**Capstone Project - The Battle of the Neighbourhoods (Week 5)**

**Introduction: Business Problem**

In this project we will try to find an optimal location for a breakfast point. Specifically, this report will be targeted to stakeholders interested in opening a **breakfast spot** near **richmond circle in Bangalore.**

Since there are lots of breakfast spots and eating joints near Richmond Circle, we will try to detect **locations that are not already crowded with breakfast spots**. We are also particularly interested in **areas with no breakfast spots in the vicinity**. We would also prefer locations **as close to richmond circle as possible**, if first two conditions are met.

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

* number of existing breakfast spots in the neighbourhood
* number of and distance to breakfast spots in the neighbourhood, if any
* distance of neighbourhood from richmond circle

We decided to use regularly spaced circular grids of locations, centred around Richmond Circle 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 **Google Maps API reverse geocoding**
* number of breakfast spots and their type and location in every neighbourhood will be obtained using **Foursquare API**
* coordinate of centre will be obtained using **Google Maps API geocoding** of well-known Richmond Town in Bangalore.

**Methodology**

In this project we will direct our efforts on detecting areas of Bangalore that have low breakfast spot density. We will limit our analysis to area ~3km around Richmond circle.

In first step we have collected the required **data: location and type (category) of every restaurant within 3km from Richmond Circle** (Under the Fly Over). We have also **identified breakfast spots** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of ‘**breakfast spot density**’ across different areas of Bengaluru - we will use **heatmaps** to identify a few promising areas close to centre with low number of breakfast spot in general (*and* less breakfast spot 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 breakfast spot in radius of 250 meters**, and we want locations **without breakfast spot 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 / 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 breakfast spot in Bangalore, there are pockets of low breakfast spot density fairly close to richmond circle centre. Highest concentration of breakfast spot was detected north and east from Richmond Circle, so we focused our attention to areas south, south-west and west, corresponding to boroughs nearby.

After directing our attention to this narrow area of interest (covering approx. 5x5km south-east from Richmond Circle) we first created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two breakfast spots in radius of 250m and those with breakfast spot 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 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 breakfast spot locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually optimal locations for a new breakfast spot! Purpose of this analysis was to only provide info on areas close to Richmond circle centre but not crowded with existing breakfast spot - it is entirely possible that there is a very good reason for small number of breakfast spot 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 Bengaluru areas close to centre with low number of breakfast spot in order to aid stakeholders in narrowing down the search for optimal location for a new breakfast spot. 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 breakfast spot. 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 breakfast spot 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.