E-commerce Furniture Sales Prediction Report

# Dataset Overview

Total Rows: 2000

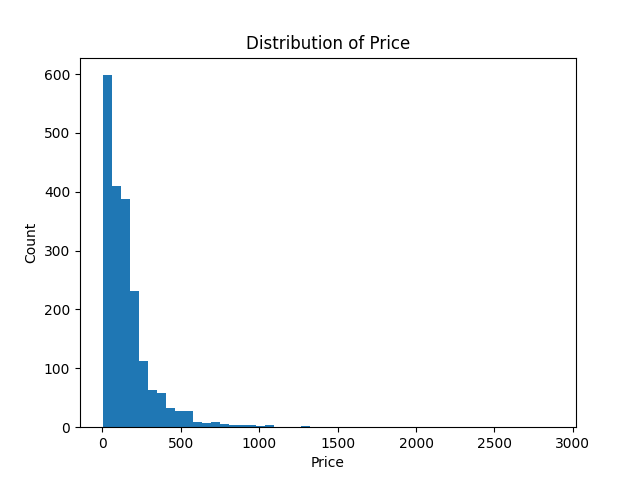
Total Columns: 10

## Missing Values per Column

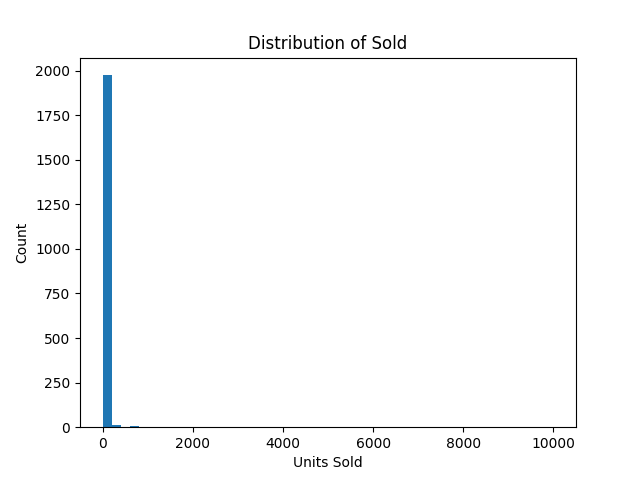
|  |  |
| --- | --- |
| Column | Missing Count |
| productTitle | 0 |
| originalPrice | 1513 |
| price | 0 |
| sold | 0 |
| tagText | 3 |
| price\_clean | 0 |
| originalPrice\_clean | 1513 |
| sold\_num | 0 |
| tagText\_simple | 0 |
| discount\_pct | 0 |

# Exploratory Data Analysis (EDA)

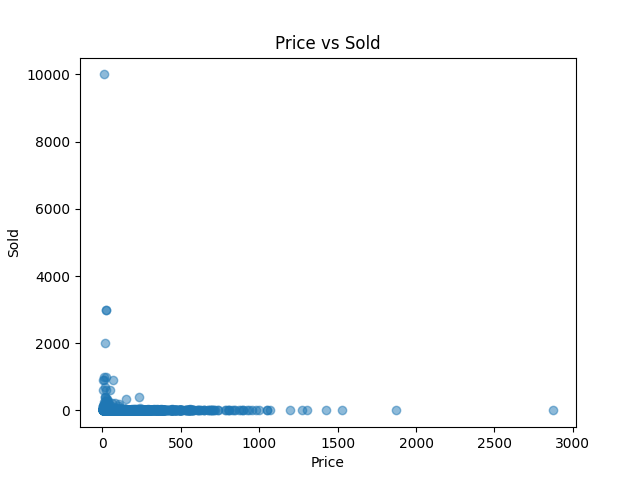
Price Distribution



Sold Distribution



Price Vs Sold



# Model Evaluation Results

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RMSE | MAE | R² |
| LinearRegression | 149.26 | 101.09 | -3.063 |
| Ridge | 114.91 | 73.59 | -1.408 (Best) |
| RandomForest | 116.66 | 28.27 | -1.482 |

# Conclusion

The Ridge performed the best on the dataset.  
It can be used to predict the number of furniture items sold based on product attributes such as price, discount percentage, product title, and shipping tags.  
  
This project demonstrates a full ML pipeline: data cleaning, exploratory analysis, feature engineering, model training, evaluation, and deployment.