PREDICTING HOUSE PRICES USING ADVANCED REGRESSION TECHNIQUES

ABDULAZEEZ SALIU

Introduction

Exploratory Data analysis

Linear Regression Model

Business recommendation

INTRODUCTION:

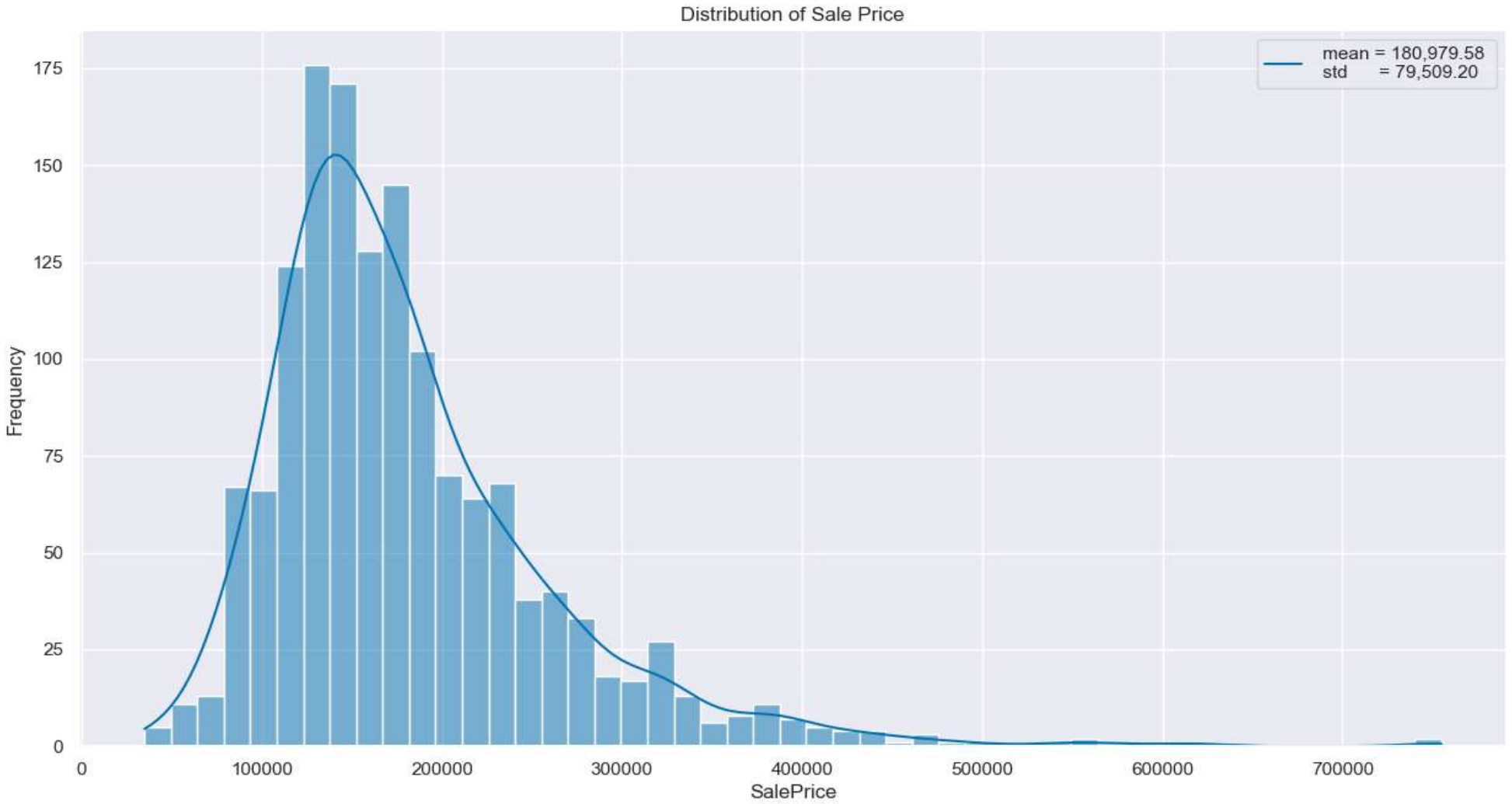
THE REAL ESTATE MARKET IS COMPLEX AND EVER-CHANGING, WITH VARIOUS FACTORS IMPACTING HOME PRICES. ACCURATELY PREDICTING THESE PRICES IS ESSENTIAL FOR BUYERS, SELLERS, REAL ESTATE AGENTS, AND INVESTORS. THE HOUSE PRICE PREDICTION MODEL IS DESIGNED TO OFFER RELIABLE PRICE ESTIMATES BASED ON A COMPREHENSIVE SET OF FEATURES, ENABLING REAL ESTATE COMPANIES TO ACCURATELY VALUE PROPERTIES FOR SALE.

PROBLEM STATEMENT:

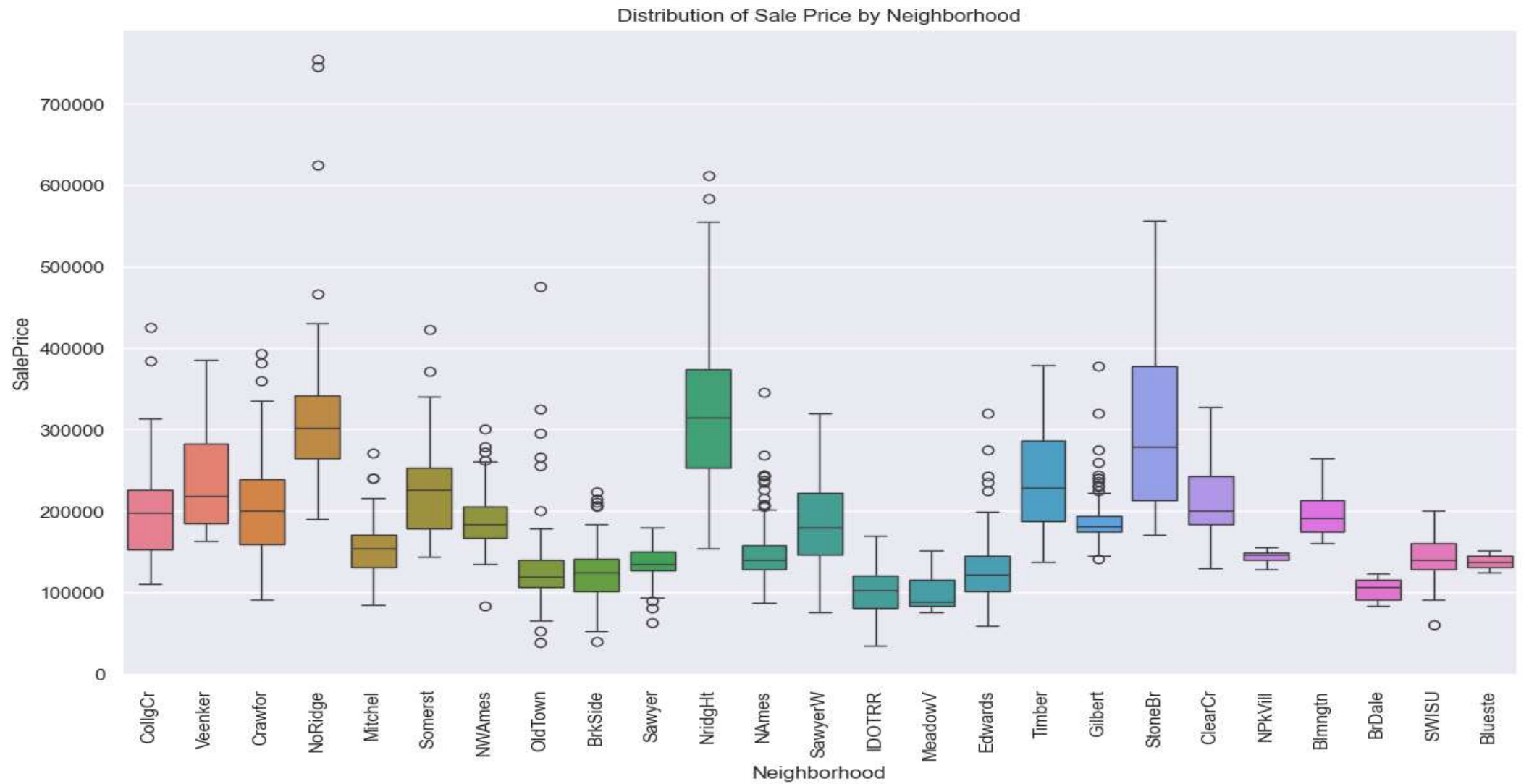
ACCURATELY PREDICTING HOUSE PRICES IS A SIGNIFICANT CHALLENGE IN THE REAL ESTATE MARKET, SHAPED BY NUMEROUS FACTORS LIKE DWELLING TYPES, ZONING CLASSIFICATIONS, LOT FEATURES, PROPERTY CONDITIONS, AND SALE CONDITIONS. TRADITIONAL PROPERTY VALUATION METHODS OFTEN DEPEND ON LIMITED DATA AND SUBJECTIVE JUDGMENTS, LEADING TO INCONSISTENCIES AND INACCURACIES. THIS UNCERTAINTY CAN IMPACT VARIOUS STAKEHOLDERS, INCLUDING HOME BUYERS, SELLERS, REAL ESTATE AGENTS, AND INVESTORS, POTENTIALLY CAUSING FINANCIAL LOSSES AND INEFFICIENCIES FOR THE REAL ESTATE COMPANY.

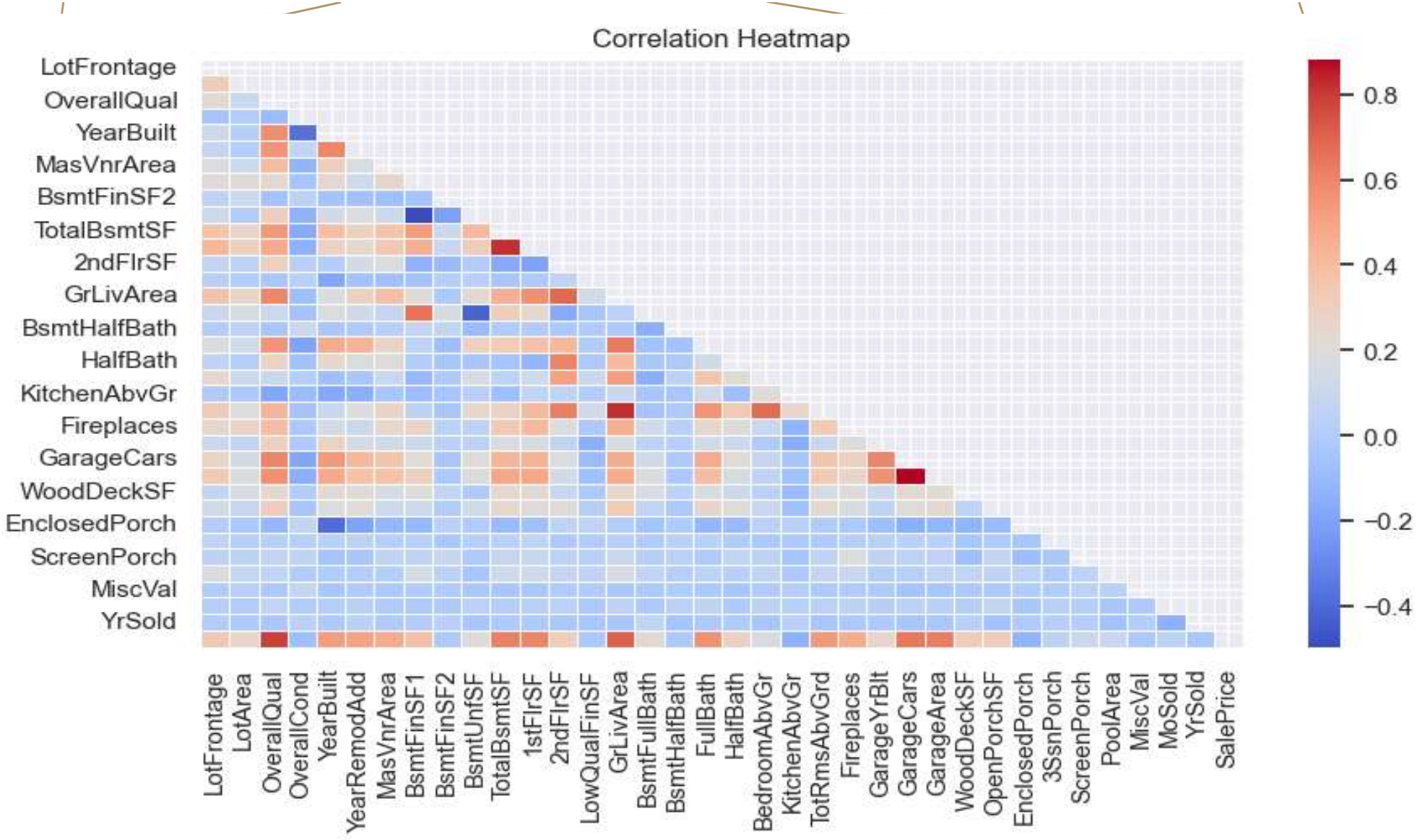
EXPLORATORY DATA ANALYSIS

TARGET VARIABLE

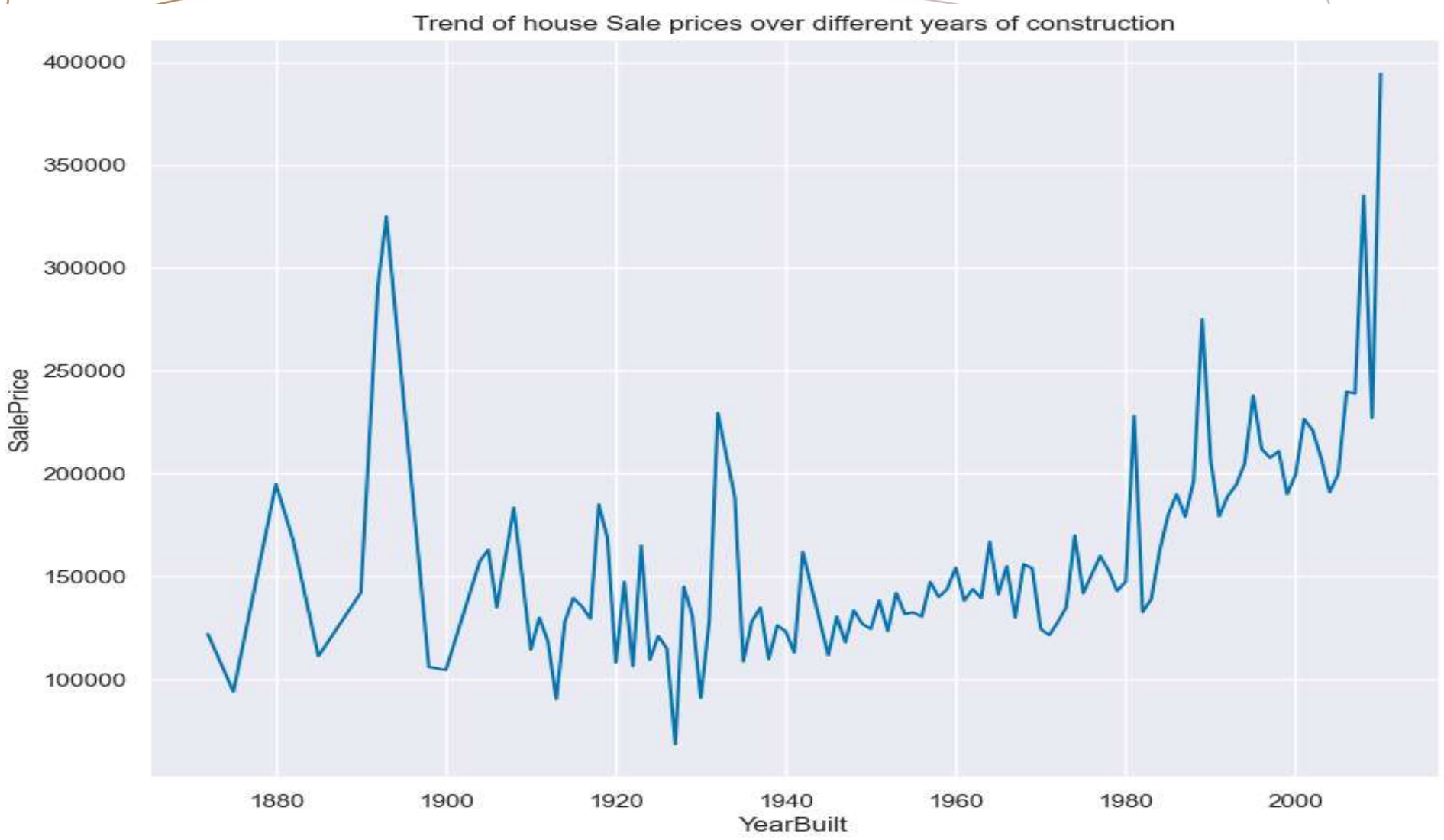


FEATURES`


Distribution of Sale Price by Neighborhood
SalePrice
0
100000
200000
300000
400000
500000
600000
700000
CollgCr
Veenker
Crawfor
NoRidge
Mitchel
Somerst
NWAmes
OldTown
BrkSide
Sawyer
NridgHt
NAmes
SawyerW
IDOTRR
MeadowV
Edwards
Timber
Gilbert
StoneBr
ClearCr
NPkVill
Blmngtn
BrDale
SWISU
Blueste
Neighborhood


Correlation Heatmap
LotFrontage
OverallQual
YearBuilt
MasVnrArea
BsmtFinSF2
TotalBsmtSF
2ndFlrSF
GrLivArea
BsmtHalfBath
HalfBath
KitchenAbvGr
Fireplaces
GarageCars
WoodDeckSF
EnclosedPorch
ScreenPorch
MiscVal
YrSold
LotFrontage
LotArea
OverallQual
OverallCond
YearBuilt
YearRemodAdd
MasVnrArea
BsmtFinSF1
BsmtFinSF2
BsmtUnfSF
TotalBsmtSF
1stFlrSF
2ndFlrSF
LowQualFinSF
GrLivArea
BsmtFullBath
BsmtHalfBath
FullBath
HalfBath
BedroomAbvGr
KitchenAbvGr
TotRmsAbvGrd
Fireplaces
GarageYrBlt
GarageCars
GarageArea
WoodDeckSF
OpenPorchSF
EnclosedPorch
3SsnPorch
ScreenPorch
PoolArea
MiscVal
MoSold
YrSold
SalePrice
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Correlation plot of sales price and Living Area
SalePrice
700000
600000
500000
400000
300000
200000
100000
0
1000
2000
3000
4000
5000
GrLivArea


Trend of house Sale prices over different years of construction
SalePrice
400000
350000
300000
250000
200000
150000
100000
1880
1900
1920
1940
1960
1980
2000
YearBuilt

Insights

Summary of Insights from EDA on Sale Prices

- Sale Price Distribution

The sale prices of houses range from $34,900 to $755,000, with an average price of $180,979 and a median of $163,000. Most houses are priced below $200,000, although there is a significant number of outliers at higher prices.

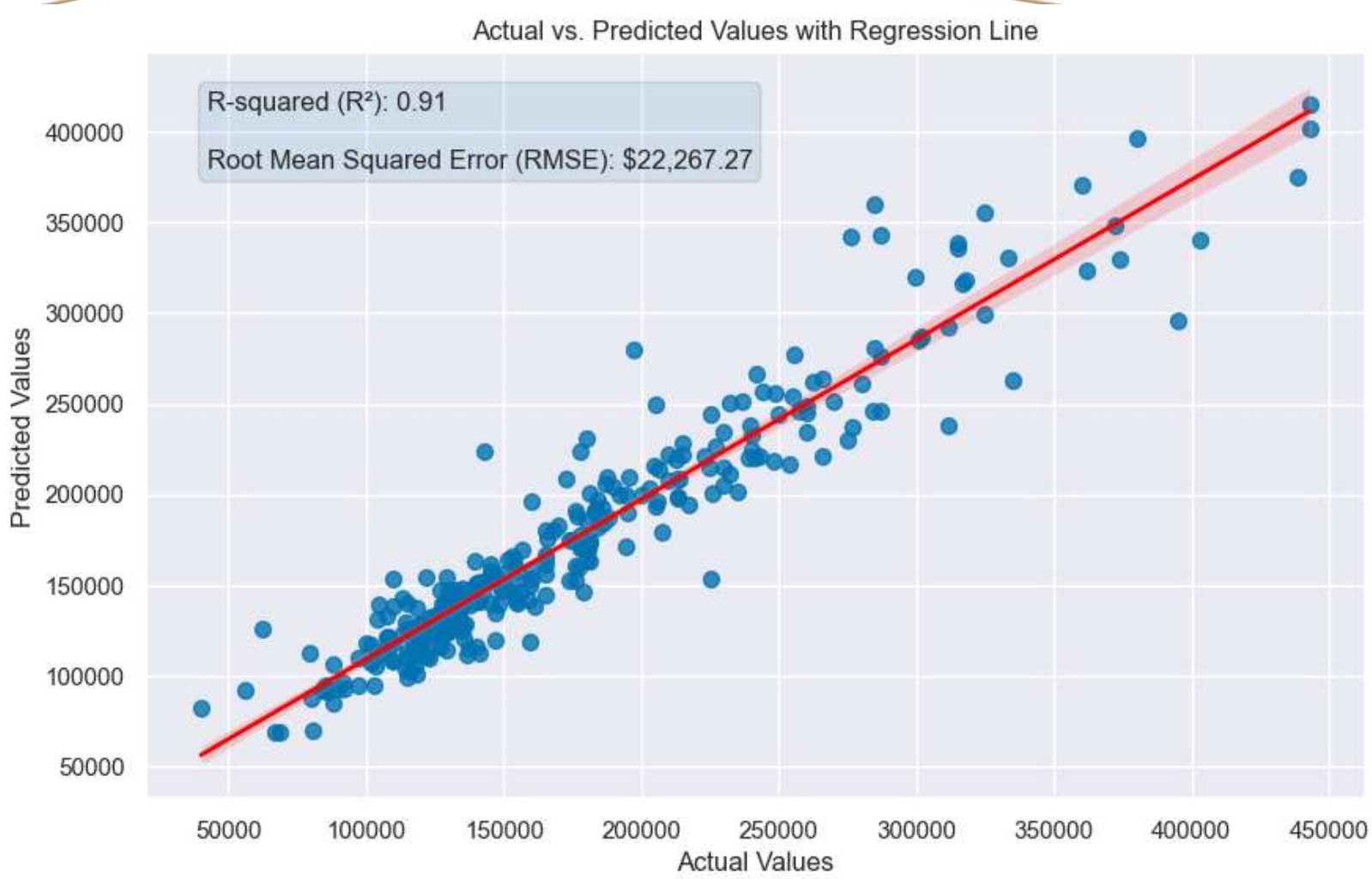
- Neighborhood Influence

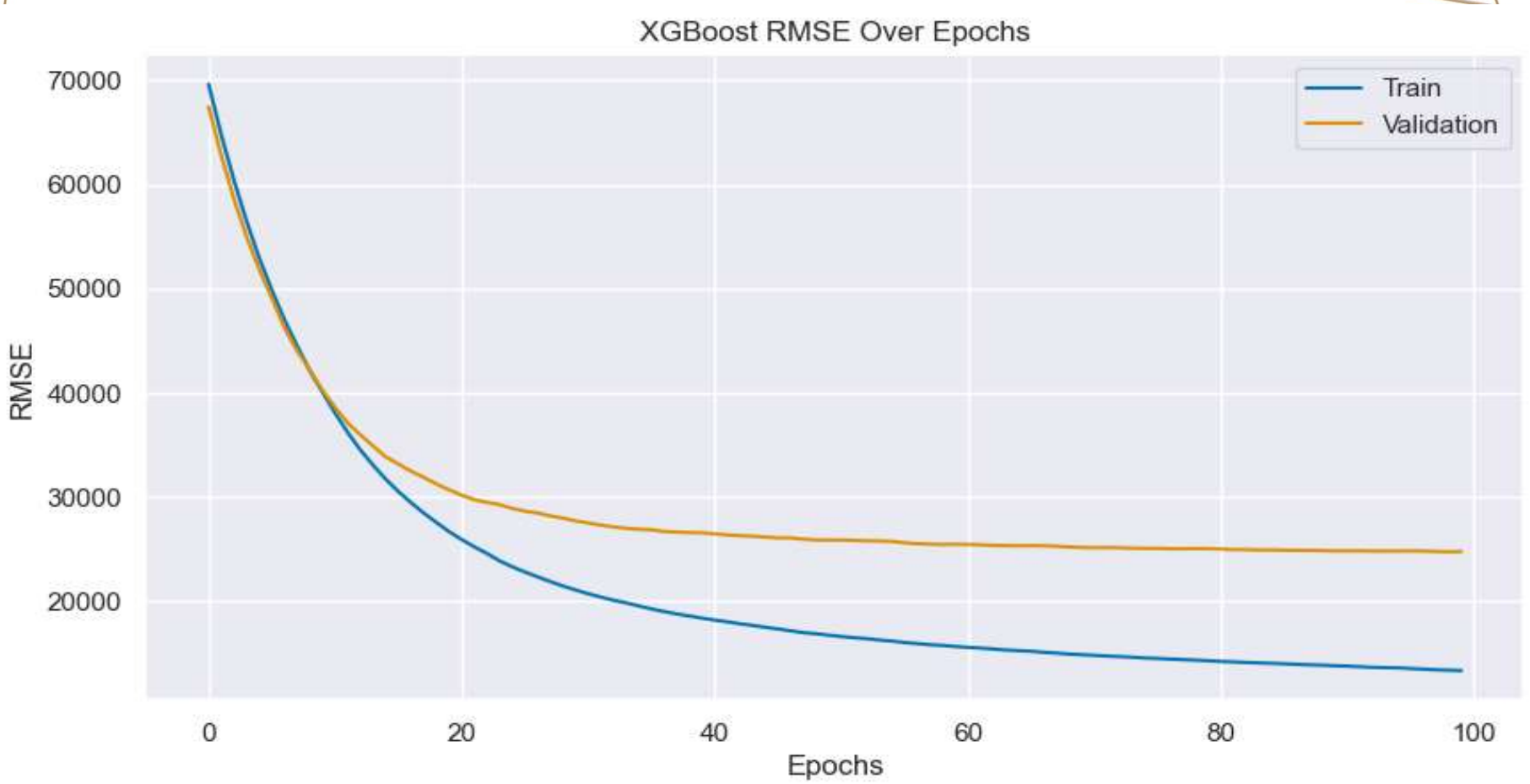
Neighbourhoods such as NridgHt, NoRidge, and StoneBr consistently appear as high-price areas, while MeadowV, IDOTRR, and BrDale are associated with lower prices.

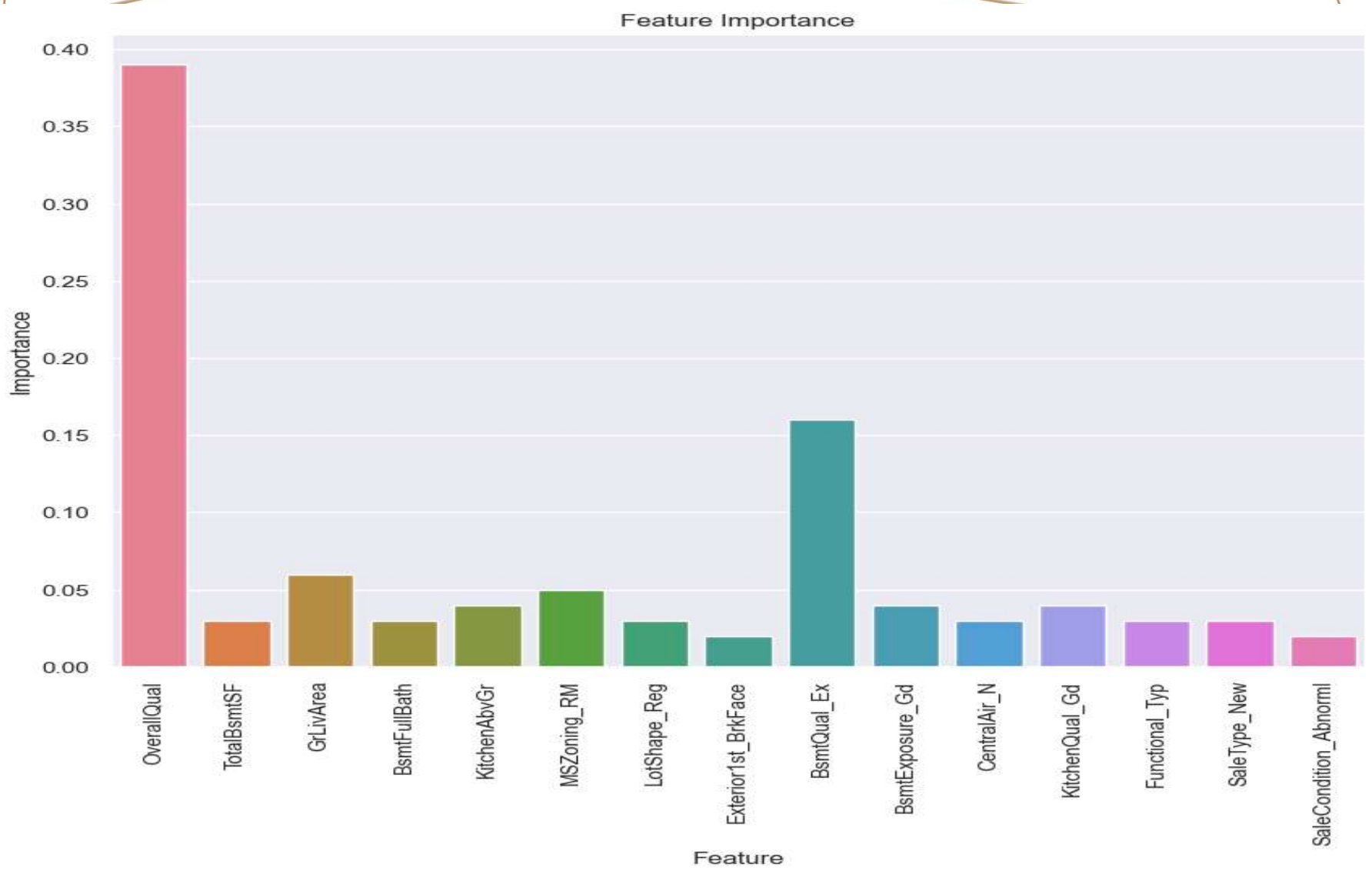
- Historical and Seasonal Trends

House sale prices have generally increased over time, with noticeable peaks around the early 1900s and a significant upward trend toward the year 2000. This reflects historical market trends and possibly improvements in construction quality and amenities. Additionally, house prices show seasonal variations, with lower prices in January, a peak in February, a dip in April, and the highest prices typically occurring in September. These trends can help both buyers and sellers determine the best time of year to buy or sell properties.

LINEAR REGRESSION MODEL (XGBOOST)


Actual vs. Predicted Values with Regression Line
R-squared (R²): 0.91
Root Mean Squared Error (RMSE): $22,267.27
Predicted Values
Actual Values
50000
100000
150000
200000
250000
300000
350000
400000
450000


XGBoost RMSE Over Epochs
Train
Validation
RMSE
70000
60000
50000
40000
30000
20000
0
20
40
60
80
100
Epochs


Feature Importance
Importance
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
OverallQual
TotalBsmtSF
GrLivArea
BsmtFullBath
KitchenAbvGr
MSZoning_RM
LotShape_Reg
Exterior1st_BrkFace
BsmtQual_Ex
BsmtExposure_Gd
CentralAir_N
KitchenQual_Gd
Functional_Typ
SaleType_New
SaleCondition_Abnorml
Feature

BUSINESS RECOMMENDATIONS

Business Recommendations

To effectively predict house prices and achieve a minimum accuracy of 85% and RMSE of $25,000, the real estate company should focus on the following recommendations for model deployment:

- Track Model Accuracy:

Regularly monitor the accuracy score of the model's predictions to assess the variance between the predicted and actual sale prices. This will help in managing the error in quoted sale prices.

- Set a Target RMSE:

Given that the XGBoost model currently has an RMSE of $23,232 and r-squared of 91%, it's advisable to set a target RMSE around $25,000. This target accounts for potential overfitting and changes in data trends once the model is in use.

- Adjust Based on Performance:

As data collection continues and the model's performance is evaluated in real-world scenarios, adjust the RMSE target as needed. If the model consistently meets or exceeds the current target, consider raising the target to align with advancements in model accuracy or evolving business needs.

THANK YOU