

# Comparing 2 cities of dreams: Mumbai and New York

*Rikki Mohanty*

31 May, 2020

## 1. Introduction

### 1.1 Background

*Let's find about the cities which are part of this project*

**Mumbai** is the second most populous city in India and the seventh most populous city in the world with a population of 19.98 million in 2018. Mumbai is the financial, commercial and entertainment capital of India.

**New York City (NYC)** is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km<sup>2</sup>). New York City has been described as the cultural, financial, and media capital of the world.

So I have decided to use these cities, explore and compare the neighborhoods and find if these cities have any similarities.

### 1.2 Target Audience

*What type of clients or a group of people/stakeholders would be interested in this project?*

1. People who are visiting these cities can make the best of city experience; also find the similar places for comfort.
2. Business personnel who want to invest. This analysis will give them an idea of where to invest.
3. People who are migrating to these cities will have better ideas where to settle down, which places have the right resource and others.

## **2. Data Description**

### **2.1 Data acquisition and cleaning**

1. I have taken the dataset of New York from the Capstone Project and found their respective coordinates.
2. For Mumbai city: the data availability is infrequent and dispersed in many places, so I've manually scraped the list of neighborhoods from this Wikipedia page <http://zipcodepincode.com/India/Maharashtra/Mumbai/Mumbai/index.html>. For this, I've used requests and BeautifulSoup4 library to create a dataframe with coordinates and pin codes which was manually scrapped from web.
3. I have used the Foursquare API to explore the neighborhoods of both the cities and segmented them.
4. These venues are then clustered using k-means. Found the most common venues (MCV) and finally compared the (MCV) of both cities to look for similarities.

### **2.3 Data with Examples**

#### **2.3.1 Programming Section (Initial Processing: Scraping from the web and getting the coordinates)**

- **Please find the notebook in github repository**