



Highlighting Hidden Gems

Breaking down how to get a better TripAdvisor ranking for Singaporean local hawker stalls/centres and restaurants

by Amelia Lui



Table of contents



01 Goal

02 Why?

03 Process

04 Results





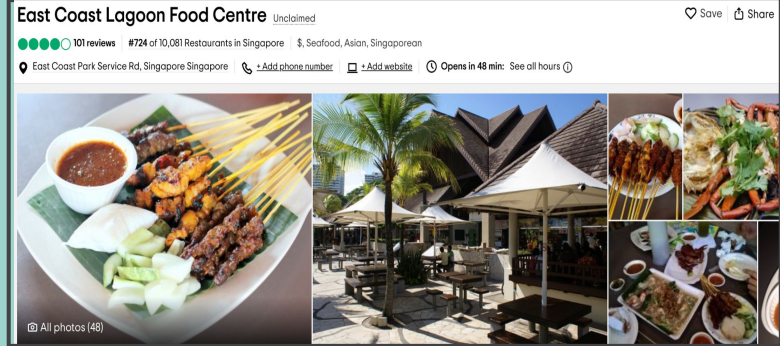
01 Goal

Turning this:

- Length of restaurant name
- Average price range (how many \$)
- Number of previous reviews
- Trip advisor ranking
- Number of photos
- Michelin?
- Travellers choice award?
- Reservation available on tripadvisor?
- Website claimed by owners?
- zipcode



Into this:



Predicted TARanking: 731
True TARanking: 724

What's for lunch?



02 Whys

Why Singapore?

- Gentrification of local foods
- Hawker stalls and small restaurant businesses are oftentimes run single family units
- Social media perhaps not accessible





Why TripAdvisor?

“Founded in 2000, Tripadvisor has grown to become **one of the most popular travel destination and accommodation websites**... As of 2020, Tripadvisor’s total number of user reviews and opinions reached **884 million**, covering listings for restaurants, hotels, vacation rentals, and attractions.”

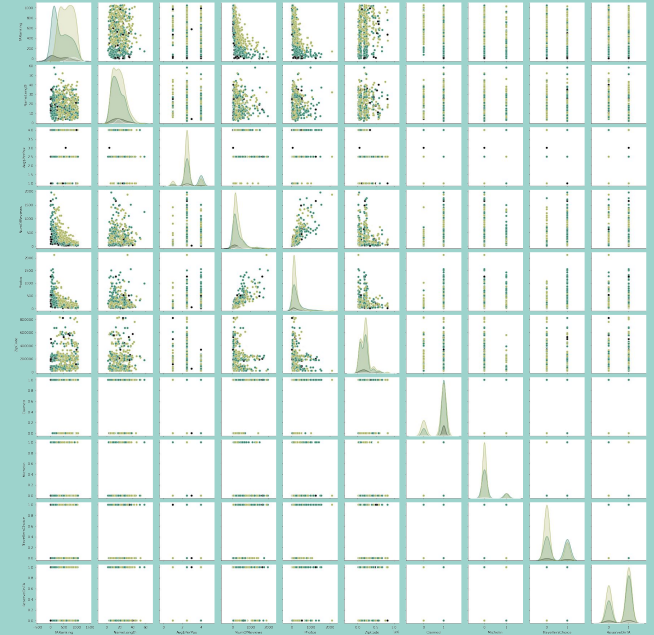
- statista

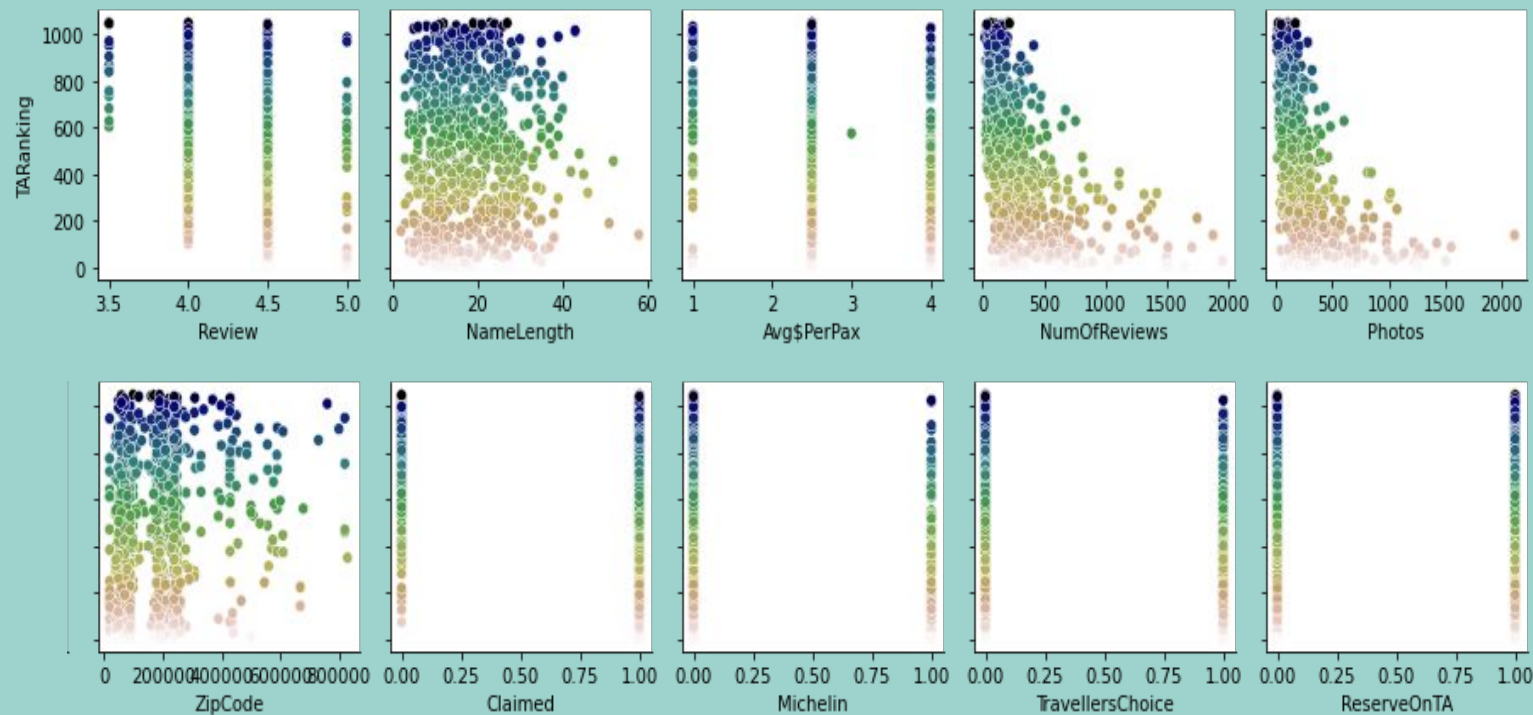


03 The Process

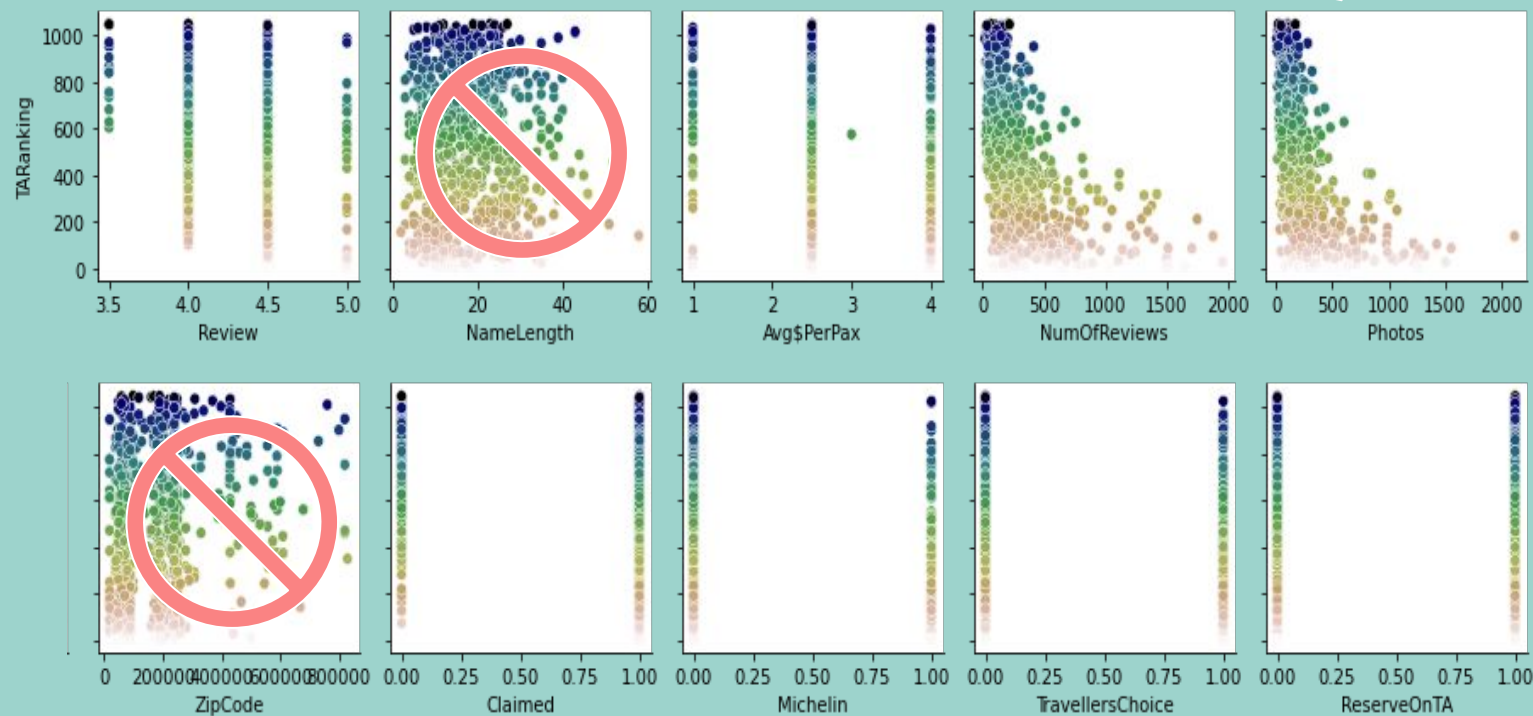
EDA

- Data cleaning: impossible data/outliers, drop rows with null data
- Split into train/test 70/30
- pairplot





heteroskedasticity



First Model Fit

- $R^2 = 0.564$
- Adjusted $R^2 = 0.557$

Plan:

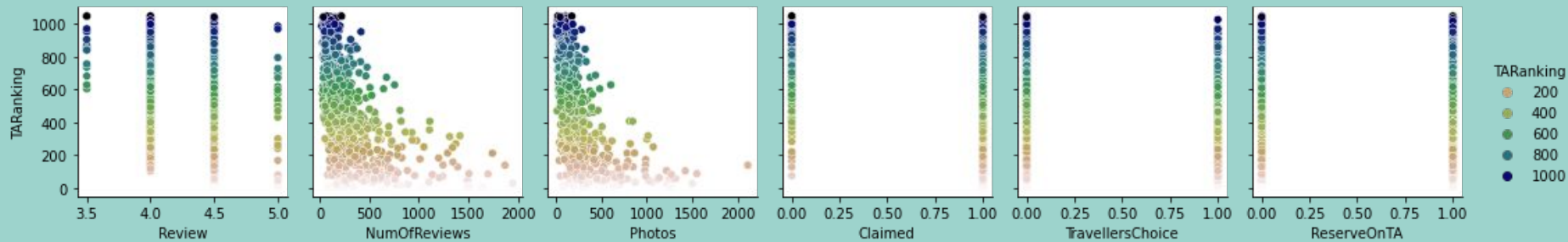
Removed features with
p-values > 0.05

Second Model Fit

- $R^2 = 0.561$
- Adjusted $R^2 = 0.557$

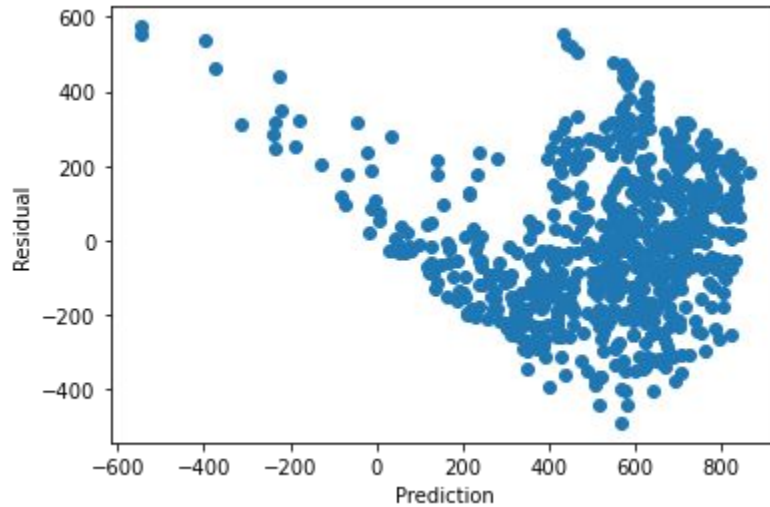
Plan:

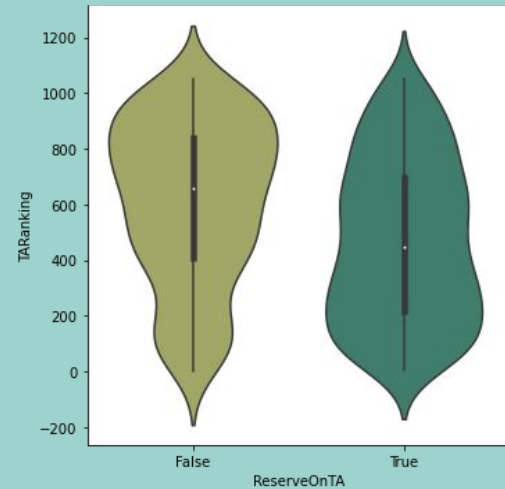
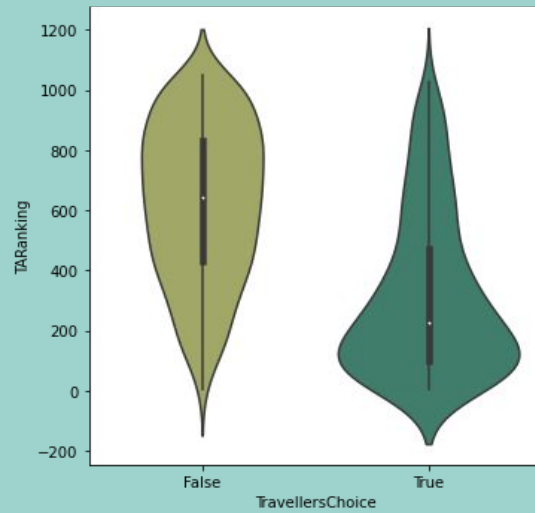
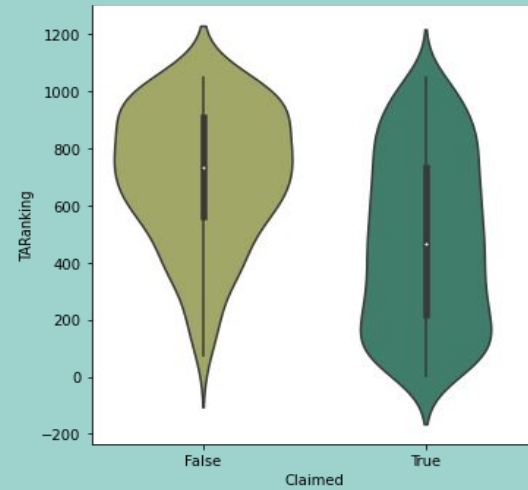
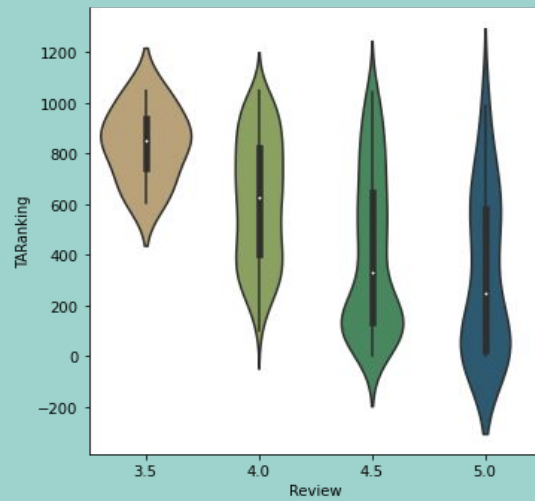
Want to check whether
heteroskedasticity and/or
categorical features are issues



What is going on here?

Residuals





Third Model Fit

(Took log of numerical features)

- $R^2 = 0.755$
- Adjusted $R^2 = 0.753$

Next:

Multicollinearity? Take -X

Fourth Model Fit

- $R^2 = 0.752$
- Adjusted $R^2 = 0.750$

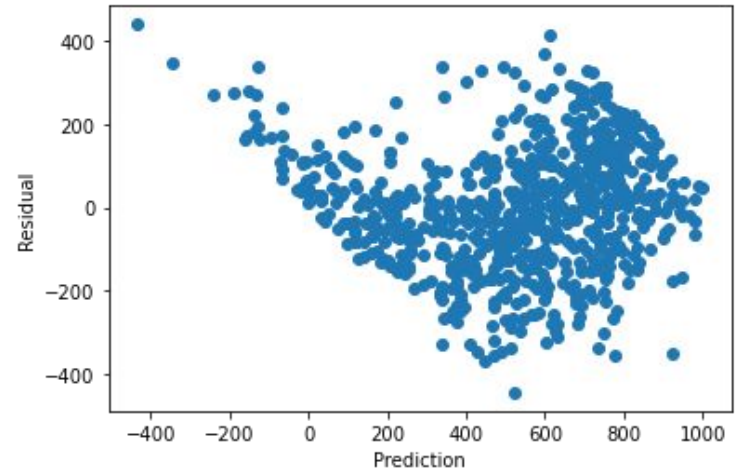
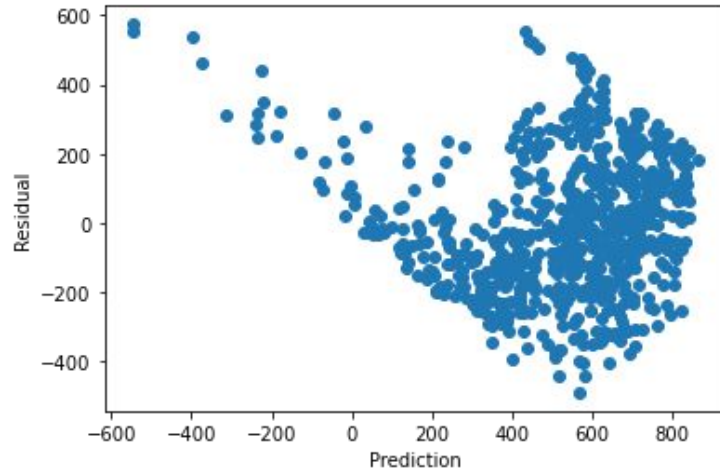
Done, chose to sacrifice higher R^2 for risk of overfitting





04 Results!

Residuals Before and After





East Coast Lagoon Food Centre

Review: 4/5

Number of Reviews: 101

Claimed: False

Travellers' Choice: False

Reservation Avail: False

Predicted: 731

True: 724



Amaravati

Review: 4/5

Number of Reviews: 101

Claimed: True

Travellers' Choice: False

Reservation Avail: False

Predicted: 658

True: 669





Next Steps

- Scrape other food review sites
- Wordcloud of popular words correlated to high-performing restaurants
- Income statements?
- Moovit - transportation

References

- “TripAdvisor,” tripadvisor.com.sg
- <https://www.statista.com/topics/3443/tripadvisor/#dossierKeyfigures>

Used Slidesgo