

Linear Regression

Predicting Rug Sale Prices

Tara Koplik



Intro

Client:

- Design Allure
 - Interior design firm located in Portland, OR
 - The client is interested in learning about potential sale prices for one of their suppliers Rugs.com in order to inform purchasing decisions.

Objective:

- Build a linear regression model to predict sale prices based on rug features. The model with the highest accuracy will be selected.

Goal:

- Predict sale prices of the rugs on Rugs.com to allow the client to identify if it is optimal to purchase the rug at the current price.

The Data:

Product information scraped from Rugs.com using Selenium

Features: Original price, Country Made In, Collection, Construction (is it handmade?), Indoor/Outdoor Use, Pile Height (height of the tufted loops), Shape, Area, Color, Materials.

Product Details

This rug is currently in stock

Free Shipping & Free Returns

Made in: Turkey

Collection: [Tuscany](#)

Pile Material: 100% Polypropylene | Pile Height: 1/2"

Brand new

The primary color is Black.

Colors include: Black, Green, Ivory, Orange, Red, Yellow, Blue.

Size FT: 5' 0" x 5' 0" | Size CM: 152 cm x 152 cm

SKU #: 6307842 | Currently in stock

Designer Sale

Ends in 15:02:17

5' x 5' Tuscany Square Rug

~~\$198~~ \$99 **50% Off**

Rugs.com



Data Cleaning & Eda

- Remove outliers
- Convert categorical data to binary
- Bucket the data to reduce complexity
- Calculate area

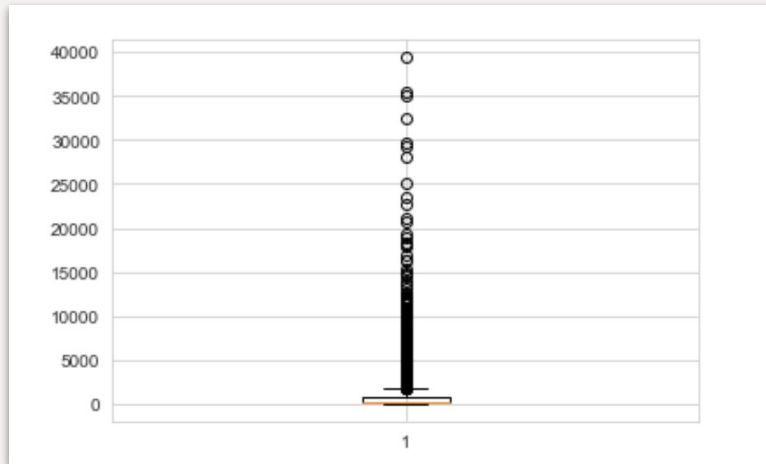


**8' x 11' 4 Hand Knotted Nain
Persian Rug**

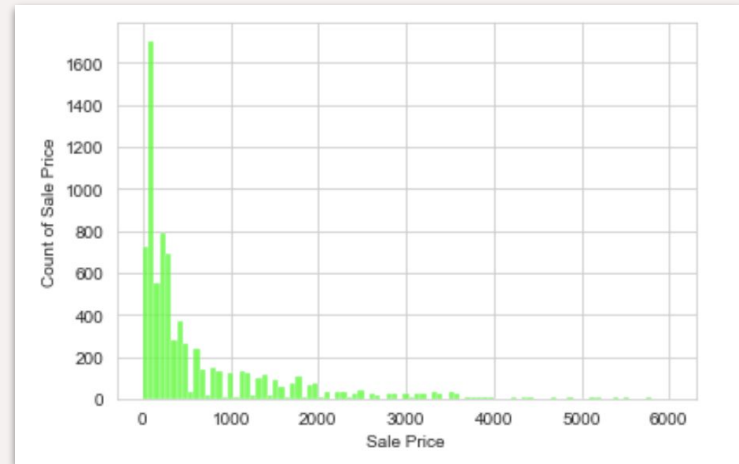
~~\$294,997~~ **\$10,278** **96% Off**

Data Cleaning & Eda

Box Plot of Rug Sale Price

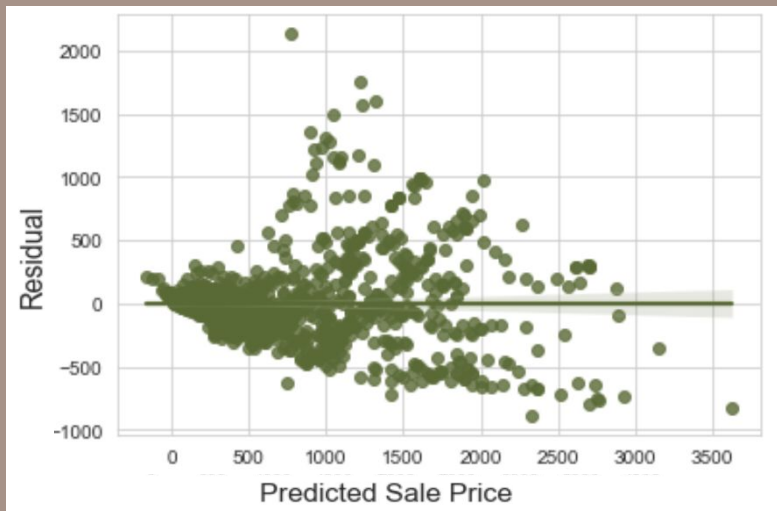


Distribution of Rug Sale Price



Baseline

Linear Regression



Train:

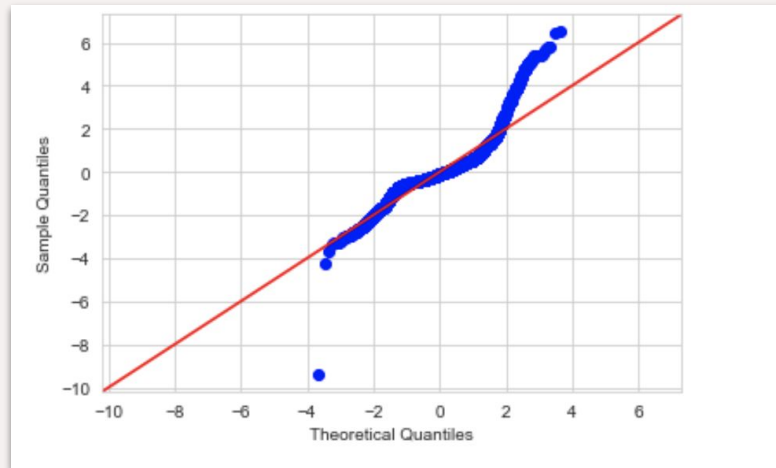
R^2 : 0.8516618639255865
RMSE: 238.7379637310733
MSE: 127.97789835244693

Test:

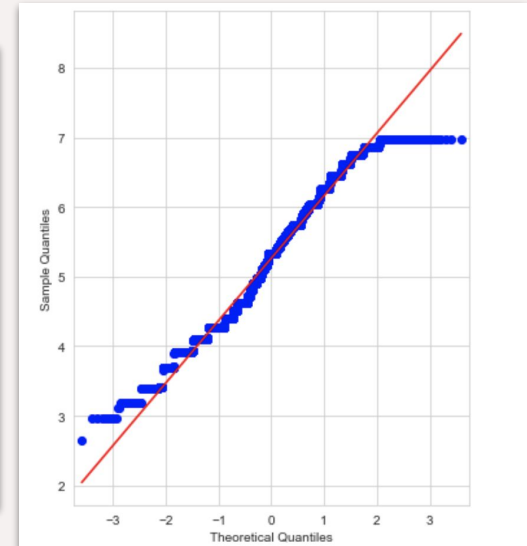
R^2 : 0.839964585810739
RMSE: 251.09783772817403
MSE: 132.29850338388272

Well that's not what we want:

Diagnostic QQ Plot



After removal of
additional outliers and
Box Cox Transformation:



Linear Regression with Box Cox Transformation of Target

Train:

R²: 0.8972325691001115
RMSE: 0.2795328447524849
MSE: 0.2148707656988781

Test:

R²: 0.8978992083456274
RMSE: 0.27978609992977815
MSE: 0.21388048742505195

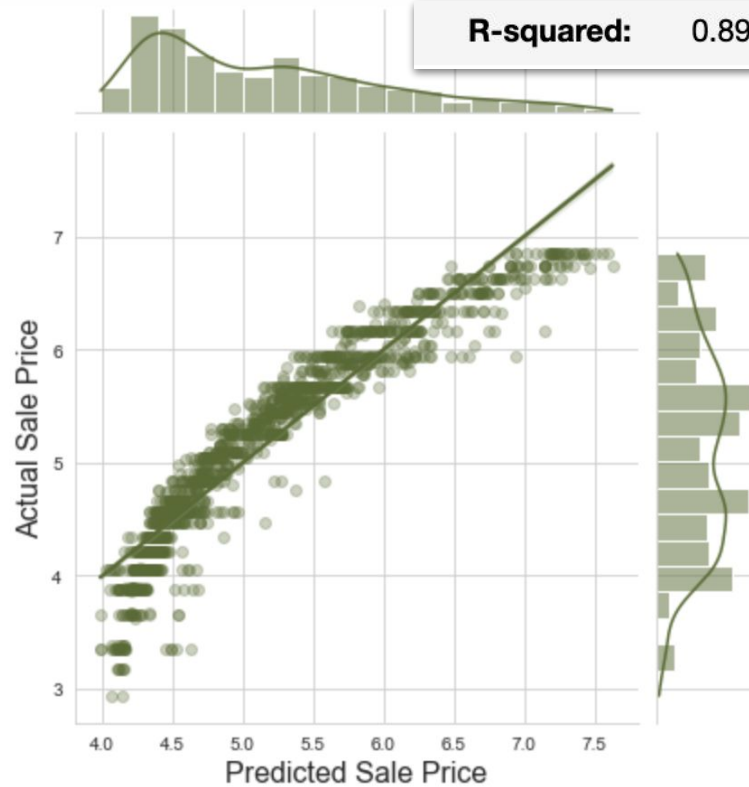


Reduce Model Complexity & Address Multicollinearity

- Remove features that are not adding value
 - Pile Height
 - Indoor/Outdoor Use
- Remove features with high VIF indicating multicollinearity issue



Final Model & Notes



The box cox transformation has the drawback of reduced interpretability.

Requires back transformation.

Future Work

- Try alternative ML models that may be a better method for predicting the data
- Additional data cleaning with a focus on handling outliers
- Add in more suppliers to more accurately reflect market prices