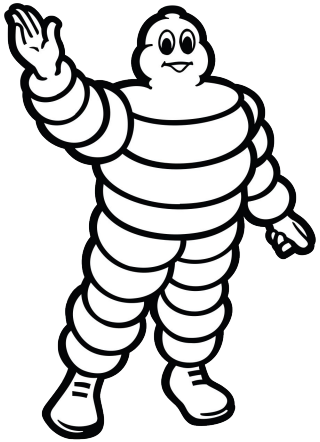


Michelin Stars & Yelp Ratings

Is there a relationship?

We're Manasvi, Erin, Heather, and Asher

Michelin Guide History

★ Starting in 1926, the first Michelin Star ratings (one star) were awarded to restaurants if they were deemed a “fine dining establishment.” They were all located in France.

★ In 1931, the rating system was expanded to become the Michelin three-star rating that it continues to be today.

★ The Michelin star rating didn’t take hold in America until 2005 and concentrated solely on fine dining in New York at the time. Today, the Michelin Guide reviews restaurants in select U.S. cities including Chicago, New York, Los Angeles, Las Vegas, and San Francisco.

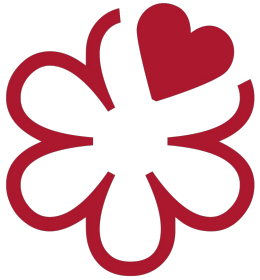
How the Michelin Guide Operates

- ★ The Michelin reviewers, also known as Inspectors, keep their identities a secret to visit restaurants and have their meals expensed by Michelin (the restaurants do not pay for their meals).

- ★ The Michelin Guide goes above and beyond to protect the anonymity of their Inspectors (they are not allowed to disclose their line of work to anyone outside Michelin).

- ★ Due to the high volume of restaurants existing, it would be impossible for inspectors to visit each one to determine whether it is Michelin-worthy or not. With this existing gap, we decided to build a machine learning model.

Why we chose Michelin?



We wanted to know:

1. Could using Yelp data and reviews help predict how many stars a known Michelin restaurant has?

2. Could Yelp data and reviews help predict if a given restaurant has any number of Michelin stars or not?

Source Data

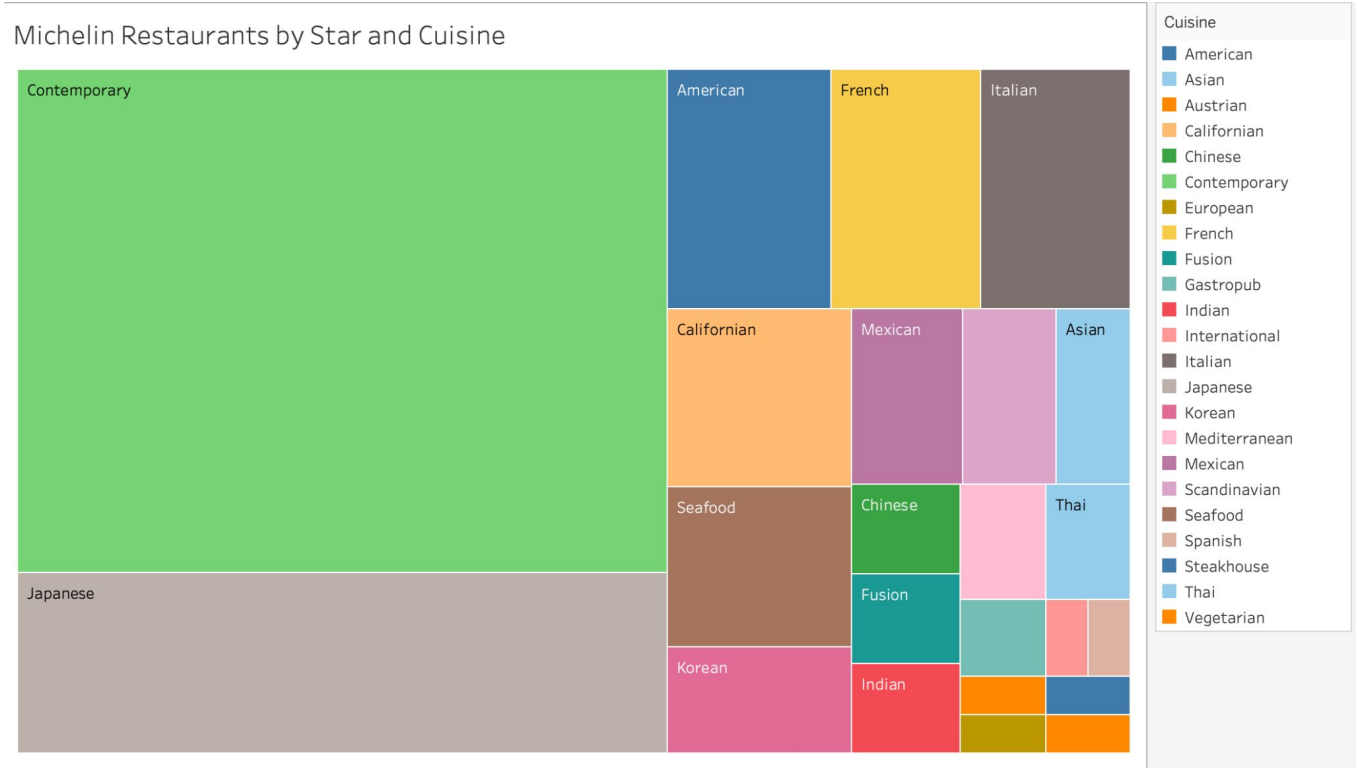
To build our models, we wanted to be able to pull as much data as possible to optimize machine learning. After reviewing various food data sites, we chose a dataset from Kaggle for Michelin star restaurant data from 2019. We gathered Yelp restaurant data using a Yelp API, and scraped Michelin restaurant data from Yelp using BeautifulSoup and Splinter. The data was cleaned and merged using Jupyter Notebook. Finally, the data was written into a CSV file for Machine Learning Analysis.

Reasoning:

- ★ Most robust data tables
- ★ Longest history of data
- ★ Low level of complexity on the Yelp website, allowing us to scrape without a web driver

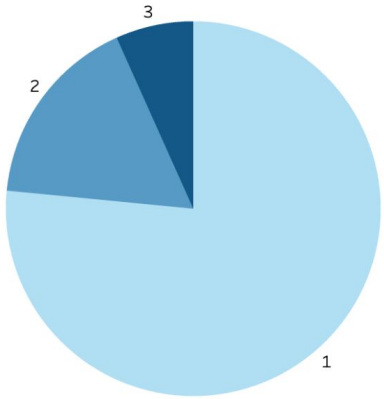
kaggle

Initial Overview of Michelin Restaurant Dataset

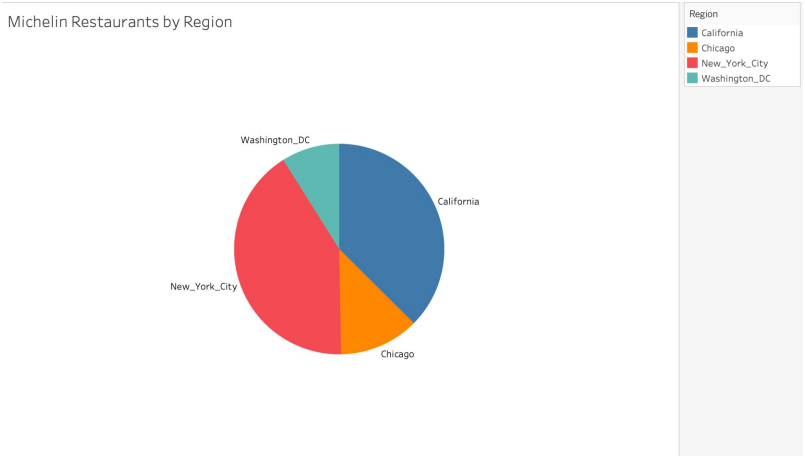


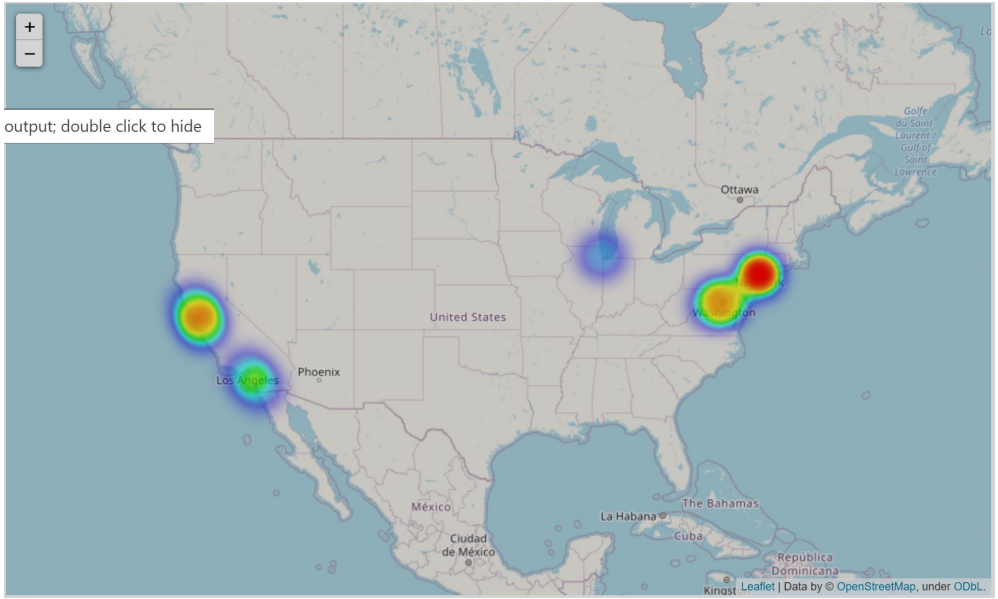
Initial Overview of Michelin Restaurant Dataset

Michelin Restaurants by # of Stars



Initial Overview of Michelin Restaurant Dataset





Building the Models

We tested several different types of models. We split the data into training and testing sets as appropriate.

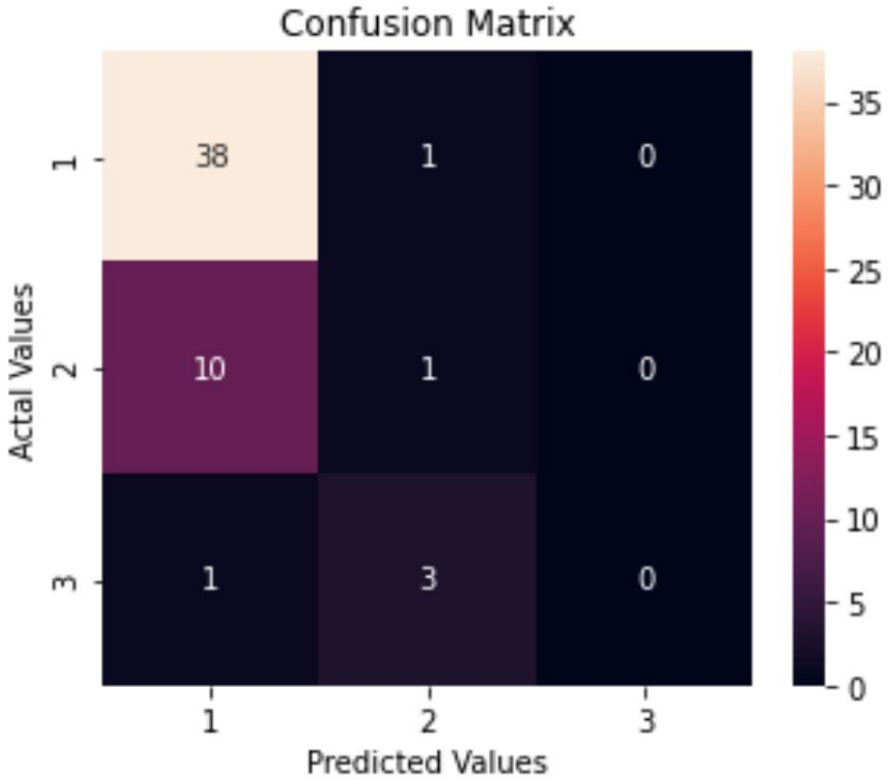
To answer our first question, *Could we build a model to classify how many stars a Michelin Starred restaurant has?* We started with a Linear Regression model and a SVM model with little success and moved onto a Deep Learning model.

To answer our second question, *Could Yelp data and reviews help predict whether a restaurant has a Michelin star?* We at first used Random Forest, Logistic Regression, and a Deep Learning Model. All three models came to the same result, and had an overall accuracy score of 97%, with very high accuracy in predicting if a restaurant did not have a Michelin star, and extremely low accuracy in predicting if it did have one. We then used Resampling Models (Oversampling, Undersampling, Combination, Balanced Random Forest, and Easy Ensemble AdaBoost).

SVM to Determine 1, 2, or 3 stars

- Kernels tested:
  - ‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’
- Most successful run was done with sigmoid kernel which is shown.
- Accuracy score: 72%

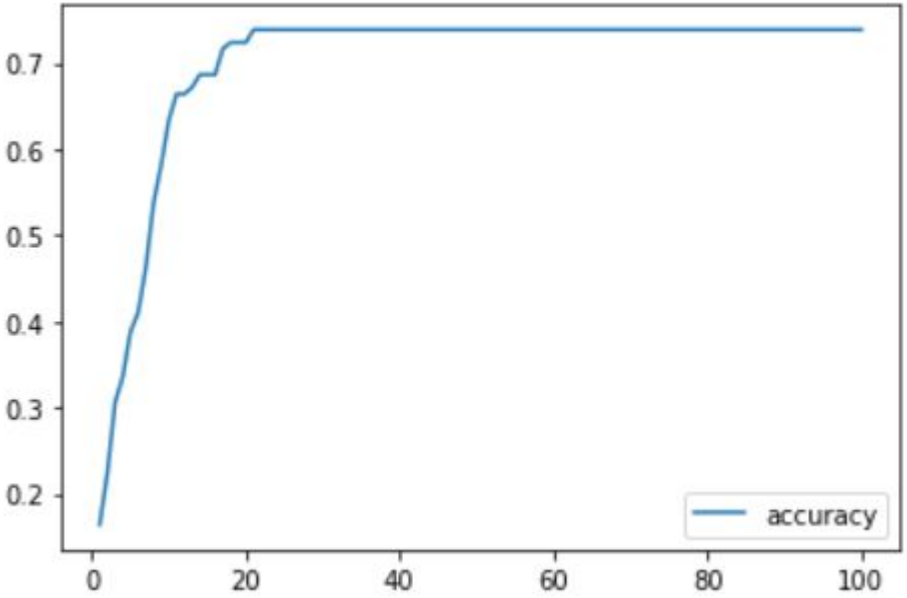
Weight	Feature
0.1630 ± 0.0363	review_count
0 ± 0.0000	lonBin
0 ± 0.0000	latBin
0 ± 0.0000	longitude
0 ± 0.0000	latitude
0 ± 0.0000	price
0 ± 0.0000	rating



Neural Network Model to Determine 1, 2, or 3 Michelin Stars

```
Loss: -11.07380199432373, Accuracy: 0.8444444537162781
```

- We used data from 179 Michelin Starred Restaurants across the United States



Model Scores to Determine Michelin Star Restaurants

Deep Learning Model

```
56/56 - 0s - loss: nan - accuracy: 0.9747 - 191ms/epoch - 3ms/step
Loss: nan, Accuracy: 0.9746906757354736
```

Undersampling

- The Undersampling model predicts 57 restaurants have Michelin Stars.
- The model predicts that 1676 Restaurants do not have Michelin Stars.
- The model predicts that 45 restaurants that have Michelin Stars.

Confusion Matrix

	Predicted no_star	Predicted star
Actual no_star	1676	57
Actual star	0	45

Imbalanced Classification Report

	pre	rec	spe	f1	geo	iba	sup
0	1.00	0.97	1.00	0.98	0.98	0.96	1733
1	0.44	1.00	0.97	0.61	0.98	0.97	45
avg / total	0.99	0.97	1.00	0.97	0.98	0.96	1778

Model Scores to Determine Michelin Star Restaurants

Resampling Models:
Naive Random Oversampling
SMOTE Oversampling
Combination (Over and Under) Sampling
Balanced Random Forest Classifier
Easy Ensemble AdaBoost Classifier

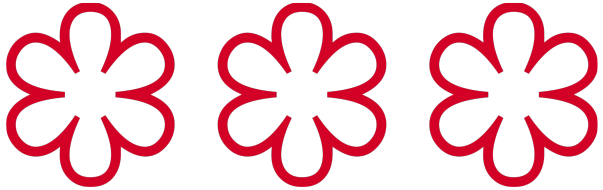
Confusion Matrix

	Predicted no_star	Predicted star
Actual no_star	1733	0
Actual star	0	45

Imbalanced Classification Report

	pre	rec	spe	f1	geo	iba	sup
0	1.00	1.00	1.00	1.00	1.00	1.00	1733
1	1.00	1.00	1.00	1.00	1.00	1.00	45
avg / total	1.00	1.00	1.00	1.00	1.00	1.00	1778

Conclusion



- ★ First Question: Can we accurately predict how many stars a Michelin starred restaurant has using data from Yelp?
  - The Logistic Regression and Support Vector Classification Models were unsuccessful predictors of our first question.
  - Although we yielded an 84% accuracy with the Deep Learning model, our dataset was far too small and we did not find meaningful results.
- ★ Second Question: Could Yelp data and reviews help predict whether a restaurant has a Michelin star?
  - Using various models of both Supervised and Unsupervised Machine Learning Models to determine whether machine learning can predict if a restaurant has a Michelin Star, we found that Deep Learning and Undersampling were the least accurate models, while all Supervised Machine Learning models were very precise in determining Michelin Star restaurants.

Next Steps

- ★ Improving the Model:
  - Include more restaurants with Michelin stars into the dataset, which would require expanding to additional countries (179 restaurants from the United States included in dataset, with only roughly 200 Michelin star restaurants in the US at a given time)
  - Include more data from Yelp:
    - scraping Yelp reviews and using Natural Language Processing to extract and analyze keywords
    - sorting Yelp reviews into elite and regular members to see if reviewer status has an impact on predicting if a restaurant is Michelin starred

- ★ Continued Testing of Accuracy:
  - Since the Michelin star restaurant data used in this analysis was from 2019, we could compare and improve the model against known 2020-2021 Michelin star restaurant data

- ★ Future Models
  - Make year 2020-2022 predictions and compare to real results

Technology & Resources Used

- ★ Python
- ★ Jupyter Notebook
- ★ Excel
- ★ Pandas
- ★ BeautifulSoup
- ★ Splinter
- ★ Tableau
- ★ R
- ★ SQLite
- ★ Kaggle
- ★ Yelp

Any Questions?

★ Thank you for listening!