

SPRING 2022

CAPSTONE 3 REPORT



PREPARED AND PRESENTED BY
RAJA RIDGWAY

PROJECT DESCRIPTION

Chicago pizza magnate Lou Malnati is looking to expand his national pizza empire. With 59 locations in Illinois, seven in Arizona, four in Wisconsin, and four in Indiana, Malnati is interested in expanding both within Arizona and Indiana and to other states. In particular, Malnati is interested in Florida, Pennsylvania, New Jersey, and Missouri.

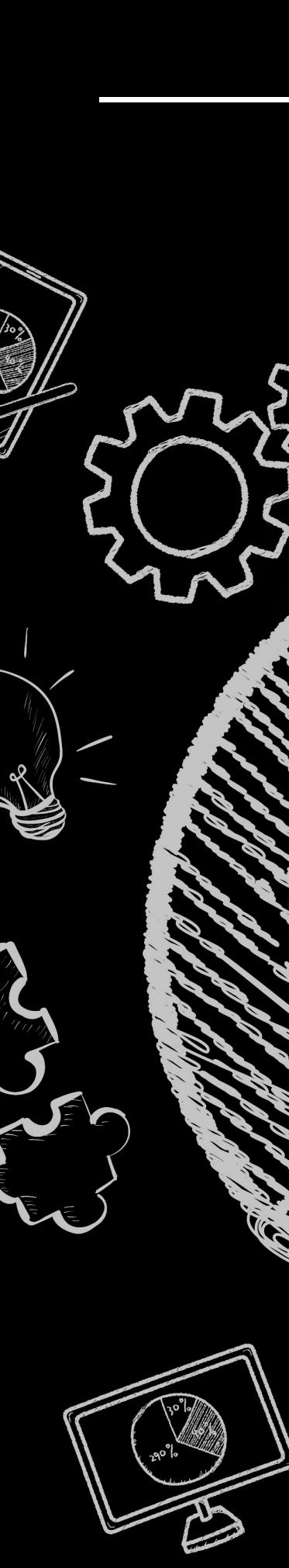
Malnati's restaurants are known for their deep dish pizza, and are looking for locations that either might not have deep dish options or locations where the pizza options are not satisfying consumers. Malnati's team believes that they can both introduce deep dish to new customers and lure currently unsatisfied customers with their nationally recognized pizza brand.

Malnati's team has requested an analysis of the existing landscape in the four new states along with Arizona and Indiana. They want to understand which state holds the most promise for more or more new locations. Ideally, they would like to open up multiple locations and want to know whether one of the new states would be a better option than continuing to open up restaurants in Arizona and Missouri.

DATASET

Yelp is a platform that, at its core, allows users to post reviews about businesses. Users can post a quantitative score and provide comments about their experience. Yelp also provides an open dataset as JSON files that can be used for data science projects. The data for this project will come from the Yelp Open Dataset, which includes approximately 7 million reviews across 150,346 businesses in 11 metropolitan areas. The data can be downloaded directly from the website in a compressed folder with JSON files. The two data sets to be included are the business and reviews files. The business file contains information about the business itself, while the review file contains the reviews - the two files both contain the business_id feature which links the businesses to their reviews.





APPROACH

Systematic and iterative

01

Access and Clean the Dataset

The dataset was downloaded from Yelp in a JSON format. The data was converted into a pandas dataframe and filtered for businesses that were categorized by the term "pizza."

The dataset was cleaned by filtering for the six specific states, converting datatypes, checking for outliers, adding a column for binary ratings, and dropping unnecessary columns from the reviews file.

02

Exploratory Data Analysis

Extensive exploratory analysis was conducted by first looking at location data on maps - both at the state level and zoomed in on the location of the businesses. Numerical data, including ratings, review counts, and open vs. closed were explored by state. Time data was also explored to understand how the reviews had changed over time. Finally text data was explored to filter by deepdish references and to process the comments for modeling.

03

Preprocess and Model

Sentiment analysis was conducted to determine the overall sentiment of the dataset. The data was split into training and testing datasets with three models run.

The three models included a Naive Bayes Classifier, a Logistic Regression Classifier, and a Random Forest Classifier.

04

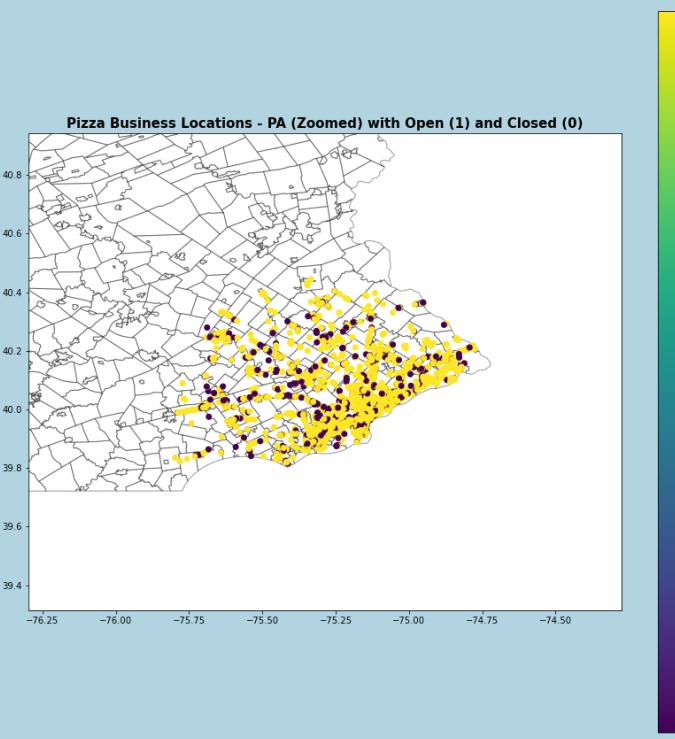
Evaluate the Models

Models were compared using accuracy, F1-scores, lift, and mean cross-validation scores. Hyper parameter tuning was then used to improve the best model.

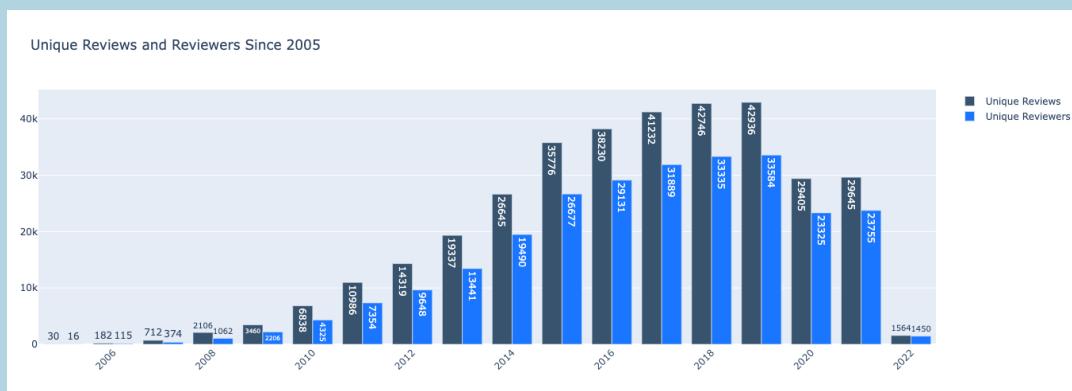
FINDINGS

Clustered communities with varying densities and similar ratings

For each state, the pizza businesses analyzed were clustered in specific parts of the state. There were no discernible patterns for which businesses were currently open or closed. For example, in Pennsylvania, the majority of the businesses were located in Philadelphia:



Pennsylvania accounted for about 40% of all the businesses in the dataset while Florida made up 20% of the businesses. A similar trend was true for the number of reviews and unique reviewers, however the distribution of ratings was similar across all states (median rating of 3.5 stars). For most states, the number of ratings had decreased since 2019, potentially due to COVID. However, Missouri, Florida, and Arizona have all seen slight rebounds in the number of reviews since 2020.



FINDINGS

Prevalence and popularity of deep dish pizza varies by state.

While **all states have businesses with reviews that mention deep dish pizza, there is variation by percentages of business** in each state. Indiana, for example, has at least one review that mentions deep dish pizza for 23% of their pizza businesses. This is contrasted with New Jersey, where only 5% of businesses have at least one review mentioning deep dish pizza:



There are also differences in how customers rate their reviews if deep dish is mentioned. While **all states see increased average review scores if deep dish is mentioned, there is again a lot of variation between the states**. In Missouri, for example, reviews mentioning deep dish were 0.33 stars more positive than the average review. This is compared to New Jersey, where reviews mentioning deep dish were only 0.047 more positive than the average review.



When reviews were analyzed to consider the most popular three-word combinations, **it's notable "Chicago" was in three of the top five word combinations**. This may indicate that people eating deep dish pizza make the comparison to Chicago.

3-Word Combinations	Frequency
deep dish pizza	2090
chicago deep dish	467
style deep dish	295
chicago style deep	258
chicago style pizza	245

RECOMMENDATIONS

Three potential market types to consider

01: Provide another option for deep dish lovers

Indiana, Missouri, and Arizona are all states where perceptions of deep dish pizza are positive and the percentage of restaurants offering deep dish pizza are relatively high. Each of these three states also has fewer pizza options in general, making it a potential option for a well-known Chicago-based chain to attract customers already excited about deep dish pizza.

02: Break into a large pizza market

Florida and Pennsylvania are large pizza markets making up 60% of the pizza businesses in the analysis. Most businesses, however, do not have reviews for deep dish pizza - although their average reviews of deep dish are relatively high. This provides an opportunity to bring the Lou Malnati brand to a state where people like pizza, but might not know much about deep dish.

03: Change opinions about deep dish pizza

New Jersey is a unique state that has a middling number of pizza restaurants but fewer serving deep dish pizza and relatively low perceptions about deep dish pizza. New Jersey thus might provide an opportunity to lure in customers who are either new to deep dish pizza or are currently unsatisfied with their options.

FUTURE RESEARCH

1

Disaggregate by national chains

Given that Lou Malnati is a national chain restaurant, it is worth understanding how perceptions differ for national chains. For example, are there locations where the positive perceptions of deep dish are influenced by whether the restaurant is a national chain or a local business?

2

Investigate closed businesses

In the entire dataset, 29% of the businesses were classified as closed. A further investigation could consider what types of businesses closed (e.g., national chains?) or whether COVID had a dramatic impact on the shuttering of businesses in specific states or cities.

3

Common words or phrases associated with deep dish

To gain further insights into the other factors that influence a customer's opinion about a pizza business, an analysis could be run to determine which words are most commonly associated with deep dish in comments. Another step would be to disaggregate by positive and negative comments.

4

Comparisons to other types of restaurants

While this analysis focused specifically on pizza businesses, another avenue for exploration could be other types of similar businesses, such as Italian restaurants. The two types of businesses could be compared to see which trends are unique to pizza businesses.