

Cuisine Vs Location

Introduction

As a popular saying goes - In the real estate business it is all about the location, location, and location. That is, the location of a property determines its value. A one storey, one bedroom house in Hamptons, New York, will cost more than a three storey, six bedroom house in Boise, Idaho. The same can be considered true in the Restaurant Business. For world-renowned chefs like Gordon Ramsey or Jamie Oliver, the location of their restaurant can be a secondary factor as customers usually flock to their restaurants for experiencing fine dining. But for someone who is starting out in the industry and doesn't have a reputation to rely on, the success of their endeavor will heavily depend on the location of their restaurant. Open a restaurant in a location with less foot traffic, and you won't have any customers. Open a restaurant in a location that has similar types of restaurants, then you are fighting for the customer share and trying to leave a mark of your own.

The mathematical approach described in this report provides a two-fold solution for the above-mentioned problem:

- At a particular location, what would be the best type of restaurant to invest in
- For a particular cuisine, where would the best location be to open a restaurant

This report will be of interest to two types of audience:

- An Investor who wants to cash in on the appeal of a location
- A Chef who is trained in a specific culinary style

Data Used

This report concentrates on the city of Toronto in Ontario Province, Canada. The data sets used in this report are as mentioned below:

- **Toronto Neighborhood Data**
 - Source – Wikipedia
 - Data Description
 - Postal codes allocated to Toronto City
 - Names of the Boroughs present in each of the postal code areas
 - Neighborhoods present in each of the boroughs
 - Screen scrapping technique is used to read the data into a data frame.
- **File Containing Latitude & Longitude Information**
 - Source - File provided by the Instructor
 - Data Description
 - Postal codes allocated to Toronto City
 - Latitude Information for all the postal codes
 - Longitude Information for all the postal codes
 - Function provided in the pandas library is used for reading the data into a data frame.
- **Restaurant Information**

- Source - Foursquare Application
- Data Description
 - Data obtained using a search query to identify all the Restaurants located in each Toronto neighborhood
 - The radius for the search criteria is set at 500 meters.
 - The search is limited to 50 restaurants per neighborhood.
 - Venue ID, Venue Name, Venue Latitude, Venue Longitude, and Venue Category information is read from the JSON returned by the search query into a data frame.
- **Restaurant Rating**
 - Source - Foursquare Application
 - Data Description
 - Data obtained using a venue details query to obtain the ratings for all the restaurants in each Toronto neighborhood.
 - Venue ID is passed as the key in the query.
 - Rating information is read from the JSON file returned by the query and added to the data frame with the venue information.

Methodology

The methodology used for this report is a variation of the content-based recommendation system. The content in this case is the existing restaurant information in a location. There are two outputs that can be derived based on this content:

- The first is the next restaurant option that might be suited for a location.
- The second is the location where a particular type of restaurant can be opened.

The reason for using the content-based recommendation system for this report is that we are analyzing the current trend of restaurants in a location and trying to predict what could be the next best options.

Pre-Processing of Data

The main data source for this report is the Toronto neighborhood information. The pre-processing performed to obtain a clean data set to start the exploratory analysis is mentioned in the following steps:

- The Toronto postal code information is read into the data frame by using the `beautifulsoup4` library function.
- Any rows without the borough information are removed from the data frame.
- The borough name is copied to the neighborhood column, in case the neighborhood value is missing.
- All the rows having the same postal code are grouped together.
- We remove any duplicate records that may have formed after the grouping exercise.
- Next we read the `.csv` file that contains the latitude and longitude information per postal code into a new data frame.
- This new data frame is merged with the Toronto neighborhood information data frame.

This gives us a data frame that contains the following set of information about a Toronto neighborhood:

- Postal Code
- Borough Name
- Neighborhood Name
- Neighborhood Latitude Value
- Neighborhood Longitude Value

This data frame will act as our input for the exploratory data analysis being carried out as part of this report.

- We create a user-defined function that can return the list of venues based on the name, latitude, and longitude information of a location.
- The function utilizes the Foursquare application search query to get the venue information.
- The search query is restricted to venues of the type restaurant.
- From the result set, the following set of information is read and stored into a new data frame:
 - Venue ID
 - Venue Name
 - Venue Latitude
 - Venue Longitude
 - Venue Category
- We also insert the Neighborhood Name, Latitude, and Longitude information into the new data set.
- If there are any neighborhoods without a restaurant, then the corresponding row is dropped from the data frame.

Exploratory Analysis

The approach taken in this report involves reviewing all the restaurant categories and then grouping them, based on the region they originate from or based on the similarity of the ingredients used. The reasoning behind this approach is as follows. If we have two nearby restaurants that offer similar types of dishes, e.g., a Thai restaurant and a Chinese restaurant, then the restaurants will be competing for the same customer base that is interested in East Asian cuisine. But instead, if we have an Italian restaurant and a Thai restaurant nearby, then such a conflict will be avoided. Hence, we follow the below approach to narrow down the restaurant categories.

- Review all the unique restaurant categories.
- Look further into individual categories that are not based on any particular cuisine type. This will give us an idea on how to group the different types of cuisines.
- Remove venues that are not truly restaurants, like movie theaters, music rooms, etc., as well as those rows that do not have any definite categorization like food, restaurant, etc.
- Next, we drop the word "restaurant" from the Category field for ease of grouping.
- The grouping exercise is performed as follows:

| Cuisine Type | Cuisine Category Group |
|--------------|------------------------|
| Chinese | East Asian |
| Thai | East Asian |
| Korean | East Asian |

| | |
|------------------|----------------|
| Asian | East Asian |
| Japanese | East Asian |
| Dim Sum | East Asian |
| Hakka | East Asian |
| Vietnamese | East Asian |
| Mongolian | East Asian |
| Cantonese | East Asian |
| Dumpling | East Asian |
| Noodle House | East Asian |
| Indian | South Asian |
| Ethiopian | African |
| African | African |
| American | North American |
| New American | North American |
| Steakhouse | North American |
| Diner | North American |
| Mexican | Latin American |
| Latin American | Latin American |
| Bar | Beverage Bar |
| Night Club | Beverage Bar |
| Wine Bar | Beverage Bar |
| Beer Bar | Beverage Bar |
| Gastropub | Beverage Bar |
| Pub | Beverage Bar |
| Brewery | Beverage Bar |
| Nightlife Spot | Beverage Bar |
| Lounge | Beverage Bar |
| Italian | European |
| Pizza Place | European |
| Spanish | European |
| Tapas | European |
| French | European |
| Portuguese | European |
| Modern European | European |
| Eastern European | European |
| Middle Eastern | Mediterranean |
| Greek | Mediterranean |
| Mediterranean | Mediterranean |
| Caribbean | Caribbean |
| Cuban | Caribbean |

- Other establishments like Breakfast Spot, Sandwich Place are retained as its own Cuisine Category group.
- A FOR loop is used to cycle through the data frame and add the New Cuisine Category group to the data frame.
- A new data frame is created by retaining only the venue-related information.
- Any records with duplicate venue information are dropped.
- A new column is added to the data frame to store the rating information for each venue.
- A user-defined function is created to retrieve the ratings information for a venue.
 - The input to this function is the Venue ID.
 - The function uses a Foursquare application premium query to get information about a venue using the id.
 - From the result set, the ratings value is read and returned to the calling statement.
- The rating value is stored in the new column created in the previous step.
- All the rows without a rating value are dropped.
- The rows are grouped based on the Cuisine Category Group value.
- The aggregate function COUNT is used to obtain the number of venues that belong to a particular cuisine category group.
- The sum of the rating per cuisine category group is also calculated.
- The weighted average for each cuisine category group is obtained by multiplying the sum of the rating for a group by the number of venues (weight) in that group and then by dividing it by the total number of venues.
- The resulting data frame is sorted based on the descending value of the Weighted Average.

This resulting data frame gives us information about the preferences of the customers in the Toronto region. This weighted average ratings data frame along with the Toronto neighborhood information data frame will be used to derive the results for the two questions we are trying to answer as part of this report.

- At a particular location, what would be the best type of restaurant to invest in.
- For a particular cuisine, where would the best location be to open a restaurant.

Results

In order to solve for the first question, a new data frame is created from one of the intermediate data frames with only the neighborhood and Cuisine Category Group information. Next, a GROUP BY operation is performed on the neighborhood value, to get one row per neighborhood and to simultaneously concatenate the Cuisine Category Group value into a TUPLE. The TUPLE is then converted into a string for ease of processing.

A user-defined function that can accept a data frame as an input variable is created. The function reads data from the input data frame neighborhood by neighborhood. For each neighborhood, the associated cuisine category grouped string is read. Next, the function reads data from the weighted average ratings data frame row by row and checks if the associated cuisine category group is present in the grouped cuisine category column value of the input data frame. If the first value read from the weighted average ratings data frame is present in the grouped cuisine category column value of the input data frame, then the function iterates to the next row of the weighted average ratings data frame until a match is not

found. At this point, the function breaks. The concept behind this approach is that if a value from the weighted average ratings data frame is not present in the grouped cuisine category value associated with a neighborhood, then the next restaurant being opened in that neighborhood should be of that category type read from the weighted average ratings group. The neighborhood name and the next category group type is stored in a new data frame, and the same is returned by the function.

In order to solve for the second question, we group the same resulting data frame returned by the above function on the next category group value. This will give us all the neighborhoods that are suitable for opening a restaurant of that particular category group type. These neighborhoods do not already have a restaurant that belongs to this group type.

Data Frame depicting the solution for the first question -

There are 56 unique neighborhood combinations grouped on the basis of the postal code and borough, and for each of these neighborhoods, we have identified the category group type under which the next restaurant can be opened.

| | Neighborhood | Next_Rest_Cat_Group |
|----|---|----------------------|
| 0 | Adelaide, King, Richmond | Molecular Gastronomy |
| 1 | Agincourt | European |
| 2 | Agincourt North, "LAmoreaux East", Milliken, S... | European |
| 3 | Alderwood, Long Branch | North American |
| 4 | Bathurst Manor, Downsview North, Wilson Heights | East Asian |
| 5 | Bedford Park, Lawrence Manor East | North American |
| 6 | Berczy Park | European |
| 7 | Brockton, Exhibition Place, Parkdale Village | European |
| 8 | Cabbagetown, St. James Town | European |
| 9 | Caledonia-Fairbanks | East Asian |
| 10 | Canada Post Gateway Processing Centre | European |
| 11 | Cedarbrae | North American |
| 12 | Central Bay Street | Beverage Bar |

Data Frame depicting the solution for the second question -

This lists all the different Category Group types and all the suitable neighborhoods for opening a restaurant of that group.

| | Next_Rest_Cat_Group | Neighborhood |
|---|----------------------|---|
| 0 | Beverage Bar | Central Bay Street, Little Portugal, Trinity |
| 1 | Breakfast Spot | Church and Wellesley |
| 2 | East Asian | Bathurst Manor, Downsview North, Wilson Height... |
| 3 | European | Agincourt, Agincourt North, "LAmoreaux East", ... |
| 4 | Mediterranean | Commerce Court, Victoria Hotel |
| 5 | Molecular Gastronomy | Adelaide, King, Richmond, Ryerson, Garden Dist... |
| 6 | North American | Alderwood, Long Branch, Bedford Park, Lawrence... |
| 7 | Sandwich Place | St. James Town, Stn A PO Boxes 25 The Esplanade |
| 8 | Vegetarian / Vegan | Design Exchange, Toronto Dominion Centre, Firs... |

Discussion

Neighborhoods

The East Asian group has the highest weighted average ratings, followed by the European and the North American groups. Let's look at the neighborhood of Agincourt; it already has restaurants belonging to the categories East Asian and Sandwich Place. Therefore, based on the analysis, the next best suited option for the locality will be a restaurant of the category European. Similarly, for the neighborhood group of Adelaide, King, & Richmond, the suitable option is a Vegetarian or Vegan Establishment, as it already has restaurants belonging to other categories.

Cuisine

If a Chef is interested in opening a restaurant serving Mediterranean dishes, the most suited location would be Commerce Court or Victoria Hotel, as they do not already have restaurants belonging to this group. Similarly, for a Breakfast Spot, the most apt location would be Church & Wellesly.

Conclusion

- This report can be used as the first step by an Investor or a Chef to begin their analysis.
- The report can be extrapolated to other cities with a similar postal identification structure (postal code, borough, and neighborhood) with ease.
- Even if the user wants to run the analysis on a city that doesn't have a similar postal identification structure, but does have other identification factors, we can make minor modifications to the code to run the same analysis.
- The report can also be extrapolated for other type of industries like clothing, retail stores, banks etc. provided we can get the location, venue and rating information from Foursquare or other applications.