Game of Michelin Stars (The Guide for Michelin Guide)

Design

For restaurants, the road to earning a Michelin star is rigorous and competitive. Due to the pandemic, Michelin has difficulty sending their employees to various restaurants throughout the world. Therefore, the current list of Michelin restaurants need adjustment in terms of whether current restaurants need to be dropped/adjusted or new restaurants should be added to the new list. In order to do this, Michelin reached out to the Data Science team to come up with a model that could predict “Michelin Score” that encompasses a fairly accurate measure of whether restaurants are suitable to be considered a Michelin without having to visit every restaurant during the pandemic.

Goal/Data Description

This project seeks to propose a data science approach to finding a solution for Michelin Guide to help them monitor whether the restaurants that have been awarded with Michelin stars are keeping up to their standards. Michelin is essentially looking for a predictive model that can give an estimate of Michelin standard score to determine whether the restaurants are still maintaining a high quality service during the pandemic.

There wasn’t an extensive dataset where it had detailed components of Michelin restaurants in terms of user rating, overall rating, exact price...etc so my initial approach is to incorporate US Health Department restaurant inspection data. Health inspection scores will be explored as it is one of the easiest public data that can be explored to determine cleanliness of the restaurant which is an essential feature that can be included in our final model.

Tools

Preliminary analysis and visualization through excel and Tableau interactive dashboard will be used to identify Michelin earned restaurants in the United States along with their corresponding health inspection score recorded by the Department of Public Health.

Data
San Francisco, Chicago, New York Public Health Inspection data will be used

Kaggle: https://www.kaggle.com/jackywang529/michelin-restaurants
Overall 50,000+ observations.

<US only>
1 star : 155 observations
2 star : 17 observations
3 star : 14 observations

MVP
First MVP is to have a graph that visualizes the health inspection score distribution among Michelin restaurants in the US.

Data science goal: To design a model that can predict Michelin standard score for restaurants.

Desired business impact: To help Michelin determine whether the restaurants are still worthy of maintaining/earning Michelin stars despite having to directly visit all restaurants during the pandemic.

Non-technical Scoping:

- Data like Customer Rating, Reviews are subjective, so the model that can predict the Michelin score may not be objective, whereas the health inspection data is fairly objective as it is provided by the health department.
- The downside of this model is that it is difficult to incorporate the data regarding the “taste” as our scenario is that Michelin employees cannot visit all the restaurants during a pandemic, but we still have to come up with a model that can predict whether the restaurant is suitable to earn a Michelin star.

Technical Scoping:

- More data will need to be scraped from website like Yelp, Zomato, OpenTable regarding review of the restaurant