Predicting the success of Starbucks locations

By: Alexis raymond

AGENDA

INTRODUCTION

DATA

EXPLORATORY DATA ANALYSIS

PREDICTIVE MODELING

CONCLUSION

INTRODUCTION

Known for its taste, quality and customer experience, the Starbucks has expanded to over 25,000 locations worldwide in the past 50 years

This project aims to predict whether a Starbucks will be successful based on its location

Being able to predict whether a location will be successful would constitute a huge competitive advantage

Data sources

Sources

Starbucks Locations Worldwide

List of United States cities by population

Foursquare Places API

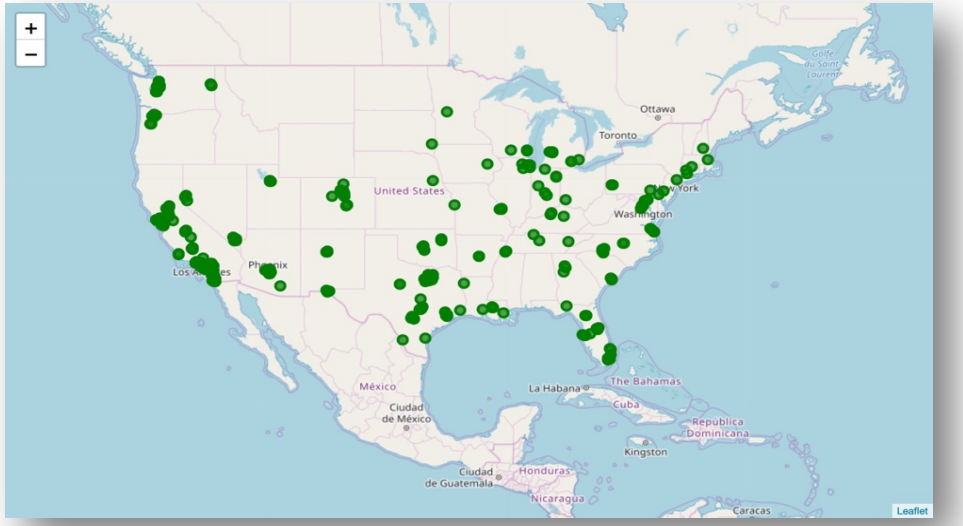
Limitations

500 quota for free version of Foursquare API

Rating only indicates customer satisfaction

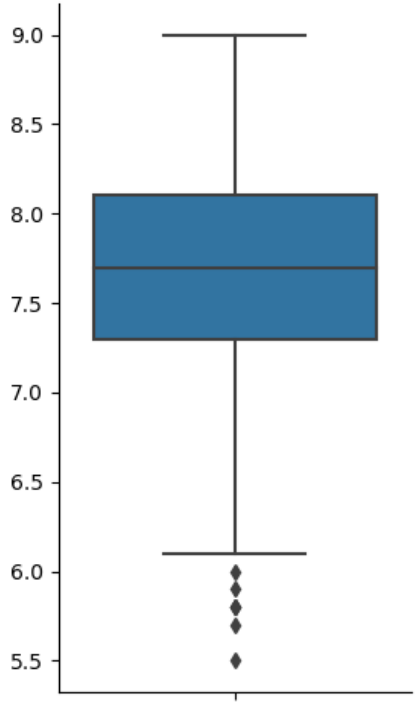
Data Cleaning

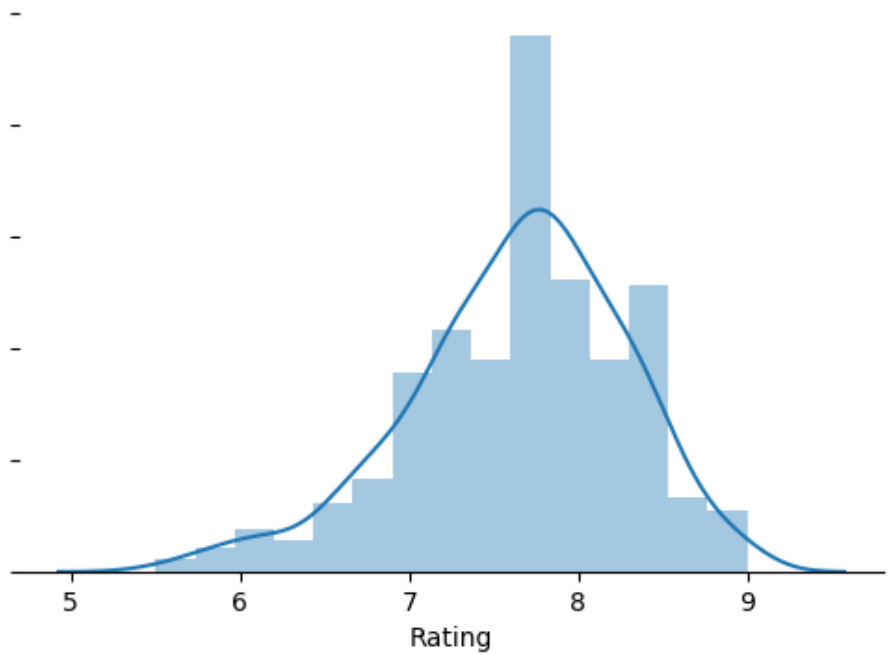
Step	Initial	Final	Removed	Reason
US cities	25,599	5,869	19,730	These locations were not in the identified US cities
Nearby venues	5,869	5,866	3	These locations did not have nearby venues in Foursquare
Foursquare ID	5,866	3,807	2 2,019	Request errors Starbucks not in Foursquare
Daily call quota	3,807	500	3,307	API daily call quota of 500 premium calls
Ratings	500	384	2 114	Rating not found Less than 10 ratings



Feature	Description
Area	Area of the city in which the Starbucks is located (in km^2)
Density	Population density of the city in which the Starbucks is located (in population per km^2)
Nearby Venues	Categories of venues located in a radius of 500 meters of the Starbucks
Distance to HQ	Distance between the Starbucks and the Starbucks head office in Seattle
Rating	Target variable: Measure of success of the Starbucks

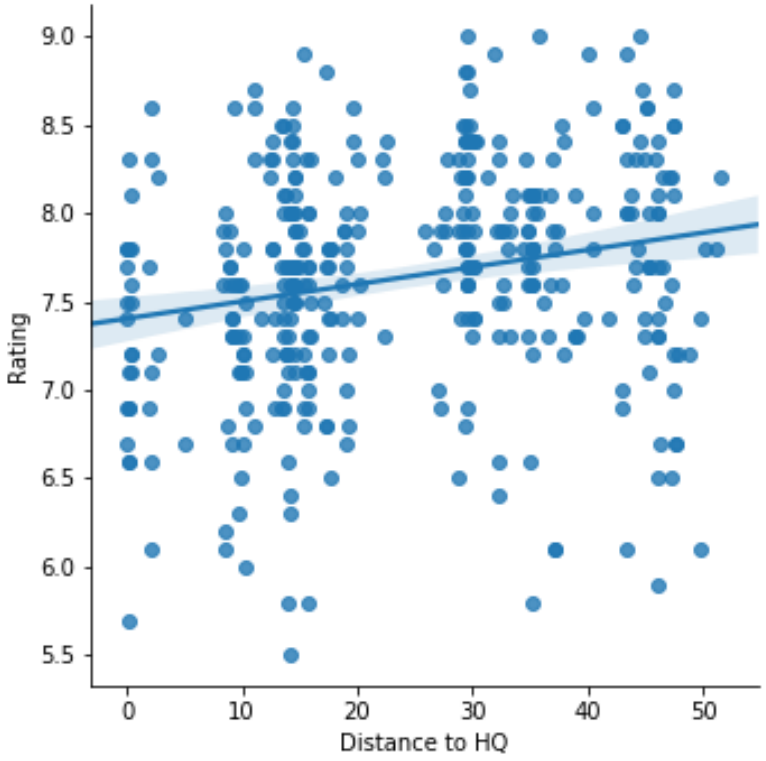
Feature Selection

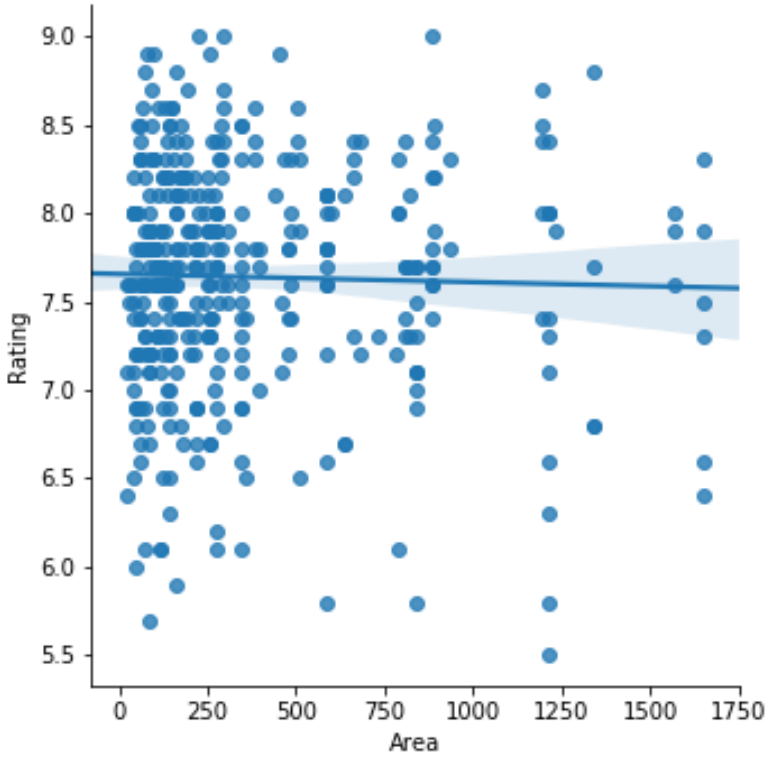


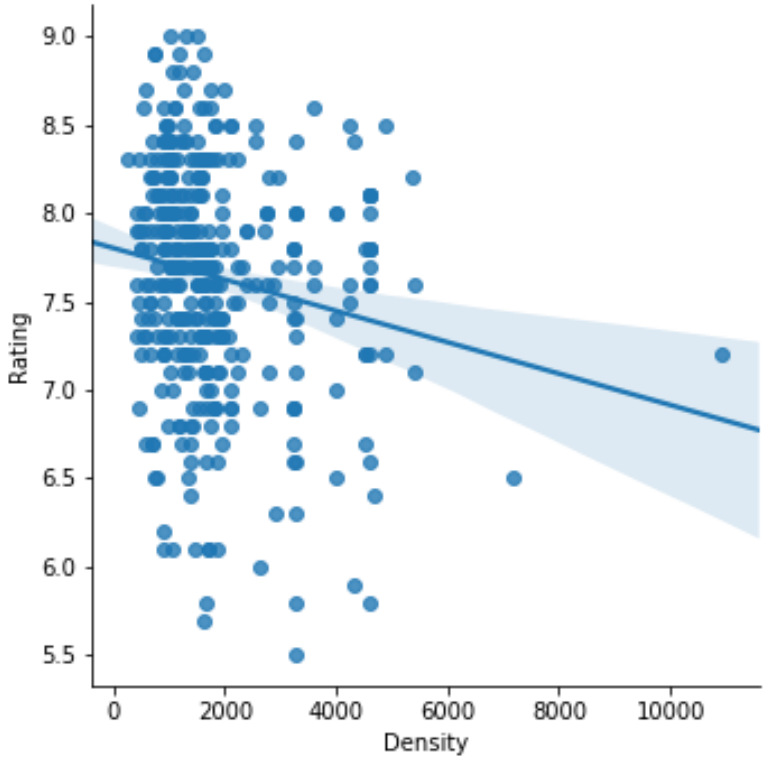


Ratings Analysis

The ratings are normally distributed around the value of 7.75 with only 5 outliers between the values of 5.5 and 6.







Absence of a correlation

Contrary to our initial hypothesis, there is no correlation between the city in which the Starbucks is located and its rating. This suggests our predictive model will rely heavily on the buildings surrounding the store.

Predictive Modeling

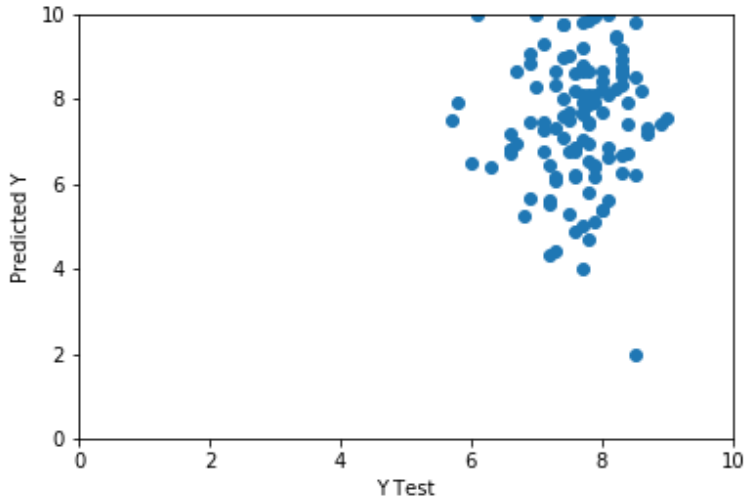
- Linear regression model was used as it is the best for predicting quantitative values with a limited dataset
- The model was trained using 617 features to predict the rating
- The following table shows the 5 features that lead to a high ranking and the 5 that lead to a low ranking

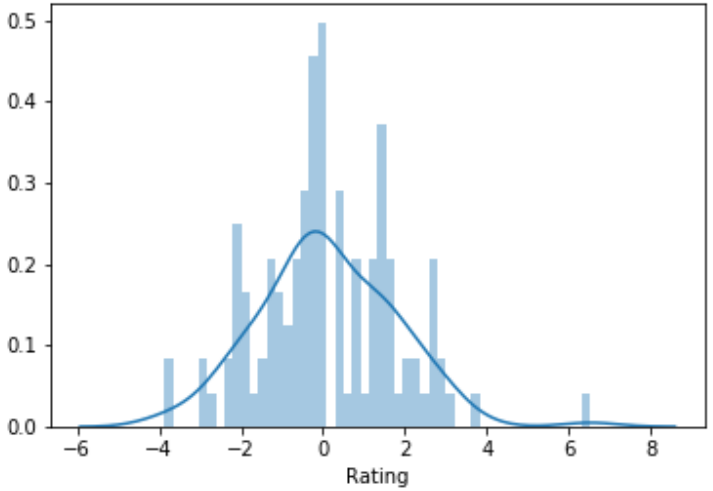
Feature	Coefficient	Feature	Coefficient
Pool Hall	42.08	Recreation Center	-43.06
Football Stadium	29.26	Beer Bar	-36.62
Garden Center	28.74	Auto Dealership	-34.84
Pier	26.60	French Restaurant	-31.34
Plaza	24.07	Shoe Repair	-29.35

Model Performance

- The model contains some outliers but in general, there is a correlation between predicted and real values
- No overfitting and performance metrics within acceptable range
- Residuals distributed normally around 0 = no bias

Metric name	Value
Mean Absolute Error	1.28
Mean Squared Error	2.80
Root Mean Squared Error	1.67





Conclusion

The result of this research was that the only good indicators of whether a Starbucks will have a good customer satisfaction rate is the type of buidings around it

Executives should use the results outlined above lightly as they do not show a complete picture of how successful the store really is

Future Directions

More Data: gather data from all around the world, analyzing ratings from more than one source and collecting more demographic data

Definition of Success: include financial data about the store vs just its rating