# Tourist Advisor

## Introduction

Whenever we are planning to visit a new City/Town as a tourist, what would we typically do?

As a first step, we will try to gather the list of tourist attractions in that city. Then we will choose the ones that are of interest to us. For example, if we are interested on history and culture then we would be planning on visiting more of Museums. On the other hand, if we are interested on Art then we would be interested on visiting more of Art Galleries in that city. Basically, amongst the most attractive places to be visited in that city, one persons list could be different from the other. Once the places of interest are sort listed, then we will plan on the details like which places can be visited on which day etc. We will very likely group the places that are located close to each other and visit them all together in the same day. This needs the knowledge of the geographic location of the places being visited.

The Tourist Advisor tool, aims at assisting the tourist with this kind of a plan.

The Tourist, here after referred as user, is requested to enter the address of the City and the number of days they would be staying on the city. Tourist Advisor, does the analysis based on this input and produces the following results.

List of Tourist attractions that can be visited on each day of their stay. For example, if the user would be staying in that City for 3 days, then it would produce Day1, Day2 and Day3 itineraries.

City Map with the locations of the tourist attractions Marked. The places are marked with colors and each color representing the day on which the places are visited. For example, places that can be visited on Day 1 are marked with Red color, places that can be visited on Day 2 are marked with Green and places that can be visited on Day 3 are marked with Purple.

## Data

### Geocoders

This data is used to identify the latitude and longitude of the City.

### Foursquare

The following are the data that are needed to achieve the functionality of the tourist advisor tool.

1. List of venues in a specified a geographic location.
2. Categories of each venue. This is required since the tool is going to produce the result based on the user choice of travel categories.
3. Location data (latitude and longitude), rating and likes information for each venue. This is required to pick up the top N places to visit and to categorize them based on the proximity to each other.

Foursquare data is the main dataset being used to get the above data. Following are the APIs being used.

1. Venues/Categories API: This API is used to get the list of Venue Categories supported in the dataset. The tourism related Categories are selected from the received response and displayed to the user. The user can then select their interested categories from this list. For example, let us say the response received from the dataset contains the following categories: Restaurants, Shopping Malls, Museums, Sports Activities, Art Galleries, Parks and Palaces. Amongst them the user could choose, Museum, Art Galleries and Parks as their interested categories. Then the tool would pick up places to visit only from these categories.
2. Venues/Explore API: This foursquare API is used to get the list of venues that matches the selected categories in the mentioned location. The fields of interest for us from the response are Venue name, Venue ID, Venue Category, Venue Latitude and Venue Longitude.
3. Venues/Venue ID API: This foursquare API is used to get the details of each venue. From the response, the fields that are of interest to us are ‘Rating’ and the ‘Likes’. The Venues are then sorted based on this ‘Rating’ and ‘Likes’ values. From this sorted list, the top N venues are picked up.

Further, the top N venues would be segmented using the K-Means clustering algorithm. The attributes used for clustering would be the latitude and longitude of the venue. Hence we would get the tourist place segmented based on the geographical location. Each segment represents the day on which the places are visited.

## Methodology

**Step 0:**  Input the City being visited and the number of days of stay

**Step 1:**  Using geocoders Nominatim MAP APIs find the latitude and longitude values of the City.

**Step 2:**  Using Foursqaure Venues/Categories API, get the list of supported venue categories.

**Step 3:**  Filter tourist attractive categories from the response.

**Step 4:**  Display the tourism categories to the user, so that he can select the categories of interest to him.

**Step 5:**  Using the Foursqaure Venues/explore API, get the list of Venues matching selected tourism categories.

**Step 6:**  Convert the response received in the json format to Dataframe with columns: ‘Venue Name’, ‘ID, ‘Category’, ‘ Latitude’ and ‘Longitude’.

**Step 7:**  Using the Foursquare Venues/VenueID API get the details of each travel venue.   Add the columns ‘Rating’, ‘Likes’ to the dataframe based on the values received from the Venue Details.

**Step 8:** Sort the dataframe based on the Rating and Likes.  Pick up top N travel venues.  The number N is derived from the duration of the stay.

**Step 9:** Create a new dataframe by dropping ‘Venue Name’, ‘Id’, ‘Category’, ‘Rating ‘ and ‘Likes’ columns.  The dataframe would be having only Latitude and Longitude columns now.

**Step 10:**  Use K-Means clustering to segment venues based on the Latitude and Longitude values.  Number of clusters is equal to the number of days of stay.

**Step 11:**  Display the list of venues in each segment.  (The itinerary for each day)

**Step 12:**  Use Folium to visualize the travel venues on the City map.  Choose a color for each segment and mark the places with the segment color.

## Results

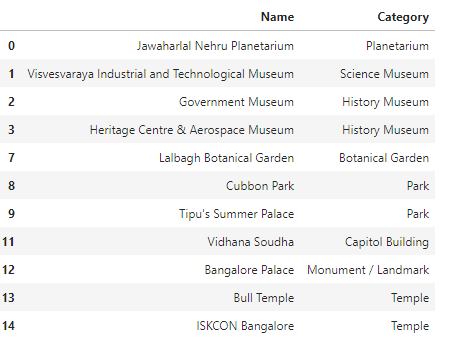
The results presented here are based on the following input.

City to be Visited: Bangalore, India

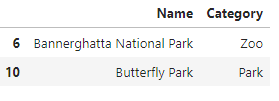
Days of stay : 3

Results from the tool:

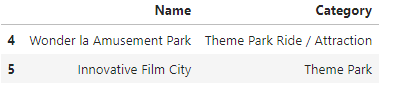
### Day 1 Itinerary



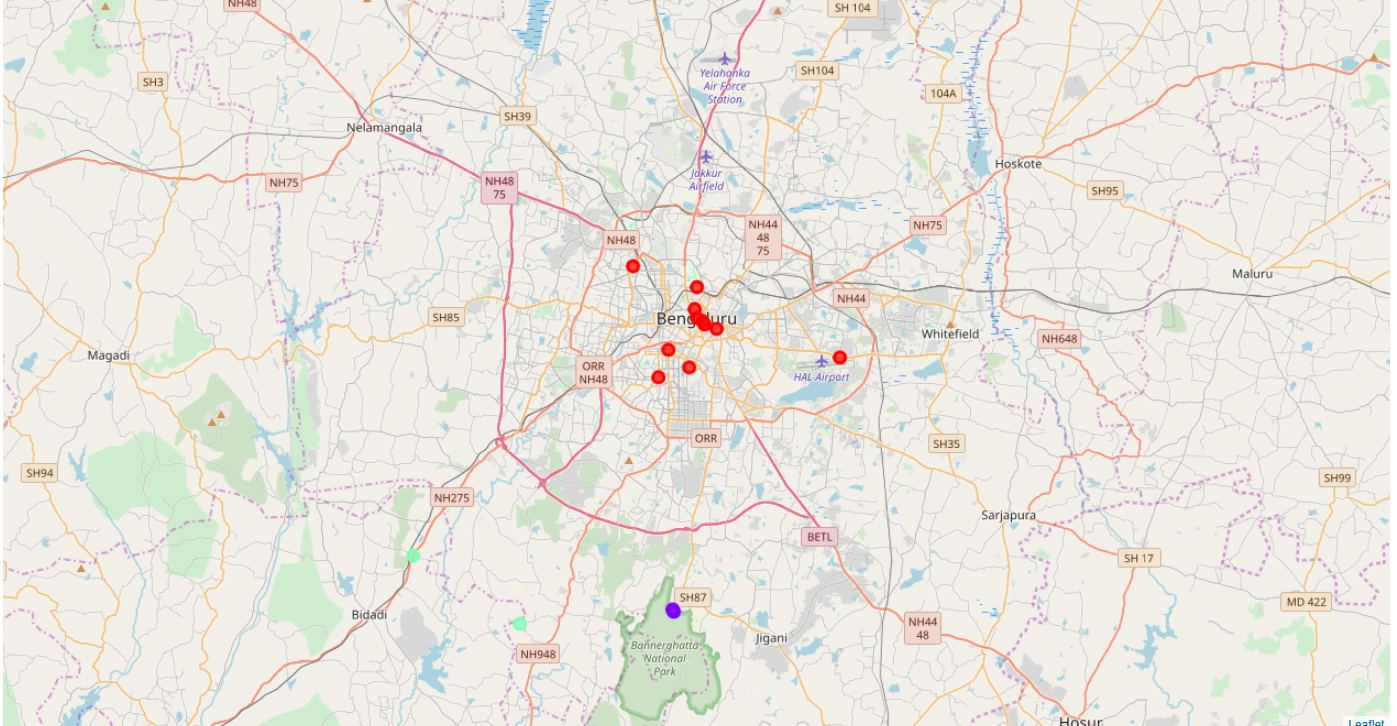
### Day 2 Itinerary



### Day 3 Itinerary



### Map



## Discussion

From the results, we can observe that the places inside the City are covered together on day1 and the places around the city are planned for the rest of the days.  Day 2 covers the areas in the South West side of the city and Day 3 covers the areas in the south part of the city.

One observation here is that, the number of places to be covered on Day 1 is long.  We have used only latitude and longitude attributes in the K-Means algorithm to do the clustering.  If we can add parameters like, popular hours to visit the place in a day, number of hours the users typically spend on each venue etc, then the algorithm could do further spilt on this list.  As these parameter are not available in the Foursquare dataset, this logic is not added.

## Conclusion

When someone is planning on vacation to an Unknown City, the planning becomes difficult due to the lack of knowledge about the city.  Hence this tool can really helps in easing that activity.  Additionally the tool takes into account, the user preference of venue categories, number of days stay, rating and likes.  Hence without the need for the user to refer multiple web pages and other related resource to arrive at the travel plan, the tool does the job for him.