

Optimal Location To open Japanese Restaurant in Houston

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a Japanese restaurant in Houston, US.

Since there are lots of restaurants in Houston, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Japanese restaurants in vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising 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.

Data

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

- 1) number of existing restaurants in the neighborhood (any type of restaurant)
- 2) number of and distance to Japanese restaurants in the neighborhood, if any distance of 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:

- 1) centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- 2) number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API
- 3) coordinate of Houston center will be obtained using Google Maps API geocoding of Downtown Houston

Neighborhood Candidates

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is approx. 40x40 kilometers centered around Houston city center.

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).

Then 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 - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific 'Japanese restaurant' category, as we need info on Japanese restaurants in the neighborhood.

Methodology

In this project we will direct our efforts on detecting areas of Houston that have low restaurant density, particularly those with low number of Japanese restaurants. We will limit our analysis to area ~20km around city center.

In first step we have collected the required data: location and type (category) of every restaurant within 20km from Houston downtown. We have also identified Japanese restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Houston - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Japanese 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 Japanese 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.

Analysis

We calculated the distance to nearest Japanese restaurant from every area candidate center (not only those within 300m - we want distance to closest one, regardless of how distant it is). To avoid less population density, we select locations with more than 2 restaurants and less than 10 restaurants. Then we 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.

Results and Discussion

Our analysis shows that although there is a great number of restaurants in Houston, there are pockets of low restaurant density fairly close to city center.

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 15 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Japanese 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 Houston Downtown but not crowded with existing restaurants (particularly Japanese) - 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.

Conclusion

Purpose of this project was to identify Houston areas close to center with low number of restaurants (particularly Japanese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Japanese restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified locations 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 decision 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.