



DTU - TECHNICAL UNIVERSITY OF DENMARK

42578 - ADVANCED BUSINESS ANALYTICS

Executive Summary - Business Analytics

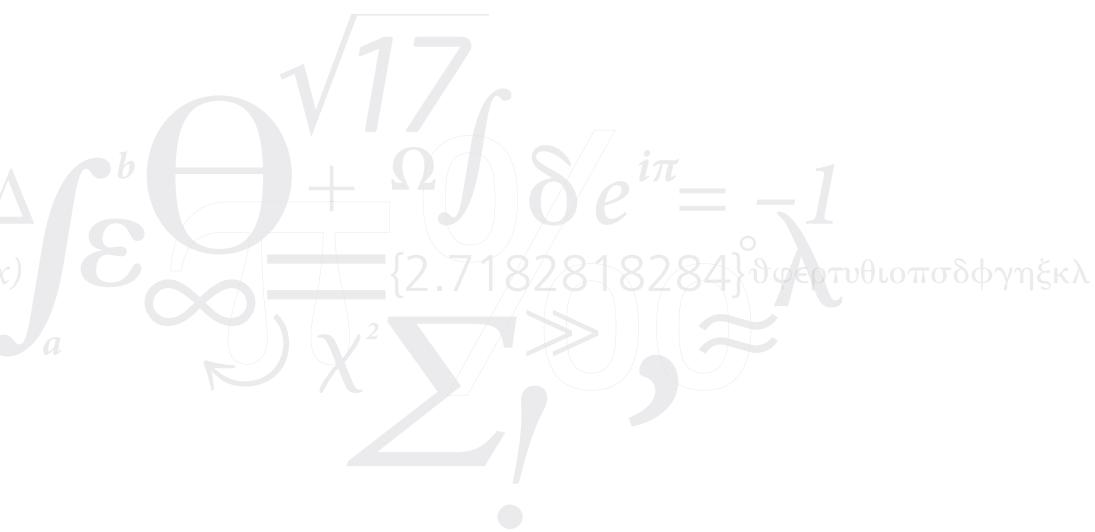
Management Report targeting restaurant businesses in Toronto

Full name:

Clara ANDERSEN
Xin LI
Long QIN
Sebastian SBIRNA
Yingrui Li

Student-number:

s164574
s192275
s165317
s190553
s171353



Contents

1	Introduction	2
2	The source of data	2
3	The target audience for this report	2
4	Business owner perspective	3
4.1	Customer analysis	3
4.1.1	Potential problem: the popularity is decreasing	3
4.1.2	Recommendation: attract returning customers to give reviews	4
4.1.3	Recommendation: use influencer marketing	5
4.2	Competitor analysis	6
4.2.1	Identifying competitors	6
4.2.2	Identifying strengths and weaknesses in relation to competitors	7
4.2.3	Prediction of customer increase by increasing opening hours	8
5	Investor perspective	10
5.1	What makes a restaurant popular?	10
5.1.1	Predicting the popularity class	10
5.1.2	Feature analysis	11
5.1.3	Recommendations-what aspects an investor should mainly take into account	12
5.2	The importance of location	13
5.2.1	Classifier using K-nearest neighbours and Multi-layer Neural Network	13
5.2.2	Recommendation to investor for choosing location	14
5.3	The dangers of opening a restaurant	14
5.3.1	Predicting which restaurants close	14
5.3.2	Recommendations - what factors to take into account when investing in restaurants	15
6	Conclusion	17

1 Introduction

Yelp data provides an effective way to analyze an industry or a specific business based on data sets of business, users and reviews. In this report, we choose the restaurant industry in Toronto to focus on. By conducting a business analysis, we will present several operational recommendations to Kinka Izakaya Original("Izakaya"), which is a Japanese restaurant in Toronto. We will also analyze the restaurant industry in Toronto as a whole, to provide recommendations for investors interested in investing in restaurants in the area..

2 The source of data

The analysis is based on the Yelp open-sourced dataset. We choose subsets which is related to this research. The dataset can be found at: <https://www.yelp.com/dataset>.

3 The target audience for this report

This report consists of two main parts. The business owner part is aimed at the owner and management team of "Kinka Izakaya Original", our business client. The investor part targets at investing companies or individual investors who has an interest in the restaurant industry in Toronto and want to start a new business there.

4 Business owner perspective

Kinka Izakaya Original is a Japanese restaurant at 398 Church St in Toronto which was set up in 2009. This restaurant has been a mature participant in the restaurant industry and has already achieved 4.0 stars on Yelp. Since it isn't the one with top rating among competitors and there is fierce competition in this industry among more than 300 business competitors, it's necessary to upgrade its service quality continuously and expand customers as much as possible to gain more market shares. This report will propose some recommendations to its future operation based on Yelp data analysis in its customers, competitors and location.

4.1 Customer analysis

Generally the more customer reviews a business get, the more popular it is. In the past ten years from 2009 to 2019, Yelp data shows Izakaya has accumulated a solid customers base because it has been one of the most popular restaurant with a high volume of reviews, 1618 in total, on Yelp. Although Izakaya is the Japanese restaurant with most reviews, its star rate is 4.0 which is not the highest in this industry in this city. Therefore the purposes of this report is to help the owner to improve its operation and to be more innovative and unique to thrive in the market.

4.1.1 Potential problem: the popularity is decreasing

By taking a look at how many reviews Izakaya gets every year since 2009 in Figure 1, overall there is a gradually increasing trend but with exceptions in 2016 and 2019. In 2019, there was a 35% drop in the quantity of reviews compared with 2018. On the other hand, shown by Figure 2, the averaged stars of 2019 was 4.00 which hit the lowest level since 2010. Figure 3 also indicates this phenomenon. Though the majority of reviews are equal or above 4.0, it's clearly presented that in 2019 the amount of 3.0, 4.0 and 5.0 which represent neutral and highly positive reviews all reduced, while the quantities of 1.0 and 2.0 which represent negative reviews increased.

Although the overall situation of Izakaya still looks good, but there is a dangerous signal when combining those two aspects to measure its operation in 2019.

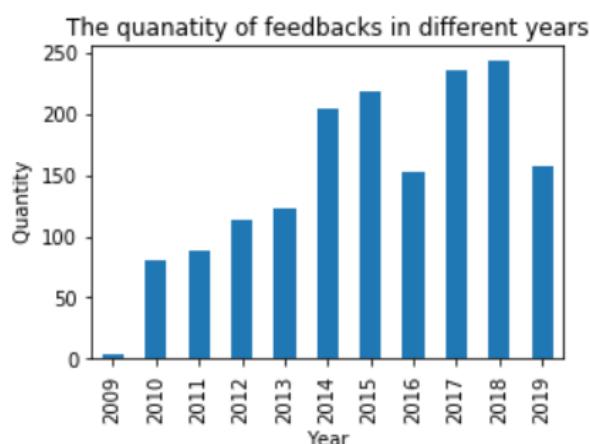


Figure 1: Yearly reviews amount

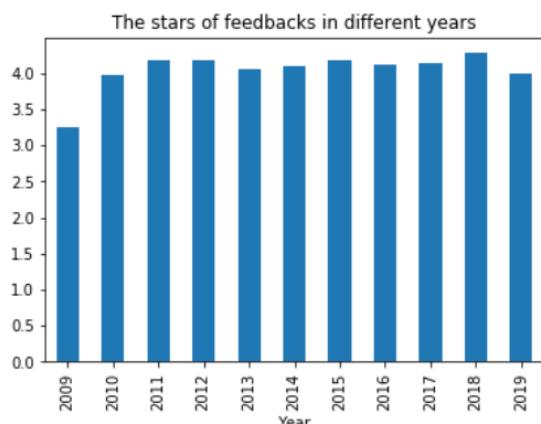


Figure 2: Yearly averaged stars

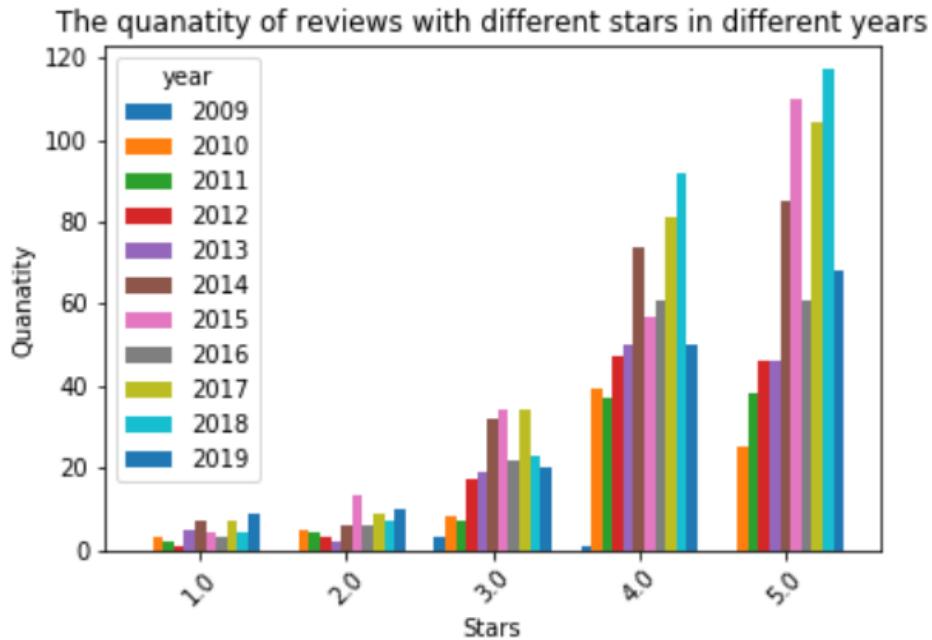


Figure 3: The distribution of historical reviews

4.1.2 Recommendation: attract returning customers to give reviews

Since Izakaya has achieved a very positive reputation regionally, the marginal benefit of traditional marketing advertisements to attract new customers are diminishing. Therefore we suggest to make a better use of present customer resources to enhance the brand loyalty.

The importance of repeat customers for restaurant industry is widely known, one huge value is that their recommendation are more convincing to the public. For example, a customer says: "I have been to this restaurant for several times, the fish there is always fresh", compared with a first-time customer says: "Good place, love their fish".

By analyzing the KPI "Returning feedback provider rate" of our client, which is the percentage of customers who gave more than one review in the past, it shows there isn't many customers who did so. The rate is only 1.61%. Below are several recommendations to maintain a good relationship with present customers and gain more positive feedback from them:

- Offer special coupons to customers and encourage them to give more positive reviews on Yelp.
- Answer every question, doubt or criticism publicly on Yelp and show the passion to welcome them again.
- Send targeted messages and emails to the present customer group regularly and remind them of any update in menu and service.
- Hang up posters or other decorations inside the restaurant to show some interactions with customers, for examples, photos chef takes with loyal customers and a printed screen shot of Yelp conversation with customers and the business owner.

4.1.3 Recommendation: use influencer marketing

By analyzing "Yelp eliteness"¹, "Yelp seniority"² and KPI "Yelp influencer rate"³ of our client's customers, which are some indicators of how business interacts with advanced social media users, it shows that our client didn't attract enough customers with a certain social media influence to give them reviews and recommendations. Because the analysis shows that, in the past 11 years, only 3.4% of customers' reviews are offered by Yelp users with more than 10 years seniority. The Yelp influencer rate is merely 0.75%. And 73.11% of reviews are given by 1-year elite users.

Considering social media marketing is playing more and more important role compared to traditional marketing tools, we highly recommended Izakaya seek cooperation with influencers both on Yelp and other social media platforms. A social media influencer is someone who wields that influence through social media[1]. Yelp is a relatively small platform, so we think the one who owns more than 100 fans is already influential on this platform. But influencers for bigger platforms like Instagram can have millions of followers. Influencer marketing is to develop cooperation with these people who can create visibility for product or service by posting on social media. The right influencer can reach the target audience, build trust, and drive engagement[1].

Here are some approaches to set up influencer marketing:

- Always find influencers whose focus aligns with food and especially Japanese food.
- Asking influencers to review the food and service on Yelp. If he / she is happy with Izakaya, it gains a great endorsement that's shared with his/her followers.[2]
- Invite influencers to the restaurant, ask them to write a detailed guest post on their website or other platform. This allows our client to show its food and service directly in front of the influencer's audience.[2]

¹Yelp's way of recognizing users who are active in the Yelp community and who are role models on Yelp site. Users are measured yearly and will be given a special sign in their account if they are admitted as elite user.

²How many years has a user used Yelp

³The percentage of customers who have more than 100 fans in Yelp

4.2 Competitor analysis

4.2.1 Identifying competitors

When trying to assess the performance of a business in its own market, a vital part in understanding how well a business is doing is by understanding its own competitors, and being able to assess which competitors are most important to keep close watch on.

In our analysis, we have created different metrics which can be used to detect what our client's competitors are. We will detail them below:

- The number of matching business categories that the client and the competitor have in common
- The popularity of different competitors, computed as the number of total reviews for a business multiplied by the average rating of those reviews
- Proximity competition, defined as the length (in kilometers) from the client's location in Toronto to that of a competitor (assuming a straight path between them)

These metrics are relevant, since they help filter out the Japanese business restaurants from Toronto to only a relevant subset. Out of those metrics, the most important one is the number of matching business categories, since having market intelligence about the products and target clientele of a business means that you can define who your **direct competitors** are.

Since our client has 7 total business categories that it tackles, we will consider as direct competitors: those businesses who match 5, 6 or 7 out of 7 categories identically with the client, since this means that they are tackling the same market base.

From a further lookup and analysis of our client on Google Maps, we can see that the client is situated in the busy Willowdale area of North Toronto, on a restaurant-packed street situated next to a commercial center. Therefore, it seems reasonable to consider the area in which our client is located for finding more potential competitors to its business. Since Willowdale is approximately 3 kilometers in radius (6 kilometers in diameter from one side to the other), we will consider any restaurant within 3 kilometers of the exact location of Kinka Izakaya to be a competitor.

Lastly, popularity is overall a difficult term to define, however it is essential to understand that a large-scale, business-class restaurant and a family-owned small-scale restaurant do not compete on the same terms, even if they tackle a similar customer market.

During our analysis, it has been revealed that our client is by far the most popular Japanese restaurant in Toronto, and even more, it is second in popularity across all businesses in Toronto (bested only by a Thai-cuisine restaurant, which falls outside our analysis scope).

The average popularity score is around 300, and our client's score is around 6300, which is **21 times higher** than the average. Clearly, at the time of the collection of dataset, the client was achieving very good review scores from a large amount of population, however perhaps there is still something to learn from other Japanese restaurants with slightly lower popularity.

Since the popularity average is 300, we will consider any Japanese restaurant with a popularity score of over 1000 to be a popularity competitor with our client, since this means that, on average, the restaurant must have had e.g. more than 250 reviews of 4.0 stars, or an equivalent number, in order to achieve this threshold. The number 1000 was selected since most of the Japanese restaurant businesses in Toronto are smaller-scale, family restaurants (according to the score calculation).

At the end of our competitor discovery part, we have found that our client is competing, directly or indirectly, with 337 out of 615 Japanese restaurants, representing a proportion of 54.8% of the total number of Japanese restaurants. Table 1 below summarizes the number of competitors found, relative to each of the different metrics mentioned above:

Competitor metric	Nr. of competitors
Any direct competitors with 7/7 client categories matching	1
Any direct competitors with 6/7 client categories matching	1
Any direct competitors with 5/7 client categories matching	32
Any competitors with similar popularity score as our client (ie. > 1000)	28
Any proximity competitors located within 1km radius of our client's location	79
Any proximity competitors located between 1 and 2 km of our client's location	134
Any proximity competitors located between 2 and 3 km of our client's location	62

Table 1: Distribution of the number of competitors found for our client, according to the different metrics defined for discovery

4.2.2 Identifying strengths and weaknesses in relation to competitors

TO gain a competitive advantage it is essential to understand the strengths and weaknesses of a business, especially in relation to its competitors. To get a better understanding of these, the reviews of Kinka Izakaya Originals itself, and its most important competitors are therefore analysed. By identifying the overall topics of the negative and positive reviews, an indication of what the customers experience as positive and negative traits are determined, and thereby also the restaurants competitive strengths. In table 2 it can be

Competitor		Positive reviews		Negative reviews	
Name	Importance	Number	Topics	Number	Topic
Ogadang	7	17	Yakitori	0	-
Yuzu Izakaya	6	26	Skewers, Salmon, Soup and salat	1	Pricing
-	5	2742	International cuisine and serving alcohol	195	Service
Client		Positive reviews		Negative reviews	
Name		Number	Topics	Number	Topic
Kinka Izakaya Original		1528	Atmosphere, service, sashimi, salmon and fried food.	55	Service, food and noise level

Table 2: The overall topics of negative and positive reviews are listed for the three most important groups of competitors, as well as for Kinka Izakaya Original.

observed that tyakitori and skewers are mentioned for both of Kinka Izakaya Originals most important competitors. This indicate that the tyakitori appear to be a popular dish amongst the customers for which Kinka Izakaya Originals is competing. To attract customers from these main competitors it can therefore be advised that tyakitori is added to the menu if not already available, or that marketing is initialized to highlight that the dish is available.

Kinka Izakaya Originals have 32 competitors with an importance score of 5. Based on the topics of the reviews for these competitors, it is clear that Kinka Izakaya Originals is primarily competing with these restaurants for customers interested in Kinka Izakaya Originals as a bar, and for customers interested in international cuisine, who are attracted by its identity as a Japanese restaurant.

At last it can be noticed that while the atmosphere of Kinka Izakaya Originals is a topic of positive reviews, and therefore can be consider as one of our clients strengths, the noise level of the restaurant is also the topic of negative reviews. By initializing initiatives to bring down the noise level within the restaurant, Kinka Izakaya Originals can further improve the experience of the customers, and thereby strengthen its competitive position further.

4.2.3 Prediction of customer increase by increasing opening hours

Since our client's business opening hours are 17:00 - 00:00 for every day of the week, we have performed a predictive analysis to understand what would the client's customer influx be, if it would open for business one hour earlier (i.e. from 16:00 to 17:00).

In order to solve this problem, we have used the sparse dataset of checkins from the Yelp dataset, and have aggregated checkins per the day of the week, in order to compute weekday-dependent proportions on what percentage of the customer influx is happening between those hours for competing restaurants which do have open during those times.

We have used multiple predictive models, including a baseline, and have recorded the errors in prediction for our models upon the existing data. Since our prediction could only include customer data about a Japanese business in Toronto, our dataset was very limited in size, and that will result in a larger margin of error that we have to work with. In Table 3 and Table 4 below, we present our results for predicting the total proportion of customers that would visit our client during the increased opening hours (which is a percentage specifying how much of the customer influx of a weekday will happen between those hours), together with a "margin of error" percentage, indicating an interval of values within which we can be most certain that the true customer influx proportion would be present in, depending on how accurate the model is.

Weekday	Dummy Regressor (pred.)	Linear Regression (pred.)	Neural Network (pred.)
Monday	17.06%	18.31%	5.95%
Tuesday	17.06%	16.64%	5.94%
Wednesday	17.06%	16.92%	5.94%
Thursday	17.06%	16.36%	5.94%
Friday	17.06%	16.39%	5.93%
Saturday	17.06%	12.35%	5.92%
Sunday	17.06%	13.38%	5.92%

Table 3: Predictions of customer influx proportion during 16:00-17:00, separated by weekday

Prediction Model	Margin of error
Dummy Regressor	$\pm 14.0\%$
Linear Regression	$\pm 10.2\%$
Neural Network	$\pm 9.8\%$

Table 4: Maximal error margins for the predictions of the specified models

The final conclusion of the predictive modeling has been that there is a clear weekday-dependent trend in the customer influx for the late-afternoon period between 16:00 and 17:00, with Mondays being the most profitable day to open up business earlier and Saturday/Sunday being the least profitable period in the week for those hours.

Our models have agreed that there is a potential increase in the number of customers that can be received by our client, however their predicted proportion amounts being quite different from each other, both due to the different predictive algorithms used, but also due to the lack of a large dataset, which could help in minimizing error margins.

5 Investor perspective

Based on the analysis of yelp reviews for restaurants in the Toronto area, we will be investigating which factors are important in determining the success of restaurants in the area. This will form the basis for a number of recommendations, regarding what you as an investor should take into account, when considering to open a restaurant in Toronto.

5.1 What makes a restaurant popular?

As an investor, the most intriguing and valuable aspect to be explored before investment is to investigate the factors that influence the popularity of a restaurant. Based on that, one can gain the business insight to some extent and make decisions more accurately.

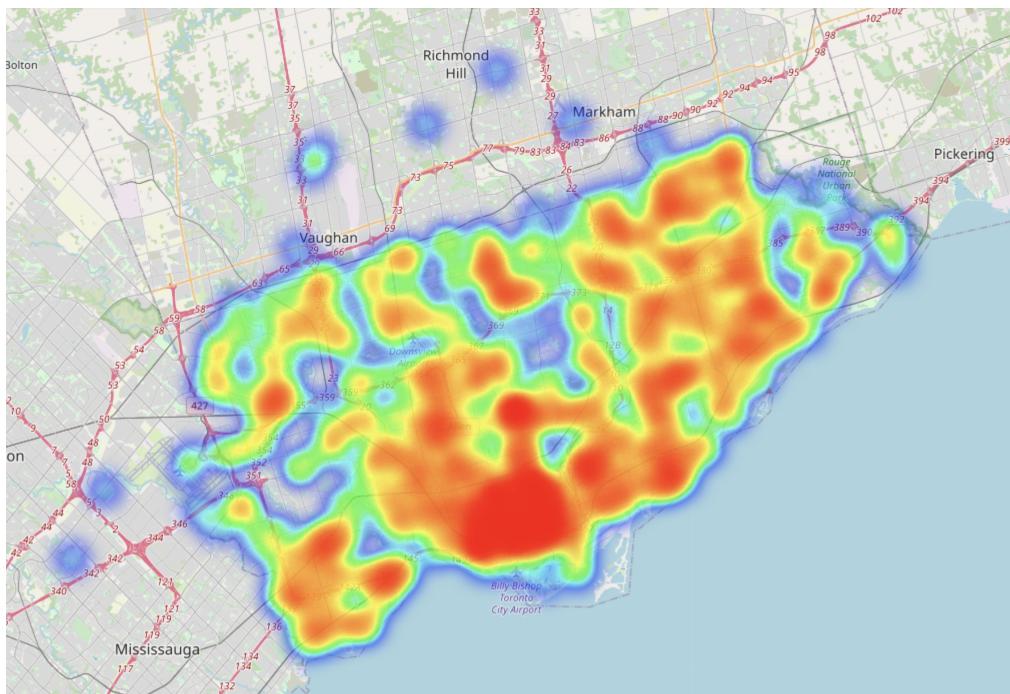


Figure 4: Heat map of restaurants density

It is natural to assume location plays a crucial role when speaking of opening a business. From the map one might not be able to notice obvious clustering except the denser distribution of restaurants close to the sea(airport and university), hence the influence of location will be discussed solely in next section whereas all the features will be considered here for analysing their contribution of popularity.

5.1.1 Predicting the popularity class

The most representative indicators of popularity of restaurants are assumed to be the rating score(reflect customer evaluation) and the amount of reviews(reflect the number of customers). Instead of combining their influences in into a continuous response for prediction, the top 15% quantile of both rating and review counts are considered as high level and left 85% are low level. In this way, four classes generated in the light of rating(high or low) and review counts(high or low) and several classification approaches are implemented for prediction.

Method	Dummy classifier (random guess)	Logistic regression (white-box)	XGBoost (black-box)
F1 score	0.236	0.496	0.515
Accuracy	0.390	0.618	0.635

Table 5: F1 score and accuracy of methods used for popularity class prediction

From table 5, one can notice that both logistic regression using one-vs-rest and XGBoost can improve the performance of prediction, while XGBoost does not outperform logistic regression to large extent. Limitation here would come from high imbalance data structure(all the original parameter inside the dictionary has been reserved leading to a sparse feature matrix), less appropriate missing value imputing(most of missing values are assumed as False which might not be consistent with reality) and manual definition of popularity class(might not be meaningful enough). However, both methods enable prediction model to output valuable results in terms of f1 score and accuracy. In next step, feature analysis will be done so that further interpretation from business perspective can be obtained.

5.1.2 Feature analysis

Known as a stable machine learning method, XGBoost lacks interpretability as a result of aggregating plenty of weak learners. SHAP is applied to tackle this problem as below:

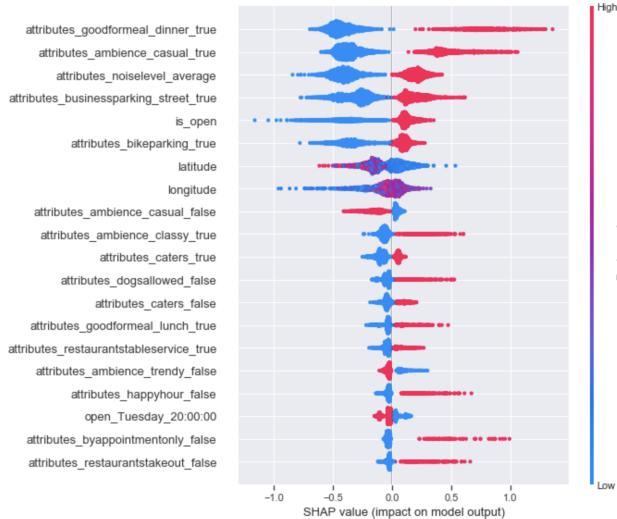


Figure 5: Summary plot of feature importance of high rating and high review counts class

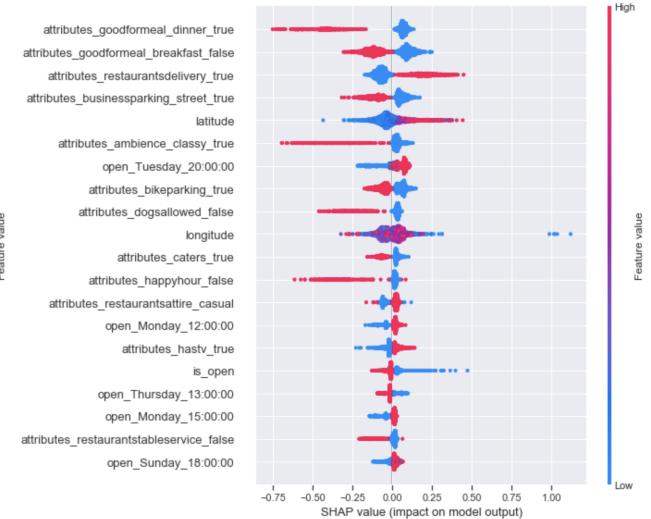


Figure 6: Summary plot of feature importance of low rating and low review counts class

The figures above list the most important features out of total 751 parameters, their contribution to prediction decreases from top to bottom. The classes high rating with high review counts and low rating with low review counts are selected for a sharp contrast when interpreting feature importance. For example, the most significant feature is representing if the restaurant has good dinner, in figure 5, the high feature value(indicating 1 or true as a binary variable) will more likely show that a restaurant belongs to high rating with high review counts class; while in figure 6 the opposite is shown: if a restaurant serves good dinner, in less chance this restaurant will be in low rating with low review counts class.

Some of the parameters seem to be not very meaningful, for instance if “attributes_happyhour_false” in figure 5 is 1 meaning happy hour is false, the positive influence will be obtained which reflects if the customers

do not spend happy hours in a restaurant, then this restaurant tends to be high rating as well as having high review counts. However, we will notice that this attribute is missing plenty of values so that it has been imputed manually(fill all the empty data points using 0) if take a further look at the original data. Hence, if only focusing on the feature importance, one might get biased conclusion since data itself is extremely sparse and based on numerous assumptions. Gathering more data or dropping parameters with too many null values might improve the performance and give better interpretation.

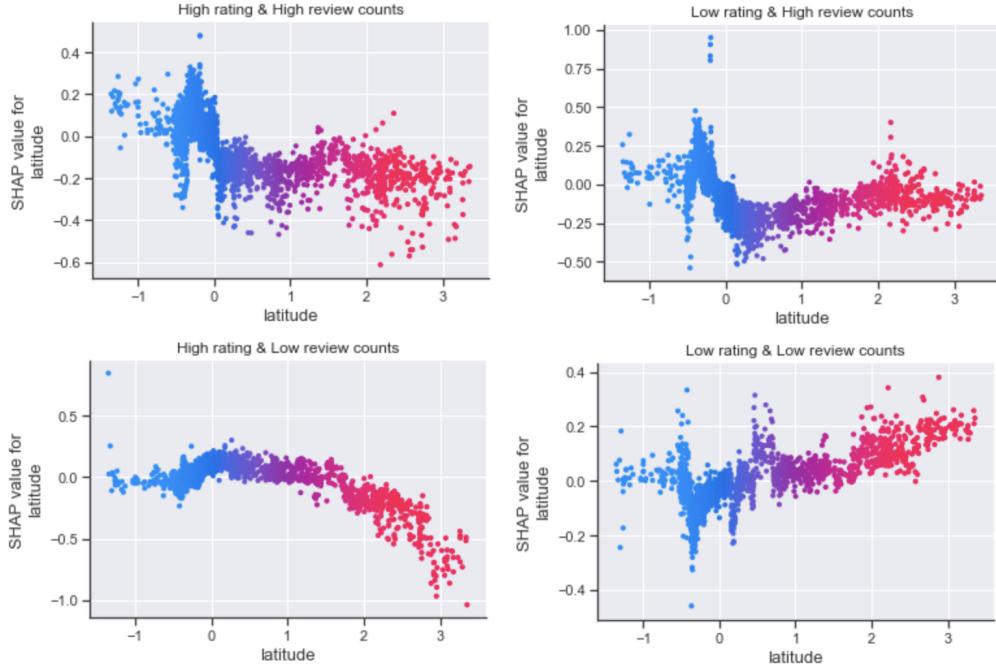


Figure 7: Partial dependence plot of latitude

Figure 7 shows the dependence plot of latitude within four classes respectively, but no clear trend can be observed since the locations have not been transformed into the proximity. Further analysis focusing on location and distance among locations will be done in next section.

5.1.3 Recommendations-what aspects an investor should mainly take into account

Probably because of the diversity of the restaurants' type(Chinese, Nordic, French, etc.), the categories of restaurant seem not to contribute much to prediction. Hence the most crucial features are basically locations and attributes that many restaurants share in common. If an investor is willing to invest in certain type of restaurant, they can dive into that aspect and do extensive analysis to gain more perspective.

Here the recommendations that we can provide are quite general to give anyone wants to open a restaurant:

1. Providing high-quality dinner will increase popularity.
2. Casual ambience seems to be attractive factor to customers.
3. Having business parking near street would probably make customers comfortable.

All the recommendations from above are given both because of the analysis and domain knowledge(or common sense sometimes), and one should always remember to take assumptions of original dataset into account to avoid biased understanding.

5.2 The importance of location

To undertake above, whenever opening a new restaurant at any area, the selection of location has to be seriously took into account. Although we have been given the geo-location data, an evaluation judgment of success is missing. Both stars of restaurants and their total value counts are not enough for overall importance as they are not directly related. In that case, a initial success score has been designed as below.

$$\text{Successscore} = \text{isOpen} * \text{avgRating} * \text{reviewCount} / 10$$

As the heat map above 4 from which may not be able to notice clustering well, we include the success score in below heat map where each circle represent a restaurant in Toronto and the color of circle stand for success score. As the color form pink to black, restaurants considered more successful.

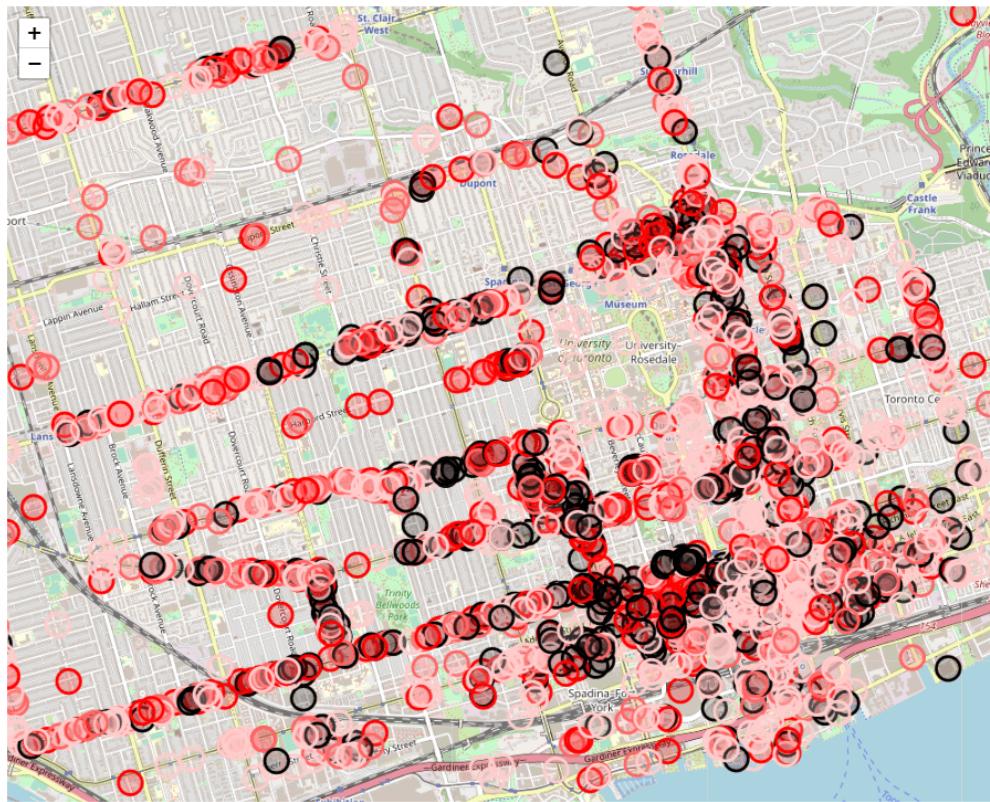


Figure 8: Heat map of circles as restaurants.

From the map we can get a hint that the most successful restaurant will located nearby coast side However, the `isOpen` attribute forced the all closed restaurants with 0 success score. It's not realistic as there could be complex reason for owner to close the restaurant. As an optimal option we would use the popularity class instead which been introduced before. More importantly, the crucial of location feature couldn't be obvious clearly.

5.2.1 Classifier using K-nearest neighbours and Multi-layer Neural Network

To give a relatively trustful suggestion for investor, we would use the 5-nearest restaurants' distance and their popularity class as features for each restaurant. This will enable us to investigate the relation of popularity of each small area. At this stage, a K-nearest neighbour(KNN) classifier will be a good tool coming into our mind as we have clearly popularity class with distance as factors.

Factors	KNN(geo only)	KNN(geo+distance)
accuracy	0.250	0.506
best_accuracy	0.282	0.545

Table 6: Result of KNN classifier for location factors

From table 6, the clear distinction from accuracy give credits for including distance of nearby restaurants as features when classifying. For the KNN of only geo-location, we only using latitude and longitude as input feature, after the execution of distance, we have 12 input feature into as latitude , longitude, distance of 1-nearest restaurant, popularity class of 1-nearest restaurant. However the result is not good enough for an investor to choose the location as it will take a risk. From that perspective, considering that we have 13 input when using distance as factors, the amount of features are probably too complicated for KNN classifier to handle. Therefore, to give a better suggestion of location prediction for investor, a simple Multi-layer Neural Network has been used.

	Neural Network
Best Accuracy	99.36%
hidden layer	2
num.classes	13

Table 7: Result of Multi-layer Neural Network for location factors

As we can see from table 7, the accuracy from multi-layer neural network classifier is pretty high as we running with a batch size 12 with 50 epochs, which is an obvious improvement comparing to K-nearest neighbours classifier. It turns out that the neural network has a better understood when input features are large.

5.2.2 Recommendation to investor for choosing location

By using different models based on geo-location data and calculated distance stats, the location is undoubtedly playing a crucial role when selecting a restaurant to investigate. An investor is recommended to using our provided Muti-layer neural network model for choosing ideal location of restuarant.

5.3 The dangers of opening a restaurant

Out of all registered restaurants in Toronto, 37% have had to close. If a restaurant is forced to close, you as an investor will suffer a financial loss. When investing in a restaurant, it is therefore important to be aware of what factors can lower the risk of having to close, thereby minimizing the risk of the investment.

In this section the focus is on identifying which features are most influential when it come to affecting whether or not a restaurant are forced to close.

5.3.1 Predicting which restaurants close

Initially it is necessary to determine whether or not any underlying structures are present, influencing if a restaurant ends up closing. To investigate this, two models was build to predict if a restaurant is closed or not based on the restaurant features. These features describe what categories a restaurant belongs in, which services are offered, which facilities are available, when the restaurant is open, the average rating received on yelp, as well as the number of received reviews.

Model	Baseline	M1	M2
Method	Random selection	Logistic regression	Decision tree
Accuracy	0.49	0.74	0.73

Table 8: Accuracy for the models predicting if a restaurant is closed.

Table 8 shows that it is possible to improve the prediction of whether or not a restaurant is closed, as the accuracy for both M1 and M2 are improved compared to random guessing. As the highest accuracy was obtained using M1, the further analysis focuses on investigating this model.

5.3.2 Recommendations - what factors to take into account when investing in restaurants

M1 are taking a total of 228 different features into account when determining whether or not a restaurant is closed.

Out of all features, the features listed in table 9 are indicated to individually lead to the highest increase in the chances of a restaurant having to be close. The higher the odds ratio are for a feature, the more the feature is expected to increase the odds of a restaurant closing. To lower the risk, when investing in a restaurant, it is recommended to avoid any specific dress-code, as all required attire is indicated to increase the chances of a restaurant having to close. More specific concepts which it is advised to avoid is waffle restaurants, ice cream and frozen yogurt shops, Peruvian restaurants as well as lounges.

Feature	Odds ratio
Formal attire	4.96
Casual attire	3.78
Divey atmosphere	3.18
International	2.84
Music venue	2.78
Ice cream and frozen yogurt	2.6
Waffles	2.53
Peruvian	2.52
Dressy attire	2.37
Lounges	2.26

Table 9: The features which by M1 is indicated to lead to the highest increase in the chances of a restaurant having to close.

In table 10, the features which are indicated to lower the odds of a restaurant closing the most are listed. Here the closer the odds ratio is to 0, the more the feature is expected to decrees the chances of a restaurant having to close. The lowest odds ratio is found for the number of reviews a restaurant receives. While this highlights the importance of encouraging customers to review, to lower the risk of a restaurant having to close, it is also an indicator of the popularity of a restaurant. This underlines the well-known fact that a restaurant unable to attract customers will be forced to close.

Similarly, the average star rating is an indication of customers satisfaction with a restaurant, and it is therefore not surprising that a higher average star rating will lower the chances of a restaurant closing. For

specific concepts breweries, donuts shops, and Filipino and fast food restaurants are recommend for potential investments, in order to lower the risk. Creating a overall hipster atmosphere appear are indicated to lower the risk of closure. At last choosing providing bike parking or having a TV available is also indicated to lower the risk of having to close, and there by the risk of the investment.

Feature	Odds ratio
Number of reviews	0.0001
Hipster atmosphere	0.11
Having a TV	0.2
Donuts	0.21
Breweries	0.27
Fast food	0.3
Filipino	0.36
Bike parking	0.51
Avg. star rating	0.78

Table 10: The features which by M1 is indicated to lead to the largest decrease in the chances of a restaurant having to close.

6 Conclusion

Data analysis is helpful to improve the operation of the business and find a new investment chance. Data could elaborate not just which segment in the market to choose and where are targeted customers but what is the present problem in the operation and how to make differentiation among competitors as well.

By doing business analytic in the case restaurants and the whole industry in Toronto, we conclude as below:

- The way of marketing events has changed a lot. Business owners should pay close attention to their word of mouth on social media platforms.
- The competition in restaurants is heavy. The secret of success is to maintain the top level quality of foods and services while keeping unique characters at the same time.
- Location is always vital. Based on the prediction model classifier, investors can choose an ideal location of restaurant to investigate.
- There are some types of restaurants that are easier to be out of business than others. So be wise to choose the right business to start.

References

- [1] Christina Newberry. *Influencer Marketing Guide: How to Work With Social Media Influencers*. 2019. URL: <https://blog.hootsuite.com/influencer-marketing/>.
- [2] Brian Sutter. *What You Need To Know About Marketing With Influencers*. 2016. URL: <https://www.forbes.com/sites/briansutter/2016/04/08/what-you-need-to-know-about-marketing-with-influencers/#1ebedf99167b>.