The Modern Food Truck Shift and Recipe for its Success

Deepti Gupta[1]

[1] Harrisburg University of Science and Technology

Author Note

Deepti Gupta, Department of Analytics, Harrisburg University of Science and Technology. This paper is in fulfillment for the course - ANLY 699 Applied Project in Analytics, due Apr 17th, 2020. Special thanks to Professor Arnie Miles for the valuable suggestions on the content of this paper.

Correspondence concerning this article should be addressed to Deepti Gupta, 326 Market St, Harrisburg, PA 17101. E-mail: DGupta2@my.harrisburgU.edu

Abstract

Street food has been gaining phenomenal preference among the present-day consumers from all sectors making it not just a convenient, affordable and prompt option, but also an escape from an ordinary food arena. The perception of the food quality and popularity of a street food business, however, can differ on various parameters and attributes under which the business is operating. This paper was intended to recommend the business oriented and demographical factors contributing to the efficiency of a food truck in achieving the maximum footfall and profit in Manhattan, New York. To accomplish this, business data for Yelp food trucks was joined with the Census demographical data for the zip codes where the food trucks were serving. Each parameter in the dataset was then assessed on its importance level in predicting the business ratings and the filtered parameters were then run on supervised machine learning classification algorithms. The study found that to be efficient and gain good business rating, a food truck should serve economic price ranged authentic cuisines, operate in a region with young consumers residing and with prominent ethnicity of Whites and Asians. It is further recommended for a food truck handler to utilize all the features provided by a business recommendation website like Yelp by adding the trivial details too, envisaging a complete street food experience.

Keywords: street food, food carts, food trucks, New York food trucks, Yelp food trucks, street food businesses, street food handlers, mobile food vendors, street vendors

Word count: 219

The Modern Food Truck Shift and Recipe for its Success

In today's world, when the cities are swarmed up with population, traffic and expanding business centers, people are moving to the outlying and sub-urban areas for a better, affordable and calmer living. The commute has increased to hours and has become a way of life. With this, Food Trucks are gaining general preference over diners and restaurants and attracting customer base from the different sectors of society, be it a factory worker with a menial wage or a white collar professional in search of a gourmet and innovative street food, a student who is covering on his limited expenses or the CEO, who is always on the move, street food enjoys everyone's attention (Alfiero, Lo Giudice, & Bonadonna, 2017).

The most likely reasons for the booming shift to the street vending and mobile food can be accredited to factors like lower cost, ubiquitous, convenience, quicker order processing and ala carte. It also seems a lucrative career start to many young culinary-school-trained chefs, who are mostly low on budget but passionate and innovative, and exhibits high potential in the market or an immigrant (Martin, 2014). This shift has brought opportunities to the small-to-medium sized business entrepreneurs and generating employment multi-folds. "It has also become a way to preserve cultural and social heritage and to stimulate tourism" (Alfiero et al., 2017. p. 3). It offers a wide variety of local and regional cuisine selections with balanced nutritional value for gourmet which is considered by tourists all over in place of diners and restaurants (Alfiero et al., 2017).

This project will help the young entrepreneur in assessing the various contributing factors like location, target audience, pricing limit, menu selection and time of operation in attaining a higher efficiency of a food truck. This problem is statistically significant as the popularity of a food truck is not likely to have occurred purely by chance and thereby has other underlying influential factors responsible which we investigated in our research. This research is extrapolated for the use cases of selecting an appropriate location to open a grocery store or a coffee shop which studied similar patterns of the customer occupancy, behavior and the effect of location matrices on the sales. It also can be used in a scenario of finding a suitable location to set up its warehouse by an e-Commerce company seeking to implement efficient and quicker delivery to its users.

The data for this project for the New York City is retrieved from the open datasets available on web and yelp, google reviews etc. for user reviews. It comprises of following details, (a) geographical location of a food truck and rating of the area based on the nearest proximity of key hot spots of the city like tourist places, parks, lakes; crowd bearing spots like colleges, business parks or shopping complexes; and main train stations, bus stations and other transport centers or junction points of the city; (b) demographic and psychographic data of the geographical location; (c) The details of a food truck with data on number of items in menu, average menu price, timings of operation and number of orders per day. Availability and processing of the data was a challenge here as it required studying and adding most of the information from web.

By performing exploratory data analysis and statistical tests, trends, patterns and correlation among the data attributes were analyzed. It helped in assessing those factors individually and their association with other factors before building the model. The model was trained, built and tested on those attributes and then, the process was reiterated to attain higher efficiency using machine learning clustering algorithms to predict the best location to maximize the foot fall and profit.

Literature Review

In order to understand the food truck shift in the modern environment, it was important to disintegrate the topic into granular set of questions. Addressing them helped us in analyzing the effect of various contributing factors and science behind the popularity of the street food vending system. The primary set of questions were, (a) what are the factors responsible for the evolution of food trucks; (b) why is the food served in a food cart preferred, is it only convenient or literally better than most diners and restaurants too ; (c) does the branding and marketing play any role; (d) what are the challenges involved with workers in the street food vending business; (e) and last but not the least, what is the generic public outlook towards a food truck.

The Trigger of the Trend

A food truck is a large vehicle; which accommodates means to store raw/half-cooked ingredients, prepare and sell food parked at the same location every day or multiple locations based on time/ days of week. The street food in New York sees its evolution from as early as fifteenth century, the trade paced after the revolution of 1848 and flourished into markets in next few decades when they delivered the food to factory workers in trucks on lower prices (Basinski, 2014; Feagan, 2008; Mokhtar, Othman, Arsat, & Bakhtiar, 2017). "In 1925... Vending provided a source of employment for new arrivals who did not have skills or could not find jobs in the formal sector. At the same time, vendors provided cheap food that impoverished immigrants were familiar with from their homelands" (Basinski, 2014, p. 2). The concept now has modernized and has gained significance since last few decades globally. It started in 2008 when the slowdown hit the global market, and several employed in food or other businesses got severely impacted. "A larger ratio of the New York street food vendors are also immigrants... who with meagre resources,... dreams big with ... entrepreneurism in their blood (Basinski, 2014, p. 3). It opened another door for those who were in direct food market but unemployed and others naïve in trade, as many of them pooled their talents with the food services guaranteed by trucks.

The Convenience and Innovation

The easy accessibility and faster meal processing that comes with low cost has proven to be a preference among the millennials. In the modern cities like New York, London etc.; food trucks are located at almost every junction point making it easily accessible for the daily commuters. Today's millennials are looking for pocket-friendly, innovating and creative with high-nutritional value. A connoisseur of food would have the opportunity to have a one on one with the chef along with a meet and discussion with other food lovers providing an open and friendly atmosphere to all (Alfiero et al., 2017). An extension to the traditional food trucks, a new concept of “Gourmet Food Trucks” have come in limelight in the last few years with a motto of “quality over hunger serving high quality food at a little higher prices by revisiting traditional recipes” (Alfiero et al., 2017; Martin, 2014). The findings from the study concluded that Gourmet Food Trucks displayed greater knowledge on hygiene, better customer satisfaction and thus higher efficiency.

A street vending or mobile food provides convenience, easy accessibility to not just to the customers but also the entrepreneur. An entrepreneur does not have to financially equipped to set up a mobile food truck, and the initial setting can be installed conveniently with a low budget along with free and low-cost marketing via platforms, like social media (Alfiero et al., 2017; Yoon & Chung, 2018).

Branding and Marketing

The most used and popular way of branding and marketing a street food business is by using the means of social media (Facebook, twitter, Instagram etc.). Many vendors update the regular consumers with their real time location, specialties and menu changes, and also use the platform to further market the consumer’s reviews to the connected virtual consumers (Mukhola, 2014).

The other crucial aspect of branding itself in an direct marketing is that a food truck should consider taking into account the consumer reviews as “they have become what is known as co-creators” (Arcese, Flammini, Lucchetti, & Martucci, 2015). Consumers share their experiences of food products, and evaluate their attributes (charm, value, acceptance, ideas, feelings, emotions and experiences) and provide feedback in a real time manner, either on the spot or later, which further reaches to a community of real and virtual consumers as precisely expressed by Arcese et al. (2015). Another intellectual way to brand and market a street food vending business is to promote a range of innovative, environment friendly, green and sustainable projects which is respected by the consumer community and they would like to contribute their bit by promoting the food truck business and spreading the word (Arcese et al., 2015; Ngo, 2012). Few examples stating the innovating ideas are (a) reducing or limiting the use of plastics; (b) recycling or upcycling the paper cups and plates after use; (c) reusing the metal cutlery; (d) managing the packaging waste; (e) promoting water harvesting and food waste reduction.

Challenges Involved

Reports says that convenience on a low price comes with a price too. Contrasting results have been published in the laboratory studies on the poor microbiological quality of the food served by the street vendors across the developed and developing countries (Castetbon, 2017; Krishnasree et al., 2018; Lues, Rasephei, Venter, & Theron, 2006; Samapundo, Climat, Xhaferi, & Devlieghere, 2015). Results confirmed that majority of the food handlers are not aware of the importance of trivial matters like proper personal hygiene and food storage and safety measures (Lues et al., 2006).

Considering these factors, the city licensing authorities in USA have heavily regulated street vending businesses with strict enforcement and steep fines. “Vendors are perhaps the most heavily regulated businesses in the city and enforcement is strict, with fines steep: as high as $1,000 per violation” (Basinski, 2014, p. 6). One of the most challenging aspect for a layman street food vendor to follow the rule book distributed by the city authorities is to understand the densely written legal terms of administrative code which is incomprehensible for many immigrants (Basinski, 2014). To help this bottleneck, a non-profit organization “street vendor project” has designed graphic brochures and video in seven languages for a better visual understanding of the rules (Basinski, 2014).

A significant challenge for a food truck is to showcase a clean and hygienic food handling and cooking practices to its consumers to conquer the generic prejudice of an ill-managed disease prone environment and poor cooking conditions in the food truck (Alfiero et al., 2017; Lues et al., 2006; Samapundo et al., 2015). If a food truck is in-housing raw food ingredients, then a special care is needed in the transportation and storage of it. As revealed in the study conducted by Sabbithi et al. (2017), "Raw foods, especially ready-to-eat vegetables, salads, sprouts, cut fruits have been often implicated in outbreaks of foodborne diseases in both developed and developing countries". To address this issue, the World Health Organization (WHO) has ruled in the Five Keys to Safer Food, i.e. keep clean, separate raw and cooked foodstuff, cook thoroughly, keep food at safe temperatures and use safe water and raw materials and integrate them into the street food sector (Alfiero et al., 2017; Donkor, Kayang, Quaye, & Akyeh, 2009; (WHO) & others, 2010).

Another primary challenge for a food truck to attract consumers is to find a perfect location as elaborated by Sevtsuk (2014) in their study. The contributing factors that drives the sales could be (a) clustering of the other retail stores – frequented or less frequented, to have an access on just residents but employees too; (b) intense "between-ness" and passing by; (c) residents having walking access to the service provider; (d) junction of multiple routes with high traffic (Sevtsuk, 2014).

A Public Outlook

Despite the popularity that the food trucks are gaining from all sets of consumers, there can be differences in perceptions, attitude and behavioral intentions that may bring down the preferences of picking a food truck among the mobile food carts, diners and restaurants (Sanlier, Sezgin, Sahin, & Yassibas, 2018; Tester, Yen, & Laraia, 2010; Yoon & Chung, 2018). Few out of many perceived risks in consumer's mind, as elaborated by Yoon and Chung (2018) that "consumers may be concerned that foods prepared by street vendors are typically nutritionally imbalanced and high in fat, because of using poor quality ingredients… including improper food storage, not fresh ingredients, unsanitary conditions, and concern for food poisoning… food waste/water contamination, excessive use of disposables" expresses the concerns on health and environmental risks. Along with risks, there are a series of perceived benefits which primarily include easy accessibility, quick order processing, authentic cooking and "escaping feeling from ordinary life", as quoted by Yoon and Chung (2018).

Hypothesis

The research question for this study is to analyze the factors contributing to the efficiency of a food truck in achieving the maximum footfall and profit. Therefore, the first hypothesis to confirm is, if the age has a correlation with the number of four and above star rating and most reviewed food trucks in a region. The second hypothesis to be tested is, if the ethnicity prominent in a county will have a positive correlation with the number of successful food trucks in that region serving the similar ethnic taste of food.

Methods

The intent of this quantitative study will be to find out whether a business will be successful or not given the influential factors of the business like categories of the cuisine offered, user rating and pricing of the business, demography of the place food truck is operating at like age (ranked) of the residents in a region and predominant ethnicity in attaining a higher efficiency of a food truck. The food truck data was retrieved from Yelp Fusion under endpoint - Business Match (Yelp, 2020a) for New York city by using the distinct API key. The city demographic data was extracted from U.S. Census Bureau, 2018 American Community Survey 1-Year Estimates (Bureau, 2020). This is a statistical study and was conducted at Harrisburg University of Science and Technology from Feb 2020 to April 2020. "One of the many challenges that social science researchers and practitioners face is the difficulty of relating United States Postal Service (USPS) ZIP codes to Census Bureau geographies" (User, 2020). The data to map the U.S. Census data with yelp business data was extracted from the website of Office of Policy Development and Research (PD&R) (User, 2020).

Participants

The participants in this study are the 2271 food trucks across the New York City County (Manhattan, NY) whose business data was extracted from an API request made to Yelp Fusion (Yelp, 2020a). The sample was retrieved by searching tags - “Street Food”, “Food Trucks”, “Food Stands”, “Streed Vendors” and “Food Court” which returned top 1000 records per API request and is a convenience sample. The criteria specified under the request is first the business vendor should be a licensed food truck or vendor, situated with in the county of New York City, i.e. Manhattan, New York, NY and business entry should exist under Yelp.

Procedure

The sample of food truck data in Yelp was provided by the Business owners who gave their consent to distribute their business information with the wider audience as mentioned in the Yelp privacy policy (Yelp, 2020b) and provided the details of their business like address, user rating, price value and business identifying metatags. The yelp business data was cleaned by removing unwanted columns, deriving dictionary column “Address” and “Category” into more specific columns and factoring columns “Pricing” and “Rating” into numeral values. The second source of data is the New York County demographic data compiled by U.S. Census Bureau (Bureau, 2020). “In 2009 and 2010, the U.S. Census Bureau commissioned the collection of two parallel data streams to monitor public reactions to the 2010 Census” (Pasek & Krosnick, 2010. p. 3) which were probability sample telephone survey data and data from non-probability sample Internet surveys and then amalgamating them. The demographic data needed a zip code joining key to connect it with yelp business data which was a challenge as it only contains Tract ID (Moore & Diez Roux, 2006). To address this issue, a third data source, Tract-ZipCode crosswalk files (User, 2020) was used to load the respective zip codes using Census Tracts in the demographic data. The demographic data preprocessing included removing margin of error estimate columns, deriving age columns for three age brackets - “Under 15 years”, “15 - 45 years” and “45 years and above”

Measures

The construct in this research study were the features provided by the business owner and it’s users since they were the decision factors to analyze their correlation with age-wise and predominant ethnicity-wise population segmentation of a region to foretell the rating of a food truck which in turn will decide the popularity of the business. The first construct is the number of reviews which defines the outreach of a food vendor to the consumers and leaving an impact. Yelp food stands were filtered to have 5 or more user reviews to be used in data modeling. The next constructs were if the food truck is operational overnight on weekends, price level of the food offered, is claimed by a business owner, has a phone number on display and offers pickup and delivery options. The attribute which describes if the food truck is operational on weekends is derived based on whether its working hours include Friday, Saturday or Sunday nights, the flag is then set to true. Price level had 5 levels, 0 refers to not mentioned, 1 as economical and 4 being exorbitant. Constructs Title1 and Title2 were the metatags which holds the categories the food cart is segregated under. The next set of constructs are the total population of the region, and its division per age bracket and ethnicity. The ethnicity population is segregated in 4 parts, White, Hispano or Black, Asian and Other races, “Hispano or Black” is combined population of “Black or African American” and “Hispanic or Latino”, and “Other races” include sum of populations of ethnicities “American Indian and Alaska Native”,“Native Hawaiian and Other Pacific Islander” and “Some other race”.

Data Analysis

The data was explored on different attributes and levels to understand the trends and patterns under feature extraction process before moving on to classification machine learning algorithms. It was profiled using python function ProfileReport() under the library pandas_profiling which provided an ability to investigate into the nature of the data. The first step to the analysis was to plot Pearson Correlations graph (refer Figure 1.) among all the attributes involving ordinal values to evaluate the relationship with each other. The correlation does not imply causation and tells how a pair of attributes vary together. The study was done to single out the attributes having a constant value or are highly correlated to another attributes.

The next step of exploratory data analysis was done by plotting bar charts of key constructs in the dataset with respect to business rating and pair grid charts for the different age brackets and ethnicity populations in the regions to understand their distributions variations with zip codes and target variable. Frequency bar charts was plotted (as depicted in Figure 2.) between the Yelp business rating which are distinguished as 1 being poor to 5 considered as excellent, and the following features, (a) price levels categorized under low, medium, high and very high priced food trucks, (b) claimed or unclaimed businesses, (c) pickup option, (d) food delivery option, (e) contact information of business on display, (f) open overnight on weekends, (g) business with more than 100 reviews, and (h) if a direct messaging option is available for the business. Further, a frequency bar chart (Figure 2, Chart I.) was plotted for top 15 food truck categories by pivoting columns - Title1, Title2 and Title3 as a single attribute Title and evaluating the variations with different business ratings. To analyze the relationship between different price levels and business ratings, scatter plots (refer Figure 3.) were plotted for the four ethnicities - White, Hispano Or Black, Asian and Other Races with respect to the total population in a region. Another pair grid chart was plotted for the population under three age brackets with respect to the total population and their distribution for 5 business ratings.

As a last step under feature extraction process, Extremely Randomized Trees Classifier (Extra Trees Classifier), a type of ensemble learning technique which aggregates the results of multiple de-correlated decision trees collected in a "forest" to output it's classification result was used to determine the most important features in the dataset. The number of trees in the forest were kept as 5 and the mathematical criteria to split the trees was used as "entropy". The normalized total reduction value in the mathematical criteria is called the "Entropy Importance" of the feature. Each feature was then ordered in descending order according to the Entropy Importance of it.

The final step before the data modelling was to prepare the data and to cover up for the under-sampling of data with respect to target variable, Over-sampling method in python library imbalanced-learn (imblearn) uses technique - Synthetic Minority Oversampling Technique (SMOTE) to create a more balanced sample of dependent and independent variables. Original target variable, Business Rating had a shape of {5: 127, 4: 545, 3: 202, 2: 21, 1: 4} which was resampled using SMOTE to {5: 200, 4: 545, 3: 300, 2: 100, 1: 50}. The data was normalized and split into train and test in 80:20 proportion to be fed to supervised classifier algorithms and deep neural network.

First, the data was modelled on XGBoost Classifier, a gradient boosting supervised machine learning algorithm using regularization with a learning rate = 0.1, tree depth = 2, Subsample ratio of columns when constructing each tree = 0.8 and 10 gradient boosted trees and the predicted output was analyzed. Second model was created using Support Vector Machine Classifier with radial basis function kernel (RBF) and parameter of the SVC learner = 1 and was fitted on trained data. The third model was Logistic Classifier using multinomial-multiclass, solver = sag which supports L2 regularization, maximum iterations = 1000 and C = 3.5. The next algorithm used was Random Forest Classifier with number of trees (n_estimator) = 100, bootstrapping enabled and max_features set as "sqrt" meaning the number of features to consider when looking for the best split should be the square root of the number of independent variables. The final analysis was done by implementing Deep Neural Network, DNN (refer Figure 5.) with 3 dense layers. The input was converted to one-hot vector using to be consumed into the DNN which used activation = tanh in first 2 dense layers and softmax in the output dense layer. The Stochastic Gradient Descent (SGD) optimizer was used with learning rate = 0.1, decay = 1e-7 and loss function = categorical_crossentropy. The model was then fitted with the training data under epoch = 50 and batch size = 100, sand was then used to predict the results from it using the test data.

Result

The first step was to discard attribute “Is_Closed” since it had a constant zero value and presented no worth in the analysis. The Pearson Correlations graph (refer Figure 1.) showed a high correlation coefficient: r = 0.92928 between “pickup” and “delivery”, r = 0.94043 between “15-44Years” and “Total population”, r = 0.95353 between “45-above” and “Total population”, r = 0.93262 between “OtherRaces” and “HispanoOrBlack” which meant that only one of the two attributes will be included in data modeling.

Under the exploratory data analysis, bar charts were plotted between business ratings and ordinal independent variables (refer Figure 2.) which are concluded as follows; (a) the food trucks with low and medium-priced food menu had better business rating than the ones with high-priced food menu, (b) businesses which were claimed and had their contact information on display had positive linear relationship with the business rating, (c) a business offering pickup, delivery options and open overnight on weekends had no impact on the target variable, (d) more user reviews lead to lowering of overall business ratings, (e) businesses with messaging option enabled are mostly under ratings 4, 4.5 and 5 making it an important feature to be included under data modeling, (f) categories - Food Trucks, Mexican, Food Stands and Food Vendors topped the list having highest business ratings and were the most popular categories among the food trucks in the region. Figure 3. displays the pairwise scatter plot of populations of residents under different age brackets with total population of the region color coded with business ratings which depicted that the regions with higher proportion of younger (i.e. residents aged under 45 years) populations have more businesses with 3 and 4 ratings than the regions dominant with older population, which demonstrated that young consumers are more inclined to try less popular restaurants than the older consumers.

Continuing to the Feature Extraction step (refer Figure 4.), furthermore the least important features in the dataset were filtered out by calculating the importance_value. The graph distinguishes between the permissible and non-permissible features by setting the threshold of importance_value as 0.04. The least important features having importance_value less than the threshold - HispanoOrBlack (0.037306) ,45-above (0.036970) ,Total population (0.033926) ,OtherRaces (0.033375) ,15-44Years (0.032913) ,is_claimed (0.030897) ,has_display_phone (0.030651) ,has_messaging_enabled (0.026274) ,is_overnight_weekend (0.019807) ,delivery (0.019208) ,pickup (0.016750) ,is_closed (0.000000) were dropped from the dataset on which modelling was performed.

After the resampling on the data, the data was modeled using 4 supervised machine learning algorithms; XGBoost, Support Vector Machine, Logistic Classifier, Random Forest Classifier, and a deep learning algorithm (refer Table 1), and the predictions from the algorithms were further compared with each other and hard voted under a custom-made Voting Predictor. The voting Predictor was an effort to make the best predictor by using the mode of predictions made by XGBoost, Support Vector Machine, Logistic Classifier and Random Forest Classifier, and if an equal vote of 2 and 2 is achieved, then it took the prediction made by Random Forest Classifier. The Root Mean Square Error (or RMSE) is lowest for Random Forest as 0.714920. Since RMSE explains about the spread of the residuals or how concentrated the data is around the line of best fit, a RMSE > 0.5 was still very high to call it a good predictive model. Accuracy is the ratio of number of correct predictions and number of total predictions, which was highest for the Random Forest Classifier with 65% accuracy. The confusion matrix provided an insight into the wrong predictions as follows, (a) 17 business ratings = 3 were predicted as 4, (b) 19 business ratings

= 4 were predicted as 3, and (c) 19 business ratings = 5 were predicted as 4. High precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as returning most of all positive results (high recall). Recall is noted to be the highest for Random Forest Classifier as 65%. The harmonic mean of precision and recall is noted as F1 Score, which is highest for again, Random Forest Classifier.

Discussion

The research on Modern Street Food concluded that the business ratings of the food trucks in New York County depend essentially on the categories of foods they are serving among which “Mexican”, “Halal”, “Breakfast and Brunch” and “Chinese” are the most preferable. The business rating also rests on the price level of the food truck, meaning the most popular food trucks are least expensive ones. The hypothesis to identify if age of the residents of a zip code had a significant impact on the business ratings of food trucks in the region is true and inferred that the zip codes with more population of younger consumers have higher number of highly rated food trucks. The second hypothesis to connect the prominent ethnicity in the region with the popularity of the food trucks is also true, and deduced that the White and Asian population is statistically more significant in promoting a better business rating than other ethnicities (Black or African American, Hispanic or Latino, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander and other race).

The outcome of the research indicating that the selection of menu items are the main factors leading to a better efficiency of a food vending business is in line with the research by other authors (Alfiero et al., 2017. p. 14; Martin, 2014. p. 14,15). The food trucks offering gourmet and innovative foods (GFTs) were rated higher than the ones selling traditional food (TFTs). “The average customer satisfaction index and the average equipment index of GFTs were higher than those of TFTs”, as concluded by Alfiero et al. (2017) concurring with the result of this research. The study discovered that millennials are expected to emerge as the most valuable critiques who can potentially influence the public outlook by adding their experiences in the form of user ratings and reviews (Yoon & Chung, 2018). Prior studies also identified that the pricing of the menu offered by a food vendor plays a major role in influencing the consumer (Martin, 2014).

However, there are few measures that could have helped the research into deriving better results and consequently, improved the classification metrics of the machine learning algorithms. First, the study had limited data of food trucks which further reduced after cleansing for data modeling which probably produced skewed or biased results. An attempt of resampling the classes of target variable could have also worked better if the businesses were more. To resolve this issue, the other dense, urban and metropolitan counties near New York County like Kings, Queens, Suffolk and Bronx Counties should have been included in the study. Also, the age population was currently divided into 3 categories, (a) less than 15 years, (b) 15 to 45 years and (c) 45 and above, but the analysis seemed to reject category b and c which were apparently quite broad for the analysis. The brackets for age feature should have been narrower in order to achieve more certainty in the age to business rating correlation and draw specific results. The study missed an attribute of total working hours of the food truck in a week, which could have proved to be a major contender in the data modeling and analysis.

The study can be extended to discover the sentiment of the user review by performing a detailed text and sentiment analysis, and figuring out the top N positive, neutral and negative adjectives used by the consumers in every food category dividing it under the hood of different business ratings. These adjectives will prove to be an essential insight for a new entrepreneur to decide on the different attributes the food truck should pertain to. The study also missed presence of certain business attributes from Yelp API which might have demonstrated superior results. First, the name of cuisines a food truck serves is missing from the dataset since the attribute “category” is more like a meta-tag making it eclectic and diverse. A specific cuisine could be among American, Thai, Mexican or Indo-Chinese, Indo-Thai etc. The average weekly orders processed by a food truck can be added under yelp dataset which will lay out an important insight on the footfall of the consumers at the food trucks.

The modern food shift is making way for convenient, quicker order processing, affordable and ubiquitous food trucks gaining more acceptance and admiration with their innovative food serving practices (Yoon & Chung, 2018). To achieve a higher customer satisfaction and better efficiency, a food truck should focus on the category of food to be served and keep an affordable price level to target more consumers.

References

Alfiero, S., Lo Giudice, A., & Bonadonna, A. (2017). Street food and innovation: The food truck phenomenon. *British Food Journal*, *119*(11), 2462–2476.

Arcese, G., Flammini, S., Lucchetti, M., & Martucci, O. (2015). Evidence and experience of open sustainability innovation practices in the food sector. *Sustainability*, *7*(7), 8067–8090.

Basinski, S. (2014). Hot dogs, hipsters, and xenophobia: Immigrant street food vendors in new york. *Social Research: An International Quarterly*, *81*(2), 397–408.

Bureau, U. S. C. (2020). *U.S. Census bureau.* Retrieved from https://data.census.gov/

Castetbon, K. (2017). Measuring food insecurity. In *Sustainable nutrition in a changing world* (pp. 35–41). Springer.

Donkor, E., Kayang, B., Quaye, J., & Akyeh, M. (2009). Application of the who keys of safer food to improve food handling practices of food vendors in a poor resource community in ghana. *International Journal of Environmental Research and Public Health*, *6*(11), 2833–2842.

Feagan, R. (2008). Direct marketing: Towards sustainable local food systems? *Local Environment*, *13*(3), 161–167.

Krishnasree, V., Nethra, P., Dheeksha, J., Madumitha, M., Vidyaeswari, R., & Lakshya, P. (2018). A pilot study on assessing the sustainability of food safety and hygienic practices in street food handling system. *Asian Journal of Dairy & Food Research*, *37*(4).

Lues, J. F., Rasephei, M. R., Venter, P., & Theron, M. M. (2006). Assessing food safety and associated food handling practices in street food vending. *International Journal of*

Environmental Health Research, *16*(5), 319–328.

Martin, N. (2014). Food fight! Immigrant street vendors, gourmet food trucks and the differential valuation of creative producers in c hicago. *International Journal of Urban and Regional Research*, *38*(5), 1867–1883.

Mokhtar, R., Othman, Z., Arsat, A., & Bakhtiar, M. F. S. (2017). Brand equity and customer behavioural intention: A case of food truck business. *Journal of Tourism, Hospitality & Culinary Arts*, *9*, 561–570.

Moore, L. V., & Diez Roux, A. V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health*, *96*(2), 325–331.

Mukhola, M. (2014). Street-food vending: Training directed at better food handling and associated environmental issues. *The Anthropologist*, *17*(1), 251–258.

Ngo, J. (2012). *Mobile food vending and the public realm: A seattle study* (PhD thesis).

Pasek, J., & Krosnick, J. A. (2010). Measuring intent to participate and participation in the 2010 census and their correlates and trends: Comparisons of rdd telephone and non-probability sample internet survey data. *Statistical Research Division of the US Census Bureau*, *15*, 2010.

Sabbithi, A., Reddi, S. L., Naveen Kumar, R., Bhaskar, V., Subba Rao, G., & Rao V, S. (2017). Identifying critical risk practices among street food handlers. *British Food Journal*, *119*(2), 390–400.

Samapundo, S., Climat, R., Xhaferi, R., & Devlieghere, F. (2015). Food safety knowledge, attitudes and practices of street food vendors and consumers in port-au-prince, haiti. *Food Control*, *50*, 457–466.

Sanlier, N., Sezgin, A. C., Sahin, G., & Yassibas, E. (2018). A study about the young consumers' consumption behaviors of street foods. *Ciencia & Saude Coletiva*, *23*, 1647–1656.

Sevtsuk, A. (2014). Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research*, *34*(4), 374–393.

Tester, J. M., Yen, I. H., & Laraia, B. (2010). Mobile food vending and the after-school food environment. *American Journal of Preventive Medicine*, *38*(1), 70–73.

User, H. (2020). *HUD usps zip code crosswalk files.* Retrieved from https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook

(WHO), W. H. O., & others. (2010). *Basic steps to improve safety of street-vended food. International food safety authorities network (infosan).* INFOSAN Information Note.

Yelp. (2020a). *Business match endpoint - yelp fusion.* Retrieved from https://www.yelp.com/developers/documentation/v3/business_match

Yelp. (2020b). *Privacy policy.* Retrieved from https://www.yelp.com/tos/privacy_en_us_20130910

Yoon, B., & Chung, Y. (2018). Consumer attitude and visit intention toward food-trucks: Targeting millennials. *Journal of Foodservice Business Research*, *21*(2), 187–199.

Table 1

RMSE and Classification Performance Metrics of Different Algorithms

S.No.	Algorithm	RMSE[1]	Accuracy[2]	Precision[3]	Recall[4]	F1 Score[5]
1	XGBoost	0.806226	0.57	0.41	0.57	0.45
2	Support Vector Machine	0.792324	0.60	0.59	0.60	0.55
3	Logistic Regression	0.875595	0.61	0.59	0.61	0.53
4	Random Forest Trees	0.714920	0.65	0.57	0.65	0.61
5	Deep Neural Network	0.819892	0.54	0.36	0.54	0.43
6	Custom Voting Predictor	0.718795	0.63	0.58	0.63	0.54

[1] RMSE or Root Mean Square Error is the standard deviation of the residuals (prediction errors); [2] Accuracy = (True Positives + True Negatives)/ (True Positives + False Positives + False Negatives + True Negatives); [3] Precision = True Positives/(True Positives + False Positives); [4] Recall = True Positives/(True Positives + False Negatives);
[5] F1 Score = 2 * (Recall * Precision) / (Recall + Precision)

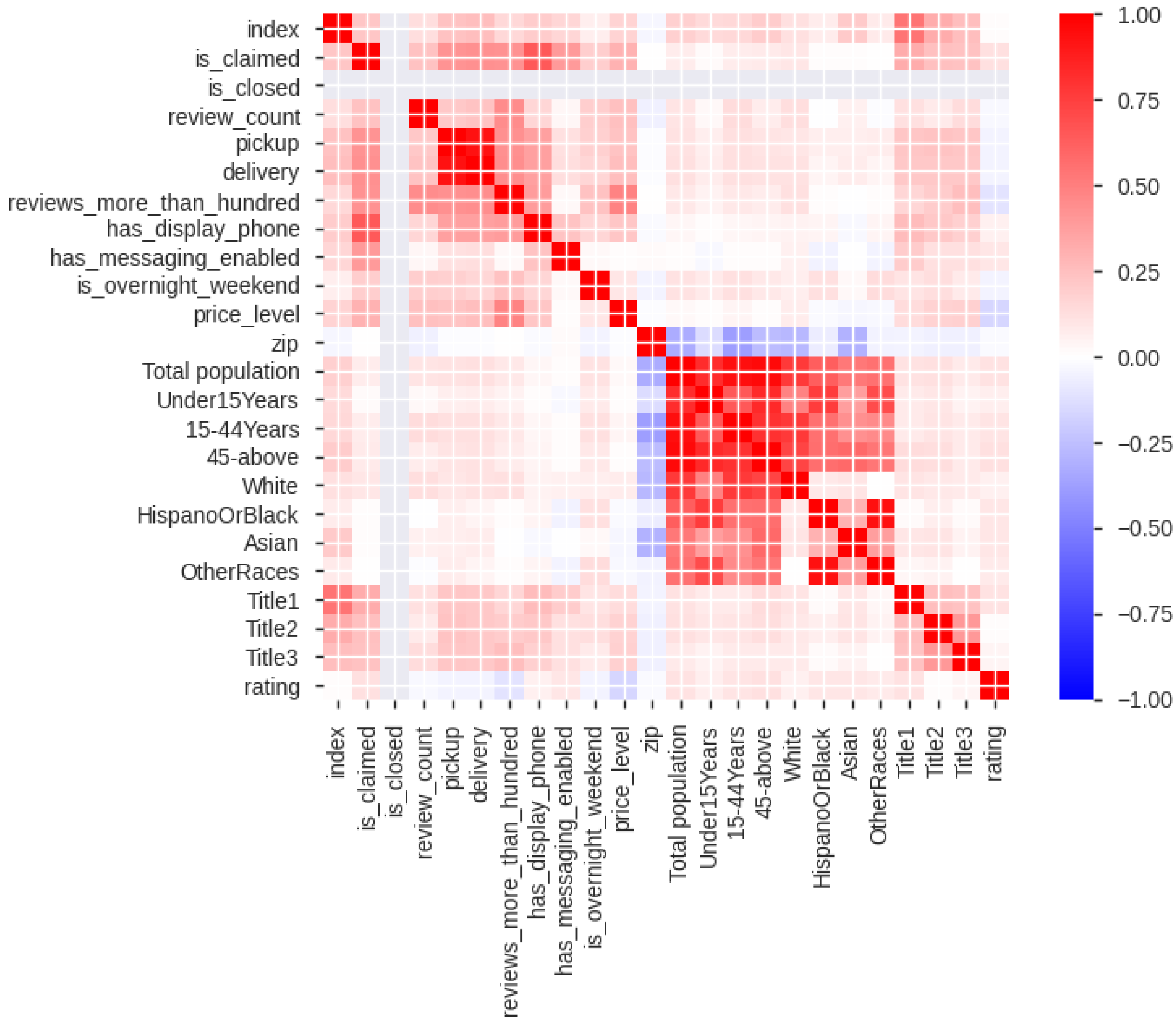


Figure 1. Pearson Correlations of the Features.

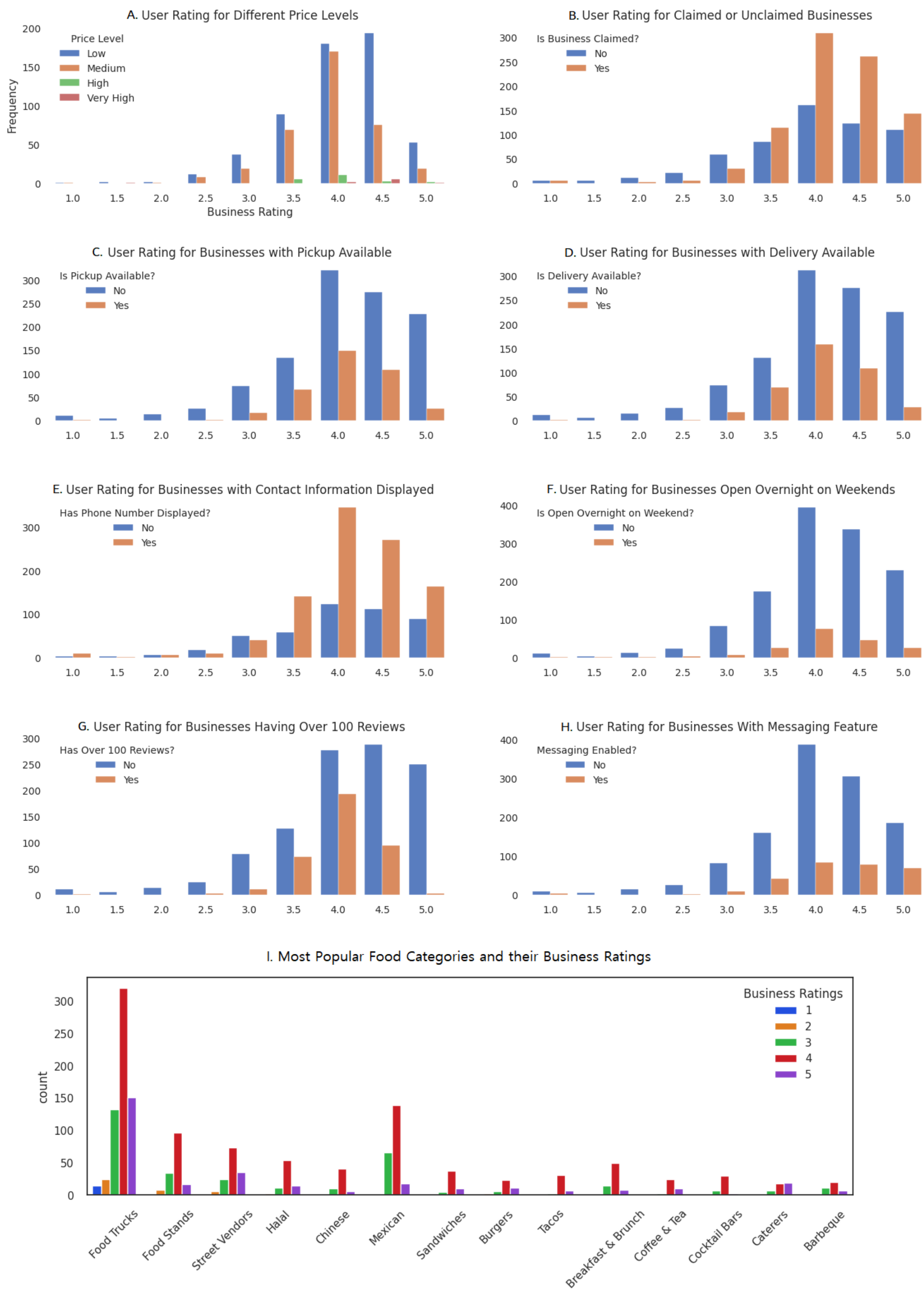


Figure 2. The Relation of Business Ratings with Key Independent Variables.

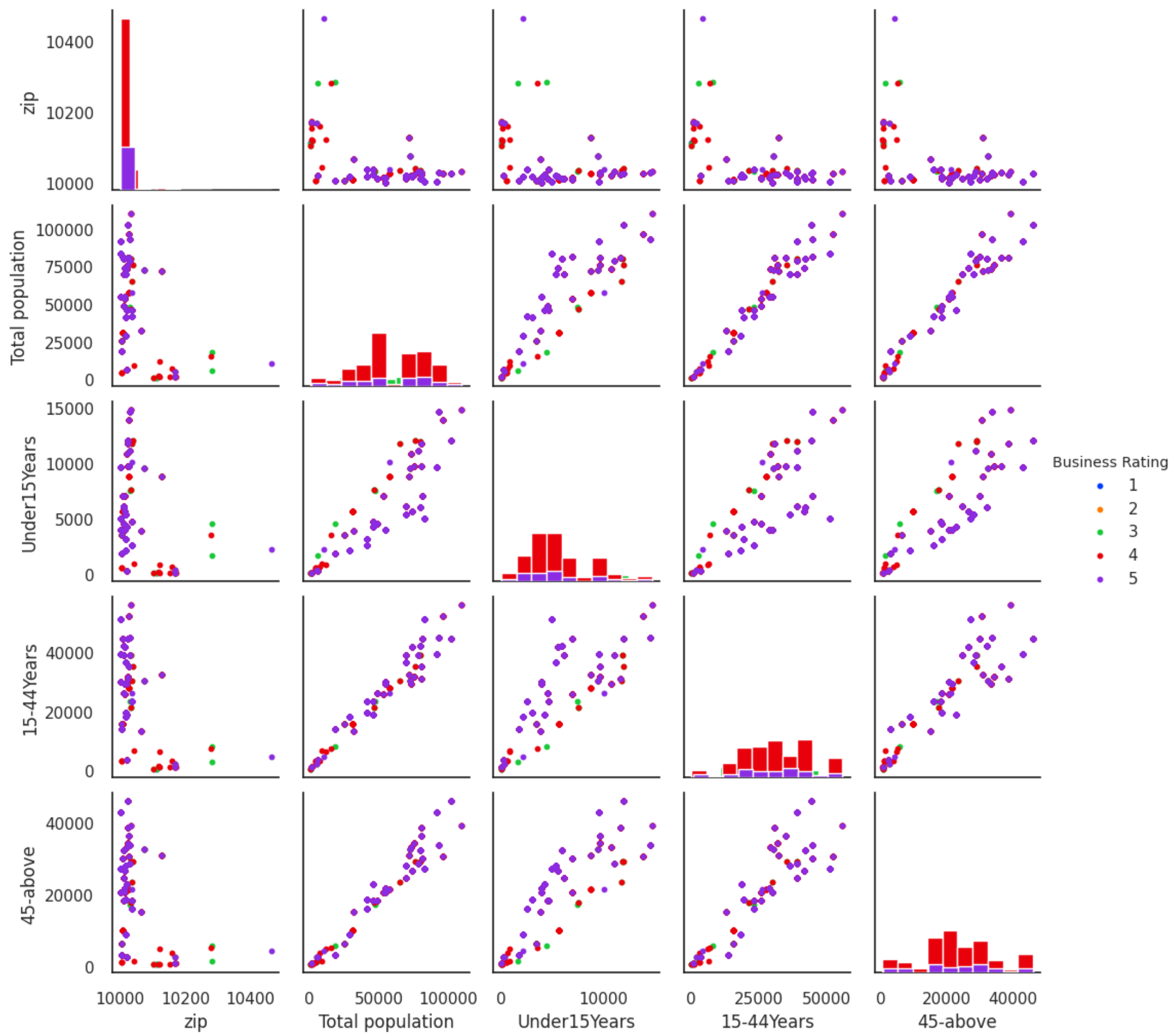


Figure 3. The Trends in Age-wise Population of a Region with the Business Ratings

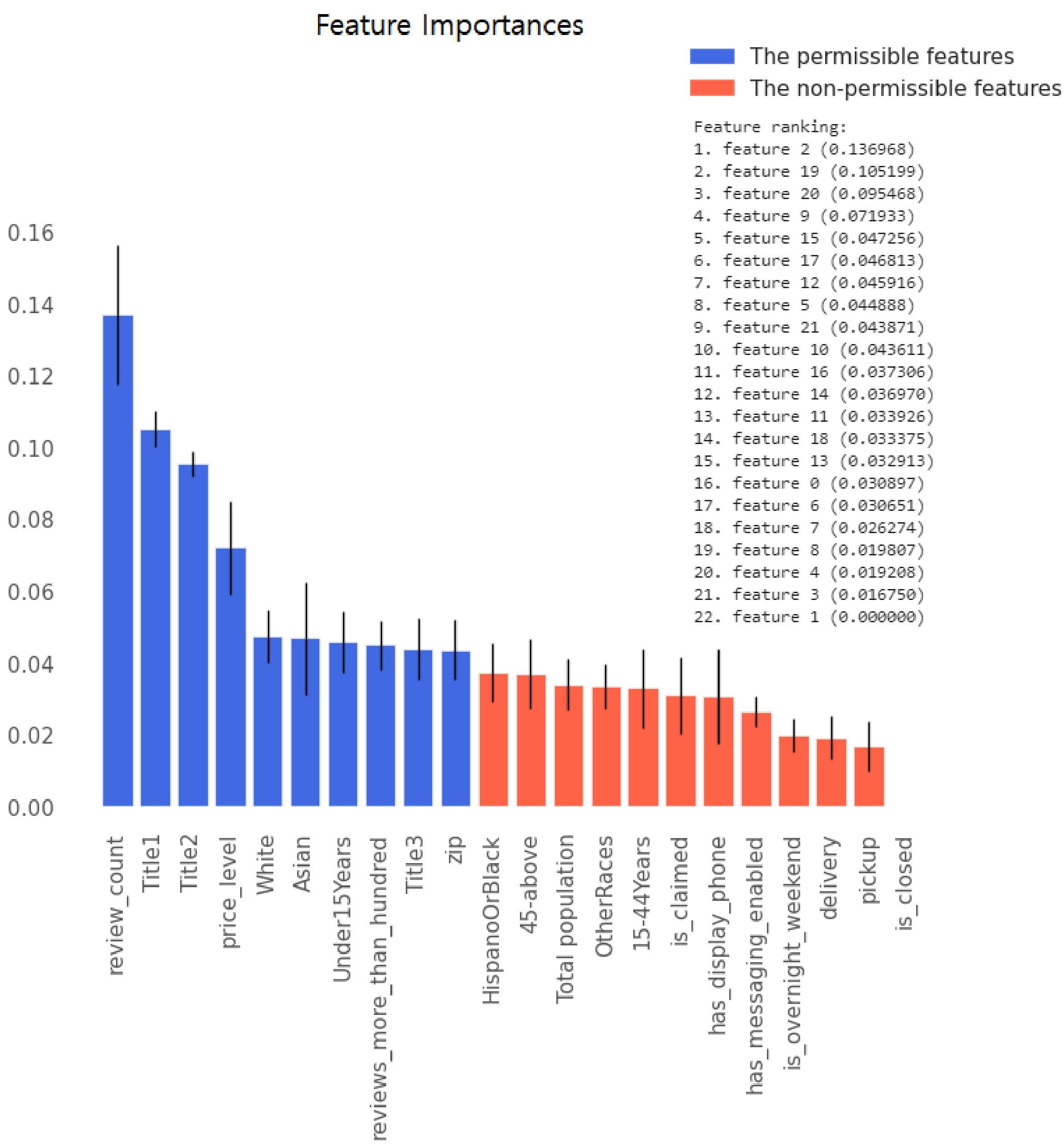
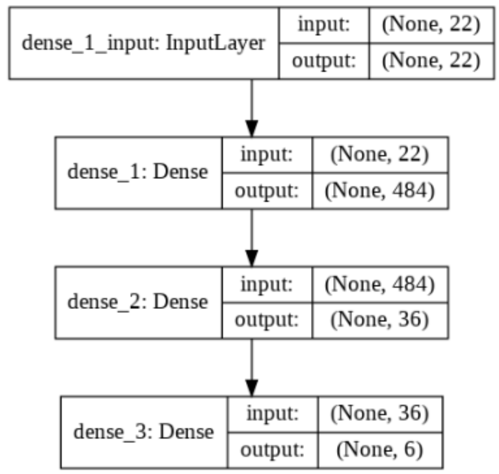


Figure 4. Feature Extraction using Extra Trees Classifier.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 484)               11132
_________________________________________________________________
dense_2 (Dense)              (None, 36)                17460
_________________________________________________________________
dense_3 (Dense)              (None, 6)                 222
=================================================================
Total params: 28,814
Trainable params: 28,814
Non-trainable params: 0
_________________________________________________________________
None
```



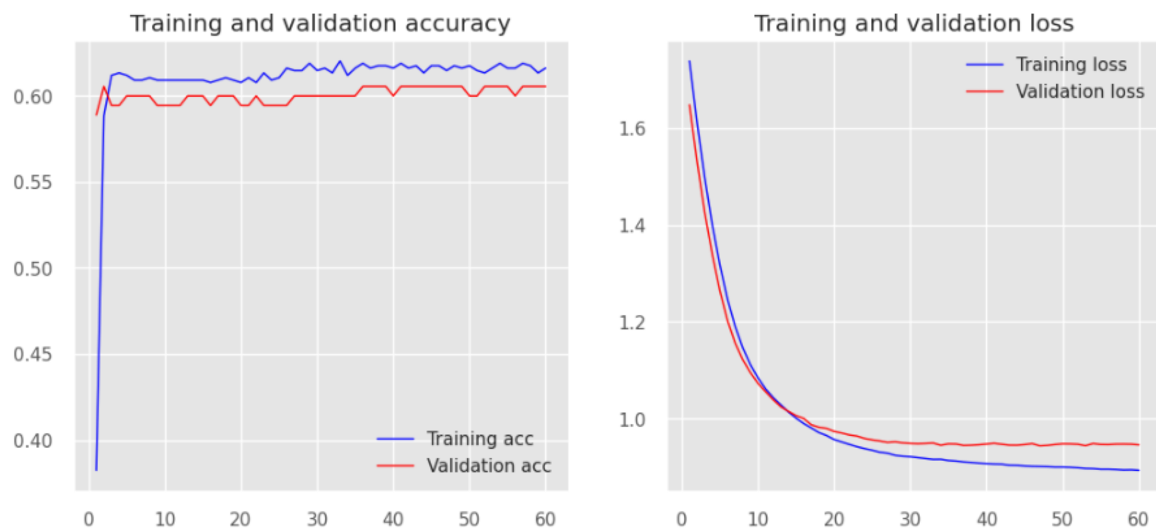


Figure 5. The Structure of the Deep Neural Network.