|  |  |
| --- | --- |
| A picture of a winding road and trees  India Growth Analysis through clustering | Abstract  What Clustering Based on Foursquare Neighbourhood Data Can Tell Us About India Growth Story?  Pranal Dongare  IBM Data Science Capstone Project |

Table of Contents

[Introduction 2](#_Toc533236820)

[Data 2](#_Toc533236821)

[Step 1 : Getting the co-ordinate details of Indian Cities 2](#_Toc533236822)

[Step 2 : Ranked List of Top 110 Indian Cities Based of Government of India’s Livability Index 2](#_Toc533236823)

[Step 3 : Data Cleanup and Manipulation 3](#_Toc533236824)

[Step 4 : Retrieval of Foursquare Neighbourhood Data 3](#_Toc533236825)

[References 3](#_Toc533236826)

## Introduction

India (Bharat, in local terms) is one of the fastest growing major, emerging economy in the world. It is pegged by many analyst to grow at more rapid pace (In Percentage Terms) than any other major economy in near future. This along with beneficial demographic dividend makes India great option for investment.

However, given it’s large size and diversity, it will be interesting to see how much different cities in India are similar or dissimilar to each other. This will also be very interesting analysis as we might get few insights about India’s growth story along the way.

Hence, in this project, I will pursue to segment major Indian cities on the basis of their Foursquare neighbourhood data and then try to identify any trends or insights from the segmentation. We might not get any particular insights after the exercise but I am sure journey is worth taking.

So, I invite you to start this journey with me! Let’s go!

## Data

Data is new oil! Of course! Let me explain the data which I will be using in this project.

### Step 1 : Getting the co-ordinate details of Indian Cities

First of all, the complete list of major Indian Cities along with details of their latitude and longitude information is created in excel file. I have referred below Github repository to create the information.

<https://gist.github.com/gsivaprabu/5336570>

The sample data from the excel sheet “Indian Cities Coordinates” is something like below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **City** | **Latitude** | **Longitude** | **State** |
| 1 | Port Blair | 11.67 N | 92.76 E | Andaman and Nicobar Islands |
| 2 | Adilabad | 19.68 N | 78.53 E | Andhra Pradesh |
| 3 | Adoni | 15.63 N | 77.28 E | Andhra Pradesh |
| 4 | Alwal | 17.50 N | 78.54 E | Andhra Pradesh |
| 5 | Anakapalle | 17.69 N | 83.00 E | Andhra Pradesh |

The link to entire dataset is [this](Indian%20Cities%20Coordinates.xlsx).

### Step 2 : Ranked List of Top 110 Indian Cities Based of Government of India’s Livability Index

Next step was to get the list of top 110 cities in India as per the Government of India releases “City Livability Index 2018”. The list was found at below location and was converted into Excel file.

<https://indianexpress.com/article/india/full-list-of-indias-ease-of-living-index-see-where-your-city-is-ranked-5306185/>

The sample data from the excel sheet “Indian Cities Livability Index” is below:

|  |  |
| --- | --- |
| **No** | **City** |
| 1 | Pune |
| 2 | Navi Mumbai |
| 3 | Greater Mumbai |
| 4 | Tirupati |
| 5 | Chandigarh |

The entire data can be found [here](Indian%20Cities%20Livability%20Index.xlsx).

A side note : Pune, my home city, has been ranked first on the list! ☺

### Step 3 : Data Cleanup and Manipulation

Now, comes the exciting part. We will start with these two datasets as base and get to the Python and Pandas for further data manipulation.

As a cleanup, we will cleanup table containing latitude and longitude details and then left join two tables on “Indian Cities Livability Index”.

This will be our input table to get the “Foursquare” neighbourhood data for the cities.

### Step 4 : Retrieval of Foursquare Neighbourhood Data

Once we have “Foursquare” data, we will cluster the cities based on the closeness to each other. For example, cities having identical weightage for features like “Pubs” or “Airport Lounge” would be clustered together.

Once we have this data ready. We can then proceed with the K-Means clustering and analysis with the help of Folium chart.

## References

1. Latitude and Longitude Data : <https://gist.github.com/gsivaprabu/5336570>
2. Indian Cities Livability Index : <https://indianexpress.com/article/india/full-list-of-indias-ease-of-living-index-see-where-your-city-is-ranked-5306185/>