Abstract

In recent years, hawker stalls in Singapore have had to compete with restaurants promising to deliver the same local foods in a more manufactured experience. While such restaurants may have a whole staff, investors, and social media managers, hawker stalls are often only run by a single family, who do not have the funds nor experience to compete with the gentrification of local foods. To help with this issue, certain features have been identified to help hawker stalls establish an online presence by raising their Trip Advisor ranking. Correlation has been identified between the following features and a higher ranking — overall review, number of reviews, number of photos, whether or not the restaurant is claimed, whether or not the restaurant has a Travellers' Choice award, and whether or not a reservation link is available on the restaurant's Trip Advisor page.

Design

Singapore is known for being a multicultural hub — its people are full of appreciation for each others' cultural practices, religions, and every year, tourists flood in to visit the country for all our peoples have to offer, especially through tasting the many local foods. However, while tourism benefits the city's economy and many citizens' lives, gentrification has become a massive problem. Tourists who don't know any better may choose to visit inauthentic and pricey restaurants, promising to serve the local food with "gastronomic experiences". While this is fair and all (and such restaurants do do what they do well!), local food stalls may end up struggling to compete with these better established restaurants with carefully manufactured online presences. Locals who grew up eating from hawker centres may see food prices rising, and many Singaporean elders end up working way past a healthy retirement age to make up for the rise in living cost. In what ways can a local Singaporean food stall/small restaurant owner raise their Trip Advisor Ranking (hence, establishing a better online presence)?

Data

There were a little over 1000 restaurants' Trip Advisors pages scraped for this project. The target in question was the ranking of the restaurant, and the features examined were the overall review, the length of the restaurant name on the website, the average price per pax, the number of reviews, the number of photos, the zipcode, whether or not the restaurant was claimed, whether or not the restaurant had a Michelin star, whether or not the restaurant had a Travellers' Choice award, and whether or not the restaurant was reservable on their Trip Advisor webpage.

Algorithms

Data Cleaning/EDA:

The data was cleaned so that outliers were eliminated, as well as some impossible datas, such as a 6 star review or a 7 digit zip code.

Feature Engineering:

1. Noticed heteroskedasticity in number of reviews and number of photos, so took log of those features
2. Converted categorical features into binary dummy variables
3. Removed number of photos because of feature collinearity

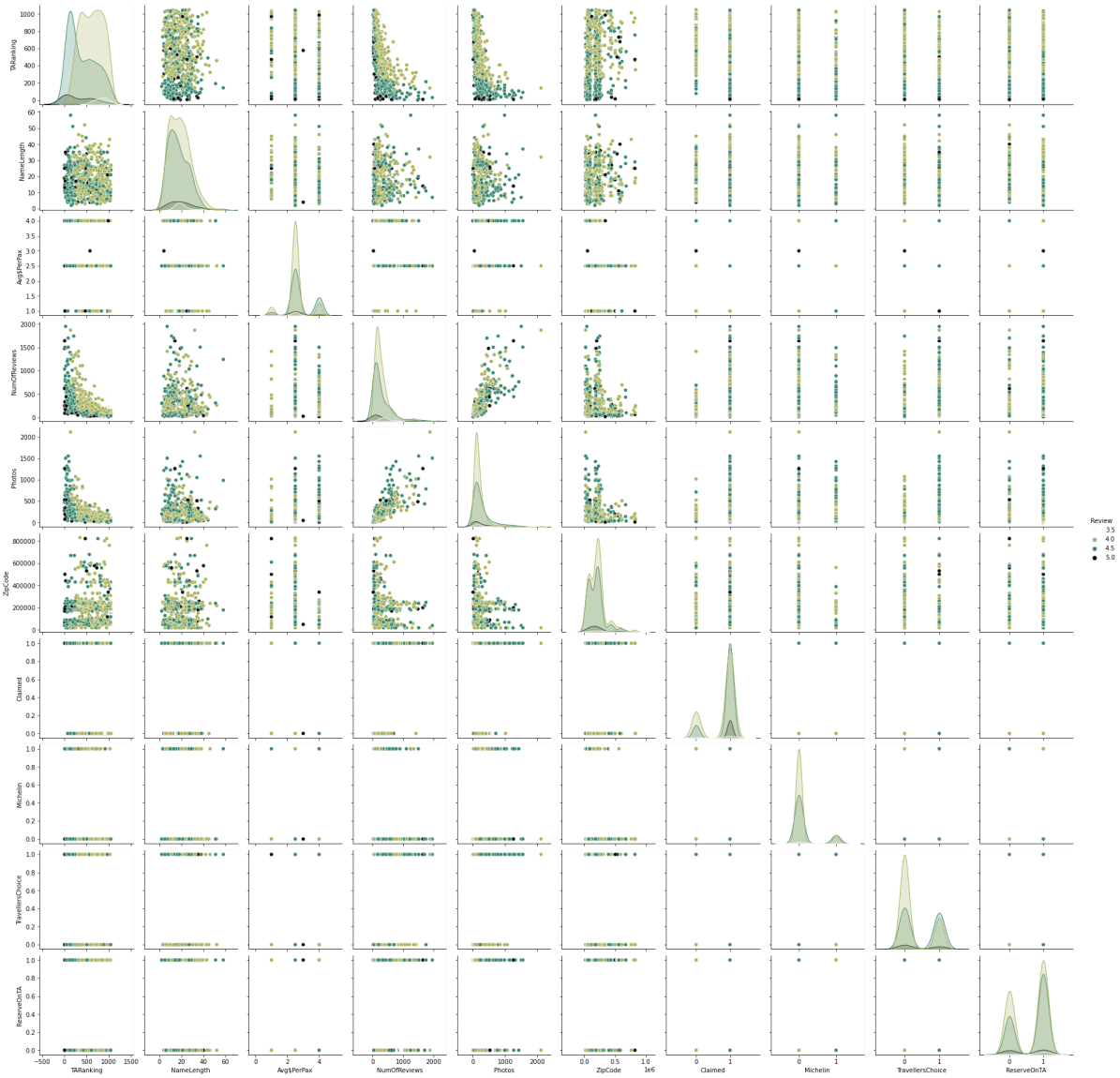
Model: Linear regression, took log of numerical features

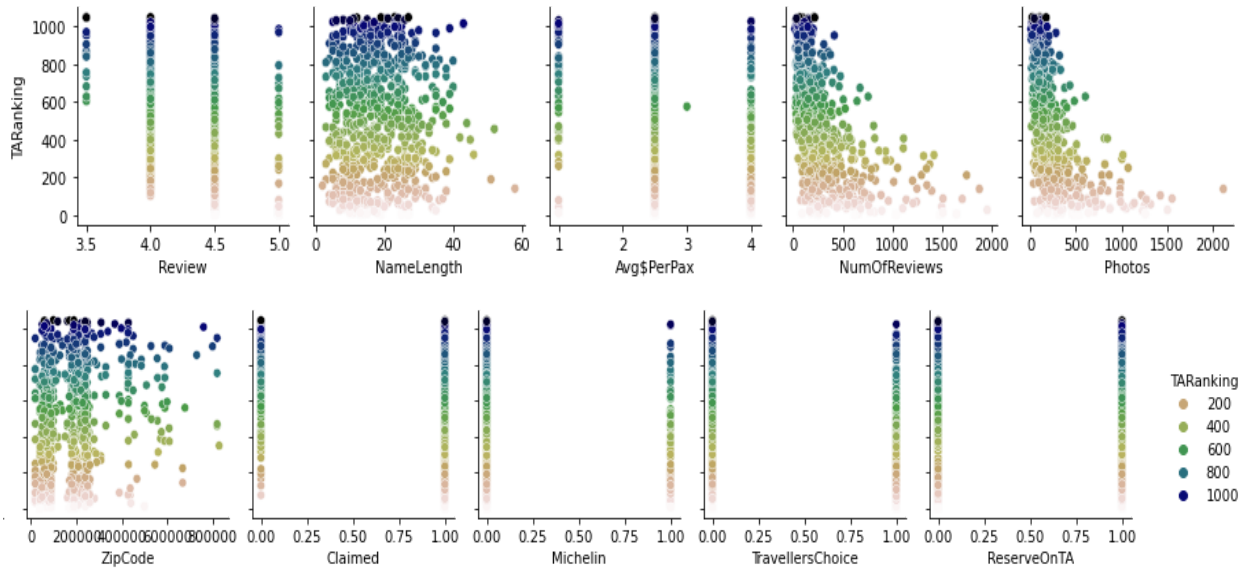
Model Evaluation and Selection: The dataset was split into a 70/30 training set and testing set. Removed features with a p-value of more than 0.05

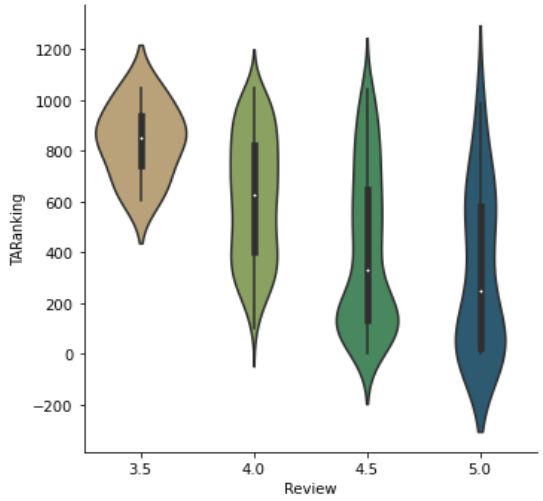
Tools

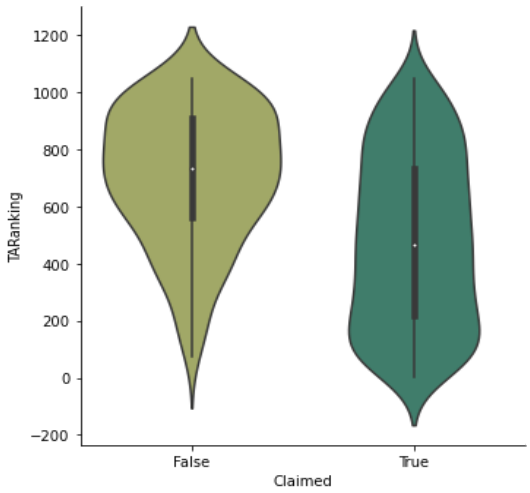
Numpy, Pandas, Matplotlib, Seaborn, Selenium, BeautifulSoup, Statsmodels

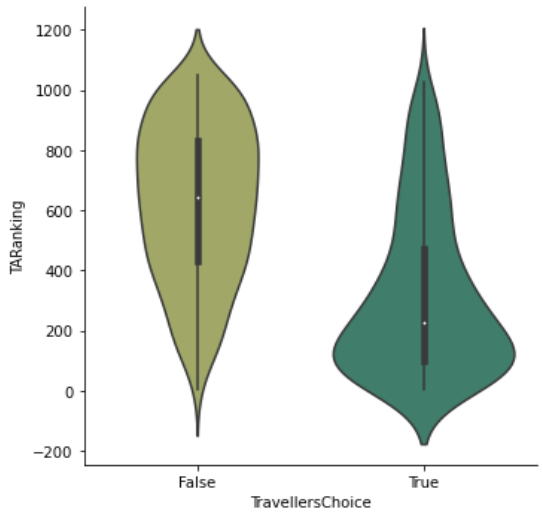
Communication

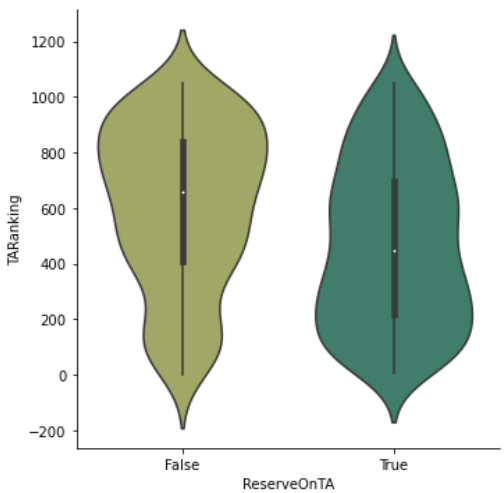

TARanking
NameLength
Avg$PerPax
NumOfReviews
Photos
ZipCode
Claimed
Michelin
TravellersChoice
ReserveOnTA
1e6
Review
3.5
4.0
4.5
5.0

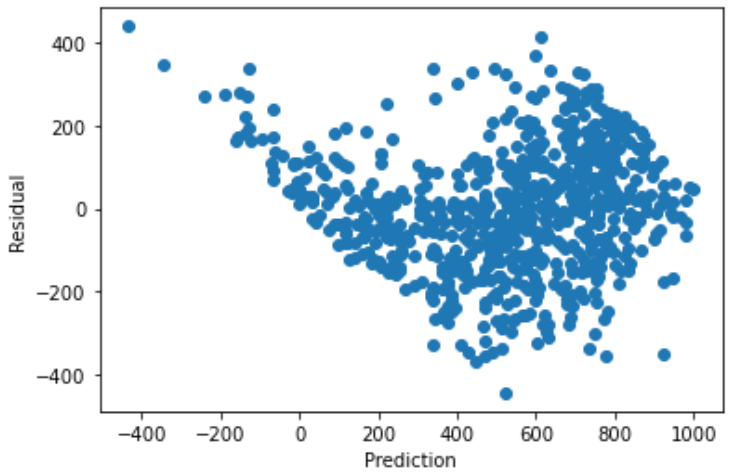

TARanking
Review
NameLength
Avg$PerPax
NumOfReviews
Photos
ZipCode
Claimed
Michelin
TravellersChoice
ReserveOnTA
TARanking
200
400
600
800
1000


TARanking
Review


TARanking
False
True
Claimed


TARanking
False
True
TravellersChoice


TARanking
False
True
ReserveOnTA


400
200
0
−200
−400
Residual
−400
−200
0
200
400
600
800
1000
Prediction