SHOULD I OPEN HERE?
PREDICTIVE MODELS FOR RESTAURANT SITE SELECTION

GUY LLEWELLYN
PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University
School of Hotel and Tourism Management

Should I Open Here?
Predictive Models for Restaurant Site Selection

Guy Llewellyn

A thesis submitted in partial fulfilment of the requirements for
the degree of Doctor of Philosophy

January 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Guy Llewellyn

Abstract

Restaurant failure can be considered from two different perspectives, the pre- and post-opening decisions. While the post-opening factors include the day-to-day operating decisions, the pre-opening decisions are the ones that a restaurateur makes in the planning and design of a restaurant; with an initial decision of the site location (Chen & Tsai, 2016; Yang Yang, Roehl, & Huang, 2017). Site locations are often made by intuition, as there is a lack of comprehensive research about which elements are the most important for a restaurant's future success. The decision of where to open a restaurant is one of the most critical pre-opening decisions for restaurateurs (Egerton-Thomas, 2005).

The current site selection method often includes engaging an expert with a 'gut feel' about a potential site's success (Clarke, Horita, & Mackaness, 2000). However, if data can be used to gain a more thorough understanding of success factors related to the site, the success rates may improve (Wood & Reynolds, 2012). Considering the knowledge gap between the pre-opening area factors and potential success of the site, the research questions are as follows:

1. Which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure?
    a. What is the most critical socio-demographic area attribute a restaurateur needs to pay attention to when selecting an area?
    b. What is the most critical restaurant site characteristic that a restaurateur needs to consider when selecting a site?
2. Considering the potential influence of the overall site characteristics, can a model be created to aid in restaurant site selection?

Prior studies on site selection for independent restaurants have looked at a limited number of area attributes or site characteristic elements. This is the first study to investigate multiple pre-opening characteristics of the area attributes and site characteristics. The model created is the first scientific method assisting restaurateurs in deciding on the site.

The study, conducted in Hong Kong, focused on the survival rates of 6,710 newly opened restaurants in 2016, 2017, and 2018. Nineteen individual variables, including eight socio-demographic elements, eight site characteristics, and three restaurant characteristics, were combined to investigate the research questions. A logistic regression model and the corresponding marginal effects were used in investigating the first research question. Locating within a mall provides the single most significant increase in potential success at 11.229 percent. The second greatest impact was the average number of individuals living in a household, as for every additional person, the potential for success increases by 10.098 percent.

Two predictive models, logistic regression and artificial neural network were used to investigate the second research question. Both models can be used to predict restaurant success with the logistic regressions accuracy rate of 71.27 percent and the artificial neural network accuracy rate of 72.55 percent. The artificial neural network is selected as the preferred model due to the marginally higher accuracy and greater area under the ROC curve.

The study expands theoretical contributions to the principle of minimum differential and central place theory. The principle of minimum differentiation showed that the clustering of restaurants is essential if the competitor is of a similar price point. The central place theory found that individuals still prefer not to travel far distances

within a dense urban environment. While transportation is ample, restaurant patrons still prefer to dine near home.

The research outlines the importance of site selection and that the mantra of ‘location, location, location’ is critically important to restaurants. Restaurateurs cannot rely solely on differentiated concepts, low fixed costs, or quality food and service. The location is a crucial ingredient in the overall success and failure of a restaurant, and the artificial neural network model created will aid all restaurant industry stakeholders in selecting the ideal site for their endeavor.

Keywords

Restaurant Success, Predictive Models, Artificial Neural Network, Area Attributes, Site Characteristics

Acknowledgements

The PhD journey would not have been possible without the support, guidance, and encouragement of many individuals. I would first like to express my sincere thanks and gratitude to my chief supervisor, Dr. Sung Gyun Mun. Your guidance, insight, critique, and comments starting from the initial research proposal was instrumental in making a robust thesis. The time spent discussing the thesis ideas, journal papers, and the hospitality industry as a whole was invaluable.

I would also like to thank my co-supervisor, Professor Haiyan Song, the BoE chair Professor Cathy Hsu, and the external examiners Professor Billy Bai and Dr. Chihyung Michael Ok. Thank you to the supervisory committee and confirmation panel members, Professor Basak Denizci, Dr. Lorenzo Masiero, and Dr. Henry Tsai. The insights and comments that were provided helped to strengthen and improve my thesis.

To those who spent time sharing a conversation and a coffee, thank you. That time was incredibly valuable in gaining a different viewpoint when I was at a crossroads and a welcome relief when the conversation went away from schoolwork and thesis writing.

To my family and friends, thank you for the ongoing support and for always being there throughout the years. To Beth, this journey and thesis would not have been possible without you. The countless hours you spent listening, reading, editing, and discussing the ideas and concepts are immensely appreciated. Your love and support was the rock I needed to complete this journey.

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **iv**

List of Tables ..... **ix**

List of Maps ..... **x**

List of Figures ..... **xi**

List of Charts ..... **xi**

Chapter 1. Introduction **1**

1.1. Chapter Introduction 1

1.2. Background of Study 1

1.3. Problem Statement 8

1.4. Research Questions 10

1.5. The Hong Kong Market 11

1.6. The Significance of the Study 14

1.6.1. Theoretical Contributions 14

1.6.2. Practical Contributions 16

1.7. Definitions 18

1.8. Structure of the Study 20

1.9. Chapter Summary 21

Chapter 2. Literature Review **22**

2.1. Chapter Introduction 22

2.2. Area Attributes 22

2.2.1. Residential Density 24

2.2.2. Household Income 27

2.2.3. Household Age 29

2.2.4. Size of Household 31

2.3. Restaurant Site Characteristics 33

2.3.1. Competitors 33

2.3.1.1. Cuisine Competitors 36

2.3.1.2. Price Competitors 37

2.3.1.3. Convenience Competitors 38

2.3.1.4. Competitors Conclusion 39

2.3.2. Distance to Mass Transit 40

2.4. Chapter Summary 42

Chapter 3. Research Methodology ... 44
3.1. Chapter Introduction ... 44
3.2. Observations ... 44
3.2.1. Restaurant Data Analysis ... 48
3.3. Dependent Variable ... 48
3.4. Independent Variables ... 49
3.4.1. Area Attributes ... 49
3.4.2. Site Characteristics ... 50
3.4.3. Control Variables ... 51
3.5. Modeling ... 57
3.5.1. Logistic Regression ... 58
3.5.1.1. Background and Reason for Selection ... 58
3.5.1.2. Assumptions ... 59
3.5.1.3. Equation ... 60
3.5.1.4. Marginal Effects ... 64
3.5.2. Artificial Neural Network ... 64
3.5.2.1. Background and Reason for Selection ... 64
3.5.2.2. Assumptions ... 66
3.5.2.3. Model Generation ... 66
3.5.2.4. Formula ... 70
3.5.2.5. Training and Validation ... 73
3.5.3. Model Selection Criteria ... 75
3.6. Chapter Summary ... 77
Chapter 4. Findings ... 79
4.1. Chapter Introduction ... 79
4.2. Success and Failure Statistics ... 80
4.2.1. Time Period Success and Failure ... 80
4.2.2. Locational Success and Failure ... 84
4.2.2.1. Regional ... 84
4.2.2.2. TPU Restaurant Openings ... 86
4.2.2.3. TPU Failure Rates ... 93
4.3. Restaurant Area Attributes and Site Characteristics Descriptives ... 100
4.3.1. Area Attributes ... 100
4.3.2. Site Characteristics ... 113

4.4. Logistic Regression........114
4.4.1. Assumptions........114
4.4.2. Model Findings........115
4.4.2.1. Area Attributes........116
4.4.2.2. Site Characteristics........119
4.5. Artificial Neural Network........124
4.5.1. Model Generation........124
4.5.2. Model Findings........124
4.6. Hypothesis Testing........131
4.7. Research Questions........133
4.7.1. The Impacts of Success........133
4.7.2. Model Selection........134
4.8. Conclusion........136
Chapter 5. Discussion........137
5.1. Introduction........137
5.2. Evaluation........137
5.2.1. Survival of Restaurants........137
5.3. Research Questions........142
5.3.1. Research Question 1........142
5.3.2. Research Question 2........145
5.4. Hypotheses........145
5.5. Implications for Hong Kong Restaurateurs........150
5.5.1. Ideal Customer........151
5.5.2. Ideal Location........155
5.5.3. Implication Conclusion........159
5.6. ANN Model Accuracy Test........162
5.7. Contributions........165
5.7.1. Theoretical Contributions........165
5.7.2. Practical Contributions........167
5.8. Limitations........168
5.9. Conclusion........170
Chapter 6. Conclusion........171
6.1. Future Research........177
6.2. Concluding Remark........178

References **179**
Appendices **198**
Appendix 1. Tertiary Planning Units 198
Appendix 2. Program to Merge XML Files for Restaurants 202
Appendix 3. Program to Merge XML Files for Convenience Competitors 207
Appendix 4. Program to Scrape the Census Data 212
Appendix 5. Program to Scrape OpenRice.com 219
Appendix 6. Program to Count Competitors with Matching Price 226
Appendix 7. Program to Count Competitors with Matching Cuisine 230
Appendix 8. Program to Count Convenience Competitors 234
Appendix 9. Program to Find if a restaurant is Close to an MTR Exit 237
Appendix 10. Program to Find if a restaurant is within a Hotel 239
Appendix 11. Program to Find if a restaurant is within a Tourist Attraction 241
Appendix 12. Program to Find if a restaurant is within a Shopping Mall 243
Appendix 13. Program to Run the Logistic Regression 245
Appendix 14. Program to Run the Artificial Neural Network 247
Appendix 15. Program to Map the Restaurants 255

List of Tables

Table 1 Statistics of Menu Price 52
Table 2 Revised Statistics of Menu Price 53
Table 3 Statistics of Restaurant Cuisine 54
Table 4 Revised Statistics of Restaurant Cuisine 54
Table 5 Dependent and Independent Variables 56
Table 6 10-Fold Cross-Validation 74
Table 7 New Restaurant Failure Statistics 2016-2019: Quarterly 81
Table 8 New Restaurant Failure Statistics 2016-2019: Quarterly and Cumulative Failure Rate 82
Table 9 Success or Failure of New Restaurants by Region 86
Table 10 Restaurant Opening Count and TPU 88
Table 11 Top TPUs by Number of New Restaurants 90
Table 12 Top TPUs by Number of Restaurant Openings 92
Table 13 Bottom TPUs by Number of Restaurant Openings 92
Table 14 Failure Rate in First 365 Days After Opening 95
Table 15 Failure Rate in First 730 Days After Opening 97
Table 16 Failure Rate in First 1,095 Days After Opening 99
Table 17 Dependent and Independent Descriptive Information 101
Table 18 Most Populated TPUs 103
Table 19 Least Populated TPUs 103
Table 20 Most Populated TPUs by Square Kilometer 106
Table 21 Least Populated TPUs by Square Kilometer 106
Table 22 TPUs with Oldest Median Age 109
Table 23 TPUs with Youngest Median Age 109
Table 24 TPUs with the Highest Median Domestic Household Income 112
Table 25 TPUs with the Lowest Median Domestic Household Income 112
Table 26 Logistic Regression Results 123
Table 27 Neural Network Design 126
Table 28 Weights from Layer 1 to Layer 2 128
Table 29 Weights from Layer 2 to Output Layer 130
Table 30 Hypothesis Findings 132
Table 31 Yearly Openings and Survival of Restaurants 138
Table 32 Restaurant Survival and Failure Rates 140
Table 33 ANN Model Accuracy Test 164

List of Maps

Map 1 Location of 20,548 Restaurants 46
Map 2 Location of 6,710 Restaurants 47
Map 3 Regions of Hong Kong 85
Map 4 Restaurant Opening Density Heat Map 87
Map 5 Top TPUs by Number of New Restaurants 89
Map 6 Number of Restaurant Openings per Square Kilometer 91
Map 7 Failure Rate in First 365 Days After Opening 94
Map 8 Failure Rate in First 720 Days After Opening 96
Map 9 Failure Rate in First 1,095 Days After Opening 98
Map 10 TPUs by Number of Residents 102
Map 11 TPUs by Number of Residents per Square Kilometer 105
Map 12 TPUs by Median Age..108
Map 13 TPUs by Median Monthly Domestic Household Income 111
Map 14 TPUs for Mean Household Size 144
Map 15 Ideal TPUs for Area Attributes 154
Map 16 Areas Within a Five-Minute Walk of MTR Station Exits 157
Map 17 Location of Shopping Malls...158
Map 18 Top TPUs for Site Characteristics 159
Map 19 Ideal TPUs 161
Map 20 Ideal TPUs in Kowloon 162

List of Figures

Figure 1 Logit Function Graph 61
Figure 2 Logistic Curve 62
Figure 3 Artificial Neural Network...68
Figure 4 Artificial Neural Network...127
Figure 5 Logistic Regression ROC Curve 135
Figure 6 Artificial Neural Network ROC Curve..135

List of Charts

Chart 1 New Restaurant Failure Statistics 2016-2019: Quarterly and Cumulative Failure Rate..83
Chart 2 Curvilinear Effect of Total Residents ('000)..116
Chart 3 Curvilinear Effect of Median Monthly Household Income ('000) 118
Chart 4 Curvilinear Effect of Price Competitors 120
Chart 5 Curvilinear Effect of Convenience Competitors 121

Chapter 1. Introduction

1.1. Chapter Introduction

This chapter introduces the background of the study, starting with a discussion of the restaurant industry overall and highlighting the key characteristics of individual restaurants. Then, the problem statement and importance of the research will be introduced as well as the research objectives. The theoretical and practical contributions will be discussed, and finally, the structure of the remaining research will be detailed.

1.2. Background of Study

Operating a restaurant is an incredibly risky endeavor, as a restaurant that opens one month might close a few months later, while another might remain in operation for years. Determining why one restaurant survives while a second fails can prove challenging. Some restaurants fail with more commonly understood reasons, such as inadequate food and service or poor management, while other restaurants fail without a clear understanding of what happened. These often have explanations of bad luck, poor timing, or that the area was not ready for a restaurant of that style. Restaurant failure can be considered from two different perspectives, the pre- and post-opening decisions. While the post-opening factors include all day-to-day decisions that are in direct control of the management, the pre-opening decisions are the preliminary ones that a restaurateur makes in the planning and design of a restaurant and are mostly out of the control of management once the restaurant begins its first day of operation. Miss-management of the pre-opening decisions can lead a restaurant to a predetermined failure, no matter how well management handles the post-opening elements. One of the initial pre-opening decisions that a restaurateur needs to consider is the physical site location (Chen & Tsai, 2016; Mazze, 1972; Schaefer, Luke, & Green, 1996; Smith, 1985; Yang Yang et al., 2017). Site location decisions are often made by intuition, as there is a lack of comprehensive research into the individual site characteristics and which elements are the most important for the future success of a restaurant. Also, there is no current model for independent restaurants to use to determine if the site has survival potential based on the impact of the socio-demographic and site characteristics elements.

The decision of where to open a restaurant is one of, if not the most critical pre-opening decisions a restaurateur needs to make (Egerton-Thomas, 2005; Olsen, Ching-Yick, & West, 1998). The site location decision needs to look past the physical size of the space or the monthly rent but also needs to consider other external elements, including the visibility (Dock, Song, & Lu, 2015; Roig-Tierno, Baviera-Puig, Buitrago-Vera, & Mas-Verdu, 2013; Wood & Tasker, 2008) affecting patrons who are menu shopping. The ease of the site accessibility (Burnaz & Topcu, 2006; Jang & Mattila, 2005; Jekanowski, Binkley, & Eales, 2001) as the majority of people navigate via the subway (Transport Department, 2018) and desire to dine close to the station exit (Njite, Dunn, & Hyunjung Kim, 2008). Operating a restaurant within a shopping mall, tourist attraction, or hotel can come with an increased operating cost (Pashigian & Gould, 1998), but could be offset by having a captive audience (Guest & Cluett, 1974; Yang Yang et al., 2017). The area competition is significant (Burnaz & Topcu, 2006; Guy, 1995; Roig-Tierno et al., 2013) as a restaurant cluster increases foot traffic (Fields, 2007). Additionally, the fixed and variable operating expenses (Nwogugu, 2006), and potential strategic advantages (Ghosh & Craig, 1983) are significant aspects. These elements will influence the short-term success by creating a base clientele that will become repeat visitors, and in turn, invite their friends, family, and colleagues, providing an avenue for long term success.

The short-term success is vital in independent restaurants as they need to have a cash inflow in a relatively short time frame to ensure available working capital. The suggested amount of working capital ranges from having enough to pay the fixed operating costs for the first twelve months (Egerton-Thomas, 2005), to a minimum of three months (Fields, 2007), or equaling four percent of sales (Mun & Jang, 2015). A restaurant with an inadequate amount of working capital faces the possibility of not being able to pay their creditors and are left to seek alternative funding. Other industries do not face this hardship while starting a new venture; financial backing can come from the head office or venture capitalists, who are willing to operate at a loss for numerous years, potentially up to a decade. Independent restaurants do not have that luxury and need to build and maintain a positive working capital, showcasing the need for rational site selection decisions.

The famous phrase attributed to Lord Harold Samuel is that “there are three things that matter in property: location, location, location.” Location is a critical factor of the success or failure of a retail shop (Craig, Ghosh, & McLafferty, 1984; Gonzlez-Benito & Gonzlez-Benito, 2005; Prayag, Landré, & Ryan, 2012; Roig-Tierno et al., 2013; Self, Jones, & Botieff, 2015; Tzeng, Teng, Chen, & Opricovic, 2002; Wendt, 1972). The significance of location was studied by Litz and Rajaguru (2008), and they found that for small retail stores, the transactional convenience of being situated close to their clientele was critical to the long-term success. Selecting a location based on the size of the shop or the amount of the rent might seem logical, but without considering the clientele, visibility, access, or nearby competition, the store owner might be dooming him or herself to failure. Selecting a site location that is not the perfect fit on paper could seem disastrous, but if the location is situated close to their ideal demographic customer, it could potentially provide a higher possibility of success than the alternate site of ideal size and rent in a less desirable location.

Among restaurateurs, the classic debate of what decision should come first, the site location, or the overall concept, is highlighted by Fields (2007) when equating it to the classic chicken and egg dilemma. No matter which comes first, to be successful, the location is paramount (Fields, 2007; Miller, 2006; Pioch & Byrom, 2004). Ghosh and Craig (1983) remarked that the location of a restaurant is the single most significant determinant of success as it can provide strategic advantages over the competition. Further, Parsa, Self, Sydnor-Busso, and Yoon (2011) recognize the importance of the decision of site location as their study remarked that the decision of location when considering the density of the area, has a direct influence on the failure of a restaurant. Although there are multiple reasons for a restaurant to fail, selecting a poor location is one of the more critical reasons (Bellini, 2016).

Restaurant studies predominantly focus on the post-opening decisions, including operational management (Ozdemir & Caliskan, 2014), financial management (Guerriero, Miglionico, & Olivito, 2014), strategic direction (Harrison, 2003), and marketing decisions (Kivela, 1997). While the studies focus on the post-opening elements, all of these elements are directly influenced by the site selection, from the ability to market the restaurant as a 'hidden gem' or 'best view in the area.' To the style of a restaurant, such as buffet, family-style, or white-glove service. Understanding the competition, what is working, or what is missing could also be used as a strategic advantage. The price point of the food and drink is affected by the site selection, operating a restaurant in an off-path area, or an area undergoing gentrification, might mean that the restaurant cannot drive prices as high as desired, but could change in years to come. Alternatively, if all the competitors are mid-priced, is there the potential to become a top-tier restaurant? Will the area support a restaurant that opens for breakfast, or is the area more suitable for lunch and dinner operations? Does the area provide a path for future expansion of either the current restaurant or growth through multiple outlets? While the post-opening research has the potential of guiding restaurant groups, corporations, and hospitality schools, the research to date overlooks the most critical influencer, the site location. This study looks to provide complimentary research and considers the importance of the site location, which will pair with the existing body of work regarding post-opening decisions.

The importance of site selection is different depending on the product involved; the site selection significance will differ for concert venues, sports arenas, car dealerships, or grocery stores. Sometimes, the vast size or uniqueness of a specific product will create a natural pull that patrons will travel to the site no matter the location. The frequency of potential visits correlates to the importance of site selection. The site selection of a car dealership that a customer might visit every five or ten years is less important than a clothing store that is visited every few months, which is less important to a restaurant that can be visited monthly or multiple times a month and relies heavily on foot traffic. The need for independent restaurants to have a constant flow of customers in a short period drives the importance of site locational decisions.

In previous studies that looked at restaurant success and socio-demographic elements, there were significant findings that population density and household size have a positive predictive component (Yang Yang et al., 2017). The site location is also critical when opening up-scale restaurants; current research shows that new up-scale restaurants should understand the site characteristics and situate themselves either by restaurants with both a similar cuisine and price or just by similar price (Jung & Jang, 2019). Studies have shown that restaurateurs need to consider both the socio-demographic factors as well as the site characteristics when selecting the site location. These two areas are vital for the long-term success of the restaurant and that a poor location decision likely correlates with a failed restaurant.

Franchise restaurant corporations understand that location decisions are critical to the long-term success, Goldberg (n.d., para. 6) called franchise location decision making a “make-or-break decision.” Restaurant franchises have years and copious amounts of data that look at the factors that determine future site selections and conduct proprietary internal research to select the ideal location (Smith, 1985). Wendy’s, one of the top ten quick-service restaurants in market capitalization, uses Esri’s software that allows them to geographically map potential sites and utilize existing customer demographics to allow the mapping software to show potential locations for future growth (Thau, 2014). Franchisors, like Wendy’s, often dictate the ideal site from their historical information and can provide information such as ideal demographics, parking, traffic, visibility, and population breakdown (Elgin, 2004; Mendelsohn, 1985; K. Park & Khan, 2006; Smith, 1983). Williams (2017, para. 2), who worked for nine years in brand and marketing at Starbucks, explained that Starbucks has “sophisticated off-the-shelf tools which show them traffic patterns, income in the area surrounding the [potential site], all sorts of demographic and psychographic information to choose the best locations.” Starbucks and other franchisee site selection decisions must be approved by the corporate office based on internal research of the area (K. Park & Khan, 2006).

Determining a formula to find ideal franchise locations is a practice not solely for restaurants; hotels also recognize that a significant factor for the overall success and failure of a hotel is site selection (Kimes & Fitzsimmons, 1990; Yang Yang, Wong, & Wang, 2012). Restaurant corporations understand the precariousness of the restaurant industry and are utilizing regression models and off the shelf software to try to maximize profitability and minimize failure. By utilizing internal modelling, the cumulative three-year failure rate of franchise restaurants is 57.22 percent, which is less than the rate of independent restaurants at 61.36 percent (Parsa, Self, Njite, & King, 2005). This study looks to create a more robust model with elements that are deemed critical to the site selection process with the objective of decreasing the failure rate of restaurants.

Independent restaurants within the United States have a median lifespan of four and a half years (Luo & Stark, 2015). While there are countless reasons why one restaurant is successful, and another fails, whether as a result of pre- or post-opening decisions, one of the first aspects that restaurateurs need to focus on, and is crucial to the overall success, is the pre-opening decision of the site location. Based on the bid rent theory, a highly sought-after location will come with higher rent (Bellini, 2016; Parsa & van der Rest, 2017), and a poor decision regarding the location is incredibly challenging to rectify due to the cost and the time frame involved of finding a new site location and moving operations (Bennison, Clarke, & Pal, 1995; Yang Yang et al., 2017). Additionally, even moving operations within a metropolitan city will be complicated as the restaurant will have to attract new clientele as their previous customers will not desire to travel to the new destination to eat the food when they have an abundant number of other restaurants from which to select.

The current method of site selection is typically done by either a veteran of the restaurant industry making a suggestion based on their experience and knowledge of the area or hiring a consultant that conducts a survey to suggest the ideal site. However, with over a sixty-percent failure rate in a three-year cumulative period (Parsa et al., 2005), the current site selection methods are not good enough to accurately select a site. This research looks to minimize the failure rate by incorporating a scientific model into the site selection method. Minimizing the failure rate of restaurants will have many benefits, from ensuring the local economy is prosperous, food and drink vendors have a continuous business to supply, providing steady employment for the workers, limiting bankruptcies and defaults on loans, and having a location for the residents and workers to enjoy. Previous research on independent restaurant site selection has only looked at a few socio-demographic or site characteristic elements in each study. To the researcher's knowledge, no study to date has tried to combine multiple elements of the area's socio-demographics and the site characteristics to understand the overall impact that they have on independent restaurants success, or to create a model to assist independent restaurants in making an informed decision on what site to open a restaurant.

1.3. Problem Statement

Previous research on restaurants has focused on either predicting financial distress on already opened restaurants (S. Y. Kim & Upneja, 2014) or investigating the post-opening elements. To date, research on the pre-opening element of independent restaurant site selection has been limited. Franchise restaurants, multi-unit stores, and hotel corporations can use historical location decision patterns and the customer demographics of those who visit their stores to determine the ideal site location for potential growth and success (Wood & Browne, 2007). Corporations can track the cross-street placements, traffic patterns, location to major thoroughfares, parking availability, future area growth, as well as the age, ethnicity, gender, income, and shopping patterns of their customers. The understanding of potential customers and area demographics allows a competitive edge when locating and opening new stores resulting in a lower failure rate (Karakaya & Canel, 1998). However, independent restaurants do not have access to historical information, and without paying for a site feasibility study, they might not understand the nuances of the road and traffic patterns.

Restaurateurs might have an ideal clientele in mind, and as Harrington, Ottenbacher, and Kendall (2011, p. 285) wrote "practitioners should consider key target market characteristics to ensure a fit between restaurant attributes and expectations of customers," but will inserting a restaurant in the middle of what is conceived to be their clientele's neighborhood bring success? What other site attributes provide the highest potential success, opening a restaurant on the main street or side street, ground floor or upper floor, in a mall or hotel, close to a tourist destination or closer to where the residents live?

Independent restaurants already face high failure rates, over sixty percent within three years (Parsa et al., 2005), and currently, do not have a way of determining what area dynamics or site characteristics are most important in the determination of a sites potential success. The current method of site selection is by having someone with experience in restaurants having a 'gut feel' about the potential success of a site (Clarke et al., 2000; Ghosh & Craig, 1983; Hernandez & Bennison, 2000), and by having a more thorough understanding will be able to have decisions based off the customer and the competitive space (Wood & Reynolds, 2012).

1.4. Research Questions

The previous studies and books about independent restaurant currently list various suggestions on how to best select a site. However, those are mere suggestions that include being aware of the access to utilities, zoning codes, visibility, accessibility, traffic patterns, and parking (Fields, 2007), but besides understanding how these independently work in a site, there is currently no way to determine the overall potential viability of a restaurant site. Considering the gap between the pre-opening area factors and potential success of the site, the research questions are as follows:

1. Which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure?
    a. What is the most critical socio-demographic area attribute a restaurateur needs to pay attention to when selecting an area?
    b. What is the most critical restaurant site characteristic that a restaurateur needs to consider when selecting a site?
2. Considering the potential influence of the overall site characteristics, can a model be created to aid in restaurant site selection?

There is an unprecedented number of post-opening independent variables that influence the success or failure of a restaurant once it opens the doors. These failing variables include, but are not limited to, having a non-differentiated concept (Fields, 2007; Parsa, Gregory, & Terry, 2011), high fixed costs (Parsa, Gregory, et al., 2011), mediocre food and service (Feloni, 2014; Parsa et al., 2005), lack of management experience (Camillo, Connolly, & Kim, 2008; Feloni, 2014; Parsa, Gregory, et al., 2011; Parsa et al., 2005), an undesirable ambiance (Parsa, Kreeger, van der Rest, Xie, & Lamb, 2019; Sulek & Hensley, 2004), inadequate amount of operating capital (J.

Everett & Watson, 1998), wrong pricing strategy (Healy & Mac Con Iomaire, 2018), and an overall lack of business strategy (Camillo et al., 2008). This study has considered these elements and will control them by including the aggregate review score for restaurants from OpenRice.com, Hong Kong's most ubiquitous and popular restaurant review website.

The second research objective will expand on the history of modeling, and especially computer-aided modeling. Cella (1968) designed one of the first computer-based restaurant site selection programs using linear regression. In 1969, computer-based site selection expanded for restaurant groups with over 30 stores (Darley & Gobar, 1969). The ability to create a model to suggest restaurant site locations based on the pre-opening factors has been available for over fifty years. However, currently, no known modeling software can predict independent restaurant success based off the pre-opening site elements, and generic modeling software is not sophisticated enough for the independent restaurant industry (Wood & Tasker, 2008).

1.5. The Hong Kong Market

The study and modeling will be focused on the restaurants located within Hong Kong, a Special Administrative Region of the People's Republic of China. Hong Kong residents enjoy dining out with restaurant receipts in 2018, totaling 119.5 billion Hong Kong dollars (Census and Statistics Department, 2019). The Hong Kong Census and Statistics Department (2016a) reported that in the latest published report, the 2014/15 Household Expenditure Survey, in a survey of 6,812 households, the mean expenditure for food, consumer goods, and services was HK$27,627 per household, an increase of 27.8 percent over the previous survey conducted in 2009/10. Out of this overall expenditure, food accounted for 27 percent, second only to housing at 36 percent. The food expenditure can be further expanded for foods consumed at home and foods consumed away from home. The food consumed away from home accounted for 66 percent of the total food expenditure. To put this in perspective, of the food expenditure in American households, 43.5 percent was away from home in the study conducted in 2017 (U.S. Department of Labor, 2018).

Hong Kong residents have more significant food away from consumption due to the domestic household living size (CBRE, 2016). The median household in 2016 was measured at 430 square feet, which equates to 161 square feet per person (Ng, 2018). In comparison, the median household in the United States is 2,467 square feet, and per person, 971 square feet (Perry, 2016). The small living quarters equate to a smaller kitchen and a small dining area, if any at all. Residents dine out at a higher frequency as it is more convenient, comfortable, and restaurant depending, cheaper. Due to the small size of the homes in Hong Kong, it is more comfortable to socially meet friends or family at a restaurant instead of preparing a home-cooked meal (CBRE, 2016). Additionally, according to a UBS report in 2016, Hong Kong workers have the longest working week of any country, working over 50 hours per week, the long hours, and the inexpensive options for dining out have increased the consumption of food away from home (CBRE, 2016).

The flat size, inexpensive dining options, and long working hours directly contribute to the growth of the restaurant industry; Hong Kong has seen year over year growth for the previous fifteen years (CBRE, 2016). When examining specific meal periods, 28.6 percent of residents over the age of 15 eat breakfast outside of the home five or more times per week, and 55 percent eat out at least once per week. During lunch, 48.9 percent of residents over the age of 15 eat out at least five times a week, and almost 70 percent eat out at least once a week. For dinner, 9.9 percent of residents over the age of 15 ate out at least five days per week, and 47.5 percent ate outside of the home at least once per week. Overall, over 80 percent of residents ate outside of the home for any meal at least once a week (Department of Health, 2017).

PricewaterhouseCoopers ranked Hong Kong as the ninth city out of the thirty that were examined for the most recent version of The Cities of Opportunity (7^{th} edition). The ranking is based on 10 indicators that included 67 independent variables (PwC, 2016). The indicators looked at the different aspects of the city, including economics, quality of life, and technology. The other global cities that are the most similar to Hong Kong in potential opportunities are London, Singapore, Toronto, Paris, Amsterdam, New York, Stockholm, San Francisco, and Sydney. As these cities have similar opportunities for starting businesses and quality of life, this research will also be able to generalize the findings to at a minimum these cities.

Hong Kong workers eat out often due to the long work hours and small flat size; globally, Hong Kong residents have the fifth-highest expenditure on food away from home, following Spain, Australia, Austria, and New Zealand (Satran, 2017). While the expenditure is high, so is the competition to attract the diners. Hong Kong has over fifteen thousand restaurants in operation for residents and tourists to dine at; proportionally, each restaurant has a potential customer base of fewer than five hundred people per meal period. With the desire for residents to purchase food away from home, the immense restaurant competition, and the ability to generalize the findings to global cities showcases why Hong Kong is an ideal location to investigate the importance of the site location. This study looks to understand the importance of the elements surrounding site location in order for restaurateurs to select a location that maximizes how many residents' patron the restaurant, in addition to capturing the business of the more than sixty-five million tourists that arrive annually in Hong Kong (Hong Kong Tourism Board, 2019a). Additionally, the Hong Kong Government provides a vast amount of readily available data to look at the overall characteristics of different restaurant site locations and how differences in area characteristics affect the survival rate.

1.6. The Significance of the Study

This thesis will be able to contribute to current knowledge both in literature and practice in terms of the theoretical and practical contributions. These contributions are outlined in the next two sections.

1.6.1. Theoretical Contributions

The research will contribute to the current theory by conducting a spatial analysis of restaurant site characteristics and by enhancing and expanding the current understanding of site selection. Utilizing multiple site characteristics including but not limited to the street location, floor level, competition, and distance to mass transit as well as eleven elements of the population dynamics of the surrounding area including potential customer density, age, gender, ethnicity, income, as well as other independent variables, the study will aim to explore which elements are critical and try to understand the reasons behind them. The researcher believes this will be the first study to investigate multiple independent pre-opening characteristics of both the area socio-demographics and the site characteristics to understand what link there is between them and potential independent restaurant success.

The study will also use the theoretical reasons behind why these characteristics are essential and attempt to create two separate models to see the impact of these elements on the potential success of the site. Parsa, Gregory, et al. (2011); Parsa, Self, et al. (2011); Parsa, van der Rest, Smith, Parsa, and Bujisic (2015) have examined some individual characteristics and restaurant success, but have not attempted to link the characteristics together to see if a model could be used to predict the possibility of success or failure. The model created can also be a blueprint for other industries and examining site locations for potential success. Industries such as clothing, furniture, general stores, or grocery would be able to modify the model utilizing their industry characteristics to understand site locations and potential success in a more thorough way.

The study will also be able to add to the understanding of the clustering of restaurants and whether selecting a site around similar competitors makes sound business sense. The principle of minimum differentiation showcases why the clustering of competitors by price is beneficial to restaurants in attracting customers who have the desire to dine at restaurants of a specific cuisine. However, when also considering the clustering of other price and cuisine differentiated restaurants as well as convenience competitors, does the clustering of all types of restaurants and food stores create a benefit or a disadvantage for a restaurant. This paper will expand the current body of knowledge on the clustering of restaurants and how different restaurant styles, prices, and food stores impacts business survival.

The study will also add to the discussion of central place theory. The central place theory proposes that there should be an area of shops that are the pull factor for those living in the nearby vicinity. Those living or working within an area will go to the closest shopping area and not travel to shopping areas farther away. This paper will look at this theory in terms of restaurants. If the central place theory is correct, then people will dine at restaurants either around their homes or places of work as that will provide the fulfillment desired. Does the density of the residents in the area impact the success or failure of restaurants within the same area, or are there areas with a small density of residents, but a high number of successful restaurants?

The bid rent theory says that the sites that are located within a central business district demand the highest rent, as they also have the highest number of potential restaurant patrons due to the density of the workforce. While the bid rent theory has been extensively looked at in terms of central business districts, this paper will consider if the theory also holds when looking at tourist attractions and do those also follow the theory. The theory should hold if there is a large number of successful restaurants surrounding tourism attractions. The paper will utilize the areas around and including the top ten list of attractions from Discover Hong Kong (n.d.) to see the impact that top tourism areas have on restaurant success.

1.6.2. Practical Contributions

The contributions of this study to society are plentiful. The five groups that will benefit the most from the research are restaurateurs, banks, investors, landlords, and consultants. Currently, only major restaurant franchises have their own in-house modeling guidelines, these guidelines have been in use for over three decades (Bowlby, Breheny, & Foot, 1984), creating a practical model that can be used by all potential owners and operators will assist in various ways.

Restaurateurs will be able to use the model to predict which site location is most suited for their restaurant. Being able to set themselves up for potential success from the onset is critical to future success. They will be able to use this knowledge when approaching and negotiating with landlords regarding rent and terms of their contract. Also, they will also be able to influence potential investors to show why one site location has a higher preference over a secondary site, and finally, they will be able to include it within their business plan when applying for a loan.

Banks and investors will be able to utilize the model when deciding to approve or deny a potential loan for a restaurateur, as improvements in the prediction accuracy of success or failure reduce their long-term risk (Wood & Tasker, 2008). They might be able to make suggestions of locations or reduce the interest rate, monetary amount of the loan, as well as the amount of risk they are willing to take based on the information the models will provide them.

Landlords strive to have long term leases to provide a steady stream of customers and revenue to their site. Knowledge about the potential success of a restaurant can allow them to accept a restaurant or decide on a different style of a tenant for the location. This study will also assist them with the amount of rent a site will demand. The knowledge that the location is ideal for a restaurant can, in turn, create higher demand and a higher price for the site.

Consultants can utilize critical variables and models when creating feasibility studies for clients. The feasibility studies can be a critical aspect of the consultation when also combining the style, ambiance, and other post-opening factors of the restaurant; being able to showcase the site potential can assist in creating confidence in their client to move forward with the project.

1.7. Definitions

The study utilizes two parameters when examining the potential restaurant location; they are the area attributes and site characteristics. Further, the study will look at separate areas of Hong Kong and how they are differentiated in the area attributes, site characteristics, and restaurant success. These are defined as follows.

Area Attributes

Area attributes include all aspects of the population demographics of those living and working around the potential site. These include the number of residents, ethnicity, salary, gender, age, and education (Arduser, 2003; Kotler, Bowen, Makens, & Baloglu, 2017). It also includes whether their residence is owned or rented, the size of the household, and the unemployment rate.

Site Characteristics

The site characteristics include the aspects the pertain directly to the physical site. These characteristics include the competitors, inclusive of the cuisine, price, convenience, in addition to the distance to mass transit, location to the street, placement within a mall or hotel, and distance to attractions.

Proximity and Vicinity

This study will consider the proximity and vicinity to be the surrounding area within a 369.95-meter radius of the individual restaurant site. A five-minute walk is the acceptable distance that people will walk before using other types of transportation (Atash, 1994), and in Hong Kong, the mean walking speed in an outdoor commercial area is 73.99 meters per minute (Lam & Cheung, 2000). Therefore, in Hong Kong, a five-minute walk equates to 369.95 meters.

Tertiary Planning Unit (TPU)

Hong Kong can be separated into multiple sections, depending on how specific or small of an area needs to be analyzed. It can be either separated into the "District Council Districts and Constituency Areas" that includes the entire area of Hong Kong, further separated into the four regions of Hong Kong Island, Kowloon, New Territories, and the outlying Islands, followed by the eighteen districts and finally to the 431 sub-districts. Alternatively, the census can be separated into the planning units. Within the 2016 census, there are nine primary planning units, fifty-two secondary planning units, and 291 TPUs (Appendix 1). While there are 291 TPUs, due to the small populations, some TPUs are combined to protect the privacy of the residents' data; as such, there are 214 TPUs available for analysis.

Restaurant Failure

The study will use the term failure for the closing of a restaurant at an individual site for both voluntary and involuntary reasons. Involuntary closures include bankruptcies or foreclosures. Voluntary closures are ones where the owner decides to close an operating business; these closures can include a desire for change to their current lifestyle, becoming burnt out, retirement, or relocating the restaurant. Although involuntary closures would be preferred in a study of restaurant success and failure, voluntary closures do occur in Hong Kong. Voluntary closures are both deemed to be limited in number and are reported as closures in the same manner as involuntary closures; additionally, they are unable to be separated. As such, both types of closures will be used in considering restaurant failure.

Restaurant Success

The study also refers to success within the research questions. As failure is considered as the discontinuance of business (J. Everett & Watson, 1998) at the specific location, the study will consider success as the continuation of the business at the individual site, no matter the amount of financial profit.

1.8. Structure of the Study

The study will be conducted in six chapters. The first chapter is the introduction, which introduces the background of the research topic, providing the research objectives as well as the contributions, both theoretical and practical, and some definitions that are used throughout the study.

The second chapter will review the current literature for the independent variables, and explain the importance of variable, the previous literature, any working theories, and the hypotheses.

Chapter three will detail the methodology and the tests used to examine the individual attributes for the independent variables. The chapter will also explain the two different models used in predicting future restaurant success. The models will have explanations of why they were selected, how the model is constructed, the individual testing of the model, and finally, how the study will determine which model is the preferred option.

The fourth chapter will list the findings based on the methodology in chapter three. The fifth chapter will provide a discussion of the findings, explanations of which independent variables are useful in restaurant site location, as well as the preferred model, the theoretical and practical applications, and the limitations of the study. The sixth and final chapter will summarize the study and discuss the potential future research topics.

1.9. Chapter Summary

This chapter has introduced restaurant failures and the importance of why a restaurant location is paramount for the survival of the restaurant. Also, some of the post-opening factors that can be attributed to restaurant failure were explained. The chapter has also emphasized the most significant pre-opening decision that restaurateurs need to make of selecting a site location that will provide them with the highest possible level of success. The research questions and sub-questions were introduced in addition to how success will be measured. The significance of the study, both the theoretical implications and practical implications were explored, and the definitions and layout of the study were introduced.

Chapter 2. Literature Review

2.1. Chapter Introduction

The attributes that make up the decision to select a specific restaurant site are numerous. Pillsbury (1987) theorizes that three aspects influence the success or failure of a restaurant: the potential site's accessibility, ambiance, and socio-economics. As this research is considering only the pre-opening elements, the ambiance will not be considered, and only the accessibility and socio-economic features are necessary. As accessibility is a small portion of the overall site characteristics, the research will use the site characteristics and utilize a higher number of elements. The socio-economic features rely heavily on the socio-demographic features; the wider-ranging area attributes that incorporate both the socio-economic and socio-demographic features will be considered. As such, the two main categories for pre-opening decisions are the area attributes that include the socio-demographic elements, and site characteristics.

The literature will focus on examining the individual attributes and highlight the importance of inclusion in the decision-making process and provide an understanding of the impact on success or failure and highlight any applicable theories.

2.2. Area Attributes

The area attributes are the socio-demographics of the restaurant's potential customers. These attributes include the ethnicity, gender, age, education, and employment of the person in addition to the net household income, number of children, household size, and whether the primary accommodation is rented or owned.

There are three types of patrons that a restaurant will have visit their establishment. The first patron is the tourist, visiting Hong Kong for either business or pleasure. In 2018, there were 65 million tourists traveling to Hong Kong, and they can be separated into same-day and overnight trips. Overnight tourists accounted for less than 30 million people. These tourists spent 14.7 percent of their total spending on meals purchased outside of the hotel, most of their spending went to the retail and hotel that accounted for 50 and 21.5 percent of their budget, respectively (Hong Kong Tourism Board, 2019b). Tourist expenditure on food outside of the hotel accounts for less than eight percent of the total restaurant receipts in 2018, while the Hong Kong residents account for the majority of the remaining 92 percent of restaurant receipts.

The second is the resident who makes the journey to the restaurant outside of their typical daily area of travel. The third type and the largest group is the patron, who lives close to the establishment. While the first two patrons are essential, they are not seen as frequently as those who live in the vicinity of the potential restaurant site; a survey of Hong Kong retailers found that 90.5 percent of their customers were local residents (Planning Department, 2005b). The socio-demographic elements of the residents around the potential site are the most critical item for potential restaurant success (Thompson, 1986). The area attributes in this research will focus on the socio-demographics of the residents; the third group of potential patrons.

Past literature has investigated who eats food away from home, including all food that is bought outside of the home and does not require any cooking or preparation before eating (Nayga Jr & Capps Jr, 1994). The literature has hypothesized various factors, including age, ethnicity, education, work status, household size, worker availability, urban or rural living, single or dual parents, number of earners, and household income (Seonok Ham, Hiemstra, & Yang, 1998; Nayga Jr & Capps Jr, 1994). While previous research has focused on these elements, some are not included in this research due to being less relevant, such as worker availability, urban or rural living, or due to the unavailability of the data, for example, the number of household earners.

This research will focus on the socio-demographic factors that are deemed essential, including residential density, income, age, number of children, and household size. These attributes are deemed the most critical, and the past literature will be examined to understand how they are essential for restaurant owners/operators to consider when examining restaurant sites.

2.2.1. Residential Density

The location of the residents' primary accommodation is significant as it is the starting and ending points of the residents’ day. The location of the accommodation can impact the potential success of a restaurant as the distance a patron needs to travel is a primary factor in selecting where to shop (Briesch, Chintagunta, & Fox, 2009; Ghosh & MacLafferty, 1987; Kahn & McAlister, 1997; Sinha & Banerjee, 2004). A study conducted in Hong Kong found that one of the top five reasons shoppers select a store to go to is the distance of that particular shop to their place of residence (Yip, Chan, & Poon, 2012).

Athey, Blei, Donnelly, Ruiz, and Schmidt (2018) looked at the travel distance for dining at restaurants within the San Francisco Bay Area, California, and found that 60 percent of visits were within two miles of the consumer’s starting point. As the area of Hong Kong is 6.14 percent of the San Francisco Bay Area, the distance Hong Kong residents will travel to go to a restaurant will be proportionately less inferring that they will visit restaurants in the close vicinity of their primary residence. A previous study shows that 34.1 percent of Hong Kong residents dine at restaurants within a ten-minute walk of their home, and 71 percent patron restaurants within the same district (Planning Department, 2005a).

While there are competing theories on distance and shopping, the central place theory is the most relevant for the restaurant industry. The central place theory was first proposed by Walter Christaller in 1933 in the book titled *Die zentralen Orte in Süddeutschland* (King, 1985). Christaller proposed that consumers would travel to the nearest location that has the goods or services that are required, and that consumer demand for goods declines as the travel distance increases (Brown, 1993; Litz & Rajaguru, 2008) until the breakpoint is reached when the next central location is closer (Kivell & Shaw, 1980).

A cluster of restaurants creates a natural central place, and restaurateurs generally prefer to open where there is a higher density of potential customers (Prayag et al., 2012). While previous studies have found that there are a higher number of restaurants operating inside a residentially dense area than within a residentially sparse area (Prayag et al., 2012; Yang Yang et al., 2017), opening in a residentially dense area does not necessarily equate to long-term success.

Opening a restaurant within a residentially dense area appears logical; a higher number of people living in the area should equate to a higher number of potential customers. However, with an increased number of people, there may be an increase in other costs. For example, in London, England, the median rent for a business located in the highly dense city center is HK$382.31 per square foot, while the rent in the mid-dense area of Cambridge is HK$225.42 per square foot, and the low-density areas have rents of HK$145.13 per square foot (McDonald, 2018). Unless the volume of customers or menu prices can be raised comparably, it may become difficult for a restaurant in a high-density district to turn a profit.

In addition to the higher rents in a residentially dense area, other expenses will also be increased. The cost of raw food and supplies will be increased in a highly dense area over low-density areas. As distributors need to either rent a warehouse close to the highly dense areas at an increased cost or locate further away from the city to have a decreased rent but an increase in the delivery cost. The increase in operating costs for the distributor means that they will increase the cost of the products being sold to the restaurant. Additionally, labor expenses are also higher in high-density areas over low-density areas (Echeverri-Carroll & Ayala, 2011).

Restaurants have a fixed number of seats, a relatively small timeframe to sell them, and a rational price point at which they can be sold; as such, the additional expenditure for locating in a highly dense area might not make financial sense. If a restaurant could locate in a medium-density area and have the same number of seats, customers, and revenue, then the restaurant could have an increased profit due to lower overall expenditures; and an even higher profit if they can retain the same revenue while locating in a low-density area.

Therefore, opening a restaurant in a central place is advantageous; however, the best choice location of the central place does not necessarily need to be in a residentially dense area. Selecting a restaurant location convenient to target customers, and not just people, in general, will yield a higher number of potential patrons (Gyaan, n.d.). Restaurants in residentially dense areas are likely to have higher operating expenses, which may not be proportionally recoverable; thus, the potential for success decreases.

H1: The residential density, in the vicinity, will impact restaurant success in an inverse U-shape.

2.2.2. Household Income

Household net income is defined as the after-tax earnings of the overall household, no matter the number of earners. The literature provides a clear relationship between a rising household net income and an increase in restaurant expenditure (Byrne, Capps, & Saha, 1998; Dong, Byrne, Saha, & Capps, 2000; Frisbee & Madeira, 1986; Sunny Ham, Hwang, & Kim, 2004; McCracken & Brandt, 1987; Nayga Jr & Capps Jr, 1993, 1994; S.-J. Yang & Magrabi, 1989). Researchers have also linked the increase of household income to an increase of expenditure in mid- and full-service restaurants and not to quick-service restaurants (Binkley, 2006; Byrne et al., 1998; E.-J. Kim & Geistfeld, 2003; Yang Yang et al., 2017). However, research has not examined whether in the aggregate restaurant success increases as a household moves from low-, medium-, or high-income areas.

The study conducted by Bundle looked at the dining habits of the top one percent of spenders living in New York City. The study found that the wealthy preferred convenience rather than new or trendy restaurants (Salmon, 2012). The Michelin rated restaurants that are perceived to be dined at by the high-earner crowd, are in fact, quite the opposite. The affluent diners account for only three percent of the patrons at Per Se, two percent at Eleven Madison Park, and one percent at Le Bernardin (Salmon, 2012). The study shows that high earners do not travel to dine at the perceived best restaurants, differentiated concepts, or trendy restaurants but prefer to stay close to their residence and patron those establishments.

Not only do the high earners not travel to eat at fine dining locations or trendy restaurants, but they are also not typically ‘foodies.’ If a differentiated concept restaurant opens around their place of residence, they would not be inclined to try it (Salmon, 2012). This does not mean that new restaurants cannot be successful in affluent areas, but the restaurateurs need to be more thoughtful about the concept and food options.

Prior research has investigated the increase of income and increase of expenditure asking whether the increase is continuously proportional or is there a flattening of expenditure when income continuously increases. The studies found that while an increase of income does increase food away from home expenditure, it does not continue to rise at the same pace indefinitely, thus following Engel’s Law that states that while food away from home expenditure increases when income increases, it does so at a slower rate (Clements & Si, 2017; Seonok Ham et al., 1998; McCracken & Brandt, 1990).

MasterCard conducted a survey in 2014 on the Consumer Purchasing Priorities and found that Hong Kong residents had the highest expenditure on food away from home out of all Asia Pacific cities (CBRE, 2016). When grouping Hong Kong residents into four quartiles based on their income, the lowest income group quartile had a monthly household expenditure for food away from home at HK$2,153 while the highest income group quartile expenditure was HK$8,420 (Census and Statistics Department, 2016b). Overall, the total expenditure for the highest income group is 6.28 times the lowest expenditure, but for food away from home, the high-income quartile is 3.91 times the lowest, showing that while expenditure is higher, it is not in the same proportion as the overall expenditure. Discretionary income was the primary reason for dining out for Hong Kong residents, with over 48 percent of people surveyed said that the main reason for dining out was that they had extra money to spend (Kivela, Inbakaran, & Reece, 2000).

There are ample studies that show that an increase of income equates to an increase of food away from home consumption; however, little has been researched regarding food away from home consumption in cities that have both areas of high and low earners. In New York City, high-income individuals' patron restaurants close to their place of residence, but the study reported no indication of the dining patterns of the low and medium-income levels. Although there is a potential income effect regarding the restaurant selection, unless income is raised or lowered a considerable amount, the effect will be negligible. In TPUs that have a higher than median income in comparison to the other TPUs one would expect the residents to have a higher expenditure for food away from home as the studies show those who have a higher income also have a higher food away from home expenditure, as such the hypothesis is:

H2: The proximity of household income will positively impact restaurant success.

2.2.3. Household Age

The age of the patron was found to be an essential variable when it comes to the socio-demographics of who eats food away from home (Harrington et al., 2011). However, while age is significant, it is not a linear relationship. Households with infants and young children are not likely to dine out, as parents of young children will find it easier and less expensive to eat at home (Sunny Ham et al., 2004). Parsa et al. (2015, p. 80) expanded the age range from not just infants and young children but that "the geographical presence of families with children under eighteen years of age does not necessarily promote restaurant longevity."

As age increases, it is positively linked to the expenditure of food purchased away from home in a restaurant (Nayga Jr & Capps Jr, 1994; Redman, 1980). E.-J. Kim and Geistfeld (2003) found that households migrate from eating at quick-service restaurants and tend to eat more at full-service restaurants as the age of the household manager increases. However, this only goes so long until a decline is seen between the elderly and expenditure at restaurants for both single-income households (Liu, Kasteridis, & Yen, 2013; Nayga Jr & Capps Jr, 1993), and dual-income households (Frisbee & Madeira, 1986). Overall, the elderly are less likely to spend money on the expenditure of food away from home (Binkley, 2006).

The age of who eats out is an inverted U-shape (Yang Yang et al., 2017), but it is not clear where the apex of the inverted U is located. Parsa et al. (2015) found that when the population was between the ages of eighteen and twenty-four, the failure rates of restaurants were the lowest. However, the study was conducted in a college town where the area that had the lowest failure rates were also highly populated by college students. In a different study, the age of 49.42 was found to be the turning point between area growth and decline (Yang Yang et al., 2017). However, within Hong Kong households with the occupants below the age of 35 spend more on food away from home with a per capita monthly spending of HK$1,943 compared to HK$1,185 for households above 35 (Cheng, n.d.).

The current trend within Hong Kong is that the young adults are eating out more often; Cheng (n.d.) argues that this trend is due to the changing of pricing in grocery stores and restaurants. Comparing both the pricing for grocery stores and restaurants in a ten-year span found that the prices in grocery stores increased by 31 percent while restaurant prices increased by ten percent (Cheng, n.d.). The smaller increase in restaurant prices is a natural pull factor for the residents to dine away from home, no matter their age.

The previous literature shows that the age of a person affects their expenditure on food away from home. Studies show that college students and young adults eat out more often than seniors or families with children; however, as the number of children is examined in the Size of Household, those who are under 15 years of age will be excluded as they are not the primary decision-makers on where to eat. All of the studies show that age impacts the food away from home expenditure, increasing while approaching adulthood and then declining when heading into retirement. While age has shown to be an inverse U-shape, there are no studies that investigate if that correlates to restaurant success in the area around their primary accommodation within an urban city. In Hong Kong, as residents prefer to eat at establishments around their home (Planning Department, 2005b), and with the understanding that age affects the food away from home in an inverse U-shape, the overall success or failure rate of restaurants will be affected.

H3: The proximity of household age will impact restaurant success in an inverse U-shape.

2.2.4. Size of Household

The size of the household is the number of individuals that live within the residence. Yang Yang et al. (2017) theorized that restaurants prefer to situate in areas where the household size is small. However, the research by Yang et al. does not speculate how the size of household corresponds to the expenditure on food away from home. Other researchers have addressed this in two ways, first by looking at households with children under the age of eighteen, and second by looking at the overall number of individuals within the household.

The literature regarding overall household size has found that an increase of overall family size decreases the number of times the household will eat away from home (Binkley, 2006; Byrne et al., 1998; Nayga Jr & Capps Jr, 1994). Mutlu and Gracia (2006) found that an increase of 1 percent in household size will decrease lunch away from home consumption by 0.30 percent. Other research has looked into the dining location of households and found that when the household size increases, there is a modification from dining at full-service restaurants to dining at quick-service restaurants (E.-J. Kim & Geistfeld, 2003).

The literature has found that an increase in overall household size will decrease the frequency of eating food away from home (Sunny Ham et al., 2004). While the frequency of eating at restaurants might decrease, the overall amount spent increases. This is found in Hong Kong as per month, a one-person household spends HK$3,638 on food away from home, while a two-person household spends HK$5,996, and a four-person household spends HK$9,256 (Census and Statistics Department, 2016b).

The studies that consider the overall household size show that while there is a decrease in the frequency of consuming food away from home, the expenditure increases due to the number of people needing to be fed. This has been shown in Hong Kong as that the household expenditure increases as household size increases.

H4: The proximity of the overall household size will positively impact restaurant success.

2.3. Restaurant Site Characteristics

The restaurant site characteristics have two sub-categories, market area characteristics, and location factors (Fisher, 1997). The market area characteristics include all aspects regarding the surrounding stores, including the cuisine competitors or homogeneous restaurants, price competitors or heterogeneous restaurants, and convenience competition. The location factors are the distance to mass transit, hotels, malls, location within the building, floor size, and if the site is within a five-minute walk to any attractions. The restaurant site characteristics, the previous literature, and the theories behind their inclusion will be discussed in further detail in the upcoming sections.

2.3.1. Competitors

Competition can be both beneficial and detrimental to the operational success of a restaurant (Fox, Postrel, & McLaughlin, 2007). Restaurants that are situated close to each other, no matter if they are price or cuisine differentiated, become a cluster. “Clusters are geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions in a particular field that compete but also cooperate” (Porter, 2000, p. 15). Clustering similar restaurants together have been a popular way to promote an area, within New York City, New York, there are famous districts of Little Italy, Little India, Chinatown, and Koreatown that work to promote the area and cuisine but are also in competition with each other. Clusters all face the concept of coopetition or engaging in both competition and cooperation simultaneously (Brandenburger & Nalebuff, 2011).

The seminal study from researcher Harold Hotelling (1929) was one of the first to investigate the concept of clustering; Hotelling classified it as the principle of minimum differentiation. The principle theorizes why businesses cluster, and the explanation was that in order to capture their fair share of the business, it is ideal for opening competing shops next to each other. The search costs for consumers decrease when similar stores are located within close vicinity to each other. The principle of minimum differentiation argues that the "proximity to rivals is more critical than the proximity to customers" (Litz & Rajaguru, 2008, p. 479).

Hotelling (1929) is the most well-known researcher who found the proximity to rivals to be the most important, and his research has been used to explain the clustering of modern big box stores, gas stations, and car dealerships. Nelson (1958, p. 58) found that "a given number of stores dealing in the same merchandise will do more business if they are located adjacent, or in proximity to each other than if they are widely scattered." Restaurants within a cluster face a more considerable amount of foot traffic, but also competition. This added competition improves overall performance (Nickell, 1996; Nickell, Nicolitsas, & Dryden, 1997). Although the cluster can expand their presence by either combined marketing, advertising, or by word of mouth, each restaurant strives to be the best performer within the cluster.

Restaurants can benefit from economies of agglomeration by generating an increase of traffic to the cluster (Teller & Reutterer, 2008; Yang Yang et al., 2017), as complementary restaurants will attract shoppers who desire to dine at a restaurant but wish to comparison shop prior to making a decision (Fischer & Harrington Jr, 1996; Fisher, 1997; Krider & Putler, 2013). Comparison shopping different menus, pricing, atmosphere, and table availability are critical to the cost-economizing shopper who wishes to have multiple options to save on the time-cost scale (Krider & Putler, 2013), comparison shopping benefits all restaurants that are situated in the vicinity (Fields, 2007).

Although there are quite a few benefits of clustering, there is also literature that is concerned about the strategy of joining a cluster. While clusters are known for increasing the traffic to the area, there is a finite number of people a cluster will draw in, and a saturation point of restaurants will eventually occur (Fields, 2007). Once the saturation point is reached, not all of the restaurants within the cluster will continue to be successful (Poole, Clarke, & Clarke, 2006). In each restaurant cluster, there are a maximum number of profitable stores that will be able to coexist (Guy, 1994), the point of saturation is difficult to define as changing economies, tourists' behaviors, and other pre-opening factors can continuously affect the number of viable restaurants.

The saturation point findings conform to the density dependence theory; the success of a restaurant is higher when an area is in a period of growth than when there are a dense number of restaurants (Kamps & Péli, 1995) — showing that there are a finite number of businesses that can operate in a highly dense area (Hannan & Carroll, 1992). When the number of restaurants is low and slowly growing, the failure rate is also low, but at a certain point above the carrying capacity, increases in the number of restaurants cause the failure rate to increase (Kamps & Péli, 1995). The denser an area is, the more it breeds competition, and with competition comes failure (Singh, 1993). Areas have a natural carrying capacity for different types of restaurants (Van Witteloostuijn, Boin, Kofman, Kuilman, & Kuipers, 2018), and in turn, the success of an area is an inverse U-shape when associated to the changes of restaurant density.

The correlation between restaurant density and restaurant success was studied in Denver, Colorado, between 2007-2013 by Parsa et al. (2019). The study found that, on average, the low-density areas have a 30.83 percent success rate, and high-density areas have a 33.09 percent success rate, while the medium-density areas have a 36.08 percent success rate. While the study tried to theorize that the type of cuisine made a difference, as low food cost cuisines showed to survive in highly dense areas, it could not generalize it across all cuisines as the study found Mexican cuisine was subject to failure in all three density areas, even with a relatively low food cost.

Until the carrying capacity is reached, restaurant clusters can be beneficial, as within a cluster, the whole is greater than the individual site (Porter, 2000; Prayag et al., 2012; Sevtsuk, 2014). However, the increase of competition has a direct relationship with the probability of bankruptcy (Nickell et al., 1997) only if they cannot differentiate from their competition (Parsa et al., 2005). Pollitt (2003, p. 14) also found that the evidence of competition is "not as overwhelming and conclusive as one might think." Although there is research showing both the positives and negatives of restaurant clusters, creating an area that is known for its diverse restaurants has been shown to increase traffic and comparison shopping.

2.3.1.1. Cuisine Competitors

Restaurant clusters may be beneficial to restaurants as they increase traffic and comparison shopping to the area, but does the additional number of identical cuisine competitors affect the possibility of restaurant success? The restaurant cuisine competitors are those who are homogeneous and are vertically differentiated. A vertical differentiation occurs when there are several similar products, but with a clear difference in quality and/or price (Piana, 2003). Take Italian cuisine as an example; in one area, there could be an osteria, a small bar that served pasta or paninis to local workers, a trattoria, a traditional family-run establishment serving rustic or regional food, or a ristorante, the more traditional restaurant serving higher-end food. While there are other types of Italian food establishments, these showcase cuisine differentiations. They all offer homogeneous cuisine, Italian food, and sometimes offer the same menu items, but the price, quality, service, and atmosphere separate them on a vertical scale. In a study of Hong Kong residents, the second most important aspect of restaurant selection is the type of food (Kivela, 1997), and not the menu price, lending support to the importance of cuisine differentiation in restaurant clustering as potential diners will have a cuisine in mind and will look for the restaurant that has the specific cuisine, no matter the price.

Cuisine differentiated restaurant clusters are often attributed in the United States to the minority groups or immigrants who desire to become entrepreneurs (Kerr & Mandorff, 2015). Immigrants are the reason behind famous restaurant clusters within New York City districts, such as Little Italy (Henderson, 2006) and Chinatown (Waxman, n.d.). Homogeneous clusters attract potential customers who have a desire for a specific cuisine but are uncertain of the price or quality that is desired (Bester, 1998).

2.3.1.2. Price Competitors

Restaurants have both cuisine competitors, who are homogeneous and are differentiated in terms of price or quality, and price competitors, who are heterogeneous. Competitors by price fall on the horizontal differentiation spectrum; these competitors are linked by differentiation in styles, or for the sake of restaurants, cuisines (Piana, 2003). In terms of restaurants, potential diners might decide how much they want to spend for the meal, and then comparison shop between multiple restaurants to decide where to eat. In this instance, the type of cuisine is not the focus, but the price.

Clusters of price differentiated restaurants are starting to become more popular than clusters of cuisine differentiated restaurants. Within New York City, clusters such as the Sandwich District, the New Midtown, and the Gastro Pub Vortex are growing in popularity (Gordinier, 2015). Research within Hong Kong has found that the price is the third most important aspect of selecting a restaurant to dine at (Kivela, 1997), supporting the importance of locating within a price differentiated restaurant cluster.

The restaurants that take advantage of positioning themselves among heterogeneous competitors benefit from potential groups of diners trying to decide where to eat, trying new cuisine, and the overall search for the ideal dinner location (Krider & Putler, 2013). These clusters benefit from incomplete information regarding cuisines and prices (Konishi, 2005), and can find it to be more profitable when situated by price differentiated restaurants rather than cuisine differentiated restaurants (Fischer & Harrington Jr, 1996).

2.3.1.3. Convenience Competitors

There is the possibility that a potential restaurant patron has the desire to dine out, but while browsing menus, they are unable to decide where to visit and instead decide to get a sandwich from a convenience store, or a pre-made salad from a supermarket. These establishments are considered convenience competitors. They do not have table service, or tables to sit down at, they might have a counter to lean on, but they are designed for eating on the go or buying and eating elsewhere.

Convenience food makes life easier for those who are eating on the go (Scholliers, 2015). This type of food is popular for those who are time-poor (Darian & Cohen, 1995), cost-conscious (Jackson & Viehoff, 2016), those who are eating alone (Candel, 2001; Verlegh & Candel, 1999), or those who live in areas further away from restaurant clusters (Lupton, 2000). Diners who eat this type of cuisine are those who do not have the time, or physical desire, associated with both deciding what restaurant to eat at, or what to cook for the next meal (Brunner, Van der Horst, & Siegrist, 2010; Jackson & Viehoff, 2016; Man & Fullerton, 1990).

Stores that sell convenience foods are not the cuisine or price competitors of restaurants. However, they do have the possibility of pulling potential restaurant customers away by the inexpensive cost, timesaving, and lack of decision making needed. These convenience stores will make an impact on the business of the restaurants and need to be considered when selecting a restaurant location.

2.3.1.4. Competitors Conclusion

Competition can be a positive aspect for restaurants; it does not allow restaurateurs to become lethargic and promotes healthy rivalries. Different types of competition, price or cuisine differentiated restaurants, provide a multitude of different elements for the restaurateur to make adaptations to improve food, service, offerings, marketing, in addition to other post-opening decisions. However, areas have a natural carrying capacity where, when exceeding that point, success is low, and failure is high. Opening a restaurant in a less dense area can provide a foothold in the market, while a high-density area will require more effort, and a differentiated style of a restaurant will be required to be successful. As the density dependence theory argues, the potential success rate is higher when entering a less dense area than when entering a highly dense area.

H5: The competitor density, in the vicinity, impacts restaurant success in an inverse U-shape.

The competitor density hypothesis can be further expanded to the three types of competitors. The expanded hypotheses are:

H5a: The density of price competitors, in the vicinity, impacts restaurant success in an inverse U-shape.

H5b: The density of cuisine competitors, in the vicinity, impacts restaurant success in an inverse U-shape.

H5c: The density of convenience competitors, in the vicinity, impacts restaurant success in an inverse U-shape.

2.3.2. Distance to Mass Transit

The human race does not like to exert any extra effort (Hull, 1943), and even Marcora (2016) argued that people are inherently lazy. This laziness is especially true when it comes to physical tasks, including walking. Research has shown that people do not want to walk far (Lorch & Smith, 1993), and a distance of a quarter mile, approximately a five-minute walk, is often used as the acceptable distance that people will walk before using other types of transportation (Atash, 1994). This was first observed in 1924 when eighty percent of shoppers showed a resistance of walking more than two blocks from where they parked their car (Longstreth, 1997). Although another study has shown that trips longer than five minutes have been seen to be common, the research only found that 16 percent of people walked, and the median distance was a half-mile, so while the participants did walk slightly farther, only a few people actually walked (Yong Yang & Diez-Roux, 2012). Overall, the effort of walking involves the distance decay effect, that the longer the distance, the increase of the reluctance to travel to the location (Reimers & Clulow, 2004).

The potential restaurant patrons seek the convenience of getting to the restaurant (Jang & Mattila, 2005), and the ease of getting to the restaurant is a critical factor of selecting where to dine (Njite et al., 2008; Sevtsuk, 2014). According to research conducted in Hong Kong, the location was the fifth most important attribute for reasons patrons select a restaurant (Kivela, 1997). Koo, Tao, and Yeung (1999) looked further into the location and found that the majority of the residents of Hong Kong prefer to dine within urban areas, over restaurants located in either rural or touristy areas. Therefore, the potential customers travel on a daily basis to and from work and desire to dine in urban environments and to get to the restaurant quickly and conveniently, in addition to the five-minute walk being the typical maximum distance.

In a metropolitan city, one of the primary ways of travel is by the subway, within Hong Kong it is called the Mass Transit Railway (MTR), as this is one of the main ways of travel throughout the city. The 2016 Census reported that over 80 percent of residents participating in the labor force had to travel to get to work, with the majority traveling via the MTR; as such, the MTR exit is a waypoint between the starting and ending places of the residents' day. The MTR transports over 90 percent of the 11 million trips that are taken on a daily basis (Lo, Tang, & Wang, 2008) and is the most popular method of transport available. The MTR has 553 exits from the 90 stations, as humans desire not to walk far, it is expected that restaurants that open within a five-minute walk of the MTR exit will have a higher number of potential customers than restaurants located outside of the radius. Within Hong Kong, the mean walking time between a person's destination and an MTR exit was found to be four minutes (Lo et al., 2008); however, this study will utilize the accepted distance of a five-minute walk and with the mean walking speed in an outdoor commercial area is 73.99 meters per minute (Lam & Cheung, 2000), a distance of 369.95 meters between the restaurant and the closest MTR exit will be used. Therefore, the location within a five-minute walk will provide greater potential success.

H6: The distance of the site to the MTR will negatively impact restaurant success.

2.4. Chapter Summary

The annual food expenditure of the sample of Hong Kong residents in the latest survey was 27 percent of their household expenses, of that 66 percent was spent on food away from home (Census and Statistics Department, 2016a). Overwhelmingly, those who view their time as more valuable also see an increase in food away from home expenditure (McCracken & Brandt, 1990; Nayga Jr & Capps Jr, 1994). The restaurants that are dined at depend on multiple factors, both the socio-demographic and the physical location.

Regarding the socio-demographic sector, "a larger supply of restaurants could be found in neighborhoods with a higher population density, higher median age, higher median household income, higher education level, larger percentage of renter-occupied housing units" (Yang Yang et al., 2017, p. 39) while a smaller number of restaurants are found in neighborhoods with larger average household size and a higher percentage of owner-occupied housing units (Yang Yang et al., 2017). Is this purely by coincidence, or do these factors help to predict the success of a restaurant? While some argue that this is due to the convenience factor of restaurant sites and not a predictor of success (Frisbee & Madeira, 1986), the literature shows differently.

Do the physical restaurant sites make any difference in the success or failure of a restaurant, and if they do, to what extent? Clustering has been a factor for centuries and helps to draw customers into an area, no matter for a price differentiated cluster or a cuisine one. How much do these clusters factor into the longevity of a restaurant? Most site characteristics seem to have a positive influence, but some have not been tested and especially not together in a model to predict restaurant success.

Chapter 3. Research Methodology

3.1. Chapter Introduction

This chapter introduces the sample selection criterion, the dependent, and independent variables. The methods that are utilized to retrieve, clean, and sort the data will be detailed. Afterward, the explanation of how the study will use logistic regression to look at the marginal effects of the different attributes and the overall impact on the restaurant's success or failure will be presented.

The chapter will then detail two predictive models, logistic regression, and the artificial neural network. These two models will be separately examined, and explanations on the reason for inclusion in the study, their assumptions, model generations, formulas, and the validation techniques used to analyze the data set are presented. The chapter will finish with an explanation of the method of selecting the ideal model for selecting future restaurant sites.

3.2. Observations

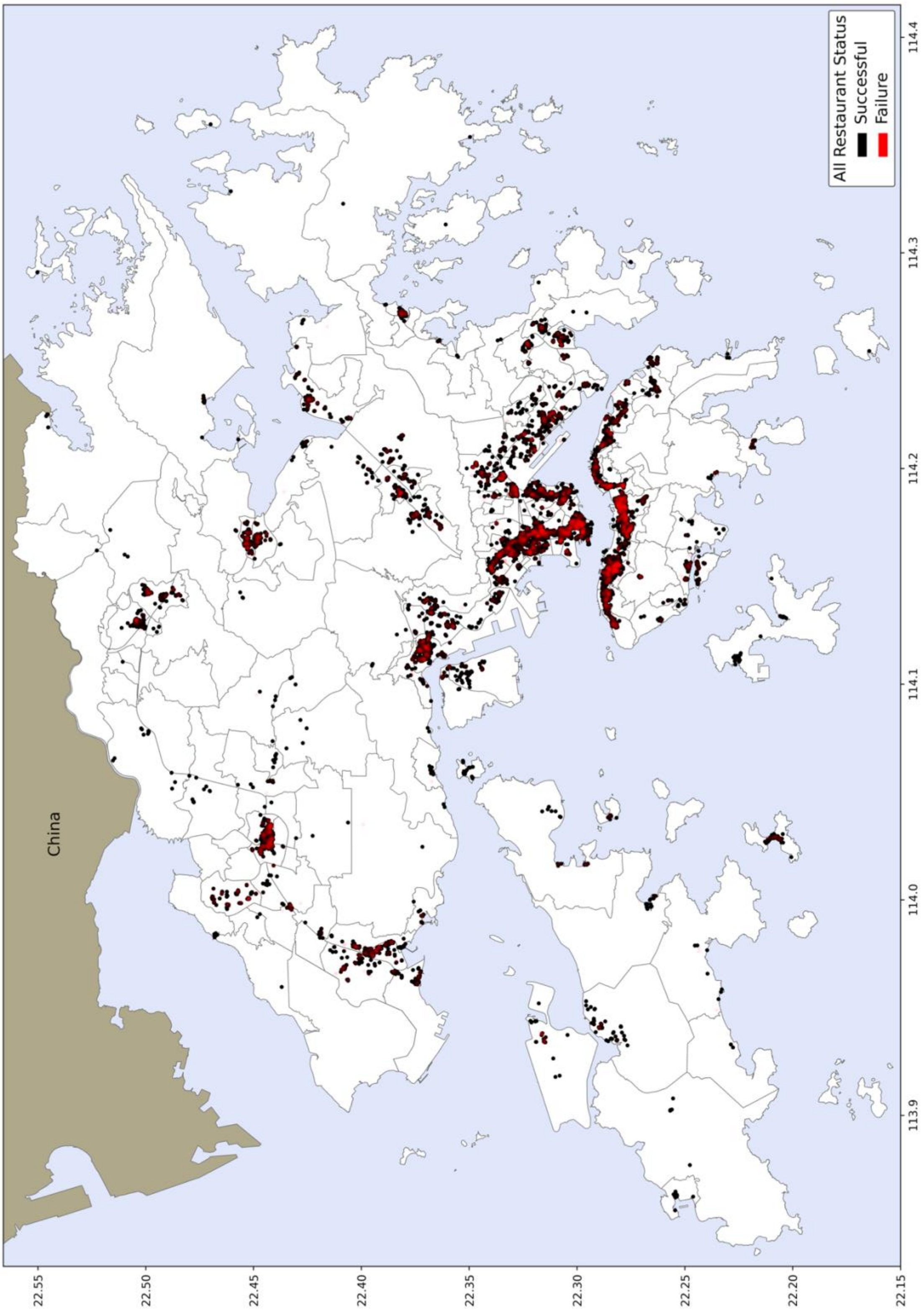
The study utilizes all legally licensed restaurants within Hong Kong with licenses issued to them into one of the three categories of general restaurant license, light refreshment restaurant license, or marine restaurant license. The general restaurant license allows the restaurant to cook and sell any food for consumption within the premise. The light refreshment restaurant license requires a restaurant to select a limited offering food category, for example noodles, cakes, or sandwiches, and may only prepare food in that category. The marine restaurant license is similar to the general restaurant license with the caveat that the restaurant is on a vessel (Food and Environmental Hygiene Department, 2018).

There were 31,073 food licenses in operation between 31 December 2015 and 31 December 2019. After removing the licenses that were either in duplicate, temporary, or non-restaurant licenses (convention centers, catering companies, or cinemas), there were 24,257 licenses. Among those licenses, 3,709 could not be verified as a restaurant, temporary, duplicate, or non-restaurant and were omitted, resulting in 20,548 licenses that could potentially be included in the study.

The study focuses on newly opened restaurants in the years 2016, 2017, and 2018; as such, the 11,738 restaurants that opened before 1 January 2016 and the 2,100 restaurants opened in the year 2019 are not included as they began operations outside of the study window. The remaining 6,710 restaurants are the main focus of the study and will be utilized in calculating the success and failure rates as well as the two predictive models. While the research focuses on 6,710 restaurants that opened in 2016, 2017, and 2018, all 20,548 restaurants are utilized in calculating the number of price and cuisine competitors. The 20,548 restaurants' locations are located in Map 1, created with the Python code in Appendix 15, while Map 2 showcases the 6,710 successful and failed restaurants included in the study, and was created in a separate, but similar, Python program. The restaurants included in the study are situated in similar positions and clusters, as the totality of restaurants.

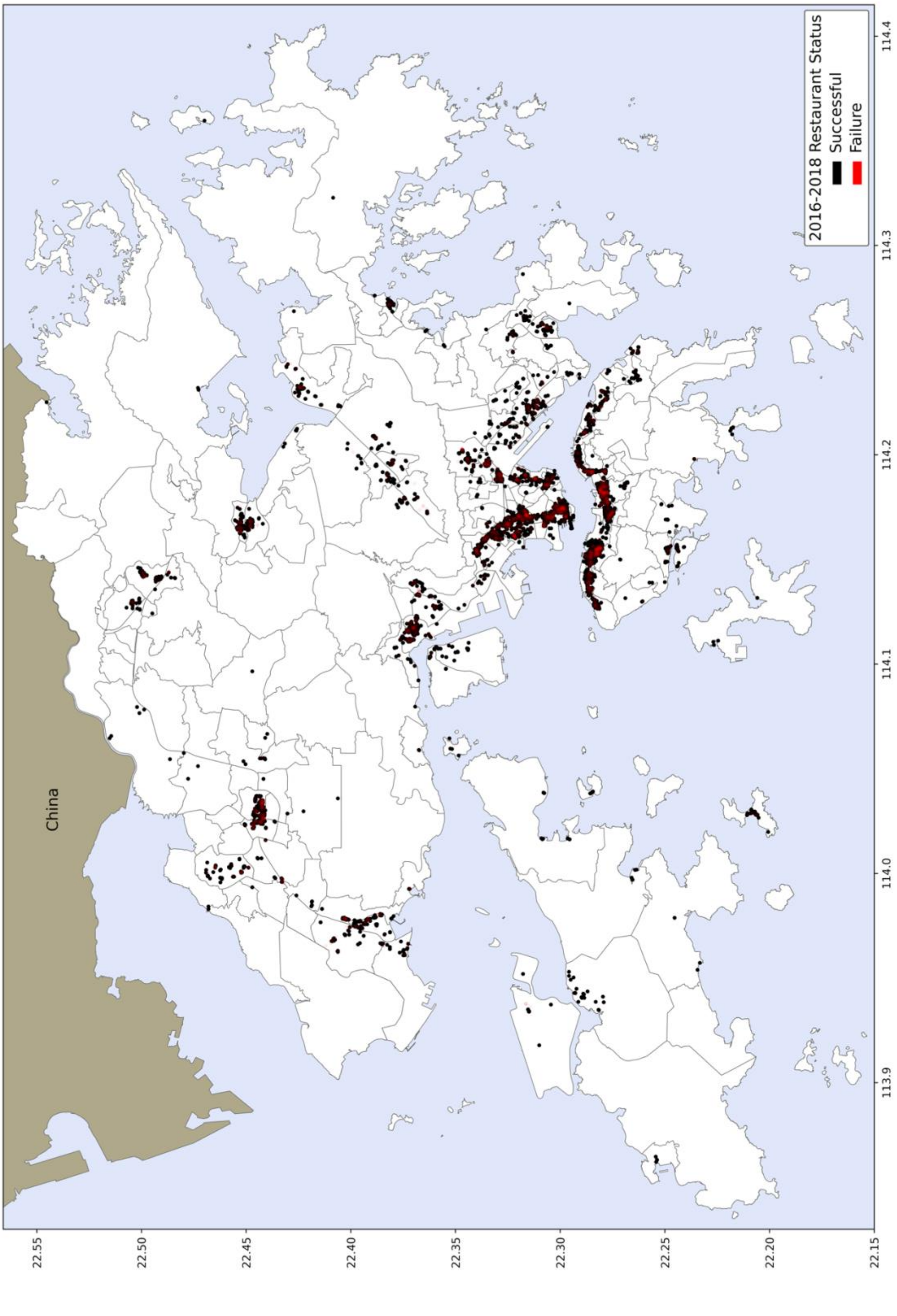
Map 1

Location of 20,548 Restaurants



Map 2

Location of 6,710 Restaurants



3.2.1. Restaurant Data Analysis

The data retrieved from the Hong Kong Government's dataset regarding the restaurant licenses are in individual files, one for each day that the restaurant list is published. To combine the daily data into an individual dataset, a custom program is written in Python to automate the process; the Python code is available in Appendix 2.

The program creates a master database of all restaurants by opening the daily restaurant license file and adding necessary information, including the date, license number, restaurant name, restaurant license, and district code. Depending on the option selected, the date can either be the first or last observation's appearance. Each time a new file is opened, it compares the restaurants in that file to the master database and only copies the restaurants that have not been previously listed. The final dataset was then created by combining the two options' outputs to create a single file that lists all restaurants, license numbers, names, addresses, and the first and last day of appearance.

3.3. Dependent Variable

The dependent variable will be the individual restaurant's success or failure. While the failure of a business can have different definitions depending on the researcher, this study will build upon S.-S. Park and Hancer (2012, p. 313), that the "term business failure is used to describe a firm's financial situation in which the amount of payment obligation exceeds its financial reservoir." In this context, this study will use failure as the establishment's closure, and restaurant success will be classified as still in operation (J. Everett & Watson, 1998).

The dependent variable is binary and will be classified into two categories, success, or a failure. A value of 1 will be given to the successful restaurants, and a value of 0 will be given to the restaurants that have failed. This comes from the final dataset created; once a restaurant stops appearing on the list, it will be considered permanently closed. Restaurants that are still appearing as of 31 December 2019 will be considered as successful.

3.4. Independent Variables

Three data sets are needed to compare the dependent variable to the influences exerted on it by the independent variables. They are the individual restaurant data, area attributes, and site characteristics. The process of collecting these three sets of variables is explained below and the table is found in Table 5.

3.4.1. Area Attributes

The area attributes are an essential portion of the independent variables relaying who lives in an area around the restaurant site, including the area's density, income, age, and household size. This data set includes the population dynamics of the TPU in which the restaurant is situated. The data utilized to compare the pre-opening factors to the dependent variable come from the 2016 Hong Kong census. The study uses the data from the TPUs to analyze the specific area attributes, as this was shown to be a successful way to analyze the geographical area of Hong Kong in the research by Fang, Li, and Li (2019).

Two different calculations will be carried out to compute the individual restaurants' independent variables for the area attributes. For the total population, the individual TPU's population will be divided by the total square meter of the TPU to provide the density per square meter. Then each restaurant observation will be the center point of a circle with a radius of 369.95 meters, as this length is used as the distance in other variables. Based on the circle's circumference, the area of the TPUs that the circle lays on top of will be calculated to the square meter. The restaurant's independent variables of the total population will be the summation of the calculated areas of the TPUs multiplied by the specific distributed variable per square meter.

The calculation for the other area attributes, household income, age, household size, ethnicity, gender, education, flat, and employment will be done by looking at the proportion of the overlapped TPU. The total square meter overlap will be divided by the area of the circle to provide the list of TPUs and the proportion of each. The attribute will be the summation of the proportion of the individual TPU multiplied by the specific attribute. A program was written in Python to examine the overlap of the individual TPUs and each restaurant; the code is in Appendix 4.

3.4.2. Site Characteristics

The site characteristic independent variables include the distance to the MTR exit, and the number of competitors, separated into the cuisine, price, and convenience competitors. The cuisine competitors have the same cuisine no matter the price, while the price competitors are the restaurants that have the same price range, no matter the cuisine. As such, all restaurants within a circle with a radius of 369.95 meters of the observed restaurant that were open concurrently are potentially included in the competitor variables. Two separate programs were written in Python to perform the count for the two competitor variables; the code is in Appendix 6 and Appendix 7.

Grocery stores, food factories, convenience stores, and other similar stores are not included in the restaurant set; however, they are included in investigating convenience competitors. These stores are excluded due to the initial requirement that the establishment's focus has to be on the preparation and serving of food. These types of establishments are included because while a potential patron is trying to decide between restaurants within a TPU, there is a potential that they are unable to decide and instead purchase a sandwich, salad, sushi, or other grab-and-go food from a convenience store.

The convenience competitors must be within the circumference of a circle with a 369.95 meters radius and the observed restaurant located in the middle. The dataset is created similarly to the restaurant dataset by combining daily license information and including those that are marked as Composite Food Shop License, Bakery License, or Siu Mei and Lo Mei Shop License. The program that creates the dataset is in Appendix 3. The program that then utilizes the convenience dataset and calculates the number of convenience competitors for each restaurant is in Appendix 8.

The variable of the distance to mass transit or, more specifically, the distance between the restaurant and the closest exit of the MTR. As of 2019, the 90 MTR stations have in total 553 exits. The study will utilize a Python program to analyze the restaurant's distance to the closest MTR exit. The results will be categorized into a binary format of close (1) or not close (0) based on the walking time of five minutes, or 369.95 meters, between the restaurant and the closest MTR exit. The Python code is available in Appendix 9.

3.4.3. Control Variables

The TPUs within Hong Kong are diverse and have different socio-demographic area attributes. To ensure that distinctive area attributes do not affect the model’s findings, ten control variables will be added based off the literature. The five area attribute control variables are the resident’s ethnicity (Tan, 2010), gender (Harrington et al., 2011), education level (Graham & Paul, 2010), flat ownership (Parsa et al., 2015), and employment status (Smith, 1983). The data will be retrieved from the 2016 census in the same manner as the other area attribute independent variables.

The five site characteristic control variables include locating within a mall (Gonzlez-Benito & Gonzlez-Benito, 2005) or hotel (Strate & Rappole, 1997), the floor level (Dock et al., 2015; Letail, 1992), situating close to a tourist attraction (Yang Yang et al., 2017), and the aggregate review score (Zhang, Ye, Law, & Li, 2010). The cuisine type, price, aggregate review score, and GPS coordinates of each restaurant are collected from OpenRice.com by utilizing a web scraper. The Python code for the scraper is in Appendix 5.

The price range of a restaurant, spending per person, as shown in OpenRice.com, has six different price ranges, Below HK$50, HK$51-100, HK$101-200, HK$201-400, HK$401-800, and above HK$801; this variable requires treatment of being converted into dummy variables before the model being constructed. Of the 6,710 restaurants, they have the frequencies of these variables as follows.

Table 1

Statistics of Menu Price

Menu Price Range	Frequency of Observation	% of Total
Below HK$50	1,970	29.359%
HK$51-100	2,008	29.925%
HK$101-200	1,749	26.066%
HK$201-400	781	11.639%
HK$401-800	138	2.057%
Above HK$800	64	0.954%
All	6,710	100%

Typically, if a categorical level has less than five percent of total observations, it should be combined with a similar level (Ray, 2015). As the HK$401-800 and Above HK$800 are both below the threshold of five percent, these two are combined with the HK$201-400 variable and are relabeled as Above HK$201. The revised frequencies of the categorical variables follow.

Table 2

Revised Statistics of Menu Price

Menu Price Range	Frequency of Observation	% of Total
Below HK$50	1,970	29.359%
HK$51-100	2,008	29.925%
HK$101-200	1,749	26.066%
Above HK$201	983	14.650%
All	6,710	100%

OpenRice.com has a potential of 66 different types of cuisines in either 37 specific countries (Italy, France, Russia, etc.), six regionals (African, International, Mediterranean, etc.), 22 Chinese types (Beijing, Hakka, Hunan, etc.), or Hong Kong Style. These 66 different cuisines were initially reduced to six based on their regional locations: Hong Kong Style, other Chinese, Japanese, other Asian, Mediterranean, and other Western; the restaurants have the following frequencies.

Table 3

Statistics of Restaurant Cuisine

Cuisine Types	Frequency of Observation	% of Total
Asian	868	12.936%
Chinese	1,166	17.377%
Hong Kong Style	1,935	28.838%
Japanese	939	13.994%
Mediterranean	295	4.396%
Western	1,507	22.459%
All	6,710	100%

The Mediterranean cuisine variable is below the five-percent threshold and was combined with the Western cuisine as that was deemed the most similar. The final frequency of cuisines is as follows:

Table 4

Revised Statistics of Restaurant Cuisine

Cuisine Types	Frequency of Observation	% of Total
Asian	868	12.936%
Chinese	1,166	17.377%
Hong Kong Style	1,935	28.838%
Japanese	939	13.994%
Western	1802	26.855%
All	6,710	100%

The price and cuisine variables are transformed into dummy variables. The categorical cuisine variable is separated into five mutually exclusive dummy variables. The Hong Kong Style is used as the base variable and will be excluded in the models. For the categorical price variable, it is separated into four mutually exclusive dummy variables. The dummy variable Below HK$50 is used as the base variable and will be excluded from the models.

The variable of the location to the street is based on the vertical location within a building. The study utilizes the address data from the final restaurant dataset to create the binary variable of ground (1) and not ground (0). If the restaurant license includes a listing of ‘ground floor,’ ‘upper ground floor,’ or ‘lower ground floor,’ or the abbreviations of the three, ‘G/F,’ ‘UG/F,’ ‘LG/F,’ then the value of 1 will be given; otherwise, the variable will receive the value of 0. The variables of MallLoc and HotelLoc will take a value of 1 if the site resides in either a mall or a hotel building; otherwise, the variable will receive a value of 0; the program code for the mall location and inside a hotel is found in Appendix 12 and Appendix 10.

The distance to attraction variable will be binary and will receive a value of 1 if the site is within 369.95 meters, a five-minute walk, to an attraction. Otherwise, it will be considered not close to an attraction and will receive a value of 0. The attractions utilized are the top list of places to go as listed on the Discover Hong Kong website. The Python program to examine if the observed restaurant is close to the tourist area is found in Appendix 11. The final control variable will be the overall review score for each restaurant. This will be used to control the restaurant's popularity or monetary success and allow the model to consider only the site factors.

Table 5

Dependent and Independent Variables

Code	Definition of Variables
Dependent Variable	
OpStatus	Current Operational Status (1 = Success)
Independent Variables	
ResDen	Number of Total People
Income	Median Monthly Household Income
Age	Median Age
NumHousehold	Mean Household Size
DistMTR	Inside 5 Minutes to an MTR Exit (1 = Close)
CuisineComp	Number of Competitors by Cuisine
PriceComp	Number of Competitors by Price
ConvComp	Number of Competitors by Convenience
Control Variables	
Ethnicity	Percentage of Chinese Ethnicity
Gender	Percentage of Males
Education	Percentage with Post-Secondary Degrees
Flat	Percentage of Owner-Occupied Flats
Employment	Percentage of Labor Force Participation
AggReview	Aggregated Review Score
StreetLoc	Location on Ground Floor (1 = Yes)
MallLoc	Located Inside a Mall (1 = Yes)
HotelLoc	Located Inside a Hotel (1 = Yes)
DisAtt	Located Inside an Attraction Area (1 = Yes)
PricePoint	Price Range of the Restaurant
Cuisine	Cuisine Type of the Restaurant

3.5. Modeling

To examine the first research question, which is,

1. Which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure?
    a. What is the most critical socio-demographic area attribute a restaurateur needs to pay attention to when selecting an area?
    b. What is the most critical restaurant site characteristic that a restaurateur needs to consider when selecting a site?

A logistic regression model will be utilized, and the marginal effects will be examined to determine the impact that each variable has on the overall success or failure of the site.

To examine the second research question, which is,

2. Considering the potential influence of the overall site characteristics, can a model be created to aid in restaurant success?

Two separate non-nested models will be utilized to see if they can predict restaurant success or failure based on the pre-opening factors. The models will be a logistic regression model and an artificial neural network.

In the following sections, each model will be introduced, and an explanation of the reasons for selection, their background, assumptions, and then the model will be presented with relevant formulas, and information about how the model will be validated. The final section will introduce the comparison methods for the two models.

3.5.1. Logistic Regression

Binary logistic regression will be utilized to examine both the marginal effects of the site characteristics and the first of the two predictive models that will be used to look at if the pre-opening factors influence the probability of a restaurant being successful. Logistic regression has the goal to find the best model to describe the relationship between the dependent variable and the multiple independent variables (Pohar, Blas, & Turk, 2004; Storey, Keasey, Wynarczyk, & Watson, 1987). The binary dependent variable is defined in Section 3.3. The independent variables will be the area attributes and site characteristics. The reason for selection, background, assumptions, and formula will be further explained in the following sections.

3.5.1.1. Background and Reason for Selection

The research conducted by Beaver (1966) was one of the first studies to use financial ratios as a predictor of failure. Using a dichotomous group of industries of bankrupt or non-bankrupt, he was able to examine the finances of both to find the trend and be able to predict the bankruptcy of future companies. Altman (1968) expanded upon Beaver's work and pioneered the research on creating a model that separated observations into two classes based on the independent variables. In his seminal paper, he argued that multiple discriminant analysis (MDA) was the most viable option due to the advantage that multiple independent variables could be grouped into a singular model, and there was no need to examine the variables individually (Altman, 1968; Gu, 2002). However, the MDA approach relies on assumptions that are typically ignored (H. Kim & Gu, 2006a, 2006b; Lennox, 1999; Ohlson, 1980), specifically failing to check the multivariate normality of the predicting variables, bringing the MDA models' accuracy into uncertainty. However, the logistic model does not have such restrictions and is free from the multivariate normal distribution assumption (H. Kim & Gu, 2006b; Ohlson, 1980; Pohar et al., 2004).

Logistic regression predicts a categorical outcome by analyzing independent variables (Field, 2013; Pohar et al., 2004; Storey et al., 1987). While the categorical outcome can be either binary or multinomial, this study will only focus on the binary logistic regression as the dependent variable is dichotomous. The use of logistic regression to predict bankruptcy was first proposed by Ohlson (1980) and soon became the preferred method over multivariate analysis (Youn & Gu, 2010).

The reason for selecting the logistic regression over the MDA is that “it has been shown that the logistic regression is often preferred over discriminant analysis in practice” (Zhang, Hu, Patuwo, & Indro, 1999, p. 25). The MDA method is now considered a dated practice, and as such, the more modern approaches are deemed to be more appropriate. Logistic regression has also been shown to be accurate in its prediction accuracy (Youn & Gu, 2010), and various applications utilize the predictive analysis (Dimitras, Zanakis, & Zopounidis, 1996).

3.5.1.2. Assumptions

Five assumptions need to be investigated before building a logistic regression model. The five assumptions that need to be checked are:

1. The dependent variable has to be in a binary format.

2. The observations need to be independent of each other.

3. There is no multi-collinearity among the independent variables.

4. There is linearity between the independent variables and log odds.

5. There is a large number of observations. For each independent variable, a minimum of 10 cases of the least frequent outcome is suggested to provide reliability. With nineteen independent variables, the minimum number of observations needs to be 1,900 (19 * 10 / 0.1).

(Lani, n.d.)

3.5.1.3. Equation

To examine the marginal effects of the site characteristics, the entire sample observations will be utilized. To start building the logistic regression model, the first step is undertaking the odds ratio. The odds ratio is based on the Bernoulli distribution of probability, which states that in a binary option:

The probability of X being 1 is P(X = 1) = p
The probability of X being 0 is P(X = 0) = 1 – p
$X \epsilon [0,1]$

The dependent variable is classified as success = 1, and failure = 0, so that it can be written as:

P(E) = Probability of success
1 – P(E) = Probability of failure
$0 \leq P(E) \leq 1$
P(E) + (1 – P(E)) = 1

As the odds ratio is the probability of something happening divided by the probability that something does not happen, this can be expressed by:

$$\frac{P(E)}{1 - P(E)} = Odds\ of\ Success \tag{1}$$

The log odd is another way of expressing the odds ratio; this is expressed as:

$$logit = \ln\left(\frac{P(E)}{1 - P(E)}\right) \tag{2}$$

The logit curve is shown as:

Figure 1

Logit Function Graph

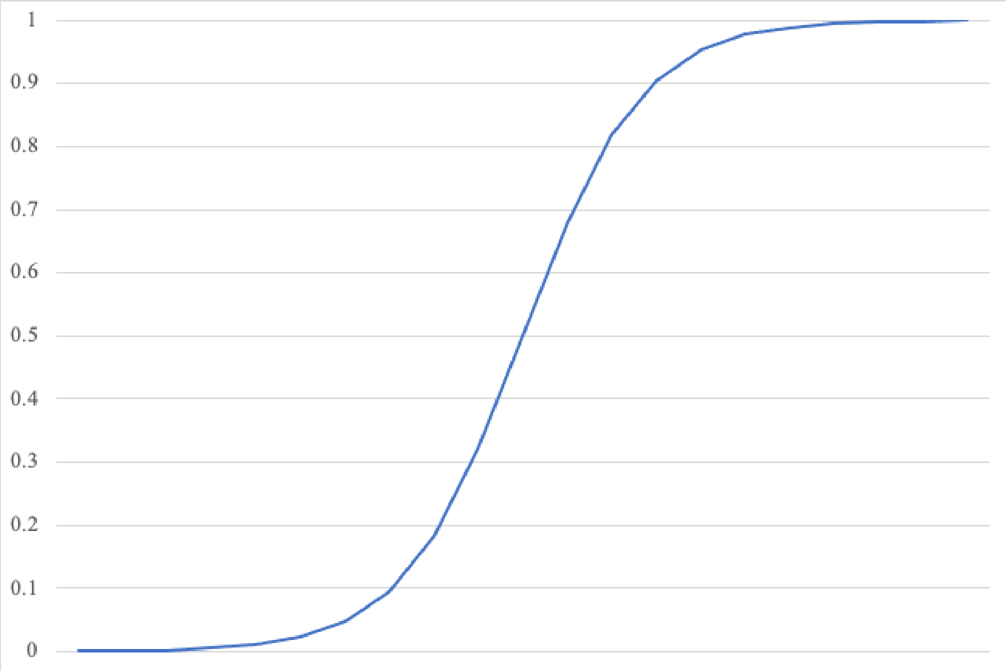
As the dependent variable is located on the y-axis and takes a value of 0 or 1, the logit function graph needs to be inverted. The inverted logit function, also considered the sigmoid function, takes the expression of:

$$logit^{-1}(X) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \tag{3}$$

This inverted logit, sigmoid, the function produces the standard logistic curve modeled in Figure 2.

Figure 2

Logistic Curve



The logistic curve has values that approach both 0 and 1 but extend infinitely and never fully reach 0 or 1. The logistic curve allows the plotting of the regression equation.

The logistic regression equation can be written as:

$$\ln\left(\frac{P(E)}{1 - P(E)}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i X_{si} \qquad (4)$$

where:

P(E)	=	Probability of a restaurant 's' success in the potential site
β_0	=	The intercept
n	=	How many independent variables of the potential site used
β_i	=	Coefficients for the i^{th} independent variable
X_{si}	=	The i^{th} independent variable of the potential site for restaurant *s* ($1 \leq i \leq n$)

The intercept and individual independent variables' coefficients are calculated using a maximum-likelihood function (Mazzocchi, 2008; Pohar et al., 2004).

As $\ln(x) = y$ can also be written as $x = e^y$ then:

$$\frac{P(E)}{1 - P(E)} = e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} \tag{5}$$

where:

e = the base of the natural logarithm (≈ 2.718281828)

As the goal is to find the estimated probability, the equation needs to be solved for P(E). The steps are as follows:

$$P(E) = e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} * \left(1 - P(E)\right) \tag{6}$$

$$P(E) = e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} - \left(e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} * P(E)\right) \tag{7}$$

$$\left(P(E) + e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}}\right) * P(E) = e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} \tag{8}$$

$$P(E) * \left(1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}}\right) = e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}} \tag{9}$$

$$P(E) = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_{si}}} \tag{10}$$

$$P(E) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i X_{si})}} \tag{11}$$

The output of P(E) will be a value between 0 and 1. To separate the outcomes into the predicted probability of failure, the study will utilize the threshold value suggested by H. Kim and Gu (2006a), that splits the outcomes at the 0.5 level. Therefore, if P(E) is greater than 0.5, then there is at least a 50 percent probability that the site will be successful. The calculations for the probability that restaurants will be successful will be based on equation eleven above. To run the logistic regression, a program was written in Python (Appendix 13) to determine which variables provide the best fitting model.

3.5.1.4. Marginal Effects

For the first research question, the logistic regression marginal effects for the independent variables will be analyzed. The marginal effects will allow the study to show what impact an increase of the independent variable would have on the probability of success. It will strengthen the argument of what variables are most important for restaurateurs to be aware of when selecting the site location.

3.5.2. Artificial Neural Network

The artificial neural network (ANN) is one of many machine learning networks considered for this study. Other methods include the decision tree, projection pursuit regression, fuzzy set theory, support vector regression, or boosted regression. While these methods were ultimately rejected, the ANN was deemed the most suitable due to the sigmoid function's influence. The following sections will look at the background, further reasons for selection, a discussion on how the ANN works with the formulas, as well as the training and validation process

3.5.2.1. Background and Reason for Selection

The ANN was initially conceptualized by McCulloch and Pitts (1943) and was inspired by the central nervous systems of animals and, more specifically, how animals process data and make decisions (Anandarajan, Lee, & Anandarajan, 2001; Jain, Mao, & Mohiuddin, 1996; Kumar & Ravi, 2007; McNelis, 2005; Sietsma & Dow, 1988; Yang Yang, Tang, Luo, & Law, 2015). ANNs are essential to machine learning algorithms designed to understand and classify the underlying patterns of non-linear data that are not evident in the dataset (Yang Yang et al., 2015; Zhang et al., 1999). The benefit of ANNs is the ability to estimate the complicated relationship between independent variables (Bishop, 1995; Bondarenko, Borisov, & Aleksejeva, 2015).

The ANN differs from other machine learning approaches in the fact that it does not run a defined set of sequential steps to try to fit the model or use inference paths that are popular in structural equation modeling; alternatively, the process happens simultaneously to determine the closest match (Denton, Hung, & Osyk, 1990). One of the pitfalls of utilizing ANN is that the effect that each independent variable has on the overall model is not clearly stated as the weights do not indicate if the variable is of importance, but that it is just a predictor of how to solve the model (Palmer, Montaño, & Franconetti, 2008).

S.-S. Park and Hancer (2012) noted that ANNs outperformed MDA, the method Altman (1968) used in his pioneering probability model, regarding accuracy and robustness. In other research dealing with bankruptcy prediction, ANN has been shown to perform at a higher rate than MDAs and other statistical methods (S. Y. Kim, 2011). Further research has shown that ANNs outperform conventional methods of model estimation, MDA, and logistic regression and are on the same quality level of other modeling techniques, including k-nearest neighbor and decision trees (S. Y. Kim, 2011; Tam & Kiang, 1992). Due to the ANN outperforming the MDA and other statistical methods, the fact that it is of the same quality of other machine learning techniques and it is based on the sigmoid function, this method was deemed appropriate for use to create the second model.

3.5.2.2. Assumptions

The most substantial benefit of utilizing an ANN for the third model is the fact that it does not have any assumptions to consider. As Fish, Barnes, and Aiken (1995, p. 433) wrote in their study, “there are no a priori assumptions made about the relationship being modeled.” Other researchers have echoed this as S. Y. Kim (2011) also noted that there are no assumptions regarding equal dispersions. Kim also went further to argue that this is what made ANNs more useful over traditional statistical studies, including logistic regression. The lack of assumptions shows that ANNs are suitable to investigate the probability of success or failure of a restaurant provided the abundance of independent variables.

3.5.2.3. Model Generation

The neural network is based around the concept of how a brain functions when dealing with a situation. The main difference is that the ANN is a much more simplistic version of the brain, consisting of non-linear computations that are all interconnected to each other (Denton et al., 1990) until the final output. While the ANN was first proposed in 1943, it was further strengthened six years later when the rules for adjusting the weights between the neurons were introduced by Hebb (1949); this pivotal research allowed the networks to learn.

The typical ANN model is organized with an unspecified number of neurons, or nodes, that are distributed into a user-defined number of layers (S. Y. Kim, 2011; Kumar & Ravi, 2007); the typical ANN model utilizes a minimum of three layers. Figure 3 shows a basic model with an input layer consisting of six nodes; the single hidden layer containing seven nodes, and the output layer consisting of one output node (Tam, 1991; Youn & Gu, 2010; Zhang et al., 1999). The input layer refers to the independent variables, in which the quantity is designated by the research being performed, the hidden layer represents the interconnections between the variables, while the output layer is the dependent variable (Goss & Ramchandani, 1995).

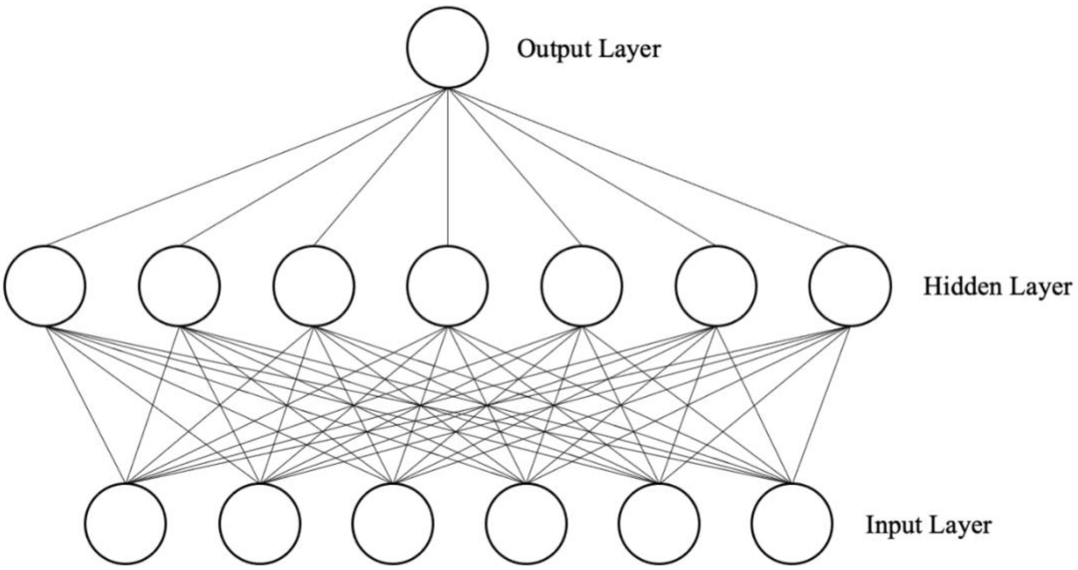
Although the research initially instructs the number of layers, there are some general suggestions for the initial ANN design. The hidden layer consists of two different aspects; the first is the number of hidden layers. While the figure only shows a single hidden layer, it is feasible to have more than one layer, but one layer has been found to be sufficient in classification models. The number of nodes within the hidden layer is difficult to determine while first building the model. Researchers have argued for several rules of thumbs, writing that the number of hidden nodes should be equal to n/2, n, n+1, or 2n+1, where n is the number of input nodes (Zhang et al., 1999). While complex models need to have multiple hidden layers and nodes, Zhang et al. (1999) cautioned that a model containing many hidden layers or nodes could memorize the connections and not learn to predict the outcome. Youn and Gu (2010) and Zhang et al. (1999) wrote that to decide the number of hidden layers and nodes it has to be conducted through trial-and-error experiments.

For the output layer, as the dependent variable is dichotomous, only one node is needed, the output will be either classified as group 1 when the value is greater or equal to 0.5 or group 0 if less than 0.5 (Zhang et al., 1999), this is in-line with the output of the logistic regression model.

In Figure 3, between each node in the input layer, a connection is made to every node in the hidden layer; 42 different connections are in the example model. There are a further seven connections between the hidden layer nodes and the output node, totaling 49 different connections. These are considered to be the communication links, and each link has a weight associated. Weights are similar to the coefficients in a logistic regression (S. Y. Kim, 2011), except there are multiple weights, and one independent variable cannot be associated with a singular weight, or interpreted in the same manner as logistic regression, as there are multiple non-linear communication links (Palmer, Montaño, & Sesé, 2006). These weights are continuously adjusting as the ANN learns the best way to predict the outcome by reducing the error between the current response and the desired response (Denton et al., 1990).

Figure 3

Artificial Neural Network



The most popular form of ANN and the one that is utilized in this study is the feed-forward method, also called the multi-layered perceptron (MLP). The MLP contains no feedback loops so that once the information moves from the input layer to the hidden layer, it does not then return to the input layer; it only moves forward through the communication links (Bishop, 1995; S. Y. Kim, 2011; S.-S. Park & Hancer, 2012; Tam, 1991; Tam & Kiang, 1992; Zhang et al., 1999).

Many other decisions need to be considered when designing an ANN; these include the initialization element optimizer, number of neurons, epochs, batches, and other elements. The first decision is the initialization, which instructs the computer how to create the initial value for the weights. There are fifteen popular methods of how to set the weights, including from an initial value of zero or one, to different randomization methods. As this research utilizes an MLP, the glorot uniform initializer is deemed most appropriate (Glorot & Bengio, 2010).

The second important decision is the optimizer, as this is the central aspect that trains the neural network; it is the aspect that adjusts the weights so that it continues to learn the best algorithm for the problem at hand. ANNs have seven popular optimizers, with the two most popular, adam and stochastic gradient descent (SGD). This study will utilize SGD as it has been found to outperform adam (Wilson, Roelofs, Stern, Srebro, & Recht, 2017). The method of using SGD is explained in the next section.

While other decisions need to be made, including the number of neurons in the hidden layer, as well as the learning rate, epochs, decay, and momentum, these are unable to be determined by statistical methods, and instead of derived from testing different values to find the one combination that provides the highest level of accuracy. This paper will test 2,560 different combinations of parameters to determine the most accurate ANN for the study. Each combination will be run through a 10-fold cross-validation method, explained in Section 3.5.2.5; as such, 2,560 tests will be conducted on the data to determine the most accurate weights and network for future site elements be inputted.

3.5.2.4. Formula

The ANN formula in a feed-forward MLP with three layers, inclusive of one input layer, one hidden layer, and one output layer providing the process moves to the next layer, never going backward is $f\colon X \rightarrow Y$ and the formula to solve for y is:

$$y = f_2\left(w_2 f_1(w_1 x)\right) \tag{12}$$

where:

f_1 = transfer functions for input to hidden layer

f_2 = transfer functions for the hidden layer to output layer

w_1 = matrix of linking weights from input layer to hidden layer

w_2 = matrix of linking weights from hidden layer to output layer

x = (x_1, x_2, …, x_n) = n-vector of independent variables

y = output

As the ANN output needs to be binary, the sigmoid function is used, which is

$$F(y) = \frac{1}{1 + e^{-y}} \tag{13}$$

It is utilized to return results between 0 and 1. Ergo, to solve for Y, the equation is:

$$Y = \frac{1}{1 + e^{-(f_2\left(w_2 f_1(w_1 x)\right))}} \tag{14}$$

The model will use a single hidden layer, however as the number of hidden nodes is an unknown number, the model will have to run multiple times to find the number of hidden nodes that provide the highest accuracy without becoming overfitting and memorizing the characteristics. The weights of the model are calculated separately.

The essential aspect of creating a useable ANN is the derivation of the weights, both between the input and hidden layer and the hidden to the output layer (Tam, 1991). To facilitate the weights, a back-propagation learning algorithm is designed and is useful for training feed-forward networks. The back-propagation learning algorithm concept is that the model keeps adjusting the individual weights through a series of iterations until they converge on the value that provides the most accurate fit of the model (S.-S. Park & Hancer, 2012; Tam & Kiang, 1992).

The idea behind the back-propagation learning algorithm is that it begins by randomly assigning values to each weight (Anandarajan et al., 2001), then inputs the independent variables and compares the dependent variable to the predicted output variable. If the actual and predicted output does not match, it slowly adjusts each connection's weights until it can determine the given dependent variable provided the independent variable (Rumelhart, Hinton, & Williams, 1988). This is followed by adding in the next set of independent variables and see if the model can correctly predict the dependent variable; further modifications are made on the weights until the training set provides the best fit for the weights on each connection.

The best fit is determined by minimizing the error at the output layer. The amount of the error is sent backward through each interconnected node, and the weights of each connection are then modified based on how much that node has contributed to the final output; the weight is then adjusted, and the model is rerun to minimize the error (Anandarajan et al., 2001).

The process of creating the weights is as follows. First, the model is run with randomly assigned weights and uses equation fourteen to calculate the Y value. This is compared to the actual dependent variable and produces the error value of:

$$e_z = Y_z - T_z \tag{15}$$

where:

z = The individual observation

e_z = The error

Y_z = The model output

T_z = Dependent variable

Therefore, the mean squared error (MSE) can be defined as the weights' function, or:

$$E(w) = \frac{1}{2}\sum_{z=1}^{n}(Y_z - T_z)^2 \tag{16}$$

where:

n = number of observations in the training set

The goal is to minimize the MSE by adjusting the individual weights. In order to correctly modify the weights, they are not changed in the same amount of each other, but instead, an optimization algorithm is used to change the weights. The weights will be changed through the gradient descent method (S.-S. Park & Hancer, 2012). The weights are changed by this equation:

$$\Delta w_{iz} = -\eta \frac{\partial E}{\partial w_{iz}} \tag{17}$$

where:

Δw_{iz} = change of the weight

η = learning factor

Equation seventeen can be thought of as the change of the error given a slight change of the observation's specific weight, given the rest of the weights and independent variables remain the same. The learning factor will be tested at values of 0.1, 0.01, 0.001, and 0.005 to ensure the steps are not too huge or too small. As this process continues, the changes are increasingly smaller in size until the ideal weight is found.

The end goal of the model is to input independent variables into the ANN, and have it run the model and produce an output value that is either 0 or 1; however, before the independent variables can be entered, the model needs to be trained and validated.

3.5.2.5. Training and Validation

To produce an accurate model, there will be both training and validation performed. The training will be on the in-sample set and the validation on the out-of-sample set (Zhang et al., 1999). The model will be validated by the out-of-sample set instead of just the in-sample set because the model will always perform better on the in-sample set. The model has the unfortunate capability to memorize the patterns and not learn (Youn & Gu, 2010). Therefore, the out-of-sample set will be able to provide the quality of the model.

The training and validation process will utilize cross-validation, which is a resampling technique that utilizes multiple in-sample and out-of-sample sets. The cross-validation technique's benefit is that every observation is used multiple times for training and once for testing, providing a higher level of validation than a single 80/20 split of the test to training (S. Y. Kim, 2011; Zhang et al., 1999).

The cross-validation method will also utilize a stratified *k*-fold method; in this method, the dataset will be split into *k* folds, or subsets, of equal size and distribution of the dependent variable. Then each fold will be split into *k* portions where the training will be on *k*-1, and the validation will be on the remaining *k* portion (Kohavi, 1995; Zhang et al., 1999). This study will classify *k* as 10, and as such, the *k*-fold method will run ten times, and each fold will use ninety percent of the data for training and ten percent for validation, as shown in Table 6.

Table 6

10-Fold Cross-Validation

	1	2	3	4	5	6	7	8	9	10
k=1	Testing	Training								
k=2	Training	Testing	Training							
k=3	Training		Testing	Training						
k=4	Training			Testing	Training					
k=5	Training				Testing	Training				
k=6	Training					Testing	Training			
k=7	Training						Testing	Training		
k=8	Training							Testing	Training	
k=9	Training								Testing	Training
k=10	Training									Testing

When running the *k*-fold method, each fold will have an error equivalent to the mean squared error (MSE):

$$MSE_k = \frac{1}{n}\sum_{z=1}^{n}\left(Y_z - \hat{f}(x_z)\right)^2 \tag{18}$$

where:

$\hat{f}(x_i)$ = The prediction that f gives for the *z* observation

Y_z = The actual dependent variable

n = The number of observations within the fold

k = The individual fold

The MSE is then calculated for each fold, and then the true error is estimated for the entire model, which is also the average error:

$$CV_{(k)} = \frac{1}{k}\sum_{z=1}^{k} MSE_z \tag{19}$$

The true error ($CV_{(k)}$) provides the overall performance estimate or the final model's accuracy. The final model will be done with all data points utilizing a feed-forward MLP technique trained with back-propagation.

The program used to perform the calculations and formulas described above was written in Python using Keras and the TensorFlow backend. The code for the program can be found in Appendix 14.

3.5.3. Model Selection Criteria

The study will have two models created based on the same dataset, and as such, a procedure for how to determine which of the two models is needed. As these models are non-nested, they are unable to be compared based on the pseudo-R^2. Instead, the study will utilize four different criteria, the overall model accuracy, the area under the Receiver Operating Characteristic (ROC) curve, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC), to compare the two models and select which one provides a better overall fit.

The overall model fit will consider the final coefficients and weights from the perspective model to see how well it can classify the data. The ROC Curve is a graphical diagram that plots the true positive and false-positive rates across a range of thresholds. The true positive rate is the number that the model identifies as positive, and in actuality, is positive. The false-positive rate, or type I error, is when the model identifies the outcome as positive when it is negative.

The AIC and BIC are used to estimate the quality of the non-nested models given the dataset. The AIC is calculated using the following formula to determine the AIC of the two non-nested models; the model with the lowest AIC will determine the better fit (Anderson & Burnham, 2004).

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln(\hat{\mathcal{L}}) \tag{20}$$

where:

k	=	Number of estimable parameters in the model
$\hat{\mathcal{L}}$	=	Maximum likelihood of the model tested

The number of parameters differs between the logistic regression model and the ANN model. In the logistic regression model, the number of parameters is the total number of independent variables plus the coefficient, with each categorical dummy variable considered separately. For the ANN, the number of parameters is the total number of weights included in the model.

The BIC formula is as follows:

$$BIC = k\log(n) - 2\log(\hat{\mathcal{L}}) \tag{21}$$

where:

n	=	Number of observations in the sample
k	=	Number of parameters in the model
$\hat{\mathcal{L}}$	=	Maximum likelihood of the model tested

The maximum likelihood is calculated by the formula:

$$\hat{\mathcal{L}} = \sum_{z=1}^{n} \left(Y_z * \ln\left(P(Y_z)\right) + (1 - Y_z) * \ln\left(1 - P(Y_z)\right)\right) \tag{22}$$

where:

Y_z	=	Actual dependent variable
$P(Y_z)$	=	The predicted outcome from the model
n	=	Number of observations in the sample

As with the AIC, the individual model's lower BIC value will represent the better fitting model.

The study will consider all four criteria, will recommend the model that proves to be most accurate across the different selection tests. If the two models split the tests equally, the model with the highest area under the ROC curve will be deemed appropriate for future use in site determination. The ROC curve test takes into consideration both the number correctly predicted as well as the type I errors that need to be reduced as much as possible.

3.6. Chapter Summary

This chapter has laid out the framework of the methods of data analysis that will be utilized. The study's main target are legally licensed newly opened restaurants in Hong Kong spanning the years 2016, 2017, and 2018, and their success and failure rates. Utilizing nineteen attributes across the area attributes and site characteristics, the study will investigate which pre-opening factors impact restaurant success.

The study will examine each of the independent variables to determine the different rate of restaurant failure within individual TPUs. This analysis will be done by utilizing a logistic regression model; the model will investigate the attributes that were laid out throughout the second chapter. The marginal effects are a crucial benefit of a logistic regression model as it provides the ability to understand how each variable contributes to the output (Neophytou & Molinero, 2004). The model will also begin the analysis of how successful or failed restaurants differ.

Two separate models will be utilized to determine if a predictive model can be created to understand the impact of the independent variables and the success or failure of a restaurant site. The first model will be a logistic regression model that will examine the different independent variables and their effect on prediction. The second model that will be analyzed is the ANN, as it has been shown to outperform logistic regression models (Anandarajan et al., 2001; Fletcher & Goss, 1993; S. Y. Kim & Upneja, 2014; Kumar & Ravi, 2007; S.-S. Park & Hancer, 2012; Youn & Gu, 2010; Zhang et al., 1999). The model will utilize the feed-forward MLP, trained with back-propagation, and undergone 10-fold cross-validation to observe the accuracy. This model should produce higher rates of accuracy due to the number of separate weights and the ability to provide quality results in making predictions (Youn & Gu, 2010).

Arguments can be made for the use of either of these models; De Andrés, Landajo, and Lorca (2005) wrote that the ANN should only be used if an unsatisfactory result is found when utilizing a logistic model, while S.-S. Park and Hancer (2012) found that both the logistic and ANN models had the same overall accuracy on the holdout sample. To further test which model would be the best to suggest for future examinations of restaurant sites, the model accuracy, ROC curve, AIC, and BIC will be used to measure the fit of each, and the model that shows most accurate across the criteria will be deemed to be the ideal model.

Chapter 4. Findings

4.1. Chapter Introduction

This chapter introduces the statistics of the 6,710 restaurants included in the study. The statistics will consist of the newly opened and failed restaurants in each quarter, and the failure rates of those restaurants, both quarterly and cumulatively for the years 2016-2019. Additionally, the success and failure statistics are counted by the number of days post-opening, through the first 1,460 days (four years).

The location of the newly opened restaurants both within the four regions and 214 TPUs will be examined, with an emphasis on the failure rates and TPU locations in the periods of 0 to 365 days, 0 to 730 days, and 0 to 1095 days. The different characteristics of the restaurants, in both the area attributes and site characteristics, will be outlined to showcase both the restaurants in the study and the aspects of the TPUs where they operate.

The first model explained is the logistic regression, starting with expanding on the five assumptions presented in Chapter 3, followed by the results and a discussion of the marginal effects. The ANN model will be presented with first the explanation of the model generation, followed by the findings. The six hypotheses and two research questions will both be assessed before showcasing which of the two models is most suited for restaurant site selection predictions.

4.2. Success and Failure Statistics

4.2.1. Time Period Success and Failure

The time period of the success and failure statistics across the 6,710 restaurants are separated into two sections. First is the quarterly breakdown of when the restaurants either opened or closed throughout the three-year period. Table 7 shows the quarterly and cumulative analysis, including the total number of restaurants in operation and the number of restaurants that either opened or closed, and the failure rate for both the specific quarter and the cumulative period.

Across the three-year study period, an average of 2,262 restaurants opened each year, with 559 opening each quarter (6.13 per day), with the greatest number of openings occurring in Quarter 3, 2017 (638 openings), and the least in Quarter 1, 2017 (511 openings). Including the failures that occurred in 2019, an average of 125 restaurants failed each quarter, resulting in 1.83 closings every day.

Utilizing the opening date as the starting point of a restaurant's existence, the lifespan of the restaurant can be analyzed. Table 8 showcases the longevity of restaurant success through the first 1,460 days after opening, and Chart 1 provides a second perspective on these results.

Table 7

New Restaurant Failure Statistics 2016-2019: Quarterly

	per Quarter			Cumulative			
	New	Failed	Failure Rate	New	Successful	Failed	Failure Rate
2016							
Quarter 1	610	3	0.49%	610	607	3	0.49% [a]
Quarter 2	531	14	1.23%	1141	1124	17	1.49% [a]
Quarter 3	519	41	2.47%	1660	1602	58	3.49% [a]
Quarter 4	526	50	2.29%	2186	2078	108	4.94% [a]
2017							
Quarter 1	511	72	2.67%	2697	2517	180	6.67% [b]
Quarter 2	540	78	2.41%	3237	2979	258	7.97% [b]
Quarter 3	638	98	2.53%	3875	3519	356	9.19% [b]
Quarter 4	616	102	2.27%	4491	4033	458	10.2% [b]
2018							
Quarter 1	522	107	2.13%	5013	4448	565	11.27% [c]
Quarter 2	526	152	2.74%	5539	4822	717	12.94% [c]
Quarter 3	602	164	2.67%	6141	5260	881	14.35% [c]
Quarter 4	569	214	3.19%	6710	5615	1095	16.32% [c]
2019							
Quarter 1		209	3.11%		5406	1304	19.43% [c]
Quarter 2		234	3.49%		5172	1538	22.92% [c]
Quarter 3		199	2.97%		4973	1737	25.89% [c]
Quarter 4		264	3.93%		4709	2001	29.82% [c]

Note:

[a] Includes restaurant that opened in 2016

[b] Includes restaurant that opened in 2016 and 2017

[c] Includes restaurant that opened in 2016, 2017, and 2018

Table 8

New Restaurant Failure Statistics 2016-2019: Quarterly and Cumulative Failure Rate

Days after Opening	Total	Number Successful	Number Failed	Failure Rate	Cumulative Failures	Cumulative Failure Rate
1 - 91 Days [a]	6,710	6,666	44	0.66%	44	0.66%
92 - 182 Days [a]	6,666	6,324	342	5.13%	386	5.75%
183 - 274 Days [a]	6,324	6,196	128	2.02%	514	7.66%
275 - 365 Days [a]	**6,196**	**5,988**	**208**	**3.36%**	**722**	**10.76%**
366 - 455 Days [a]	5,988	5,795	193	3.22%	915	13.64%
456 - 546 Days [a]	5,795	5,621	174	3.00%	1,089	16.23%
547 - 638 Days [a]	5,621	5,424	197	3.50%	1,286	19.17%
639 - 730 Days [a]	**5,424**	**5,261**	**163**	**3.01%**	**1,449**	**21.59%**
731 - 820 Days [b]	5,261	5,160	101	1.92%	1,550	23.10%
821 - 911 Days [b]	5,160	5,057	103	2.00%	1,653	24.63%
912 - 1,003 Days [b]	5,057	4,947	110	2.18%	1,763	26.27%
1,004 - 1,095 Days [b]	**4,947**	**4,839**	**108**	**2.18%**	**1,871**	**27.88%**
1,096 - 1,187 Days [c]	4,839	4,786	53	1.10%	1,924	28.67%
1,186 - 1,276 Days [c]	4,786	4,751	35	0.73%	1,959	29.20%
1,277 - 1,368 Days [c]	4,751	4,731	20	0.42%	1,979	29.49%
1,369 - 1,460 Days [c]	**4,731**	**4,709**	**22**	**0.47%**	**2,001**	**29.82%**

Note:

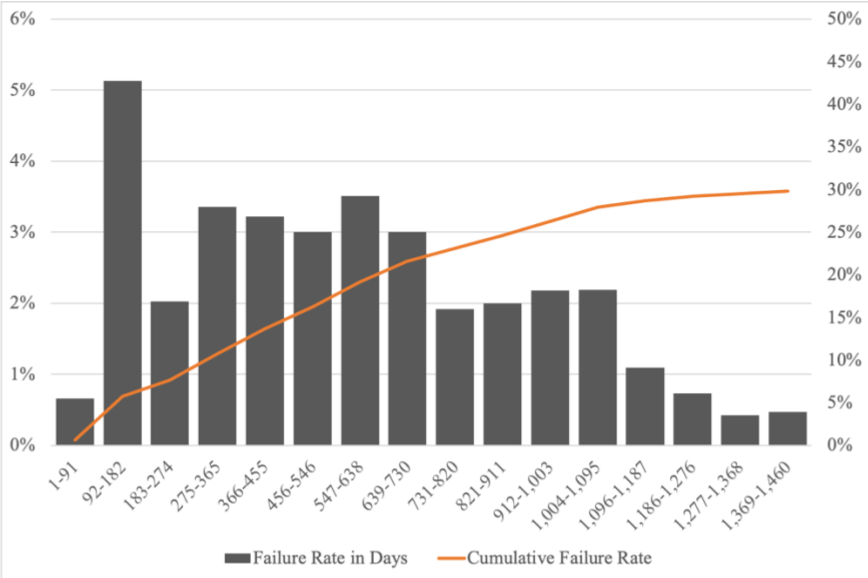
[a] Includes restaurants that opened in 2016, 2017, and 2018

[b] Includes restaurants that opened in 2016 and 2017

[c] Includes restaurants that opened in 2016

Chart 1

New Restaurant Failure Statistics 2016-2019: Quarterly and Cumulative Failure Rate



Six thousand seven hundred ten (6,710) restaurants opened in the years 2016, 2017, and 2018; 4,709 (70.18 percent) of which were successful as measured by remaining in operation as of 31 December 2019; conversely, 2,001 (19.82 percent) restaurants failed (Table 8). Regarding the success and failure, the highest period of failure fell between 92-182 days' post opening with 342 failures, accounting for 17.09 percent of all restaurant failures. This concentration may be due to the Food and Beverage Licensing Board not granting a full restaurant license after the original provisional restaurant license, or secondarily, it may indicate an underestimate of required startup capital.

The restaurants within the study have a failure rate in the first 182 days after opening of 5.75 percent, 386 closures out of 6,710 restaurants. Although the number of restaurants that failed in the second 182-day period decreased over the first 182 days, the failure rate increased to 10.76 percent, with a total of 722 restaurants failing. The period of 366 to 730 days had 727 restaurants close, five more than in the first period of 0 to 365 days, combining to total 1,449 failures (21.59 percent). The third 365-day period, 731 to 1,095 days, had 422 failures; however, this only includes those restaurants that opened in the years 2016 and 2017. There is a total of 1,871 restaurants that failed by the 1,095th day or 27.88 percent of the newly opened restaurants. By the end of the study period, three out of ten restaurants that opened in 2016, 2017, and 2018, had failed.

4.2.2. Locational Success and Failure

4.2.2.1. Regional

Hong Kong is separated into four main regions, Hong Kong Island, Kowloon, New Territories, and the outlying Islands ("What You'll See in Hong Kong's Areas," n.d.). Due to the nature of Hong Kong, less than 25 percent of the land is incredibly dense, high-rise filled city districts, while over 75 percent of areas are more rural areas with village houses (Civil Engineering and Development Department, 2019). The four regions encompass all of the 214 TPUs that are utilized throughout the analysis; the regions and TPU boundary lines are found in Map 3.

Out of the four regions of Hong Kong, Kowloon is the most popular region for opening restaurants with the number of new restaurants accounting for 2,577 of the 6,710 restaurants, 38.41 percent of newly opened restaurants included in the study; however, they also have the highest failure rate at 31.86 percent (Table 9), over Hong Kong Island (30.72 percent), New Territories (27.37 percent) and the outlying Islands (11.11 percent). The four regions are broken down into the 214 TPUs, while the borders of the TPUs generally align with the regional boundaries, there are slight incursions based on the natural geography into a neighboring region.

Map 3

Regions of Hong Kong

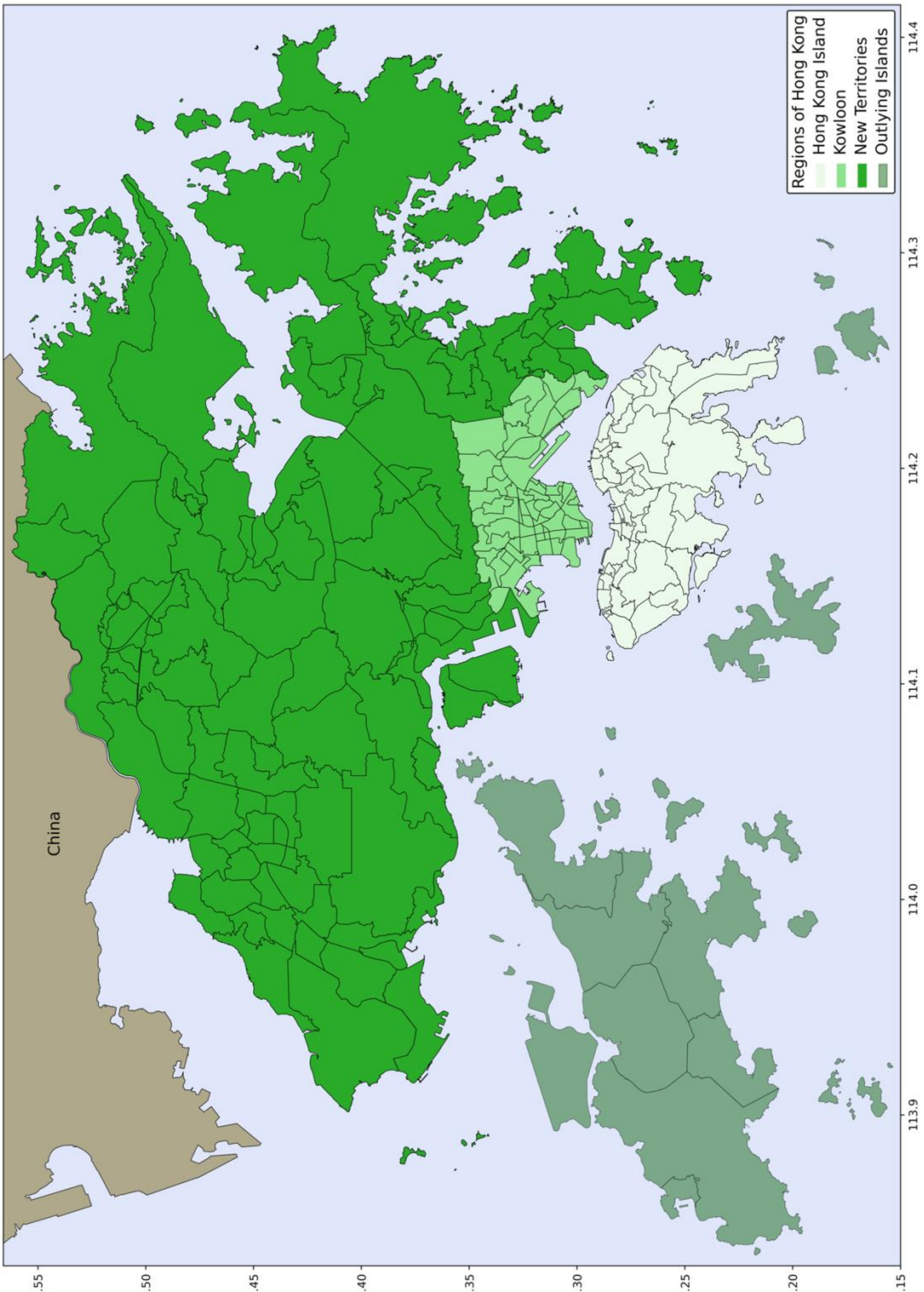


Table 9

Success or Failure of New Restaurants by Region

Location	Number of Restaurants [a]	Success	Success %	Failed [b]	Failed %
Hong Kong Island	2,025	1,403	69.28%	622	30.72%
Kowloon	2,577	1,756	68.14%	821	31.86%
New Territories	1,991	1,446	72.63%	545	27.37%
Outlying Islands	117	104	88.89%	13	11.11%
Total	6,710	4,709	70.18%	2001	29.82%

Note:

[a] Includes restaurants that opened in 2016, 2017, and 2018

[b] Includes restaurants that failed between 2016 and 2019

4.2.2.2. TPU Restaurant Openings

Considering the TPU locations of the newly opened restaurants in the three-year study period, there are 50 TPUs that did not have any new restaurant openings (Map 4 and Table 10). The most common range of new restaurants in a TPU is 1 to 49 openings, occurring in 117 of the 214 TPUs (54.673 percent). In order to investigate the most popular TPUs to open a restaurant, Map 5 showcases the ten TPUs that saw the greatest number of new restaurants during the study period, with five of those TPUs located on Hong Kong Island, three in Kowloon, and two in the New Territories (Table 11).

Map 4

Restaurant Opening Density Heat Map

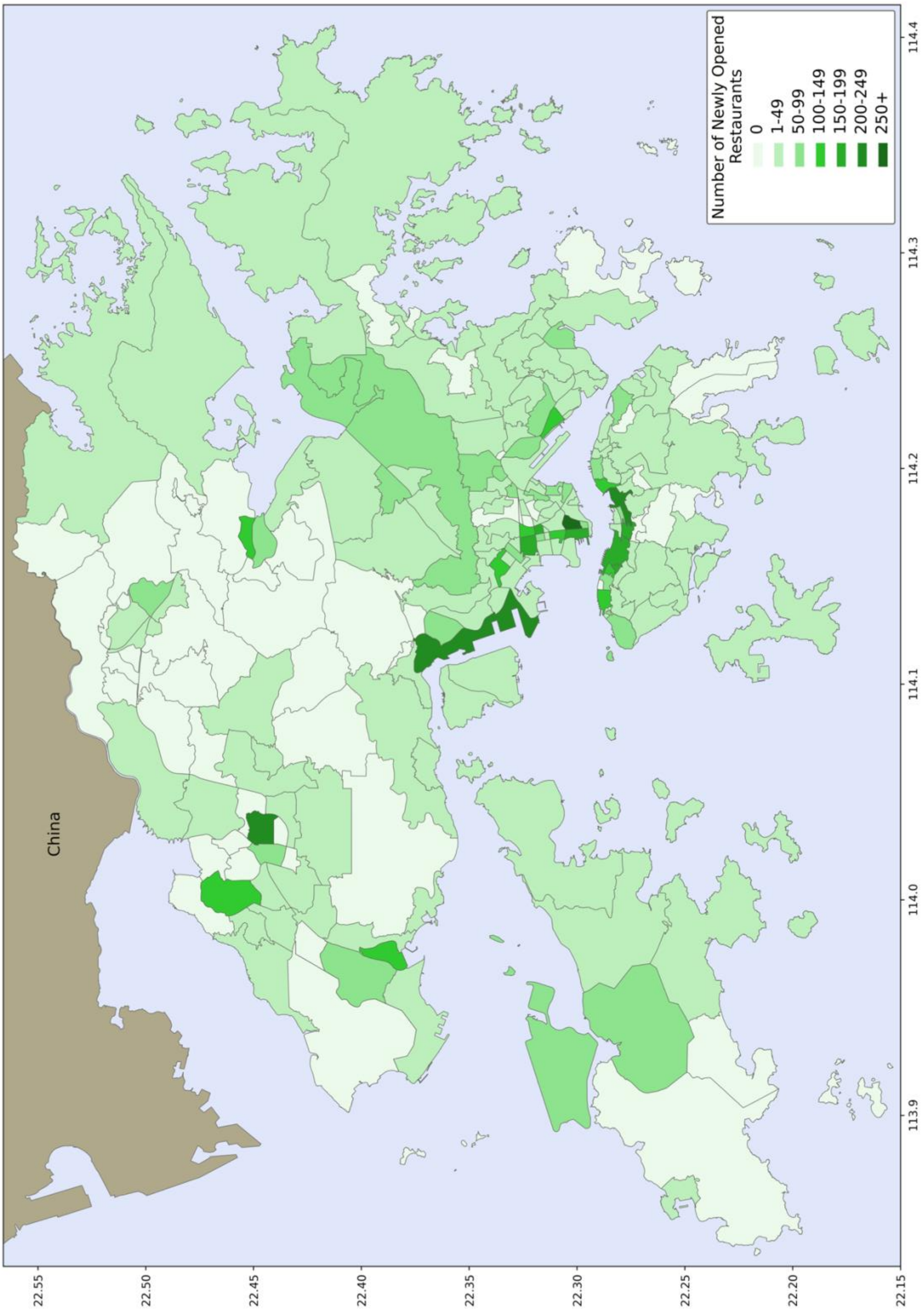


Table 10

Restaurant Opening Count and TPU

Number of Restaurants [a]	Number of TPUs	% of Total
0	50	23.364%
1 – 49	117	54.673%
50 – 99	27	12.617%
100 – 149	9	4.206%
150 – 199	6	2.804%
200 – 249	4	1.869%
250 +	1	0.467%

Note:

[a] Includes restaurants that opened in 2016, 2017, and 2018

Map 5

Top TPUs by Number of New Restaurants

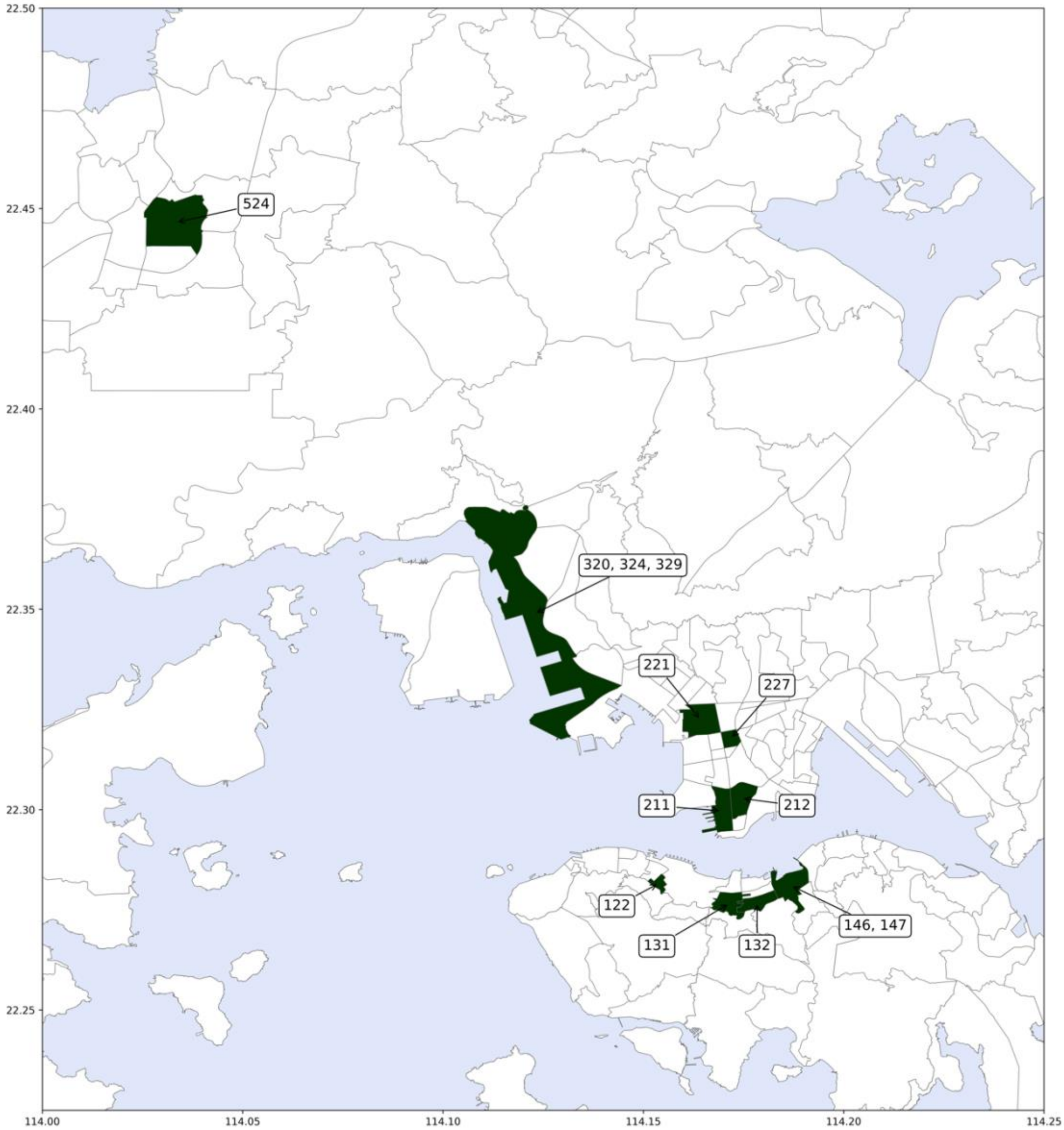


Table 11

Top TPUs by Number of New Restaurants

Location	Region	Number of Restaurants [a]
212	Kowloon	254
320, 324, 329	New Territories	233
524	New Territories	223
146, 147	HK Island	211
132	HK Island	200
221	Kowloon	197
131	HK Island	179
211	Kowloon	161
227	HK Island	161
122	HK Island	151

Note:

[a] Includes restaurants opened in 2016, 2017, and 2018

Among the TPUs with the greatest number of openings, TPU 212, located in Kowloon, had 254 restaurants open within the three-year time span. The greatest number of new openings across the TPUs are predominately on Hong Kong Island (5) and Kowloon (3).

Due to the different land areas of the TPUs, the number of restaurants is standardized by calculating the number of restaurants per square kilometer of the TPU. The most and least number of restaurant openings are found in Map 6 and Table 12 and Table 13.

Map 6

Number of Restaurant Openings per Square Kilometer

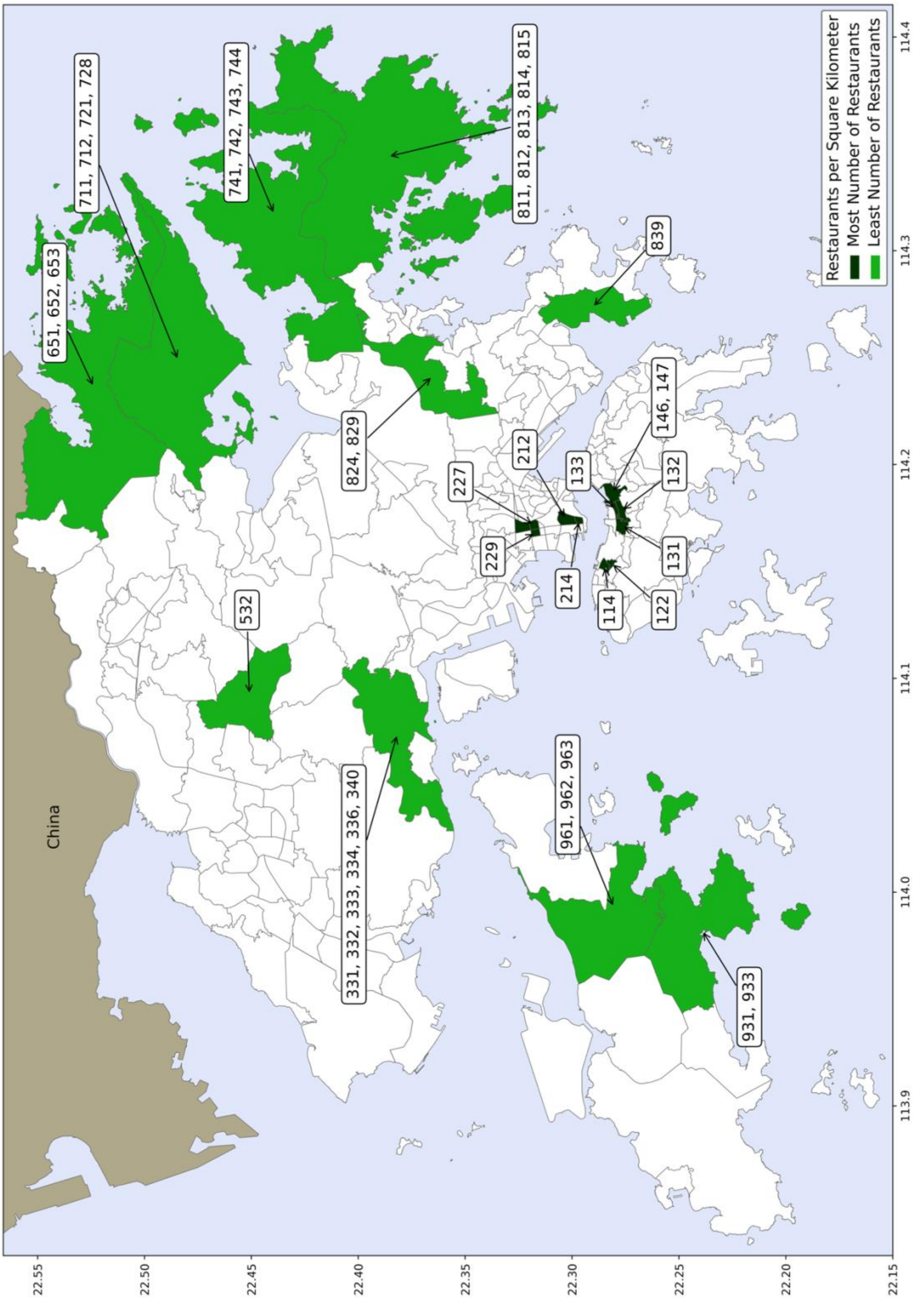


Table 12

Top TPUs by Number of Restaurant Openings

Location	Region	TPU Size Sq. Km	Total Number of Restaurants [a]	Number of Restaurants per Sq. Km
122	HK Island	0.1529	151	987.739
227	Kowloon	0.2402	161	670.332
114	HK Island	0.2699	142	526.045
214	Kowloon	0.1527	78	510.962
132	HK Island	0.4077	200	490.521
212	Kowloon	0.5938	254	427.732
229	Kowloon	0.1914	79	412.855
133	HK Island	0.1929	74	383.581
131	HK Island	0.5135	179	348.616
146, 147	HK Island	0.8129	211	259.555
222	Kowloon	0.4815	116	240.938

Note:

[a] Includes restaurants that opened in 2016, 2017, and 2018

Table 13

Bottom TPUs by Number of Restaurant Openings

Location	Region	TPU Size Sq. Km	Total Number of Restaurants [ab]	Number of Restaurants per Sq. Km
811, 812, 813, 814, 815	New Territories	88.889	1	0.011
651, 652, 653	New Territories	71.538	1	0.014
741, 742, 743, 744	New Territories	64.747	2	0.031
711, 712, 721, 728	New Territories	54.644	2	0.037
331, 332, 333, 334, 336, 340	New Territories	22.604	1	0.044
824, 829	New Territories	14.267	1	0.070
532	New Territories	13.866	1	0.072
931, 933	Outlying Islands	29.524	3	0.102
839	New Territories	8.713	1	0.115
961, 962, 963	Outlying Islands	30.073	4	0.133

Note:

[a] Includes Restaurants opened in 2016, 2017, and 2018

[b] Minimum 1 Restaurant

The greatest number of restaurant openings per square kilometer are found in Kowloon and Hong Kong Island; these locations also have some of the greatest numbers of residents per square kilometer. The TPUs that have the least number of restaurants per square kilometer are located in some of the least populous areas of the New Territories and the outlying Islands, showing that restaurateurs have a preference to open close to potential customers.

4.2.2.3. TPU Failure Rates

Restaurateurs have shown a preference for opening in areas that have a high number of residents. However, the failure rate of restaurants in the TPUs has to be examined to understand if that is a wise opening decision. The failure rates across 365 day periods within the TPUs are illustrated in Map 7, Map 8, and Map 9 and corresponding Table 14, Table 15, and Table 16. The maps look at the individual TPUs failure rate and the number of days until restaurant failure, and not calendar years. The period of 1,096 to 1,460 days is not included as only the restaurants that opened in 2016 are included in that time span.

Map 7

Failure Rate in First 365 Days After Opening
(Minimum 25 Restaurants)

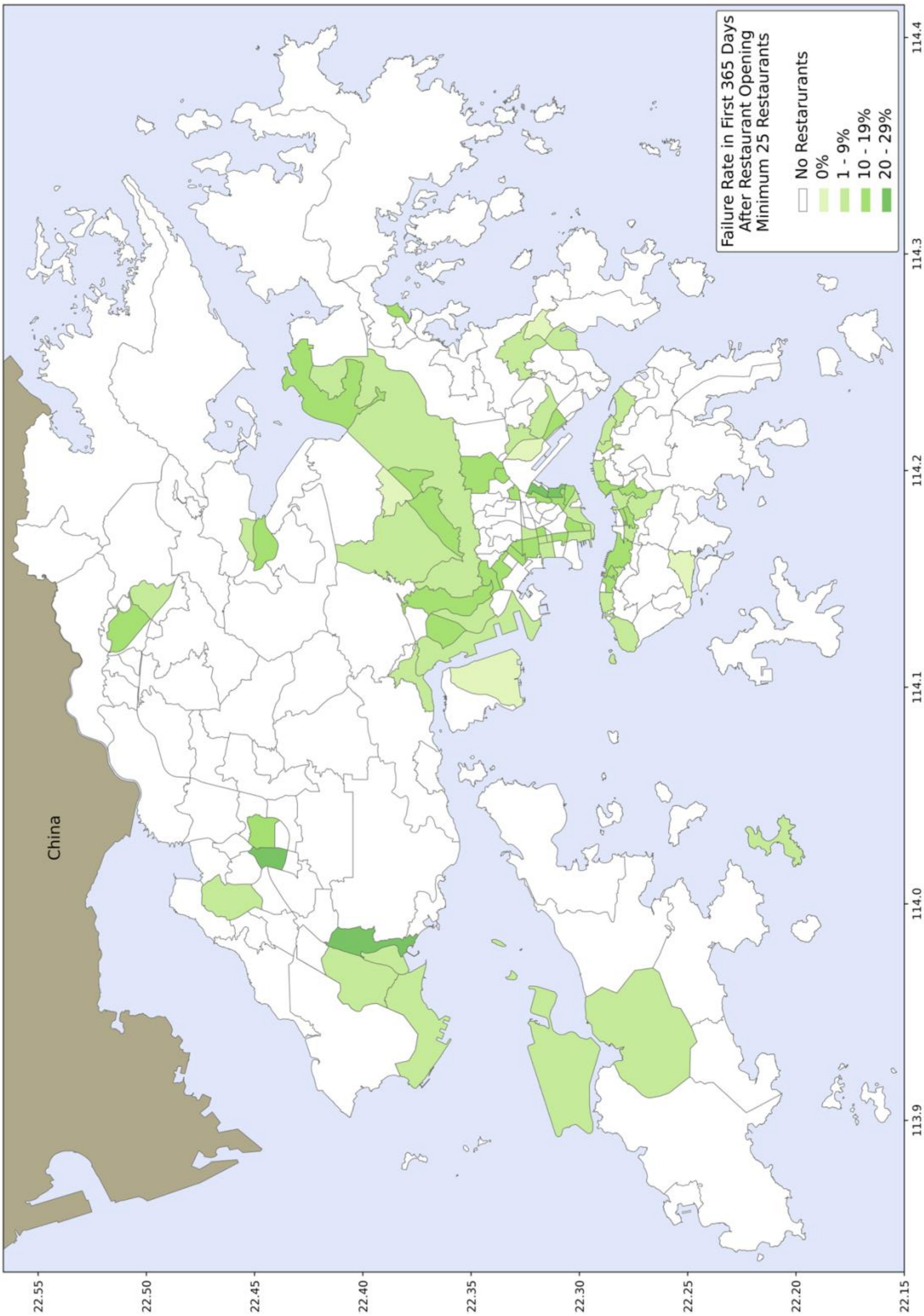


Table 14

Failure Rate in First 365 Days After Opening

TPU	Total Restaurants [ab]	Failures in 365 Days	Failure Percentage
425	36	9	25.00%
527	86	19	22.09%
242	35	7	20.00%
244	50	10	20.00%
221	197	39	19.80%
243	36	7	19.44%
285	98	19	19.39%
225	99	19	19.19%
261	42	8	19.05%
723	60	11	18.33%

Note:

[a] Includes restaurants opened in 2016, 2017, and 2018

[b] Minimum 25 Restaurants

Map 8

Failure Rate in First 720 Days After Opening
(Minimum 25 Restaurants)

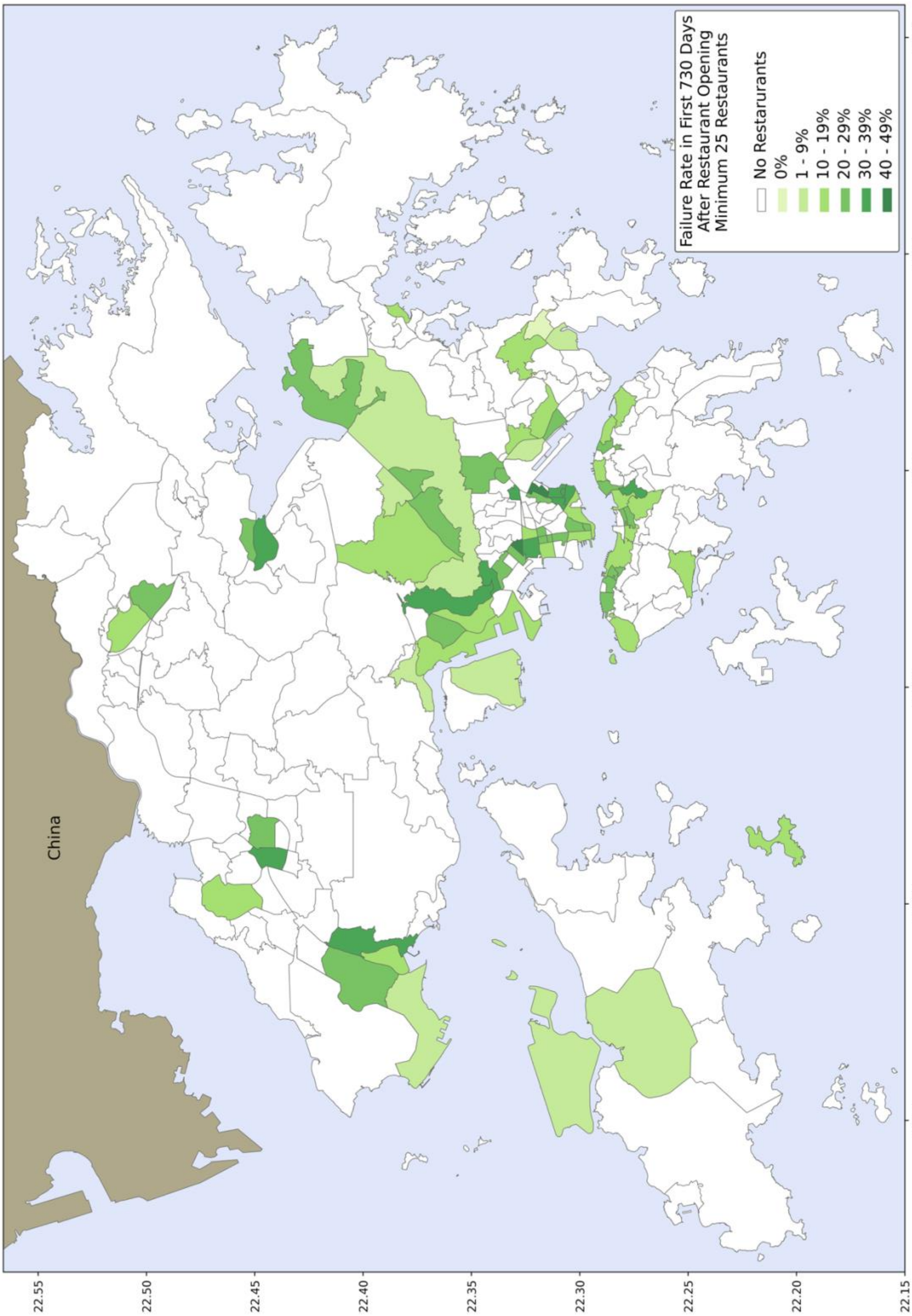


Table 15

Failure Rate in First 730 Days After Opening

TPU	Total Restaurants [ab]	Failures in 730 Days	Failure Percentage
242	35	15	42.86%
267	39	16	41.03%
425	36	13	36.11%
244	50	18	36.00%
221	197	68	34.52%
527	86	29	33.72%
285	98	33	33.67%
327	34	11	32.35%
145	25	8	32.00%
723	60	19	31.67%

Note:

[a] Includes restaurants opened in 2016, 2017, and 2018

[b] Minimum 25 Restaurants

Map 9

Failure Rate in First 1,095 Days After Opening
(Minimum 25 Restaurants)

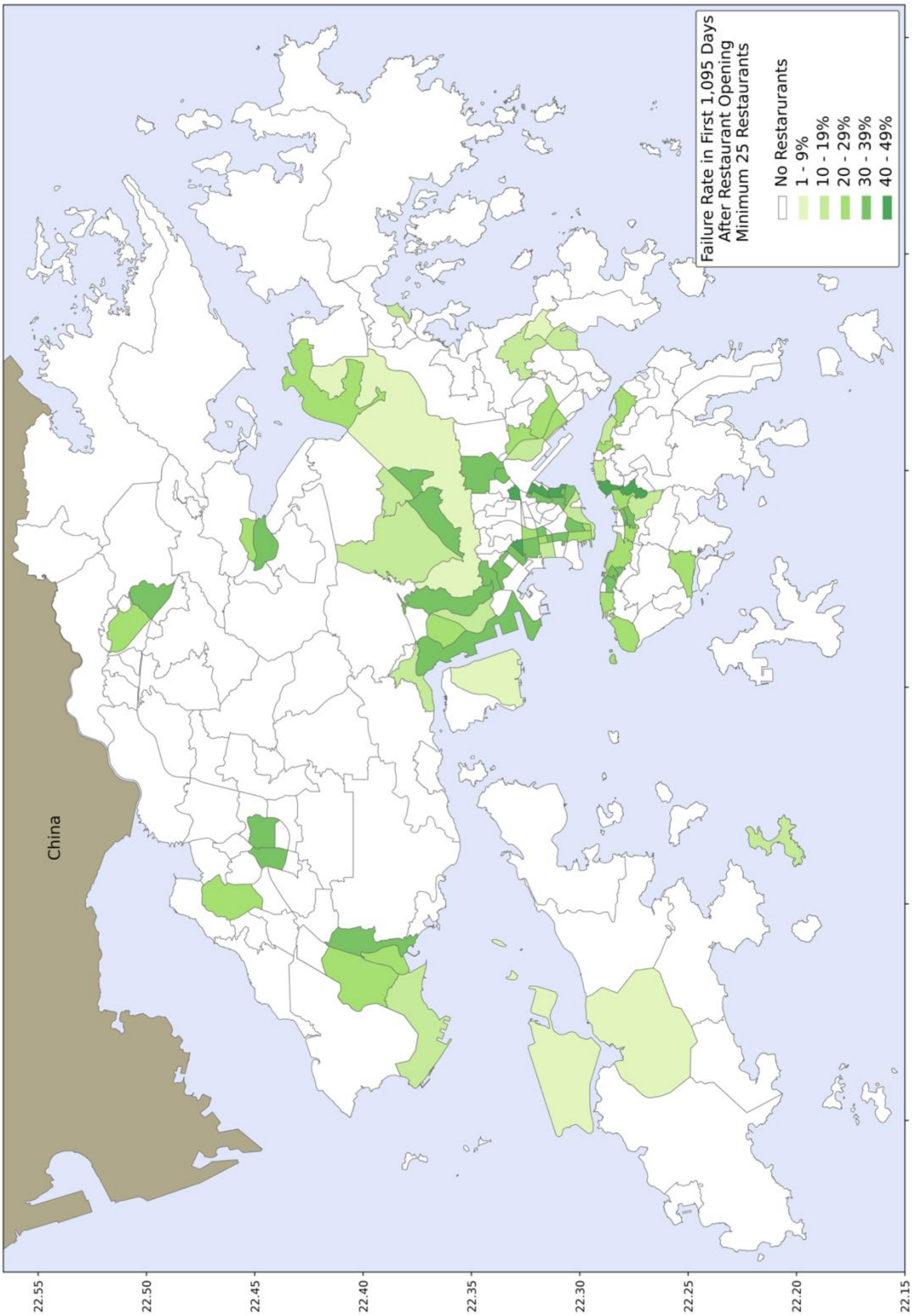


Table 16

Failure Rate in First 1,095 Days After Opening

TPU	Total Restaurants [ab]	Failures in 1,095 Days	Failure Percentage
267	39	19	48.72%
242	35	17	48.57%
145	25	12	48.00%
151	116	50	43.10%
285	98	40	40.82%
244	50	20	40.00%
527	86	34	39.53%
221	197	77	39.09%
243	36	14	38.89%
261	42	16	38.10%

Note:

[a] Includes restaurants opened in 2016, 2017, and 2018

[b] Minimum 25 Restaurants

Kowloon has the overall highest number of failed TPUs across the three-year span. In the first 365 days, out of the top ten TPUs with both the highest failure rate and a minimum of 25 restaurants, seven are located within Kowloon. TPUs 221, 225, 242, 243, 244, 261, and 285 all have a higher than 19 percent failure rate. The period of 366 to 730 days had five of the ten top TPUs located in Kowloon with a greater than 30 percent failure rate. The final period of 731 to 1,095 days had seven of the top ten TPUs in Kowloon and a failure rate greater than 38 percent. Across the three-year period, four TPUs in Kowloon were in all the top ten lists, TPUs 221, 242, 244, and 285. Apart from the TPUs in Kowloon, there is a single other TPU, located in the New Territories, that is on all three lists of top failure rates, TPU 527.

4.3. Restaurant Area Attributes and Site Characteristics Descriptives

4.3.1. Area Attributes

The 6,710 restaurants in the study have the descriptive attributes displayed in Table 17. The differences between dense and rural areas of Hong Kong can be appreciated by considering the number of people living within a five-minute walk of a restaurant. To understand the difference this makes to a restaurant, consider that a restaurant located in a remote area may have as few as 20.896 people living within a five-minute walk, while a restaurant located in one of the denser TPUs has up to 64,087.077 people living within a five-minute walk. Overall, the average restaurant located in Hong Kong has 20,802.211 people living within a five-minute walk.

Considering the individual TPUs, the areas with the most and least number of residents is found in Map 10 and Table 18 and Table 19. The TPUs with the greatest number of residents are mostly found in the New Territories, with eight of the ten being found in the region; this is due to the large size area of the TPUs in the New Territories. The TPUs with the least number of residents are found in either remote sections, TPU 941, 942, and 943, TPU 623, TPU 629, TPU 821, or in the central business districts, TPU 134 and TPU 135.

Table 17

Dependent and Independent Descriptive Information

Variables (n = 6,710)	Mean / %	Minimum	Maximum	Standard Deviation
% Currently in Operation	70.18%			
Census				
Number of Residents	20,802.21	20.90	64,087.08	12,471.65
Age of Residents	43.61	35.64	55.90	1.98
Monthly Household Income (HKD)	29,305.10	9,924.79	155,920.00	10,876.46
Residents within Household	2.68	1.60	4.10	0.25
Post-Secondary Degrees	31.19%	8.50%	56.30%	9.33%
Labor Force Participation	62.66%	46.70%	72.10%	3.42%
Chinese Ethnicity	86.93%	44.90%	99.80%	9.81%
Owner/Occupied Flats	50.14%	0.10%	82.90%	11.87%
Males	45.92%	39.60%	58.20%	1.64%
Number of Competitors				
Cuisine	63.92	0	539	74.17
Price	67.94	0	321	60.85
Convenience	8.84	0	33	5.26
Aggregate Review Score	3.29	0	5	1.27
Location				
On the Ground Floor	66.65%			
Inside 5 Minutes to an MTR Exit	66.83%			
Inside a Hotel	8.79%			
Close to an Attractions Area	16.68%			
Inside a Mall	13.98%			

Note:

All variables represent the statistics within the surrounding area within a 369.95-meter radius of the individual restaurant site (Proximity)

Map 10

TPUs by Number of Residents

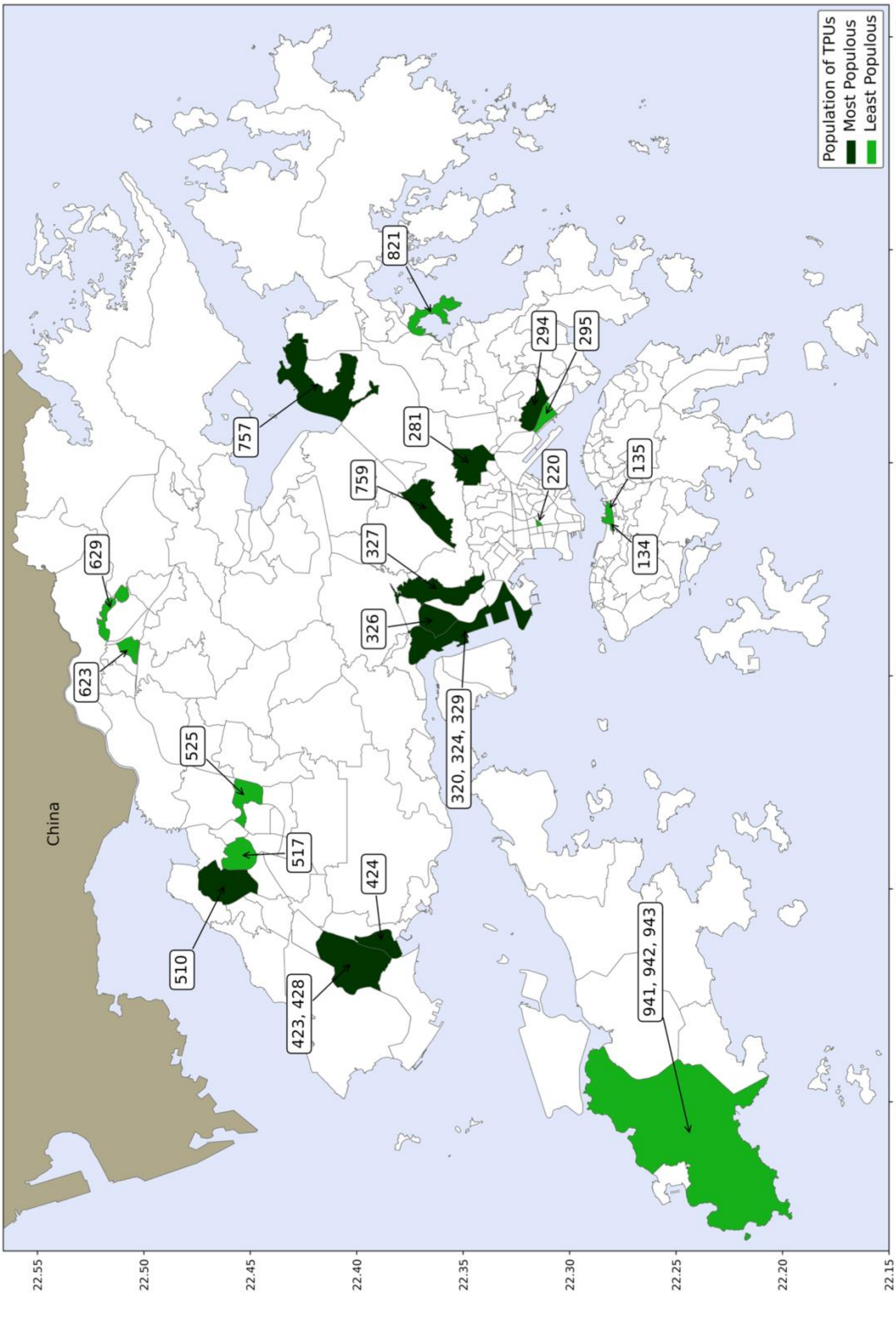


Table 18

Most Populated TPUs

Location	Region	Population [a]
510	New Territories	286,232
757	New Territories	209,714
281	Kowloon	206,097
423, 428	New Territories	170,753
326	New Territories	167,455
294	Kowloon	161,884
759	New Territories	127,917
320, 324, 329	New Territories	116,263
327	New Territories	113,647
424	New Territories	108,898

Note:

[a] Population as of 2016 Census

Table 19

Least Populated TPUs

Location	Region	Population [a]
623	New Territories	1,298
525	New Territories	1,259
134	HK Island	1,200
941, 942, 943	Outlying Islands	1,184
220	Kowloon	1,139
517	New Territories	1,115
135	HK Island	1,076
821	New Territories	1,068
295	Kowloon	1,060
629	New Territories	1,032

Note:

[a] Population as of 2016 Census

Due to the different land areas of the TPUs, the number of residents is standardized by calculating the number of residents per square kilometer of the TPU. When considering the sizes of the TPUs and the number of residents, there is a shift from the New Territories having the most populous TPUs to Kowloon. Map 11 and corresponding Table 20 and Table 21 showcases the number of residents when controlled for the size of the TPU per square kilometer. Nine of the ten TPUs with the greatest number of residents when controlled for the area are in Kowloon, while the least populous areas are the distant parts of the New Territories or outlying Islands.

Map 11

TPUs by Number of Residents per Square Kilometer

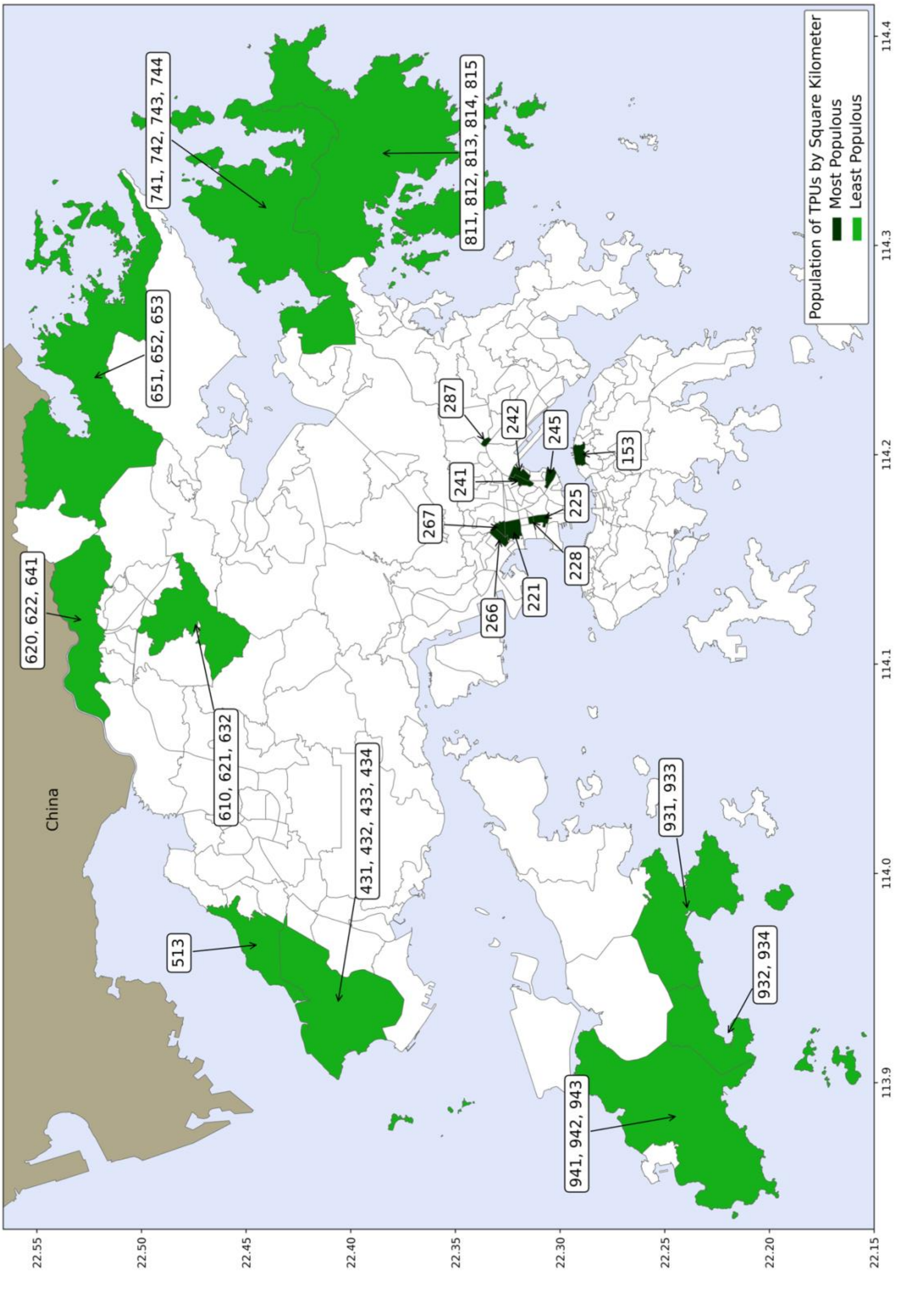


Table 20

Most Populated TPUs by Square Kilometer

Location	Region	TPU Size Sq. Km	Population [a]	Population per Sq. Km
228	Kowloon	0.116	15,738	135,173.695
267	Kowloon	0.336	42,488	126,368.269
266	Kowloon	0.519	60,161	115,872.151
225	Kowloon	0.429	48,272	112,488.622
287	Kowloon	0.165	18,551	112,338.876
245	Kowloon	0.436	48,069	110,326.752
242	Kowloon	0.488	53,800	110,226.475
241	Kowloon	0.442	46,095	104,246.661
153	HK Island	0.746	73,686	98,816.056
221	Kowloon	0.927	87,291	94,148.267

Note:

[a] Population as of 2016 Census

Table 21

Least Populated TPUs by Square Kilometer

Location	Region	TPU Size Sq. Km	Population [a]	Population per Sq. Km
620, 622, 641	New Territories	20.904	3,992	190.972
513	New Territories	9.976	1,844	184.849
610, 621, 632	New Territories	17.655	3,131	177.347
932, 934	Outlying Islands	16.625	2,526	151.941
651, 652, 653	New Territories	71.538	9,383	131.162
931, 933	Outlying Islands	29.524	3,144	106.489
741, 742, 743, 744	New Territories	64.747	6,576	101.564
431, 432, 433, 434	New Territories	31.739	1,892	59.611
811, 812, 813, 814, 815	New Territories	88.889	3,474	39.083
941, 942, 943	Outlying Islands	57.311	1,184	20.659

Note:

[a] Population as of 2016 Census

Hong Kong restaurants have a surrounding population with the average median age of 43.611. When considering the median ages of the TPUs in Hong Kong, the oldest median age TPUs are spread across multiple regions, while the youngest median age TPUs are concentrated in Kowloon and two portions of the New Territories, Map 12 and Table 22 and Table 23.

Map 12

TPUs by Median Age

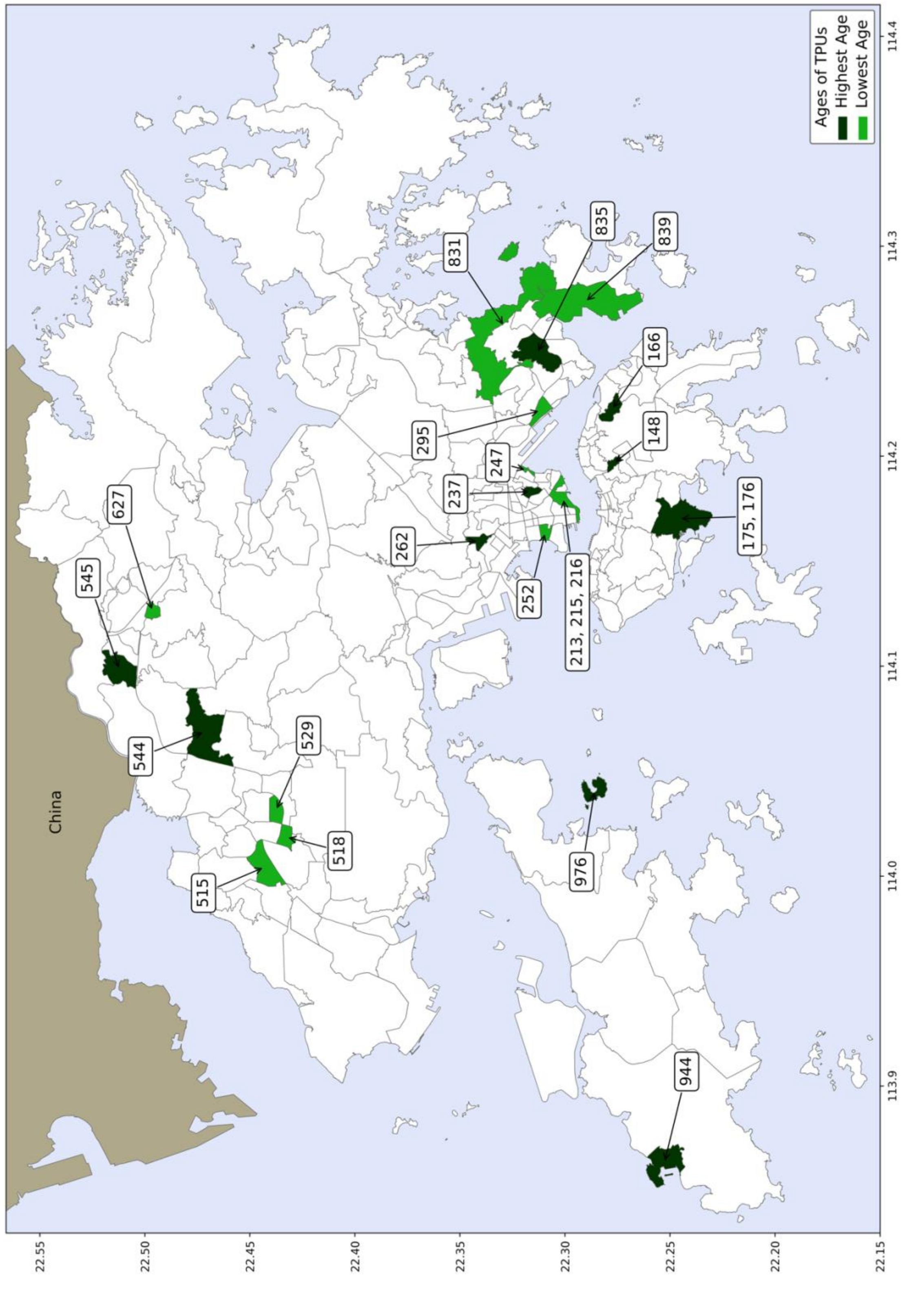


Table 22

TPUs with Oldest Median Age

Location	Region	Median Age [a]
944	Outlying Islands	55.9
545	New Territories	52.0
237	Kowloon	51.0
148	HK Island	50.1
262	Kowloon	50.1
835	New Territories	50.1
976	Outlying Islands	50.1
175, 176	HK Island	49.6
166	HK Island	49.4
544	New Territories	49.3

Note:

[a] Median Age as of 2016 Census

Table 23

TPUs with Youngest Median Age

Location	Region	Median Age [a]
831	New Territories	39.2
515	New Territories	38.5
627	New Territories	38.4
839	New Territories	38.3
518	New Territories	38.1
529	New Territories	38.1
247	Kowloon	38.0
252	Kowloon	36.1
295	Kowloon	34.9
213, 215, 216	Kowloon	34.7

Note:

[a] Median Age as of 2016 Census

Hong Kong restaurants have a surrounding population monthly household income of HK$29,305.10. However, the income levels of the TPUs range from a low of HK$9,520 in TPU 295 in Kowloon, to a high of HK$155,920 in TPU 181 and 182 (Map 13, and Table 24 and Table 25). These TPUs are located on HK Island in an area known as The Peak, which is known to have some of the most expensive real estate due to the view of Victoria Harbour.

Map 13

TPUs by Median Monthly Domestic Household Income

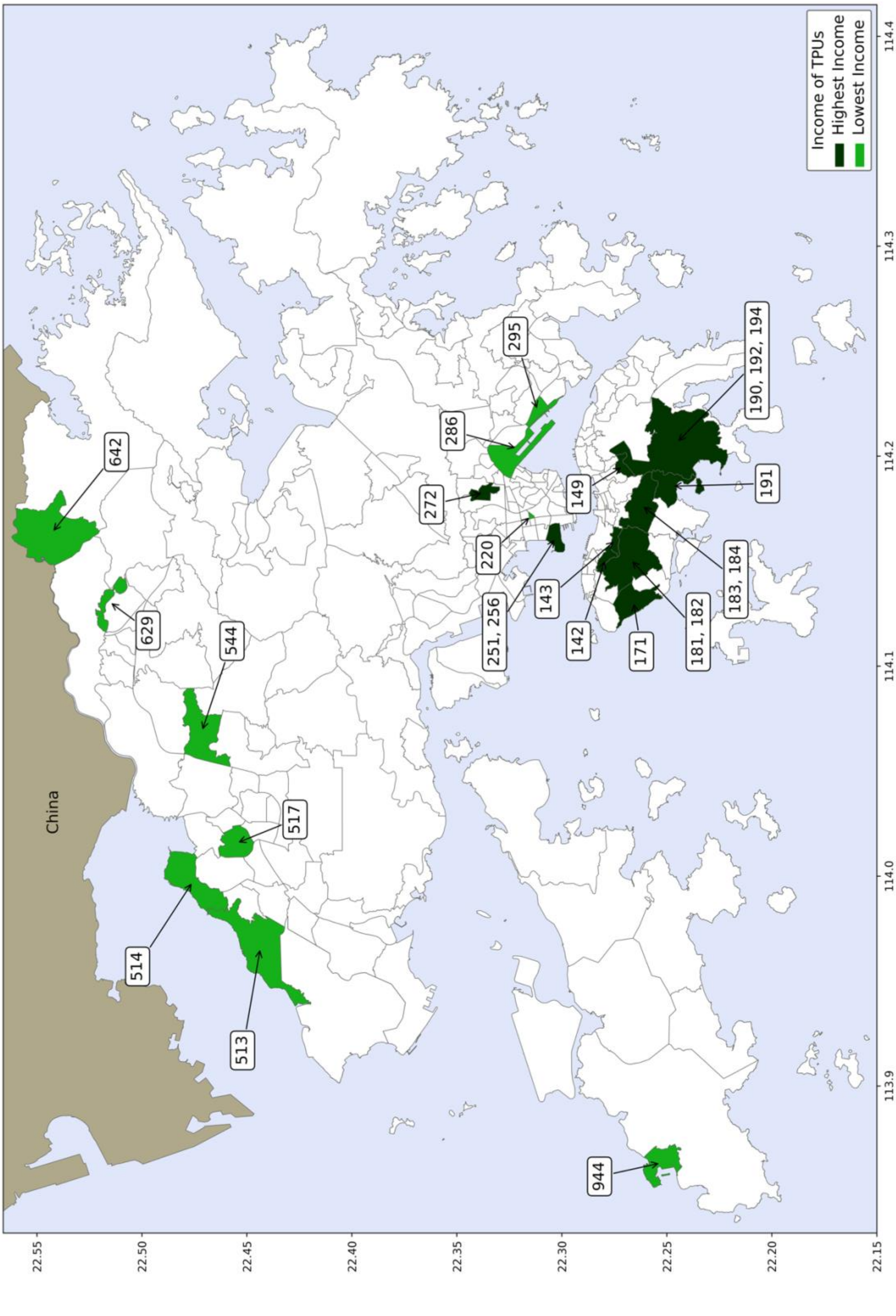


Table 24

TPUs with the Highest Median Domestic Household Income

Location	Region	Income (HK$) [a]
181, 182	HK Island	$ 155,920
171	HK Island	$ 114,210
149	HK Island	$ 113,160
183, 184	HK Island	$ 106,210
191	HK Island	$ 106,000
190, 192, 194	HK Island	$ 104,250
143	HK Island	$ 100,000
251, 256	Kowloon	$ 89,000
142	HK Island	$ 85,500
272	Kowloon	$ 85,320

Note:

[a] Median Income as of 2016 Census

Table 25

TPUs with the Lowest Median Domestic Household Income

Location	Region	Income (HK$) [a]
629	New Territories	$ 14,690
513	New Territories	$ 14,500
544	New Territories	$ 14,500
286	Kowloon	$ 14,000
517	New Territories	$ 13,950
642	New Territories	$ 13,000
220	Kowloon	$ 12,490
514	New Territories	$ 12,000
944	Outlying Islands	$ 12,000
295	New Territories	$ 9,520

Note:

[a] Median Income as of 2016 Census

Regarding other domestic characteristics, the area surrounding the restaurants on average have 2.675 individuals living in the household. Further, the population is primarily of Chinese ethnicity (86.935 percent), 62.664 percent participate in the labor force, and 31.190 percent have a post-secondary degree.

4.3.2. Site Characteristics

The number of competitors also varies significantly from the rural to the urban areas, with a range of 0 cuisine competitors in more rural areas to 539 in urban areas. The range of the number of price competitors is significant, though not as vast as the cuisine competitors. The range of price competitors is between 0 and 321; the convenience competitors have an even smaller span, with restaurants having anywhere from 0 to 33 competitors.

The five cuisine variables show that the local Hong Kong Cuisine is the most prevalent, with 28.838 percent of restaurants preparing that style of cuisine. Followed closely by Western style (26.855 percent), Chinese (17.377 percent), Japanese (13.994 percent), and Asian cuisine (12.936 percent). Regarding the price of the restaurants, the highest number of restaurants, 29.925 percent, have a price point of between HK$51-100, followed closely by Below HK$50 (29.359 percent), then HK$101-200 (26.066 percent), and finally, the most expensive category of Above HK$201 (14.650 percent).

Considering the location of the restaurants relative to the building in which it sits, the majority opened on the ground floor (66.647 percent). The majority of restaurants selected a location within a five-minute walk to an MTR station exit (66.826 percent). A small percentage of restaurants opened inside a hotel (8.793 percent), mall (13.979 percent), or close to an attraction area (16.677 percent).

4.4. Logistic Regression

The demographic variables used in the analysis were taken from the 2016 Hong Kong Census; a separate logistic regression model was conducted for only the restaurants that opened in 2016 and compared to the model that incorporated all 2016-18 restaurants. The results were similar, and it was deemed permissible by the researcher to link the 2016 census data to all restaurants that opened throughout 2016-18 with the understanding of potential fluctuations due to the natural movement of Hong Kong citizens.

4.4.1. Assumptions

The five assumptions of logistic regression, outlined in Section 3.5.1.2, were investigated, with the first ensuring the dependent variable was binary, with the values of 0 and 1 when 1 equates to restaurant success. Throughout the data cleaning process, the second assumption was met by ensuring that the observations are independent of each other, and there were no duplications or repeated observations of an individual restaurant.

To investigate the third assumption, the collinearity between the variables was checked. Using the variance inflation factor (VIF), the scores were examined to determine whether there were any scores higher than 10, thus indicating a correlation (Myers, 1990). Two variables had VIF scores higher than 10, Median Monthly Household Income, and Percent with Post-Secondary Degrees; to cure the issue, the Percent with Post-Secondary Degrees variable was removed. All other variables had VIF scores less than the threshold of 10, providing a decreased cause of concern (Bowerman & O'Connell, 1990) and remained in the model.

The fourth assumption regarding the linearity of the logit was investigated, and utilizing the Box-Tidwell Transformation Test, four variables were significant at a 0.05 level, indicating the assumption of linearity is not met. The four variables received a polynomial term as a curve was present (Wuensch, n.d.); the variables are Total Population, Median Monthly Household Income, Price Competitors, and Convenience Competitors. These four variables were grand mean centered (GMC) to reduce the multicollinearity, and then the squared term was added (Hofmann & Gavin, 1998; Kreft, 1995). The final assumption of large sample size was met as the model had 19 variables, meaning that the minimum number of observations should be 1,900, and the dataset has 6,710 entries.

4.4.2. Model Findings

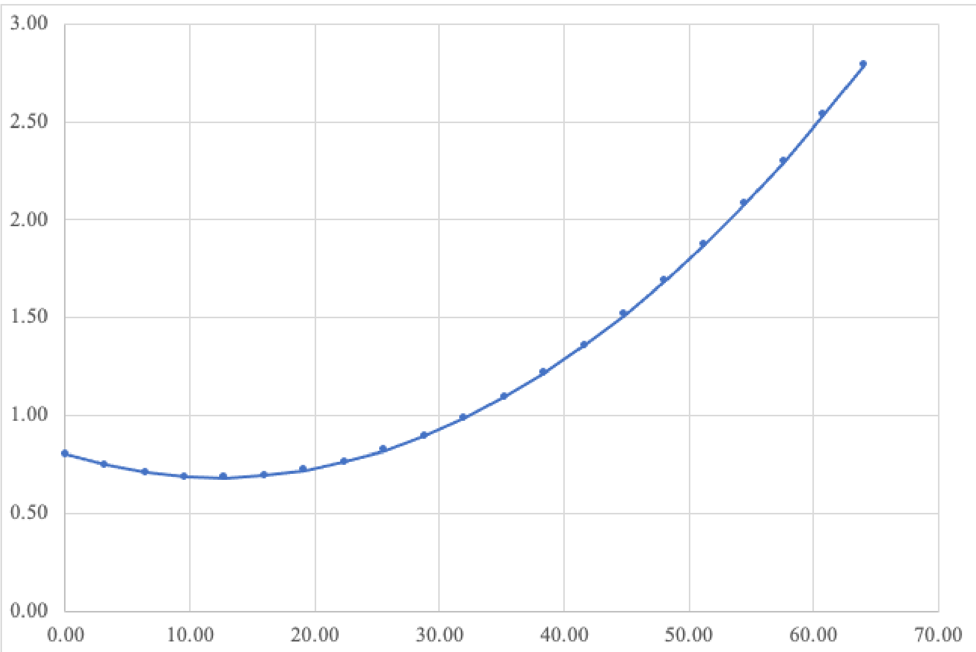
According to the logistic regression model (Table 26) created by utilizing the StatsModels package in a Python program (Seabold, Skipper, & Perktold, 2010), the model was able to predict the outcomes with a 71.27 percent accuracy. Out of the expanded 24 variables, 17 variables were significant. The model found that ten variables, Total Population, Mean Household Size, Percentage of Owner/Occupied Flats, Price Competitors, Convenience Competitors, Aggregate Review Score, Located on the Ground Floor, Inside a 5 Minute Walk to the MTR Exit, Inside a Mall, and Price HK$101-200, were highly significant at a 99 percent confidence level. Four variables, Median Age, Median Monthly Household Income, Percentage of Chinese Ethnicity, and Price – HK$51-100, were significant at a 95 percent confidence level. Three variables, Percent Participating in the Labor Force, Cuisine – Asian, and Price – Above $201, were moderately significant at a 90 percent confidence level. Utilizing a confidence level of 90 percent is appropriate when conduction exploratory research (Warner, 2012, p. 89).

4.4.2.1. Area Attributes

Considering the area attribute variables, the first highly significant variable was the total population. The relationship between the total population and the operational status of restaurants is a significant quadratic effect with the beta of 0.0008 ($p < 0.001$). The relationship between population density and restaurant success is described by a single curve in a hockey-stick shape (Chart 2). The curve is very shallow between 20.896 residents and the mean total population, 20,802.211, with the probability of success decreasing from the fewest number of residents around a restaurant site, 20.896, and the bottom of the curve at 12,344.735 residents. Then the potential success of a restaurant site continuously increases as the total population within a five-minute walk also increases, with the highest population around a site of 64,087.077 individuals.

Chart 2

Curvilinear Effect of Total Residents ('000)



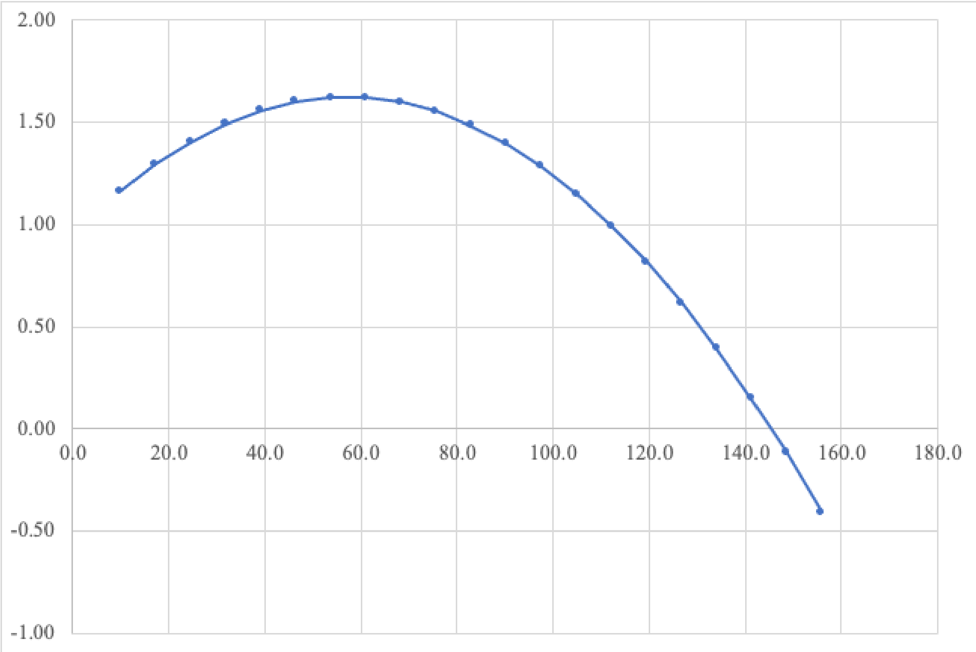
The median age was significant at a 95 percent confidence level, with a coefficient of 0.045 ($p = 0.013$). The logistic regression model shows that when the median age around the site increases, the probability of success also increases. The mean, median age across the 6,710 restaurant sites is 43.611 years old, and the marginal effect shows that for each additional year of age of the surrounding population, the probability of success increases by 0.868 percent.

The mean household size was highly significant at a 99 percent confidence level, with a coefficient of 0.582 ($p = 0.008$). When the household size increases, the potential success also increases. The mean household size of those living around the restaurant sites in the study is 2.675, and the marginal effect shows that for each additional person in the household increases the probability of success by 10.098 percent.

The relationship between Median Monthly Household Income and the success or failure of restaurants shows a significant quadratic effect with the beta of -0.0002 ($p = 0.025$); the relationship is described by a single curve in an inverse U-shape (Chart 3). The minimum observed income within a five-minute walk of the restaurants in the study is HK$9,924.79. While the median monthly household income increases, the potential for restaurant success also increases until the income reaches HK$57,031.53; afterward, the potential restaurant success decreases with the continuous increase of monthly household income.

Chart 3

Curvilinear Effect of Median Monthly Household Income ('000)



The percentage participating in the labor force is marginally significant at a 90 percent confidence level, with a coefficient of -3.224 ($p = 0.051$). The restaurants in the study had a minimum percentage of labor force participants of 46.70 percent and a maximum of 72.10 percent, with a mean percentage of 62.66 percent. For every percentage increase of people participating in the labor force, the probability of restaurant success decreases by 0.635 percent.

The percentage of Chinese ethnicity is significant at a 95 percent confidence level, with a coefficient of 1.390 ($p = 0.012$). An average of 86.93 percent of residents within a five-minute walk of the restaurants in the study are of Chinese ethnicity. The marginal effect shows that between the minimum percentage of Chinese ethnicity of the population (44.90 percent) to the maximum percentage (99.80 percent), for a single percentage increase of Chinese ethnicity, the probability of success increases by 0.271 percent.

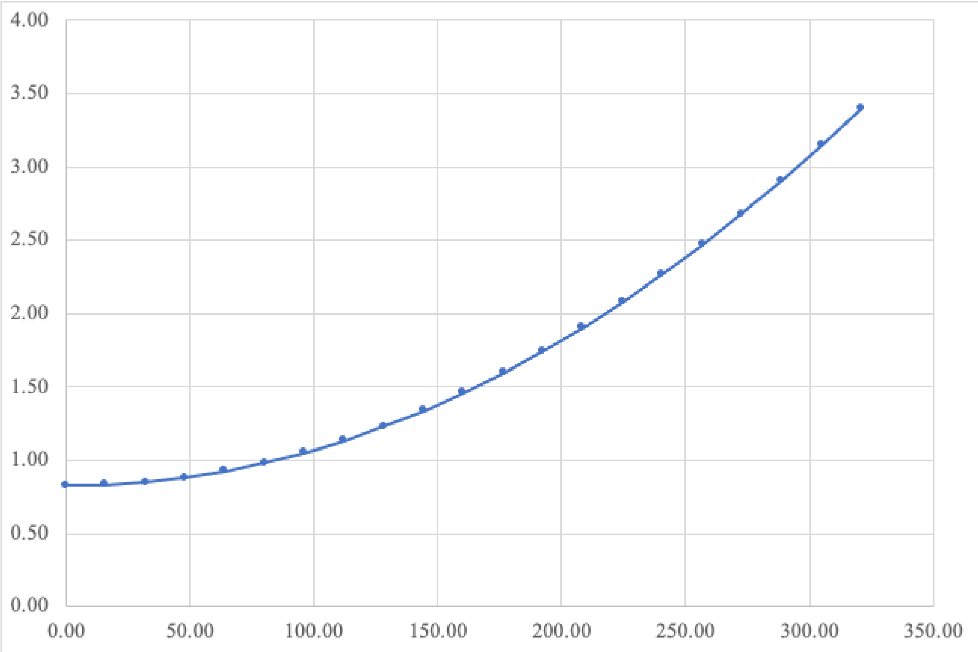
The percentage of the owner/occupied flats is highly significant at a 99 percent confidence level, with a coefficient of -1.575 ($p < 0.001$). The descriptive restaurant details show that the minimum percentage of owner/occupied flats of 0.10 percent, and a maximum rate of 82.90 percent, while the mean is 50.14 percent; the marginal effects show that for every percentage increase of owner/occupied flats decreases the probability of restaurant success by 0.309 percent.

4.4.2.2. Site Characteristics

The relationship between the number of Price Competitors and the operational status of restaurants shows a significant quadratic event with the beta of 0.00003 ($p = 0.002$); the relationship is described by a single curve in an upward arc-shape (Chart 4). The restaurants in the study had a minimum number of price competitors within a five-minute walk of 0, and a maximum of 321, with the mean number of price competitors of 67.94. The model shows that the probability of restaurant success very slightly decreases between 0 competitors and the bottom of the curve of 3.152; after the number of price competitors passes the bottom of the curve, the potential of restaurant success continually increases with additional price competitors.

Chart 4

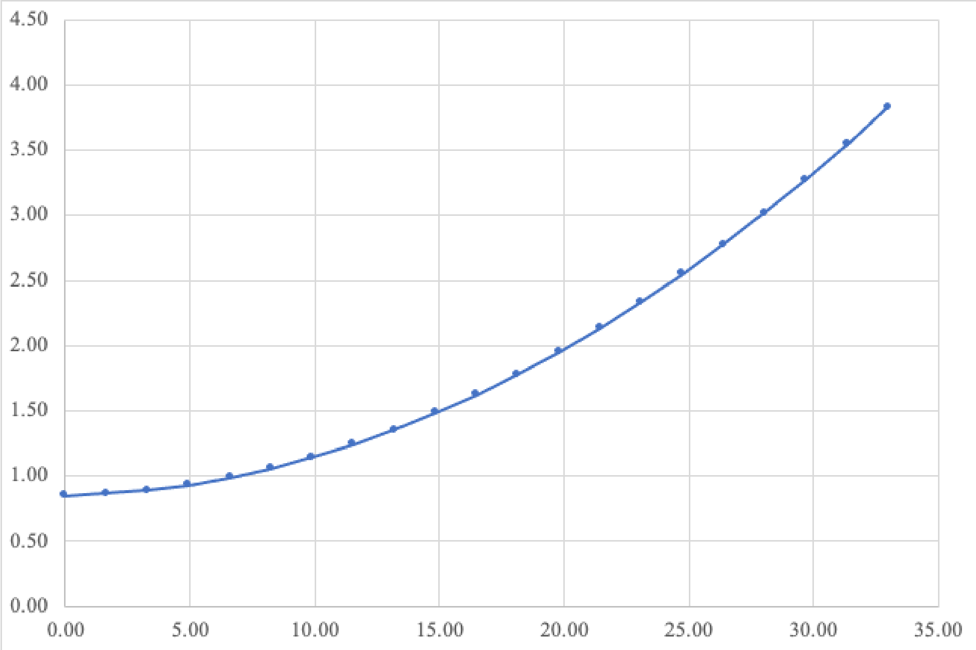
Curvilinear Effect of Price Competitors



Convenience Competitors and the operational status of restaurants has a significant quadratic effect with the beta of 0.00263 ($p = 0.004$); the relationship is a single curve in an upward arc-shape (Chart 5). The mean number of convenience competitors is 8.84, while the minimum number of observed convenience competitors is 0, and a maximum of 33. The bottom of the curve sits at -0.642, as a restaurant cannot have a negative number of competitors, the probability of restaurant success increases as the number of convenience competitors increases, slowly at first, and then at a faster rate.

Chart 5

Curvilinear Effect of Convenience Competitors



The model found that the variable of the ground floor was highly significant at a 99 percent confidence with a coefficient of -0.248 ($p < 0.001$). Although 66.647 percent of restaurants open on the ground floor, the probability of success decreases by 4.953 percent when opening on the ground floor over opening on any other floor.

Across the 6,710 restaurants in the study, 66.826 percent opened within a five-minute walk to an MTR station exit. The model found opening within a five-minute to the MTR station exit is highly significant, with a 99 percent confidence level and a coefficient of 0.259 ($p < 0.001$). The marginal effect shows that opening close to the MTR station exit increases the probability of success by 5.361 percent over opening outside of a five-minute walk.

Opening inside a mall was also highly significant at a 99 percent confidence level, with a coefficient of 0.614 ($p < 0.001$). There are 13.979 percent of restaurants that opened within a shopping mall in the 2016-2018 period. The marginal effects show that opening in a mall increases the probability of success by 11.229 percent over opening in any other location.

The cuisine of the restaurant is only partially significant; when compared to Hong Kong cuisine, only Asian cuisine is marginally significant at a 90 percent confidence level with a coefficient of -0.178 ($p = 0.080$). The marginal effects show that opening an Asian cuisine restaurant over a Hong Kong cuisine restaurant decreases the probability of success by -3.720 percent. Chinese, Japanese, and Western cuisines were not significant when compared to Hong Kong cuisine.

The price of the restaurant was significant when compared to the base price point of below HK$50. Operating a restaurant with a price point of between HK$51-100 was significant at a 95 percent confidence level with a coefficient of -0.172 ($p = 0.030$). The price point of HK101-200 was highly significant at a 99 percent confidence level, with a coefficient of -0.285 ($p = 0.001$). Restaurants with prices greater than HK$200 were marginally significant at a 90 percent confidence level with a coefficient of -0.199 ($p = 0.084$). The marginal effects show that the probability of success decreases by -3.552 percent when opening a restaurant priced between HK$51-100, -5.960 percent for restaurants priced between HK$101-200, and -4.173 percent for price points above HK$201 when compared to restaurants priced below HK$50.

Table 26

Logistic Regression Results

Variables	Coefficient	Standard Error	Sig.	Marginal Effects
Constant	-1.6439	2.2589	0.4668	
Total Population ('000) - GMC	-0.0194	0.0031	0.0000	
Total Population ('000) - GMC - Squared	0.0008***	0.0002	0.0000	
Median Household Income ('000) - GMC	0.0237***	0.0071	0.0008	
Median Household Income ('000) - GMC - Squared	-0.0002**	0.0001	0.0249	
Median Age	0.0447**	0.0180	0.0128	0.868%
Mean Household Size	0.5821***	0.2177	0.0075	10.098%
Cuisine Competitors	-0.00005	0.0007	0.9478	-0.001%
Price Competitors - GMC	-0.00016	0.0013	0.8987	
Price Competitors - GMC - Squared	0.00003***	0.0000	0.0020	
Convenience Competitors - GMC	0.0034	0.0084	0.6860	
Convenience Competitors - GMC - Squared	0.0026***	0.0009	0.0040	
Inside a 5 Minute Walk to MTR Exit	0.2589***	0.0672	0.0001	5.361%
% of Chinese Ethnicity	1.3901**	0.5559	0.0124	0.271%
% of Males	-0.6424	2.1096	0.7607	-0.126%
% of Owner/Occupied Flats	-1.5750***	0.3188	0.0000	-0.309%
% Participating in Labor Force	-3.2244*	1.6513	0.0509	-0.635%
Aggregate Review	0.2616***	0.0216	0.0000	4.868%
Located on the Ground Floor	-0.2481***	0.0696	0.0004	-4.953%
Inside a Mall	0.6142***	0.1034	0.0000	11.229%
Inside a Hotel	-0.1271	0.0993	0.2006	-2.643%
Close to an Attraction Area	-0.1197	0.0952	0.2087	-2.477%
Price - $51-100	-0.1721**	0.0792	0.0298	-3.552%
Price - $101-200	-0.2850***	0.0877	0.0011	-5.960%
Price - Above $201	-0.1993*	0.1155	0.0844	-4.173%
Cuisine - Asian	-0.1779*	0.1015	0.0796	-3.720%
Cuisine - Chinese	-0.1178	0.0874	0.1776	-2.435%
Cuisine - Japanese	-0.0865	0.1039	0.4052	-1.783%
Cuisine - Western	-0.0648	0.0838	0.4394	-1.326%
% of Correct Prediction	71.27%			

Note:

Dependent variable is a dummy variable (success (1) or failure (0) of new restaurants)

Standard errors are calculated by utilizing the White method.

*** Highly Significant ($p < 0.01$), ** Significant ($p < 0.05$), * Moderately Significant ($p < 0.1$)

4.5. Artificial Neural Network

4.5.1. Model Generation

The final model was created by first normalizing the data and running the dataset through 2,560 combinations with the potential number of neurons in the hidden layer being 24, 12, 25, or 49. The learning rate ranged from 0.1, 0.01, 0.001, to 0.005 while the momentum had the potential values of 0.0, 0.5, 0.9, and 0.99. The values considered for the decay were 0.0, 0.1, 0.01, and 0.001, and the number of epochs was either 1, 10, 50, 100, or 500. The value of the Nesterov was tested at the value of both True and False. The different combinations were tested utilizing a 10-fold cross-validation model, and the model accuracy was used to determine the best combination to create the final model.

4.5.2. Model Findings

The final model has an accuracy rate of 72.55 percent, with the parameters of:

Neurons:	25
Learning Rate:	0.005
Momentum:	0.99
Decay:	0.00
Epochs:	50
Nesterov:	True

The total number of weights in layer one is 600 and layer two 25 for a total of 625. The model is presented in Figure 4. The weights for both layers are found following in Table 28 and Table 29. To consider future restaurants' site locations with the model, the restaurant data will be normalized for the numerical variables utilizing the original data parameters, the categorical variables will be left at a value of 0, and 1, based on the structures explained above. The resulting values will be placed in the first layer neurons in the order listed in Table 27.

The values in the first node layer will be multiplied by the weights listed in Table 28; each node is multiplied by 25 specific weights progressing towards the second layer. The second layer will be a summation of the first layer values that have been multiplied by the weights. The second layer node values will then be multiplied by the weights found in Table 29 and added together to form the output value.

Table 27

Neural Network Design

Variables	Node Number	Minimum	Maximum
Total Population	1	20.896	64,087.077
Median Age	2	35.636	55.9
Median Household Income	3	9,924.792	155,920
Mean Household Size	4	1.596	4.1
% Participating in Labor Force	5	0.467	0.72104
% of Chinese Ethnicity	6	0.449	0.998
% of Owner/Occupied Flats	7	0.001	0.829
% of Males	8	0.396	0.582
Price Competitors	9	0	321
Cuisine Competitors	10	0	539
Convenience Competitors	11	0	33
Aggregate Review	12	0	5
Located on the Ground Floor	13		
Inside a 5 Minute Walk to MTR Exit	14		
Inside a Hotel	15		
Close to an Attraction Area	16		
Inside a Mall	17		
Cuisine - Asian	18		
Cuisine - Chinese	19		
Cuisine - Japanese	20		
Cuisine - Western	21		
Price - $51-100	22		
Price - $101-200	23		
Price - Above $201	24		

Figure 4

Artificial Neural Network

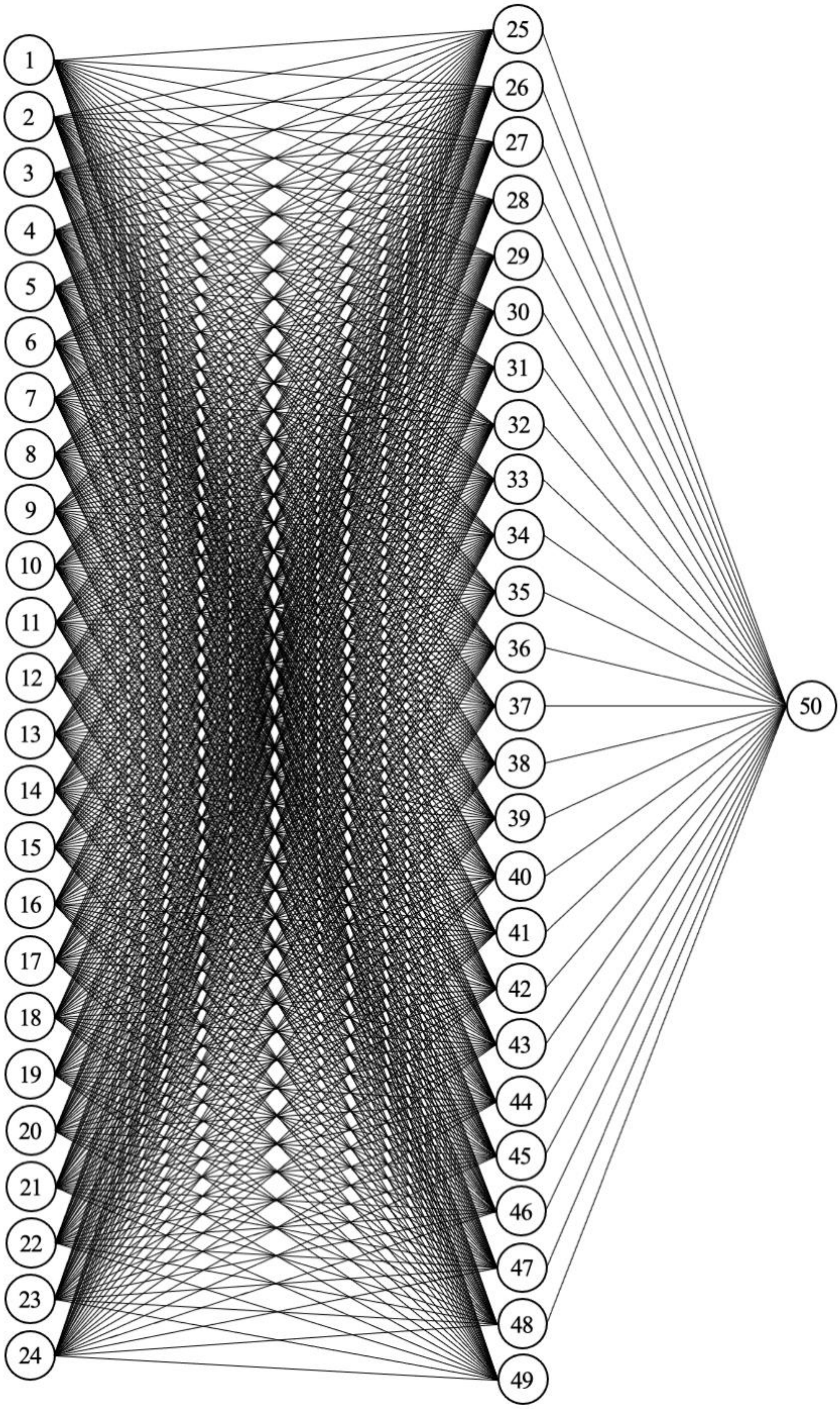


Table 28

Weights from Layer 1 to Layer 2

	25	26	27	28	29	30	31	32	33	34	35	36
1	0.308	-0.207	0.102	-1.532	0.310	0.188	-0.207	-0.459	0.722	-0.214	0.414	-2.440
2	-0.333	0.136	-0.337	1.085	0.627	-0.268	0.096	0.463	0.593	-0.334	-0.733	1.062
3	0.226	0.045	-0.420	0.322	0.250	0.175	-0.120	-0.125	-0.212	-0.376	0.600	0.198
4	1.193	-0.148	-0.477	0.241	-0.708	-0.275	-0.163	1.489	0.499	-0.469	0.160	-1.106
5	-0.502	0.377	-0.566	-0.104	0.035	-0.373	0.026	-0.030	-0.908	0.068	-0.287	-0.018
6	-0.321	-0.565	-0.343	-0.030	-0.464	-0.154	-0.101	0.800	-0.215	-0.142	-0.376	-1.041
7	-0.588	0.151	-0.351	0.134	-0.256	0.454	-0.320	-1.437	-0.308	-0.285	0.666	0.701
8	-0.558	-0.111	-0.240	0.039	-0.383	-0.442	-0.483	0.441	-0.699	0.051	-0.238	0.567
9	1.173	0.494	1.221	-0.794	-0.824	-0.031	0.211	-1.342	0.842	-0.098	-0.299	-0.782
10	0.133	-0.086	0.281	-0.013	0.238	-0.187	0.112	-0.990	0.206	0.077	-0.936	0.091
11	1.105	0.149	0.814	-1.142	0.642	0.132	-0.334	-1.594	1.692	0.174	-0.102	-0.373
12	0.244	-0.535	-0.022	-0.432	-0.313	0.012	-0.598	0.332	0.800	-0.074	-1.189	0.411
13	-0.317	0.047	-0.096	-0.461	0.056	-0.150	-0.299	-0.232	-0.281	-0.455	0.210	0.704
14	0.586	-0.603	-0.263	0.439	-0.153	-0.416	-0.327	-0.264	-0.378	-0.416	0.163	-0.439
15	-0.567	0.126	0.751	-0.793	0.078	0.677	-0.254	-0.639	0.073	0.224	-0.169	0.489
16	-0.091	-0.055	0.127	0.138	0.937	-0.107	-0.049	-0.191	-0.315	-0.329	-0.255	-0.045
17	0.397	-0.300	-0.117	-0.965	0.072	-0.190	0.033	0.619	0.332	-0.073	0.068	0.008
18	0.164	-0.339	-0.324	-0.118	0.076	-0.066	0.064	0.354	-0.163	-0.149	0.513	0.017
19	-0.301	0.523	-0.150	-0.038	0.246	0.493	0.231	-1.095	0.769	-0.366	0.187	0.229
20	0.229	0.476	0.576	-1.007	-0.534	0.383	-0.352	0.024	0.359	-0.305	-0.081	-0.106
21	-0.298	-0.517	-0.166	0.144	0.329	-0.038	-0.043	-0.575	0.411	-0.361	0.696	-0.188
22	0.330	-0.317	0.296	0.540	0.273	-0.187	-0.207	0.367	-0.753	-0.159	0.533	-0.479
23	-0.337	0.482	-0.449	-0.251	0.050	0.097	0.177	0.188	-0.584	0.002	-0.734	-0.116
24	0.130	-0.056	-0.600	-0.311	-0.445	-0.481	0.186	-0.337	-0.186	0.226	-0.068	0.551

	37	38	39	40	41	42	43	44	45	46	47	48	49
1	-0.509	-1.080	-0.373	-0.106	-1.037	0.086	-0.440	0.062	1.051	0.616	-0.408	0.201	0.617
2	0.046	0.116	0.047	0.111	0.424	-0.665	-0.050	0.034	0.435	-0.528	-0.054	-0.683	-1.123
3	-0.353	0.241	0.019	-0.472	0.282	0.486	0.283	0.200	-0.651	0.271	-0.130	-0.404	-0.194
4	-0.491	1.221	0.047	0.200	0.547	0.483	1.424	0.849	0.287	0.267	-0.116	-0.819	-0.136
5	-0.111	-0.359	-0.205	0.473	-0.317	-0.463	-0.723	-0.445	-0.322	-0.651	-0.095	0.121	0.152
6	0.364	0.059	-0.278	0.013	0.073	0.393	-0.467	-0.030	0.212	0.039	-0.436	0.095	0.240
7	-0.554	-0.244	-0.202	0.130	-0.573	-0.004	-0.138	-0.065	-0.808	-0.770	-0.513	-0.169	0.334
8	-0.092	-0.576	-0.324	0.749	0.214	-0.167	-0.304	-0.675	-0.223	-0.683	0.157	-0.158	0.012
9	0.434	-0.257	0.034	0.829	-0.948	-1.631	-0.047	0.589	1.573	1.703	0.419	-0.658	-0.631
10	-0.111	-0.284	-0.141	-0.423	-0.832	-0.648	0.098	0.156	0.460	0.920	0.668	-0.113	-0.470
11	-0.687	1.074	-0.005	-0.845	-0.420	-0.727	-0.329	0.441	0.714	0.549	0.136	0.073	-0.430
12	-0.108	0.102	-0.169	-0.928	-0.389	0.487	0.894	0.188	-1.047	0.293	-0.187	-0.640	-1.783
13	0.240	-0.015	-0.045	-0.522	0.811	-1.135	0.595	-0.108	-0.791	-0.027	-0.607	0.566	1.004
14	-0.630	0.608	-0.229	-0.468	-0.185	0.346	0.722	0.241	-0.813	-0.129	0.818	0.077	-0.151
15	-0.424	0.482	-0.045	0.635	0.021	0.325	-0.392	-0.775	0.820	0.005	-0.349	0.581	0.423
16	0.500	-0.529	0.210	-0.346	-0.801	0.388	-0.262	0.121	-0.391	-0.042	-0.529	0.105	0.161
17	0.819	-0.198	0.000	-0.223	0.746	-0.265	0.079	0.653	-0.792	-0.864	-0.166	-0.709	-0.596
18	0.007	-0.775	-0.228	-0.162	-0.709	-0.415	-0.094	0.386	0.738	0.103	0.745	0.793	0.244
19	-0.452	0.135	0.227	-0.535	-0.137	0.413	0.091	0.389	-0.463	-0.237	-0.023	0.698	-0.272
20	0.151	0.011	-0.424	0.545	-0.047	0.197	-0.142	0.119	0.174	0.266	-0.368	0.197	0.840
21	-0.109	0.039	-0.441	-0.641	-0.736	0.320	0.218	-0.297	0.668	-0.476	-0.209	-0.601	-0.437
22	0.363	-0.008	0.191	-0.499	-0.337	-1.073	-0.127	-0.753	-0.111	0.388	0.123	0.472	-0.269
23	0.537	0.026	0.023	0.512	-0.447	0.021	0.136	-0.229	-0.639	-0.296	-0.611	0.471	-0.612
24	-0.092	-0.062	-0.104	0.735	-0.662	-0.161	-0.301	0.193	-0.050	-0.168	0.247	-0.152	-0.257

Table 29

Weights from Layer 2 to Output Layer

	50
25	0.547
26	-0.815
27	1.019
28	1.050
29	-0.666
30	-0.902
31	0.008
32	0.680
33	0.540
34	-0.158
35	-1.037
36	1.055
37	0.743
38	0.435
39	0.405
40	-0.860
41	1.132
42	1.030
43	0.386
44	0.365
45	1.478
46	0.609
47	-0.695
48	-0.399
49	-0.718

4.6. Hypothesis Testing

Four area attributes and two site characteristics were deemed critical attributes for restaurant success and were hypothesized as listed in Table 30. There are four area attribute hypotheses; the first is the residential density and is a highly significant finding. However, the shape was in a U-shape and not in an inverse U-shape; as such, the first hypothesis is not supported. The second hypothesis that was investigated is the median monthly household income. The findings were significant, and the shape was an inverse U-shape; as such, the hypothesis is not supported. The proximity of household ages was also significant, and the hypothesis that it impacts restaurant success in an inverse U-shape is not supported as no curvilinear effect was present. The final hypothesis regarding the area attributes is the household size, and the model found that it was highly significant and positively impacts restaurant success; as such, the hypothesis is supported.

The site characteristic hypotheses regarding restaurant density are separated into three sub-hypotheses regarding the individual cuisine, price, and convenience competitor density, all impacting in an inverse U-shape. The cuisine competitors are not significant, both in a curved stature or in a straight line, and as such, the hypothesis is not supported. Regarding the cuisine and convenience competitors, they are both found to be significant in the model; however, both curves are U-shaped and not as an inverse U-shape; as such, the overall hypothesis is not supported. The final hypotheses regarding the distance of a five-minute walk to the MTR exit was highly significant and is a positive influence on restaurant success, and the hypothesis is supported.

Table 30

Hypothesis Findings

	Hypotheses	Finding		Result
		Shape	Significance	
Area Attributes				
H1	The residential density, in the vicinity, will impact restaurant success in an inverse U-shape.	Hockey Stick	Highly Significant ($p < 0.01$)	Not Supported
H2	The proximity of household income will positively impact restaurant success.	Inverse U-Shape	Significant ($p = 0.025$)	Not Supported
H3	The proximity of household age will impact restaurant success in an inverse U-shape.	Positive Increase	Significant ($p = 0.013$)	Not Supported
H4	The proximity of the overall household size will positively impact restaurant success.	Positive Increase	Highly Significant ($p < 0.01$)	Supported
Site Characteristics				
H5	The competitor density, in the vicinity, impacts restaurant success in an inverse U-shape.			Not Supported
H5a	The density of price competitors, in the vicinity, impacts restaurant success in an inverse U-shape.	Upward Arc	Highly Significant ($p < 0.01$)	Not Supported
H5b	The density of cuisine competitors, in the vicinity, impacts restaurant success in an inverse U-shape.		Not Significant	Not Supported
H5c	The density of convenience competitors, in the vicinity, impacts restaurant success in an inverse U-shape.	Upward Arc	Highly Significant ($p < 0.01$)	Not Supported
H6	The distance of the site to the MTR will negatively impact restaurant success.	Negative Increase	Highly Significant ($p < 0.01$)	Supported

4.7. Research Questions

The study also proposed two research questions to investigate. The research questions are as follows:

1. Which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure?
2. Considering the potential influence of the overall site characteristics, can a model be created to aid in restaurant site selection?

4.7.1. The Impacts of Success

The socio-demographic attribute that has the highest impact on success is the Mean Household Size, a single additional person in the household increases a restaurant's probability of success by 10.098 percent. The current Mean Household Size is 2.675, so an additional person creates a big difference. The second most important factor is the Median Age, as with each additional year of those living in the area surrounding the site, the probability of success increases by 0.868 percent.

The most significant impact in increasing the probability of success regarding site characteristics is opening a restaurant within a mall. Opening in the mall increases the probability of success by 11.229 percent over opening outside of the mall. The second attribute that has the most significant amount of impact on restaurant success is locating within a five-minute walk to an MTR exit; locating close to the exit increases the probability of success by 5.361 percent. Finally, locating on the ground floor is a close third in terms of impact, as those that open on the ground floor decrease the probability of success by 4.953 percent over those that open on any other floor.

4.7.2. Model Selection

Considering the second research question of whether a model could be created to aid in restaurant site selection. Two models were successfully created to aid in restaurant site selection. In determining which model is better at predicting site success, four elements are examined, model accuracy, the ROC curve, AIC, and BIC values.

The logistic regression model had a model accuracy of 71.267 percent, and the area under the ROC curve (Figure 5) is calculated at 0.654. The model has an AIC value of 7,792.185 and a BIC value of 7,989.714. The ANN has a model accuracy of 72.55 percent, and the area under the ROC curve (Figure 6) is calculated at 0.706. The ANN has an AIC value of 8,641.700 and a BIC value of 12,898.797.

The logistic regression model has both the lower AIC and BIC values in comparison to the ANN model; however, the ANN model has better model accuracy and a greater area under the ROC curve. While the AIC and BIC values showcase the overall model fit, the ROC value "provides an estimate of the predictive power of [the] models" (B. G. Everett, Rehkopf, & Rogers, 2013). While both models could be utilized due to having similar overall accuracies, the predictive power of the model is the most critical aspect in predicting future site success; as such, the ANN is deemed to be the preferred model.

Figure 5

Logistic Regression ROC Curve

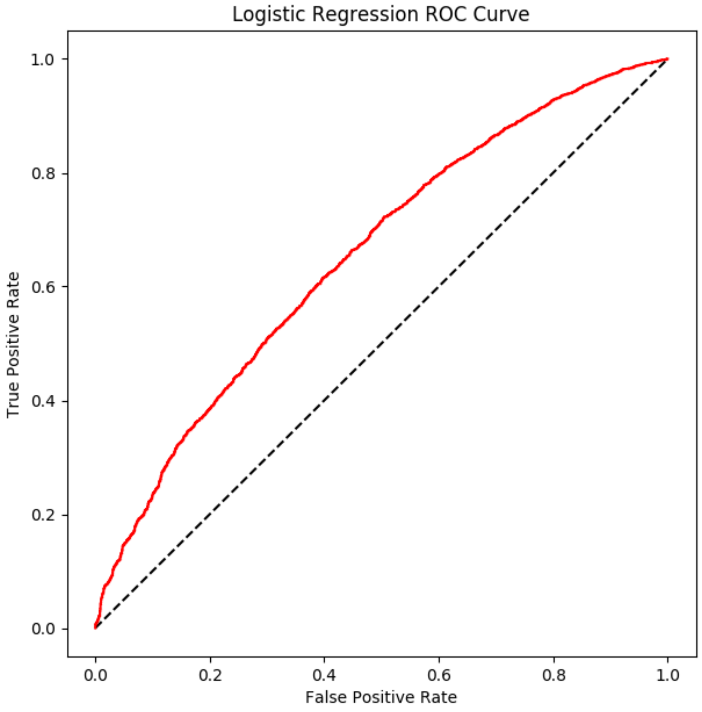
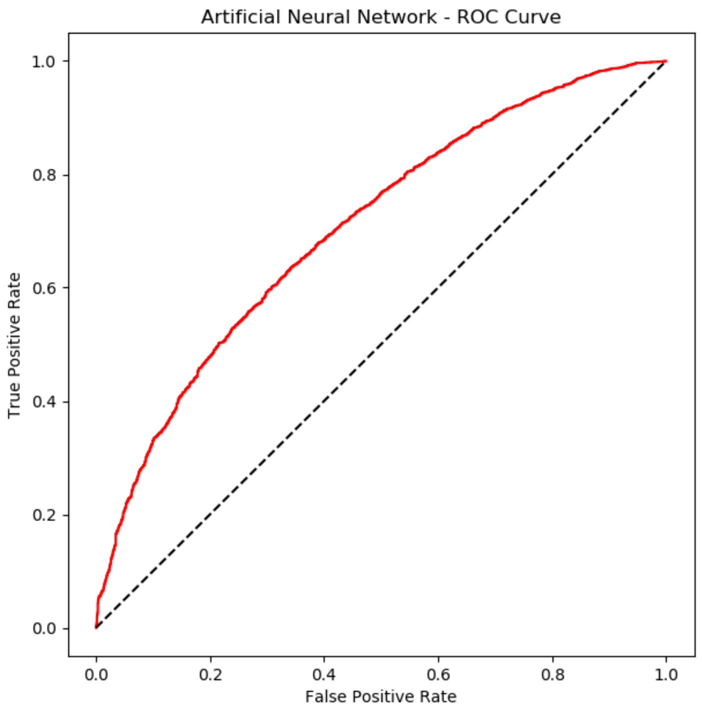


Figure 6

Artificial Neural Network ROC Curve



4.8. Conclusion

The study investigated 6,710 restaurants that opened in the years of 2016, 2017, and 2018, and examined 19 variables between the area attributes and the site characteristics. Of these 19 variables, the dummy variables were separated out, and a total of 24 variables were included; there are 17 variables that were significant above a 90 percent confidence level. There were six hypotheses presented; the study found that two hypotheses were supported while four were not supported. The study looked at two potential models, and between the logistic regression and the artificial neural network both were able to predict restaurant success. The model that had both the greater accuracy and area under the ROC curve, and deemed the ideal model for predicting the success or failure of a potential restaurant site is the ANN.

Chapter 5. Discussion

5.1. Introduction

This chapter examines the findings and seeks to understand why Hong Kong restaurants' survival rates differ from previous research on restaurant survival and failure. A discussion of the two research questions and hypotheses follow, and a connection to the previous literature is presented. Based on the findings, the research will investigate which TPUs showcase the ideal characteristics for potential restaurants both in the significant individual variables, the area attributes, site characteristics, and all variables combined.

An ANN Model Accuracy Test will be presented utilizing 188 new restaurants that opened in January of 2019 to examine the ANN's accuracy. Following the ANN Model Accuracy Test will be a discussion regarding the theoretical and practical contributions that are presented in Chapter 1. The final part of this chapter discusses the limitations faced in the overall research.

5.2. Evaluation

5.2.1. Survival of Restaurants

The predominant research that has been conducted on the survival of restaurants was conducted in Columbus, Ohio, in the United States in the years spanning 1996-1999 (Parsa et al., 2005). The research utilized 2,439 restaurants that opened in 1996 and monitored them over a three-year period (Table 31). The study did not add any additional new restaurants that opened in the subsequent years of the research and only utilized those that opened in the first year, considered the vintage year.

Table 31

Yearly Openings and Survival of Restaurants

Columbus, Ohio	1996	1997	1998	1999
Openings in Vintage Year	2,439	-	-	-
Survival One Year from Vintage Year	-	1,801	-	-
Survival Two Years from Vintage Year	-	-	1,332	-
Survival Three Years from Vintage Year	-	-	-	982

Hong Kong, SAR	2016	2017	2018	2019
Openings in Vintage Year	2,186	2,305	2,219	-
Survival One Year from Vintage Year	-	2,078	2,226	2,116
Survival Two Years from Vintage Year	-	-	1,807	1,950
Survival Three Years from Vintage Year	-	-	-	1,549

The study's original findings highlighted that the first-year failure rate was 26.16 percent, decreasing in year two to 19.23 percent, and further reducing in year-three to 14.35 percent. However, these failure rates only look at the failures compared to the total number of restaurants in the study, not the number that were still open at the beginning of the specific year. Utilizing the vintage year and removing previous years' failures, the new failure rate of year-two was 26.04 percent and year-three of 26.28 percent—the cumulative failure rate of 59.74 percent throughout the longitudinal study (Table 32).

The modified failure rates show a constant failure rate each year. Unlike the original findings, survival from one year to the next does not predict an increase in potential future survival. The revised calculations are done in the method suggested by Healy and Mac Con Iomaire (2018) and contains the limitation that the analysis holds a restaurant that opens in January and a restaurant that opens in December in the same year equal when considering year-over-year failure rates.

This research utilizes the number of days post-opening to investigate the failure rates in 365-day periods, creating a more robust investigation into the failure rates. The robustness can be shown that for Hong Kong restaurants that opened in 2016, the first-calendar year failure rate, those who closed before 2017, is 4.94 percent, for newly opened 2017 restaurants, the first-calendar year failure rate is 3.43 percent, and for the 2018 restaurants, the first-calendar year failure rate is 4.64 percent. Comparing these three calendar year failure rates, all of which are below five percent, to the failure rate for the first period of 0 - 365-days post-opening, which was 10.76 percent, more than double, showing that focusing solely on calendar years does not lead to the true nature of restaurant failures.

Overall, Hong Kong restaurants' three-year failure rate (27.88 percent) is lower than the rate conducted in Columbus, Ohio, between 1996-1999 (59.74 percent). Hong Kong also has a lower failure rate than in the study by English (1996) that found in El Paso, Texas, between 1990-1993, 40 percent of the restaurants failed. However, the Hong Kong failure rate is higher than the five-year study (2003-2008) done in Los Angeles County that out of the 8,205 restaurants, 20.9 percent failed (Self et al., 2015).

The lower rate of failure in Hong Kong can be explained due to the upfront capital requirements of finding and securing the correct location. The overall capital required to start a restaurant in Hong Kong is greater than most cities, as Hong Kong has the second-highest cost per square foot for commercial locations worldwide (Munish, 2017). The greater capital risk is one factor that lowers the risk tolerance of independent restaurateurs in Hong Kong (Jogaratnam, 2002). With restaurateurs ensuring their business plans are more concrete to entice financial institutions, investors, or partners to invest, they face a higher potential of success than previous studies.

Table 32

Restaurant Survival and Failure Rates

Columbus, Ohio	Year 1		Year 2		Year 3	
Yearly Failure Rate	Failed	Survived	Failed	Survived	Failed	Survived
1996	26.16%	73.84%	26.04%	73.96%	26.28%	73.72%
Cumulative Failure Rate						
1996	26.16%	73.84%	45.39%	54.61%	59.74%	40.26%

Hong Kong, SAR	Year 1		Year 2		Year 3	
Yearly Failure Rate	Failed	Survived	Failed	Survived	Failed	Survived
2016	4.94%	95.06%	13.04%	86.96%	14.28%	85.72%
2017	3.43%	96.57%	12.40%	87.60%	15.03%	84.97%
2018	4.64%	95.36%	15.41%	84.59%		
Average	4.34%	95.66%	13.62%	86.38%	14.65%	85.35%
Cumulative Failure Rate						
2016	4.94%	95.06%	17.34%	82.66%	29.14%	70.86%
2017	3.43%	96.57%	15.40%	84.60%	28.11%	71.89%
2018	4.64%	95.36%	19.33%	80.67%		

	0 – 365 Days		366 – 730 Days		731 - 1,095 Days	
Period Failure Rates	Failed	Survived	Failed	Survived	Failed	Survived
	10.76%	89.24%	12.14%	87.86%	8.02%	91.98%

Cumulative Period Failure Rates	0 – 365 Days		0 – 730 Days		0 - 1,095 Days	
	Failed	Survived	Failed	Survived	Failed	Survived
	10.76%	89.24%	21.59%	78.41%	27.88%	72.12%

While there is a lower failure rate for newly opened restaurants in Hong Kong, restaurateurs might still underestimate the nuances of undertaking a new restaurant venture. The underestimation shows in the number of restaurant failures in the period ranging from 92-180 days post opening. This ninety-day period has the greatest number of restaurant failures in the study (342 failures of 6,710 restaurants), a 5.097 percent rate. Two potential explanations are capital requirements and licensing changes.

While the capital requirements are more significant in Hong Kong than in previous restaurant survival studies, there may still be a miscalculation in the capital required for construction, decoration, soft and hard wares (Parsa, Gregory, & Terry; Parsa et al., 2005). Potentially there is a miscalculation in early business success, while the recommendation for restaurateurs to have a year of operating capital available after opening (Turiace, 2015), some restaurateurs could have considerably less and rely on early success to cover costs. Slow business growth could create a situation that the restaurateur runs out of capital. These potential errors could mean that restaurants face an early demise.

The second reason for the high failure rate in 92 - 180 days post opening is the temporary provisional license. The provisional license issued by the Food and Environmental Hygiene Department division only lasts 180 days. Both voluntary and involuntary reasons can explain the high number of failures between 92 - 180 days. The involuntary reason could be that the restaurateurs were not given a full license due to poor inspection reports. There is also the potential that the restaurateur voluntarily changes their mind due to underestimating the amount of effort or time needed to manage the restaurant (Camillo et al., 2008; Parsa et al., 2005). Additionally, if the restauranter is denied a liquor license, they could decide that long-term success is not feasible on food sales alone and not pursue a full license. Finally, the restaurateur could reconsider the overall concept and start anew by canceling the license application and reapplying with a new floorplan.

5.3. Research Questions

5.3.1. Research Question 1

The first research question, *which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure*, utilized the logistic regression model and identified seventeen of the twenty-four variables that significantly affect the probability of restaurant success. The first sub-question asks what is the most critical socio-demographic area attribute? The marginal effects of the logistic regression found that the most vital area-attribute is the mean household size. Hong Kong has a mean household size of 2.675 individuals within a household; for each additional person living in the household, there was an increase in the probability of success by 10.098 percent.

Although the literature found that an increase of overall family size decreases the number of times the household will eat away from home (Binkley, 2006; Byrne et al., 1998; Sunny Ham et al., 2004; Nayga Jr & Capps Jr, 1994); the amount that the larger households spent on dining out surpasses the expenditure of smaller households. This is in line with the most recent census that found a one-person household spent HK$3,638, and a four-person household spent HK$9,256 (Census and Statistics Department, 2016b) on food away from home. While the expenditure amount per household increased, the amount per person decreased. This can be due to a culture of family-style eating and supporting E.-J. Kim and Geistfeld (2003), who found that larger households dine at quick-service restaurants rather than full-service restaurants.

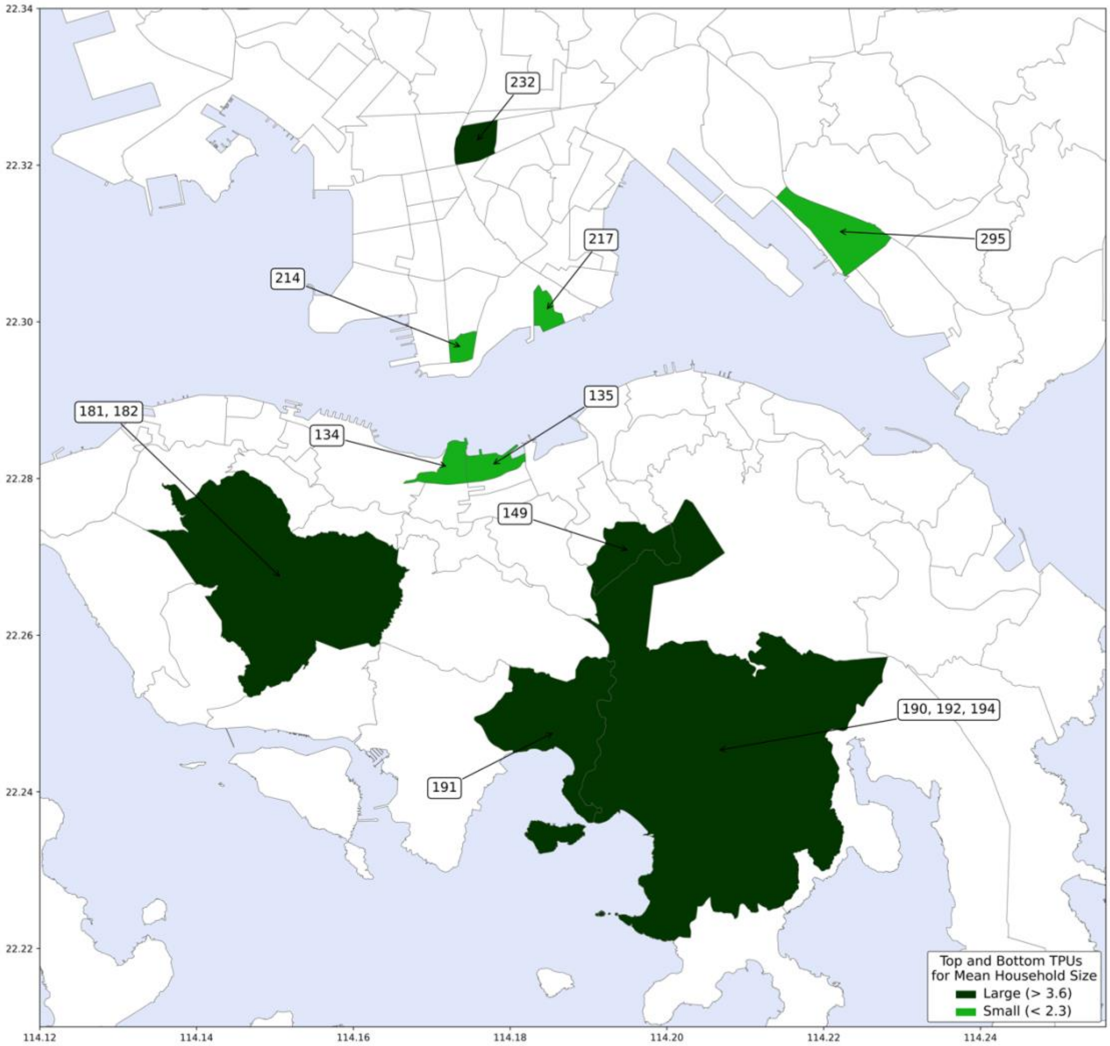
Restaurateurs seeking to open restaurants with a lower price point per diner should focus on opening in TPUs with a more significant number of mean persons per household. Out of the top five TPUs with a mean household size greater than 3.6 individuals, four are located on Hong Kong Island, in TPUs 149, 181/182, 190/192/194, and 191, while TPU 232 is in Kowloon (Map 14). Restauranters seeking a higher price point per diner should consider TPUs with smaller mean household size, TPUs 134 and 135 on Hong Kong Island and TPUs 214, 217, and 195 have a mean household size less than 2.3 individuals.

The second sub-question of the first research question looked to answer what is the most critical restaurant site characteristic that a restaurateur needs to consider when selecting a site. The logistic regression's marginal effects found that the most vital attribute is locating the restaurant within a shopping mall. Restaurants that open within a mall increase their probability of success by 11.229 percent over restaurants that open in other buildings.

The increased potential success of mall sites supports the concept that malls are retail nuclei (Guest & Cluett, 1974; Johnston & Kissling, 1971). Those who patron a mall become captive shoppers from which both the retail shops and food and beverage establishments can benefit. However, restaurateurs need to be aware that mall patrons do not necessarily visit the entire mall and potentially focus on one floor or a specific area (Brown, 1991, 1992). If malls do not have a central food area, then further site surveys of the most popular retail stores and situating close to them might provide an even more significant benefit than solely opening within a mall.

Map 14

TPUs for Mean Household Size



Opening a restaurant within a shopping area offers the single greatest increase in potential site success at 11.229 percent; however, the second most significant increase in possible site success is locating within a 5-minute walk to one of the MTR station exits. Opening in short walking time to an MTR exit increases the potential of restaurant success by 5.361 percent. Restaurateurs should consider one of the malls that contain an exit from the MTR station directly into the shopping area.

The logistic regression has indicated that the mean household size and locating in a mall are the most critical factors that restaurateurs need to consider for the area attributes and the site characteristics. While these are the two greatest influences, the other fifteen significant variables need to be considered. The ideal customer and location are discussed below in Sections 5.5.1 and 5.5.2.

5.3.2. Research Question 2

The second research question asks whether it is possible to create a model to aid in restaurant site selection based on the overall site characteristics' influences. The logistic regression and ANN models predicted restaurant site success with accuracy greater than 70 percent, with the logistic regression predicting at 71.27 percent and the ANN at a 72.55 percent accuracy. Either of these models standing by itself is enough to conclude that it is feasible to create a model aiding in restaurant site selection. However, the ANN's higher accuracy and the greater area under the ROC curve showcase why the ANN is the preferred model over the logistic regression model.

This research is the first to use both the area attributes and site characteristics within machine learning tools to predict future restaurant site success. The benefit of utilizing logistic regression and ANN models is the ability to adapt to new data. The models created can be continually updated, and patterns will be observed with the adoption of the additional data. For example, while restaurateurs seem to currently benefit by opening within a shopping mall, this benefit might change with potential customers moving to online shopping. The modeling techniques will be able to provide the updated importance of attributes or characteristics.

5.4. Hypotheses

Six hypotheses were presented within the literature review, and of the six, four are not supported, and two are supported. Considering the four hypotheses in the section of area attributes, the first hypothesis that residential density would impact restaurant success in an inverse U-shape was not supported.

The modeling found that while residential density was a significant variable, the shape was, in effect, a hockey stick. The curve is very shallow between 20.896 residents and 24,229 with the bottom of the curve located at 12,344.735 residents, but the likelihood of success increases as the density continues to grow. Fourteen of the twenty-five restaurants with the lowest number of residents surrounding them are located on Hong Kong’s outlying Islands. The remaining are based in locations close to beaches in the New Territories in areas such as Lau Fau Shan, Sha Tau Kok, Tai Po, Ting Kau Village, and Sai Kung. The initial flat end of the stick demonstrates that there is some benefit to opening in areas that residents day-trip to on their rest days; however, an immense benefit is to locate in the most residentially dense locations.

When the number of residents increases past 24,344.735, the probability of site success increases, the research did not control for the day-trip activities of the residents of Hong Kong due to the unavailability of data. However, the results show that while restaurants can succeed in low residentially dense areas of beaches, residentially dense areas provide a greater probability of success—offering support for the central place theory that the residents prefer to dine close to their primary residence. Restaurateurs should focus on potential sites that provide a high number of residents within a five-minute walk to the restaurant site.

The New Territories is both home to seven of the top ten least populated TPUs per square kilometer and the highest success rate at 72.63 percent. Of the top ten TPUs by success at the end of the 1,095 days with a minimum of 10 restaurants, two are in the New Territories, TPUs, 837, and 350. While the finding seems contradictory to the correlation between an increase of residential density and increased success, looking at the two TPUs shows that the findings hold. TPU 837 is the 25th most populous TPU with 81,238 residents, and its success rate is 94.12 percent. TPU 350 is the 12th highest in total population across the TPUs with 106,497 residents, and a restaurant success rate of 90 percent. While the overall region of the New Territories has a low density, there are TPUs that are incredibly dense and support the findings of greater residential density and restaurant success.

The second hypothesis stated that restaurants in areas with higher household income would see a positive impact on restaurant success; this is ultimately not supported. While income is a significant variable, the shape is an inverse U. Considering the grouping of the residents into four quartiles, while the highest income group has a greater food away from expenditure HK$8,420, over the lowest income group HK$2,153 (Census and Statistics Department, 2016b), they are not spending it around where they reside.

The curvilinear effects of income show that potential restaurant success increases until the median monthly income reaches HK$57,031.53; after that, the probability of success decreases rapidly. Showing that while the highest income group is spending more money on food away from home, they are not necessarily spending it around where they reside; these findings conflict with Salmon (2012) findings of credit card spending habits of the wealthy in New York City.

The individuals in the highest income groups are also most likely dining at restaurants with a per person price point above HK$201; this price point is found at 14.65 percent of restaurants, and not necessarily in the area surrounding their primary residence. The high-income diners are traveling to other TPUs to dine at restaurants with a higher price point than dining at local restaurants with a lower price point. The trend in Hong Kong is that the higher-income individuals have private cars; in 2017, there were 600,433 private cars registered (Transport Department, 2019), for a population of 7,336,585 (8.18 percent). Owning a car allows them to drive themselves or have their driver take them to one of the 68 Michelin-starred restaurants. Wealthy Hong Kong residents are also members of private member clubs, totaling over 500 across Hong Kong, which may also account for their food away from home expenditure. Restaurateurs in Hong Kong should look at potential sites in areas with a median household income between HK$52,000 to HK$62,000.

The third hypothesis looked at the household age and stated that household age proximity would impact restaurant success in an inverse U-shape. The household age variable was found to be significant; however, it was not an inverse U-shape, and the hypothesis is not supported. Instead, the marginal effects show that for every additional year that household age increases, the probability of success also increases by 0.868 percent, supporting the research by Harrington et al. (2011). Nayga Jr and Capps Jr (1994); Redman (1980) which found that household age is linked to the expenditure of food purchased away from home and, in turn, restaurant success.

The previous conclusions found that households with children up to the age of eighteen are not likely to dine out (Sunny Ham et al., 2004; Parsa et al., 2015) or that the elderly are less likely to spend money on food away from home (Binkley, 2006). Resulting in the correlation that restaurant success and household age is an inverted U-shape (Yang Yang et al., 2017). Looking at the restaurants in this study, the lowest mean household age in the surrounding area was 35.636, and the mean maximum household age was 55.900. Increasing age resulted in higher levels of restaurant success; however, there is not enough spread in the data to understand whether at a particular age, below 35.636 or above 55.900, the success drops off, and an inverted U-shape appears.

The final hypothesis from the area attributes considered the household size and stated that the proximity of the overall household size would positively impact restaurant success. As detailed above, this hypothesis is not only supported; it has the highest impact on potential restaurant success of all the area attributes.

Two site characteristics hypotheses are considered; the first looked at restaurant density and hypothesized that restaurant density, in the vicinity, impacts restaurant success in an inverse U-shape. As mentioned in Section 3.4.3, restaurant competitors are separated into three classes, cuisine, price, and convenience. While the variable of price competitors and convenience competitors were both highly significant variables, they were in an upward arc shape, and not the hypothesized inverse U-shape; the cuisine competitor variable was also found not to be significant. As all three competitor types did not follow the stated hypothesis, the hypothesis is not supported.

The curvilinear effects of both the price and convenience competitors are arc-shaped, curving upwards from the y-axis. The bottom of the arc in the price competitors is located at 3.152, while the convenience competitors bottom is at -0.642. Considering the bottom of the curve for both are close to zero competitors, and then arc upwards, this lends support to the concept of clustering and the principle of minimum differentiation (Hotelling, 1929).

The one drawback to clustering is the saturation point, which states that there is a maximum number of successful stores that a cluster can support (Guy, 1994). While Fields (2007) argues that the saturation point of restaurants will eventually occur, the threshold has not yet been met in either the price or convenience competitors in the Hong Kong market. Restaurateurs should look for locations that are close to similarly priced restaurant competitors as well as convenience competitors. This will provide them a greater probability of success than trying to be the only restaurant of a specific price in an area.

The second hypothesis in the site characteristics considers the distance of the site to the MTR exit stating that the distance of the site to the MTR will negatively impact restaurant success. As mentioned above, the probability of success negatively increases by 5.361 percent when locating a restaurant within a five-minute walk of an MTR exit; thus, the hypothesis is supported and further buttressed by the findings from Jang and Mattila (2005); Njite et al. (2008); Sevtsuk (2014) who found that potential customers prefer restaurants that are convenient and easy to access.

The findings also corroborate the previous research showing that people do not want to walk far (Lorch & Smith, 1993). A quarter-mile, approximately a five-minute walk, is a distance that people will walk before using other types of transportation (Atash, 1994). Restaurateurs should focus on finding a location that is close to an MTR exit, as that simple act will increase the probability of success.

5.5. Implications for Hong Kong Restaurateurs

The implications of this research look to present the ideal TPUs to start a restaurant based on both the ideal customer, utilizing the significant variables of the area attributes and ideal potential sites based on the significant site characteristics.

5.5.1. Ideal Customer

The ideal customer profile surrounding the restaurant site looks at the area attribute results of the logistic regression. The findings show that the total residential density surrounding the restaurant site is a hockey-stick shape, with the form having a shallow curve at the lower end of the density of residents and then increasing, the ideal site has a greater number of people residing around it. While the restaurant site considers a five-minute walk radius that can encompass multiple TPUs, the TPUs with the most significant residential density will provide the best potential number of customers. TPU 153 on Hong Kong Island and the nine TPUs, 221, 225, 228, 241, 242, 245, 266, 267, and 287 on Kowloon are the areas that have the most significant amount of potential success due to their residential density.

The findings show that the increased age of the resident surrounding the area provides a higher level of potential success than younger areas. The areas that should be considered are TPU 944 and 976 on the outlying Islands, TPU 544, 545, and 835 in the New Territories, TPUs 237 and 262 in Kowloon, and TPU 148, 166, and 175/176 on Hong Kong Island. These ten TPUs have the oldest median age, ranging from 49.3 to 55.9 years of age.

The residents with a median income of HK$59,250 provide the highest potential success. The TPUs that have around that level of income includes some on Hong Kong Island, TPUs 144, 145, 196, two in Kowloon, TPU 213/215/216 and 231, four in the New Territories in TPUs 720, 729, 839, and 832/834, and the final one on the outlying Islands in TPU 971/972/973/974. These areas provide the most significant likelihood of success when considering the median income.

The findings show that the greater the household size, the higher the level of potential success, the five areas that have the largest household size are on Hong Kong Island and the TPUs of 181/182, 190/192/194, and 191, and in Kowloon with the TPUs of 232 and 271. The lower the number of individuals participating in the labor force, the better the chance of success; the five areas with the lowest labor force participation rate are on the outlying Islands with TPU 941/942/943, Kowloon with TPUs 514, 544, and 545, and the New Territories with TPU 651/652/653.

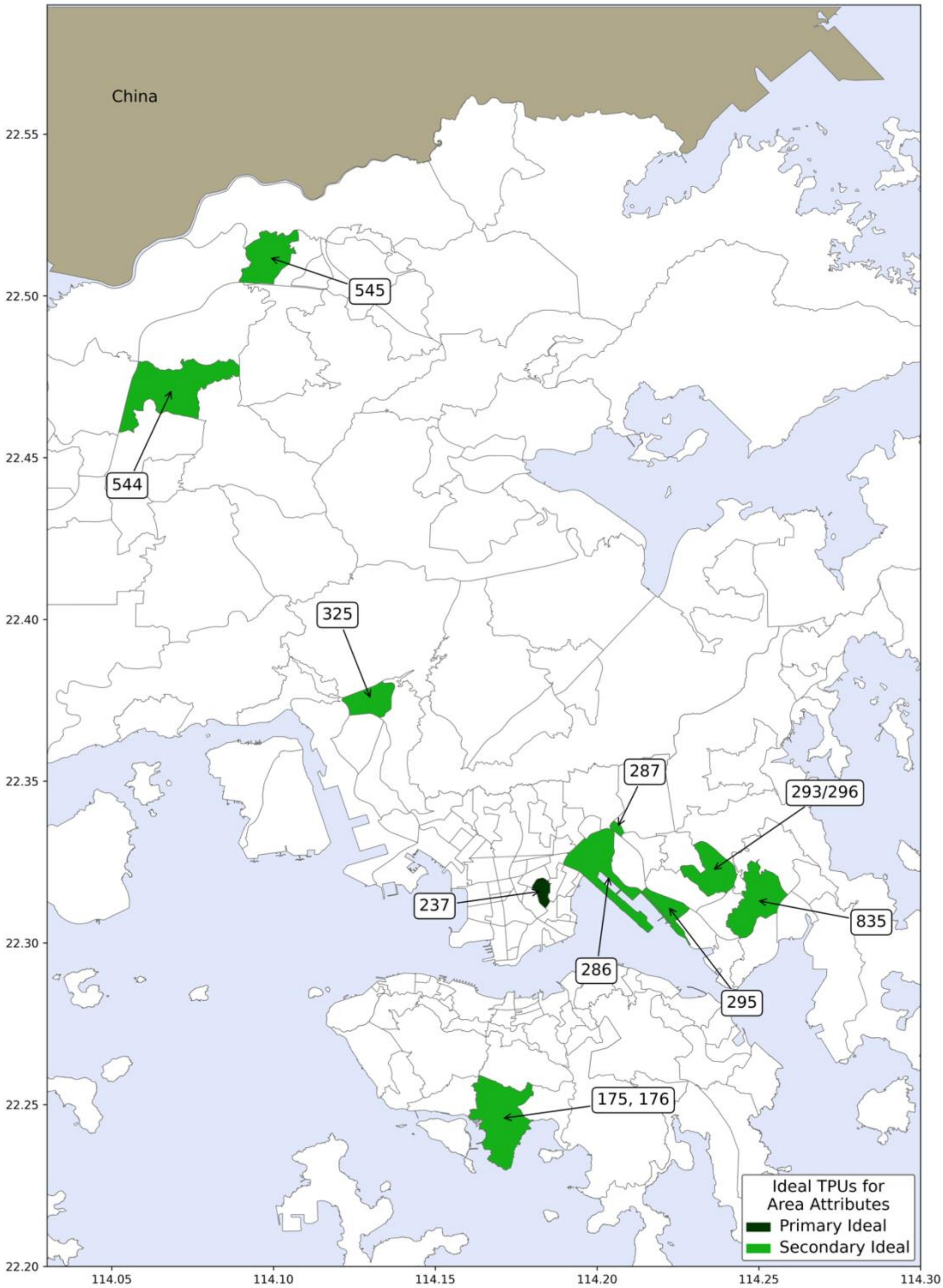
The greater the percentage of residents that are of Chinese ethnicity, the greater the potential success. The five TPUs with the highest percentage are in Kowloon 293/296, 286, and 295, and in the New Territories in 626 and 627. The lower the percentage of the owner/occupied flats, the greater the success. The lowest percentage areas are in Kowloon in TPUs 263, 286, 287, and 292, and in TPU 325 in the New Territory.

When considering the totality of the area attributes' importance, one TPU, 237 in Kowloon, is featured on the three separate lists and has the ideal residents for potential restaurant success. There were nine TPUs that were in the top ten of two categories the TPUs are in Hong Kong Island 175/176, in Kowloon in TPU 286, 287, 293/296, and 295, and four in Kowloon in TPUs 325, 544, 545, 835, and should be considered for potential restaurant sites. The locations of the ten TPUs is located in Map 15.

Considering the type of residents that are featured living in either a highly-populated area, with medium income, increased age, increased family size, decreased labor force, and increased number of renters, one type of residents that encompass those categories. By deductive reasoning, restaurateurs may therefore seek out areas where it is common to have a non-working family member, partner, spouse, or relative who contributes to the operation of the home and children as the presence of these individuals has led to increased restaurant success.

Map 15

Ideal TPUs for Area Attributes



5.5.2. Ideal Location

The ideal location is an amalgam of significant site characteristic variables realized from the logistic regression model. The significant variables include both price and convenience competitors, distance to an MTR exit, location within a shopping area, and the location within the building. One of the highly significant findings of the logistic regression model that is anathema to conventional wisdom is that opening on the ground floor decreases the potential success by 4.953 percent. This may be due to the percent of rent premiums for locating on the ground floor that may exceed the percentage increase of relative revenue. As the option of opening on the ground floor or opening not on the ground floor is available in any TPU, there is no ideal TPUs to be considered. However, the ideal TPUs for the other significant variables are presented next, starting with the price competitors.

The significant variable of price competitors showcased a curvilinear relationship. The vortex of the curve is at 3.152 and then arcing upwards; this positive increase in competitors and restaurant success is not only both surprising and unexpected but a meaningful finding for the site selection. Looking at the TPUs that restaurants operated in that had the most significant number of competitors, ten TPUs showcases this variable's ideal location. The TPUs are located in two of the four areas of Hong Kong, with six located on Hong Kong Island, TPUs 114, 121/123/124, 122, 132, 133, and 146/147, and four in Kowloon in TPUs 211, 212, 214, and 227.

The convenience competitors also showed a curvilinear relationship in an upwards arc, with having a high number of competitors being ideal. When investigating the TPUs that restaurants operated in that had the highest number of convenience competitors, ten TPUs showed to be the ideal location for a potential restaurant. The

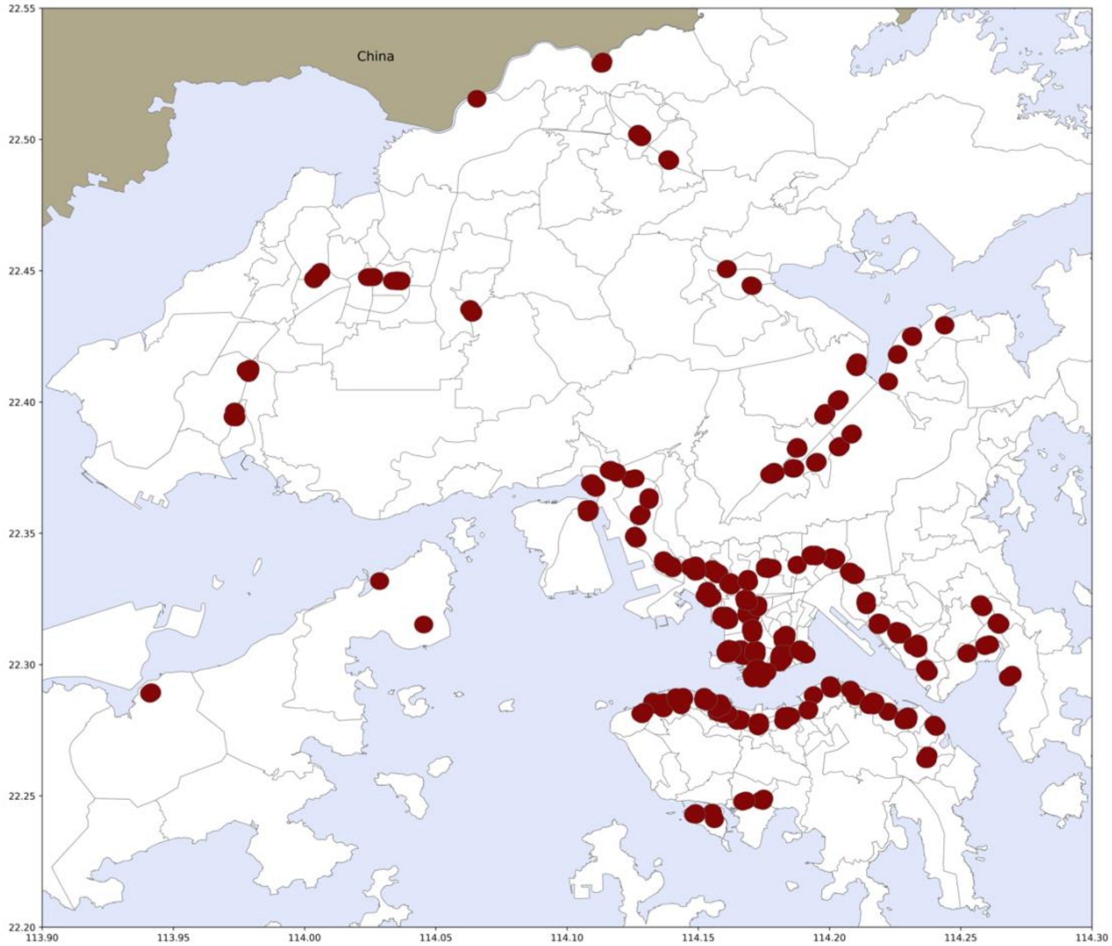
TPUs are located in Hong Kong Island in TPUs 122 and 153, in Kowloon TPUs 221, 222, 227, 229, 241, and 246, and the New Territories in TPUs 320/324/329 and 524.

The logistic regression findings show that opening within a five-minute walk of an MTR station increases the potential for success by 5.361 percent. While the five-minute walk from an MTR station only encompasses 64.859 square kilometers or 4.688 percent of the total land area of Hong Kong, the distance decay effect is shown to hold, that the longer the distance, the increase of the reluctance to travel to the location (Reimers & Clulow, 2004).

Considering the TPUs, 141 of the 214 have coverage from MTR exits (Map 16). Of the 90 MTR stations, which have a total of 553 exits, the ten that have the highest coverage of station exits are four on Hong Kong Island, including TPUs 112, 114, 121/123/124, and 122, and six in Kowloon including TPUs 211, 222, 225, 227, 229, and 245.

Map 16

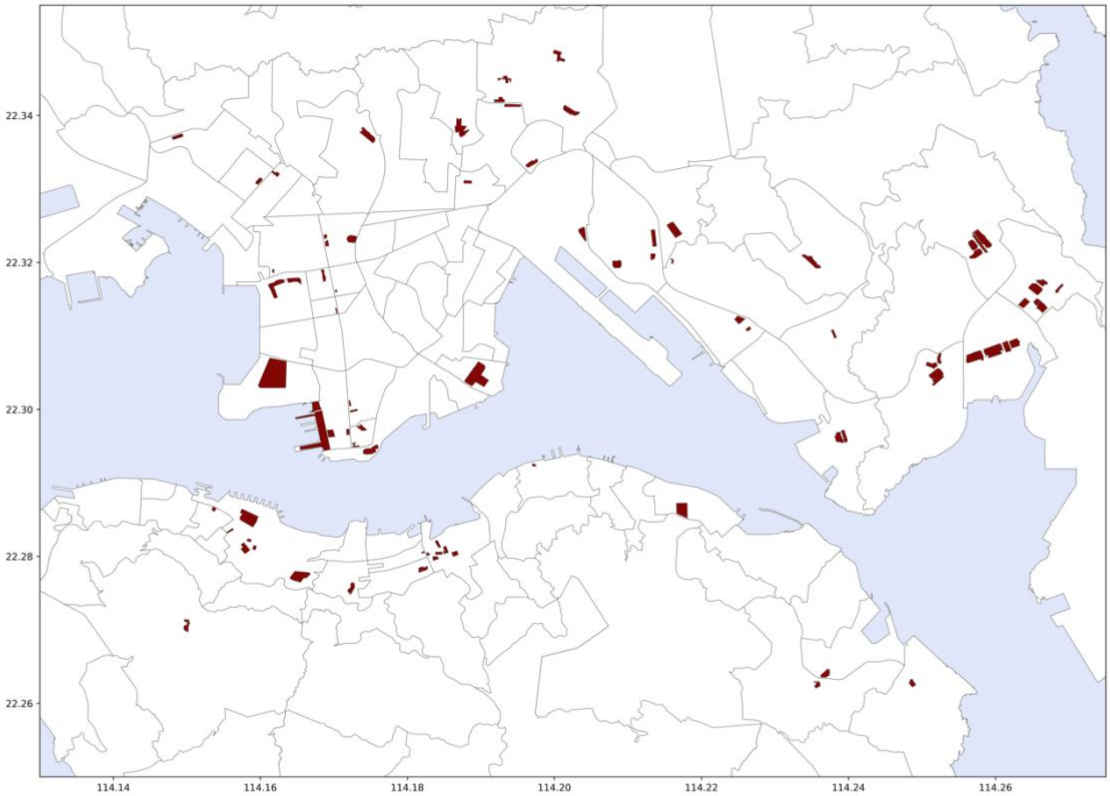
Areas Within a Five-Minute Walk of MTR Station Exits



The probability of restaurant success increases by 11.229 percent, when locating within a shopping area. Of the 214 TPUs, 67 encompass the 127 shopping areas (Map 17). The TPUs with the most shopping areas by square meters are Hong Kong Island in the TPU 121/123/124, Kowloon in the four TPUs 211, 214, 245, 251/256, and 253, and the New Territories in the TPUs 726, 758, 837, and 838. While opening in any shopping area provides a greater level of success than not opening in one, the availability of potential spaces is limited, which is why these ten TPUs are preferred to those that just have a singular shopping area.

Map 17

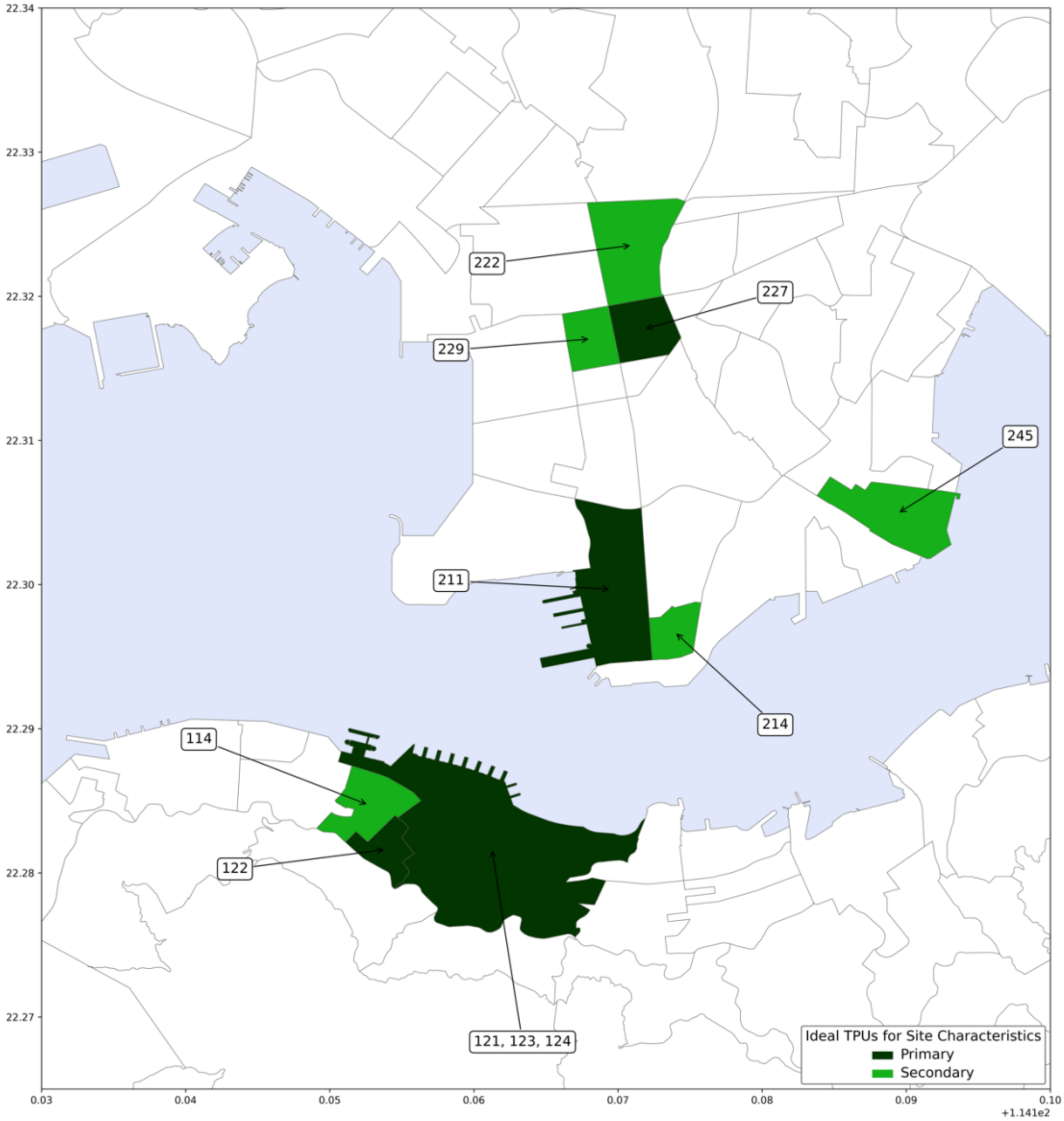
Location of Shopping Malls



When considering the totality of the site characteristics importance, four TPUs were featured on the lists three times; the TPUs of 121/123/124, 122, 211, and 227 have the ideal characteristics for potential restaurant success. There are five TPUs in the top ten of two categories, the TPUs of 114, 214, 222, 229, and 245. While there are eighteen on a list once, no one TPU showed to be a better area than others; as such, the top nine TPUs are located in Map 18.

Map 18

Top TPUs for Site Characteristics



5.5.3. Implication Conclusion

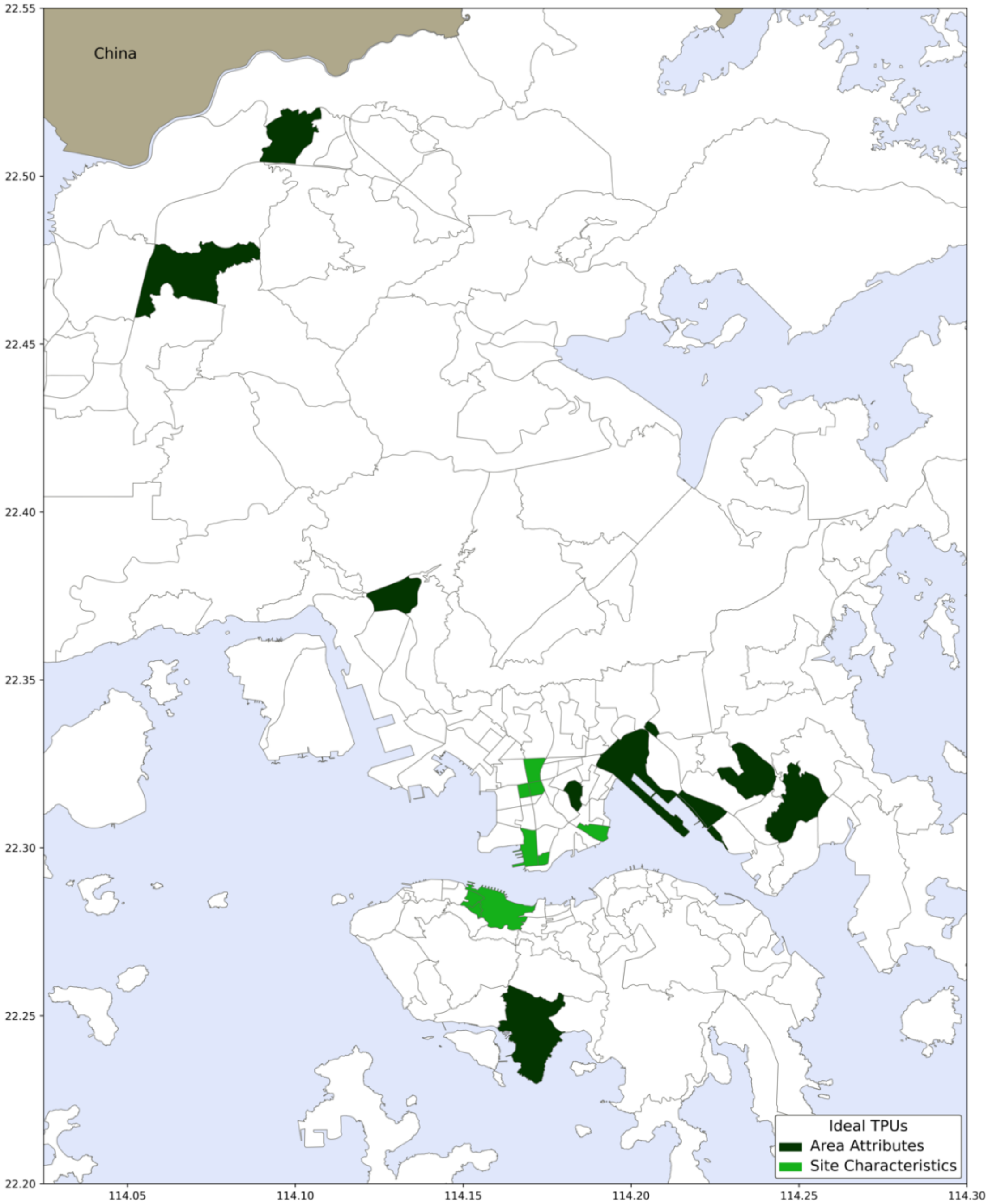
The ideal TPUs for area attributes and the ideal TPUs for site characteristics did not overlap for the ideal location to open a restaurant. However, when looking at the location of the TPUs (Map 19), the district of Kowloon contained eleven of the nineteen TPUs, with Hong Kong Island and the New Territories, both containing four.

The locations on Hong Kong Island are spread out between the north and south side of the island and are not close enough to walk between. However, the eleven TPUs located in Kowloon are within a close distance and opening in one of the following TPUs: 211, 214, 222, 227, 229, 237, 245, 286, 287, 293/296, and 295 provides the location that will give the best potential for success (Map 20).

The failure rate of TPUs, outlined in Section 4.3, conveyed that four TPUs located in Kowloon were consistently in the top ten list of the highest failure rate over the three years. However, the ideal TPUs, while also located in Kowloon, are different from the TPUs that have the highest failure rate. The Kowloon region contains both the most ideal TPUs and the least ideal TPUs, and success or failure depends on selecting the optimum site.

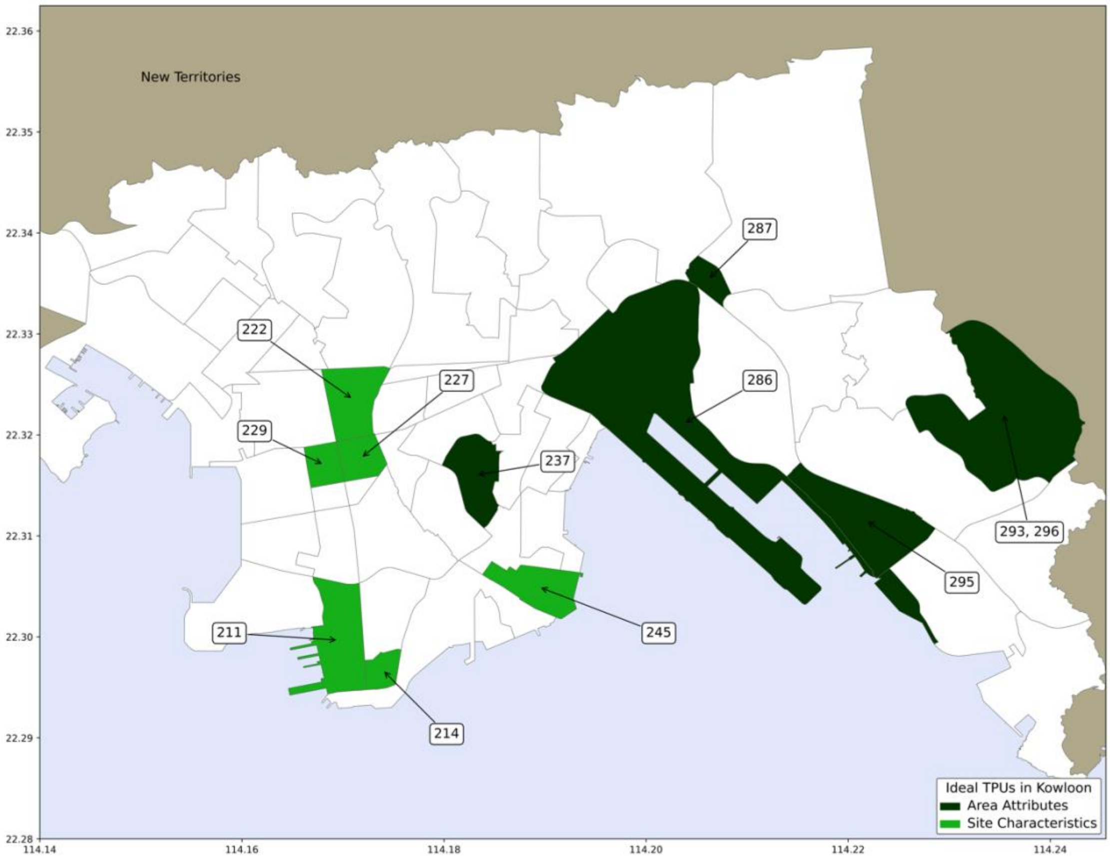
Map 19

Ideal TPUs



Map 20

Ideal TPUs in Kowloon



5.6. ANN Model Accuracy Test

The ANN Model Accuracy Test investigated 188 restaurants that opened in January 2019, having all the requisite variables. The restaurant data were normalized based on values in Table 27, located in Section 4.5.2, to ensure consistency between the model's training set and the new prediction set. A Python program was utilized to load the previously trained model and weights to predict the new restaurant data. The ANN model had an accuracy rate of 82.447 percent of correctly predicting all 188 restaurants' restaurant sites' success or failure in the first year of operation. Five randomly selected restaurants opened in the first week of January 2019 are investigated to examine the model's predictive potential.

Two of the five restaurants used in the ANN Model Accuracy Test opened on Hong Kong Island in TPUs 146/147 and 122, and three opened in Kowloon in TPUs 211, 226, and 291. Of the five restaurants, Restaurants 1, 2, 3, and 5 were successful as measured by still in operation as of 31 December 2019, while Restaurant 4 failed as it closed in mid-December 2019. The characteristics of the five restaurants and the ANN predictions are in Table 33.

Utilizing the ANN model that was deemed more accurate in Section 4.7.2, the ANN correctly predicted the outcome of four out of five restaurants; Restaurants 1, 2, and 3 are both predicted to be, and are, successful. Restaurant 4 was both predicted to fail, and the restaurant closed in mid-December 2019. The singular incorrect prediction is Restaurant 5, which, while it was predicted to fail, survived. The actual prediction value of Restaurant 5 was 0.4890, so while the prediction was incorrect, it was a borderline failure prediction based on a 0.5 cutoff between success and failure. The ANN Model Accuracy Test results showcase the ANN's ability to predict the success and failure of restaurant sites.

Table 33

ANN Model Accuracy Test

	Restaurant 1	Restaurant 2	Restaurant 3	Restaurant 4	Restaurant 5
Operation Status	Successful	Successful	Successful	Failed	Successful
Census					
Total Population	13,476.96	13,546.07	27,808.59	10,972.31	28,947.83
Median Age	39.451	47.084	43.031	41.902	41.723
Median Household Income	23,775.01	43,983.17	19,971.33	25,729.78	29,625.59
Mean Household Size	2.03	2.303	2.531	2.439	2.92
% Participating in Labor Force	0.593	0.67	0.601	0.598	0.65
% of Chinese Ethnicity	0.722	0.699	0.963	0.662	0.704
% of Owner/Occupied Flats	0.423	0.498	0.319	0.451	0.467
% of Males	0.407	0.489	0.461	0.446	0.448
Competitors					
Price Competitors	86	159	42	79	59
Cuisine Competitors	93	366	16	47	32
Convenience Competitors	8	8	8	9	8
Aggregate Review	4	4.5	4	4	4
Cuisine	Hong Kong Style	Western	Chinese	Asian	Western
Price	Below $50	Above $201	Below $50	$51-100	$51-100
Site Characteristics					
Located on the Ground Floor	Yes	No	No	Yes	Yes
Inside a 5 Minute Walk to MTR Exit	Yes	Yes	Yes	Yes	Yes
Inside a Hotel	No	No	No	Yes	Yes
Close to an Attraction Area	No	Yes	No	Yes	Yes
Inside a Mall	No	No	No	No	No
Artificial Neural Network					
Class Prediction	Successful	Successful	Successful	Failed	Failed

5.7. Contributions

The research contributes theoretically and practically to restaurants and site selection, which may be expanded and adjusted to markets beyond Hong Kong. Additionally, it can also be utilized by other hospitality or retail environments. This section will describe what the contributions are, first the theoretical, then the practical side.

5.7.1. Theoretical Contributions

The research contributes to the current theories by enhancing and expanding the current understanding of site selection utilizing spatial analysis of the restaurant characteristics. Three different theories regarding site selection have been considered and have been expanded upon, the principle of minimum differentiation, central place theory, and the bid rent theory.

Harold Hotelling initially conceived the principle of minimum differentiation in 1929; the principle theorizes why businesses cluster. The explanation was that to capture their fair share of the business; it is ideal for opening competing shops next to each other. The theory argues that the clustering of competitors offering similar prices helps attract customers who desire to browse multiple options before making a decision.

Previous research has suggested that restaurants can benefit from economies of price agglomeration (Teller & Reutterer, 2008; Yang Yang et al., 2017), as potential diners desire to comparison shop before making a decision (Fischer & Harrington Jr, 1996; Fisher, 1997; Krider & Putler, 2013). This study has confirmed the previous suggestions and that the principle of minimum differentiation holds in the restaurant's realm. However, the agglomeration of restaurants by cuisine is not critical, and famous food areas known for their cuisine, such as Little Italy, are only meaningful if the restaurants have comparable prices.

The central place theory proposes that residents prefer to patron shops that are in close vicinity to their place of residence—first proposed by Walter Christaller in 1933. The premise behind the theory is that there is a centrally located area to support the surrounding population. Christaller proposed that consumers travel to the nearest location with the required goods or services. While it was initially proposed when considering rural areas and one town's pull factor over a second, the research has found that the central place theory holds when considering restaurants in Hong Kong. Holding all variables constant, residents prefer to dine at restaurants around their primary residences and that restaurants face a more significant potential success by locating in dense residential areas. The shopping malls findings also confirm the central place theory, as individuals do not stray far from where they are, and shopping malls are self-containing areas that provide the shoppers with everything they need.

The bid rent theory says that the sites located within a central business district demand the highest rent, as they also have the highest number of potential restaurant patrons due to the workforce's density. The research looked to see if this holds when considering tourist attractions as those are areas that draw a lot of tourists and day-tripping residents. However, the study has found that the bid rent theory does not hold to attraction areas as it does for central business districts. Not only were the attraction areas not a significant variable in the logistic regression results, but the restaurants locating in a tourism area also had a greater failure rate than those not in a tourism area (32.887 percent vs. 29.208 percent). The bid rent theory cannot be expanded past the current central business districts.

5.7.2. Practical Contributions

The groups that will benefit from the practical contributions are restaurateurs, banks, investors, landlords, and consultants. Before this three-year longitudinal study, in house modeling techniques were mainly only available to major restaurant franchises. The research shows that not only is it feasible to have probability site modeling for independent restaurants, but the model can also predict the potential success or failure accurately over 70 percent of the time.

The second practical contribution that this research provides is the ability to showcase what variables are most critical to restaurants' overall success. The attributes related to the potential customers residing around the potential site in addition to the characteristics of the site itself. The increased potential success of a restaurant by selecting a site that not only has a high population density but also has households with multiple-individuals residing provides a guideline on what restaurateurs need to consider for their potential customers.

The knowledge to select a site with a cluster of restaurants with a similar price point provides more advantages than a site with a cluster of similar cuisine is critical for those either opening or financing a restaurant. While the clustering is crucial, the simple step of opening within a shopping mall increases the probability of success by over 11 percent. The captive nature of shopping malls provides immense benefits over sites located on the street or in multi-story buildings.

The contributions of the importance of specific area attributes or site characteristics are critical to restaurateurs trying to decide between two locations that the ANN has ranked as a potential successful location. The ANN should be utilized in conjunction with the logistic regression results to select a site that provides the most significant potential for success.

5.8. Limitations

The research is not without limitations; while the logistic regression and ANN models utilized nineteen unique variables; some variables were not included. The omitted variables were either considered a non-factor in Hong Kong or were not included due to the unavailability of the data. Two variables that were part of previous restaurant studies, but omitted in this study, were parking availability and staff availability.

Within Hong Kong, around eight percent of the population own a private car, while the remaining residents rely on public transportation to maneuver across the city; thus, the variable of parking spaces around a potential restaurant site was deemed a non-factor. The second variable of staffing availability was also mentioned in multiple pieces of literature. While this variable might be included in site-feasibility studies in more rural areas, the size and availability of transportation within Hong Kong deemed this variable unnecessary to include in the modeling.

Four other variables were not included due to the unavailability of the data, the day-tripping activities of Hong Kong residents, the restaurant floor size, the monthly rent of potential restaurant sites, if the restaurant was a franchise or independent restaurant, and the service style of the restaurant. Each potential restaurant site's floor size was not readily available for the thousands of potential restaurant sites. Although some floor plans are available online, either at the building's website, or on commercial realtors' websites, the majority was not available, and the variable was not included in the modeling.

The rent for the potential site was also desired to be included, but as with the floor size, few sites had data available. The commercial realtors' websites listed the asking price of currently available restaurant sites; of course, those sites already occupied are not included. Additionally, the prices listed were the asking price and not the negotiated price. The data were from the time of the research conducted and not the time that a lease would have been executed, as such was deemed both unreliable and did not provide enough data points to be included.

The type of restaurant, franchise or independent, was considered in previous studies of restaurant failure. This study did not include a separate variable due to the data not being available. While some franchise restaurants are known, not every restaurant could be reliably separated into the dichotomous group, and as such it was deemed more appropriate to exclude the variable. Previous studies on dining out characteristics have also utilized a variable of either limited or full-service styles of restaurants. For the restaurants included in the study, the restaurant style data was not readily available. This study instead utilized the price point of the restaurant to differentiate types of restaurants.

The research considers a three-year longitudinal study; the researcher desired to have a more extended study period to include more restaurants in the modeling; however, the data from prior years was not available from the Hong Kong Government. The Food and Environmental Hygiene Department denied the request for the restaurant license database going back further years.

In this study, restaurant success is defined as remaining open at the end of the period. There is no profitability hurdle or even a break-even threshold. Some restaurants might start having a positive net profit quicker, or at a higher rate than other restaurants;

unfortunately, independent restaurant financials are not available to the public. The lack of availability to the financial documents also limits the understanding on which TPUs provide a greater financial return than solely remaining in business. While some restaurateurs will have higher returns on their investment, this study focuses on restaurants remaining in operation, no matter the amount of profit.

The civil unrest that occurred throughout Hong Kong at the end of 2019 could have potentially caused some restaurants to close prematurely. However, the data of which restaurants closed due to the unrest or other reasons were not available; all restaurants that closed were treated the same.

5.9. Conclusion

The discussion has concluded that the risk-averse restaurateur and the more significant upfront capital expenses are why restaurants in Hong Kong have a greater success rate over previous studies on restaurant survival. The two research questions posed in Section 1.4 have been answered, with the mean household size and locating in the shopping mall being the most critical aspects for the area attribute and site characteristic variables. The seven hypotheses laid out in the literature review have been discussed, and the connections to previous studies have been linked.

The ANN Model Accuracy Test introduces five new restaurants and has shown that the trained ANN model can predict the success and failure of restaurant sites in a real-world setting. The contributions to the three existing theories dealing with site selection elements were presented and further explaining how the research benefits the practical side of the industry. The limitations, including both the omission of specific variables and the longitudinal study's shortness, were explained.

Chapter 6. Conclusion

When restaurants fail, there are two avenues of inquiry to understand the root cause; decisions made before opening and decisions made post-opening, during the operation of the restaurant. This study addresses the most significant pre-opening choice, the decision of where to open a restaurant (Church, 2002; Egerton-Thomas, 2005; Fields, 2007; Miller, 2006; Olsen et al., 1998; Pioch & Byrom, 2004). Ghosh and Craig (1983) found that the location decision for a restaurant is the single most significant determinant of success. The location decision is critical as the short-term success is vital for independent restaurants as they need to have a cash inflow in a relatively short time frame to ensure available working capital. Additionally, all post-opening decisions are influenced by the restaurant's initial site location.

To understand the variables that are more likely to lead to restaurant success, the researcher looked to scientific models to analyze industry data from both successful and failed restaurants. Hong Kong is known as a dining city, with residents spending, on average, over 327 million Hong Kong Dollars per day at restaurants in 2018 (Census and Statistics Department, 2019). The Census and Statistics Department (2016a) reported in the annual food away from home expenditures of Hong Kong residents, that dining at restaurants accounted for more than 16 percent of their overall expenses. Hong Kong residents have a significant expenditure at restaurants due to the small domestic household size and having the longest average work week of any country, over 50 hours per week (CBRE, 2016); as such, Hong Kong was deemed to be an appropriate city to conduct the study.

The study focuses on 6,710 newly opened restaurants in the years 2016, 2017, and 2018. Nineteen individual variables of the area attributes and site characteristics were combined to answer the research questions and hypotheses. To answer the first research question and the six hypotheses, a logistic regression model was implemented, and the marginal effects were investigated. The study uses two prediction models, logistic regression and ANN, to answer the second research question of whether an algorithm could be constructed to aid in independent restaurant site selection. The model that had both greater model accuracy and area under the ROC curve and deemed the ideal model for predicting the success or failure of a potential restaurant site is the ANN.

Hong Kong had a smaller failure rate than in previous restaurant survival studies in other markets. Having a three-year failure rate of 27.88 percent of restaurants that opened in 2016-18; compared to Columbus, Ohio that had a 59.74 percent failure rate of restaurants that opened in 1996 and monitored for three years (Parsa et al., 2005). Hong Kong restaurants had a failure rate of 10.76 percent in the first 365 days post opening, increasing to 21.59 percent in the next 365-day period, and ending with 27.88 precent at the end of 1,095 days. The failure rate of restaurants in Hong Kong is lower than in previous restaurant studies, arguably due to the higher capital requirements, restaurateurs are more risk-averse than in other areas where the potential losses are decreased.

The logistic regression model found that of the variables, expanded to twenty-four with the inclusion of the dummy variables, seventeen significantly affect the probability of restaurant success. The model found that ten variables, Total Population, Mean Household Size, Percentage of Owner/Occupied Flats, Price Competitors, Convenience Competitors, Aggregate Review, Located on the Ground Floor, Inside a 5 Minute Walk to the MTR Exit, Inside a Mall, and Price HK$101-200, were highly significant at a 99 percent confidence level. Four variables, Median Age, Median Household Income, Percentage of Chinese Ethnicity, and Price – HK$51-100, were significant at a 95 percent confidence level. Three variables, Percent Participating in the Labor Force, Cuisine – Asian, and Price – Above $201, were moderately significant at a 90 percent confidence level.

The first research question that was presented, *which socio-demographic attributes and site characteristics have the highest impact on restaurant success or failure*, examined which variables were crucial to restaurant site success. The most significant implications for the area attributes were the number of individuals living within a household, and for the site characteristics, locating within malls. Locating within a mall provides the most significant increase in potential success, more than any of the other seventeen significant variables.

The logistic regression and ANN models answered the second research question, *considering the potential influence of the overall site characteristics, can a model be created to aid in restaurant site selection*. The findings indicate that it is not only feasible but highly recommended to develop a model assisting in independent restaurant site selection. The logistic regression model had an accuracy of 71.27 percent and the ANN, 72.55 percent While the post-opening aspects are critical to restaurant success, the pre-opening site location decision is essential, more than previously thought.

Two of the six hypotheses were supported; although four were ultimately not supported, three of them, the residential density, income, household age, were all found to be significant. The impact of these three variables on restaurant success is different than hypothesized. The residential density is graphed in line with the hockey stick effect, that while there is a slight increase of limited residents due to the residential day-trippers or weekend travel, the expansion of potential success grows significantly when approaching residentially denser areas.

Previous research suggests that a rise in income equates to greater food away from the home expenditure (Byrne et al., 1998; Dong et al., 2000; Frisbee & Madeira, 1986; Sunny Ham et al., 2004; McCracken & Brandt, 1987; Nayga Jr & Capps Jr, 1993, 1994; S.-J. Yang & Magrabi, 1989). The findings showed an initial correlation with increased income equating to increased food away from home expenditure. However, the overall results found that those that have a higher median household income of over HK$57,031.530 will start to dine at either higher priced establishments, have a domestic helper prepare dinner at their home, or dine away from their place of residence, potentially due to having private cars, being members of private clubs, or dining in a business district rather than near home.

The age of surrounding residents was found to affect the potential success positively (Harrington et al., 2011; Nayga Jr & Capps Jr, 1994; Redman, 1980); the older the mean age of the surrounding population, the greater the potential for success. However, there is not enough spread in the data to understand whether, at a particular age, below 35.636 or above 55.900, the anticipated lower expenditure on food away from home would occur. While the literature found that both younger families (Sunny Ham et al., 2004; Parsa et al., 2015) and retirees (Liu et al., 2013; Nayga Jr & Capps Jr, 1993), have a decreased expenditure on food away from home, this could not be collaborated.

The hypothesis related to restaurant density was ultimately not supported. Although two of the three competitor categories, price, and convenience, were significant, they were in an upward arc and not in the inverse U-shape that was hypothesized. The upward arc-shape of the price competitors provides support for theory the that clustering similarly priced establishments increases the potential success. The third type, the cuisine competitors, was not a significant factor in restaurant success.

The remaining two hypotheses examining the household size and distance to the MTR exit, were both supported by the logistic regression modeling. The study found that when it comes to mean household size, for each additional person living in the household, there was an increase in the probability of success by 10.098 percent. The site characteristic of distance to the MTR station MTR found that a site within a five-minute walk increases the potential of restaurant success by 5.361 percent.

The Discussion chapter laid out the ideal customer and site location, noting that out of the nineteen ideal TPUs across the different categories, the region of Kowloon contained eleven of the nineteen. While Kowloon does have the TPUs with highest failure rate of any region, different TPUs provides the greatest increase of potential success, showcasing the importance of site selection. At the same time, both Hong Kong Island and the New Territories had four each of the nineteen ideal TPUs.

An ANN Model Accuracy Test of five new restaurants that were not included in the training model was presented and found that the ANN correctly predicted, even within a one-year period, the success or failure of the site in four out of five cases. The fifth restaurant, which is successful, the ANN predicted failure with the prediction score of 0.4890, with the cutoff being 0.5. Considering all 188 restaurants that opened in January 2019, the ANN model was able to predict the outcome with an 82.447 percent accuracy, confirming the applicability of the model for predicting site success.

The theoretical and practical contributions were presented within the Discussion chapter. Overall, there were three theories that the study was able to add to, the principle of minimum differential (Hotelling, 1929), the central place theory (King, 1985), and the bid rent theory (Bellini, 2016; Parsa & van der Rest, 2017). The principle of minimum differentiation showed that clustering of restaurants is only essential if the competitor is of a similar price point, or if they are a convenience competitor. Competitors of similar cuisines, no matter the price point, receive no additional benefits from clustering.

The central place theory found that within a dense urban environment, individuals still prefer not to travel far differences. While transportation is ample, restaurant patrons still prefer to dine close to their place of residence and not travel to go to restaurants. The bid rent theory could not be expanded past central business districts. The tourist areas showed that they were not a significant factor in restaurant success; they performed worse than restaurants not in tourist areas.

The five groups that will benefit from the findings of the study are restaurateurs, banks, investors, landlords, and consultants. As the research shows that it is feasible to model potential site success for independent restaurants over 70 percent of the time; individuals that take part in either funding or site selecting can receive an added benefit in understanding the potential of the site before the final selection.

The limitations of the study included the exclusion of five variables that were not included either due to being considered non-factor variables or as the data being unavailable. The civil unrest that occurred in Hong Kong at the end of 2019 could have potentially impacted the success or failure of restaurants; however, separating the restaurants that closed due to the unrest or due to the ordinary course of business was not feasible.

The research has shown that it is possible to utilize the variables from already established independent restaurants to create an algorithm aiding in the site selection process of new entrants. Future research may expand on these findings.

6.1. Future Research

Future studies may investigate if the reliability of the ANN model is feasible in cities similar to Hong Kong, or how it can be modified to include location-specific variables. As mentioned in Section 1.5, London, Singapore, Toronto, Paris, Amsterdam, New York, Stockholm, San Francisco, and Sydney are similar to Hong Kong by PricewaterhouseCoopers most recent list of The Cities of Opportunity (7^{th} edition) (PwC, 2016). Conducting secondary research in one of these cities will be able to expand the ability of the ANN model outside of Hong Kong.

Secondary potential studies may also include other types of hospitality establishments and how they could use a similar algorithm, customized for their clienteles, to select a site to operate. Establishments such as hotels, bars, night clubs, and other entertainment venues could benefit from having an expanded algorithm that includes both the area attributes and site characteristics.

There are seventeen variables that the research found significant in the current site selection process; however, the significant variables can potentially change over time. For instance, the COVID-19 pandemic could place a greater degree of importance in sites that allow for either socially distant dining or sites that have a smaller indoor footprint but allow for a greater emphasis on takeaway, delivery, or outdoor dining. As long as there is continual input into the algorithm, it can learn over time what characteristics are rising to prominence.

6.2. Concluding Remark

The research outlines the importance of site selection and that the mantra of 'location, location, location' is critically important to restaurants. Restaurateurs cannot rely solely on having a differentiated concept, low fixed costs, quality food and service, an experienced manager, chef, or servers, a strong pricing strategy, or adequate starting capital. The location is a crucial ingredient in the overall success and failure of a restaurant, and the ANN model created will aid all stakeholders in the restaurant industry in selecting the ideal site for their endeavor.

Site selection algorithms are no longer only for large corporations that rely on their proprietary historical database in deciding where to open their next shop or franchise. The benefits of creating relevant site selection models can, and should, be expanded to all small business owners within the hospitality industry, retail, and customer forward shops. The use of industry-specific site selection algorithms has the overwhelming potential to avoid losses and enhance the probability of long-term success.

References

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance, 23*(4), 589-609.

Anandarajan, M., Lee, P., & Anandarajan, A. (2001). Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. *Intelligent Systems in Accounting, Finance & Management, 10*(2), 69-81.

Anderson, D., & Burnham, K. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*.

Arduser, L. (2003). *Restaurant site location: Finding, negotiating & securing the best food service site for maximum profit* (Vol. 1): Atlantic Publishing Company.

Atash, F. (1994). Redesigning suburbia for walking and transit: emerging concepts. *Journal of Urban Planning and Development, 120*(1), 48-57.

Athey, S., Blei, D., Donnelly, R., Ruiz, F., & Schmidt, T. (2018). *Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data.* Paper presented at the AEA Papers and Proceedings.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111.

Bellini, J. (2016). The No. 1 thing to consider before opening a restaurant. Retrieved from https://www.cnbc.com/2016/01/20/heres-the-real-reason-why-most-restaurants-fail.html

Bennison, D., Clarke, I., & Pal, J. (1995). Locational decision making in retailing: an exploratory framework for analysis. *International Review of Retail, Distribution and Consumer Research, 5*(1), 1-20.

Bester, H. (1998). Quality uncertainty mitigates product differentiation. *The RAND Journal of Economics*, 828-844.

Binkley, J. K. (2006). The Effect of Demographic, Economic, and Nutrition Factors on the Frequency of Food Away from Home. *Journal of Consumer Affairs, 40*(2), 372-391. doi:10.1111/j.1745-6606.2006.00062.x

Bishop, C. M. (1995). *Neural networks for pattern recognition*: Oxford university press.

Bondarenko, A., Borisov, A., & Aleksejeva, L. (2015). *Neurons vs weights pruning in artificial neural networks.* Paper presented at the Proceedings of the 10th International Scientific and Practical Conference. Volume III.

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Belmont, CA: Duxbury.

Bowlby, S., Breheny, M., & Foot, D. (1984). Store location: problems and methods 1: is locating a viable store becoming more difficult? *Retail and Distribution Management, 12*(5), 31-33.

Brandenburger, A. M., & Nalebuff, B. J. (2011). *Co-opetition*: Crown Business.

Briesch, R. A., Chintagunta, P. K., & Fox, E. J. (2009). How does assortment affect grocery store choice?(Report). *Journal of Marketing Research, 46*(2), 176. doi:10.1509/jmkr.46.2.176

Brown, S. (1991). Shopper circulation in a planned shopping centre. *International Journal of Retail & Distribution Management, 19*(1).

Brown, S. (1992). Tenant mix, tenant placement and shopper behaviour in a planned shopping centre. *Service Industries Journal, 12*(3), 384-403.

Brown, S. (1993). Retail location theory: evolution and evaluation. *International Review of Retail, Distribution and Consumer Research, 3*(2), 185-229.

Brunner, T. A., Van der Horst, K., & Siegrist, M. (2010). Convenience food products. Drivers for consumption. *Appetite, 55*(3), 498-506.

Burnaz, S., & Topcu, Y. I. (2006). A multiple-criteria decision-making approach for the evaluation of retail location. *Journal of Multi-Criteria Decision Analysis, 14*(1-3), 67-76.

Byrne, P. J., Capps, O., & Saha, A. (1998). Analysis of quick-serve, mid-scale, and up-scale food away from home expenditures. *International Food and Agribusiness Management Review, 1*(1), 51-72. doi:10.1016/S1096-7508(99)80028-7

Camillo, A. A., Connolly, D. J., & Kim, W. G. (2008). Success and failure in Northern California: Critical success factors for independent restaurants. *Cornell Hospitality Quarterly, 49*(4), 364-380.

Candel, M. J. (2001). Consumers' convenience orientation towards meal preparation: conceptualization and measurement. *Appetite, 36*(1), 15-28.

CBRE. (2016). *Hunger for Growth - Unlocking Opportunities in Hong Kong Food & Beverage Retail*. Retrieved from https://www.cbre.com.hk/zh-hk/global/research-and-reports/apac-unlocking-opportunities-in-hong-kong-food-and-beverage-retail

Cella, F. R. (1968). Computer evaluation of restaurant sites. *Cornell Hotel and Restaurant Administration Quarterly, 9*(3), 25-31.

Census and Statistics Department. (2016a). *2014/15 Household Expenditure Survey and the Rebasing of the Consumer Price Indices*. Retrieved from https://www.statistics.gov.hk/pub/B10600082015XXXXB0100.pdf

Census and Statistics Department. (2016b). *Results of the 2014/15 Household Expenditure Survey*. Retrieved from https://www.statistics.gov.hk/pub/B71608FB2016XXXXB0100.pdf

Census and Statistics Department. (2019). Report on Quarterly Survey of Restaurant Receipts and Purchases [Press release]. Retrieved from https://www.statistics.gov.hk/pub/B10800022019QQ01B0100.pdf

Chen, L.-F., & Tsai, C.-T. (2016). Data mining framework based on rough set theory to improve location selection decisions: A case study of a restaurant chain. *Tourism Management, 53*, 197-206.

Cheng, J. L. (n.d.). What were the changes in consumption choices among Hong Kong’s young households in recent years? Retrieved from https://www.statistics.gov.hk/wsc/CPS006-P3-S.pdf

Church, R. L. (2002). Geographical information systems and location science. *Computers and Operations Research, 29*(6), 541-562. doi:10.1016/S0305-0548(99)00104-5

Civil Engineering and Development Department. (2019). *Land Usage Distribution in Hong Kong*. Retrieved from https://www.cedd.gov.hk/filemanager/eng/content_954/Info_Sheet2.pdf

Clarke, I., Horita, M., & Mackaness, W. (2000). The spatial knowledge of retail decision makers: capturing and interpreting group insight using a composite cognitive map. *The International Review of Retail, Distribution and Consumer Research, 10*(3), 265-285.

Clements, K. W., & Si, J. (2017). Engel’s law, diet diversity, and the quality of food consumption. *American Journal of Agricultural Economics, 100*(1), 1-22.

Craig, C., Ghosh, A., & McLafferty, S. (1984). Models of the Retail Location Process: A Review. *Journal of Retailing, 60*(1), 5.

Darian, J. C., & Cohen, J. (1995). Segmenting by consumer time shortage. *Journal of Consumer Marketing, 12*(1), 32-44.

Darley, R. D., & Gobar, A. J. (1969). Restaurant Site Selection: A Properly Designed Computer Model Can Pinpoint Most Profitable Sites and Provide Useful Operating Guides. *Cornell Hotel and Restaurant Administration Quarterly, 10*(3), 61-69.

De Andrés, J., Landajo, M., & Lorca, P. (2005). Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case. *European journal of operational research, 167*(2), 518-542.

Denton, J. W., Hung, M. S., & Osyk, B. A. (1990). A neural network approach to the classification problem. *Expert Systems with Applications, 1*(4), 417-424.

Department of Health. (2017). *Report of Population Health Survey 2014/2015*. Retrieved from Surveillance and Epidemiology Branch: https://www.chp.gov.hk/files/pdf/dh_hps_2014_15_full_report_eng.pdf

Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European journal of operational research, 90*(3), 487-513.

Discover Hong Kong. (n.d.). Top 10 Attractions. Retrieved from http://www.discoverhongkong.com/eng/see-do/highlight-attractions/top-10/index.jsp

Dock, J. P., Song, W., & Lu, J. (2015). Evaluation of dine-in restaurant location and competitiveness: Applications of gravity modeling in Jefferson County, Kentucky. *Applied Geography, 60*, 204-209.

Dong, D., Byrne, P. J., Saha, A., & Capps, O. (2000). Determinants of Food-Away-From-Home (FAFH) Visit Frequency: A Count-Data Approach. *Journal of Restaurant & Foodservice Marketing, 4*(1), 31-46. doi:10.1300/J061v04n01_04

Echeverri-Carroll, E. L., & Ayala, S. G. (2011). Urban wages: Does city size matter? *Urban Studies, 48*(2), 253-271.

Egerton-Thomas, C. (2005). *How to open and run a successful restaurant*: John Wiley & Sons.

Elgin, J. (2004). Choosing a Great Franchise Location. Retrieved from https://www.entrepreneur.com/article/71508

English, W. (1996). Restaurant attrition: a longitudinal analysis of restaurant failures. *International Journal of Contemporary Hospitality Management*.

Everett, B. G., Rehkopf, D. H., & Rogers, R. G. (2013). The nonlinear relationship between education and mortality: an examination of cohort, race/ethnic, and gender differences. *Population research and policy review, 32*(6), 893-917.

Everett, J., & Watson, J. (1998). Small business failure and external risk factors. *Small Business Economics, 11*(4), 371-390.

Fang, L., Li, H., & Li, M. (2019). Does hotel location tell a true story? Evidence from geographically weighted regression analysis of hotels in Hong Kong. *Tourism Management, 72*, 78-91.

Feloni, R. (2014). Food Network Chef Robert Irvine Shares The Top 5 Reasons Restaurants Fail. Retrieved from https://www.businessinsider.com/why-restaurants-fail-so-often-2014-2

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: sage.

Fields, R. (2007). *Restaurant Success by the Numbers: A Money-guy's Guide to Opening the Next New Hot Spot*: Ten Speed Press.

Fischer, J. H., & Harrington Jr, J. E. (1996). Product variety and firm agglomeration. *The RAND Journal of Economics*, 281-309.

Fish, K. E., Barnes, J. H., & Aiken, M. W. (1995). Artificial neural networks: a new methodology for industrial market segmentation. *Industrial Marketing Management, 24*(5), 431-438.

Fisher, D. P. (1997). Location, Location, Location: Ensuring a Franchisee's Success. *Hospitality Review, 15*(1), 4.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information & Management, 24*(3), 159-167.

Food and Environmental Hygiene Department. (2018). Guide on Types of Licences Required. Retrieved from https://www.fehd.gov.hk/english/licensing/Guide_on_Types_of_Licences_Required.html

Fox, E. J., Postrel, S., & McLaughlin, A. (2007). The impact of retail location on retailer revenues: An Empirical investigation. *Unpublished manuscript, Edwin L. Cox School of Business, Southern Methodist University, Dallas, TX*.

Frisbee, W. R., & Madeira, K. (1986). Restaurant meals-convenience goods or luxuries? *The Service Industries Journal, 6*(2), 172-192.

Ghosh, A., & Craig, C. S. (1983). Formulating retail location strategy in a changing environment. *Journal of Marketing, 47*(3), 56-68.

Ghosh, A., & MacLafferty, S. L. (1987). *Location Strategies for Retail and Service Firms*: Lexington Books.

Glorot, X., & Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks.* Paper presented at the Proceedings of the thirteenth international conference on artificial intelligence and statistics.

Goldberg, E. (n.d.). Site Selection: Choosing the Best Location. Retrieved from https://www.franchising.com/guides/site_selection_choosing_the_best_location.html

Gonzlez-Benito, S., & Gonzlez-Benito, J. (2005). The role of geodemographic segmentation in retail location strategy. *International Journal of Market Research, 47*(3), 295-316.

Gordinier, J. (2015). The Latest Wave of New York's Food Districts. Retrieved from https://www.nytimes.com/interactive/2015/12/16/dining/food-new-york-neighborhoods.html

Goss, E. P., & Ramchandani, H. (1995). Comparing classification accuracy of neural networks, binary logit regression and discriminant analysis for insolvency prediction of life insurers. *Journal of Economics and Finance, 19*(3), 1.

Graham, B., & Paul, C. (2010). Does higher education really lead to higher employability and wages in the RMI. *US Census Bureau Report*.

Gu, Z. (2002). Analyzing bankruptcy in the restaurant industry: A multiple discriminant model. *International Journal of Hospitality Management, 21*(1), 25-42.

Guerriero, F., Miglionico, G., & Olivito, F. (2014). Strategic and operational decisions in restaurant revenue management. *European journal of operational research, 237*(3), 1119-1132.

Guest, A. M., & Cluett, C. (1974). Metropolitan retail nucleation. *Demography, 11*(3), 493-507.

Guy, C. (1994). Grocery store saturation: Has it arrived yet? *International Journal of Retail & Distribution Management, 22*(1), 3-11.

Guy, C. (1995). Retail store development at the margin. *Journal of Retailing and Consumer services, 2*(1), 25-32.

Gyaan, R. (n.d.). Top 10 Reasons Why Restaurants Fail Within the First Year of Operations. Retrieved from https://www.posist.com/restaurant-times/restro-gyaan/top-10-reasons-restaurants-fail.html

Ham, S., Hiemstra, S. J., & Yang, I.-S. (1998). Modeling US household expenditure on food away from home (FAFH): Logit regression analysis. *Journal of Hospitality & Tourism Research, 22*(1), 15-24.

Ham, S., Hwang, J. H., & Kim, W. G. (2004). Household profiles affecting food-away-from-home expenditures: a comparison of Korean and US households. *International Journal of Hospitality Management, 23*(4), 363-379.

Hannan, M. T., & Carroll, G. R. (1992). *Dynamics of organizational populations: Density, legitimation, and competition*: Oxford University Press.

Harrington, R. J., Ottenbacher, M. C., & Kendall, K. (2011). Fine-dining restaurant selection: Direct and moderating effects of customer attributes. *Journal of foodservice business research, 14*(3), 272-289.

Harrison, J. S. (2003). Strategic analysis for the hospitality industry. *Cornell Hotel and Restaurant Administration Quarterly, 44*(2), 139-152.

Healy, J., & Mac Con Iomaire, M. (2018). Calculating restaurant failure rates using longitudinal census data. *Journal of Culinary Science & Technology*, 1-23.

Hebb, D. O. (1949). *The organization of behavior*: na.

Henderson, M. A. (2006). Little Italy. In G. M. Bakken & A. Kindell (Eds.), *Encyclopedia of Immigration and Migration in the American West*.

Hernandez, T., & Bennison, D. (2000). The art and science of retail location decisions. *International Journal of Retail & Distribution Management, 28*(8), 357-367.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623-641.

Hong Kong Tourism Board. (2019a). *Monthly Report - Visitor Arrival Statistics : Dec 2018*. Retrieved from https://partnernet.hktb.com/filemanager/intranet/pm/VisitorArrivalStatistics/ViS_Stat_E/VisE_2018/Tourism%20Statistics%2012%202018_R1.pdf

Hong Kong Tourism Board. (2019b). *Visitor Profile Report - 2018*. Retrieved from https://securepartnernet.hktb.com/filemanager/intranet/ir/ResearchStatistics/paper/Visitor-Pro/Profile2018/Visitor%20Profile%202018.pdf

Hotelling, H. (1929). Stability in Competition. *The Economic Journal, 39*(153), 41-57.

Hull, C. L. (1943). Principles of behavior: An introduction to behavior theory.

Jackson, P., & Viehoff, V. (2016). Reframing convenience food. *Appetite, 98*, 1-11.

Jain, A. K., Mao, J., & Mohiuddin, K. (1996). Artificial neural networks: A tutorial. *Computer*(3), 31-44.

Jang, D., & Mattila, A. S. (2005). An examination of restaurant loyalty programs: what kinds of rewards do customers prefer? *International Journal of Contemporary Hospitality Management, 17*(5), 402-408.

Jekanowski, M. D., Binkley, J. K., & Eales, J. S. (2001). Convenience, accessibility, and the demand for fast food. *Journal of Agricultural and Resource Economics, 26*(1835-2016-148749), 58-74.

Jogaratnam, G. (2002). Entrepreneurial orientation and environmental hostility: an assessment of small, independent restaurant businesses. *Journal of Hospitality & Tourism Research, 26*(3), 258-277.

Johnston, R., & Kissling, C. (1971). Establishment use patterns within central places. *Australian Geographical Studies, 9*(2), 116-132.

Jung, S. S., & Jang, S. S. (2019). To cluster or not to cluster?: Understanding geographic clustering by restaurant segment. *International Journal of Hospitality Management, 77*, 448-457.

Kahn, B. E., & McAlister, L. (1997). *Grocery revolution*: Addison-Wesley.

Kamps, J., & Péli, G. (1995). Qualitative reasoning beyond the physics domain: The density dependence theory of organizational ecology. *Proceedings of QR95*, 114-122.

Karakaya, F., & Canel, C. (1998). Underlying dimensions of business location decisions. *Industrial management & data systems, 98*(7), 321-329.

Kerr, W. R., & Mandorff, M. (2015). *Social networks, ethnicity, and entrepreneurship*. Retrieved from

Kim, E.-J., & Geistfeld, L. V. (2003). Consumers' Restaurant Choice Behavior and the Impact of Socio-Economic and Demographic Factors. *Journal of foodservice business research, 6*(1), 3-24. doi:10.1300/J369v06n01_02

Kim, H., & Gu, Z. (2006a). A logistic regression analysis for predicting bankruptcy in the hospitality industry. *The Journal of Hospitality Financial Management, 14*(1), 17-34.

Kim, H., & Gu, Z. (2006b). Predicting restaurant bankruptcy: A logit model in comparison with a discriminant model. *Journal of Hospitality & Tourism Research, 30*(4), 474-493.

Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal, 31*(3), 441-468.

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling, 36*, 354-362.

Kimes, S. E., & Fitzsimmons, J. A. (1990). Selecting profitable hotel sites at La Quinta motor inns. *Interfaces, 20*(2), 12-20.

King, L. J. (1985). Central place theory. *Regional Research Institute, West Virginia University Book Chapters*, 1-52.

Kivela, J. (1997). Restaurant marketing: selection and segmentation in Hong Kong. *International Journal of Contemporary Hospitality Management, 9*(3), 116-123.

Kivela, J., Inbakaran, R., & Reece, J. (2000). Consumer research in the restaurant environment. Part 3: analysis, findings and conclusions. *International Journal of Contemporary Hospitality Management, 12*(1), 13-30.

Kivell, P. T., & Shaw, G. (1980). The study of retail location. *Retail Geography. RLE Retailing and Distribution: New York*.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Paper presented at the Ijcai.

Konishi, H. (2005). Concentration of competing retail stores. *Journal of Urban economics, 58*(3), 488-512.

Koo, L., Tao, F. K., & Yeung, J. H. (1999). Preferential segmentation of restaurant attributes through conjoint analysis. *International Journal of Contemporary Hospitality Management, 11*(5), 242-253.

Kotler, P., Bowen, J. T., Makens, J., & Baloglu, S. (2017). Marketing for hospitality and tourism.

Kreft, I. G. (1995). The Effects of Centering in Multilevel Analysis: Is the Public School the Loser or the Winner?

Krider, R. E., & Putler, D. S. (2013). Which birds of a feather flock together? Clustering and avoidance patterns of similar retail outlets. *Geographical Analysis, 45*(2), 123-149.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research, 180*(1), 1-28.

Lani, J. (n.d.). Assumptions of Logistic Regression. Retrieved from https://www.statisticssolutions.com/assumptions-of-logistic-regression/

Lam, W. H., & Cheung, C.-y. (2000). Pedestrian speed/flow relationships for walking facilities in Hong Kong. *Journal of transportation engineering, 126*(4), 343-349.

Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of economics and Business, 51*(4), 347-364.

Letail, R. (1992). Site attributes in retail leasing: an analysis of a fast-food restaurant market. *The Appraisal Journal*.

Litz, R. A., & Rajaguru, G. (2008). Does small store location matter? A test of three classic theories of retail location. *Journal of Small Business & Entrepreneurship, 21*(4), 477-492.

Liu, M., Kasteridis, P., & Yen, S. T. (2013). Who are consuming food away from home and where? Results from the Consumer Expenditure Surveys. *European Review of Agricultural Economics, 40*(1), 191-213. doi:10.1093/erae/jbs012

Lo, H. K., Tang, S., & Wang, D. Z. W. (2008). Managing the accessibility on mass public transit

The case of Hong Kong. *Journal of Transport and Land Use, 1*(2), 23-49. Retrieved from www.jstor.org/stable/26201613

Longstreth, R. W. (1997). *City center to regional mall : architecture, the automobile, and retailing in Los Angeles, 1920-1950*. Cambridge, Mass., London, England: MIT Press.

Lorch, B. J., & Smith, M. J. (1993). Pedestrian movement and the downtown enclosed shopping center. *Journal of the American Planning Association, 59*(1), 75-86.

Luo, T., & Stark, P. B. (2015). Nine out of 10 restaurants fail? Check, please. *Significance, 12*(2), 25-29.

Lupton, D. (2000). The heart of the meal: food preferences and habits among rural Australian couples. *Sociology of Health & Illness, 22*(1), 94-109.

Man, D., & Fullerton, E. (1990). Single drop depositors–an aid to consistent production of chilled ready meals. In R. W. Field & J. A. Howell (Eds.), *Process Engineering in the Food Industry-2: Convenience Foods and Quality Assurance* (pp. 75-87). London.

Marcora, S. (2016). Can doping be a good thing? Using psychoactive drugs to facilitate physical activity behaviour. In: Springer.

Mazze, E. M. (1972). Identifying the key factors in retail store location. *Journal of Small Business Management, 10*(1), 17-20.

Mazzocchi, M. (2008). *Statistics for marketing and consumer research*. London: Sage.

McCracken, V. A., & Brandt, J. A. (1987). Household consumption of food-away-from-home: total expenditure and by type of food facility. *American Journal of Agricultural Economics, 69*(2), 274-284.

McCracken, V. A., & Brandt, J. A. (1990). Time value and its impact on household food expenditures away from home. *Home Economics Research Journal, 18*(4), 267-285.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5*(4), 115-133.

McDonald, R. (2018). The geography of demand for residential and commercial space. Retrieved from https://www.centreforcities.org/reader/city-space-race-balancing-need-homes-offices-cities/geography-demand-residential-commercial-space/

McNelis, P. D. (2005). *Neural networks in finance: gaining predictive edge in the market*: Academic Press.

Mendelsohn, M. (1985). *The guide to franchising* (4th ed.). Oxford [Oxfordshire], New York: Pergamon Press.

Miller, D. (2006). *Starting a Small Restaurant: How to Make Your Dream a Reality*: Houghton Mifflin Harcourt.

Mun, S. G., & Jang, S. S. (2015). Working capital, cash holding, and profitability of restaurant firms. *International Journal of Hospitality Management, 48*, 1-11.

Munish, M. (2017). Most expensive retail property in Asia? Hong Kong's Causeway Bay. *SCMP*. Retrieved from https://www.scmp.com/magazines/style/news-trends/article/2122067/most-expensive-retail-property-asia-hong-kongs-causeway

Mutlu, S., & Gracia, A. (2006). Spanish food expenditure away from home (FAFH): by type of meal. *Applied Economics, 38*(9), 1037-1047.

Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2): Duxbury press Belmont, CA.

Nayga Jr, R. M., & Capps Jr, O. (1993). Analysis of socio-economic and demographic factors affecting food away from home consumption: A synopsis. *Journal of Food Distribution Research, 24*(856-2016-57672), 69.

Nayga Jr, R. M., & Capps Jr, O. (1994). Impact of socio-economic and demographic factors on food away from home consumption: Number of meals and type of facility. *Journal of Restaurant & Foodservice Marketing, 1*(2), 45-69.

Nelson, R. L. (1958). *The selection of retail locations*: FW Dodge Corporation.

Neophytou, E., & Molinero, C. M. (2004). Predicting corporate failure in the UK: a multidimensional scaling approach. *Journal of Business Finance & Accounting, 31*(5-6), 677-710.

Ng, J. (2018). LCQ15: Improving average living floor area per person [Press release]. Retrieved from https://www.info.gov.hk/gia/general/201806/20/P2018062000367.htm

Nickell, S. (1996). Competition and corporate performance. *Journal of political economy, 104*(4), 724-746.

Nickell, S., Nicolitsas, D., & Dryden, N. (1997). What makes firms perform well? *European economic review, 41*(3-5), 783-796.

Njite, D., Dunn, G., & Hyunjung Kim, L. (2008). Beyond good food: what other attributes influence consumer preference and selection of fine dining restaurants? *Journal of foodservice business research, 11*(2), 237-266.

Nwogugu, M. (2006). Site selection in the US retailing industry. *Applied mathematics and computation, 182*(2), 1725-1734.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.

Olsen, M. D., Ching-Yick, T., & West, J. J. (1998). *Strategic management in the hospitality industry*: John Wiley and Sons.

Ozdemir, B., & Caliskan, O. (2014). A review of literature on restaurant menus: Specifying the managerial issues. *International Journal of gastronomy and food science, 2*(1), 3-13.

Palmer, A., Montaño, J. J., & Franconetti, F. (2008). Sensitivity analysis applied to artificial neural networks for forecasting time series. *Methodology, 4*(2), 80-86.

Palmer, A., Montaño, J. J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781-790.

Park, K., & Khan, M. A. (2006). An exploratory study to identify the site selection factors for US franchise restaurants. *Journal of foodservice business research, 8*(1), 97-114.

Park, S.-S., & Hancer, M. (2012). A comparative study of logit and artificial neural networks in predicting bankruptcy in the hospitality industry. *Tourism Economics, 18*(2), 311-338.

Parsa, H., Gregory, A., & Terry, M. Reasons Restaurants Fail.

Parsa, H., Gregory, A., & Terry, M. (2011). Why do restaurants fail? Part III: An analysis of macro and micro factors.

Parsa, H., Kreeger, J. C., van der Rest, J.-P., Xie, L. K., & Lamb, J. (2019). Why Restaurants Fail? Part V: Role of Economic Factors, Risk, Density, Location, Cuisine, Health Code Violations and GIS Factors. *International Journal of Hospitality & Tourism Administration*, 1-26.

Parsa, H., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why restaurants fail? Part II-The impact of affiliation, location, and size on restaurant failures: Results from a survival analysis. *Journal of foodservice business research, 14*(4), 360-379.

Parsa, H., Self, J. T., Njite, D., & King, T. (2005). Why restaurants fail. *Cornell Hotel and Restaurant Administration Quarterly, 46*(3), 304-322.

Parsa, H., & van der Rest, J.-P. (2017). Business Failure in the U.S. Restaurant Industry. In J. Adriaanse & J.-P. van Der Rest (Eds.), *Turnaround Management and Bankruptcy: A Research Companion*.

Parsa, H., van der Rest, J.-P. I., Smith, S. R., Parsa, R. A., & Bujisic, M. (2015). Why restaurants fail? Part IV: The relationship between restaurant failures and demographic factors. *Cornell Hospitality Quarterly, 56*(1), 80-90.

Pashigian, B. P., & Gould, E. D. (1998). Internalizing Externalities: the Pricing of Space in Shopping Malls 1. *The Journal of Law & Economics, 41*(1), 115-142. doi:10.1086/467386

Perry, M. (2016). *New US homes today are 1,000 square feet larger than in 1973 and living space per person has nearly doubled*. Retrieved from http://www.aei.org/publication/new-us-homes-today-are-1000-square-feet-larger-than-in-1973-and-living-space-per-person-has-nearly-doubled/

Piana, V. (2003). Product Differentiation. Retrieved from http://economicswebinstitute.org/glossary/product.htm

Pillsbury, R. (1987). From Hamburger Alley to Hedgerose Heights: Toward a Model of Restaurant Location Dynamics. *The Professional Geographer, 39*(3), 326-344. doi:10.1111/j.0033-0124.1987.00326.x

Pioch, E., & Byrom, J. (2004). Small independent retail firms and locational decision-making: outdoor leisure retailing by the crags. *Journal of Small Business and Enterprise Development, 11*(2), 222-232.

Planning Department. (2005a). Shopping Survey (Household Survey) [Press release]. Retrieved from https://www.pland.gov.hk/pland_en/p_study/comp_s/shopping/e_household_survey.htm

Planning Department. (2005b). Shopping Survey (Retail Survey) [Press release]. Retrieved from https://www.pland.gov.hk/pland_en/p_study/comp_s/shopping/e_retailer_survey.htm

Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodoloski zvezki, 1*(1), 143.

Pollitt, C. (2003). *The essential public manager*: McGraw-Hill Education (UK).

Poole, R., Clarke, G. P., & Clarke, D. B. (2006). Competition and saturation in West European grocery retailing. *Environment and Planning A, 38*(11), 2129-2156.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly, 14*(1), 15-34.

Prayag, G., Landré, M., & Ryan, C. (2012). Restaurant location in Hamilton, New Zealand: clustering patterns from 1996 to 2008. *International Journal of Contemporary Hospitality Management, 24*(3), 430-450.

PwC. (2016). *Cities of Opportunity 7*. Retrieved from https://www.pwc.com/us/en/cities-of-opportunity/2016/cities-of-opportunity-7-report.pdf

Ray, S. (2015). Simple Methods to deal with Categorical Variables in Predictive Modeling. Retrieved from https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/

Redman, B. J. (1980). The impact of women's time allocation on expenditure for meals away from home and prepared foods. *American Journal of Agricultural Economics, 62*(2), 234-237.

Reimers, V., & Clulow, V. (2004). Retail concentration: a comparison of spatial convenience in shopping strips and shopping centres. *Journal of Retailing and Consumer services, 11*(4), 207-221.

Roig-Tierno, N., Baviera-Puig, A., Buitrago-Vera, J., & Mas-Verdu, F. (2013). The retail site location decision process using GIS and the analytical hierarchy process. *Applied Geography, 40*, 191-198.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Learning representations by back-propagating errors. *Cognitive modeling, 5*(3), 1.

Salmon, F. (2012). Gastronomics: Where the One Percent Eats. *New York Magazine*. Retrieved from http://www.grubstreet.com/2012/01/one-percent-eating-habits-analysis.html

Satran, J. (2017). 18 Countries That Love To Eat, Drink & Smoke More Than The U.S. *The Huffington Post*. Retrieved from https://www.huffpost.com/entry/countries-that-love-to-eat-drink-smoke_n_3880324

Schaefer, A. D., Luke, R. H., & Green, J. (1996). Attitudes of restaurant site selection executives toward various people magnets. *Journal of Restaurant & Foodservice Marketing, 1*(3-4), 1-14.

Scholliers, P. (2015). Convenience foods. What, why, and when. *Appetite, 94*, 2-6.

Seabold, Skipper, & Perktold, J. (2010). *statsmodels: Econometric and statistical modeling with python.* Paper presented at the Proceedings of the 9th Python in Science Conference.

Self, J. T., Jones, M. F., & Botieff, M. (2015). Where restaurants fail: A longitudinal study of micro locations. *Journal of foodservice business research, 18*(4), 328-340.

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.

Sietsma, J., & Dow, R. J. F. (1988). Neural net pruning-why and how. In (pp. 325-333 vol.321).

Singh, J. V. (1993). Density dependence theory-Current issues, future promise. In: University of Chicago Press.

Sinha, P. K., & Banerjee, A. (2004). Store choice behaviour in an evolving market. *International Journal of Retail & Distribution Management, 32*(10), 482-494.

Smith, S. L. (1983). Restaurants and dining out: geography of a tourism business. *Annals of tourism research, 10*(4), 515-549.

Smith, S. L. (1985). Location patterns of urban restaurants. *Annals of tourism research, 12*(4), 581-602.

Storey, D. J., Keasey, K., Wynarczyk, P., & Watson, R. (1987). The Performance of Small Firms: Profits, Jobs and Failures. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Strate, R. W., & Rappole, C. L. (1997). Strategic alliances between hotels and restaurants. *Cornell Hotel and Restaurant Administration Quarterly, 38*(3), 50-61. doi:10.1016/S0010-8804(97)89508-0

Sulek, J. M., & Hensley, R. L. (2004). The relative importance of food, atmosphere, and fairness of wait: The case of a full-service restaurant. *Cornell Hotel and Restaurant Administration Quarterly, 45*(3), 235-247.

Tam, K. Y. (1991). Neural network models and the prediction of bank bankruptcy. *Omega, 19*(5), 429-445.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science, 38*(7), 926-947.

Tan, A. K. (2010). Demand for food-away-from-home in Malaysia: a samples election analysis by ethnicity and gender. *Journal of foodservice business research, 13*(3), 252-267.

Teller, C., & Reutterer, T. (2008). The evolving concept of retail attractiveness: what makes retail agglomerations attractive when customers shop at them? *Journal of Retailing and Consumer services, 15*(3), 127-143.

Thau, B. (2014). How Big Data Helps Chains Like Starbucks Pick Store Locations -- An (Unsung) Key To Retail Success. Retrieved from https://www.forbes.com/sites/barbarathau/2014/04/24/how-big-data-helps-retailers-like-starbucks-pick-store-locations-an-unsung-key-to-retail-success/#5d2847a116db

Thompson, J. S. (1986). *Site selection*. New York: Lebhar-Friedman Books.

Transport Department. (2018). Hong Kong Facts: Transport [Press release]. Retrieved from https://www.gov.hk/en/about/abouthk/factsheets/docs/transport.pdf

Transport Department. (2019). *Vehicles Registration & Licensing*. Retrieved from https://www.td.gov.hk/filemanager/en/content_4883/table41a.pdf

Turiace, B. (2015). *Restaurant Adventures: Think or Sink-Do You Have What it Takes?* : Page Publishing Inc.

Tzeng, G.-H., Teng, M.-H., Chen, J.-J., & Opricovic, S. (2002). Multicriteria selection for a restaurant location in Taipei. *International Journal of Hospitality Management, 21*(2), 171-187.

U.S. Department of Labor. (2018). *Consumer Expenditures in 2017*. Retrieved from https://www.bls.gov/opub/reports/consumer-expenditures/2017/home.htm

Van Witteloostuijn, A., Boin, A., Kofman, C., Kuilman, J., & Kuipers, S. (2018). Explaining the survival of public organizations: Applying density dependence theory to a population of US federal agencies. *Public Administration*, *96*(4), 633-650.

Verlegh, P. W., & Candel, M. J. (1999). The consumption of convenience foods: reference groups and eating situations. *Food Quality and Preference, 10*(6), 457-464.

Warner, R. M. (2012). *Applied statistics: From bivariate through multivariate techniques*. Sage Publications.

Waxman, S. (n.d.). The History of New York City's Chinatown. Retrieved from https://www.ny.com/articles/chinatown.html

Wendt, P. F. (1972). Deciding on location for a small business. *Journal of Small Business Management (pre-1986), 10,* 1.

What You'll See in Hong Kong's Areas. (n.d.). Retrieved from https://www.hongkong.net/guide

Williams, P. (2017). Re: How does Starbucks choose locations for its stores? Message posted to https://www.quora.com/How-does-Starbucks-choose-locations-for-its-stores

Wilson, A. C., Roelofs, R., Stern, M., Srebro, N., & Recht, B. (2017). *The marginal value of adaptive gradient methods in machine learning.* Paper presented at the Advances in Neural Information Processing Systems.

Wood, S., & Browne, S. (2007). Convenience store location planning and forecasting–a practical research agenda. *International Journal of Retail & Distribution Management, 35*(4), 233-255.

Wood, S., & Reynolds, J. (2012). Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. *Geoforum, 43*(6), 1076-1087. doi:10.1016/j.geoforum.2012.06.014

Wood, S., & Tasker, A. (2008). The importance of context in store forecasting: the site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing, 16*(2), 139-155.

Wuensch, K. L. (n.d.). Independent Samples T Tests versus Binary Logistic Regression. Retrieved from http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm

Yang, S.-J., & Magrabi, F. M. (1989). Expenditures for services, wife's employment, and other household characteristics. *Home Economics Research Journal, 18*(2), 133-147.

Yang, Y., & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American journal of preventive medicine, 43*(1), 11-19.

Yang, Y., Roehl, W. S., & Huang, J.-H. (2017). Understanding and projecting the restaurantscape: The influence of neighborhood sociodemographic characteristics on restaurant location. *International Journal of Hospitality Management, 67*, 33-45.

Yang, Y., Tang, J., Luo, H., & Law, R. (2015). Hotel location evaluation: A combination of machine learning tools and web GIS. *International Journal of Hospitality Management, 47*, 14-24.

Yang, Y., Wong, K. K., & Wang, T. (2012). How do hotels choose their location? Evidence from hotels in Beijing. *International Journal of Hospitality Management, 31*(3), 675-685.

Yip, T. C., Chan, K., & Poon, E. (2012). Attributes of young consumers' favorite retail shops: A qualitative study. *Journal of Consumer Marketing, 29*(7), 545-552.

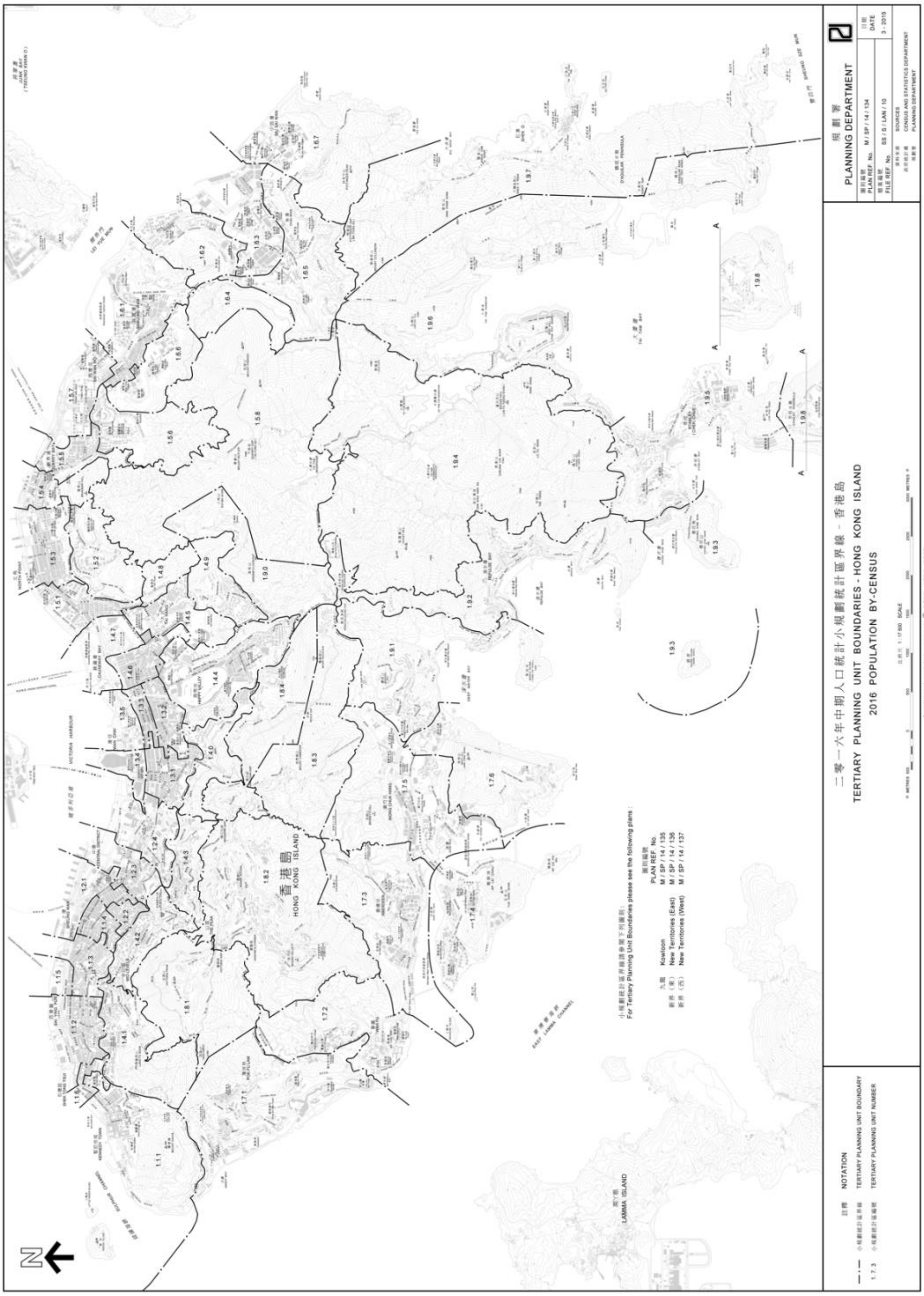
Youn, H., & Gu, Z. (2010). Predict US restaurant firm failures: The artificial neural network model versus logistic regression model. *Tourism and Hospitality Research, 10*(3), 171-187.

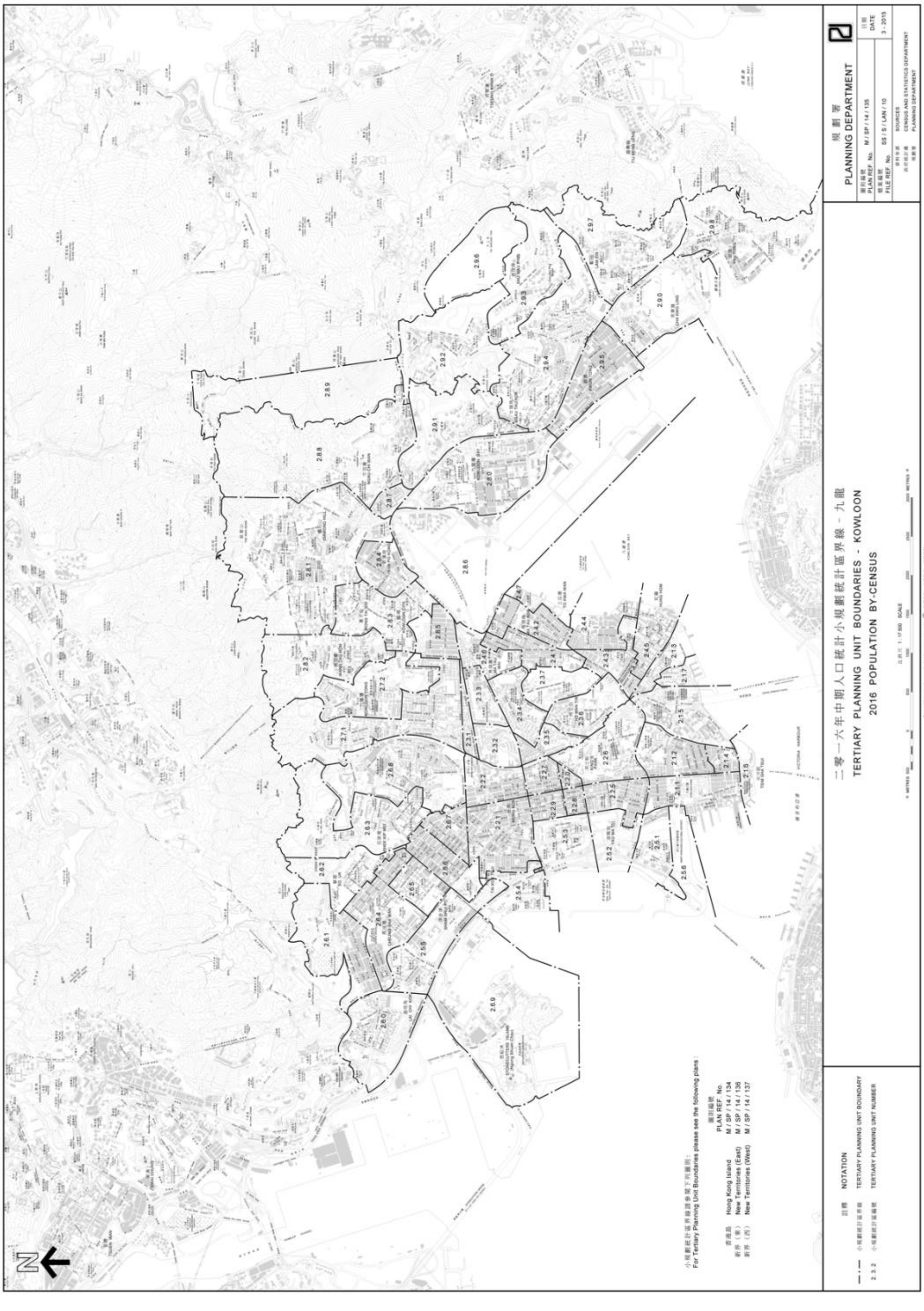
Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European journal of operational research, 116*(1), 16-32.

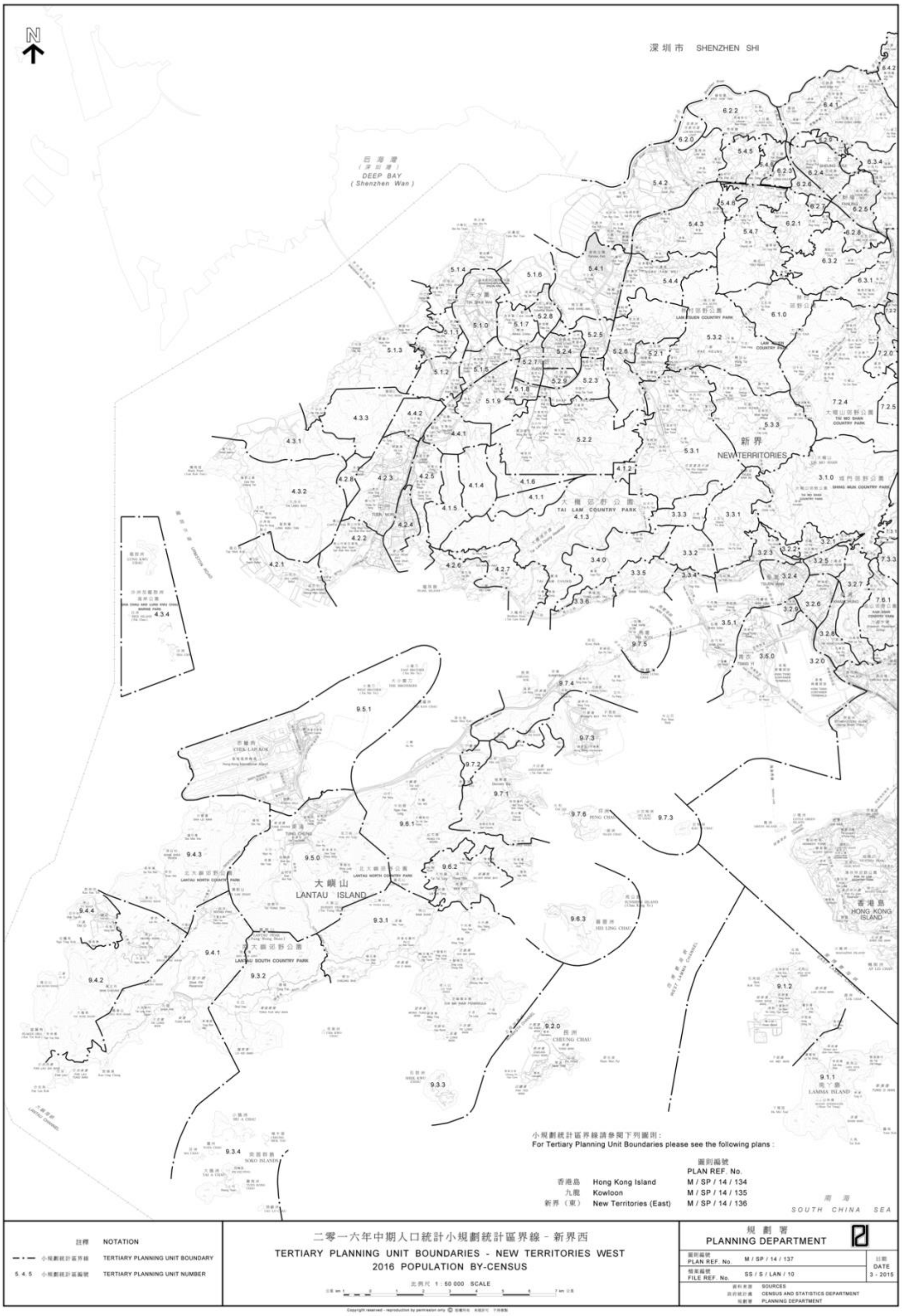
Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management, 29*(4), 694-700. doi:https://doi.org/10.1016/j.ijhm.2010.02.002

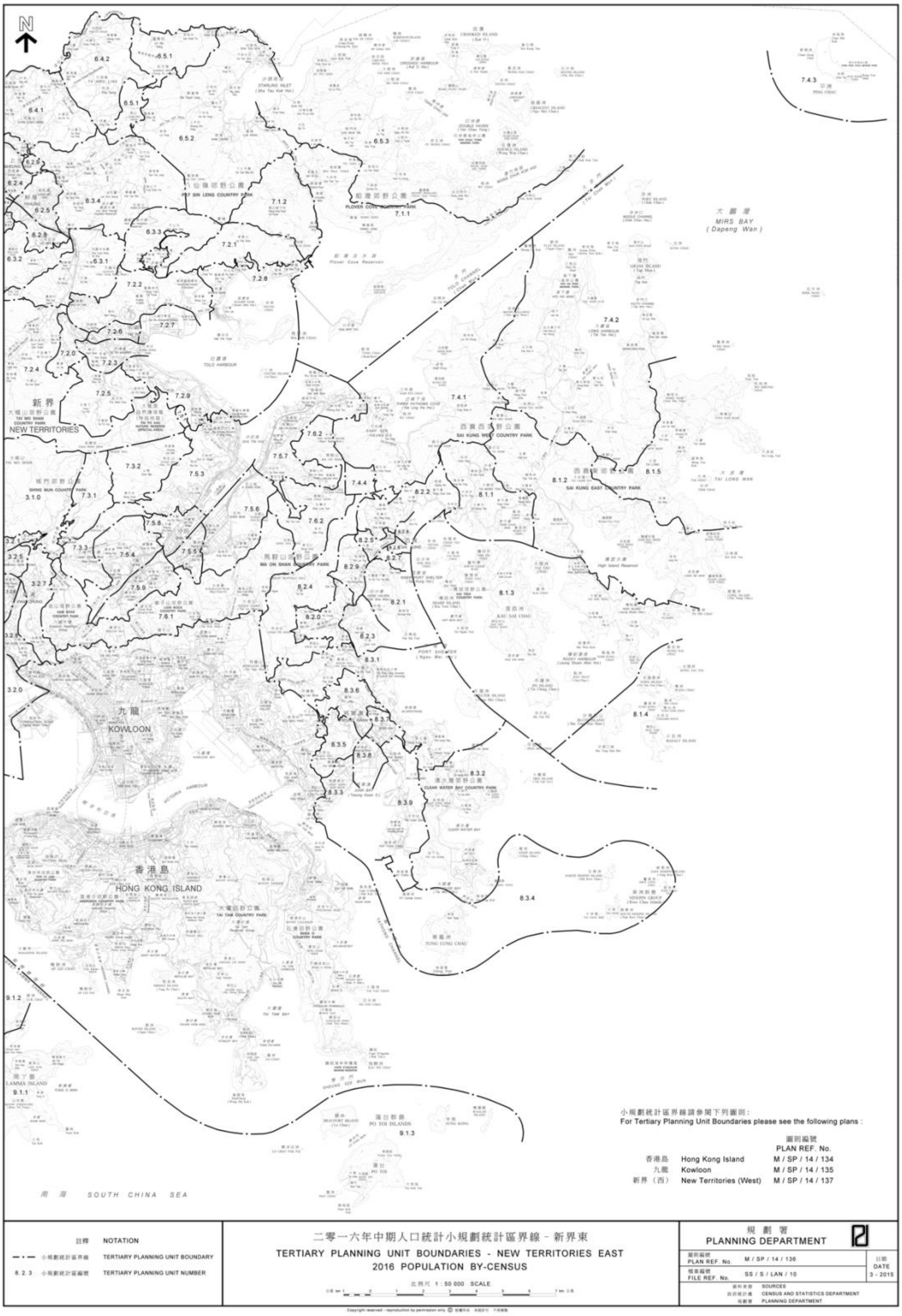
Appendices

Appendix 1. Tertiary Planning Units




二零一六年中期人口統計小規劃統計區界線 - 九龍
TERTIARY PLANNING UNIT BOUNDARIES - KOWLOON
2016 POPULATION BY-CENSUS
規劃署
PLANNING DEPARTMENT
圖則編號 PLAN REF. No. M / SP / 14 / 135
檔案編號 FILE REF. No. SS / S / LAN / 10
日期 DATE 3 - 2015
資料來源 SOURCES
政府統計處 CENSUS AND STATISTICS DEPARTMENT
規劃署 PLANNING DEPARTMENT
註釋 NOTATION
小規劃統計區界線 TERTIARY PLANNING UNIT BOUNDARY
2.3.2 小規劃統計區編號 TERTIARY PLANNING UNIT NUMBER
小規劃統計區界線請參閱下列圖則：
For Tertiary Planning Unit Boundaries please see the following plans :
圖則編號 PLAN REF. No.
香港島 Hong Kong Island M / SP / 14 / 134
新界（東） New Territories (East) M / SP / 14 / 136
新界（西） New Territories (West) M / SP / 14 / 137
比例尺 1 : 17 500 SCALE
N


N
深圳市 SHENZHEN SHI
后海灣 (深圳灣) DEEP BAY (Shenzhen Wan)
新界 NEW TERRITORIES
大欖郊野公園 TAI LAM COUNTRY PARK
大帽山郊野公園 TAI MO SHAN COUNTRY PARK
大嶼山 LANTAU ISLAND
北大嶼郊野公園 LANTAU NORTH COUNTRY PARK
南大嶼郊野公園 LANTAU SOUTH COUNTRY PARK
長洲 CHEUNG CHAU
南丫島 LAMMA ISLAND
香港島 HONG KONG ISLAND
索罟群島 SOKO ISLANDS
小規劃統計區界線請參閱下列圖則：
For Tertiary Planning Unit Boundaries please see the following plans :
圖則編號
PLAN REF. No.
香港島 Hong Kong Island M / SP / 14 / 134
九龍 Kowloon M / SP / 14 / 135
新界（東） New Territories (East) M / SP / 14 / 136
南海
SOUTH CHINA SEA
註釋 NOTATION
小規劃統計區界線 TERTIARY PLANNING UNIT BOUNDARY
5. 4. 5 小規劃統計區編號 TERTIARY PLANNING UNIT NUMBER
二零一六年中期人口統計小規劃統計區界線－新界西
TERTIARY PLANNING UNIT BOUNDARIES - NEW TERRITORIES WEST
2016 POPULATION BY-CENSUS
比例尺 1 : 50 000 SCALE
規劃署
PLANNING DEPARTMENT
圖則編號 PLAN REF. No. M / SP / 14 / 137
檔案編號 FILE REF. No. SS / S / LAN / 10
日期 DATE 3 - 2015
資料來源 SOURCES
政府統計處 CENSUS AND STATISTICS DEPARTMENT
規劃署 PLANNING DEPARTMENT


N
7.4.3
大鵬灣
MIRS BAY
(Dapeng Wan)
船灣郊野公園
PLOVER COVE COUNTRY PARK
八仙嶺郊野公園
PAT SIN LENG COUNTRY PARK
吐露港
TOLO HARBOUR
新界
NEW TERRITORIES
大帽山郊野公園
TAI MO SHAN COUNTRY PARK
城門郊野公園
SHING MUN COUNTRY PARK
西貢西郊野公園
SAI KUNG WEST COUNTRY PARK
西貢東郊野公園
SAI KUNG EAST COUNTRY PARK
馬鞍山郊野公園
MA ON SHAN COUNTRY PARK
獅子山郊野公園
LION ROCK COUNTRY PARK
清水灣郊野公園
CLEAR WATER BAY COUNTRY PARK
九龍
KOWLOON
香港島
HONG KONG ISLAND
大潭郊野公園
TAI TAM COUNTRY PARK
南海 SOUTH CHINA SEA
小規劃統計區界線請參閱下列圖則:
For Tertiary Planning Unit Boundaries please see the following plans :
圖則編號
PLAN REF. No.
香港島 Hong Kong Island M / SP / 14 / 134
九龍 Kowloon M / SP / 14 / 135
新界（西） New Territories (West) M / SP / 14 / 137
註釋 NOTATION
小規劃統計區界線 TERTIARY PLANNING UNIT BOUNDARY
8.2.3 小規劃統計區編號 TERTIARY PLANNING UNIT NUMBER
二零一六年中期人口統計小規劃統計區界線－新界東
TERTIARY PLANNING UNIT BOUNDARIES - NEW TERRITORIES EAST
2016 POPULATION BY-CENSUS
比例尺 1 : 50 000 SCALE
規劃署
PLANNING DEPARTMENT
圖則編號 PLAN REF. No. M / SP / 14 / 136
檔案編號 FILE REF. No. SS / S / LAN / 10
日期 DATE 3 - 2015
資料來源 SOURCES
政府統計處 CENSUS AND STATISTICS DEPARTMENT
規劃署 PLANNING DEPARTMENT

Appendix 2. Program to Merge XML Files for Restaurants

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program opens the individual XML file and creates a list of restaurants and
either the first or last appearance, as well as license number, name, and address
'''

import os
import sys
import xml.etree.ElementTree as ET
import openpyxl
import pandas as pd

##read each file to pull out date and license numbers
def readXML(xmlfile, File_Name, Master_Date, Master_License,
   Master_Name, Master_Address):

    countchecker = len(Master_License)
    print("Master list is", countchecker, "in length")

    Temp_List = []
    Shop_License = []
    Shop_Name = []
    Shop_Address = []

    tree = ET.parse(xmlfile)
    root = tree.getroot()
    date = root[1].text #get date

    for License in root.iter('LICNO'): #add license numbers to list
        Shop_License.append(License.text)
    for SName in root.iter('SS'): #add restaurant name to list
        Shop_Name.append(SName.text)
    for SAddress in root.iter('ADR'): #add address to list
        Shop_Address.append(SAddress.text)
    length = len(Shop_License) #find out how many licenses numbers were found

    counter = 0
    tempcount = 0
    length2 = len(Master_License)

    print("  ", date, "has", length, "entries")

    while counter < length: #Go through each entry in list
        ShouldAdd = True
        counter2 = 0
```

```
        #See if license number already exists in the master license number list
        if (Shop_License[counter] in Master_License):
           #go through all licenses to see if all data points match
            while counter2 < length2:
                if ((Shop_License[counter] == Master_License[counter2]) and
                   (Shop_Name[counter] == Master_Name[counter2]) and
                   (Shop_Address[counter] == Master_Address[counter2])):

                    ShouldAdd = False
                    break
                counter2 += 1

        #if the restaurant data is unique, add to the lists.
        if ShouldAdd == True:
            tempcount += 1
            Master_Date.append(date)
            Master_License.append(Shop_License[counter])
            Master_Name.append(Shop_Name[counter])
            Master_Address.append(Shop_Address[counter])

        counter += 1
    print("  ", date, "should add:", tempcount)

#Main Program
if __name__ == "__main__":
    inputList = []
    inputList2 = []
    Master_Date = []
    Master_License = []
    Master_Name = []
    Master_Address = []

    print("XML Scraper Master List v.4")

    ## Should the program find the first appearance or last appearance
    print("Select one:")
    print("1. Oldest to Newest")
    print("2. Newest to Oldest")
    varinput = input("Selecct 1 or 2 ")
    varinput = int(varinput)

    if (varinput == 1): #If going oldest to newest in appearance
        #Output Directory
        File_Name = os.path.expanduser("~/Desktop/Old to New Competition.csv")

        ##Input Directory
        #Directory to move finished xml files to
        File_Added = os.path.expanduser(
           "~/Downloads/Restaurant List - Old to New/Added/")
```

```
    #Directory of xml files still to add
    File_toAdd = os.path.expanduser(
      "~/Downloads/Restaurant List - Old to New/To Add/")

    #Read all file names in folder
    inputList = os.listdir(File_toAdd)
    inputList.sort() #reorder list
    inputLength = len(inputList)

#On a Mac this file can show up on loading list, rewrite to exclude
    if inputList[0] == ".DS_Store":
        counter = 1
        inputLen = inputLength - 1
        while inputLen > 0:
            inputList2.append(inputList[counter])
            inputLen -= 1
            counter += 1
        inputList = inputList2

    inputLength = len(inputList) #get length of list
    counter = 0 #counter for the array
    while counter < inputLength:
      #get next file ready
        xmlfile = File_toAdd + inputList[counter]
        #get file name to move on complete
        xmlmove = File_Added + inputList[counter]

    #Get license numbers
        readXML(xmlfile, File_Name, Master_Date, Master_License,
           Master_Name, Master_Address)

        #Move file that is added to a new directory
        os.rename(xmlfile, xmlmove)
        counter += 1 #add to the counter for the next item in the list

    df = pd.DataFrame({
      'First_Appearance' : Master_Date,
      'License' : Master_License,
      'Name' : Master_Name,
      'Address' : Master_Address
      })

elif varinput == 2: #If going newest to oldest in appearance
    #Output Directory
    File_Name = os.path.expanduser("~/Desktop/New to Old Competition.csv")

    ##Input Directory
    #directory to move finished xml files to
```

```
    File_Added = os.path.expanduser(
      "~/Downloads/Restaurant List - New to Old/Added/"
      )

    #Directory of xml files still to add
    File_toAdd = os.path.expanduser(
      "~/Downloads/Restaurant List - New to Old/To Add/"
      )

    #Read all file names in folder
    inputList = os.listdir(File_toAdd)
    inputList.sort() #reorder list
    inputLength = len(inputList)

#On a mac this file can show up on loading list, rewrite file to exclude
    if inputList[0] == ".DS_Store":
        counter = 1
        inputLen = inputLength - 1
        while inputLen > 0:
            inputList2.append(inputList[counter])
            inputLen -= 1
            counter += 1
        inputList = inputList2

    inputLength = len(inputList)-1 #get length of list
    while inputLength >= 0:
        #Get next file ready
        xmlfile = File_toAdd + inputList[inputLength]

        #Get file name to move on complete
        xmlmove = File_Added + inputList[inputLength]

    #get license numbers
        readXML(xmlfile, File_Name, Master_Date, Master_License,
            Master_Name, Master_Address)

        #move file that is added to a new directory
        os.rename(xmlfile, xmlmove)

        inputLength -= 1 #add to the counter for the next item in the list

    df = pd.DataFrame({
      'Last_Appearance' : Master_Date,
      'License' : Master_License,
      'Name' : Master_Name,
      'Address' : Master_Address
      })

countchecker = len(Master_License)
```

```
  print("Final master list is", countchecker, "in length")

#Save the csv file
  df.to_csv(File_Name, encoding='utf_8_sig', index=False)
  print("Finished extracting data from the XML files.")
  input("Press Enter to exit")
```

Appendix 3. Program to Merge XML Files for Convenience Competitors

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program that opens the individual XML file and creates a list of indirect competitors
and either the first or last appearance, as well as license number, name and address
'''

import os
import sys
import xml.etree.ElementTree as ET
import openpyxl
import pandas as pd

##Read each file to pull out date and license numbers
def readXML(xmlfile, File_Name, Master_Date, Master_License,
    Master_Name, Master_Address):

    countchecker = len(Master_License)
    print("Master list is", countchecker, "in length")

    Temp_List = []
    Shop_License = []
    Shop_Name = []
    Shop_Address = []
    Shop_Type = []

    tree = ET.parse(xmlfile)
    root = tree.getroot()
    date = root[1].text #get date

    for License in root.iter('LICNO'): #add license numbers to list
        Shop_License.append(License.text)
    for SName in root.iter('SS'): #add restaurant name to list
        Shop_Name.append(SName.text)
    for SAddress in root.iter('ADR'): #add address to list
        Shop_Address.append(SAddress.text)
    for SType in root.iter('TYPE'):
        Shop_Type.append(SType.text)
    length = len(Shop_License) #find out how many licenses numbers were found

    counter = 0
    tempcount = 0
    length2 = len(Master_License)

    print("  ", date, "has", length, "entries")
    while counter < length: #Go through each entry in list
```

```
      if((Shop_Type[counter] == "CL") or
        (Shop_Type[counter] == "FB") or
        (Shop_Type[counter] == "FS")):

          ShouldAdd = True
          counter2 = 0

          #See if license number already exists in the master license list
          if (Shop_License[counter] in Master_License):
              #go through all licenses to see if all data points match
                while counter2 < length2:
                    if ((Shop_License[counter] == Master_License[counter2]) and
                      (Shop_Name[counter] == Master_Name[counter2]) and
                      (Shop_Address[counter] == Master_Address[counter2])):

                        ShouldAdd = False
                        break
                    counter2 += 1

        #If the competitors' data is unique, add to the lists.
          if ShouldAdd == True:
                tempcount += 1
                Master_Date.append(date)
                Master_License.append(Shop_License[counter])
                Master_Name.append(Shop_Name[counter])
                Master_Address.append(Shop_Address[counter])

      counter += 1
  print("  ", date, "should add:", tempcount)

#Main Program
if __name__ == "__main__":
  inputList = []
  inputList2 = []
  Master_Date = []
  Master_License = []
  Master_Name = []
  Master_Address = []

  print("XML Scraper Convenience Competitor Master List v.4")

 ## Should the program find the first appearance or last appearance
  print("Select one:")
  print("1. Oldest to Newest")
  print("2. Newest to Oldest")
  varinput = input("Selecct 1 or 2 ")
  varinput = int(varinput)
```

```
if varinput == 1: #If going oldest to newest in appearance
    #Output Directory
    File_Name = os.path.expanduser(
        "~/Desktop/Convenience Competitors Old to New Competition.csv"
        )

    ##Input Directory
    #Directory to move finished xml files to
    File_Added = os.path.expanduser(
        "~/Downloads/Convenience Competitors List - Old to New/Added/"
        )

    #Directory of xml files still to add
    File_toAdd = os.path.expanduser(
        "~/Downloads/Convenience Competitors List - Old to New/To Add/"
        )

    #Read all file names in folder
    inputList = os.listdir(File_toAdd)
    inputList.sort() #reorder list
    inputLength = len(inputList)

#On a mac this file can show up on loading list, rewrites to exclude
    if inputList[0] == ".DS_Store":
        counter = 1
        inputLen = inputLength - 1
        while inputLen > 0:
            inputList2.append(inputList[counter])
            inputLen -= 1
            counter += 1
        inputList = inputList2

    inputLength = len(inputList) #get length of list
    counter = 0 #counter for the array
    while counter < inputLength:
        #get next file ready
        xmlfile = File_toAdd + inputList[counter]
        #get file name to move on complete
        xmlmove = File_Added + inputList[counter]

        #get license numbers
        readXML(xmlfile, File_Name, Master_Date, Master_License,
            Master_Name, Master_Address)

        #move file that is added to a new directory
        os.rename(xmlfile, xmlmove)

        counter += 1 #add to the counter for the next item in the list
```

```
    df = pd.DataFrame({
        'First_Appearance' : Master_Date,
        'License' : Master_License,
        'Name' : Master_Name,
        'Address' : Master_Address
        })

elif varinput == 2: #If going newest to oldest in appearance
    #Output Directory
    File_Name = os.path.expanduser(
        "~/Desktop/Convenience Competitors New to Old Competition.csv"
        )

    ##Input Directory
    #Directory to move finished xml files to
    File_Added = os.path.expanduser(
        "~/Downloads/Convenience Competitors List - New to Old/Added/"
        )

    #Directory of xml files still to add
    File_toAdd = os.path.expanduser(
        "~/Downloads/Convenience Competitors List - New to Old/To Add/"
        )

    #Read all file names in folder
    inputList = os.listdir(File_toAdd)
    inputList.sort() #reorder list
    inputLength = len(inputList)

#On a mac this file can show up on loading list, rewrites to exclude
    if inputList[0] == ".DS_Store":
        counter = 1
        inputLen = inputLength - 1
        while inputLen > 0:
            inputList2.append(inputList[counter])
            inputLen -= 1
            counter += 1
        inputList = inputList2

    inputLength = len(inputList)-1 #get length of list
    while inputLength >= 0:
        #Get next file ready
        xmlfile = File_toAdd + inputList[inputLength]
        #get file name to move on complete
        xmlmove = File_Added + inputList[inputLength]

        #get license numbers
        readXML(xmlfile, File_Name, Master_Date, Master_License,
            Master_Name, Master_Address)
```

```
            #move file that is added to a new directory
            os.rename(xmlfile, xmlmove)

            inputLength -= 1 #add to the counter for the next item in the list

    df = pd.DataFrame({
        'Last_Appearance' : Master_Date,
        'License' : Master_License,
        'Name' : Master_Name,
        'Address' : Master_Address
        })

countchecker = len(Master_License)
print("Final convenience competitor list is", countchecker, "in length")

df.to_csv(File_Name, encoding='utf_8_sig', index=False) #save the csv file
print("Finished extracting data from the XML files.")
input("Press Enter to exit")
```

Appendix 4. Program to Scrape the Census Data

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program to use the latitude and longitude of each restaurant to get the census data
'''

import numpy as np
import json
import geojson
import os
from geojson import Feature
import geog
import pyproj
import shapely.geometry
import pandas as pd
from shapely.geometry import shape, Polygon
from shapely.ops import transform
from functools import partial
from osgeo import ogr
from osgeo import osr

if __name__ == "__main__":
    counter_1 = 0
    counter_2 = 0

    ##Set the lists for each variable
    Save_Total_Int_Pop = []
    Save_Total_Int_Age = []
    Save_TotalInt_Education = []
    Save_TotalInt_Employ = []
    Save_TotalInt_Ethnicity = []
    Save_TotalInt_Flat = []
    Save_TotalInt_Gender = []
    Save_TotalInt_Income = []
    Save_TotalInt_NumChild = []
    Save_TotalInt_NumHouse = []
    Save_TotalInt_NumAdult = []

    print("Census Information for Restaurants")

    #Initialize Source and Target
    source = osr.SpatialReference()
    source.ImportFromEPSG(4326)
    target = osr.SpatialReference()
    target.ImportFromEPSG(5243)
    transform = osr.CoordinateTransformation(source, target)
```

```
#Read the list of restaurants
file_name = os.path.expanduser("~/Desktop/Hong Kong Restaurants.csv")
#Read the list of restaurants
Restaurant_List = pd.read_csv(file_name, encoding='utf-8')

#Read the list of TPU numbers
TPU_file = os.path.expanduser(
   "~/Documents/Thesis/Data Lists/TPU Number.csv"
   )

TPU = pd.read_csv(TPU_file, encoding='utf-8')

#Read the list of TPU Census Data
Census_file = os.path.expanduser("~/Documents/Thesis/TPU Census Data.csv")
Census = pd.read_csv(Census_file, encoding='utf-8')

#Get Counts
Restaurant_Counter = Restaurant_List['First_Appearance'].count()
TPU_Count = TPU['TPU Number'].count()

#Get Final Column to save data to
Restaurant_Column = Restaurant_List.shape[1]

while(counter_1 < Restaurant_Counter):
    counter_2 = 0
    counter_3 = 0
    Count_Intersect = 0
    Int_Pop = 0
    Total_intersect = 0
    Total_Int_Age = 0
    TotalInt_Education = 0
    TotalInt_Employ = 0
    TotalInt_Ethnicity = 0
    TotalInt_Flat = 0
    TotalInt_Gender = 0
    TotalInt_Income = 0
    TotalInt_NumChild = 0
    TotalInt_NumHouse = 0
    TotalInt_NumAdult = 0
    Total_Int_Pop = 0
    Inter_size = []
    Inter_list = []
    Inter_num = []
    xx = 0
    Number_TPU_Overlap = 0

    ##Restaurant Information
    #restaurant number
```

```
Restaurant_License = Restaurant_List["License_Number"][counter_1]
#latitude of restaurant
Restaurant_Latitude = Restaurant_List["Latitude"][counter_1]
#longitude of restaurant
Restaurant_Longitude = Restaurant_List["Longitude"][counter_1]

##Create a circle with the mid of the restaurant, and radius of 369.95 meters

Restaurant_Center = shapely.geometry.Point([
    Restaurant_Longitude,
    Restaurant_Latitude
    ])

n_points = 360 #how many datapoints in the circle
d = 369.95 #radius in meters
angles = np.linspace(0, 360, n_points)

#create polygon
polygon = geog.propagate(
    Restaurant_Center,
    angles,
    d,
    deg = True,
    bearing = False
    )

temaa = shapely.geometry.Polygon(polygon)

rest_polygon2 = shapely.geometry.Polygon(polygon)
rest_polygon = geojson.Feature(geometry=temaa, properties=[[]])

print(counter_1+1, "of", Restaurant_Counter)
print("  Restaurant Number:", Restaurant_License)

#Restaurant Area
rest_polygon3 = ogr.CreateGeometryFromJson(str(rest_polygon['geometry']))
rest_polygon3.Transform(transform)
Restaurant_area = rest_polygon3.GetArea()

while (counter_3 < TPU_Count):
    TPU_Number = TPU['TPU Number'][counter_3]

    #Open TPU File
    file_open = os.path.expanduser(
        "~/Documents/Thesis/GeoJson/TPU GeoJson Files/Entire Hong Kong/"
        + str(TPU_Number) + ".geojson")

    with open(file_open) as f:
        TPU_polygon = json.load(f)
```

```
    TPU_P = shape(TPU_polygon['geometry']) #Get the polygon of the TPU

    #Get the TPU Area
    poly = ogr.CreateGeometryFromJson(str(TPU_polygon['geometry']))
    poly.Transform(transform)
    TPU_area = poly.GetArea()

    #Intersection Area
    #create a polygon of the intersect points
    Poly_Intersect = rest_polygon2.intersection(TPU_P)

    if (Poly_Intersect.is_empty == False): #If they do intersect
        Poly_feat = geojson.Feature(
            geometry = Poly_Intersect,
            properties = [[]]
            )

        poly = ogr.CreateGeometryFromJson(str(Poly_feat['geometry']))
        poly.Transform(transform)
        intersect_area = poly.GetArea()
        Inter_num.append(counter_3)
        Inter_list.append(TPU_Number)
        Inter_size.append(intersect_area)
        print(" ", TPU_Number,
              "overlaps: %.2f%%" %((intersect_area/Restaurant_area) * 100))

        Total_intersect = Total_intersect + intersect_area
    counter_3 += 1

Number_TPU_Overlap = len(Inter_list)

while xx < Number_TPU_Overlap:
    #get percentage of the land (exclude water) area the TPU intersects
    nTPUmeter = Inter_size[xx] / Total_intersect

    TPU_Number = TPU['TPU Number'][Inter_num[xx]]

    #find the row of TPU that intersects in Census File
    RowNum = Census.loc[Census['TPU Number']==TPU_Number].index[0]

    #Use per square meter
    #multiply census per sq. meter with square meter
    Int_Pop = Census['ResDen'].iloc[RowNum] * Inter_size[xx]
    #add all TPU intersects together
    Total_Int_Pop = Total_Int_Pop + Int_Pop
    #multiply census per sq. meter with square meter
    Int_NumChild = Census['NumChild'].iloc[RowNum] * Inter_size[xx]
    #add all TPU intersects together
```

```
        TotalInt_NumChild = TotalInt_NumChild + Int_NumChild
        #multiply census per sq. meter with square meter
        Int_NumAdult = Census['NumAdult'].iloc[RowNum] * Inter_size[xx]
        #add all TPU intersects together
        TotalInt_NumAdult = TotalInt_NumAdult + Int_NumAdult

        #use proportion of overlap
        #multiply census per sq. meter with Percentage overlap
        Int_Income = Census['Income'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Income = TotalInt_Income + Int_Income
        #multiply census per sq. meter with Percentage overlap
        IntersectAge = Census['Age'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        Total_Int_Age = Total_Int_Age + IntersectAge
        #multiply census per sq. meter with Percentage overlap
        Int_NumHouse = Census['NumHouse'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_NumHouse = TotalInt_NumHouse + Int_NumHouse
        #multiply census per sq. meter with Percentage overlap
        Int_Ethnicity = Census['Ethnicity'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Ethnicity = TotalInt_Ethnicity + Int_Ethnicity
        #multiply census per sq. meter with Percentage overlap
        Int_Gender = Census['Gender'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Gender = TotalInt_Gender + Int_Gender
        #multiply census per sq. meter with Percentage overlap
        Int_Education = Census['Education'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Education = TotalInt_Education + Int_Education
        #multiply census per sq. meter with Percentage overlap
        Int_Flat = Census['Flat'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Flat = TotalInt_Flat + Int_Flat
        #multiply census per sq. meter with Percentage overlap
        Int_Employ = Census['Employ'].iloc[RowNum] * nTPUmeter
        #add all TPU intersects together
        TotalInt_Employ = TotalInt_Employ + Int_Employ

        xx += 1



    Save_Total_Int_Pop.append(Total_Int_Pop)
    Save_Total_Int_Age.append(Total_Int_Age)
    Save_TotalInt_Education.append(TotalInt_Education)
    Save_TotalInt_Employ.append(TotalInt_Employ)
    Save_TotalInt_Ethnicity.append(TotalInt_Ethnicity)
```

```
        Save_TotalInt_Flat.append(TotalInt_Flat)
        Save_TotalInt_Gender.append(TotalInt_Gender)
        Save_TotalInt_Income.append(TotalInt_Income)
        Save_TotalInt_NumChild.append(TotalInt_NumChild)
        Save_TotalInt_NumHouse.append(TotalInt_NumHouse)
        Save_TotalInt_NumAdult.append(TotalInt_NumAdult)

        counter_1 +=1

#Save the file
#insert the Population
Restaurant_List.insert(
    Restaurant_Column,
    "Population",
    Save_Total_Int_Pop
    )

#insert the Age
Restaurant_List.insert(
    Restaurant_Column + 1,
    "Age",
    Save_Total_Int_Age)

#insert the Education
Restaurant_List.insert(
    Restaurant_Column + 2,
    "Education",
    Save_TotalInt_Education
    )

#insert the Employ
Restaurant_List.insert(
    Restaurant_Column + 3,
    "Employ",
    Save_TotalInt_Employ
    )

#insert the Ethnicity
Restaurant_List.insert(
    Restaurant_Column + 4,
    "Ethnicity",
    Save_TotalInt_Ethnicity
    )

#insert the Flat
Restaurant_List.insert(
    Restaurant_Column + 5,
    "Flat",
    Save_TotalInt_Flat
```

```
    )

#insert the Gender
Restaurant_List.insert(
    Restaurant_Column + 6,
    "Gender",
    Save_TotalInt_Gender
    )

#insert the Income
Restaurant_List.insert(
    Restaurant_Column + 7,
    "Income",
    Save_TotalInt_Income
    )

#insert the Number of Children
Restaurant_List.insert(
    Restaurant_Column + 8,
    "NumChild",
    Save_TotalInt_NumChild
    )

#insert the Num Household
Restaurant_List.insert(
    Restaurant_Column + 9,
    "NumHouse",
    Save_TotalInt_NumHouse
    )

#insert the Number of Adults
Restaurant_List.insert(
    Restaurant_Column + 10,
    "NumAdult",
    Save_TotalInt_NumAdult
    )

#save the csv file
Restaurant_List.to_csv(file_name, encoding='utf_8_sig', index=False)
```

Appendix 5. Program to Scrape OpenRice.com

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program that opens up the unique URL for each restaurant and downloads the data
from OpenRice.com
'''

from selenium import webdriver
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.options import Options
from selenium.common.exceptions import TimeoutException
from bs4 import BeautifulSoup
from sys import argv
import multiprocessing
import time
import csv
import unicodedata
import numpy as np
import re
import os
import pathlib
import sys
import math
import json
import pandas as pd

def allItems(OutputFile, url_list):
    #Set basic variables
    url_Length = 0
    countx = 0

    #Load the browser settings
    browser = webdriver.Firefox()
    url_Length = len(url_list)

    #go to each individual url for the Item_Search topic and scrape the recipe
    while countx < url_Length:
        Finished2 = 0
        counter = 0
        xx = url_list[countx]
        while Finished2 == 0:
            try:
                browser.get(xx) #load the specific webpage
                Finished2 = 1
                time.sleep(5)
            except:
```

```
            #if it goes to except 5 times, close the browser and reopen
             if counter == 5:
                 browser.close()
                 time.sleep(1)
                 browser = webdriver.Firefox()
                 counter = 0
             counter += 1
             time.sleep(1)
        print(countx+1, "of", url_Length)

        #Get the information from the page
        scrape_restaurant(browser, OutputFile, xx)
        countx +=1

    browser.quit() #close the browser


#Get the data from the HTML code
def scrape_restaurant(browser, OutputFile, xx):
    print("Scraping:", xx)
    rtype = []
    shouldinclude = 0
    cx = 0
    isclosed = 99
    ismoved = 99
    issold = 99
    isbuilding = 99
    isrenamed = 99
    isdelivery = 99
    AggReview = 99999
    POI_Name = ""
    numseats = 99999
    NumTried = 0

    while NumTried < 5:
        try:
           #Try to open the HTML Code
            soup = BeautifulSoup(browser.page_source, 'html.parser')

            try: #look for the json information
                js = json.loads(soup.find(
                    "script",type="application/ld+json").text
                    )
            except:
                try:
                    tempfile = soup.find(
                        "script", type="application/ld+json").text
                    js = json.loads(tempfile, strict=False)
                except:
```

```
        pass

try: #Try to find the information from the json data
    address = js["address"]
    name = js["name"]
    name = name.replace("&#039;", "'")
    name = name.replace("&#39;", "'")
    name = name.replace("&amp;", "&")
    type = js["@type"]
    address_type = address["@type"]
    address_street = address["streetAddress"]
    address_street = address_street.replace("&#039;", "'")
    address_street = address_street.replace("&#39;", "'")
    address_street = address_street.replace("&amp;", "&")
    address_locality = address["addressLocality"]
    address_region = address["addressRegion"]
    address_postal = address["postalCode"]
    address_country = address["addressCountry"]
    pricerange = js["priceRange"]
    cuisine = js["servesCuisine"]
except: #if json file doesn't work, look elsewhere for data
    name = soup.find('span', itemprop='name', class_='name').text
    for x in soup.find('div', text = re.compile('Address'),
        attrs = [['class':'title']]).find_next_siblings():

        address_street = x.text
        address_street = address_street.rstrip().strip()

    address_type = "None Listed"
    address_locality = ""
    address_region = ""
    address_postal = ""
    address_country = "HK"
    pricerange = soup.find('div',
        class_="header-poi-price dot-separator").text

    pricerange = pricerange.rstrip().strip()

    cuisine = soup.find('div',
        class_="header-poi-categories dot-separator").text

    cuisine = cuisine.rstrip().strip()

try: #Try to get the latitude and longitude
    geo = js["geo"]
    latitude = geo["latitude"]
    longitude = geo["longitude"]
except:
    try:
```

```
        secondgeo = soup.find("div", class_="mapview-container")
        latitude = secondgeo.attrs.get('data-latitude')
        longitude = secondgeo.attrs.get('data-longitude')
    except:
        latitude = 0.00
        longitude = 0.00

##get restaurant categories
try:
    for i in soup.select("div.header-poi-categories > a"):
        rest_type = i.text
        if(rest_type not in cuisine):
            rtype.append(rest_type)
except:
    rtype.append("Failed")

##Find POI Name
try:
    POI_Name = soup.find('span', itemprop='name', class_='name').text
except:
    POI_Name = "Failed"

##Find Aggregate Review Score
try:
    AggReview = soup.find("div", class_="header-score").text
except:
    AggReview = 99999

##Permanently closed, moved, new owner, under renovation, or renamed
try:
    if(soup.find("span",
       class_="poi-with-other-status").text == "(Closed)"):

        isclosed = 1
    elif(soup.find("span",
       class_="poi-with-other-status").text == "(Moved)"):

        ismoved = 1
    elif(soup.find("span",
       class_="poi-with-other-status").text == "(Changed Hand)"):

        issold = 1
    elif(soup.find("span",
       class_="poi-with-other-status").text == "(Under Renovation)"):

        isbuilding = 1
    elif(soup.find("span",
       class_="poi-with-other-status").text == "(Renamed)"):
```

```
                isrenamed = 1
            elif(soup.find("span",
                class_="poi-with-other-status").text == "(Delivery Only)"):

                isdelivery = 1
        except:
            pass

        ##Number of Seats
        try:
            for x in soup.find(
                'div',
                text = re.compile('Number of Seats'),
                attrs = [['class':'title']]).find_next_siblings():

                numseats = x.text
        except:
            numseats = 99999

        ##Get Review Scores
        for el in soup.find_all("div", class_="score-div"):
            if(cx == 0): smile = el.get_text() #first reference is smile
            elif(cx == 1): ok = el.get_text() #second reference is ok
            elif(cx == 2): cry = el.get_text() #third reference is cry
            cx +=1
        NumTried = 5
    except:
        print("Failed to scrape")
        NumTried += 1
        name = "Failed"
        address_type = "Failed"
        address_street = "Failed"
        address_locality = "Failed"
        address_region = "Failed"
        address_postal = "Failed"
        address_country = "Failed"
        latitude = 0.00
        longitude = 0.00
        pricerange = "Failed"
        cuisine = "Failed"
        POI_Name = "Failed"
        numseats = 99999
        AggReview = 99999
        isclosed = 99
        ismoved = 99
        issold = 99
        isbuilding = 99
        isrenamed = 99
        isdelivery = 99
```

```
            smile = 99
            ok = 99
            cry = 99

    print("Name:", name)
    print("POI Name:", POI_Name)
    print("Address Type:", address_type)
    print("Street Address:", address_street)
    print("Locality:", address_locality)
    print("Region:", address_region)
    print("Postal:", address_postal)
    print("Country:", address_country)
    print("Latitude:", latitude)
    print("Longitude:", longitude)
    print("Price Range:", pricerange)
    print("Cuisine:", cuisine)
    print("Restaurant Type:", rtype)
    print("Number of Seats:", numseats)
    print("Aggregated Review:", AggReview)
    print("Review Score:")
    print("  Smile:", smile)
    print("  OK:", ok)
    print("  CRY:", cry)
    print("----")

    #Write the individual recipe to the csv file
    with open(
       OutputFile, 'a',
       newline = '\n',
       encoding = 'utf_8_sig'
       ) as Overview_file:

        writer = csv.writer(
           Overview_file,
           delimiter = ',',
           quoting = csv.QUOTE_MINIMAL
           )

        writer.writerow([xx, name, POI_Name, address_type, address_street,
           address_locality, address_region, address_postal, address_country,
           latitude, longitude, numseats, isclosed, ismoved, issold,
           isbuilding, isrenamed, isdelivery, cuisine, pricerange, rtype,
           AggReview, smile, ok, cry])

#Main Program
if __name__ == "__main__":
    url = []

    print("Welcome to OpenRice Scraper")
```

```
#File paths to file
InputFile = os.path.expanduser(
    "~/Desktop/OpenRice/URLs to Scrape.csv"
    )

OutputFile = os.path.expanduser(
    "~/Desktop/OpenRice/OpenRice Scraped Results.csv"
    )

#if files do not exist, create them and write headings
if not os.path.exists(OutputFile):
      with open(OutputFile, 'w',
          newline = "\n",
          encoding = 'utf_8_sig'
          ) as Overview_file:

            writer = csv.writer(
                Overview_file,
                delimiter = ',',
                quoting = csv.QUOTE_MINIMAL
                )

            writer.writerow(["Searched URL", "Name", "POI Name",
                "Address Type","Street Address","Locality", "Region",
                "Postal", "Country", "Latitude", "Longitude", "Number of Seats",
                "Closed", "Moved", "New Owner", "Under Renovation", "Renamed",
                "Delivery Only", "Cuisine", "Price", "Type", "Aggregated Review",
                "Smile Rating", "OK Rating", "Cry Rating"])

List = pd.read_csv(InputFile, encoding='utf-8')
url_list = List['URL']

#Go to following section to perform operations
allItems(OutputFile, url_list)

input("Press Enter to exit") #use to keep window open
sys.exit() #exit the program
```

Appendix 6. Program to Count Competitors with Matching Price

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*

'''
See how many restaurants price competitors are within a radius of 369.95 meters and
were open at the same time.
'''

from datetime import datetime
from collections import namedtuple
import pandas as pd
import geopy.distance
import os

#Main Program
if __name__ == "__main__":
    Range = namedtuple('Range', ['start', 'end'])
    competitors = 0
    rest_len = 0
    comp = []

    file_name1 = os.path.expanduser(
      "~/Desktop/Hong Kong Restaurants 2016-2018.csv"
      )

    pilot1 = pd.read_csv(file_name1, encoding='utf-8')
    file_name2 = os.path.expanduser("~/Desktop/Entire Hong Kong Restaurants.csv")
    pilot2 = pd.read_csv(file_name2, encoding='utf-8')

    #Get number of rows
    pilot_len1 = pilot1["First_Appearance"].count()
    pilot_len2 = pilot2["First_Appearance"].count()

    pilot_col_to_save = pilot1.shape[1] #Get number of columns

    while(rest_len < pilot_len1):
        competitors = 0
        counter = 0

        r1_license = pilot1["License_Number"][rest_len]
        r1_name = pilot1["Name"][rest_len]
        r1_address = pilot1["Address"][rest_len]

        r1_start = pilot1["First_Appearance"][rest_len]
        r1_end = pilot1["Last_Appearance"][rest_len]
```

```
#Extract Day Month Year from Start Date
r1_start_month = int(r1_start.split('/')[0])
r1_start_day = int(r1_start.split('/')[1])
r1_start_year = int(r1_start.split('/')[2])

#Extract Day Month Year from End Date
r1_end_month = int(r1_end.split('/')[0])
r1_end_day = int(r1_end.split('/')[1])
r1_end_year = int(r1_end.split('/')[2])

#Get coordinates of original restaurant
r1_coord_lat = pilot1["Latitude"][rest_len]
r1_coord_lon = pilot1["Longitude"][rest_len]
r1_coords = (r1_coord_lat, r1_coord_lon)

#Get Date Range
r1_range = Range(
    start = datetime(
        r1_start_year,
        r1_start_month,
        r1_start_day
        ),
    end = datetime(
        r1_end_year,
        r1_end_month,
        r1_end_day
        )
    )

#Get Price
r1_price = pilot1["Price"][rest_len]

while(counter < pilot_len2):
    overlap = 0
    m_dist = 0

    r2_name = pilot2["Name"][counter]
    r2_license = pilot2["License_Number"][counter]
    r2_address = pilot2["Address"][counter]

    #Make sure the restaurant being checked isn't the original restaurant
    if(r1_license == r2_license and
        r1_name == r2_name and
        r1_address == r2_address):

        if((counter+1) == pilot_len2):
            break
        else:
            counter += 1
```

```
   r2_name = pilot2["Name"][counter]
   r2_license = pilot2["License_Number"][counter]

   #Get coordinates of potential competitor
   r2_coord_lat = pilot2["Latitude"][counter]
   r2_coord_lon = pilot2["Longitude"][counter]
   r2_coords = (r2_coord_lat, r2_coord_lon)

#get distance between restaurants
   m_dist = geopy.distance.distance(r1_coords, r2_coords).m

#Is the restaurant within a close distance to the one being checked
   if(m_dist < 369.95):
      r2_price = pilot2["Price"][counter] #get competitor price

  #if price is the same then...
      if(r1_price == r2_price):
         #Extract First and Last Appearance
         r2_start = pilot2["First_Appearance"][counter]
         r2_end = pilot2["Last_Appearance"][counter]

         #Extract Day Month Year from Start Date
         r2_start_month = int(r2_start.split('/')[0])
         r2_start_day = int(r2_start.split('/')[1])
         r2_start_year = int(r2_start.split('/')[2])

         #Extract Day Month Year from End Date
         r2_end_month = int(r2_end.split('/')[0])
         r2_end_day = int(r2_end.split('/')[1])
         r2_end_year = int(r2_end.split('/')[2])

         #Set date range
         r2_range = Range(
           start = datetime(
             r2_start_year,
             r2_start_month,
             r2_start_day
             ),
           end = datetime(
             r2_end_year,
             r2_end_month,
             r2_end_day
             )
           )

         if(r1_range.start <= r2_range.end and
           r2_range.start <= r1_range.end):
```

```
                    competitors += 1

            counter += 1 #Go to next next restaurant

        comp.append(competitors)
        print(rest_len+1, "of", pilot_len1, '\t',
            "Number of price competitors for", r1_license, ":", competitors)
        rest_len +=1

#add competitors to dataframe
  pilot1.insert(pilot_col_to_save, "Number of Price Competitors", comp)
  pilot1.to_csv(file_name1, encoding='utf_8_sig', index=False) #save dataframe
```

Appendix 7. Program to Count Competitors with Matching Cuisine

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*

'''
See how many restaurants cuisine competitors are within a radius of 369.95 meters
and were open at the same time.
'''

from datetime import datetime
from collections import namedtuple
import pandas as pd
import geopy.distance
import os

#Main Program
if __name__ == "__main__":
    Range = namedtuple('Range', ['start', 'end'])
    competitors = 0
    rest_len = 0
    comp = []

    file_name1 = os.path.expanduser(
      "~/Desktop/Hong Kong Restaurants 2016-2018.csv"
      )

    pilot1 = pd.read_csv(file_name1, encoding='utf-8')
    file_name2 = os.path.expanduser(
      "~/Desktop/Entire Hong Kong Restaurants.csv"
      )

    pilot2 = pd.read_csv(file_name2, encoding='utf-8')

    #Get number of rows
    pilot_len1 = pilot1["First_Appearance"].count()
    pilot_len2 = pilot2["First_Appearance"].count()

    pilot_col_to_save = pilot1.shape[1] #Get number of columns

    while(rest_len < pilot_len1):
        competitors = 0
        counter = 0

        r1_license = pilot1["License_Number"][rest_len]
        r1_name = pilot1["Name"][rest_len]
        r1_address = pilot1["Address"][rest_len]

        r1_start = pilot1["First_Appearance"][rest_len]
```

```
r1_end = pilot1["Last_Appearance"][rest_len]

#Extract Day Month Year from Start Date
r1_start_month = int(r1_start.split('/')[0])
r1_start_day = int(r1_start.split('/')[1])
r1_start_year = int(r1_start.split('/')[2])

#Extract Day Month Year from End Date
r1_end_month = int(r1_end.split('/')[0])
r1_end_day = int(r1_end.split('/')[1])
r1_end_year = int(r1_end.split('/')[2])

#Get coordinates of original restaurant
r1_coord_lat = pilot1["Latitude"][rest_len]
r1_coord_lon = pilot1["Longitude"][rest_len]
r1_coords = (r1_coord_lat, r1_coord_lon)

#Get Date Range
r1_range = Range(
  start = datetime(
    r1_start_year,
    r1_start_month,
    r1_start_day
    ),
  end = datetime(
    r1_end_year,
    r1_end_month,
    r1_end_day
    )
  )

#Get Cuisine
r1_cuisine = pilot1["Cuisine"][rest_len]

while(counter < pilot_len2):
   overlap = 0
   m_dist = 0

   r2_name = pilot2["Name"][counter]
   r2_license = pilot2["License_Number"][counter]
   r2_address = pilot2["Address"][counter]

   #Make sure the restaurant being checked isn't the original restaurant
   if(r1_license == r2_license and
     r1_name == r2_name and
     r1_address == r2_address):

      if((counter+1) == pilot_len2):
         break
```

```
            else:
                counter += 1

        r2_name = pilot2["Name"][counter]
        r2_license = pilot2["License_Number"][counter]

        #Get coordinates of potential competitor
        r2_coord_lat = pilot2["Latitude"][counter]
        r2_coord_lon = pilot2["Longitude"][counter]
        r2_coords = (r2_coord_lat, r2_coord_lon)

    #get distance between restaurants
        m_dist = geopy.distance.distance(r1_coords, r2_coords).m

    #Is the restaurant within a close distance to the one being checked
        if(m_dist < 369.95):
            r2_cuisine = pilot2["Cuisine"][counter] #get competitor cuisine

            if(r1_cuisine == r2_cuisine):  #if they match the same cuisine:
                #Extract First and Last Appearance
                r2_start = pilot2["First_Appearance"][counter]
                r2_end = pilot2["Last_Appearance"][counter]

                #Extract Day Month Year from Start Date
                r2_start_month = int(r2_start.split('/')[0])
                r2_start_day = int(r2_start.split('/')[1])
                r2_start_year = int(r2_start.split('/')[2])

                #Extract Day Month Year from End Date
                r2_end_month = int(r2_end.split('/')[0])
                r2_end_day = int(r2_end.split('/')[1])
                r2_end_year = int(r2_end.split('/')[2])

                #Set date range
                r2_range = Range(
                    start = datetime(
                        r2_start_year,
                        r2_start_month,
                        r2_start_day
                        ),
                    end = datetime(
                        r2_end_year,
                        r2_end_month,
                        r2_end_day
                        )
                    )

                if(r1_range.start <= r2_range.end and
                    r2_range.start <= r1_range.end):
```

```
                    competitors += 1

            counter += 1 #Go to next restaurant

        comp.append(competitors)
        print(rest_len+1, "of", pilot_len1, '\t',
            "Number of cuisine competitors for",
            r1_license, ":", competitors)

        rest_len +=1

#add competitors to dataframe
  pilot1.insert(
      pilot_col_to_save,
      "Number of Cuisine Competitors",
      comp
      )

  #save dataframe
  pilot1.to_csv(file_name1, encoding='utf_8_sig', index=False)
```

Appendix 8. Program to Count Convenience Competitors

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*

'''
Program to look for how many convenience competitors a restaurant has that are
within 369.95-meter radius and were in operation at the same time.
'''

from datetime import datetime
from collections import namedtuple
import pandas as pd
import geopy.distance

#Main Program
if __name__ == "__main__":
    Range = namedtuple('Range', ['start', 'end'])
    competitors = 0
    counter = 0
    rest_len = 0
    comp = []

    file_name = "~/Desktop/ Hong Kong Restaurants 2016-2018.csv"
    pilot = pd.read_csv(file_name, encoding='utf-8')

    pilot_len = pilot['First_Appearance'].count() #get number of rows
    pilot_col_to_save = pilot.shape[1] #Get number of columns

    #Convenience competitors:
    IC_file_name = "~/Desktop/Convenience Competitor List.csv"
    IC_list = pd.read_csv(IC_file_name, encoding='utf-8')

    IC_len = IC_list['Name'].count() #Get number of potential competitors

    while(rest_len < pilot_len):
        competitors = 0
        counter = 0

        rest_license = pilot["License_Number"][rest_len]

        r1_start = pilot['First_Appearance'][rest_len]
        r1_end = pilot['Last_Appearance'][rest_len]

        #Get coordinates of original restaurant
        r1_coord_lat = pilot['Latitude'][rest_len]
        r1_coord_lon = pilot['Longitude'][rest_len]
        r1_coords = (r1_coord_lat, r1_coord_lon)
```

```
#Extract Day Month Year from Start Date
r1_start_month = int(r1_start.split('/')[0])
r1_start_day = int(r1_start.split('/')[1])
r1_start_year = int(str("20") + str(r1_start.split('/')[2]))

#Extract Day Month Year from End Date
r1_end_month = int(r1_end.split('/')[0])
r1_end_day = int(r1_end.split('/')[1])
r1_end_year = int(str("20") + str(r1_end.split('/')[2]))


while(counter <= IC_len):
    overlap = 0
    m_dist = 0

    #Get coordinates of potential convenience competitor
    r2_coord_lat = IC_list['Latitude'][counter]
    r2_coord_lon = IC_list['Longitude'][counter]
    r2_coords = (r2_coord_lat, r2_coord_lon)

    #get distance between restaurants
    m_dist = geopy.distance.distance(r1_coords, r2_coords).m

    #Is the restaurant within a close distance to the indirect competitor
    if(m_dist < 369.95):
        #Extract First and Last Appearance
        r2_start = IC_list['First_Appearance'][counter]
        r2_end = IC_list['Last_Appearance'][counter]

        #Extract Day Month Year from Start Date
        r2_start_month = int(r2_start.split('/')[0])
        r2_start_day = int(r2_start.split('/')[1])
        r2_start_year = int(str("20") + str(r2_start.split('/')[2]))

        #Extract Day Month Year from End Date
        r2_end_month = int(r2_end.split('/')[0])
        r2_end_day = int(r2_end.split('/')[1])
        r2_end_year = int(str("20") + str(r2_end.split('/')[2]))

        r1_range = Range(
                start=datetime(
                      r1_start_year,
                      r1_start_month,
                      r1_start_day
                      ),
                end=datetime(
                      r1_end_year,
                      r1_end_month,
                      r1_end_day
```

```
                )
            )

        r2_range = Range(
            start=datetime(
                r2_start_year,
                r2_start_month,
                r2_start_day
                ),
            end=datetime(
                r2_end_year,
                r2_end_month,
                r2_end_day
                )
            )

        latest_start = max(r1_range.start, r2_range.start)
        earliest_end = min(r1_range.end, r2_range.end)
        delta = (earliest_end - latest_start).days + 1
        overlap = max(0, delta)

        #count how many competitors are overlapped
        if(overlap > 0):
            competitors +=1

      counter += 1 #Go to next competitor

    comp.append(competitors)
    print("Number of Convenience Competitors for", rest_license, ": ", competitors)
    rest_len +=1

#add competitors to dataframe
pilot.insert(pilot_col_to_save, "Number of Convenience Competitors", comp)

#save dataframe
pilot.to_csv(file_name, encoding='utf_8_sig', index=False)
```

Appendix 9. Program to Find if a restaurant is Close to an MTR Exit

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program looks at the individual restaurant's latitude and longitude and checks the
distance of every MTR exit. If the MTR exit is less than 369.95 meters away, it marks
the restaurant as 1 (close) and  then saves the new column to the existing file.
'''

import pandas as pd
import geopy.distance
import os

#Main Program
if __name__ == "__main__":
    MTR = 0
    counter = 0
    rest_len = 0
    MTR_X = []
    file_name = os.path.expanduser(
        "~/Desktop/Hong Kong Restaurants 2016-2018.csv"
        )

    #Read the csv files
    pilot = pd.read_csv(file_name, encoding='utf-8')
    MTR_file = os.path.expanduser(
        "~/Documents/Thesis/Data Lists/List of Coordinates of MTR Exits.csv"
        )

    MTR_list = pd.read_csv(MTR_file, encoding='utf-8')

    pilot_len = pilot['First_Appearance'].count() #get number of rows
    pilot_col_to_save = pilot.shape[1] #Get number of columns
    MTR_len = MTR_list['Station'].count()-1 #Get number of stations

    while(rest_len < pilot_len):
        counter = 0
        MTR = 0

        rest_license = pilot["License_Number"][rest_len]

        #Get restaurants coordinates
        r1_coord_lat = pilot['Latitude'][rest_len]
        r1_coord_lon = pilot['Longitude'][rest_len]
        r1_coords = (r1_coord_lat, r1_coord_lon)

        while(counter <= MTR_len):
```

```
            #Get MTR Exit Coordinate
            r2_coord_lat = MTR_list['Latitude'][counter]
            r2_coord_lon = MTR_list['Longitude'][counter]
            r2_coords = (r2_coord_lat, r2_coord_lon)

            #Get distance between restaurant and coordinate
            m_dist = geopy.distance.distance(r1_coords, r2_coords).m

            if(m_dist < 369.95):
                MTR = 1 #close to MTR Exit
                break #if one MTR station is close, go to next restaurant
            counter += 1
        MTR_X.append(MTR) #add 1 or 0 to the list

        print("Close to MTR Exit ", rest_license, ": ", MTR, " (1 = Close)")

        rest_len +=1

    #insert the 1 or 0 into the dataframe
    pilot.insert(pilot_col_to_save, "MTR Exit (1 = Close)", MTR_X)

    #save the csv file
    pilot.to_csv(file_name, encoding='utf_8_sig', index=False)
```

Appendix 10. Program to Find if a restaurant is within a Hotel

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program looks at the individual restaurant's latitude and longitude and checks the
distance of every Hotel. If the Hotel is less than 25 meters away, it marks the
restaurant as 1 (close) and  then saves the new column to the existing file.
'''

import pandas as pd
import os
import geopy.distance

#Main Program
if __name__ == "__main__":
    Hotel = 0
    counter = 0
    rest_len = 0
    Hotel_X = []
    file_name = os.path.expanduser(
        "~/Desktop/Hong Kong Restaurants 2016-2018.csv"
        )

    #Read the csv files
    pilot = pd.read_csv(file_name, encoding='utf-8')

    Hotel_file = os.path.expanduser(
        "~/Documents/Thesis/Data Lists/List of Hotels and Guesthouses.csv"
        )

    Hotel_list = pd.read_csv(Hotel_file, encoding='utf-8')

    pilot_len = pilot['First_Appearance'].count() #get number of rows
    pilot_col_to_save = pilot.shape[1] #Get number of columns
    Hotel_len = Hotel_list['To Search'].count()-1 #Get number of Hotels

    while(rest_len < pilot_len):
        counter = 0
        Hotel = 0

        rest_license = pilot["License_Number"][rest_len]

        #Get restaurants coordinates
        r1_coord_lat = pilot['Latitude'][rest_len]
        r1_coord_lon = pilot['Longitude'][rest_len]
        r1_coords = (r1_coord_lat, r1_coord_lon)
```

```
    while(counter <= Hotel_len):
        #Get Hotel Coordinate
        r2_coord_lat = Hotel_list['Latitude'][counter]
        r2_coord_lon = Hotel_list['Longitude'][counter]
        r2_coords = (r2_coord_lat, r2_coord_lon)

        #Get distance between restaurant and coordinate
        m_dist = geopy.distance.distance(r1_coords, r2_coords).m

        if(m_dist < 25):
            Hotel = 1 #close to Hotel
            break #if one Hotel is close, go to next restaurant

        counter += 1

    Hotel_X.append(Hotel) #add 1 or 0 to the list

    print("Close to Hotel ", rest_license, ": ", Hotel, " (1 = Close)")
    rest_len +=1

#insert the 1 or 0 into the dataframe
pilot.insert(pilot_col_to_save, "Hotel (1 = Close)", Hotel_X)

#save the csv file
pilot.to_csv(file_name, encoding = 'utf_8_sig', index = False)
```

Appendix 11. Program to Find if a restaurant is within a Tourist Attraction

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program looks at the individual restaurant's latitude and longitude and checks to see if
it sits within a tourist attraction. If the restaurant is within, it marks the value as 1 (close)
and then saves the new column to the existing file.
'''

import os
import json
import pandas as pd
from shapely.geometry import shape, Point, LinearRing
import geopy.distance

countx = 0
coutner = 0
Tourist_x = []

file_name = os.path.expanduser("~/Desktop/Hong Kong Restaurants 2016-2018.csv")

#Read the list of restaurants
List = pd.read_csv(file_name, encoding='utf-8')

tourist_file = os.path.expanduser(
    "~/Documents/Thesis/GeoJson/Tourist Area GeoJson/TouristList.csv"
    )

Tourist_List = pd.read_csv(tourist_file, encoding='utf-8')

List_len = List['First_Appearance'].count() #get number of rows
List_col_to_save = List.shape[1] #Get number of columns
Tourist_Len = Tourist_List['Tourist Area'].count() #Get number of Tourism Areas

#while the counter is less than the number of restaurants...
while(countx < List_len):
    counter = 0
    found = 0

    rest_license = List["License_Number"][countx] #get the license number

    r1_coord_lat = List['Latitude'][countx] #get the latitude of the restaurant
    r1_coord_lon = List['Longitude'][countx] #get the longitude of the restaurant
    r1_coords = (r1_coord_lat, r1_coord_lon)

    # construct point based on longitude/latitude
    point = Point(r1_coord_lon, r1_coord_lat)
```

```
    print("Checking", countx+1, "out of", List_len)
    while(counter < Tourist_Len):
        #get the Tourist number from list
        Tourist_Number = Tourist_List['Tourist Number'][counter]

        #Open GeoJson file of the TPU
        file_open = os.path.expanduser(
            '~/Documents/Thesis/GeoJson/Tourist Area GeoJson/' +
            str(Tourist_Number) +
            '.geojson')

        with open(file_open) as f:
            TourismLoad = json.load(f)

        polygon = shape(TourismLoad['geometry'])

        pol_ext = LinearRing(polygon.exterior.coords)
        d = pol_ext.project(point)
        p = pol_ext.interpolate(d)
        closest_point_coords = list(p.coords)[0]

        r2_coord_lon = closest_point_coords[0]
        r2_coord_lat = closest_point_coords[1]
        r2_coords = (r2_coord_lat, r2_coord_lon)

        m_dist = geopy.distance.distance(r1_coords, r2_coords).m

        if(m_dist < 369.95):
            found = 1
            Tourist_x.append(1) #add the TPU of where it was found to the list
            break

        counter += 1

    if(found == 0):
        Tourist_x.append(0)
    countx += 1

#insert the 1 or 0 into the dataframe
List.insert(List_col_to_save, "Tourist Attraction (1 = Close)", Tourist_x)

#save the csv file
List.to_csv(file_name, encoding='utf_8_sig', index=False)
```

Appendix 12. Program to Find if a restaurant is within a Shopping Mall

```
#!/usr/bin/env python3.7
# -*- coding: utf-8 -*-

'''
Program looks at the individual restaurant's latitude and longitude and checks to see if
it sits within a shopping mall. If the restaurant is within, it marks the value as 1 (close)
and then saves the new column to the existing file.
'''

import json
import os
import pandas as pd
from shapely.geometry import shape, Point

countx = 0
coutner = 0
Mall_x = []

file_name = "~/Desktop/Hong Kong Restaurants 2016-2018.csv"
file_name = os.path.expanduser(file_name)

#Read the list of restaurants
List = pd.read_csv(file_name, encoding='utf-8')

mall_file_path = os.path.expanduser(
    "~/Documents/Thesis/GeoJson/Shopping Area/Mall List.csv"
    )

Mall_List = pd.read_csv(mall_file_path, encoding='utf-8')

List_len = List['First_Appearance'].count() #get number of rows
List_col_to_save = List.shape[1] #Get number of columns
Mall_Len = Mall_List['Mall Area'].count() #Get number of Tourism Areas

print("Checking", List_len, "restaurants")

#while the counter is less than the number of restaurants...
while(countx < List_len):
    counter = 0
    found = 0
    print("Checking Restaurant", countx+1, "out of", List_len)
    rest_license = List["License_Number"][countx] #get the license number
    r1_coord_lat = List['Latitude'][countx] #get the latitude of the restaurant
    r1_coord_lon = List['Longitude'][countx] #get the longitude of the restaurant
    r1_coords = (r1_coord_lat, r1_coord_lon)
```

```
    # construct point based on longitude/latitude
    point = Point(r1_coord_lon, r1_coord_lat)

    while((counter < Mall_Len) and (found == 0)):
        #get the Mall number from list
        Mall_Number = Mall_List['Mall Number'][counter]

        #Open GeoJson file of the TPU
        file_open = os.path.expanduser(
            "~/Documents/Thesis/GeoJson/Shopping Area/" +
            str(Mall_Number) +
            ".geojson"
            )

        with open(file_open) as f:
            MallLoad = json.load(f)

        for feature in MallLoad['features']:
            polygon = shape(feature['geometry'])

            if polygon.contains(point):
                found = 1
                Mall_x.append(1)
                break
        counter += 1

    if(found == 0):
        Mall_x.append(0)
    countx += 1

print("Done, Total Restaurants:", List_len)

#insert the 1 or 0 into the dataframe
List.insert(List_col_to_save, "Inside Mall (1 = Yes)", Mall_x)
#save the csv file
List.to_csv(file_name, encoding='utf_8_sig', index=False)
```

Appendix 13. Program to Run the Logistic Regression

```
#!/usr/local/bin/python3.7
# -*- coding: utf-8 -*

import pandas as pd
import statsmodels.api as sm
import numpy as np
from sklearn.metrics import roc_curve, roc_auc_score
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure

df = pd.read_csv(
    "~/Desktop/csv files/Entire Hong Kong Restaurants - 2016-18 - LR Min.csv"
    )

train_cols = df.columns[1:] #separate x variables from the y var.

#fit model with y and x variables
logit = sm.Logit(df['OpStatus'], sm.add_constant(df[train_cols]))
#use White (1980) for robust standard errors
result = logit.fit(cov_type='HC0')

print(result.summary2())
print(result.pred_table())

## Calculate the AIC and BIC values
y = df['OpStatus']
y2 = result.predict(sm.add_constant(df[train_cols]))
n = len(y)
k = len(train_cols) + 1

#Calculate the log likelihood
LogL = sum(y * np.log(y2) + (1 - y) * np.log(1 - y2))

AIC = -2 * (LogL - k)
BIC = -2 * LogL + np.log(n) * k

print("AIC:", AIC)
print("BIC:", BIC)

##Calculate the ROC score
auc = roc_auc_score(df['OpStatus'], y2)
print("AUC: %.6f" %auc)

##Plot the Roc Curve
FPR, TPR, _ = roc_curve(df['OpStatus'], y2)
figure(num=1, figsize=(7, 7), edgecolor = 'w')
plt.plot([0,1], [0,1], '--', color='k') #create a diagonal dashed linestyle
```

```
plt.plot(FPR, TPR, marker='.', ms=0.05, color='r') #create the roc curve
plt.title('Logistic Regression ROC Curve')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.show()
```

Appendix 14. Program to Run the Artificial Neural Network

```
#!/usr/local/bin/python3.7
# -*- coding: utf-8 -*

import os
from os import system, name
import time
import pandas as pd
import json
import numpy as np
from matplotlib import pyplot as plt
from matplotlib.pyplot import figure
from keras import optimizers
from keras.models import Sequential #used for model building
from keras.layers import Dense #used for creating layers
from keras.wrappers.scikit_learn import KerasClassifier #To group classifications
from sklearn.model_selection import StratifiedKFold #Stratified K Fold
from sklearn.model_selection import GridSearchCV #Grid Searching
from sklearn.metrics import confusion_matrix #import confusion matrix
from sklearn.metrics import roc_curve, roc_auc_score #import ROC curve tools
from itertools import product

os.environ['TF_CPP_MIN_LOG_LEVEL'] = '2'

neurons2 = 0
learn_rate2 = 0
decay2 = 0
momentum2 = 0
nesterov2 = True
epochs2 = 0
counterx = 1

def ANN_model(neurons, learn_rate, decay, momentum, nesterov, init, epochs, IV):
    global neurons2, learn_rate2, decay2, momentum2
    global nesterov2, epochs2, counterx

    if((neurons != neurons2) or (learn_rate != learn_rate2) or
        (decay != decay2) or (momentum != momentum2) or
        (nesterov != nesterov2) or (epochs != epochs2)):

        if counterx <= testcomblen:
            print(counterx, "of", testcomblen, '\t',
                "Neurons:", neurons,
                "-- Learn Rate:", learn_rate,
                "-- Decay:", decay,
                "-- Momentum:", momentum,
                "-- Nesterov:", nesterov,
                "-- Epochs:", epochs)
```

```
        neurons2 = neurons
        learn_rate2 = learn_rate
        decay2 = decay
        momentum2 = momentum
        nesterov2 = nesterov
        epochs2 = epochs
        counterx += 1

##Create the model
#make a sequential model
model = Sequential()

#Create the hidden layer
model.add(Dense(
    neurons,
    kernel_initializer = init,
    use_bias = False,
    activation = 'relu',
    input_dim=IV)
    )

#Create the output layer
model.add(Dense(1,
    kernel_initializer = init,
    use_bias = False,
    activation = 'sigmoid')
    )

#Compile Optimizer
optimizer = optimizers.SGD(
    lr = learn_rate,
    momentum = momentum,
    decay = decay,
    nesterov = nesterov
    )

#Compile Model
model.compile(
    optimizer = optimizer,
    loss = 'binary_crossentropy',
    metrics = ['accuracy']
    )

#return model
return model
```

```
if __name__ == "__main__":
    tempmean = []
    tempstd = []
    IV = []

    #Open the data file
    df = pd.read_csv(
        "~/Desktop/csv files/Entire Hong Kong Restaurants - 2016-18 - ANN.csv"
        )

    #Directory on where to save model on completion
    desktop = os.path.expanduser("~/desktop/ANN Results/")
    if not os.path.exists(desktop): #if folder does not exist, create it
        os.makedirs(desktop)


    ##File paths
    final_model_params              = desktop + "final_model_params.json"
    final_model_json                = desktop + "final_model_json.json"
    final_model_weights             = desktop + "final_model_weights.h5"
    final_model_complete            = desktop + "final_model_complete.h5"
    final_model_layer_1_weights     = desktop + "final_model_layer_1_weights.csv"
    final_model_layer_2_weights     = desktop + "final_model_layer_2_weights.csv"
    final_model_mean_stdev          = desktop + "final_model_mean_stdev.csv"
    final_model_probabilities       = desktop + "final_model_probabilities.csv"

    column_headings = list(df.columns.values) #get the column headers


    ##Get Mean and Standard Deviation of Dataset
    tempmean.clear()
    tempstd.clear()

    for k in column_headings:
        tempmean.append(df[k].mean())
        tempstd.append(df[k].std())

    d = {
        'Variable' : column_headings,
        'Mean' : tempmean,
        'StDev' : tempstd
        }

    Mean_Stdev = pd.DataFrame(d)
```

```
## Splitting the data into IV and DV
#Split into independent variable, ccopy all comlumns except Outcome
X = df.loc[:, df.columns != 'OpStatus']

#Split off the dependent variable, copy only entire column of outcome
y = df.loc[:, 'OpStatus']

num_X = X.shape[1] #Get number of independent varaibles
IV.append(num_X)


##Get start time
start_time = time.time()


##Find Ideal Parameters
estimator = KerasClassifier(build_fn = ANN_model, verbose = 0)

init            = ['glorot_uniform'] #how to set the initial random weights
nesterov        = [True, False]
neurons         = [24, 12, 25, 49] #hidden layer set for n, n/2, n+1, 2n+1
learn_rate      = [0.1, 0.01, 0.001, 0.005]
momentum        = [0.1, 0.5, 0.9, 0.99]
decay           = [0.0, 0.1, 0.01, 0.001]
epochs          = [1, 10, 50, 100, 500]

#Find the number of different combinations
testcomb = product(neurons, learn_rate, momentum, nesterov, decay, epochs)
testcomblen = len(list(testcomb))

print("Running Stratified K Fold Testing...")
param_grid = dict(
    neurons = neurons,
    decay = decay,
    learn_rate = learn_rate,
    momentum = momentum,
    nesterov = nesterov,
    epochs = epochs,
    init = init,
    IV = IV
    )

SKfold= StratifiedKFold(
    n_splits = 10,
    shuffle = True,
    random_state = None
    )
```

```
#cv cross-validation generator to use a 10x StatifiedKFold
grid = GridSearchCV(
    estimator = estimator,
    param_grid = param_grid,
    cv = SKfold
    )

#Find the data to the model
grid_result = grid.fit(X, y)


##Get finished time
finish_time = time.time()
total_time = finish_time - start_time
minutes, seconds = divmod(total_time, 60)
hours, minutes = divmod(minutes, 60)
days, hours = divmod(hours, 24)

print("\n" + "Finished in %.0f" %days + " days %.0f" %hours +
    " hours %.0f" %minutes + " minutes %.0f" %seconds + " seconds")

print(
    "\n" +
    "Best Parameter Accuracy: %.2f%%" %(grid_result.best_score_ * 100)
    )

bp_dict = grid_result.best_params_

bp_init         = bp_dict["init"]
bp_neurons      = bp_dict["neurons"]
bp_learn_rate   = bp_dict["learn_rate"]
bp_momentum     = bp_dict["momentum"]
bp_decay        = bp_dict["decay"]
bp_nesterov     = bp_dict["nesterov"]
bp_epochs       = bp_dict["epochs"]
bp_IV           = num_X

print("\n" + "Top Parameters For Final Model:")
print("  Neurons" + '\t' + ": " + str(bp_neurons))
print("  Learn Rate" + '\t' + ": " + str(bp_learn_rate))
print("  Momentum" + '\t' + ": " + str(bp_momentum))
print("  Decay" + '\t' + '\t' + ": " + str(bp_decay))
print("  Nesterov" + '\t' + ": " + str(bp_nesterov))
print("  Epochs" + '\t' + ": " + str(bp_epochs))
```

```
##Set final model parameters based on testing and run model
sk_params = {
    'init': bp_init,
    'epochs': bp_epochs,
    'neurons': bp_neurons,
    'learn_rate': bp_learn_rate,
    'decay': bp_decay,
    'momentum':bp_momentum,
    'nesterov':bp_nesterov,
    'IV':bp_IV
    }

#Set the parameters based off of the best found
final_estimator = KerasClassifier(
    build_fn = ANN_model,
    **sk_params,
    verbose = 0
    )

#Fit the final model based off of the X and y
history = final_estimator.fit(X, y)
scores = final_estimator.model.evaluate(X, y, verbose = 0)


##Get the final weights used
#place weights in numpy array
weights_layer_1 = final_estimator.model.layers[0].get_weights()[0]
weights_layer_2 = final_estimator.model.layers[1].get_weights()[0]

#convert numpy array to DataFrame
weights_layer_1 = pd.DataFrame(weights_layer_1)
weights_layer_2 = pd.DataFrame(weights_layer_2)


##Get the actual scores of the model and the classes
#actual predicted scores of the DV based on the IV
pred_actual = final_estimator.model.predict(X)
#place predicted results in a DataFrame
pred_actual = pd.DataFrame(pred_actual, columns=['Predicted Probability'])

#classes of the predicted scores of the DV based on the IV
pred_class = final_estimator.model.predict_classes(X)
#place predicted results in a DataFrame
pred_class = pd.DataFrame(pred_class)

print("\n" + "Model Evaluation: %.2f%%" % (scores[1]*100))
```

```
##Save files
#JSON files
b_file = open(final_model_params, "w")
json.dump(sk_params, b_file)
b_file.close()

final_json = final_estimator.model.to_json()
with open(final_model_json, "w") as json_file:
    json_file.write(final_json)

#H5 files
final_estimator.model.save_weights(final_model_weights)
final_estimator.model.save(final_model_complete)

#CSV files
weights_layer_1.to_csv(final_model_layer_1_weights, index = False)
weights_layer_2.to_csv(final_model_layer_2_weights, index = False)
Mean_Stdev.to_csv(final_model_mean_stdev, index = False)
pred_actual.to_csv(final_model_probabilities, index = False)

print("\n" + "Saved 8 files." + "\n")


##Calculate AIC & BIC
n = len(y)
k = IV[0] * bp_neurons + bp_neurons
y2 = pred_actual.T.squeeze()

LogL = sum(y * np.log(y2) + (1 - y) * np.log(1 - y2))
AIC = -2 * (LogL - k)
BIC = -2 * LogL + np.log(n) * k

print("AIC:", AIC)
print("BIC:", BIC, "\n")

##Confusion Matrix
tn, fp, fn, tp = confusion_matrix(y, pred_class).ravel()

print("Confusion Matrix")
print("\t", "\t", "Predicted")
print("\t", "\t", "0", "\t", "1", "\t", "Total")
print("Actual", "\t", "0", "\t", tn, "\t", fp, "\t", tn+fp)
print("\t", "1", "\t", fn, "\t", tp, "\t", fn+tp)
print("\t", "Total", "\t", tn+fn, "\t", fp+tp)

##Calculate the AUROC score
auc = roc_auc_score(y, pred_actual)

print('\n' + "AUC: %.6f" %auc)
```

```
##Create ROC curve
FPR, TPR, _ = roc_curve(y, pred_actual) #get the rates for the curve
figure(num=1, figsize=(7, 7), edgecolor = 'w')
plt.plot(FPR, TPR, marker='.', ms=0.05, color='r')
plt.plot([0,1], [0,1], '--', color='black') #create a diagonal dashed line
plt.title('Artificial Neural Network - ROC Curve')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.show()
```

Appendix 15. Program to Map the Restaurants

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import geopandas as gpd
from shapely.geometry import Point#, Polygon
import os
import platform

plat = platform.system()

if plat == "Windows":
    director =
        "C:\\Users\\SHTM\\Documents\\Thesis\\Shape Files\\Hong Kong\\"
    Input_File_Shenzhen =
        "C:\\Users\\SHTM\\Documents\\Thesis\\GeoJson\\Shenzhen\\Shenzhen.shp"
    df2 =
        pd.read_csv("C:\\Users\\SHTM\\Documents\\Thesis\\Data Mapping\\Hong
        Kong Restaurants.csv", encoding='utf-8')
else:
    director =
        "/Users/guyllewellyn/Google Drive/Thesis/Shape Files/Hong Kong/"
    Input_File_Shenzhen =
        "/Users/guyllewellyn/Google Drive/Thesis/GeoJson/Shenzhen/Shenzhen.shp"
    df2 =
        pd.read_csv("/Users/guyllewellyn/Google Drive/Thesis/Data Mapping/Hong
        Kong Restaurants.csv", encoding='utf-8')

all_files = []
crs = [['init': 'epsg:4326']]

TPU_111 = director + "111.shp"
TPU_111 = director + '111.shp'
TPU_112 = director + '112.shp'
TPU_113 = director + '113.shp'
TPU_114 = director + '114.shp'
TPU_115 = director + '115.shp'
TPU_116 = director + '116.shp'
TPU_121_123_124 = director + '121_123_124.shp'
TPU_122 = director + '122.shp'
TPU_131 = director + '131.shp'
TPU_132 = director + '132.shp'
TPU_133 = director + '133.shp'
TPU_134 = director + '134.shp'
TPU_135 = director + '135.shp'
```

```
TPU_140 = director + '140.shp'
TPU_141 = director + '141.shp'
TPU_142 = director + '142.shp'
TPU_143 = director + '143.shp'
TPU_144 = director + '144.shp'
TPU_145 = director + '145.shp'
TPU_146_147 = director + '146_147.shp'
TPU_148 = director + '148.shp'
TPU_149 = director + '149.shp'
TPU_151 = director + '151.shp'
TPU_152 = director + '152.shp'
TPU_153 = director + '153.shp'
TPU_154 = director + '154.shp'
TPU_155 = director + '155.shp'
TPU_156_158 = director + '156_158.shp'
TPU_157 = director + '157.shp'
TPU_161 = director + '161.shp'
TPU_162 = director + '162.shp'
TPU_163 = director + '163.shp'
TPU_164_165 = director + '164_165.shp'
TPU_166 = director + '166.shp'
TPU_167 = director + '167.shp'
TPU_171 = director + '171.shp'
TPU_172 = director + '172.shp'
TPU_173 = director + '173.shp'
TPU_174 = director + '174.shp'
TPU_175_176 = director + '175_176.shp'
TPU_181_182 = director + '181_182.shp'
TPU_183_184 = director + '183_184.shp'
TPU_190_192_194 = director + '190_192_194.shp'
TPU_191 = director + '191.shp'
TPU_193_195_198 = director + '193_195_198.shp'
TPU_196 = director + '196.shp'
TPU_197 = director + '197.shp'
TPU_211 = director + '211.shp'
TPU_212 = director + '212.shp'
TPU_213_215_216 = director + '213_215_216.shp'
TPU_214 = director + '214.shp'
TPU_217 = director + '217.shp'
TPU_220 = director + '220.shp'
TPU_221 = director + '221.shp'
TPU_222 = director + '222.shp'
TPU_225 = director + '225.shp'
TPU_226 = director + '226.shp'
TPU_227 = director + '227.shp'
TPU_228 = director + '228.shp'
TPU_229 = director + '229.shp'
TPU_231 = director + '231.shp'
TPU_232 = director + '232.shp'
```

```
TPU_233 = director + '233.shp'
TPU_234 = director + '234.shp'
TPU_235 = director + '235.shp'
TPU_236 = director + '236.shp'
TPU_237 = director + '237.shp'
TPU_241 = director + '241.shp'
TPU_242 = director + '242.shp'
TPU_243 = director + '243.shp'
TPU_244 = director + '244.shp'
TPU_245 = director + '245.shp'
TPU_246 = director + '246.shp'
TPU_247 = director + '247.shp'
TPU_251_256 = director + '251_256.shp'
TPU_252 = director + '252.shp'
TPU_253 = director + '253.shp'
TPU_254 = director + '254.shp'
TPU_255_269 = director + '255_269.shp'
TPU_260 = director + '260.shp'
TPU_261 = director + '261.shp'
TPU_262 = director + '262.shp'
TPU_263 = director + '263.shp'
TPU_264 = director + '264.shp'
TPU_265 = director + '265.shp'
TPU_266 = director + '266.shp'
TPU_267 = director + '267.shp'
TPU_268 = director + '268.shp'
TPU_271 = director + '271.shp'
TPU_272 = director + '272.shp'
TPU_280 = director + '280.shp'
TPU_281 = director + '281.shp'
TPU_282 = director + '282.shp'
TPU_283 = director + '283.shp'
TPU_284 = director + '284.shp'
TPU_285 = director + '285.shp'
TPU_286 = director + '286.shp'
TPU_287 = director + '287.shp'
TPU_288_289 = director + '288_289.shp'
TPU_290 = director + '290.shp'
TPU_291 = director + '291.shp'
TPU_292 = director + '292.shp'
TPU_293_296 = director + '293_296.shp'
TPU_294 = director + '294.shp'
TPU_295 = director + '295.shp'
TPU_297 = director + '297.shp'
TPU_298 = director + '298.shp'
TPU_310_321 = director + '310_321.shp'
TPU_320_324_329 = director + '320_324_329.shp'
TPU_322 = director + '322.shp'
TPU_323 = director + '323.shp'
```

```
TPU_325 = director + '325.shp'
TPU_326 = director + '326.shp'
TPU_327 = director + '327.shp'
TPU_328 = director + '328.shp'
TPU_331_332_333_334_336_340 = director + '331_332_333_334_336_340.shp'
TPU_335 = director + '335.shp'
TPU_350 = director + '350.shp'
TPU_351 = director + '351.shp'
TPU_411_412_413_414_415_416_427 =
    director + '411_412_413_414_415_416_427.shp'
TPU_421_422 = director + '421_422.shp'
TPU_423_428 = director + '423_428.shp'
TPU_424 = director + '424.shp'
TPU_425 = director + '425.shp'
TPU_426 = director + '426.shp'
TPU_431_432_433_434 = director + '431_432_433_434.shp'
TPU_441 = director + '441.shp'
TPU_442 = director + '442.shp'
TPU_510 = director + '510.shp'
TPU_511 = director + '511.shp'
TPU_512 = director + '512.shp'
TPU_513 = director + '513.shp'
TPU_514 = director + '514.shp'
TPU_515 = director + '515.shp'
TPU_516 = director + '516.shp'
TPU_517 = director + '517.shp'
TPU_518 = director + '518.shp'
TPU_519 = director + '519.shp'
TPU_521 = director + '521.shp'
TPU_522 = director + '522.shp'
TPU_523 = director + '523.shp'
TPU_524 = director + '524.shp'
TPU_525 = director + '525.shp'
TPU_526 = director + '526.shp'
TPU_527 = director + '527.shp'
TPU_528 = director + '528.shp'
TPU_529 = director + '529.shp'
TPU_531 = director + '531.shp'
TPU_532 = director + '532.shp'
TPU_533 = director + '533.shp'
TPU_541 = director + '541.shp'
TPU_542 = director + '542.shp'
TPU_543_546 = director + '543_546.shp'
TPU_544 = director + '544.shp'
TPU_545 = director + '545.shp'
TPU_547 = director + '547.shp'
TPU_548 = director + '548.shp'
TPU_610_621_632 = director + '610_621_632.shp'
TPU_620_622_641 = director + '620_622_641.shp'
```

```
TPU_623 = director + '623.shp'
TPU_624 = director + '624.shp'
TPU_625 = director + '625.shp'
TPU_626 = director + '626.shp'
TPU_627 = director + '627.shp'
TPU_628 = director + '628.shp'
TPU_629 = director + '629.shp'
TPU_631_633 = director + '631_633.shp'
TPU_634 = director + '634.shp'
TPU_642 = director + '642.shp'
TPU_651_652_653 = director + '651_652_653.shp'
TPU_711_712_721_728 = director + '711_712_721_728.shp'
TPU_720 = director + '720.shp'
TPU_722_727 = director + '722_727.shp'
TPU_723 = director + '723.shp'
TPU_724 = director + '724.shp'
TPU_725 = director + '725.shp'
TPU_726 = director + '726.shp'
TPU_729 = director + '729.shp'
TPU_731_733_754 = director + '731_733_754.shp'
TPU_732_751_753 = director + '732_751_753.shp'
TPU_741_742_743_744 = director + '741_742_743_744.shp'
TPU_755 = director + '755.shp'
TPU_756_761_762 = director + '756_761_762.shp'
TPU_757 = director + '757.shp'
TPU_758 = director + '758.shp'
TPU_759 = director + '759.shp'
TPU_811_812_813_814_815 = director + '811_812_813_814_815.shp'
TPU_820 = director + '820.shp'
TPU_821 = director + '821.shp'
TPU_822 = director + '822.shp'
TPU_823 = director + '823.shp'
TPU_824_829 = director + '824_829.shp'
TPU_825 = director + '825.shp'
TPU_826_828 = director + '826_828.shp'
TPU_827 = director + '827.shp'
TPU_831 = director + '831.shp'
TPU_832_834 = director + '832_834.shp'
TPU_833 = director + '833.shp'
TPU_835 = director + '835.shp'
TPU_836 = director + '836.shp'
TPU_837 = director + '837.shp'
TPU_838 = director + '838.shp'
TPU_839 = director + '839.shp'
TPU_911_912_913 = director + '911_912_913.shp'
TPU_920 = director + '920.shp'
TPU_931_933 = director + '931_933.shp'
TPU_932_934 = director + '932_934.shp'
TPU_941_942_943 = director + '941_942_943.shp'
```

```
TPU_944 = director + '944.shp'
TPU_950_951 = director + '950_951.shp'
TPU_961_962_963 = director + '961_962_963.shp'
TPU_971_972_973_974 = director + '971_972_973_974.shp'
TPU_975 = director + '975.shp'
TPU_976 = director + '976.shp'

Map_Shenzhen = gpd.read_file(Input_File_Shenzhen)

Map_TPU_111 = gpd.read_file(TPU_111)
Map_TPU_112 = gpd.read_file(TPU_112)
Map_TPU_113 = gpd.read_file(TPU_113)
Map_TPU_114 = gpd.read_file(TPU_114)
Map_TPU_115 = gpd.read_file(TPU_115)
Map_TPU_116 = gpd.read_file(TPU_116)
Map_TPU_121_123_124 = gpd.read_file(TPU_121_123_124)
Map_TPU_122 = gpd.read_file(TPU_122)
Map_TPU_131 = gpd.read_file(TPU_131)
Map_TPU_132 = gpd.read_file(TPU_132)
Map_TPU_133 = gpd.read_file(TPU_133)
Map_TPU_134 = gpd.read_file(TPU_134)
Map_TPU_135 = gpd.read_file(TPU_135)
Map_TPU_140 = gpd.read_file(TPU_140)
Map_TPU_141 = gpd.read_file(TPU_141)
Map_TPU_142 = gpd.read_file(TPU_142)
Map_TPU_143 = gpd.read_file(TPU_143)
Map_TPU_144 = gpd.read_file(TPU_144)
Map_TPU_145 = gpd.read_file(TPU_145)
Map_TPU_146_147 = gpd.read_file(TPU_146_147)
Map_TPU_148 = gpd.read_file(TPU_148)
Map_TPU_149 = gpd.read_file(TPU_149)
Map_TPU_151 = gpd.read_file(TPU_151)
Map_TPU_152 = gpd.read_file(TPU_152)
Map_TPU_153 = gpd.read_file(TPU_153)
Map_TPU_154 = gpd.read_file(TPU_154)
Map_TPU_155 = gpd.read_file(TPU_155)
Map_TPU_156_158 = gpd.read_file(TPU_156_158)
Map_TPU_157 = gpd.read_file(TPU_157)
Map_TPU_161 = gpd.read_file(TPU_161)
Map_TPU_162 = gpd.read_file(TPU_162)
Map_TPU_163 = gpd.read_file(TPU_163)
Map_TPU_164_165 = gpd.read_file(TPU_164_165)
Map_TPU_166 = gpd.read_file(TPU_166)
Map_TPU_167 = gpd.read_file(TPU_167)
Map_TPU_171 = gpd.read_file(TPU_171)
Map_TPU_172 = gpd.read_file(TPU_172)
Map_TPU_173 = gpd.read_file(TPU_173)
Map_TPU_174 = gpd.read_file(TPU_174)
Map_TPU_175_176 = gpd.read_file(TPU_175_176)
```

```
Map_TPU_181_182 = gpd.read_file(TPU_181_182)
Map_TPU_183_184 = gpd.read_file(TPU_183_184)
Map_TPU_190_192_194 = gpd.read_file(TPU_190_192_194)
Map_TPU_191 = gpd.read_file(TPU_191)
Map_TPU_193_195_198 = gpd.read_file(TPU_193_195_198)
Map_TPU_196 = gpd.read_file(TPU_196)
Map_TPU_197 = gpd.read_file(TPU_197)
Map_TPU_211 = gpd.read_file(TPU_211)
Map_TPU_212 = gpd.read_file(TPU_212)
Map_TPU_213_215_216 = gpd.read_file(TPU_213_215_216)
Map_TPU_214 = gpd.read_file(TPU_214)
Map_TPU_217 = gpd.read_file(TPU_217)
Map_TPU_220 = gpd.read_file(TPU_220)
Map_TPU_221 = gpd.read_file(TPU_221)
Map_TPU_222 = gpd.read_file(TPU_222)
Map_TPU_225 = gpd.read_file(TPU_225)
Map_TPU_226 = gpd.read_file(TPU_226)
Map_TPU_227 = gpd.read_file(TPU_227)
Map_TPU_228 = gpd.read_file(TPU_228)
Map_TPU_229 = gpd.read_file(TPU_229)
Map_TPU_231 = gpd.read_file(TPU_231)
Map_TPU_232 = gpd.read_file(TPU_232)
Map_TPU_233 = gpd.read_file(TPU_233)
Map_TPU_234 = gpd.read_file(TPU_234)
Map_TPU_235 = gpd.read_file(TPU_235)
Map_TPU_236 = gpd.read_file(TPU_236)
Map_TPU_237 = gpd.read_file(TPU_237)
Map_TPU_241 = gpd.read_file(TPU_241)
Map_TPU_242 = gpd.read_file(TPU_242)
Map_TPU_243 = gpd.read_file(TPU_243)
Map_TPU_244 = gpd.read_file(TPU_244)
Map_TPU_245 = gpd.read_file(TPU_245)
Map_TPU_246 = gpd.read_file(TPU_246)
Map_TPU_247 = gpd.read_file(TPU_247)
Map_TPU_251_256 = gpd.read_file(TPU_251_256)
Map_TPU_252 = gpd.read_file(TPU_252)
Map_TPU_253 = gpd.read_file(TPU_253)
Map_TPU_254 = gpd.read_file(TPU_254)
Map_TPU_255_269 = gpd.read_file(TPU_255_269)
Map_TPU_260 = gpd.read_file(TPU_260)
Map_TPU_261 = gpd.read_file(TPU_261)
Map_TPU_262 = gpd.read_file(TPU_262)
Map_TPU_263 = gpd.read_file(TPU_263)
Map_TPU_264 = gpd.read_file(TPU_264)
Map_TPU_265 = gpd.read_file(TPU_265)
Map_TPU_266 = gpd.read_file(TPU_266)
Map_TPU_267 = gpd.read_file(TPU_267)
Map_TPU_268 = gpd.read_file(TPU_268)
Map_TPU_271 = gpd.read_file(TPU_271)
```

```
Map_TPU_272 = gpd.read_file(TPU_272)
Map_TPU_280 = gpd.read_file(TPU_280)
Map_TPU_281 = gpd.read_file(TPU_281)
Map_TPU_282 = gpd.read_file(TPU_282)
Map_TPU_283 = gpd.read_file(TPU_283)
Map_TPU_284 = gpd.read_file(TPU_284)
Map_TPU_285 = gpd.read_file(TPU_285)
Map_TPU_286 = gpd.read_file(TPU_286)
Map_TPU_287 = gpd.read_file(TPU_287)
Map_TPU_288_289 = gpd.read_file(TPU_288_289)
Map_TPU_290 = gpd.read_file(TPU_290)
Map_TPU_291 = gpd.read_file(TPU_291)
Map_TPU_292 = gpd.read_file(TPU_292)
Map_TPU_293_296 = gpd.read_file(TPU_293_296)
Map_TPU_294 = gpd.read_file(TPU_294)
Map_TPU_295 = gpd.read_file(TPU_295)
Map_TPU_297 = gpd.read_file(TPU_297)
Map_TPU_298 = gpd.read_file(TPU_298)
Map_TPU_310_321 = gpd.read_file(TPU_310_321)
Map_TPU_320_324_329 = gpd.read_file(TPU_320_324_329)
Map_TPU_322 = gpd.read_file(TPU_322)
Map_TPU_323 = gpd.read_file(TPU_323)
Map_TPU_325 = gpd.read_file(TPU_325)
Map_TPU_326 = gpd.read_file(TPU_326)
Map_TPU_327 = gpd.read_file(TPU_327)
Map_TPU_328 = gpd.read_file(TPU_328)
Map_TPU_331_332_333_334_336_340 =
    gpd.read_file(TPU_331_332_333_334_336_340)
Map_TPU_335 = gpd.read_file(TPU_335)
Map_TPU_350 = gpd.read_file(TPU_350)
Map_TPU_351 = gpd.read_file(TPU_351)
Map_TPU_411_412_413_414_415_416_427 =
    gpd.read_file(TPU_411_412_413_414_415_416_427)
Map_TPU_421_422 = gpd.read_file(TPU_421_422)
Map_TPU_423_428 = gpd.read_file(TPU_423_428)
Map_TPU_424 = gpd.read_file(TPU_424)
Map_TPU_425 = gpd.read_file(TPU_425)
Map_TPU_426 = gpd.read_file(TPU_426)
Map_TPU_431_432_433_434 = gpd.read_file(TPU_431_432_433_434)
Map_TPU_441 = gpd.read_file(TPU_441)
Map_TPU_442 = gpd.read_file(TPU_442)
Map_TPU_510 = gpd.read_file(TPU_510)
Map_TPU_511 = gpd.read_file(TPU_511)
Map_TPU_512 = gpd.read_file(TPU_512)
Map_TPU_513 = gpd.read_file(TPU_513)
Map_TPU_514 = gpd.read_file(TPU_514)
Map_TPU_515 = gpd.read_file(TPU_515)
Map_TPU_516 = gpd.read_file(TPU_516)
Map_TPU_517 = gpd.read_file(TPU_517)
```

```
Map_TPU_518 = gpd.read_file(TPU_518)
Map_TPU_519 = gpd.read_file(TPU_519)
Map_TPU_521 = gpd.read_file(TPU_521)
Map_TPU_522 = gpd.read_file(TPU_522)
Map_TPU_523 = gpd.read_file(TPU_523)
Map_TPU_524 = gpd.read_file(TPU_524)
Map_TPU_525 = gpd.read_file(TPU_525)
Map_TPU_526 = gpd.read_file(TPU_526)
Map_TPU_527 = gpd.read_file(TPU_527)
Map_TPU_528 = gpd.read_file(TPU_528)
Map_TPU_529 = gpd.read_file(TPU_529)
Map_TPU_531 = gpd.read_file(TPU_531)
Map_TPU_532 = gpd.read_file(TPU_532)
Map_TPU_533 = gpd.read_file(TPU_533)
Map_TPU_541 = gpd.read_file(TPU_541)
Map_TPU_542 = gpd.read_file(TPU_542)
Map_TPU_543_546 = gpd.read_file(TPU_543_546)
Map_TPU_544 = gpd.read_file(TPU_544)
Map_TPU_545 = gpd.read_file(TPU_545)
Map_TPU_547 = gpd.read_file(TPU_547)
Map_TPU_548 = gpd.read_file(TPU_548)
Map_TPU_610_621_632 = gpd.read_file(TPU_610_621_632)
Map_TPU_620_622_641 = gpd.read_file(TPU_620_622_641)
Map_TPU_623 = gpd.read_file(TPU_623)
Map_TPU_624 = gpd.read_file(TPU_624)
Map_TPU_625 = gpd.read_file(TPU_625)
Map_TPU_626 = gpd.read_file(TPU_626)
Map_TPU_627 = gpd.read_file(TPU_627)
Map_TPU_628 = gpd.read_file(TPU_628)
Map_TPU_629 = gpd.read_file(TPU_629)
Map_TPU_631_633 = gpd.read_file(TPU_631_633)
Map_TPU_634 = gpd.read_file(TPU_634)
Map_TPU_642 = gpd.read_file(TPU_642)
Map_TPU_651_652_653 = gpd.read_file(TPU_651_652_653)
Map_TPU_711_712_721_728 = gpd.read_file(TPU_711_712_721_728)
Map_TPU_720 = gpd.read_file(TPU_720)
Map_TPU_722_727 = gpd.read_file(TPU_722_727)
Map_TPU_723 = gpd.read_file(TPU_723)
Map_TPU_724 = gpd.read_file(TPU_724)
Map_TPU_725 = gpd.read_file(TPU_725)
Map_TPU_726 = gpd.read_file(TPU_726)
Map_TPU_729 = gpd.read_file(TPU_729)
Map_TPU_731_733_754 = gpd.read_file(TPU_731_733_754)
Map_TPU_732_751_753 = gpd.read_file(TPU_732_751_753)
Map_TPU_741_742_743_744 = gpd.read_file(TPU_741_742_743_744)
Map_TPU_755 = gpd.read_file(TPU_755)
Map_TPU_756_761_762 = gpd.read_file(TPU_756_761_762)
Map_TPU_757 = gpd.read_file(TPU_757)
Map_TPU_758 = gpd.read_file(TPU_758)
```

```
Map_TPU_759 = gpd.read_file(TPU_759)
Map_TPU_811_812_813_814_815 = gpd.read_file(TPU_811_812_813_814_815)
Map_TPU_820 = gpd.read_file(TPU_820)
Map_TPU_821 = gpd.read_file(TPU_821)
Map_TPU_822 = gpd.read_file(TPU_822)
Map_TPU_823 = gpd.read_file(TPU_823)
Map_TPU_824_829 = gpd.read_file(TPU_824_829)
Map_TPU_825 = gpd.read_file(TPU_825)
Map_TPU_826_828 = gpd.read_file(TPU_826_828)
Map_TPU_827 = gpd.read_file(TPU_827)
Map_TPU_831 = gpd.read_file(TPU_831)
Map_TPU_832_834 = gpd.read_file(TPU_832_834)
Map_TPU_833 = gpd.read_file(TPU_833)
Map_TPU_835 = gpd.read_file(TPU_835)
Map_TPU_836 = gpd.read_file(TPU_836)
Map_TPU_837 = gpd.read_file(TPU_837)
Map_TPU_838 = gpd.read_file(TPU_838)
Map_TPU_839 = gpd.read_file(TPU_839)
Map_TPU_911_912_913 = gpd.read_file(TPU_911_912_913)
Map_TPU_920 = gpd.read_file(TPU_920)
Map_TPU_931_933 = gpd.read_file(TPU_931_933)
Map_TPU_932_934 = gpd.read_file(TPU_932_934)
Map_TPU_941_942_943 = gpd.read_file(TPU_941_942_943)
Map_TPU_944 = gpd.read_file(TPU_944)
Map_TPU_950_951 = gpd.read_file(TPU_950_951)
Map_TPU_961_962_963 = gpd.read_file(TPU_961_962_963)
Map_TPU_971_972_973_974 = gpd.read_file(TPU_971_972_973_974)
Map_TPU_975 = gpd.read_file(TPU_975)
Map_TPU_976 = gpd.read_file(TPU_976)

color_fill = "white"
color_edge = "#64645e"

geometry2 = [Point(xy) for xy in zip( df2["Longitude"], df2["Latitude"])]
geometry2[:3]
geo_df2 = gpd.GeoDataFrame(df2, crs=crs, geometry=geometry2)

fig,ax = plt.subplots(figsize = (20,20))
ax.margins(y=0.0005)

ax.set_facecolor('#e0e7fb')

Map_Shenzhen.plot(ax=ax,
                  alpha=1,
                  color="#afa88b",
                  edgecolors=color_edge,
                  linewidth=0.5)
```

```
Map_TPU_111.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='111')
Map_TPU_112.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='112')
Map_TPU_113.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='113')
Map_TPU_114.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='114')
Map_TPU_115.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='115')
Map_TPU_116.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='116')
Map_TPU_121_123_124.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='121_123_124')
Map_TPU_122.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='122')
Map_TPU_131.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='131')
Map_TPU_132.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='132')
Map_TPU_133.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='133')
Map_TPU_134.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='134')
Map_TPU_135.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='135')
Map_TPU_140.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='140')
Map_TPU_141.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='141')
Map_TPU_142.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='142')
Map_TPU_143.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='143')
Map_TPU_144.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='144')
Map_TPU_145.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='145')
Map_TPU_146_147.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='146_147')
Map_TPU_148.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='148')
Map_TPU_149.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='149')
Map_TPU_151.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='151')
Map_TPU_152.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='152')
```

```
Map_TPU_153.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='153')
Map_TPU_154.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='154')
Map_TPU_155.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='155')
Map_TPU_156_158.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='156_158')
Map_TPU_157.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='157')
Map_TPU_161.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='161')
Map_TPU_162.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='162')
Map_TPU_163.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='163')
Map_TPU_164_165.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='164_165')
Map_TPU_166.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='166')
Map_TPU_167.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='167')
Map_TPU_171.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='171')
Map_TPU_172.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='172')
Map_TPU_173.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='173')
Map_TPU_174.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='174')
Map_TPU_175_176.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='175_176')
Map_TPU_181_182.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='181_182')
Map_TPU_183_184.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='183_184')
Map_TPU_190_192_194.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='190_192_194')
Map_TPU_191.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='191')
Map_TPU_193_195_198.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='193_195_198')
Map_TPU_196.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='196')
Map_TPU_197.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='197')
Map_TPU_211.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='211')
```

```
Map_TPU_212.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='212')
Map_TPU_213_215_216.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='213_215_216')
Map_TPU_214.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='214')
Map_TPU_217.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='217')
Map_TPU_220.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='220')
Map_TPU_221.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='221')
Map_TPU_222.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='222')
Map_TPU_225.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='225')
Map_TPU_226.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='226')
Map_TPU_227.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='227')
Map_TPU_228.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='228')
Map_TPU_229.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='229')
Map_TPU_231.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='231')
Map_TPU_232.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='232')
Map_TPU_233.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='233')
Map_TPU_234.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='234')
Map_TPU_235.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='235')
Map_TPU_236.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='236')
Map_TPU_237.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='237')
Map_TPU_241.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='241')
Map_TPU_242.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='242')
Map_TPU_243.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='243')
Map_TPU_244.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='244')
Map_TPU_245.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='245')
```

```
Map_TPU_246.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='246')
Map_TPU_247.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='247')
Map_TPU_251_256.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='251_256')
Map_TPU_252.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='252')
Map_TPU_253.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='253')
Map_TPU_254.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='254')
Map_TPU_255_269.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='255_269')
Map_TPU_260.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='260')
Map_TPU_261.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='261')
Map_TPU_262.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='262')
Map_TPU_263.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='263')
Map_TPU_264.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='264')
Map_TPU_265.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='265')
Map_TPU_266.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='266')
Map_TPU_267.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='267')
Map_TPU_268.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='268')
Map_TPU_271.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='271')
Map_TPU_272.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='272')
Map_TPU_280.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='280')
Map_TPU_281.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='281')
Map_TPU_282.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='282')
Map_TPU_283.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='283')
Map_TPU_284.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='284')
Map_TPU_285.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='285')
```

```
Map_TPU_286.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='286')
Map_TPU_287.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='287')
Map_TPU_288_289.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='288_289')
Map_TPU_290.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='290')
Map_TPU_291.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='291')
Map_TPU_292.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='292')
Map_TPU_293_296.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='293_296')
Map_TPU_294.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='294')
Map_TPU_295.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='295')
Map_TPU_297.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='297')
Map_TPU_298.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='298')
Map_TPU_310_321.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='310_321')
Map_TPU_320_324_329.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='320_324_329')
Map_TPU_322.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='322')
Map_TPU_323.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='323')
Map_TPU_325.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='325')
Map_TPU_326.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='326')
Map_TPU_327.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='327')
Map_TPU_328.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='328')
Map_TPU_331_332_333_334_336_340.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='331_332_333_334_336_340')
Map_TPU_335.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='335')
Map_TPU_350.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='350')
Map_TPU_351.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='351')
Map_TPU_411_412_413_414_415_416_427.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='411_412_413_414_415_416_427')
```

```
Map_TPU_421_422.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='421_422')
Map_TPU_423_428.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='423_428')
Map_TPU_424.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='424')
Map_TPU_425.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='425')
Map_TPU_426.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='426')
Map_TPU_431_432_433_434.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='431_432_433_434')
Map_TPU_441.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='441')
Map_TPU_442.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='442')
Map_TPU_510.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='510')
Map_TPU_511.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='511')
Map_TPU_512.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='512')
Map_TPU_513.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='513')
Map_TPU_514.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='514')
Map_TPU_515.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='515')
Map_TPU_516.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='516')
Map_TPU_517.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='517')
Map_TPU_518.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='518')
Map_TPU_519.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='519')
Map_TPU_521.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='521')
Map_TPU_522.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='522')
Map_TPU_523.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='523')
Map_TPU_524.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='524')
Map_TPU_525.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='525')
Map_TPU_526.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='526')
```

```
Map_TPU_527.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='527')
Map_TPU_528.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='528')
Map_TPU_529.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='529')
Map_TPU_531.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='531')
Map_TPU_532.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='532')
Map_TPU_533.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='533')
Map_TPU_541.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='541')
Map_TPU_542.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='542')
Map_TPU_543_546.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='543_546')
Map_TPU_544.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='544')
Map_TPU_545.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='545')
Map_TPU_547.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='547')
Map_TPU_548.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='548')
Map_TPU_610_621_632.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='610_621_632')
Map_TPU_620_622_641.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='620_622_641')
Map_TPU_623.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='623')
Map_TPU_624.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='624')
Map_TPU_625.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='625')
Map_TPU_626.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='626')
Map_TPU_627.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='627')
Map_TPU_628.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='628')
Map_TPU_629.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='629')
Map_TPU_631_633.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='631_633')
Map_TPU_634.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='634')
```

```
Map_TPU_642.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='642')
Map_TPU_651_652_653.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='651_652_653')
Map_TPU_711_712_721_728.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='711_712_721_728')
Map_TPU_720.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='720')
Map_TPU_722_727.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='722_727')
Map_TPU_723.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='723')
Map_TPU_724.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='724')
Map_TPU_725.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='725')
Map_TPU_726.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='726')
Map_TPU_729.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='729')
Map_TPU_731_733_754.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='731_733_754')
Map_TPU_732_751_753.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='732_751_753')
Map_TPU_741_742_743_744.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='741_742_743_744')
Map_TPU_755.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='755')
Map_TPU_756_761_762.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='756_761_762')
Map_TPU_757.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='757')
Map_TPU_758.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='758')
Map_TPU_759.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='759')
Map_TPU_811_812_813_814_815.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='811_812_813_814_815')
Map_TPU_820.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='820')
Map_TPU_821.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='821')
Map_TPU_822.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='822')
Map_TPU_823.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='823')
Map_TPU_824_829.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='824_829')
```

```
Map_TPU_825.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='825')
Map_TPU_826_828.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='826_828')
Map_TPU_827.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='827')
Map_TPU_831.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='831')
Map_TPU_832_834.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='832_834')
Map_TPU_833.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='833')
Map_TPU_835.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='835')
Map_TPU_836.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='836')
Map_TPU_837.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='837')
Map_TPU_838.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='838')
Map_TPU_839.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='839')
Map_TPU_975.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='975')
Map_TPU_911_912_913.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='911_912_913')
Map_TPU_920.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='920')
Map_TPU_931_933.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='931_933')
Map_TPU_932_934.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='932_934')
Map_TPU_941_942_943.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='941_942_943')
Map_TPU_944.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='944')
Map_TPU_950_951.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='950_951')
Map_TPU_961_962_963.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='961_962_963')
Map_TPU_971_972_973_974.plot(ax=ax, alpha=1, color='#ffffff',
edgecolors=color_edge, linewidth=0.5, label='971_972_973_974')
Map_TPU_976.plot(ax=ax, alpha=1, color='#ffffff', edgecolors=color_edge,
linewidth=0.5, label='976')

geo_df2[geo_df2['OpStatus'] == 1].plot(
                                        ax = ax, alpha = 1, markersize = 25, color =
                                        'black', marker = '.', label = 'Success'
                                        )
```

```
geo_df2[geo_df2['OpStatus'] == 0].plot(
                                        ax = ax, alpha = .05, markersize = 25, color =
                                        'red', marker = '.', label = 'Failed'
                                        )

ax.text(114.02, 22.54, "China", fontsize=14)
ax.set_ylim(22.15, 22.566) #bottom, top
ax.set_xlim(113.83, 114.415) #left, right

patch1  = mpatches.Patch(color='black', label='Successful')
patch2  = mpatches.Patch(color='red', label='Failure')

legend = plt.legend(loc='lower right',
        facecolor='white',
        framealpha=1,
        edgecolor=color_edge,
        title='All Restaurant Status',
        handles = [patch1, patch2]
        )

legend.get_title().set_fontsize('14')
plt.setp(plt.gca().get_legend().get_texts(), fontsize='14')

fig.savefig("1 All Restaurant Locations.png",
            format='png',
            dpi=600,
            bbox_inches='tight'
            )
```