Iowa Housing Price Projection

By Kevin Comaduran

Content overview

The Problem

1. Understand the Ames Housing dataset through EDA
2. Perform any data cleaning
3. Prepare a clean dataset for modeling
4. Make prediction using the model

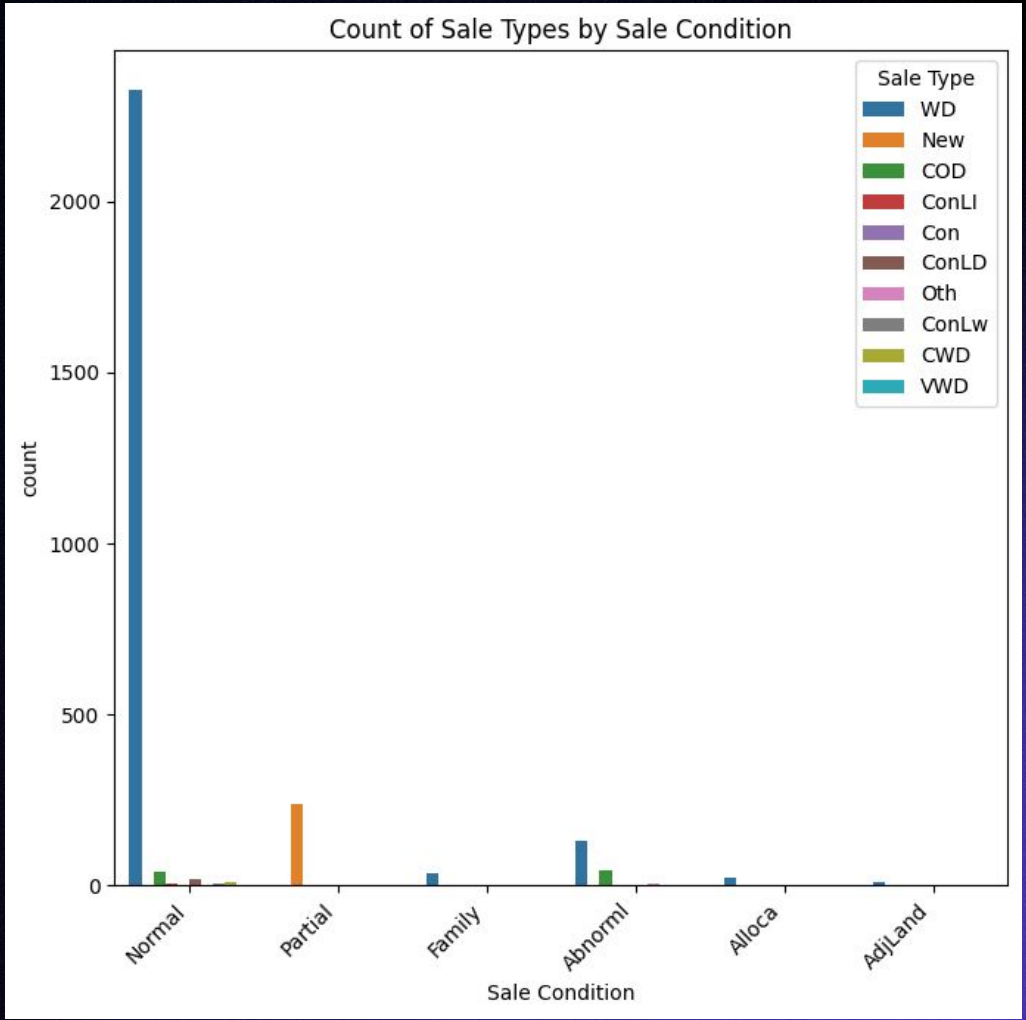
Data Processing

- Removing and filling NaN values
- Converting numeric columns to float
- Removed top 5 Outliers from Ground Living Area column

Insights

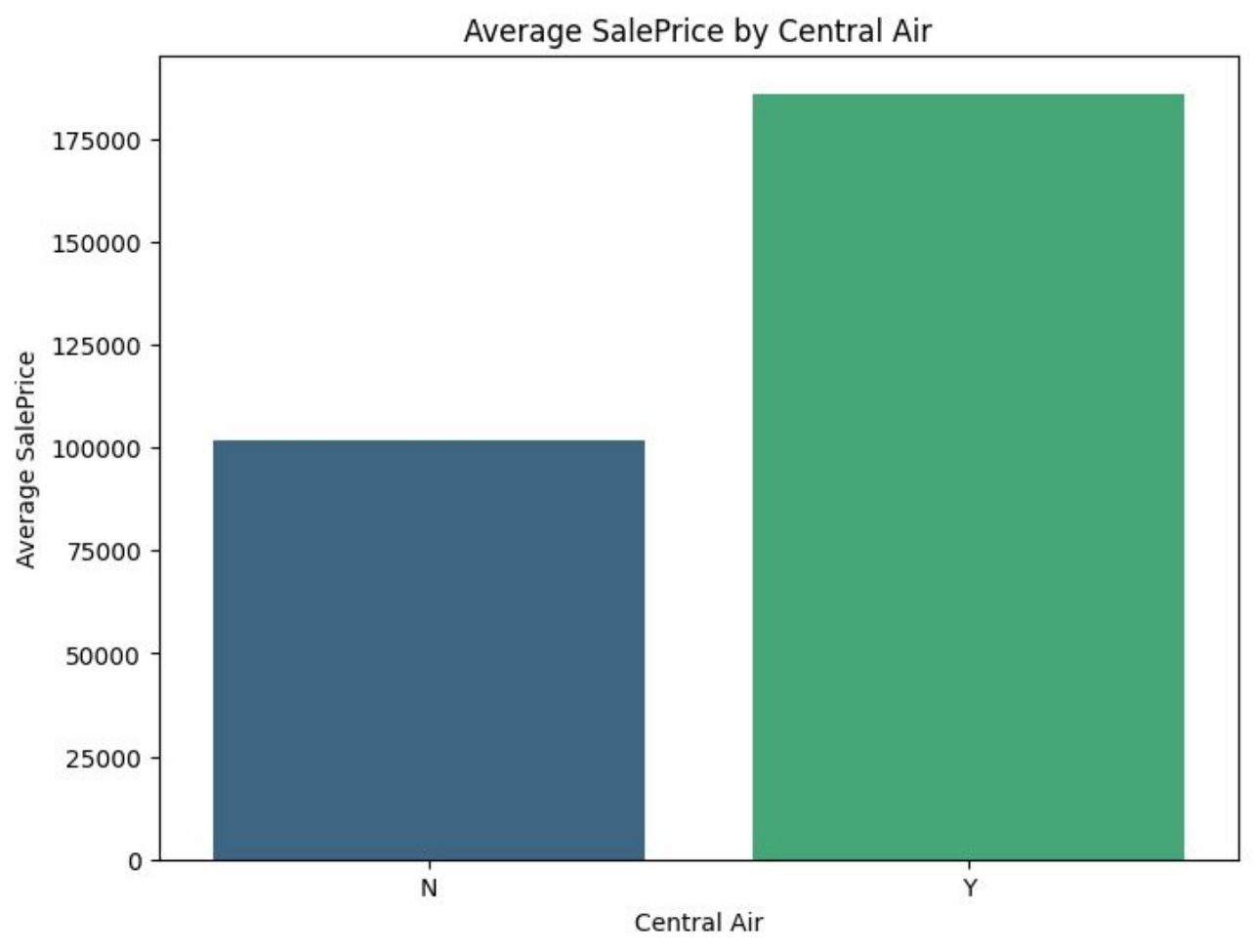
First Observation

- Starting around 1940s, houses built in later years sold more quickly, compared to houses built in the 1900s.
- Majority of the house were sold in normal condition


Count of Sale Types by Sale Condition
Sale Type
WD
New
COD
ConLI
Con
ConLD
Oth
ConLw
CWD
VWD
count
2000
1500
1000
500
0
Normal
Partial
Family
Abnorml
Alloca
AdjLand
Sale Condition

Second Observation

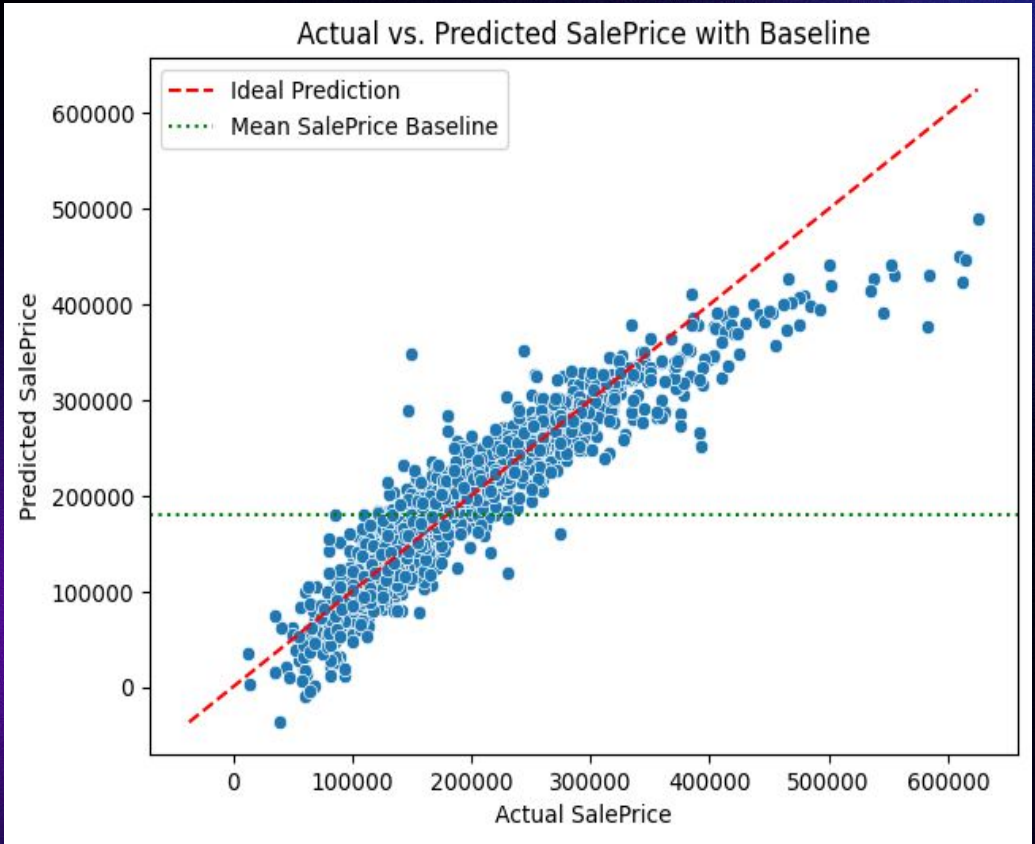
- Houses built from 1958 and onwards have more modern A/C system and have higher price tag
- House that were built in 1910 to 1940s that have natural cooling elements.
- There is a significant difference in average sale price based on sale price base on central air.


Average SalePrice by Central Air
Average SalePrice
175000
150000
125000
100000
75000
50000
25000
0
N
Y
Central Air

What Is A Baseline

- A Baseline a simple model that serves a point of reference for evaluating the performance of the model.

Linear Model



- Used Linear Regression to build the prediction model.
- The SalePrice column was removed from the dataset.
- The remaining columns were utilized as the model's features.
- This approach helped focus the predictive analysis efforts.

Model Metrics

- R2 Score : 87.5%
- Mean Square Error : $28,367
- The model explains 87.5% of the house price variation and predicts and sale price within an average of $28,367.

Thank you!