

# ***Predicting Ratings***

***on***



**Tripadvisor**

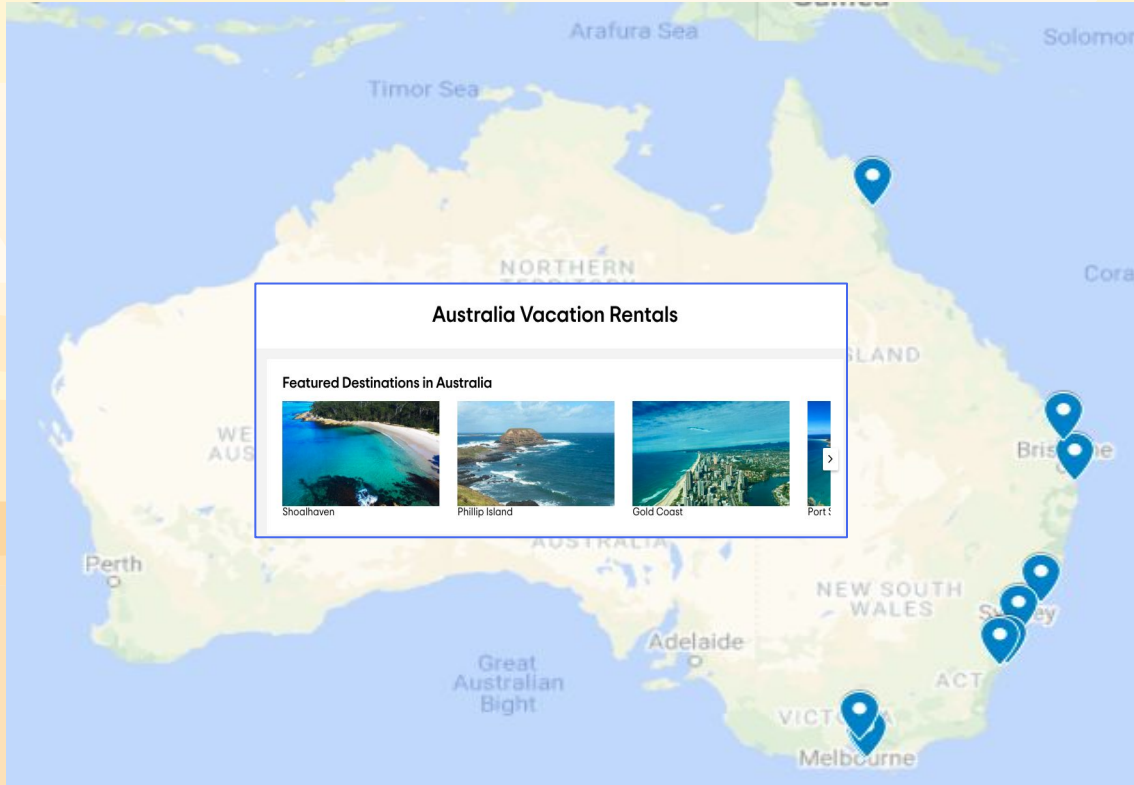
Zoe Liao  
12/15/21

# Goals



- Tripadvisor is a popular site for hotel booking
- Can overall Tripadvisor review score for a hotel be predicted by certain features of the hotel?
- If so, which features are most important?

# Data



## "Featured Destination" locations:

- Shoalhaven
- Gold Coast
- Phillip Island
- Melbourne
- Cowes
- Sydney
- Noosa
- Surfer's Paradise
- Port Douglas
- Port Stephens

1309 hotels

# Data point example: 1 hotel

## Pleasant Way River Lodge



- Tripadvisor Rating: 4.0
- Location: Shoalhaven, New South Wales 2541 Australia
- 419 ratings
- \$88/night



53

**Somewhat walkable**  
Grade: 53 out of 100

54

**Restaurants**  
within 0.75 miles

9

**Attractions**  
within 0.75 miles

### Property amenities

- Free parking
- Free High Speed Internet (WiFi)
- Pool
- Breakfast in the room
- BBQ facilities
- Baggage storage
- Laundry service
- Self-serve laundry

Show more

### Room features

- Soundproof rooms
- Air conditioning
- Iron
- Kitchenette
- Microwave
- Refrigerator
- Flatscreen TV

### Room types

- Non-smoking rooms

### Good to know

HOTEL CLASS



LANGUAGES SPOKEN

English

HOTEL STYLE

Charming

# Process



## What makes a hotel highly rated?

- Attractions/Mile
- Hotel Style: "Budget", nice views
  - Drop "Centrally Located"

### Baseline results:

- Train  $R^2$ : 0.331
- Val  $R^2$ : 0.231
- MAE: 0.371

## Feature Expansion

- Covid precautions
- Price per night
- Top 15 property amenities + "Other"
  - Drop "Non smoking hotel"
- Top 15 room features + "Other"
  - Drop "Complimentary toiletries"

## Tried but dropped

- Restaurants/Mile
- Hotel Star class
- Number of reviews

More after regularization...

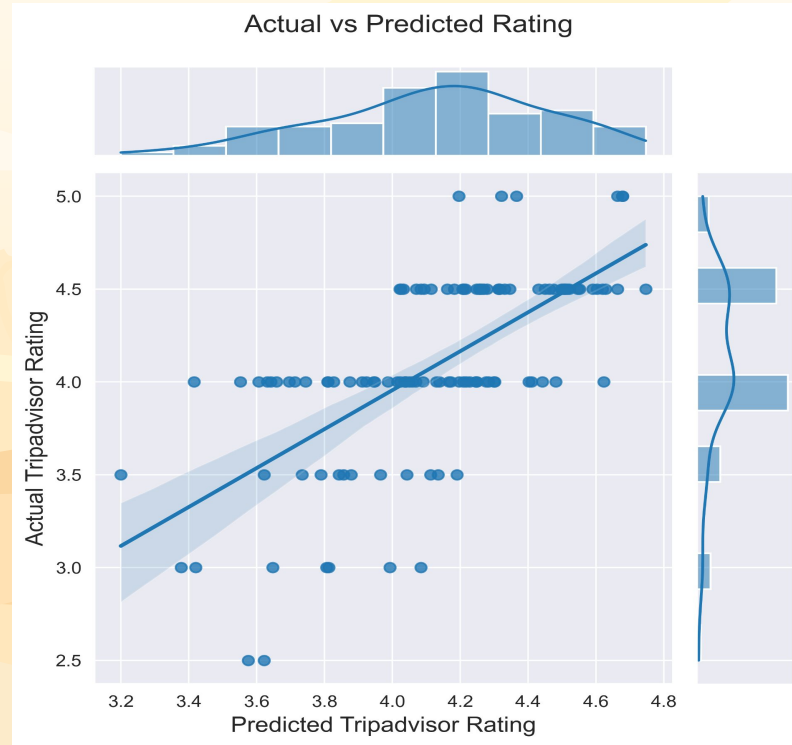
# Final Model Performance

## Elastic Net Model

- L1 ratio: 0.01
- Alpha: 0.1

## 5-fold cross-validation

- Mean  $R^2$ : 0.365
- Mean NMAE: -0.316



# Results: Feature Interpretation

**Y =**

**Intercept:  
3.634**

+

| Features                    | Coefficient |
|-----------------------------|-------------|
| Featured Price (percentile) | 0.678       |
| Attractions/Mi (percentile) | -0.123      |

+

| Features          | Coefficient |
|-------------------|-------------|
| Budget            | -0.311      |
| Covid Precautions | 0.257       |
| Family Resort     | -0.225      |
| Modern            | 0.202       |
| Charming          | 0.198       |
| Free Wi-Fi        | 0.193       |
| BBQ facilities    | 0.102       |
| Bath/Shower       | -0.082      |

If numerical features = 0 and:  
"non-smoking" &  
"complimentary toiletries"  
& "centrally located"

Median:

\$156/night  
30 attractions/mile

# ***Conclusion***

- Linear regression may not be best approach
- Training data had limited low rating scores

## **Future Work:**

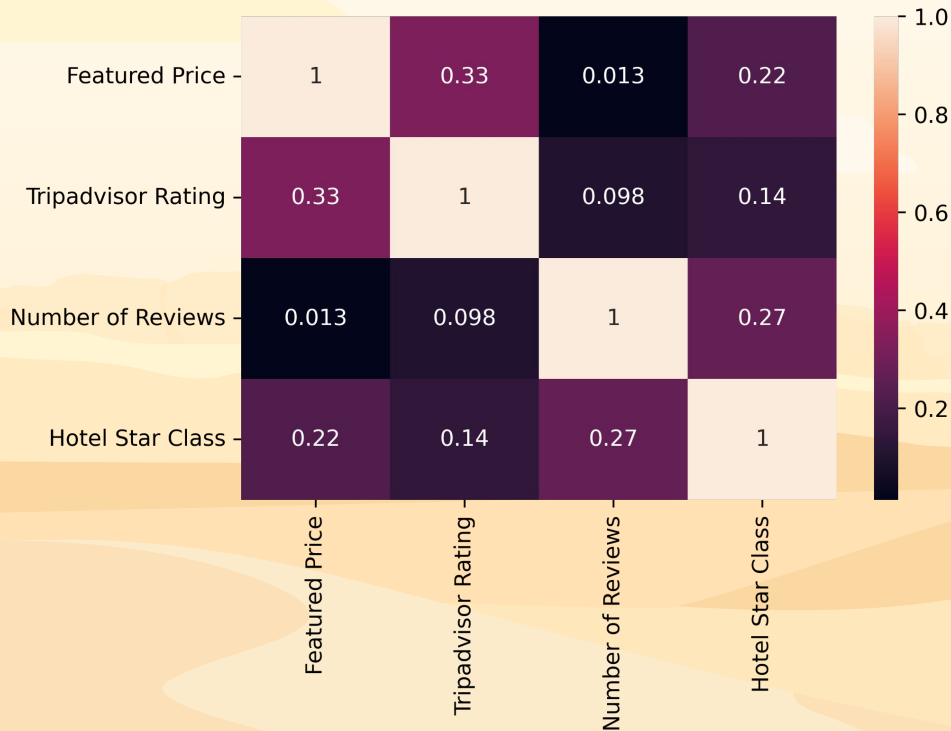
- More data (down to 571 hotels in final model)
- Did not use location or hotel type
  - Limited pricing data, which had strongest correlation among features
  - Could use location/zip code to help impute price
- Other regression approaches



# ***Appendix***



# Baseline: numerical features

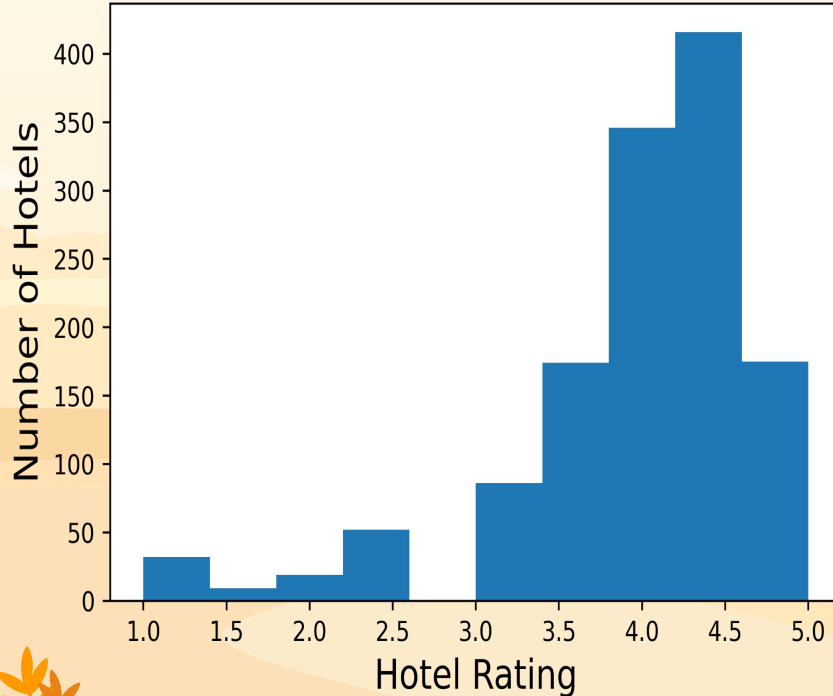


Baseline with numerical features:

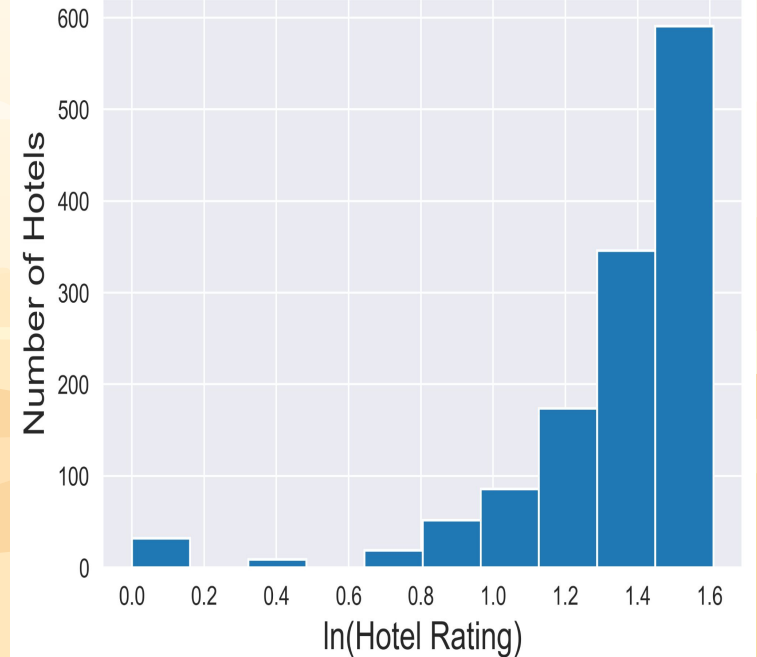
OLS using LR-imputed price  
Train: 0.098  
Val: 0.151

# ***Transforming target using natural log***

Distribution of Tripadvisor Ratings



Histogram of  $\ln(\text{Tripadvisor Rating})$



# Example of Covid-19 Precautions



## Travel safe during COVID-19 ⓘ

### What you can expect during your stay

- All linens sanitized in high-temperature wash
- Floors marked for social distancing
- Hand sanitizer available to guests & staff
- Stay-at-home policy for staff with symptoms
- Regularly sanitized high-traffic areas
- Staff required to regularly wash hands

Read more ▾

Updated 06/15/20 Safety precautions are provided by the property and are not checked by Tripadvisor for accuracy.

### A note from Quest Nowra

The Quest Stay Safe Promise is our ongoing commitment to the health and safety of all guests.

Read more ▾

### More questions?

☎ 1800 334 033

✉ Email hotel

### Community

💬 Ask a question

✍ Write a review

## ***Regularization: Cutting down from 52 features***

### ***Lasso***

**$R^2$ : 0.375; NMAE: -0.305**

Alpha = 0.01

### ***Ridge***

**$R^2$ : 0.390; NMAE: -0.308**

Alpha = 6.22

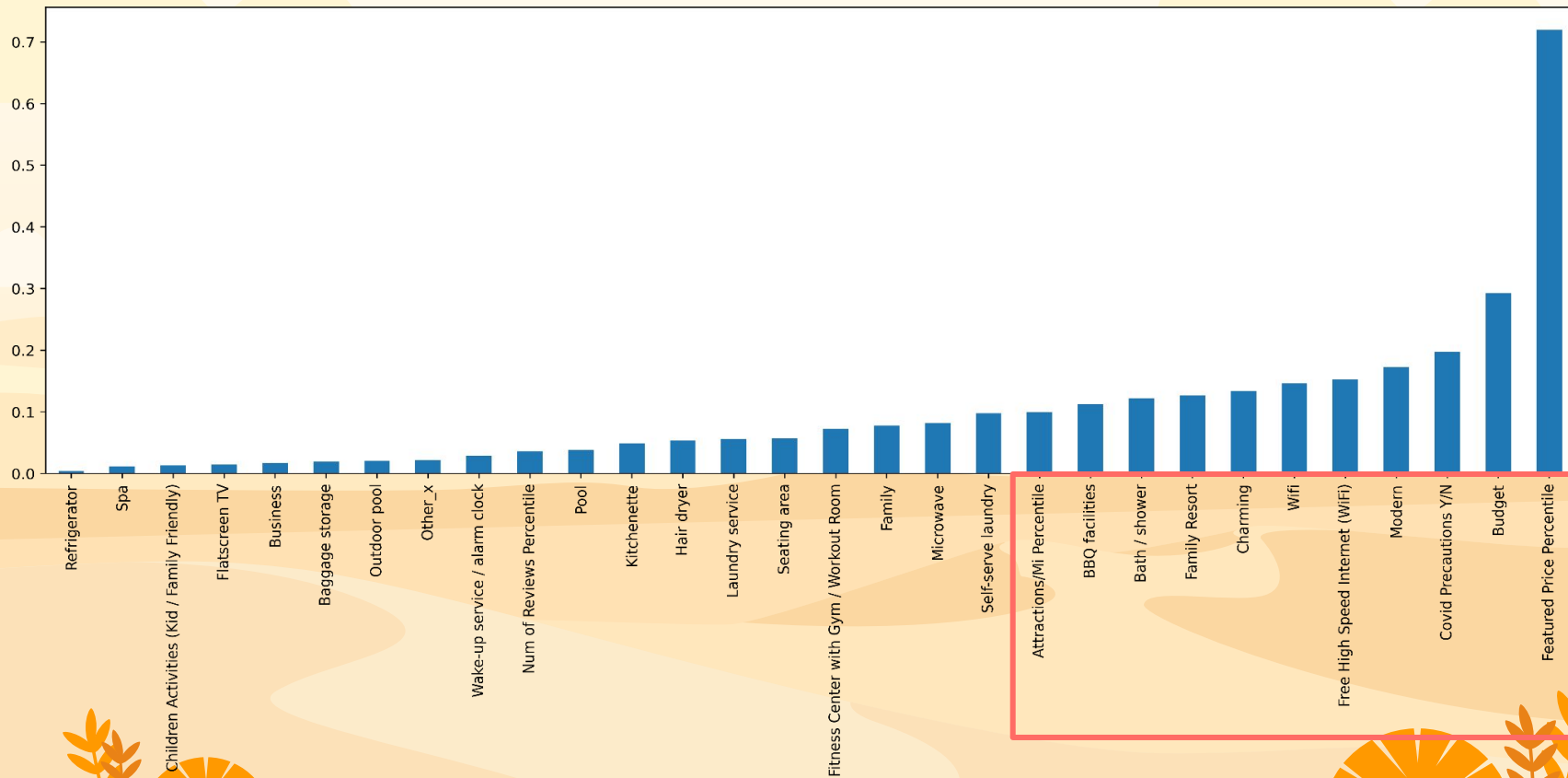
### ***Elastic Net***

**$R^2$ : 0.397; NMAE: -0.305**

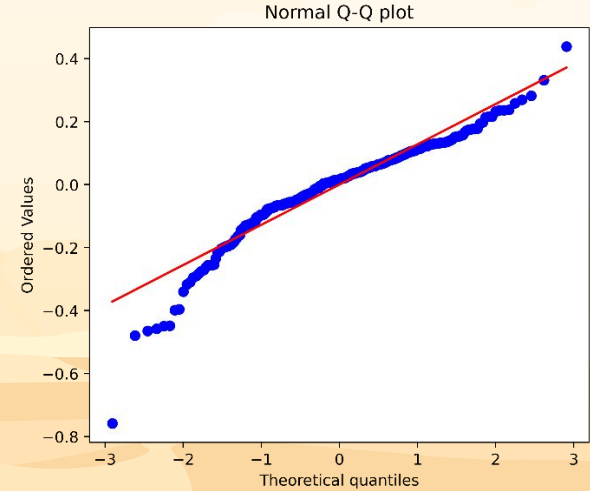
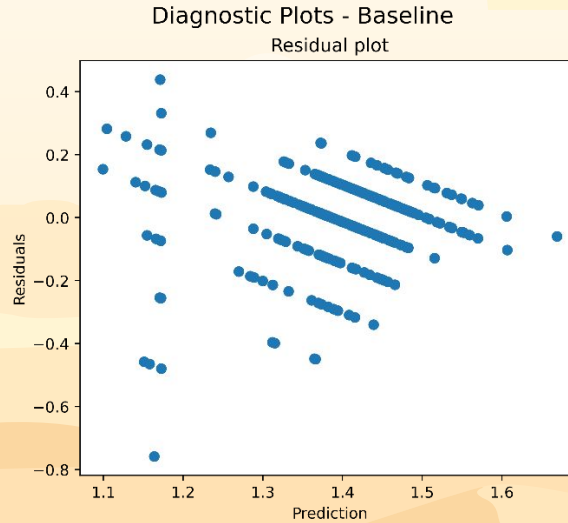
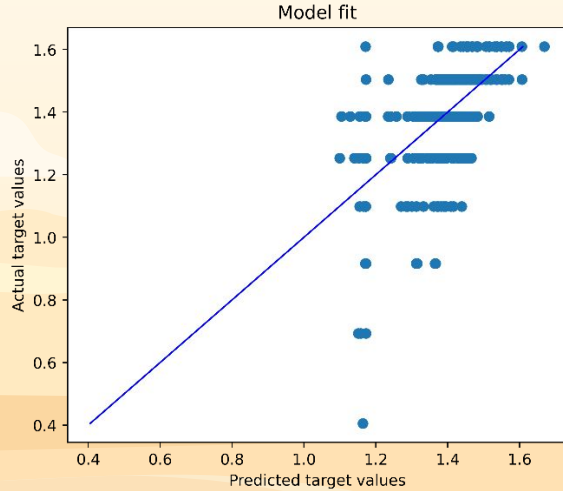
Alpha = 0.01; L1 = 0.5

# Trimming down features

Coefficient size of features in Elastic Net Model



# ***Diagnostic plot - Baseline Model***



# ***Residual distribution of final model***

