**Exploring the Restaurant scenario in the Neighbourhood of Tirupathi**

Sree Ramakrishna Patanjali Molugu

May 25, 2020

1. **Introduction**
   1. **Background**

In this project we will put efforts in analysing the restaurant scenario based on the location and cuisines located in the temple city Tirupathi .Tirupathi is India’s top pilgrimage centre and has a floating population averaging from 55,000 to 1,00,000 everyday. There are various types of restaurants catering to the needs of the floating population. Most of the restaurants are clustered around the places where likelihood of foot fall is high (Temples, Pilgrim complexes, Travel complexes etc). Urban population has a very little choice for food options in the city, as most of the restaurants in the city are of Standard Indian Cuisine Keeping in the mind to cater to the needs of Pilgrims. In this project we will try to propose various categories of restaurants around a defined central location, based on the ratings, price range and distance from the central location of the City. We will fetch the data from Foursquare API and Zomato API ‘s and analyse the data .We will also plot various Interactive maps which will help understand the dynamics of a restaurants profile.

* 1. **Target Audience**

Entrepreneurs interested in investing in the business of Food & Beverages in Tirupati will benefit in understanding the different dimensions to be factored in while setting up business. We can recommend stakeholders by exploring the neighbourhoods based on the ratings, price range and proximity for each categories. We can recommend which category of Restaurant will stand out uniquely in the city.

We will use the data science tools and techniques to understand or weigh in the pros and cons of a location. We provide an analysis for the stakeholders to take a data driven decision to choose the best category/location/price range in the city about most promising and viable option.

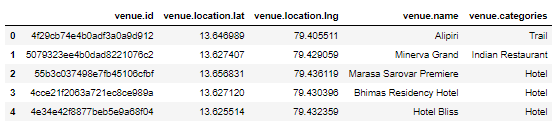
1. **Data Acquisition and Cleaning**
   1. **Data Sources**

We have sourced data using APIs to fetch various venues scattered around the city of Tirupathi within radius of 20 kms. We have used Foursquare API to fetch various venues irrespective of category. Using Zomato API, we have fetched all the restaurants located around 3 km radius of each of the venues and collected the restaurant names, cuisines, locations, rating and average price of restaurants.

**2.2 Data Description**

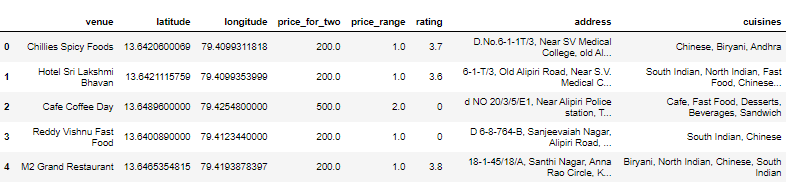
From Foursquare API (<https://api.foursquare.com/v2/venues/>) , I retrieved the following for each venue:

* **Name:** The name of the venue.
* **Category:** The category type as defined by the API.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.



From Zomato API (<https://developers.zomato.com/api>), I retrieved the following for each venue:

* **Venue Name:** The name of the venue.
* **Address:** The complete address of the venue.
* **Rating:** The ratings as provided by many users.
* **Price range:** The price range the venue belongs to as defined by Zomato.
* **Price for two:** The average cost for two people dining at the place. I later convert the same to average price per person by dividing by 2.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.
* **Cuisines:** Cusines of the venue



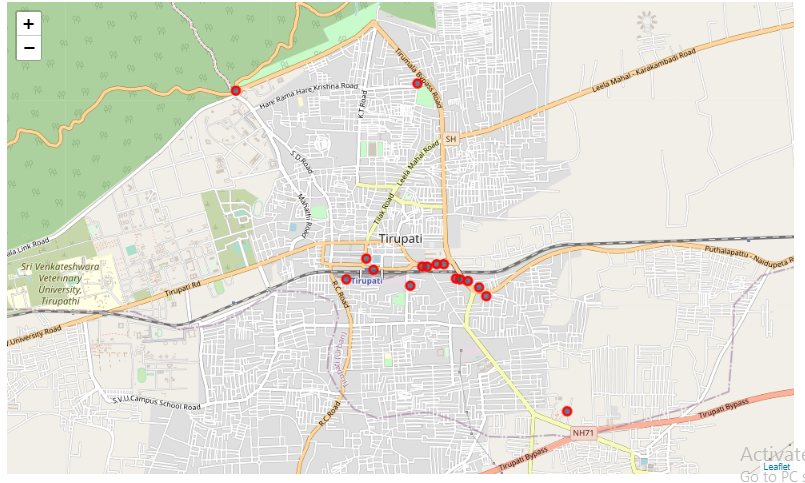


Figure 1:Venues fetched from Foursquare API

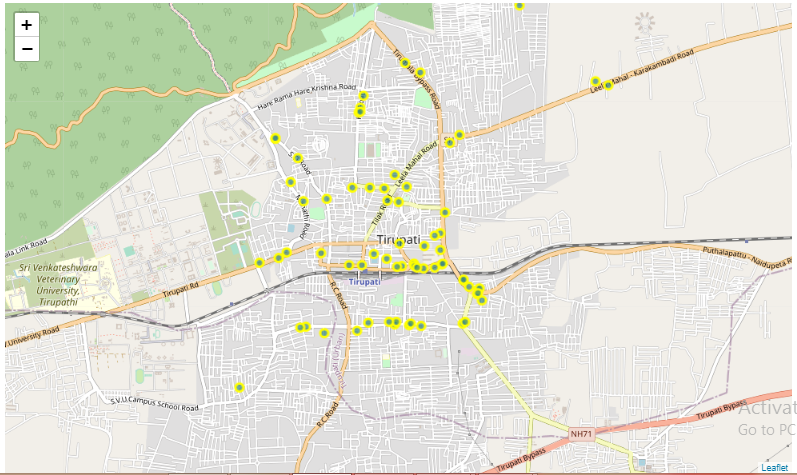


Figure 2:Restaurant venues fetched from Zomato API

* 1. **Data Cleaning**

Extracted Restaurant venues resulted in 419 venues. We have discovered from the data collected that there are many duplicate venues and venues from other cities. We have dropped the duplicates and filtered out the restaurants by the city name ‘Tirupati’ from the address. Final dataset has 80 restaurants from the city Tirupati without any redundant data.

1. **Methodology**

In this project we will try to propose various profiles of restaurants around a defined central location, based on the ratings, price range and distance from the central location of the City.   
We have selected a central location i.e NTR circle as the central location.

Steps followed for this project:

Step1: Get all the locations from foursquare api

Step2: Get all the restaurants with user ratings, passing the venues of four square api as an input

Step3: Calculate distance from the defined central location to each of the restaurants

Step4: Cluster the area based on ratings,distance,price range within 3 kms range from the central location for profiling the restaurants

1. **Exploratory Data Analysis**

**4.1 Rating**

We will explore the data based on the rating. From the above plots we can understand that the ratings of the restaurants are highly ranging from 3.2 to 4.1.About 13 restaurants has the rating '0'. Lets explore the subset of the dataset which has poor rating of 0.

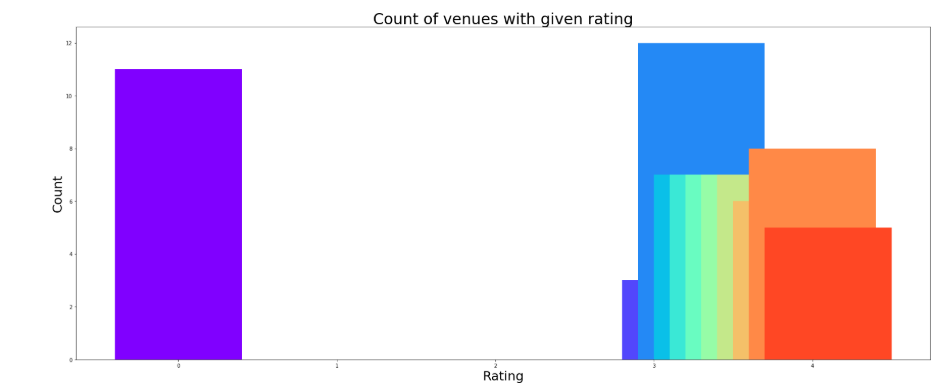


Figure 3 Rating vs count of Venues

**4.2 Cuisines for poor rating**

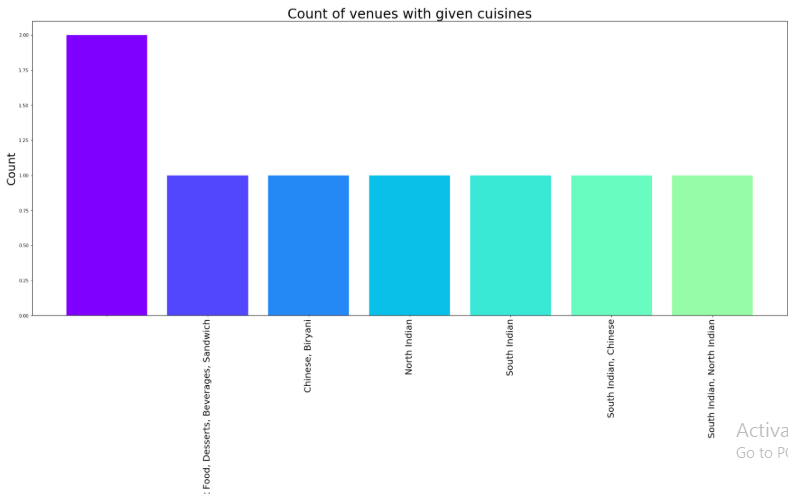


Figure 4: Cuisines of poor rated restaurants

From the above plot we can undestand that restaurants providing more or less same kind of cuisines have 0 rating.We can't assess which cuisine/category in particular has poor rating from the above plot.

**4.3 Price Range**

Price range for the distribution is pocket friendly in the city of Tirupathi.Zomato Price\_range explains cost range from 1 to 4(1 being the pocket friendly to 4 being expensive).

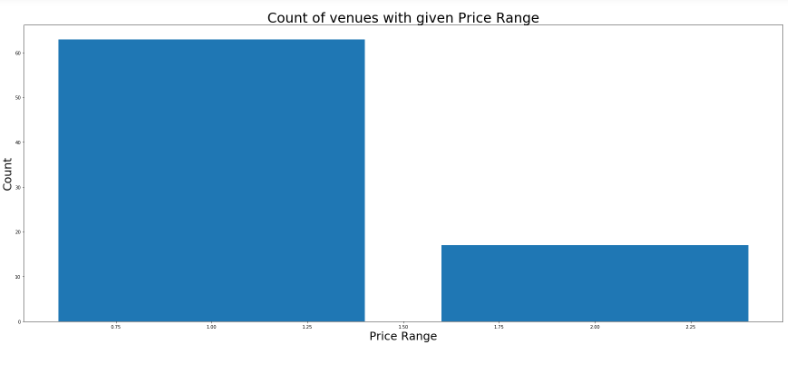


Figure 5Price Range of Restaurants in Tirupathi

**4.4 Price**

We will take a look at cost of two across all the venues.We can understand the range of prices in the city.

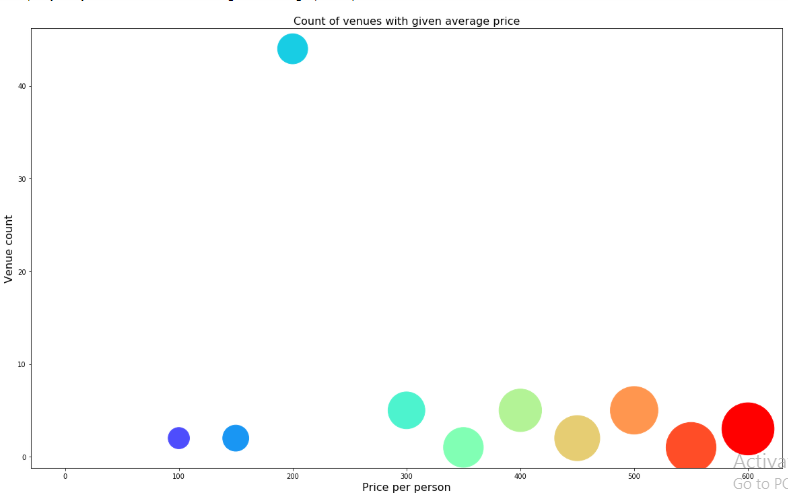


Figure 6 Price per person vs Number of Venues

Lowest price range for the restaurants in tirupathi is 100.Maximum price range for the restaurants is 600.Maximum number of restaurants has the price range of 200.

**5. Clustering and Analysis**

**5.1 Clustering**

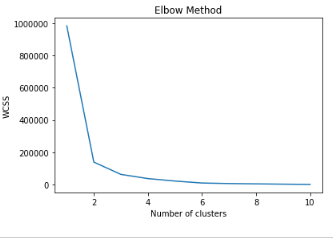
We are clustering based on the price,rating,price\_range and distance.We are using sklearn library ,k means clustering algorithm.To Find the best parameter we are iterating through a range of K values with the criteria of Within cluster sum of squares.

Figure 7 Parameter tuning by iterating through k values to find the best fit

We have selected 4 clusters as it has very less within the cluster variation.To avoid over fitting we have selected 4 clusters to build the clustering algorithm.

**5.2 Analysis**

**5.2.1 Cluster 0**

* Mean distance of a distance from central location of cluster 0: 1.4575 km
* Mean rating of cluster 0: 3.8375
* Mean price rate for an individual of cluster 0: 271.875
* There are 8 restaurants in cluster 0
* 14.29 % of restaurants are in cluster 0

**5.2.2 Cluster 1**

* Mean distance of a distance from central location of cluster 1: 1.5399999999999998
* Mean rating of cluster 1: 3.6315789473684212
* Mean price rate for an individual of cluster 1: 97.36842105263158
* There are 38 restaurants in cluster 1
* 67.86 % of restaurants are in cluster 1

**5.2.3 Cluster 2**

* Mean distance of a distance from central location of cluster 2: 1.4283333333333335
* Mean rating of cluster 2: 3.8166666666666664
* Mean price rate for an individual of cluster 2: 208.33333333333334
* There are 6 restaurants in cluster 2
* 10.71 % of restaurants are in cluster 2

**5.2.4 Cluster 3**

* Mean distance of a distance from central location of cluster 3 1.315
* Mean rating of cluster 3: 3.8
* Mean price rate for an individual of cluster 3: 156.25
* There are 4 restaurants in cluster
* 7.14 % of restaurants are in cluster 3

Interactive maps are generated with different color mappings for each of the cluster.

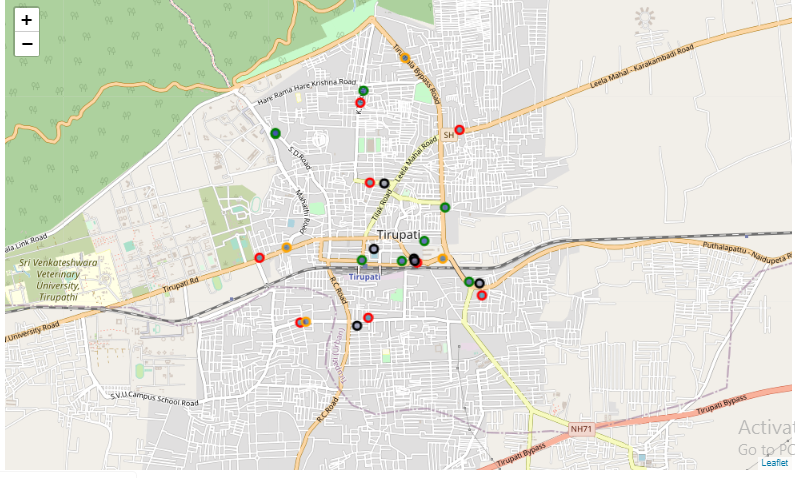


Figure 8 Interactive plot with different clusters

**6. Results and Discussion**

Based on the analysis we can draw many conclusions. We have profiled the clusters based on the price\_Range,rating,average price for two and distance from the centre.

From cluster 0 ,we can understand that the 14.29% of the restaurants lies within the mean distance of 1.67 kms(We have calculated the distance using geographical co ordinates(however traffic and road travel distance has to be considered for precise analysis).Ratings of the venues in this cluster are at 3.8375 which explains the quality of the restaurants located in this cluster. Price range for individuals is high with the mean price range at rs./271.875.Distance from the central location is very fairly closer . We can profile this cluster as Mid price range profile with high quality food/Customer service.

From cluster 1, 67.86% of the restaurants lies within the mean distance of 1.53 kms.Ratings of the venues in this cluster is 3.63.Price range of this cluster is at rs/-97 ~ 100.We can profile this cluster as Midrange and Pocket friendly with cuisines like Biryani,South indian,North Indian and Fast foods having high number restaurants. Pizza and Burger Joints are comparitively lesser in this cluster with rating of 3.72.None of the Pizza and Burger Joints in this cluster has a Franchisee of its own and doesn’t offer home delivery.

From cluster 2, we can understand that 10.71% lies within the mean distance of 1.42 kms. Ratings of the venues in this cluster is 3.81. Price Range of this cluster is Rs/-200 approximately. We can see that this cluster has pizza franchise like dominos and KFC.KFC and dominos offers home delivery.

From cluster 3, Mean distance from central location is close to 1.315 kms and has good rating of 3.8.Price range of cluster 3 is 156.25 which can be deemed as short distance mid range profile.

**7. Conclusion**

The purpose of the project is to explore the neighborhood of Tirupathi and understand the restaurant scenario. The venues have been plotted with the clusters marked in interactive maps. Restauranteers can understand the various profiles in the city and thus weigh in the pros and cons of various profiles so that they can take informed decisions.