# The Absolute City Guide for Chennai, India

## By: Aniruth R

## November 2019

### Introduction

For many people visiting Chennai for a variety of reasons, they often find it struggling to know where to visit in the city for there isn't a proper channel available to see the specialities of the area they are currently in and what can they expect in the area or which area is better suited for their needs. This assignment tries to identify this issue and come up with a solution for the same.

### Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Chennai, India to explore several areas of interest. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the
question: In the city of Chennai, India, if a visitor is looking for a particular requirement, where would you recommend that they visit?

### Target Audience of this project

This project is particularly useful for local travellers and tourists looking to different amenities in the city of Chennai.

### Data

To solve the problem, we will need the following data:
• List of neighbourhoods in Chennai. This defines the scope of this project which is confined to the city of Chennai
• Latitude and longitude coordinates of those neighbourhoods. This is required in order to
plot the map and also to get the venue data.
• Venue data, particularly data related to user interests. We will use this data to perform clustering on the neighbourhoods.

**Sources of data and methods to extract them**

This Geonames.org page (http://www.geonames.org/export/zip/) contains a list of places in Chennai, with a total of 65 places. We will not use web scraping as the data is readily available in the required format. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the
Neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the venues data in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.