Predicting the target locations to open a Chinese Restaurant in Pune, India

## 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 Chinese restaurant in Pune, India.

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

We will use our data science powers to generate a few most promising neighbourhoods based on this criterion. 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:

* number of existing restaurants in the neighbourhood (any type of restaurant)
* number of and distance to Italian restaurants in the neighbourhood, if any
* distance of neighbourhood from city centre

We decided to use regularly spaced grid of locations, centred around city centre, to define our neighbourhoods.

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

* centres of candidate areas will be generated algorithmically and approximate addresses of centres of those areas will be obtained using **Geopy Nominatim**
* number of restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**
* coordinate of Pune centre will be obtained using **Geopy Nominatim** of well-known Pune location (Shivajinagar)

## Methodology

In this project we will direct our efforts on detecting areas of Pune that have low restaurant density, particularly those with low number of Chinese restaurants. We will limit our analysis to area ~10km around city centre.

In first step we have collected the required **data: location and type (category) of every restaurant within 10km from Pune centre (Shivajinagar).** 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 Pune - we will use **heat maps** to identify a few promising areas close to centre 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 500 meters**, and we want **locations without Italian restaurants in radius of 800 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## Results and Discussion

Our analysis shows that although there is a great number of restaurants in Pune (~600 in our initial area of interest which was 20x20km around Shivajinagar), there are pockets of low restaurant density fairly close to city centre. Highest concentration of restaurants was detected east and west from Shivajinagar, so we focused our attention to areas south and north, corresponding to boroughs Swargate and south-east of Shivajinagar. Another borough was identified as potentially interesting (Mandalay Lane, north from Shivajinagar), but our attention was focused on Swargate which offer a combination of popularity among tourists, closeness to city centre, strong socio-economic dynamics and a number of pockets of low restaurant density.

After directing our attention to this narrower area of interest (covering approx. 5x5km south from Shivajinagar) we first created a dense grid of location candidates (spaced 300m apart); those locations were then filtered so that those with more than two restaurants in radius of 500m and those with a Chinese restaurant closer than 800m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centres 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 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 Pune centre 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.

## Conclusion

Purpose of this project was to identify Pune areas close to centre 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 (Swargate), 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 centres 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 neighbourhoods 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 neighbourhood etc.