



How to Identify Hotel Deals in Las Vegas

Bentley Ou

10/10/2019

Approach

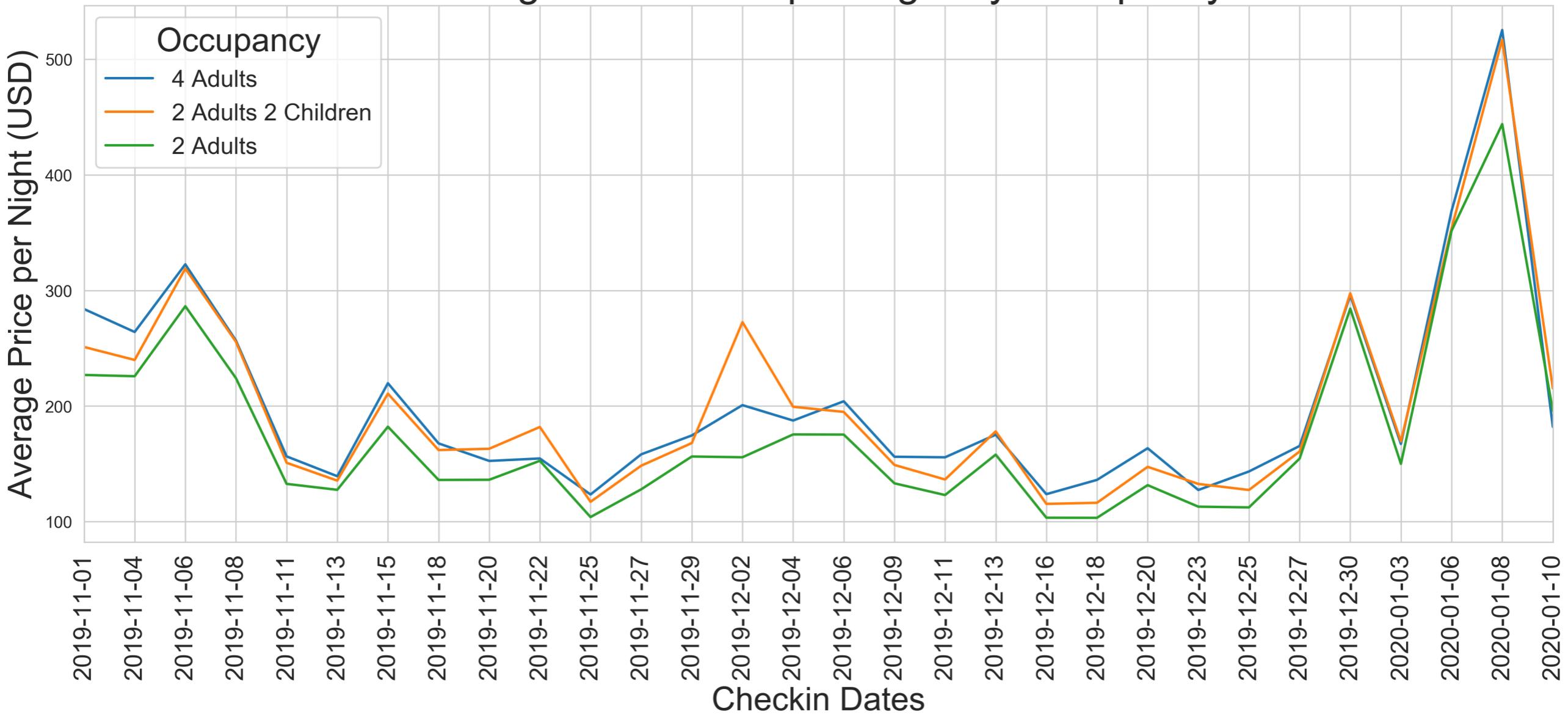
- ✿ Scrape hotel listing data from hotel.com

Approach

- ✿ Scrape hotel listing data from hotel.com
- ✿ Preprocess the data

Hotel Price Trend

Average Hotel Price per Night by Occupancy



Hotel Price Trend

Approach

- ✿ Scrape hotel listing data from hotel.com
- ✿ Preprocess the data
- ✿ Train machine learning model to predict the expected hotel price

Approach

- ✿ Scrape hotel listing data from [hotel.com](#)
- ✿ Preprocess the data
- ✿ Train machine learning model to predict the expected hotel price
- ✿ Evaluate model performance using Mean Absolute Error (MAE) metric

Approach

- ❖ Scrap hotel listing data from [hotel.com](#)
- ❖ Preprocess the data
- ❖ Train machine learning model to predict the expected hotel price
- ❖ Evaluate model performance using Mean Absolute Error (MAE) metric
- ❖ Use the best model to predict the expected hotel price of a listing

Model Preparation

- ✿ Input Features:
 - ✿ eg. Length of stay, Checkin/Checkout dates, Distances to city center, Hotel Reviews/Review counts, Location, Amenities, TripAdvisor Ratings,
- ✿ Output Target: Hotel Price (USD per night)
- ✿ Total Observations (15897 samples)
- ✿ Use Statsmodel to determine initial decision making
 - ✿ Remove features which are not significant P-Val > 0.05

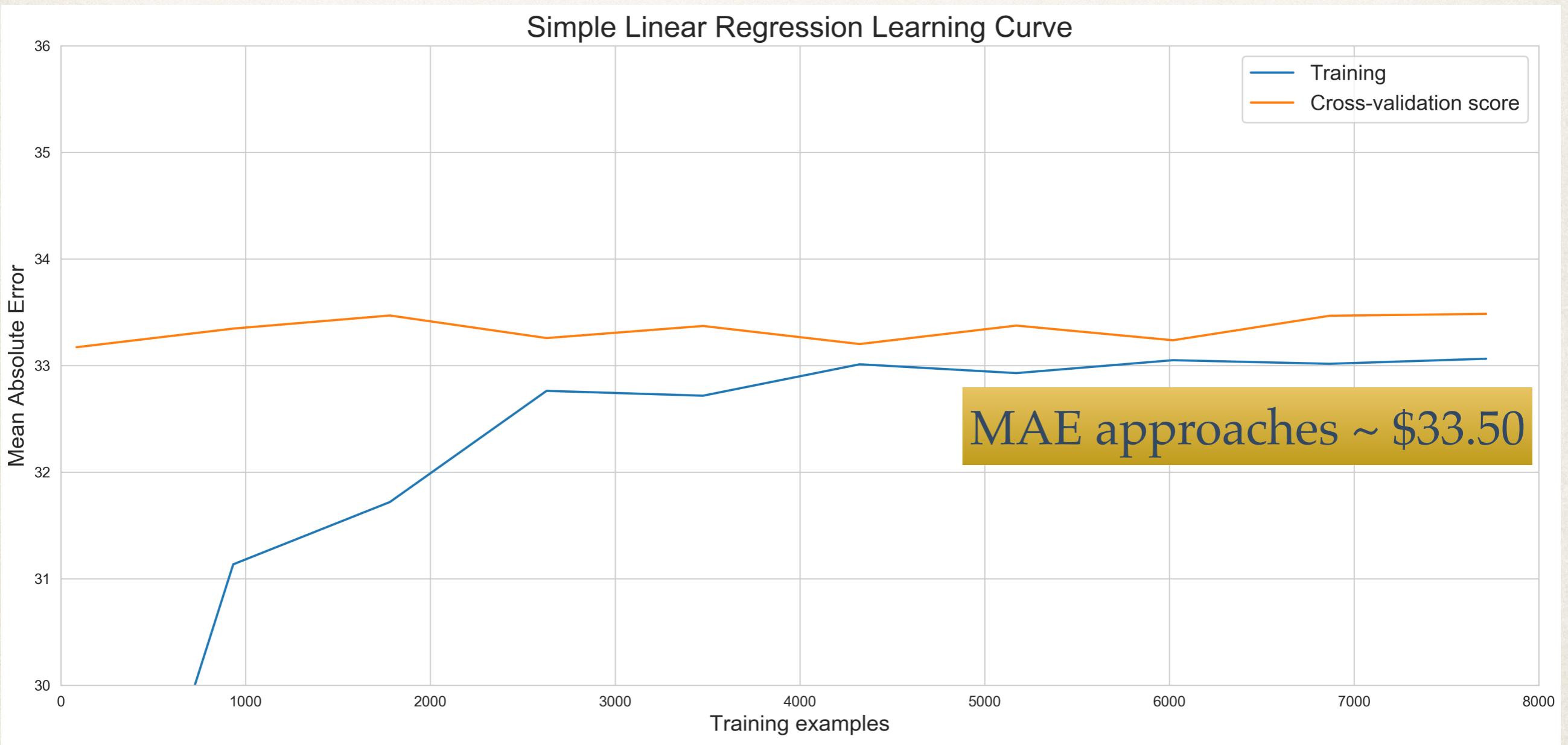
Model Selection

- ❖ Train the data using Simple Regression()
- ❖ Train the data using LassoCV() finding the optimal lambda for feature selection to simplify the model
 - ❖ Lambda: 0.0001
- ❖ Train the data using RidgeCV() finding the optimal lambda to deal with feature collinearity
 - ❖ Lambda: 13.970
- ❖ Train the data using RidgeCV() using non-zero beta coefficients determined by Lasso()
 - ❖ Lambda: 9.011

Cross Validation Scores

Regression Model	MAE
Simple	\$33.22
Lasso	\$33.20
Ridge	\$33.22
Ridge with non-zero Betas	\$33.20

Learning Curve



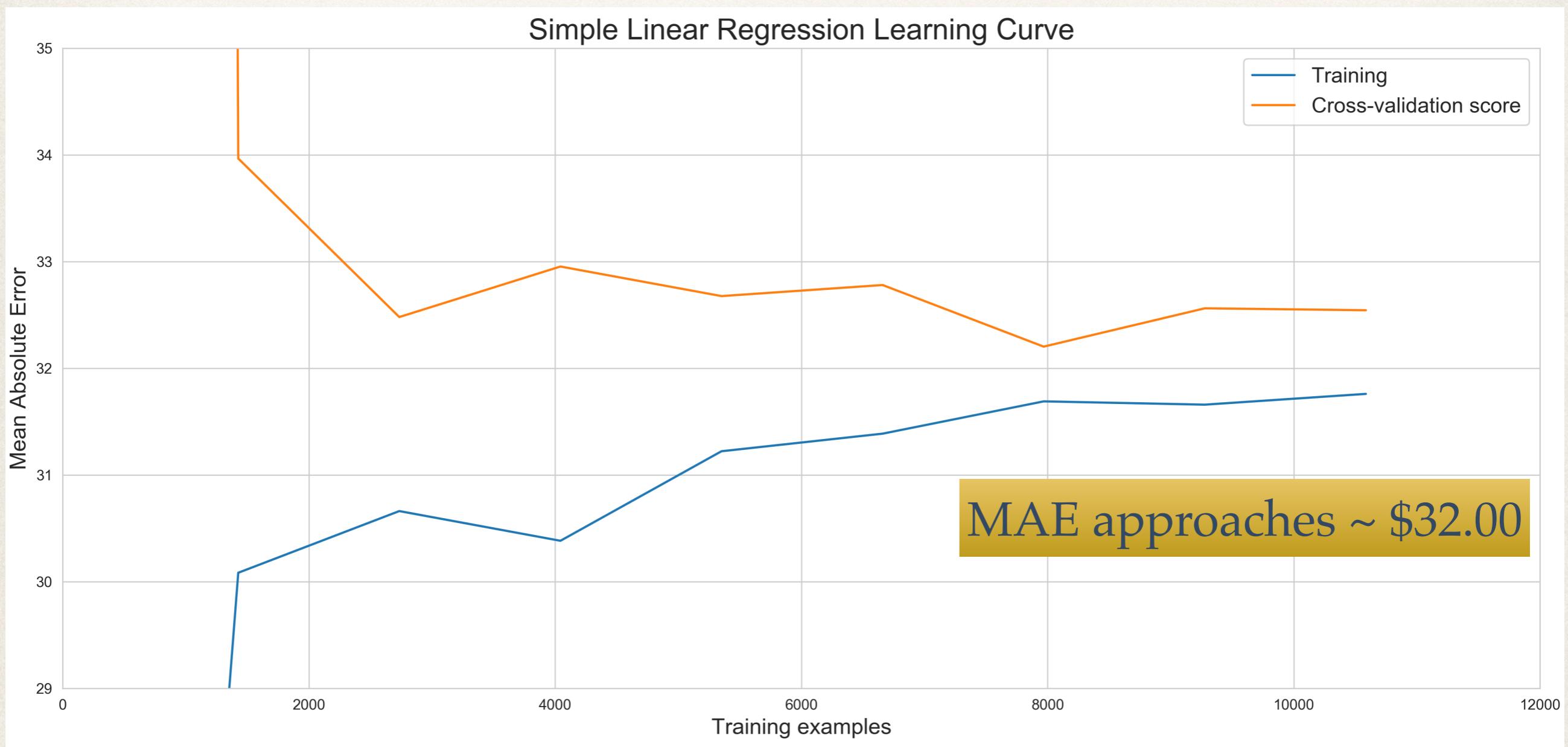
Feature Engineer and Model Selection #2

- ✿ Apply Polynomial Features to degree 2 to all numerical features
 - ✿ Created feature interactions
 - ✿ Improved model complexity
- ✿ Repeat the model selection process again
 - ✿ Simple Regression, LassoCV, RidgeCV
 - ✿ RidgeCV with non-zero betas

Cross Validation Scores 2

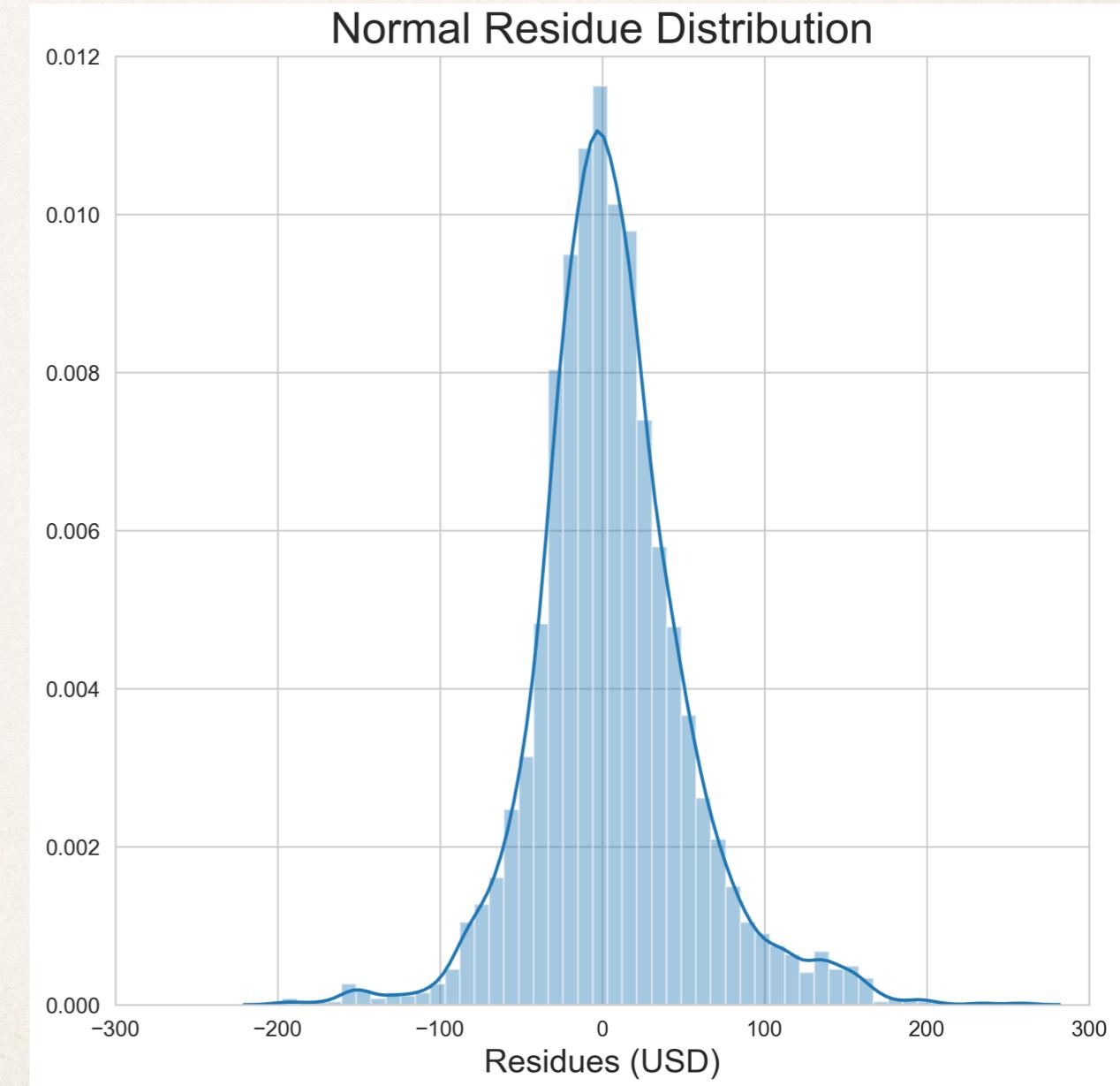
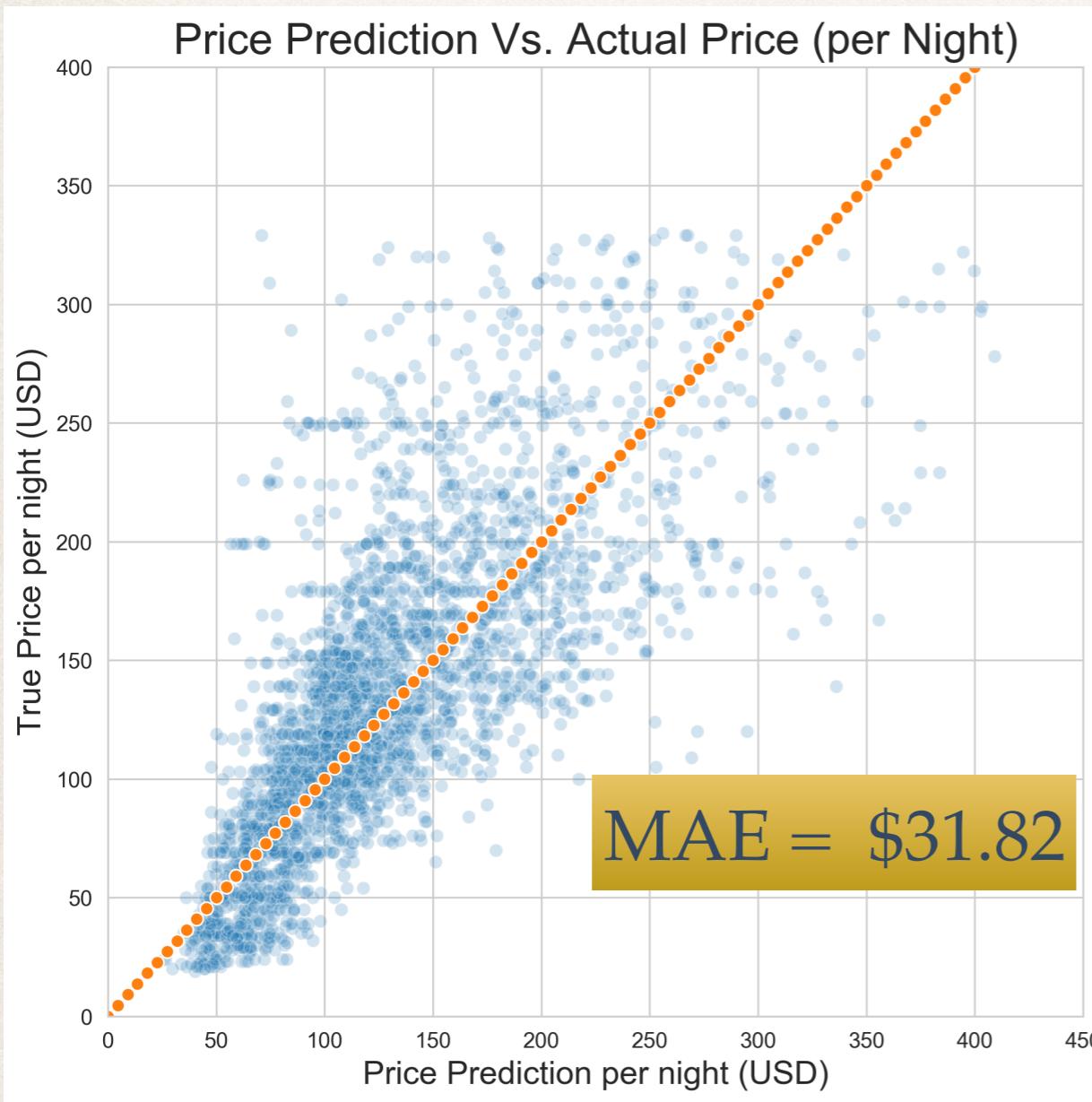
Regression Model	MAE
Simple	\$31.82
Lasso	\$31.92
Ridge	\$31.81
Ridge_nonZero_Beta	\$32.00

Learning Curve



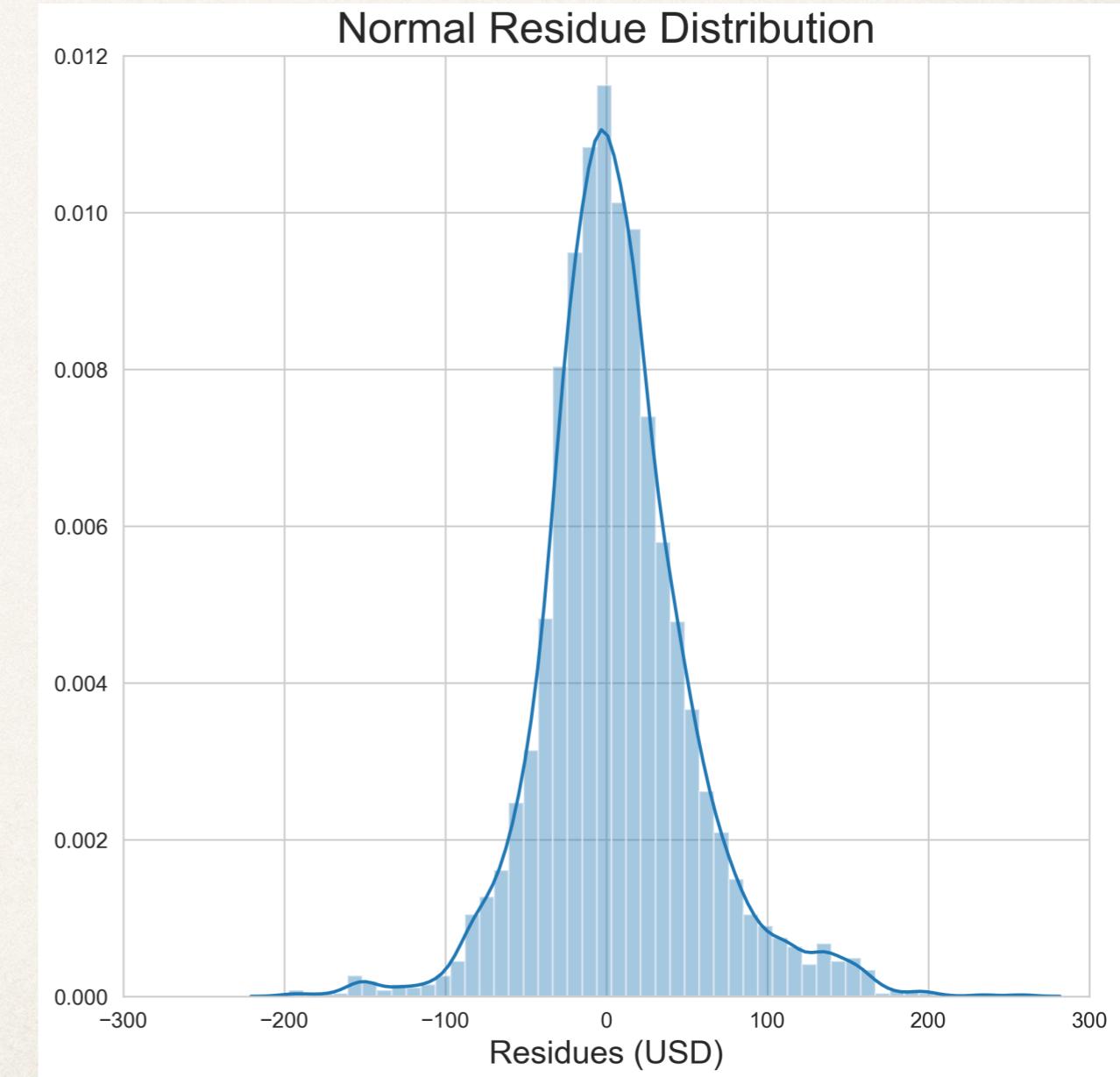
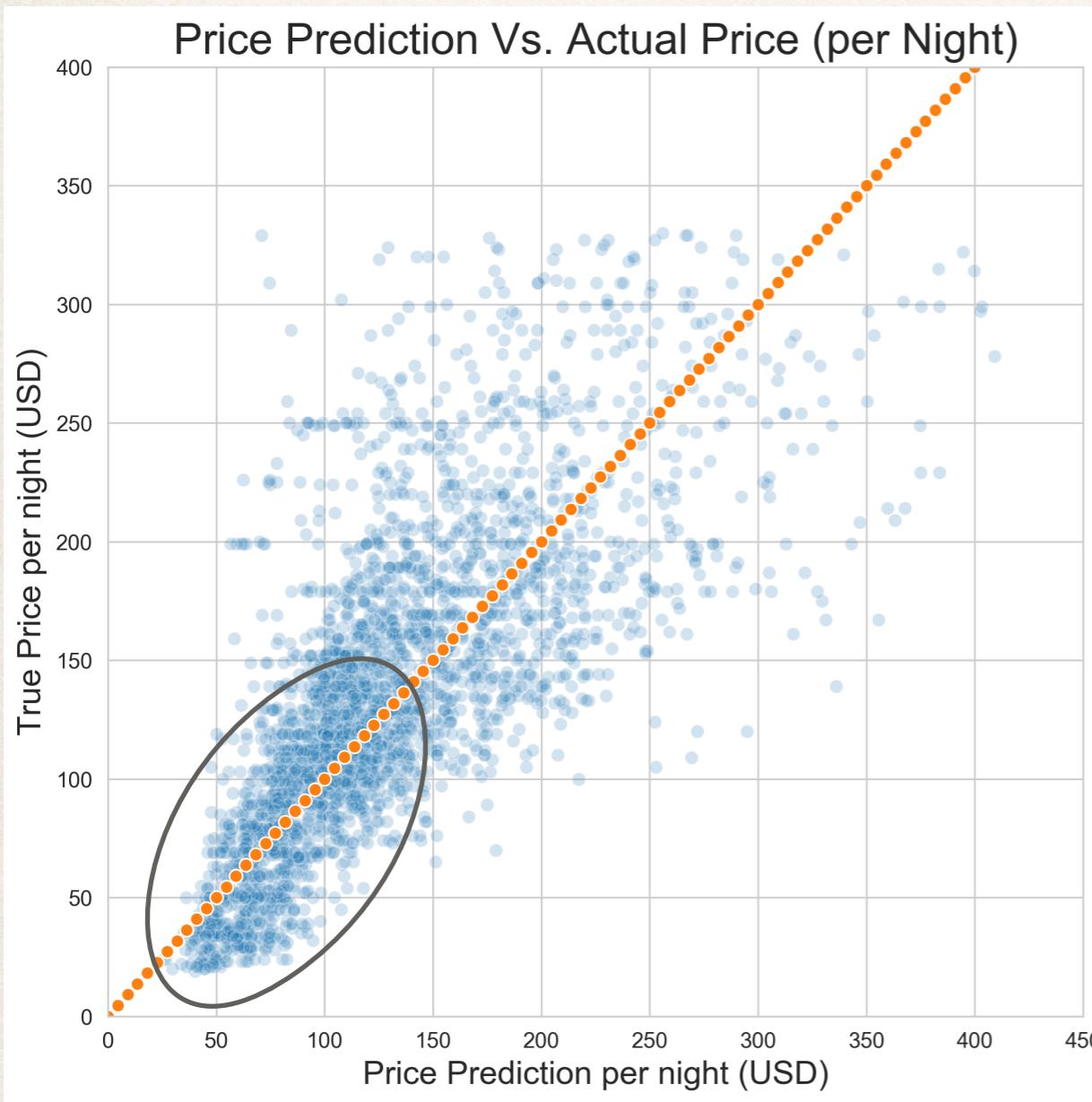
Evaluation on Test Set

Simple Linear Regression



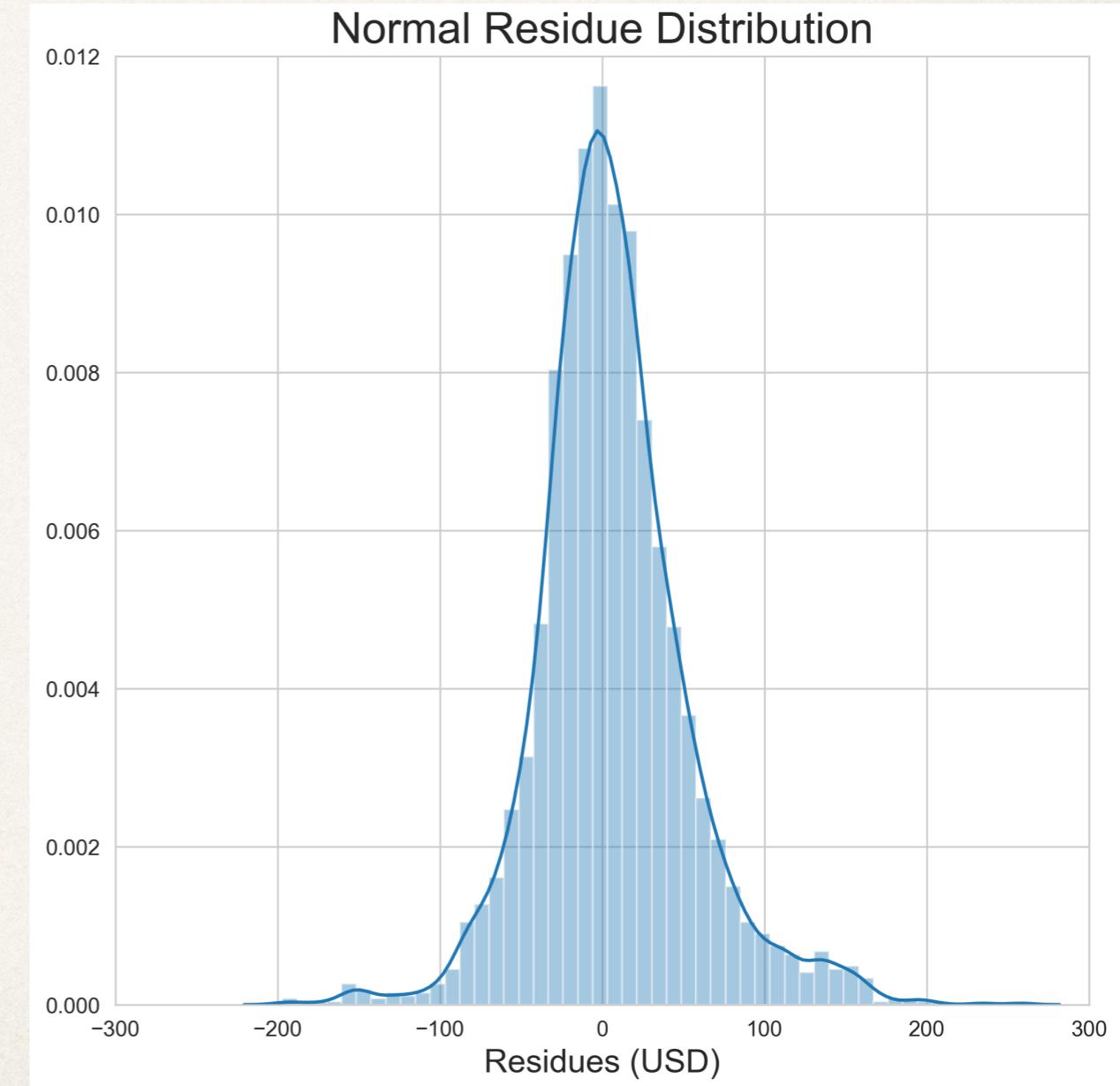
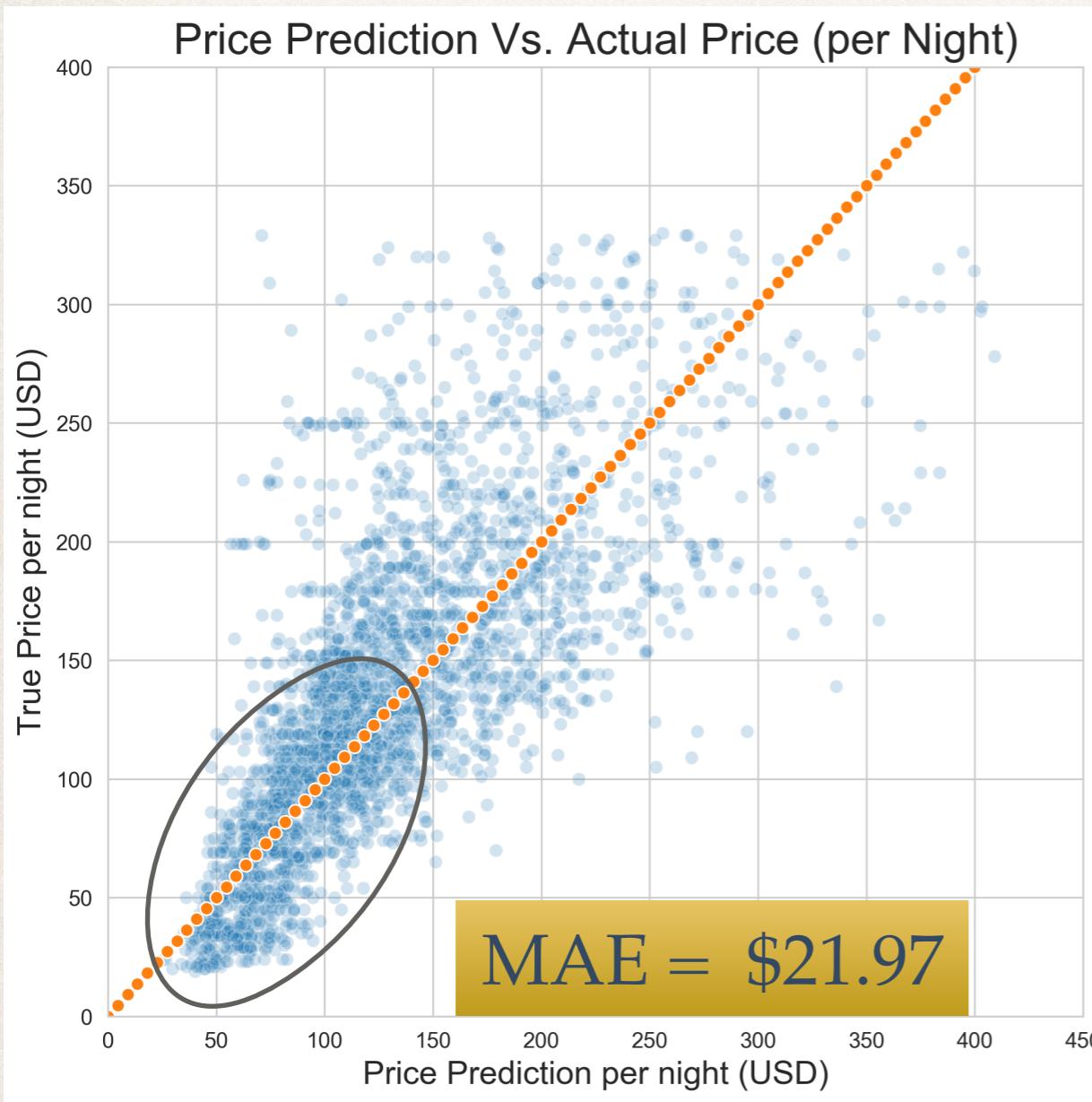
Evaluation on Test Set

Simple Linear Regression



Evaluation on Test Set

Simple Linear Regression



Questions

- ✿ How to identity hotel deals on hotel.com checking into Las Vegas?



Practical Example

- ✿ You are looking book a hotel for a getaway in Las Vegas in the upcoming months.

Practical Example

- ✿ You are looking book a hotel for a getaway in Las Vegas in the upcoming months.
- ✿ You think you are being offered a great deal on hotel.com

Practical Example

- ✿ You are looking book a hotel for a getaway in Las Vegas in the upcoming months.
- ✿ You think you are being offered a great deal on hotel.com
- ✿ Based on the listing attributes, use the model to predict the hotel price

Practical Example

- ✿ You are looking book a hotel for a getaway in Las Vegas in the upcoming months.
- ✿ You think you are being offered a great deal on hotel.com
- ✿ Based on the listing attributes, use the model to predict the hotel price
- ✿ If the predicted price is above than the listed price, you've found a deal!

Improvements

- ❖ Incorporate the time delta between booking and checkin
- ❖ Collect data on room type
- ❖ Extract more specific location details
- ❖ Vary the number of days of stays
- ❖ Spread out data collection in multiple days/weeks if time permits

Thank You!

-Bentley Ou

Appendix

Background

- ❖ Hotel prices can fluctuate as frequently as every minute using today's dynamic pricing technology
- ❖ Main cause for the price fluctuations are:
 - ❖ Supply and demand
 - ❖ Competition
 - ❖ Seasonality
 - ❖ Location
 - ❖ Day of the week

Additional Insights

- ❖ Which dates and hotels to avoid when booking Las Vegas during the holiday season?
- ❖ Largest beta values for the Zip Code and Date Features

Zip Codes	Dates
89135	Checkin: 01/08/20 Checkout: 01/10/20
89012	Checkin: 12/30/19 Checkout: 01/01/19
89144	Checkin: 01/06/20 Checkout: 01/08/20
89183	Checkin: 11/06/19 Checkout: 11/08/19
89177	Checkin: 11/04/19 Checkout: 11/06/19

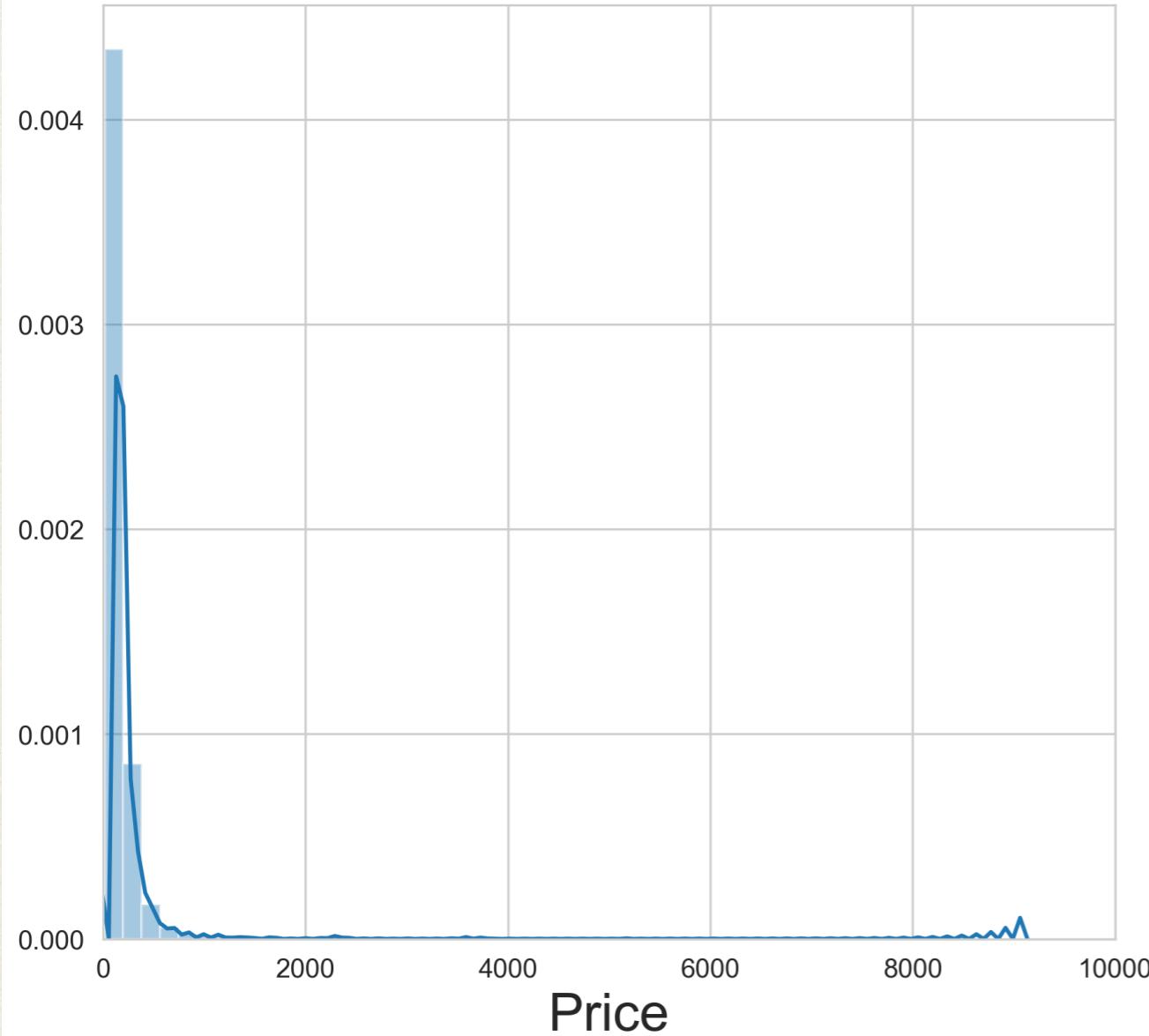
Beta Coefficients

- ✿ Features that contributes the least to the total cost of

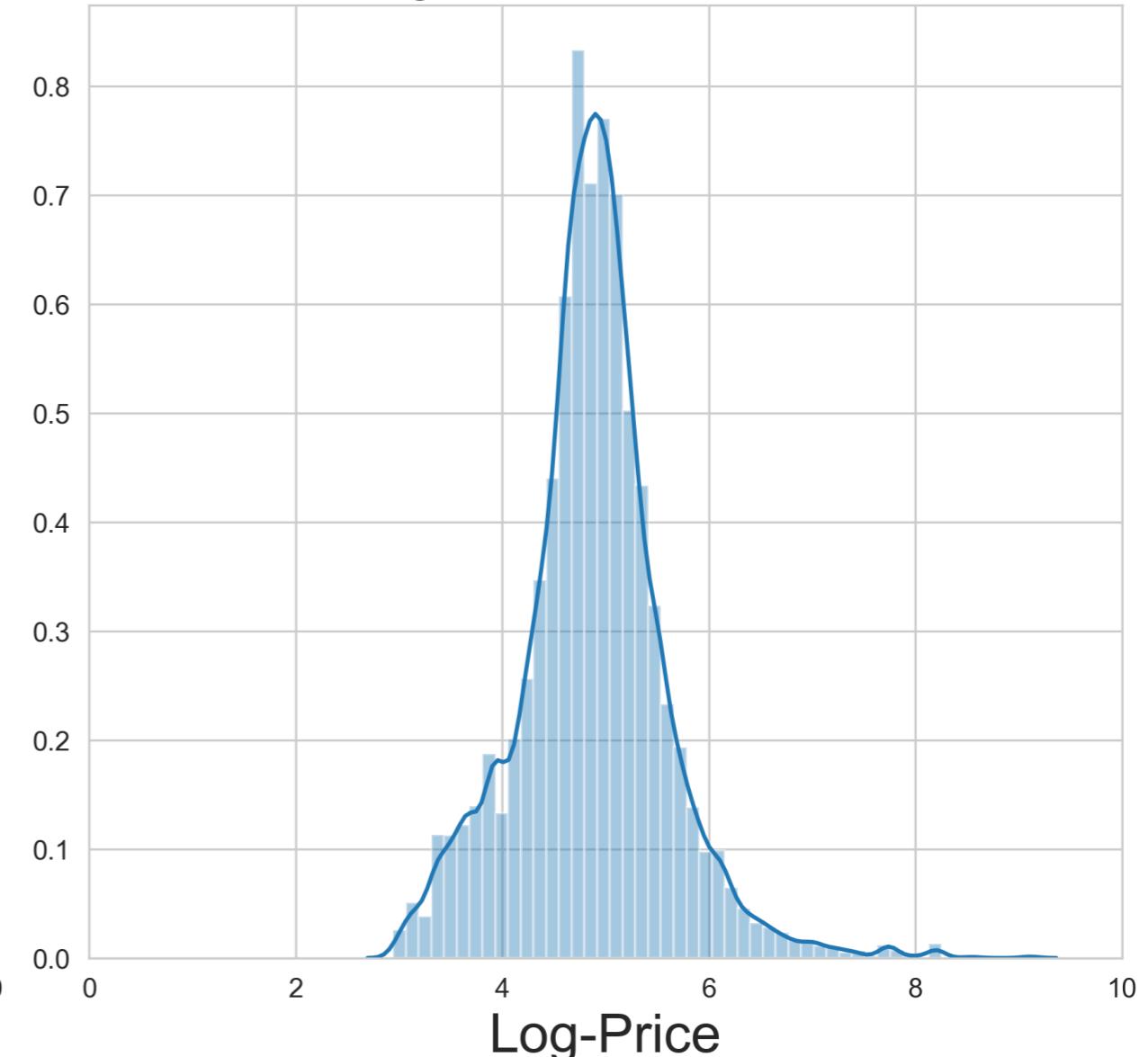
Features	Beta Coefficients	Beta Intercept
Zip Code 89310	\$0.77	
Zip Code 89107	\$0.80	
From 12/20/19 to 12/22/19	\$1.56	\$46.23
From 01/03/20 to 01/05/20	\$1.64	
From 11/25/19 to 11/27/19	\$1.65	

Data Exploration

Price Distribution



Log-Price Distribution



Logarithmic Transformation

End of Slides