**Capstone Project - The Battle of the Neighbourhoods (Week 2)**

**Applied Data Science Capstone by IBM/Coursera**

## Table of contents

* [Introduction: Business Problem](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#introduction)
* [Foursquare API Explore Function](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#item4)
* [Data](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#data)
* [Methodology](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#methodology)
* [Analysis](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#analysis)
* [Results and Discussion](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#results)
* [Conclusion](https://render.githubusercontent.com/view/ipynb?commit=7be05557e3487a41f4185befd204b95f13b98227&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f73757262686973617261776167692f757262616e2d63616c6c2f376265303535353765333438376134316634313835626566643230346239356631336239383232372f46696e616c25323043617073746f6e6525323050726f6a6563742532302831292e6970796e62&nwo=surbhisarawagi%2Furban-call&path=Final+Capstone+Project+%281%29.ipynb&repository_id=195231523&repository_type=Repository#conclusion)

Introduction: Business Problem 

In this project we will try to find an optimal location for opening a restaurant in Mumbai, India near very famous Taj Hotel located near 'Gateway of India'. Our sole objective is to find location that is surrounded by least number of restaurants and is nearest to Taj hotel that is a heritage ,five-star ,luxury hotel situated next to 'Gateway of India' an arch monument which attracts almost 6million tourists per year, making it the 30th most visited location worldwide. Since there are lots of restaurants in this location, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no restaurants in vicinity**. We would also prefer locations **as close to Taj or Gateway of India 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. 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.

**1.3 Interest**

Obviously, stakeholder would be very interested in accurate prediction location for new ventures and for getting their objectives fulfilled i.e. Attract most tourists visiting Gateway of India & hotel Taj Palace, for competitive advantage and business values.

**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)
* distance of neighbourhood from Taj hotel

We decided to use regularly spaced grid of locations, cantered around Taj hotel, 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 restaurants and their type and location in every neighbourhood will be obtained using **Foursquare API**

* coordinate of Taj hotel will be obtained using **Google Maps API geocoding**.

Firstly, we will do a general exploration using foursquare Api search query of food.

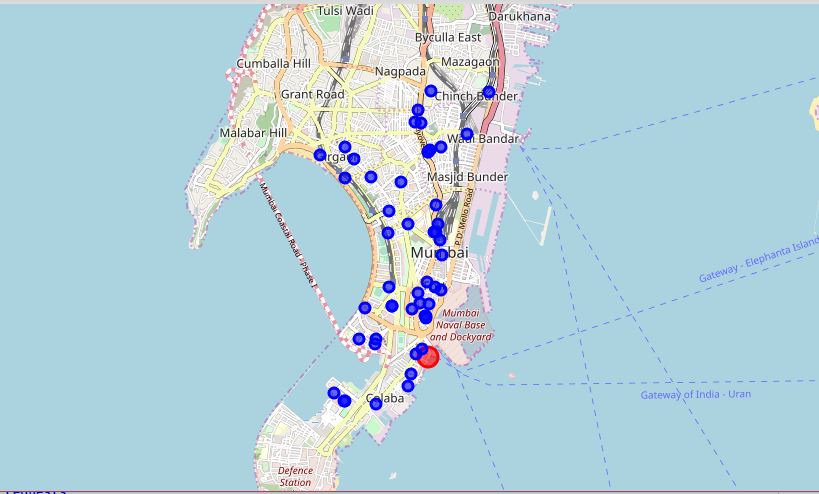
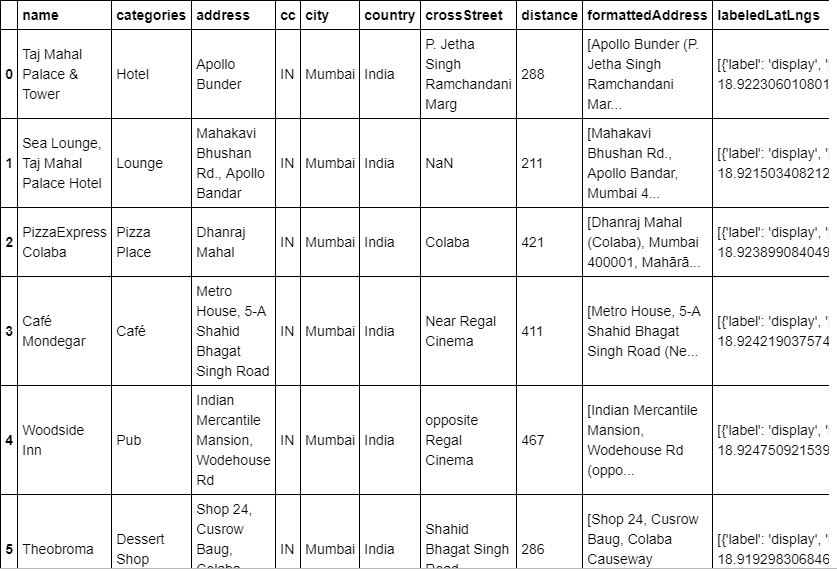
Geographical Co-ordinates of Taj, Mumbai: 18.9205619 72.831554

**Around 6km radius from Taj**

Total number of restaurants: 91

Average number of restaurants in neighbourhood: 0.497

**MAP representation of restaurants in 6km radius of Taj, Mumbai**

**General Exploration around the location of Taj and getting important venues.**

**Creating** **hexagonal grid of cells**: we offset every other row, and adjust vertical row spacing so that **every cell center is equally distant from all its neighbors**.

364 neighborhoods are created in 6km radius around Taj hotel.

Since around 4km in radius Taj there many restaurants so will try to find most optimal location for it. We would also prefer locations **as close to Taj as possible**. We will divide area around Taj into grids up to 6km radius and try to find are with least number of restaurants in those grids. 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. 12x12 kilometers centered around Taj.

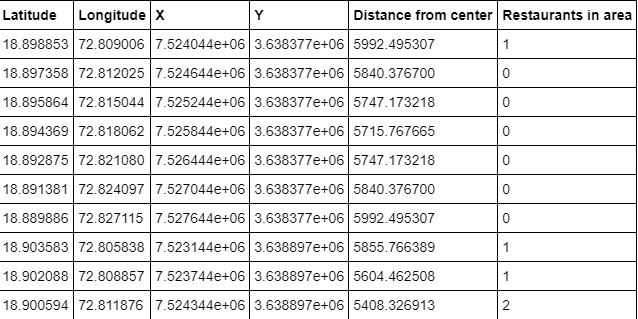
### Foursquare

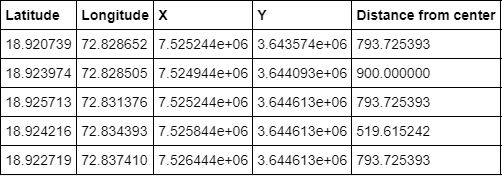
Now that we have our location candidates, we use Foursquare API to get info on restaurants in each neighbourhood.

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.

## **Analysis**

## Performing some basic explanatory data analysis and derive some additional info from our raw data. First let's count the **number of restaurants in every area candidate**: Around 300m radius of neighborhoods.

****

**Preferable locations found for setting up restaurant near Taj which are nearest and least surrounded by another restaurant.**

***Result***

Our analysis shows that there are many restaurants near Taj Hotel and we were mainly concerned about finding best suitable location that is not already crowded with restaurants and is nearest to Taj and very famous 'Gateway of India'. Total number (~91 in our initial area of interest which was 12x12km around Taj), there are large pockets of low restaurant density fairly close to Taj. Highest concentration of restaurants was detected north from Taj, so we found out the neighbourhoods with no restaurants and distance from Taj less than 1km and found out 5 such suitable neighbourhoods.

Result of all these zones containing largest number of potential new restaurant location. Purpose of this analysis was to provide information on areas close to Taj. 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 Mumbai areas close to Taj and Gateway of India with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. By calculating restaurant density distribution from Foursquare data we have first identified co-ordinates and neighbourhoods that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. 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.