

Alice Vadney

DS 4002

<https://github.com/alicevadney/CS3/tree/main>

Do Charlottesville Restaurant Reviews Match Their Sentiment?



Consistently ranking among cities with the highest number of restaurants per capita, Charlottesville, VA boasts a robust and lively restaurant scene. From dive bars on the Corner to nicer establishments on the Downtown Mall, the city offers something for every palate.

With reservations filling up weeks in advance and a wide variety of cuisines to choose from, online reviews are one of the best tools to use when deciding where to eat. A quick glance at a restaurant's average star rating often determines whether it is worth fighting for a reservation. But how much can you really trust those stars?

While ratings offer a convenient summary, they might not tell the full story. Some five-star reviews are barely more than a single sentence, if at all. On the other hand, some two-star reviews might be filled with praise about the food, service, and ambiance, but drop the score due to a minor issue like parking or a long wait.

This is where you come in. You are a data scientist tasked with finding the correlation between ratings and the sentiment of the written content of restaurant reviews in Charlottesville. Your goal is to determine whether the sentiment expressed in the text of a review matches the number of stars it was given.

To do so, you will perform a simple linear regression analysis using OLS (Ordinary Least Squares) to extract sentiment scores from review texts and compare them to the corresponding star ratings. You'll analyze how closely the sentiment aligns with the given rating, and thus if numerical ratings are a good indicator of a restaurant's quality.