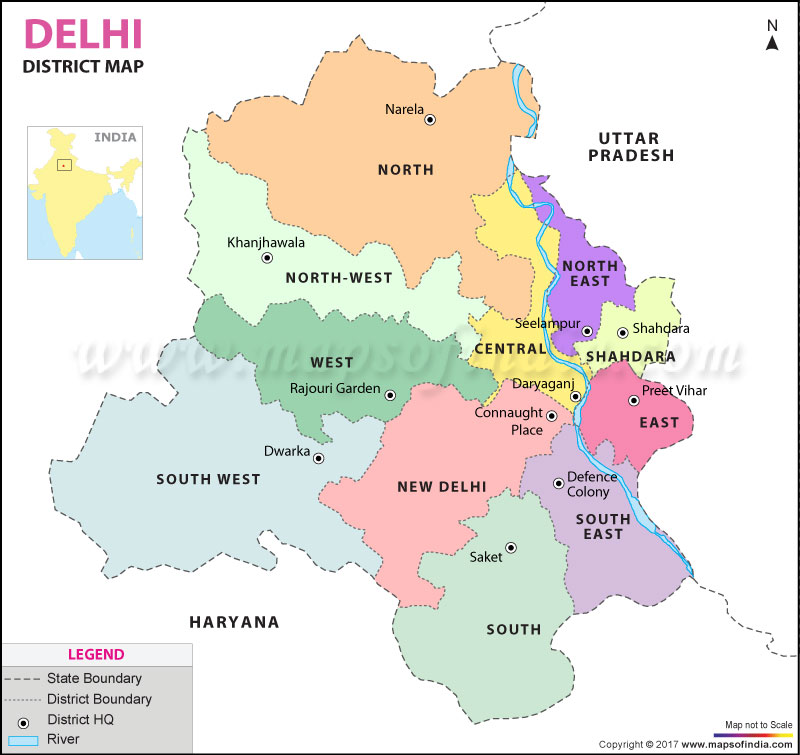
**Location Preference Search in Delhi**

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwi53JmPktngAhXMMo8KHcfwC3UQjRx6BAgBEAU&url=/url?sa%3Di%26rct%3Dj%26q%3D%26esrc%3Ds%26source%3Dimages%26cd%3D%26ved%3D%26url%3Dhttps://www.mapsofindia.com/delhi/districts/%26psig%3DAOvVaw1lZGVuuQfcd8nNUqkY2XQe%26ust%3D1551261651345440&psig=AOvVaw1lZGVuuQfcd8nNUqkY2XQe&ust=1551261651345440)

# Introduction

## Background

Delhi, India’s capital territory, is a massive metropolitan area in the country’s north. Delhi, the capital of India, is situated in northern India and stands on the west bank of Yamuna River bounded by Uttar Pradesh and on the north, west and south by Haryana. Delhi is spread over an area of 1483 sq. kilometers, 216 meters above sea level and has a population of around 14 million.

Delhi is a lively city reflecting a perfect blend of modernization and traditional architectures. Being a capital city of the Republic India, Delhi is the center of Government's legislature and judiciary systems. Set on the both sides of River Yamuna, Delhi is seen as one of the fastest growing cities in India. Delhi, being a multi-cultured and multi-linguistic city, lures a large number of people not just from India but abroad as well.

## Problem

Until 2030, the population is supposed to grow around 8.5%, predominantly because of inflowing migration. A lot of professionals, students, government officers move to Delhi for various reasons. Before moving to a new place, in particular in Delhi, people explore new places in several respects. Like student chose to stay in cheap places. Being a metropole in India, Delhi consists of 9 zones. All of them share commonalities and have regional distinctions. That requires a systematic neighborhood analysis of various factors. By considering multiple and heterogeneous factors, this analysis supports the decision of people, e.g. student, professional who move to a new place in Delhi.

## Interest

The project demonstrates how instruments of data science support a systematic neighborhood analysis by processing and jointly analyzing heterogeneous data sources. The respective notebook provides a structured data analysis in order to make the informed decision where to move in Delhi.

# Data acquisition and cleaning

## Data Sources

From the data point of view this project contains various data sources. It is very difficult to get spatial data in India. An effort was made to collect the data manually from various sites. Data collected manually was.

* District
* Neighborhood
* Pin code
* Latitude
* Longitude

The sites used to collect manual data are as follows:

* https://indiamapia.com/
* http://www.diva-gis.org/

To get the venue related data, Four Square API are used. Four Square provides the location intelligence data.

* <https://foursquare.com>

Location and venue data was collected using the RESTful APIs of Foursquare

## Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There are several problems with the datasets. Same neighborhood was having different names with minor spelling changes.

Data collected manually had multiple neighborhood in the same pin code. This was resulting in data duplication at pin code level. To remove data delicacy, only one Neighborhood was selected for each pin code. And search radius is increased to provide the whole coverage of area in a neighborhood.

After fixing these problems, I checked for outliers in the data. After all cleaning, I plot the graph for all locations. I found there are some neighborhood which belonged to nearby states like Haryana. This may be due to wrong spatial data. Those data were removed.

# Exploratory Analysis

The analysis consists of three main steps:

First, at the central coordinates of neighborhood in Delhi, more than 13,000 surrounding venues are collected for a systematic assessment. The venues in a venue category, such as hotel, park, supermarket, allow a ranking of the top ten most common venue categories at each neighborhood. This information describes the structures of venues within the boroughs.

Second, this analysis provides an aggregated view on the structure of boroughs by employing a cluster analysis as part of unsupervised machine learning.

Third, "top rated venues" show places in the boroughs that attract the attention of Delhi inhabitants.

These steps support the analysis of neighborhoods regarding personal preferences, extract interesting places and can result in a selection of potential locations. The approach requires the implementation of external data sources that provide the features, needs a comparative analysis of these features as well the visualization of geographic information in maps.

# Conclusion

# Cluster Color Index

|  |  |
| --- | --- |
| **Cluster** | **Colour Index** |
| Cluster 1 | RED |
| Cluster 2 | BLUE |
| Cluster 3 | INDIGO |
| Cluster 4 | GREEN |
| Cluster 5 | ORANGE |

# Modes of Venue Categories

|  |  |
| --- | --- |
| **Cluster** | **Venue Category** |
| Cluster 1 | Train Transport, Restaurant, Pizza Stores |
| Cluster 2 | ATM, Women Store, Fabric Shop |
| Cluster 3 | Hotel, Snack Place, Cafe |
| Cluster 4 | Botanical Garden, Women’s Store, Fabric Shop |
| Cluster 5 | Various Types of Restaurants, Stadium, Tourist Spots |

# Analysis

|  |  |
| --- | --- |
| **Cluster** | **Comparison** |
| Cluster 1 | Good if someone need public transport services, budget outings |
| Cluster 2 | Basic amenities with family daily needs |
| Cluster 3 | More like a High class outing |
| Cluster 4 | Good for who likes to stay near park and for nature refresh |
| Cluster 5 | Good for those who are interested in sports activities |