US Restaurant Survival Tool – The QuaFoodies Application

Godson Adejumo, Tiffany Bui, Katherine Graessle, Zhi Yang, Christine Yao, Emily Zakas

Introduction

When evaluating restaurant viability, entrepreneurs often lean on intuition or "experience", and **while they may take a data-driven approach to support decision-making, it's often simplistic and multiple sources of data need to be leveraged as opposed to having a platform where data integration and modeling has been done for decision making**. There are many factors that contribute to restaurant success, but there are few tools that integrate and make sense of the data in an accessible and easy-to-use manner.

The COVID-19 pandemic has presented significant challenges to the restaurant industry, resulting in widespread closures. As a result, there is a lot of uncertainty for restaurant stakeholders – *will my business succeed given the current economic landscape?*

Integrating data from various sources, the QuaFoodies web app has been developed to analyze factors that contribute to restaurant success by predicting whether a prospective business will succeed or fail.

Experiments

<u>Data Attribute Evaluation:</u> To verify that the data attributes we chose were relevant and suitable for predicting restaurant closure, we conducted several exploratory analyses, including:

- Data profiling to further understand the predictor and target variables
- Outlier analysis to identify potential outliers and apply needed treatment to records as necessary
- Statistical analysis to verify modeling assumptions such as normality and multicollinearity checks

<u>Model Evaluation:</u> To determine whether our model predicted restaurant closure accurately, we split our final dataset into 70% training data, 15% validation data, and 15% test data. We then used this data to test five different classification models: Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, and XGBoost. We then evaluated and compared these models using metrics such as accuracy score, confusion matrix, and area under the curve (AUC).

<u>Web Application Evaluation/Testing:</u> We conducted user experience testing leveraging the "think aloud protocol". This strategy involved someone sitting down with us for 30 minutes while we recorded them talking out loud as they used our application. This helped us gather unbiased feedback, first impressions, and assumptions the user made given our visualizations and UI interface. The results were then synthesized by leveraging the "dump and sort" method to sort the feedback into categories that helped us improve the quality and usability of our app.

Results

<u>Model Evaluation Results</u>

- From the test result in Table 1, we were able to determine that the XGBoost model performed the best. These tests were able to help us conclude to the first experiment question, that the champion model we selected, does predict restaurant closure relatively accurately.
  - **Compared to other model selection methods**, the method we adopted of leveraging the F1 score, AUC and accuracy is most optimal as
    - F1 score provides the most balanced metric between sensitivity and precision
    - AUC and accuracy scores provide quick and easy, industry –accepted metrics to compare our model

<u>Web Application User Testing Results</u>

Post User Testing, please see the feedback we received in Figure 2. Please note that the yellow stickies indicate one user's feedback, while the green stickies indicate the second interviewee's feedback. **Compared to other methods of collecting user feedback** like surveys and questionnaires, this method provides real time feedback from the users. All the feedback provided were implemented to arrive at the final solution seen in Figure 1.

Approach

To help prospective restaurant owners gain insights needed to make better location decisions for starting up restaurants, the Quafoodies solution was developed. In building this solution, our approach was to:

- Research factors that drive restaurant opening and closure (i.e. rental price, zip code, etc.)
- Collect and integrate data from multiple sources
- Analyze data and develop a predictive model based on data collected
- Develop a web application for data visualization and restaurant location viability result generation via a choropleth map and user inputs

We strongly believe this approach would help alleviate the problem of prospective restaurant owners not being able to effectively determine location viability, as the approach:

- Is data-driven and captures data from multiple relevant sources to determine restaurant viability
- Integrates COVID-19 related data and correlates its impact to restaurant closure (**this data is a relatively new** but very relevant data attribute)
- Provides a centralized solution for restaurant owners to help determine whether to open a restaurant in a given location or not.

Data

The data used in the web app is stored as CSV; these CSV files are parsed and input into the choropleth map component via JSON.

The main dataset we worked with was sourced from **Yelp**'s open dataset, containing over 100,000 business ids from March 2020. To get current business information such as open/closed status, we developed a Python script to pull data from the Yelp Fusion API. After filtering the data for US restaurants and additional cleaning, we were left with **34,118** datapoints. Note that we were limited to the business ids from March 2020 as the Yelp Fusion API offers no way to explicitly search for closed businesses, and if we searched for other businesses, it is likely that only open restaurants would be returned and skew our data.

To build our predictive models, the Yelp data was integrated with variables from several different data sources, namely, real estate rental price data from **Zillow and Redfin,** COVID-19 transmission rate data from the **Centers for Disease Control (CDC)**, and unemployment rate and median household income from the **USDA Unemployment Report**. All datasets were keyed and joined off zip code, FIPS code, or state.

YELP + ZILLOW + REDFIN + CDC + USDA
31,000+ datapoints and 40+ attributes

Figure 1: The QuaFoodies Web Application

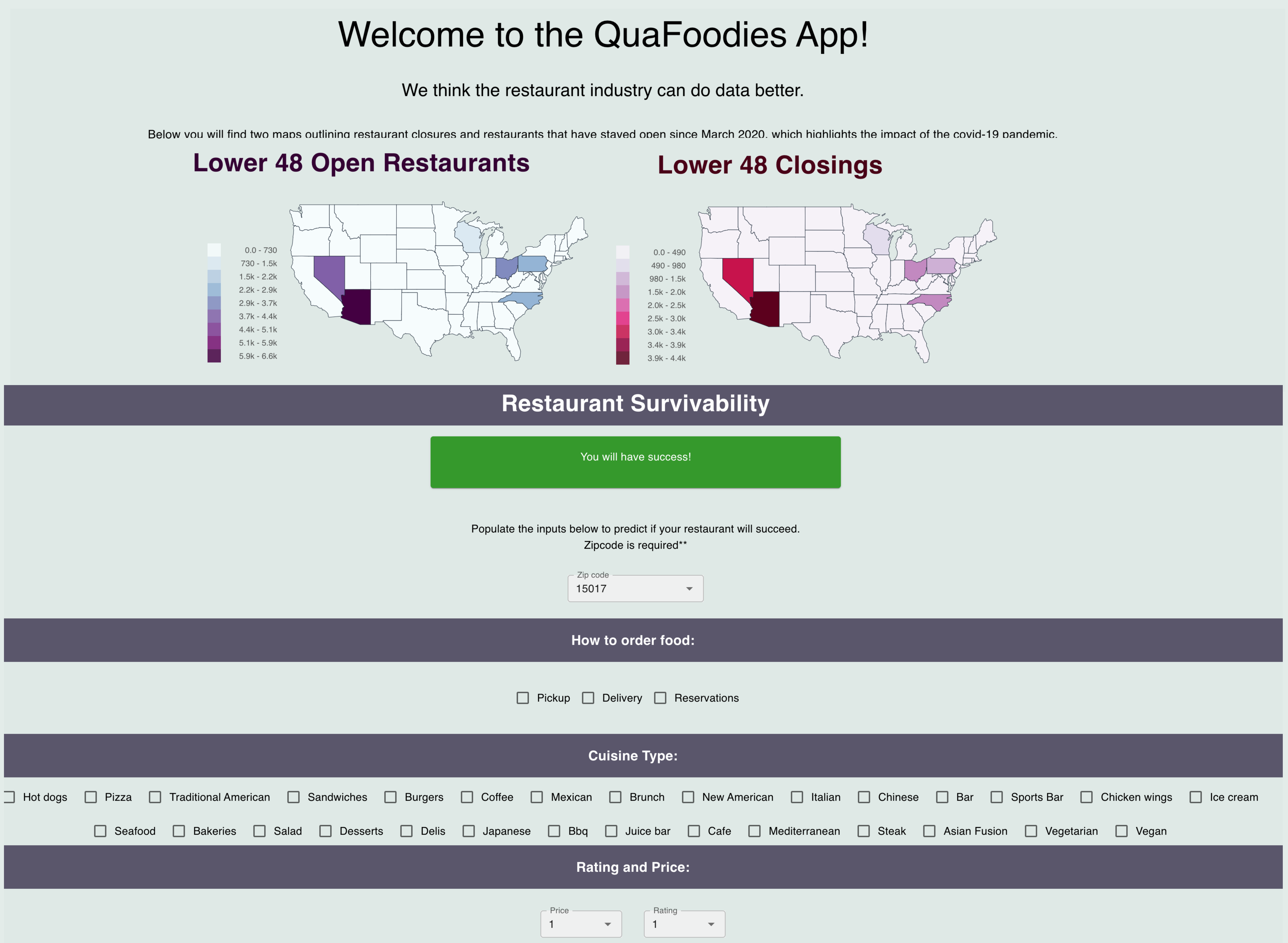

Table 1: Model Evaluation Matrix

Model	Accuracy	f1-score (not closed)	f1-score (closed)	AUC
Logistic Regression	60.0%	0.74	0.11	51.0%
Decision Tree	74.4%	0.80	0.66	72.2%
Random Forest	75.2%	0.81	0.66	72.6%
Naïve Bayes	60.3%	0.73	0.22	52.8%
XGBoost *(CHAMPION MODEL)*	75.9%	0.81	0.67	73.3%

Figure 2: User Acceptance Testing Think Aloud Interviews