Improving Success Rates of Restaurants

A Study in Data Science by Juichia Holland

15,000,000 restaurants

around the world are serving food to customers!

Failed restaurants are costly to the restaurateurs, to their financial sponsors, to the industry and to the diners. This study in data science aims to help restaurateurs to incorporate more success factors into their strategies when opening new restaurants.

1. The Data

In order to make inferences about restaurants in general, a decent sample of restaurant data is needed for analysis. Yelp provides just the dataset needed at https://www.yelp.com/dataset.

➔ **Businesses**

192,609 businesses with over 1.2 million attributes

➔ **Check Ins**

Check ins over time for each of the 192,609 businesses

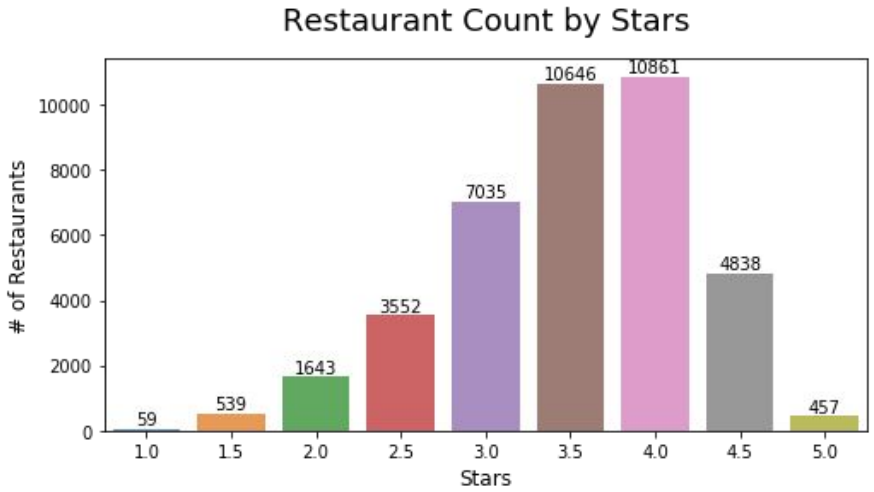
➔ **Reviews**

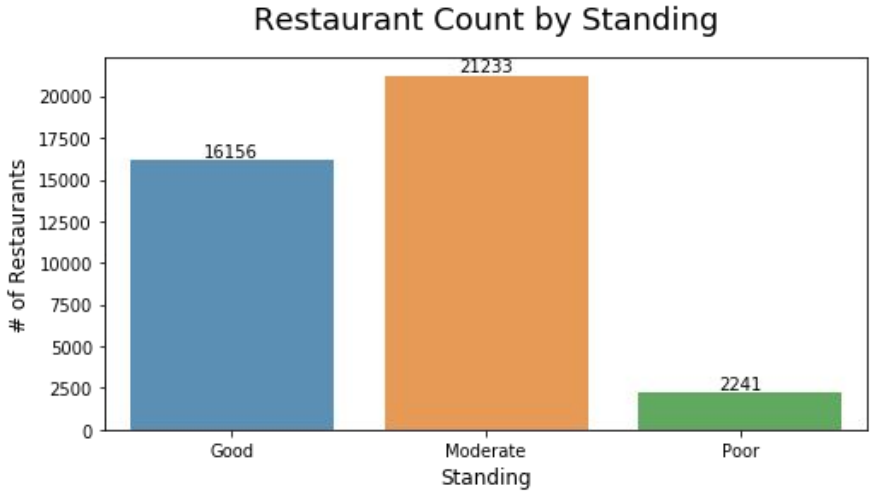
6 million reviews over time for the 192,609 businesses

How difficult is it to have a rating of 5 stars on yelp?

Not as difficult as having a 1 star rating! Relax.

There are almost 10 times more restaurants with 5 star ratings than those with 1 star ratings.





Yelp restaurants average 3.5 stars in rating. While higher ratings do not guarantee business success, low ratings increase the chance of failure. Therefore, it is important to find out what contributes to higher ratings.

Good restaurants experience higher consistency in ratings over time, so starting on the right foot matters.

Restaurant category affects restaurant rating. Location, features and price don't matter as much.

2. The Analysis

Initial findings through exploratory data analysis include:

➔ **Restaurant Category**

Choosing the right type of restaurant to open matters! Whether it is because of trends in food culture or something else, some categories have more restaurants with better ratings.

➔ **Ratings Volatility**

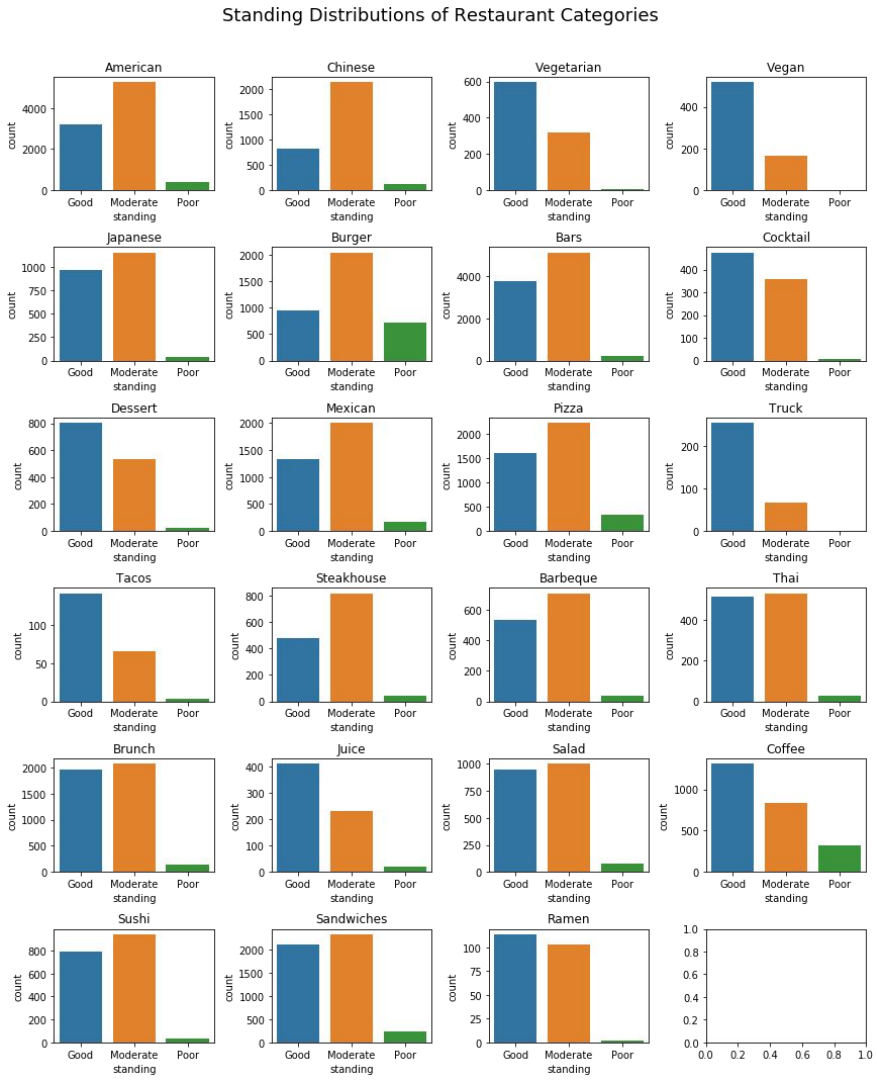
Starting on the right foot matters! Good restaurants experience higher consistency in ratings over time than moderate and poor restaurants.

➔ **Features and Price**

Price does not affect ratings. No restaurant feature was found significant in relationship to ratings.

➔ **Location and Density**

The location of the restaurant makes no difference on the ratings, and neither does restaurant density in the surrounding area.

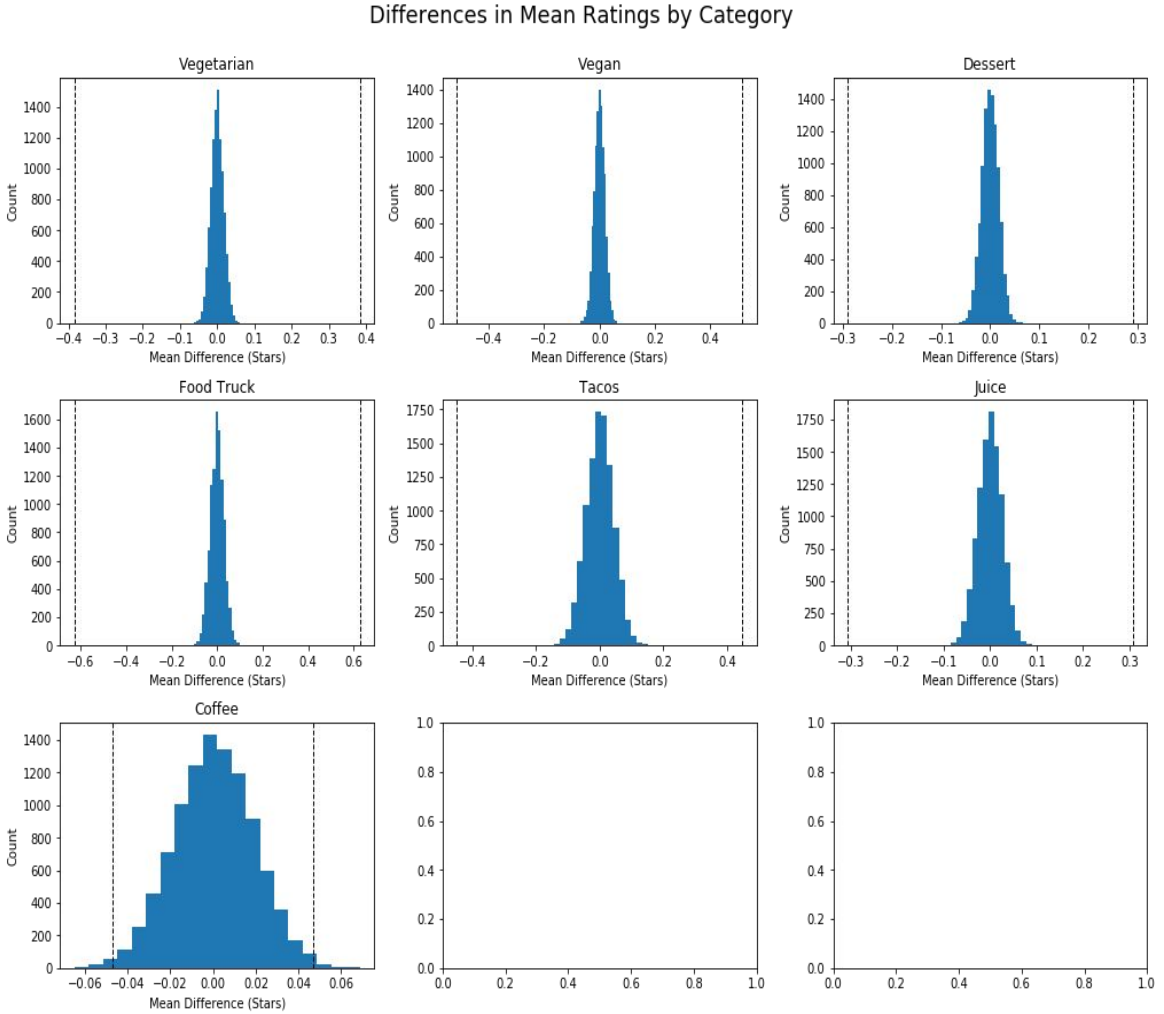


Distribution by Category

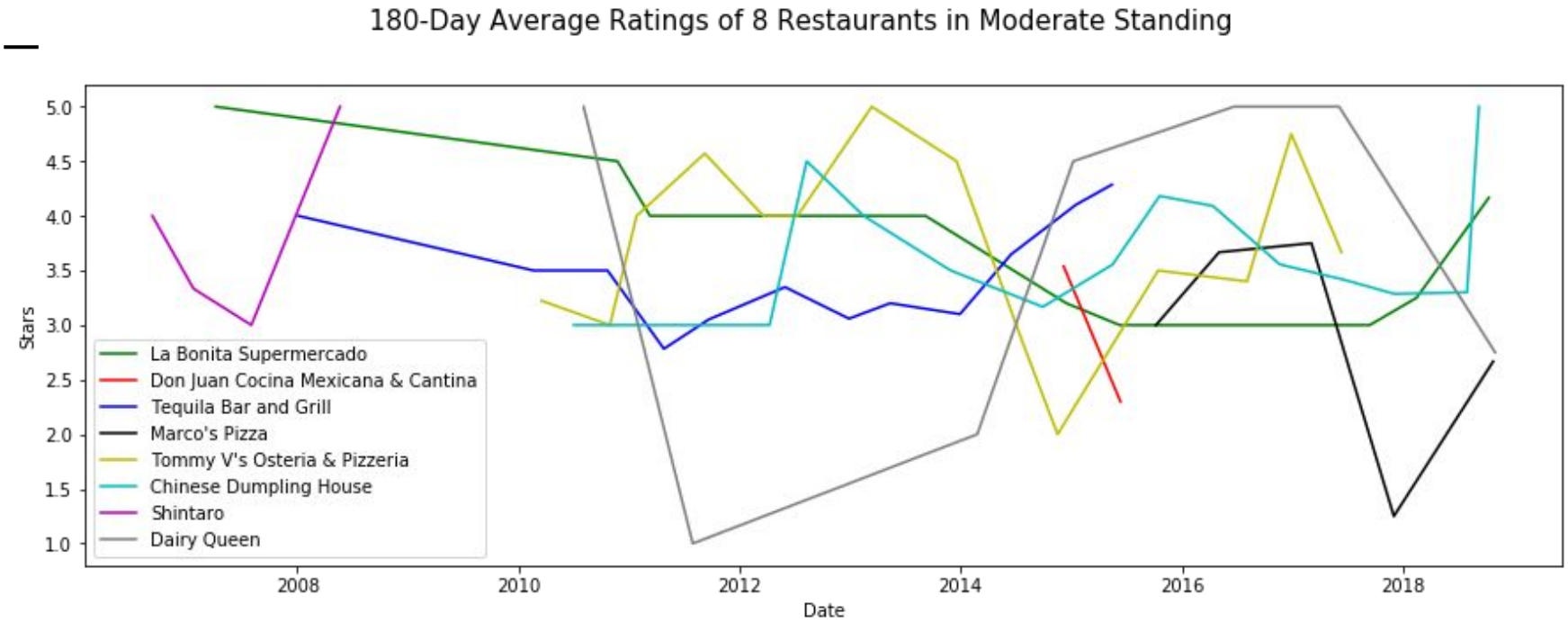
Restaurants of type Vegetarian, Vegan, Dessert, Food Truck, Tacos, Juice, and Coffee have noticeably more restaurants in good standing.

Can ratings be improved by changing restaurant category?

Yes. Using bootstrap inference for the hypothesis test, a p-value of 0 suggests that there is significant difference in ratings depending on restaurant category.



Changing restaurant category changes restaurant rating.

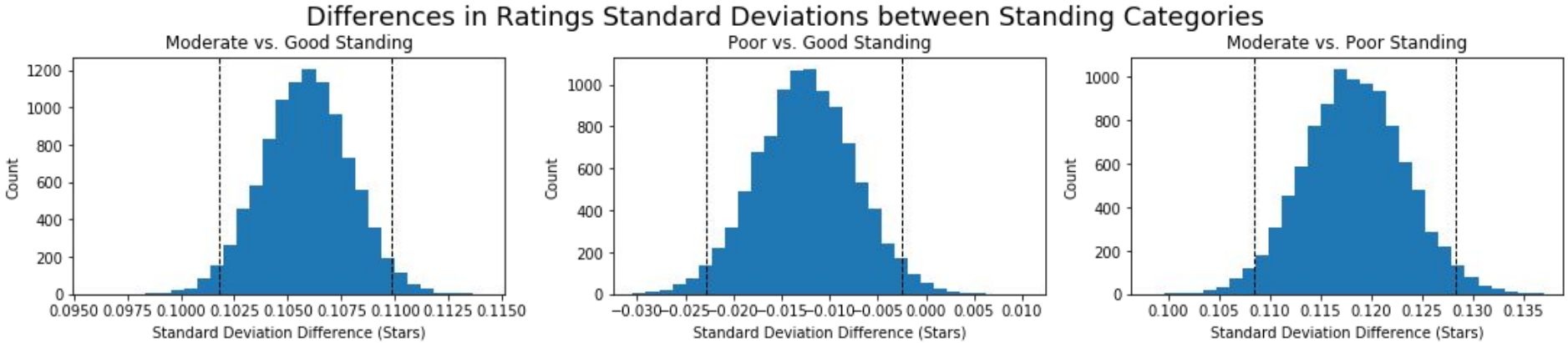


Ratings Standard Deviations

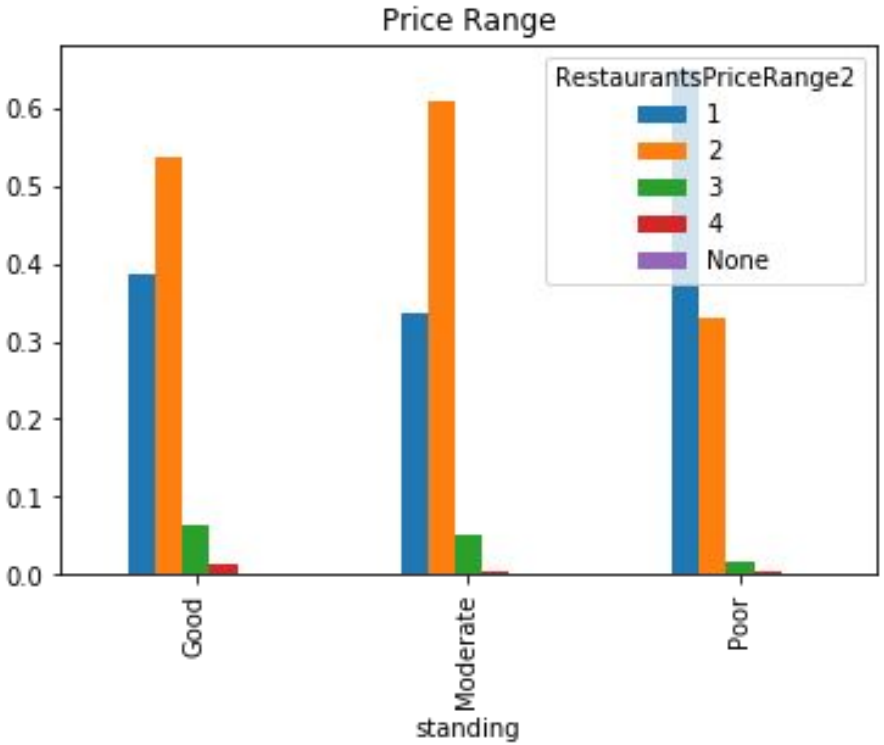
Between the standing categories, moderate restaurants have the largest standard deviation in ratings.

Does restaurant standing make a difference in ratings volatility?

Yes. Using bootstrap inference for the hypothesis test, a p-value of 0 suggests that there is significant difference in ratings standard deviations depending on restaurant standing.




Restaurants in moderate standing experience significantly higher volatility in ratings.
CHARLOTTE BRONTË
JANE EYRE

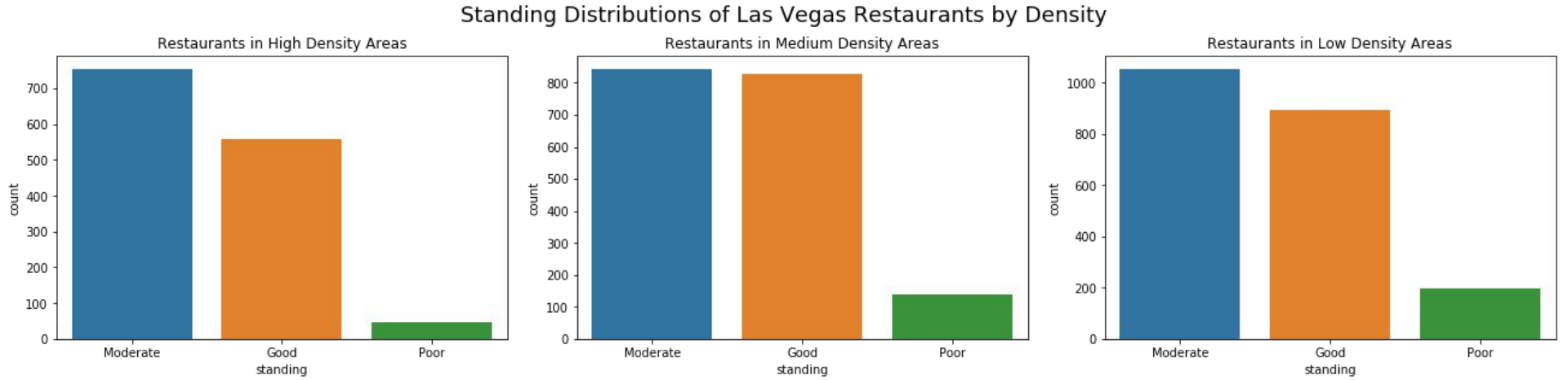


Price and other features make no difference on ratings.

A restaurant is equally likely to be poor, moderate or good regardless of price range. The same is found in other features such as alcohol, reservations and outdoor seating.

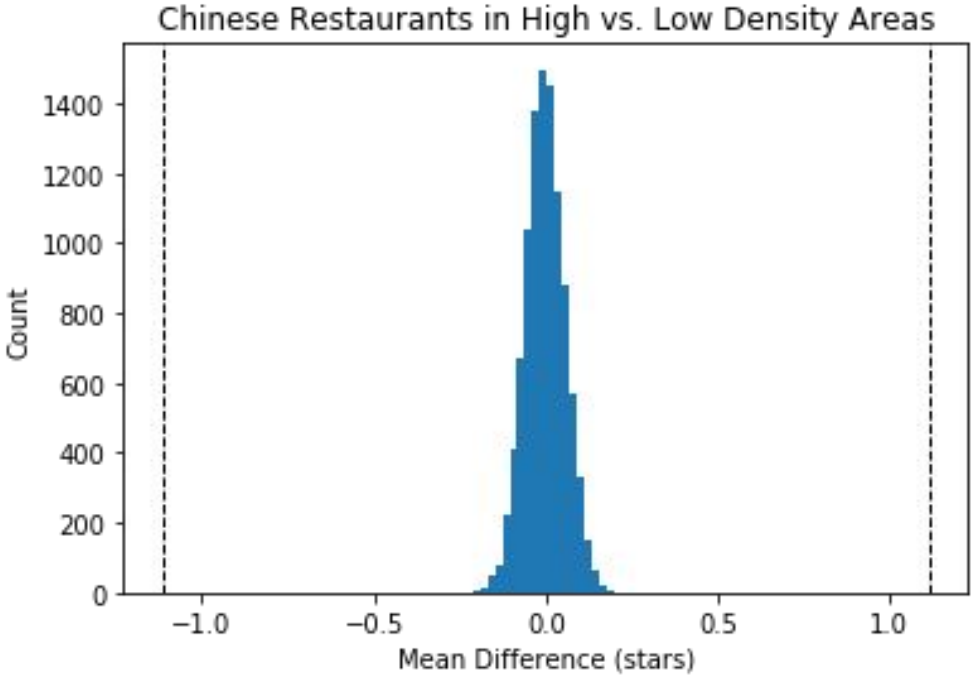
Location and Density

Restaurants in Las Vegas show similar standing distributions regardless of restaurant density. Restaurant density around the location of a restaurant makes no difference on restaurant ratings.



Can ratings for a Chinese restaurant be improved by moving towards or away from Toronto's Chinatown?

No. Using bootstrap inference for the hypothesis test, a p-value of 0.3 suggests that there is no significant difference in ratings depending on restaurant density around a Chinese restaurant.




Price, features, location, and density make no difference on ratings.

The Yelp restaurants dataset contains 7690 restaurants that have closed down permanently. While open restaurants in the sample may not represent success, finding the **important predictors for restaurant closure** may help reduce the risk of business failure.

The average change in stars over time is an important predictor for restaurant closure, as is **restaurant density**. Restaurant category and ratings volatility are less important predictors in restaurant closure.

3. In-Depth Analysis

Findings using supervised learning to build predictive models include:

➔ **Restaurant Category**

Restaurant type is not an important predictor for restaurant closure.

➔ **Ratings Volatility**

Consistency in rating is not an important predictor for restaurant closure.

➔ **Restaurant Density**

Closed restaurants have a higher average in restaurant count within one square kilometer of their locations than restaurants that are open.

➔ **Change in Stars**

Closed restaurants experience on average 3 times more of a drop in stars over time than restaurants that are open.

Feature Importances

Using supervised learning with a decision tree algorithm for predictive modeling, the features were reduced from 7 to 5, and the resulting model provides an accuracy of 0.67 on training data and 0.65 on test data. The most important predictors for restaurant closure are the average change in stars over time and restaurant density.

Can the accuracy for predicting restaurant closure be improved with ensemble methods?

Yes. Using a voting algorithm that combined the decision tree, logistic regression and K nearest neighbors, an accuracy score of 0.7 was achieved on test data.


Closed restaurants experience on average 3 times more of a drop in ratings over time and are located in areas with higher restaurant density.

4. Summary

In this data science study, the Yelp restaurants dataset was used to uncover factors that may help improve the success rates of restaurants.

➔ Choose a Trending Category

Statistical data analysis shows that choosing the right type of restaurant to open matters!

➔ Monitor Customer Experience

Restaurants that have closed down experience on average 3 times more of a drop in ratings over time. Monitoring the change in customer experience periodically for early warning signs may help prevent restaurant closure.

➔ Start Off on the Right Foot

Statistical data analysis shows that good restaurants experience higher consistency in ratings.

➔ Avoid High Density Areas

Closed restaurants are located in areas with higher restaurant density. This could be due to those locations having a higher operational cost (rent, labor) and a variety of other reasons. Further analysis using rent and other data may reveal the underlying causes.