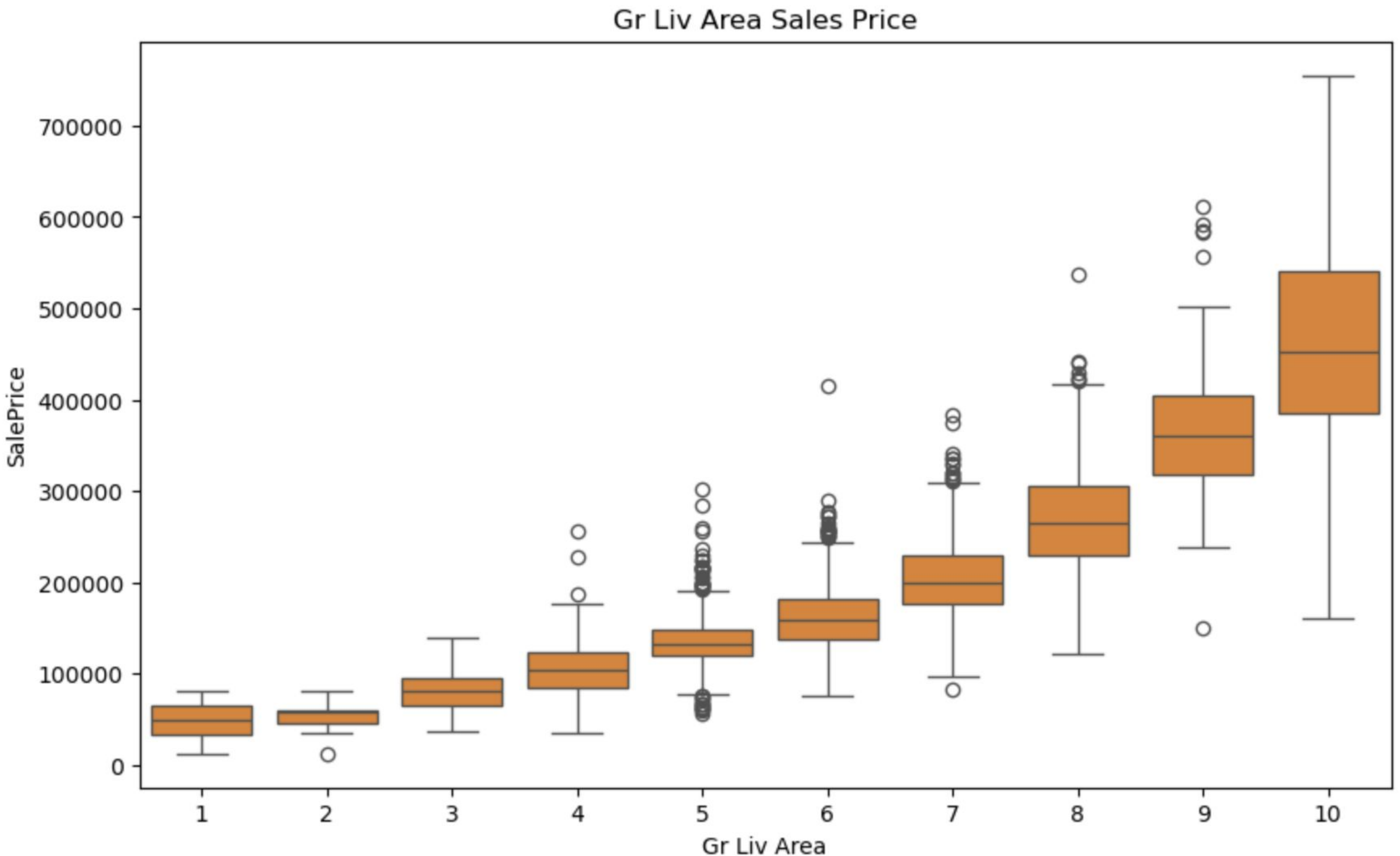
Exploratory Data Analysis (EDA)

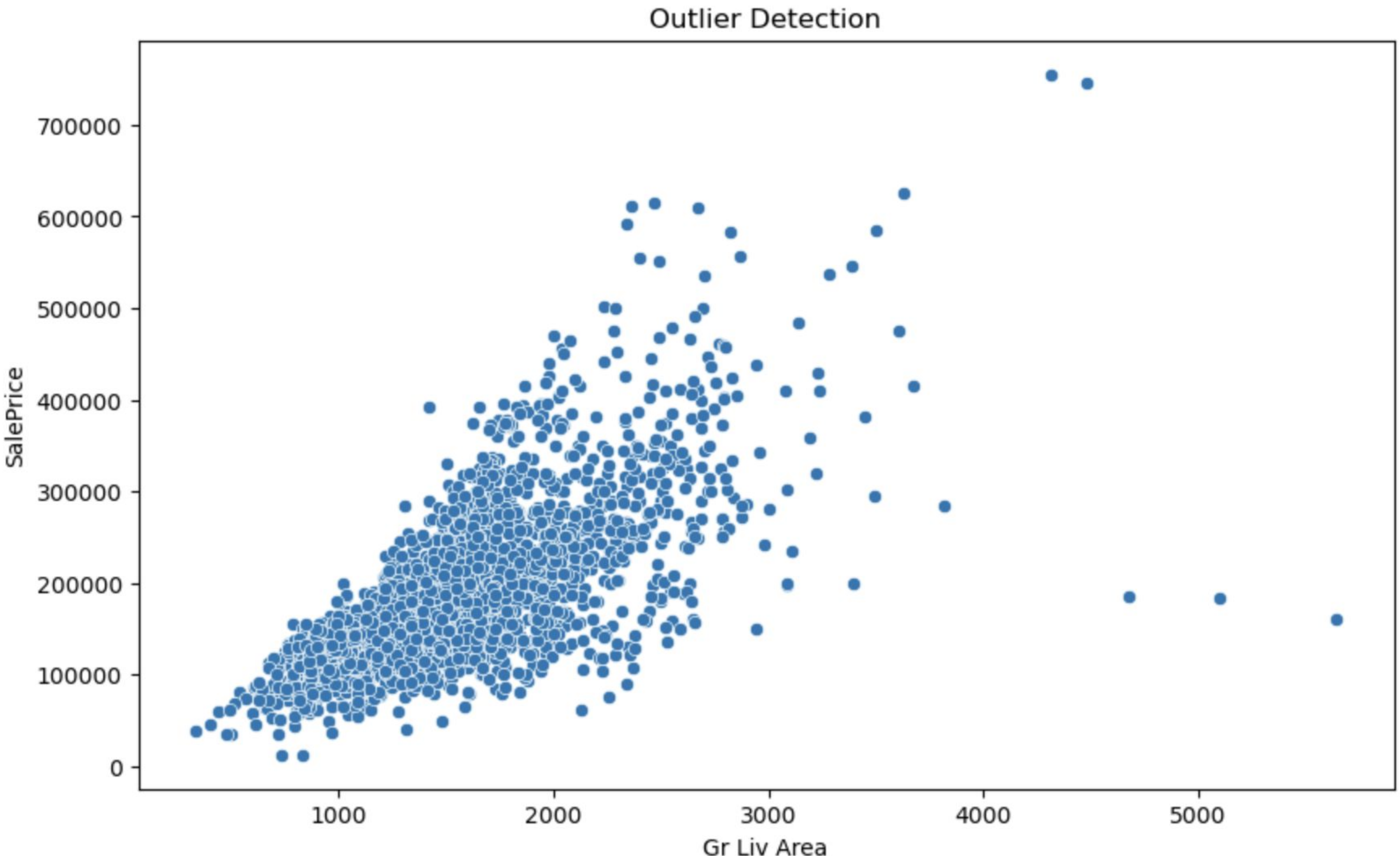
Identified missing data and handled it through imputation.

Visualized relationships between features like `Gr Liv Area`, `Overall Qual`, `TotRms AbvGrd`, etc., with the target `SalePrice`.

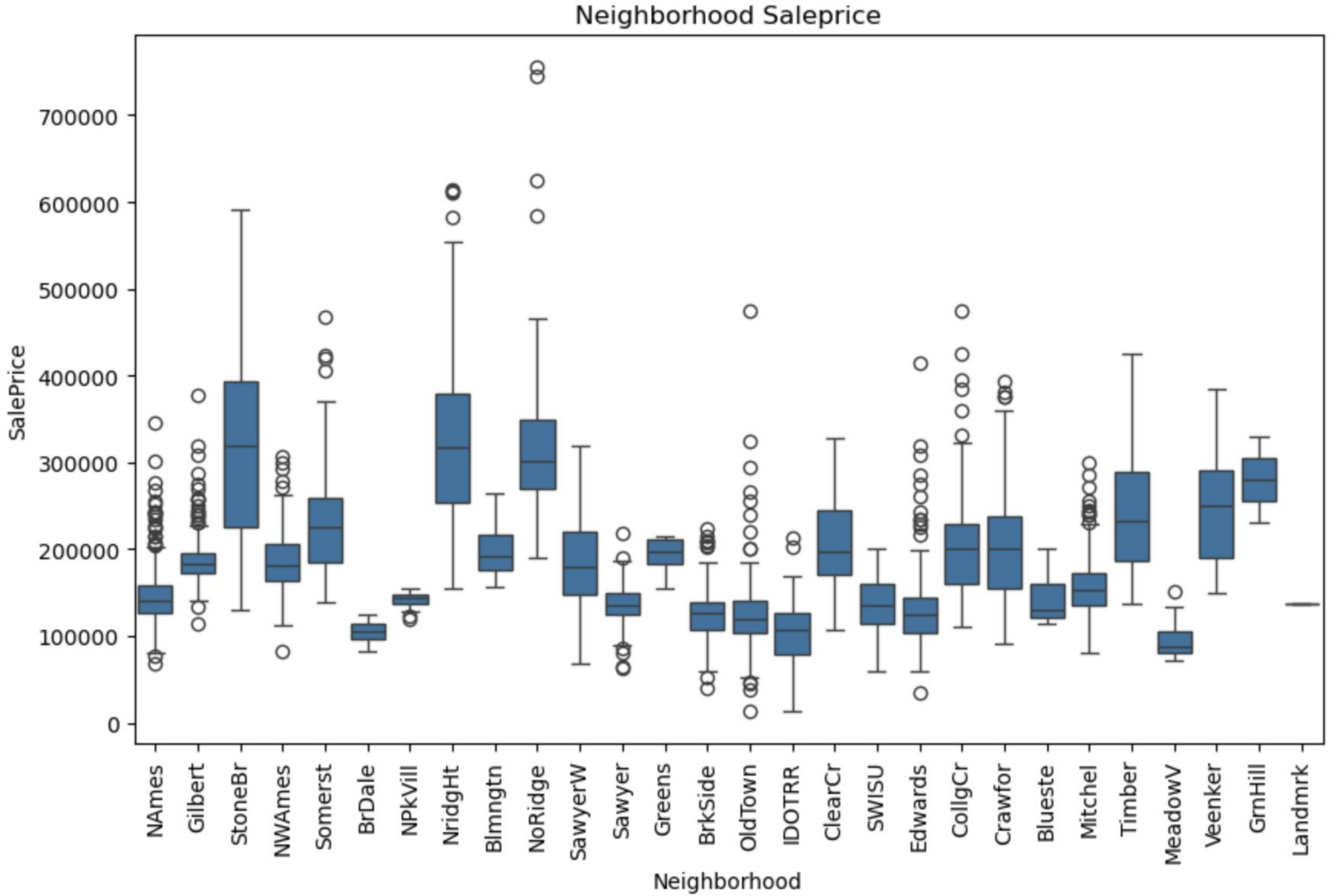
Used correlation heatmap to identify key features influencing sale prices.

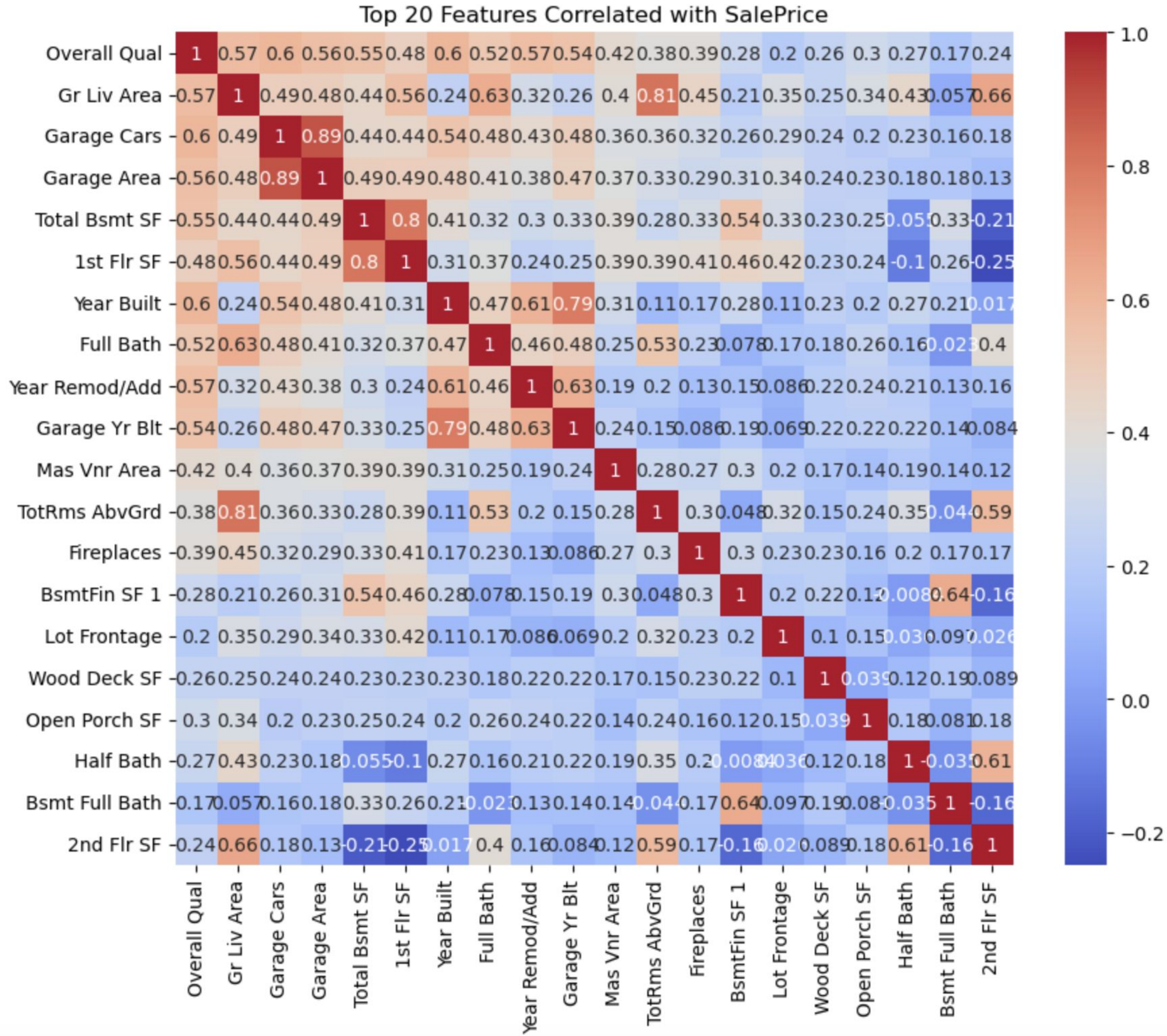

SalePrice Distribution
Count
SalePrice
0
50
100
150
200
250
300
0
100000
200000
300000
400000
500000
600000
700000


Gr Liv Area Sales Price
SalePrice
0
100000
200000
300000
400000
500000
600000
700000
1
2
3
4
5
6
7
8
9
10
Gr Liv Area



Outlier Detection
SalePrice
Gr Liv Area
0
100000
200000
300000
400000
500000
600000
700000
1000
2000
3000
4000
5000


Neighborhood Saleprice
SalePrice
0
100000
200000
300000
400000
500000
600000
700000
NAmes
Gilbert
StoneBr
NWAmes
Somerst
BrDale
NPkVill
NridgHt
Blmngtn
NoRidge
SawyerW
Sawyer
Greens
BrkSide
OldTown
IDOTRR
ClearCr
SWISU
Edwards
CollgCr
Crawfor
Blueste
Mitchel
Timber
MeadowV
Veenker
GrnHill
Landmrk
Neighborhood


Top 20 Features Correlated with SalePrice
Overall Qual
Gr Liv Area
Garage Cars
Garage Area
Total Bsmt SF
1st Flr SF
Year Built
Full Bath
Year Remod/Add
Garage Yr Blt
Mas Vnr Area
TotRms AbvGrd
Fireplaces
BsmtFin SF 1
Lot Frontage
Wood Deck SF
Open Porch SF
Half Bath
Bsmt Full Bath
2nd Flr SF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Feature Engineering and Model Selections

Transformations:

- Handled categorical features with OneHotEncoding.
- Standardized numerical features using `StandardScaler`.
- Split dataset into training and testing sets to validate model performance.

Model Choice:

- Selected Ridge regression for its regularization properties to avoid overfitting.
- Also tested Lasso regression to compare performance with Ridge.

Model Training and Evaluation

Model Training

Modeling Pipeline:

- Used `Pipeline` for streamlined preprocessing and model fitting.
- Ridge and Lasso models were trained on the dataset with the features and targets.

Model Evaluation (Ridge)

Performance:

- **Ridge RMSE:** 28,909.72
- **Ridge R^2:** 0.8958
- **Lasso RMSE:** 29,230.17
- **Lasso R^2:** 0.8934
- Demonstrated that Ridge regression is slightly better in performance compared to Lasso.

Prediction Accuracy

Top 5 Most Accurate Predictions:

- Small errors between actual and predicted prices.
- Examples:
  - Actual: $390,000 | Predicted: $389,873.61 | Error: $126.39
  - Actual: $175,900 | Predicted: $176,045.30 | Error: $145.30

Bottom 5 Predictions:

- Larger discrepancies in predicted vs actual prices.
- Example:
  - Actual: $611,657 | Predicted: $436,617.86 | Error: $175,039.14

Model Insights, Conclusion & Next Steps

Summary:

- Ridge model is effective for **mid-range properties**, showing high accuracy with smaller errors.
- Struggles with **high-value properties**, suggesting potential improvements with advanced models for outlier prediction.

Key Takeaways:

- The Ridge model performs well on most properties but faces challenges with high-value homes.
- Potential to explore more complex models (e.g., Random Forest, Gradient Boosting) to improve predictions for outliers.
- Further tuning and feature engineering could enhance model performance.

Next Steps:

- Experiment with different algorithms.
- Hyperparameter tuning and model ensembling.

Thank you