Analysis of opening a new resturant around Dundas Square area

Xin Gu

December 23

1. Introduction: Business Problem

Business analytics for opening restaurants is integral to understanding the inner workings of the business and being aware of how the owner can improve it to foster a sustainable level of success. By working with relevant data, the owner should be able to track, monitor, and measure the most valuable business insights in a way that is clear, concise, and digestible, pulling from past, present, and predictive data. This will enable sustainable KPI management processes that will ultimately increase productivity and save money.

In this project I will try to utilize various kinds of data to find an optimal location for opening a Chinese restaurant in **Toronto Downtown, around Dundas Square**.
Since there are lots of restaurants in Downtown Toronto we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Chinese restaurants**. We would also prefer locations **as not close to Chinatown as possible**, assuming that first two conditions are met.
We will use our research to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Data

2.1 Data acquistion

Based on definition of our problem, factors that will influence our decision are:

- Geographic data
- Competitors data
- Outside sources, such as social media posts, locational data, weather data, traffic data
- Neighbourhood data, such as the community category, living condition(work type, living status, etc.)
- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Chinese restaurants in the neighborhood
- distance of any existing Chinese resturant neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **python packages**
- acquiring the data of Dundas Square through **google API**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

2.2 Data acqusition and data cleaning

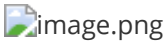
2.2.1 Google API

We are using Google map api to get the coordinates of Toronto downtown. First we create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is aprox. 12x12 killometers centered around Toronto Downtown.

Firstly, to find the latitude & longitude of Toronto Downtown, using specific, well known address (Dundas Square, for example) and Google Maps geocoding API.
Then, create a grid of area candidates, equaly spaced, centered around city center and within ~6km from Dundas Square. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

After we get the coordinates of the city enter, then let's create a **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all it's neighbors**. We totally get 364 candidate neighborhood centers generated. Each one is equally distant from the center we identified "Dundas Square". Then we want to visualize the data we got on the map, using the Folium map to show how it looks like.

 image.png

After we have the coordinates of centers of neighborhoods/areas to be evaluated, equally spaced (distance from every point to it's neighbors is exactly the same) and within ~6km from Dundas Square. Then we use Google Maps API to get approximate addresses of those locations, and transfer those addresses into Pandas dataframe and save to local file in order to further analysis.

2.2.2 Foursquare API

What is Foursquare API?
Foursquare is a social location service that allows users to explore the world around them. Users can download the Foursquare application to their iPhone, Blackberry, or Android phone and sign up for free. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests.

We use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only those that are proper restaurants - coffe shops, fast food, bakeries etc. are not direct competitors so we don't care about those. So we will include in out list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Chinese restaurant' category, as we need info on Chinese restaurants in the neighborhood.

Firstly, get the category IDs corresponding to Chinese restaurants were taken from Foursquare web site (https://developer.foursquare.com/docs/resources/categories).

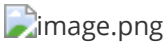
Secondly, find the 'Root' category for all food-related venues: --'4d4b7105d754a06374d81259' --.

Then, search all of the Chinese realated categories IDs, manually wirrten into a data line.

2.2.3 Cleaning and transformation

After acquiring the data we need, and written them into the dataframe, we need to transfer the data into the format we are able to manipulate and analysis. First we have to set up a filter to auto skip some raw data we really don't need, for example, some categiries contain key word like "fast food", "bakery", etc. At the same time, to keep some we do need, "resturant", "dinner", and some key words only for Chinese resturants "fired rice", "fried noodle", etc.
Go over our neighborhood locations and get nearby restaurants; we'll also maintain a dictionary of all found restaurants and all found Chinese restaurants
Then define a function to auto formate the address, like we do not really need country within the address, for Toronto is Canada, we set up to replace the country with space.
Also define a function for getting the venues of each neighborhoods by giving the Foursquare client ID. Define a function to append the address to each resturant, not only Chinese resturant, all of them around 12*12 KM of Dundas Square.
Load the data into Pandas dataframes, then get a visualization on the map.

Blue dots indicates all the restrants aroud Dundas Square.
Red dots indicates all the Chinese resturants around Dundas Square.

 image.png

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Chinese restaurant!

3. Exploratory Data Analysis

3.1 Methodology

In this project we will direct our efforts on detecting areas of Toronto downtown that have low restaurant density, particularly those with low number of Chinese restaurants. We will limit our analysis to area ~6km around city center.

In first step we have collected the required **data: location and type (category) of every restaurant within 6km from Toronto city center** (Dundas Square). We have also **identified Chinese restaurants** (according to Foursquare categorization).

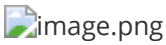
Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Toronto - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (*and* no Chinese restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two restaurants in radius of 250 meters**, and we want locations **without Chinese restaurants in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

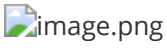
3.2 Analysis

First let's perform some basic explanatory data analysis and derive some additional info from our raw data. Count the **number of restaurants in every area candidate**:

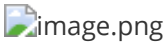
Average number of restaurants in every area with radius=300m: 4.91208791209

 image.png

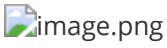
Then calculate the **distance to nearest Chinese restaurant from every area candidate center** (not only those within 300m - we want distance to closest one, regardless of how distant it is).
Average distance to closest Chinese restaurant from each area center: 1300.779399600793



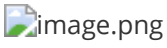
Now create a map showing **heatmap / density of restaurants** and try to extract some meaningfull info from that.



Through the Heatmap, looks like a few pockets of low restaurant density closest to city center can be found **south, south-east and east from city center**.
Then we create another heatmap map showing **heatmap/density of Chinese restaurants** only. By having these two maps, we can have a better insights that where has a lower density of resturants and also Chinese resturants.



Comparing the maps, although we could not get too much information, but we can see that at the south, south-east and east side of city center, there are lots of opportunity to start up new business, there has low density of resturants and Chinese resturants specifically. Especially we can explore the Financial Distrcit and near Union Station. So next step, let's explore those neighborhoods a little bit further.



Within 2500 meters, how the resturants distrbuted around south, south-east and east of toronto center

Now we want to see how the resturants are distributed around the new region that we are interested. First, we calculate the total number of candidate neighborhoods is 2261. Then we calculate two most important things for each location candidate: **number of restaurants in vicinity** (we'll use radius of **250 meters**) and **distance to closest Chinese restaurant**. By filtering those locations, we only interested in locations that with no more than two restaurants in radius of 250 meters**, and** no Chinese restaurants in radius of 400 meters**.

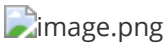
Locations with no more than two restaurants nearby: 1413

Locations with no Chinese restaurants within 400m: 1626

Locations with both conditions met: 1222

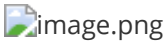
Showing the data on the map

Visualize on the heat map



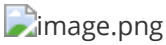
We now have a bunch of locations fairly close to Dundas Square and we know that each of those locations has no more than two restaurants in radius of 250m, and no Chinese restaurant closer than 400m. Any of those locations is a potential candidate for a new Chinese restaurant, at least based on nearby competition.
And we also have a clear indication of zones with low number of restaurants in vicinity, and *no* Chinese restaurants at all nearby.

For next step would be **cluster** those locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.



After clustering the area into 15 clusters, with the equal radius of 500m.

This concludes our analysis. We have created 15 addresses representing centers of zones containing locations with low number of restaurants and no Chinese restaurants nearby, all zones being fairly close to city center (all less than 4km from Dundas Square, and some of those less than 2km from Dundas Square). Although zones are shown on map with a radius of ~500 meters (green circles), their shape is actually very irregular and their centers/addresses should be considered only as a starting point for exploring area neighborhoods in search for potential restaurant locations.



4. Results and analysis

Our analysis shows that although there is a great number of restaurants in Toronto Downtown (~2000 in our initial area of interest which was 12x12km around Dundas Square), there are pockets of low restaurant density fairly close to Square center.

Highest concentration of restaurants was detected north and west from Dundas Square, so we focused our attention to areas south, south-east and east. our attention was focused on Financial District and east which offer a combination of popularity among tourists, closeness to Square center, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this more narrow area of interest (covering approx. 5x5km south-east from Dundas Square) we first created a dense grid of location candidates (spaced 100m appart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Chinese restaurant closer than 400m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 25 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Chinese restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Toronto center but not crowded with existing restaurants (particularly Chinese) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

5. Conclusion

Purpose of this project was to identify Toronto areas close to Dundas Square center with low number of restaurants (particularly Chinese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Chinese restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.