IBM CAPSTONE PROJECT  
Housing prices in Chennai

By Uma Sankar G

Contents

[Introduction 4](#_Toc74339832)

[Business Problem 4](#_Toc74339833)

[Data 4](#_Toc74339834)

[Strategy for problem solving 4](#_Toc74339835)

[Data Cleaning 5](#_Toc74339836)

[Visualization 6](#_Toc74339837)

[Foursquare API 6](#_Toc74339838)

[Data Transformation 7](#_Toc74339839)

[Elbow method 7](#_Toc74339840)

[KMeans Clustering 8](#_Toc74339841)

[Clustered points 8](#_Toc74339842)

[Conclusion 9](#_Toc74339843)

[Figure 1 5](#_Toc74339813)

[Figure 2 5](#_Toc74339814)

[Figure 3 6](#_Toc74339815)

[Figure 4 6](#_Toc74339816)

[Figure 5 7](#_Toc74339817)

[Figure 6 7](#_Toc74339818)

[Figure 7 8](#_Toc74339819)

[Figure 8 8](#_Toc74339820)

# Introduction

Chennai is the capital of Indian state of Tamil Nadu. It is the one of the largest cultural, economic and educational centres of south India. It is the six most populous city in India.

Chennai has the fifth largest urban economy of India.

It is also called Detroit of South Asia due to large number of automobile industries.

# Business Problem

When you buy a house, a lot of factors come into play, important points of these are the price per square feet and the locality in which the house is situated. The accessibility it is has to essential services also plays an important role too.

In this project, we are going to use Kaggle dataset <https://www.kaggle.com/sonukiller99/>[indian-house-price-combined](https://www.kaggle.com/sonukiller99/indian-house-price-combined) to determine affordable housing for in the price range up to Rs. 100000 in the city of Chennai.

We are going to consider affordable houses having 3 bedrooms or more.

Clustering neighbourhoods based on venues. Such that venues that are mostly frequented are found.

# Data

We will be using data set from <https://www.kaggle.com/sonukiller99/indian-house-price-combined>.

It contains data of India’s metropolitan cities consisting of price of each neighbourhood area with details containing location, plot area, number of bedrooms, availability of lift, swimming pool, etc. are mentioned along with the price of the house.

This data set contains prices of houses from 6 major metropolitan areas.

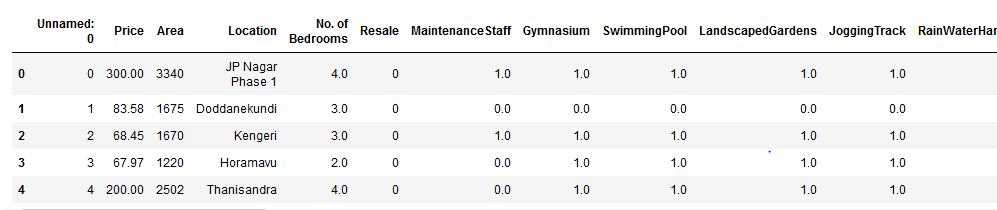
To determine the value of locality we will explore the revenues using Foursquare API ([https://api.foursquare.com](https://api.foursquare.com/)) to explore various revenues near each locality in Chennai

# Strategy for problem solving

* We will collect the housing data from Kaggle set.
* We will explore venues for each locality using Foursquare API.
* We will merge venues with the housing data from Kaggle set.
* We will find best localities for affordable housing.
* With Folium library we will analyse the results.

# Data Cleaning

* Download the Kaggle data set and select city Chennai as a pandas DataFrame as shown in Figure 1



Figure

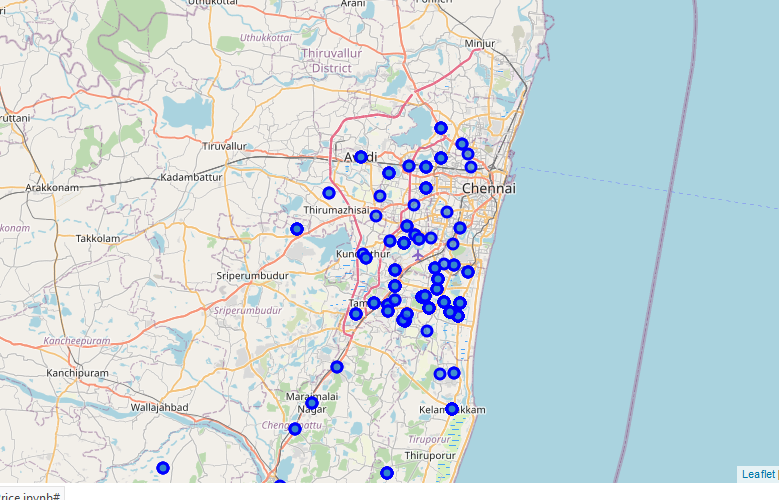
* Drop the columns that are irrelevant and set the No of Bedrooms to 3 and Price to equal or lesser than 100 lakhs as shown in Figure 2.



Figure

# Visualization

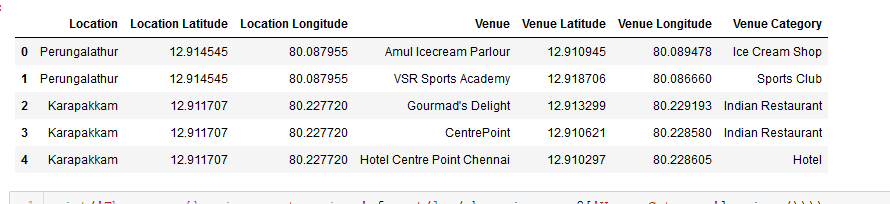
* Cluster the areas using Folium package and the data was created as per Figure 3



Figure

# Foursquare API

* The Foursquare company provides location data. Their API allows developers to extract the venues from the location with its latitude and longitudes and find the venues name with its latitudes and longitudes as shown in Figure 4



Figure

# Data Transformation

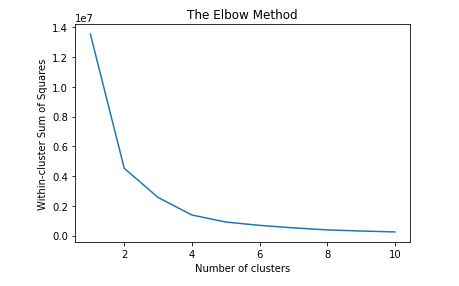
Now transfer the data in the below format as shown in Figure 5



Figure

# Elbow method

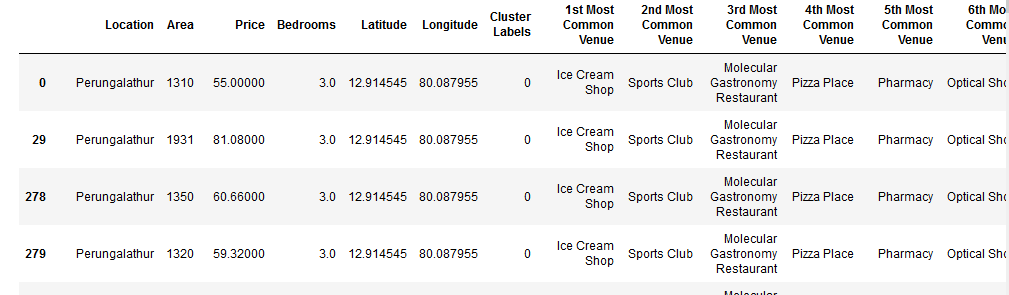
* The Elbow method is a heuristic that is used to determine the number of clusters in a dataset. Using the elbow method, we find the number of clusters to run the KMeans clustering. As per the Figure 6, K = 6 is the optimal point the data’s can be clustered.



Figure

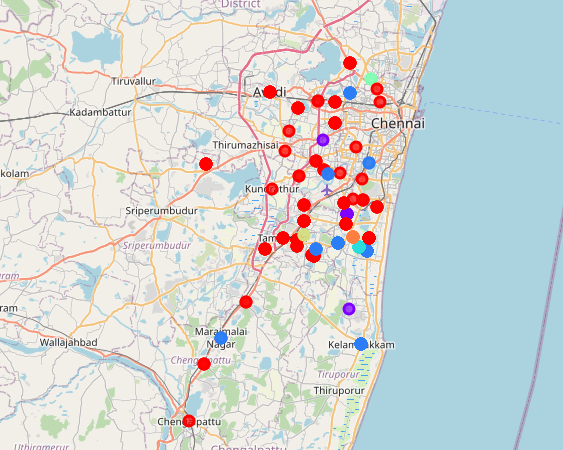
# KMeans Clustering

* Using the k value find the clusters for the localities. The clusters are labelled from 0 to 6. The Figure 7 shows Locations that are clustered labelled with 0.



Figure

# Clustered points



Figure

# Conclusion

* In Cluster 0, there are lot of ice cream shops and restaurants. There is a pharmacy. In some place there is train station will be useful for daily commuters. The
* In Cluster 1, there are lot of Grocery shops and restaurants suitable for all kinds of people
* In Cluster 2, there are lot of restaurants.
* In Cluster 3, the most frequent place used is Pizza shop.
* In Cluster 4, the Smoke shop is the most frequent place.
* In Cluster 5, the most frequent place is lake
* In Cluster 6, the place is nearer to farmers market.
* I will recommend Cluster 6 because it has Farmers market, Pharmacy, Multiplex and Movie Theater. Also, the price of the housing is affordable. With time the value of the house can increase.