Recruit Restaurant Visitor Forecasting

Final Report - Executive Summary

6th March, 2018

Foodie Analytics

Shipra Sethi
Dong Bing
Daniel Colvin
Subba Muthurangan
Erik Platt

Executive Overview

Recruit Holdings has unique access to key data sources on the restaurant industry in Japan. These data sources track reservation and visitation data for any participating restaurant and have amassed a significant amount of historical data over time. If properly mined, this data would not only offer insights on the number of customers expected for any existing or new restaurant owner but also provide overall customer flow in the restaurant industry as a whole.

Recruit Holdings has appointed Foodie Analytics, a leading data consulting firm headquartered in Chicago, IL, to deliver data insight through a predictive modeling solution. Through extensive EDA and multiple models/techniques, Foodie Analytics has identified the best performing forecasting model to predict the number of visitors for a specified future period. This has a tremendous value proposition as understanding future visitor traffic will allow the restaurant owner to:

1. Reduce operational uncertainty and ensure more efficient resource allocation
2. Cut costs by procuring the proper amount of ingredients and adequate staffing
3. Facilitate growth by identifying the most optimal area to expand
4. Better understand the impact of visitation seasonality and holidays on sales
5. Gain competitive intelligence

In order to fulfill these needs and bring real value through our solution, Foodie Analytics focused on delivering the following deliverables:

- **A machine learning model/technique -** This will optimize resource allocation and increase operating efficiency by allowing restaurant owners to predict the number of customers ahead of time.
- **A data visualization platform -** This web portal will provide the capability of visually exploring all relevant data backed by predictive analytics in an interactive environment. Through rich charts and maps, restaurant owners can monitor and track daily customer flow as well as spot patterns and unearth hidden opportunities.
- **A mobile application -** An alternative delivery channel to our standard web portal, it will provide the flexibility to utilize our product for those who are constantly on the move.

Problem Statement

Restaurant start-ups face a notoriously high failure rate. While the often advertised metric stating 90% of restaurants fail within the first year is exaggerated, it is still far from an easy industry to find success. There are many coordinating factors to drive customers in; an appetizing menu simply isn't enough to guarantee traffic. However, increasing utilization of online reservation systems has offered an opportunity to gain a competitive advantage through customer data.

Recruit Holdings' partnership with Foodie Analytics will leverage this customer data. Using existing datasets within Hot Pepper Gourmet and AirREGI platforms, Foodie Analytics will build detailed analytics and predictive models to inform customer traffic patterns. Understanding when and where customers plan to dine allows the opportunity for efficient staffing, fresh ingredients, and an idealized experience.

Our Approach

Foodie Analytics utilized R as the primary analytic tool for Exploratory Data Analysis (EDA) and model development. The web-based dashboard and visualizations are built with Tableau, while the mobile application is built utilizing Android, HTML, Java and CSS Media Queries.

Data Description

The primary datasets which will be used for data insight generation and modeling purposes come from 2 separate data sources:

- Hot Pepper Gourmet (HPG): Similar to Yelp, users can search restaurants and make a reservation online. This data source contains 2 datasets.
- AirREGI / Restaurant Board (Air): Similar to Square, a reservation control and cash register system. This data source contains 3 datasets.

In addition, 2 more datasets are provided, one identifying common restaurants between HPG and Air data sources and another indicating when Japanese holidays occur. We also used historical weather data from the Japan Meteorological Agency (JMA).

Category	Description
Number of Sources	2
Total Number of Datasets	8
Number of Unique Records	963,705
Number of Fields	10
Number of Observations with Missing Values	57
Location of Restaurants	Japan
Time Frame of Observations	1/1/2016 - 4/23/2017

Table 1: Summary of data

Data Merging and Transformation

Even though the provided data came from two different sources, HPG and Air, there were commonalities in the fields which made it possible to merge them together.

Using “store_id”, we were able to combine these 7 datasets into one coherent object. For the weather data, we used the nearest location to merge with all observations. In the training dataset, we only kept the records which contained a value for the “visitors” variable. The final object contained the following columns: “air_store_id”, “hpg_store_id”, “visit_datetime”, “visitors”, “reserve_datetime”, “reserve_visitors”, “genre_name”, “area_name”, “latitude”, “longitude”, “holiday_flg”, “precipitation” and “avg_temperature”.

Exploratory Data Analysis (EDA)

This assessment includes determining the number of missing values, detection of any outliers and their potential impacts, as well as any anomalies within the historical data. The goal is to address and resolve any data quality issues leading to significant adverse impacts on model feasibility. Table 2 below shows a summary of our EDA findings.

EDA Objectives	Insight
Outliers Identification	Distribution of visitors show there are no extreme outliers present
Find missing data points	A small number of missing values are removed and data quality is preserved
Spot relationships in the data	Weak relationships exist between reservation and actual visitors
Analysis of historical trend for business insight	Holidays and seasonality have a significant impact on the number of visitors

Table 2: Summary of Exploratory Data Analysis

Model Development

Foodie Analytics followed the standard model building process through model selection, model fitting and model validation. We determined machine learning is the preferred approach given a large number of records as well as non-linear association with the response variable.

Multiple models/techniques were developed and tested. Variations of each were created using different parameter combinations and values. Models were compared using the Root Mean Squared Logarithmic Error (RMSLE). Table 3 shows the 5 models which yielded the highest accuracy: eXtreme Gradient Boosting (XGB), H2O, LightGBM, and K-Means with eXtreme Gradient Boosting (XGB) and Random Forest. Ultimately, we proceeded with a K-means with XGBoost ensemble method because its score had the highest accuracy in the validation dataset for deployment.

Model Name	RMSLE (Validation Set)	RMSLE (Test Set)
XGBoost	0.52	0.52
H2O	0.512	0.512
LightGBM	0.525	0.525
K-Means with XGBoost	0.47	0.47
Random Forest	0.48	0.48

Table 3: Model Accuracy

Dashboard and Mobile Application

The primary delivery channel is a web-based dashboard, providing a thorough understanding of how to transform our data insights into actionable decisions. Additionally, Foodie Analytics has found much utility on mobile platforms. Ensuring clients can access data while away from their desk has proven to be quite valuable. As such, an interface was developed for on-the-go analytics.

The intent of the dashboard is to give restaurateurs, such as Recruit Holdings, the power to understand what is happening within the competitive market. This will be accomplished via three fully interactive pages within the dashboard, providing different levels of aggregations from a prefecture/national view down to the single restaurant level. The mobile application development is a two-phase approach, first building a native application for Android devices and then developing a responsive design web application for iOS devices.

To view the dashboard in a browser window, visit this link:
https://public.tableau.com/profile/erik.platt#!/vizhome/FoodieAnalyticsDashboardv2/LocationDashboard

To view the responsive web application from a browser window, visit this link:
http://foodieapp-env.us-east-2.elasticbeanstalk.com/

To download the Android mobile application, see section 12.3 of the Appendix in the final report.

Conclusions

Given the highly competitive nature of the restaurant industry, there are many challenges in maintaining a successful enterprise. Foodie Analytics hopes to resolve some of these challenges through leveraging the vast and detailed data collected by Recruit Holdings via Hot Pepper Gourmet and AirREGI platforms.

Foodie Analytics has completed an in-depth exploratory data analysis to ensure all the intricacies and nuances of the data have been captured and understood. We have identified and analyzed the response variable and each predictor variable used for data modeling. We identified commonalities within the datasets, merged them into training and testing datasets, and iterated through several rounds of modeling. After identifying 5 promising models/techniques, we chose to proceed with a K-Means with XGBoost model given its high accuracy of the Root Mean Squared Logarithmic Error. Finally, Foodie Analytics has developed a dashboard and mobile application to help Recruit Holdings derive immediate value from our efforts.

Furthermore, identifying Japanese holidays provided key insights into visitor predictions. In addition, derivations of actual visitors (such as mean and median visitors), and bringing in external data on temperature and precipitation further improved our ability to predict. With a complex time-series problem such as this, we found the inputs needed for accurate predictions were relatively simple. These findings will likely help Foodie Analytics in producing actionable results on related future projects.

Recommendations

Our Tableau dashboard and mobile/web application will provide direct insight into current and potential restaurant owners. Although the final decision for starting/optimizing a restaurant should take into account the professional expertise of the restaurant owner, we suggest using the tool as follows:

For new restaurant owners:

- Identify locations where your particular genre would fit in best. Consider whether a restaurant should be opened in an area where the potential owner's preferred genre is already popular because there is an existing demand for it or whether the restaurant should be started in an area with low density because it is missing in the local market.
- Specify the genre in the dashboard or mobile/web application and look at predicted visitors across competitors. Are there one or two restaurants monopolizing the market in the genre? If so, it may be more advantageous to avoid this area.
- Take into account overall visitor predictions over time to ensure consistent visitation and analyze the impact of visitation on holidays.

For current restaurant owners:

- Analyze average visitation rates vs predictions for the owner's current restaurant. Is it competitive with the average restaurant? If not, why?
- Use the predicted visitor rates over time to compare when competitors are receiving the most visitors.
- Compare predicted visitors for the restaurant owner's preferred genre vs predicted visitors for other genres to see if the restaurant would be more competitive by adding some more trendy options to the menu.

Combining these recommendations with professional knowledge of the restaurant business will help restaurant owners make more money, be more responsive to changes in restaurant trends, and better analyze the competition.