

Finding a Suitable Location for a Restaurant

By Tim Strebel

Introduction: Business Problem

In this project, I'm going to apply Data Science tools and methodologies that I've learned throughout the IBM Professional Data Science Course to find an optimal location for an Asian fusion restaurant in my hometown of Salt Lake City, Utah. This report will be targeted toward stakeholders interested in opening an Asian fusion or any other type of Asian restaurant in Salt Lake City.

This charming city is home to nearly 201,000 citizens and contains many attractions that draw tourism from all over the world. Home to the world's first Mormon temple, it contains stunning architecture, a unique culture and an abundance of shopping and walkways to accommodate pedestrian traffic. It is truly an ideal area for restaurants to feed people famished from a day's worth of site-seeing shopping. I chose Asian fusion as the theme for my restaurant because it is my favorite food and growing up in Salt Lake City, I wished there were more of these restaurants located downtown.

Based on my experience growing up in Salt Lake City combined with the Data Science methods I have learned on from this course, I have ascertained that ideal candidate locations for a restaurant are ones closest to city center (Temple Square) with the fewest number of restaurants to compete with, especially Asian restaurants which would be direct competition. In this project, I will explore downtown Salt Lake City using various geospatial, mathematical and clustering tools to check my assumptions and lay the groundwork for my stakeholders in finding an optimal location to open a new restaurant.

Methodology

I. Location Data Collection

Find the city center and a set of candidate locations spaced equally apart withing a 6 Kilometer radius of city Center. Identify all restaurants surrounding each candidate locations and of those restaurants, identify which are Asian restaurants.

II. Explore Restaurant Density of Each Candidate Area

Use heatmaps to identify promising areas close to center with a low number of restaurants and with no Asian restaurants in a relatively close proximity.

III. Use K-means Clustering to Find Optimal Location Centroids

Once viable candidate locations are identified, use k-means clustering to find the center points that contain the greatest number of viable areas to for our restaurant.

Description of the Data

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

- Number of existing restaurants in the candidate area of any type.
- Number of and distance to any other Asian restaurants

- Distance of candidate area from city center

I decided to use evenly spaced locations around city center, to define my candidate areas. The following data sources will be needed to extract/generate the required information:

- Latitude and Longitude coordinates of Salt Lake City Center will be obtained using geolocation information from **AmericanTowns.com**
- Once I obtain city center coordinates, I will generate equally spaced candidate areas in a 6x6 Kilometer grid that will surround city center. To achieve this, I will use **PyProj** to transform lat-long coordinates to UTM, use mathematical algorithms to generate the areas geospatially and finally transform back to lat-long from UTM.
- Approximate addresses of candidate areas will be obtained using an open-source geocoding library called **Nominatim**.
- Finally, the number of restaurants and their type and location in every candidate area will be obtained using **Foursquare API**

I. Finding City Center

The first step in this process is to find coordinates that represent city center. To achieve this, I've web scraped data from **AmericanTowns.com** that I have found sufficient for this task. I have used the `pandas.read_html` function to quickly obtain lat-long coordinates for each zip code located within Salt Lake City. Once I obtained these data points, I used folium to display them geospatially so I could choose the zip code coordinates closes to city center. Zip code 84133 has panned out to be the closest location.

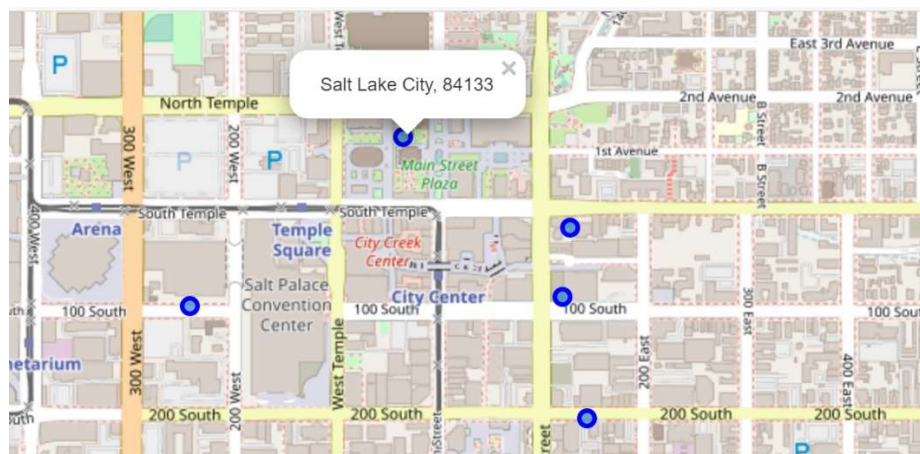


Figure 1 Salt Lake City Zip Code Coordinates

II. Generating Candidate Areas

Using the **PyProj** library, I created algorithms to generate equally spaced candidate areas that are 300 meters apart and equally spaced across a 6x6 kilometer grid surrounding city center. In order to achieve this, I created a function to convert lat-long to Universal Transverse Mercator (UTM) so that locations could be offset from city center and from each other in metric units to allow for greater precision. Also, while in UTM, I can calculate the distance between two locations in meters to obtain the distance from the candidate area to the city center. Once my candidate areas were set, I transformed the coordinates back to lat-long so they could be used later for geocoding and **Foursquare**.



The **Foursquare API** is an open-source tool that allows a limited number of free calls to obtain information about venues the surrounding area. Given a pair of latitude-longitude coordinates, you can obtain the name, address, categorical and other information about venues in the surrounding area. I used this API to obtain a list of all restaurants surrounding my candidate areas. My interest was in venues from the food category, but only ones that are restaurants, not coffee shops, bakeries, convenience stores, etc. Additionally, out of all these restaurants I made sure to label Asian restaurants since these would be direct competition.

Below is a map of these restaurants. Asian restaurants are denoted by red dots.



Analysis

Once I had my data, I conducted data and heat-map analysis to obtain a sense of restaurant Using the information obtained from Foursquare, I was able to reveal that the average distance from each candidate location was approximately 747 meters. Looking at the average distance was not all that informative but the heat-maps did a great job of showing restaurant density in the surrounding areas.



Figure 4 Salt Lake City Restaurant Density



Figure 5 Salt Lake City Asian Restaurant Density

After viewing the heat maps, I began to wonder why some areas were denser than others. Particularly why the locations to north and west of city center contained areas of restaurant sparsity.

Concerning the northern part, you can see from the map that residential and city roadways end where the base of the mountain begins. Also, due north is a part of Salt Lake called the Avenues. This is a historic part of town that is dense with residential real-estate. Lastly, there are many land-mark locations such as temple square and the capital building that. Considering these factors, there are likely few locations suitable for a restaurant due to a lack of available real-estate.

Regarding the west side, I have found a significant rates of crime in the areas west of city center. Crime rate data from website called **neighborhood scout** reveals a trend in crime that I found unsavory and let me to rule out the west side of town.

After my analysis I decided to narrow my scope to the south-east section of the city which still contained many viable areas suitable for my restaurant. Below are heatmaps representing restaurant density in these areas.

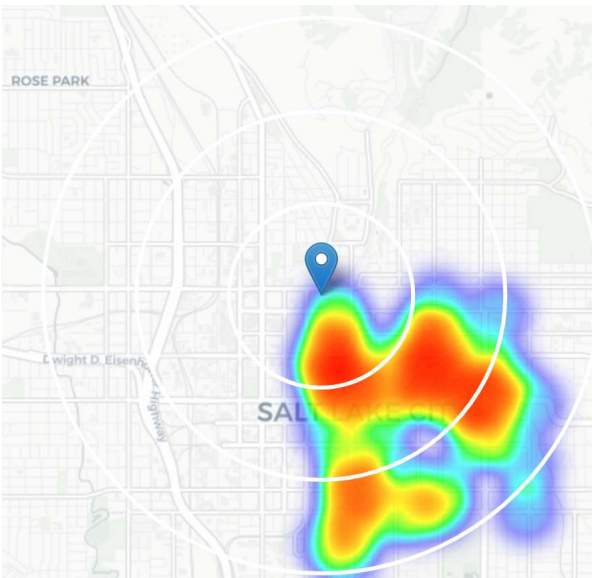


Figure 6 Southeast SLC Restaurant Density

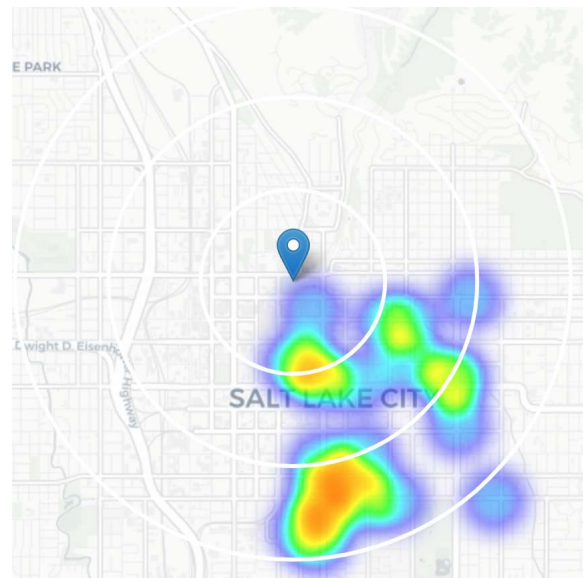


Figure 7 Southeast SLC Asian Restaurant Density

To narrow my scope even further, I looked at areas that contained two or less restaurants in the area that did not contain an Asian restaurant within 400 meters of the vicinity. I found 17 areas that had two or less restaurants, 13 areas that did not contain an Asian restaurant within 400 meters and 12 areas that met both criteria.

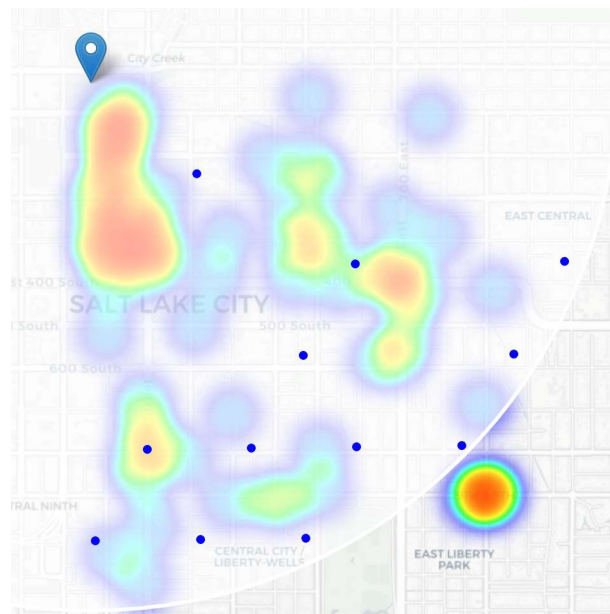


Figure 8 Areas that met Criteria

For the final part of my analysis I used k-means clustering to come up with four clearly defined areas that contain the greatest number of suitable areas for my restaurant. The clusters were derived using the **Sci-Kit Learn** k-means clustering algorithm on the UTM coordinates of the 12 locations that met both of my criteria for suitability. Using the cluster centroids, I was able to map out the four best areas based on my criteria represented by green circles below.

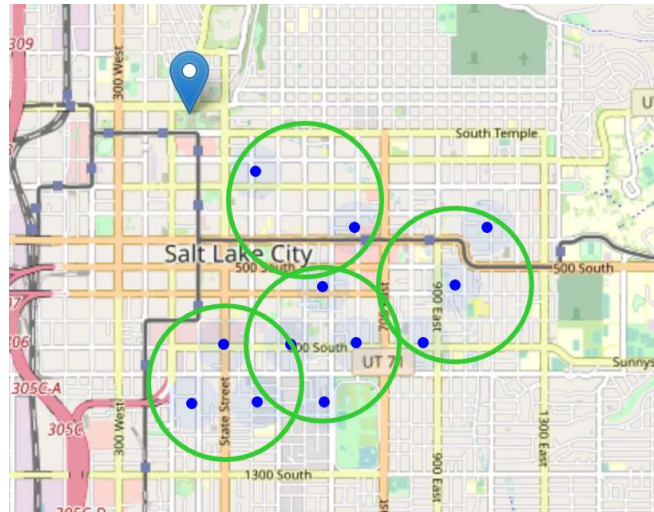


Figure 9 Optimal Areas based on Criteria

Results and Discussion

Most of the restaurants in Salt Lake City are located approximately ~1000 meters south of city center. The reason many restaurants are not located to the north is due to the location of many landmark sites, mountainous terrain and a high density of residential real estate. Due west of city center, there is a higher rate of crime hence a lower rate of tourism and foot traffic. Located south east of city center there are plenty of locations with a lower density of restaurants that are far away from other potential competing Asian restaurants making it a most suitable area for an Asian fusion restaurant.

Narrowing in on the south-east side of the city, we clustered the candidate locations to create zones of interest which contain the greatest number of location candidates. We extracted addresses at the center of these zones using reverse-geocoding so that we had clear reference points.

The result was four clearly defined areas suitable for starting Asian fusion restaurant in which the restaurant will have low competition. This analysis is not final because there are likely other factors that will play into locating an optimal location; however, this is a great start. It is entirely possible that there are reasons that these areas have a low number of restaurants that make them unsuitable for my proposed restaurant. I suggest further research by other like-minded stake holders to derive an optimal decision.

Conclusion

The purpose for this project was to identify desirable areas in Salt Lake City Utah to start a new Asian fusion restaurant. The scope of the analysis included areas that contain a low density of restaurants distant from other Asian restaurants. Limitations of my analysis are as follows: limited real estate and property availability information, limited population density information and while some analysis was done on crime, further analysis should be conducted at a greater level of granularity to validate my findings.

Due to the limitations of this analysis, this information should not be used to make a final decision but can serve as a starting point to narrow the scope of time and resource dependent in-depth analysis. Someone using this research to open a new Asian restaurant can be rest assured that these areas will at least be free of existing competition that can pose as a barrier to entry in the local restaurant economy.