Bon appétit! In this project, we aim to identify the key features people in different countries look for in their dining experience. Are Americans more inclined toward a late-night snack than their German counterparts? Do Canadians value a take-out option more than those who live in the United Kingdom? These findings will help businesses make informed decisions when expanding globally. To achieve this goal, we tackled the problem in three phases:

1. Obtain and preprocess restaurant data from around the world.
2. For each country, perform feature selection to pick the attributes that correspond to high star ratings.
3. Run various classification and regression models to evaluate the strength of selected feature sets.

Dataset and Preprocessing
We selected Yelp, one of the largest and most popular platforms for crowd-sourced reviews about restaurants, to be the primary source of our data. The main dataset comes from the Yelp Dataset Challenge, and contains a total of 77K businesses, 2.2 million reviews by 552K users, and 566K business attributes. Restaurants come from one of four countries: the United Kingdom, Germany, Canada, and the United States.

We ran a Python script to convert the raw JSON data into CSV files. We converted categorical feature values, such as casual, dressy, or formal for restaurant attire, into numerical values, such as 0, 1, and 2, respectively. For feature values that are true and false, we converted the Boolean into an integer so that 1 represents true and 0 represents false. To account for missing data attributes, we used a method similar to sci-kit’s Imputer - we replaced empty feature values with the average of existing values in that feature column.

Feature Selection
We considered two different methods of restaurant classification based on their star ratings: binomial and multinomial. In the binomial case, restaurants with a star rating below 4.0 are classified as 0, and restaurants with a star rating of 4.0 and above are classified as 1. The machine learning models we used to train and predict the data are Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Gaussian Discriminant Analysis. In the multinomial case, restaurants are classified from 0 to 5 based on the nearest integer value to their star rating. The machine learning models used are Multinomial Logistic Regression, Decision Tree, and Random Forest.

Classification Methods
We conducted a variety of binomial and multinomial classification models after selecting 20 features, we discovered that Anova F-value is a more consistent scoring function. The results of modeling using features selected by Anova F-value show that Naïve Bayes is generically the best model for restaurants in all four countries. The Multinomial Decision Tree is the worst. This suggests that features have low dependency upon each other. Also, Naïve Bayes could also be a robust model in this case because there are fewer data samples for most countries. Finally, it appears that test accuracy is generally higher for binomial models than for multinomial ones, which makes sense since multi-class classification usually yields lower accuracy.

Discussion
We found that in both the Chi-squared and Anova test results, there are 6 features that are highly weighted in all four countries: parking street, takes reservations, review count, casual ambience, noise levels, and attire. Other features that were generally correlated (which we quantified as appearing on three of the four countries’ feature selected lists) include: hipster ambience, garage parking lot, Wi-Fi, intimate ambience, good for kids, good for groups, allows smoking, and has TV.

From the most significant word features retrieved from sentiment analysis, we note that Americans reviewers tend to be more negative in their reviews, with rampant capitalization and emphasis on lack of personal attention (acknowledge, acknowledgement).

Future Work and Acknowledgements
One of the neat aspects that our project focuses on is identifying any differences in the success of a restaurant across multiple countries. One of the challenges that we faced in doing so was performing sentiment analysis on German restaurant reviews, since the majority of the reviews were written in German. While this proved difficult to process, we were still able to train our models on the features given in the dataset; in the future, however, we would like to improve our accuracy through leveraging human-annotated multilingual sentiment datasets and explore language-independent sentiment analysis.

We conducted Binary Naïve Bayes (unigram, tokenized, stemmed) using NLTK’s built-in classifier to determine the word features in restaurant reviews that were most indicative of restaurant success.