

Predict Restaurants Rating

Caroline Troude: cct65 ; Diego Carrasco: dbc86 ; Juan Felipe Gonzalez: jg2265

September 21, 2017

1 Problem Specification

What makes a restaurant successful? For restaurateurs, it is crucial to understand their business in order to improve their performance and increment their clientele. However, in the restaurant industry, making data driven decisions can be challenging. Fortunately, in recent years, thousands of websites offer services to their users and allow them to have free access to restaurants reviews. This can be either a great advantage or a great disadvantage for restaurateur. People are relying more and more on sites such as Tripadvisor to find places to have a meal and they are highly influenced by ratings, prices and public comments.

With this project, we want to explore all the information contained in TripAdvisor and possibly other data bases to answer the following questions:

- Based on available public data, Can we help restaurant owners to understand their industry?
- Can we extract insights to improve the rating of a restaurant that already exists?
- Based on comments from a focus group, can we predict the future rating of a restaurant before opening ?

2 Description of Dataset

We are going to study the source code of TripAdvisor website in order to create a parser (a Python program) that will collect useful input data. We are going to focus on New-York restaurants; the corresponding URL is the following one: <https://www.tripadvisor.com/Restaurants-g28953-NewYork.html>

Our dataset will have the following fields: Name, Address, City, State, Country, Neighbourhood, General ratings, Service Ratings, Value Ratings, Food Ratings, Atmosphere ratings and comments. The fields might evolve during our project according to the questions that will arise and according to how well our model meet our expectancies. Once the data has been collected, we have to analyze it and determine the connection between the user's ratings and other factors such as comments on the restaurant location, atmosphere, food etc.