

## **Executive Summary**

### **Introduction**

A person owning a restaurant might get positive or negative feedback from their customers on Yelp. But, what specific reason is prompting their customers to give such feedback? This is a question that permeates the mind of restaurant owners everywhere. If this question is addressed, it would give the restaurant owners direction and insights into the specific experience factor(s) they should focus on. To accomplish this goal, firstly we hypothesize that topics such as food quantity and quality, service, ambience are the most important criteria that matters to the customer's satisfaction and ultimately, the ratings they provide. We perform topic modelling and sentiment analysis on the reviews obtained from Yelp for restaurants in Chicago area. Then through a regression analysis, we aim to derive insights about what dining experience topics carry the most weight when explaining restaurant star rating reviews. This is important because even a one-star increase can led to ~9% increase in revenue of independent restaurants.

### **Analysis**

We start off our analysis by first building a relevant data set using BeautifulSoup and recognizing various patterns in the URL and HTML structure. Once we obtain the review text and other relevant data, we preprocess it to make it ready for aspect extraction using topic modelling. Our purpose for this step was to obtain relevant categories. We utilized LDA for the same and obtained categories such as food, service, place and price. Under the hood, topic modeling methods find the most common things people are talking about in reviews and aggregates them into common categories. We do this because a high level categorization of reviews into relevant categories can help user to understand why the reviewer rated the restaurant as "high" or "low". From the topics

obtained, we can deduce that they are important to customers as a large portion of the customers mention these topics in their reviews. But, our alternate hypothesis was that these experience factors impact the rating they provide. To find out whether these topics impact the ratings or not, we perform linear regression. By this we can conclusively prove or disprove our hypothesis. In order to build a model, we assigned overall polarity scores to each topic at a restaurant level. Polarity scores basically describe whether the reviews expressed is positive, negative, or neutral. We build our model by training it on the polarity scores of every topic for each restaurant. The linear regression model we obtained works well because the beta coefficients provide information about the relationship between each topic and the rating. Also by analyzing the beta coefficients, we can conclusively prove that a particular topic has a non – significant effect on the rating and thus we proved our hypothesis.

### **Application**

To further establish our hypothesis, we took Ballast Point restaurant in Chicago as an example. We analyzed their present polarities and found that there is a definite room for improvement in the quality of food, service provided and also in pricing. By utilizing our linear regression model, we recommended several measures Ballast Point can take to significantly improve their revenue.

### **Conclusion**

Our model has certain shortcomings with regards to computational constraints and also bias due to non-segmentation of restaurants by cuisines. But, despite these shortcomings there can now be a better understanding of what customers are talking about and how they influence the restaurant's overall Yelp review ratings from the model we developed.