**Quality of Living assessment for Indian Cities**

**Introduction**

An assessment of quality of life in any place plays an important role in many important decisions people make. One of the key reasons of wanting to know this information is to make right career choice and ask for a compensation that is as per the standards of the city. While assessment of quality of life across various major cities in the worlds is available, courtesy efforts of international agencies like Mercer, such that is not available for smaller cities at country level. In this project our focus will be on assessing the quality of living across major cities in India.

As per Mercer’s quality of living reports, the assessment of quality of living in any place depends on the following availability of factors :

* Consumer goods
* Economic environment
* Housing
* Medical and health considerations
* Natural environment
* Political and social environment
* Public services and transport
* Recreation
* Schools and education
* Socio-cultural environment

In a population heavy country like India, another factor is per person availability of the above mentioned facilities. Thus another factor that may play an important role in defining quality of life would be population of the city.

In this project our objective is to :

**“Classify major Indian cities based on different levels of quality of living offered by them to a citizen”**

**Data Requirement**

To achieve this objective the data required would be as follows :

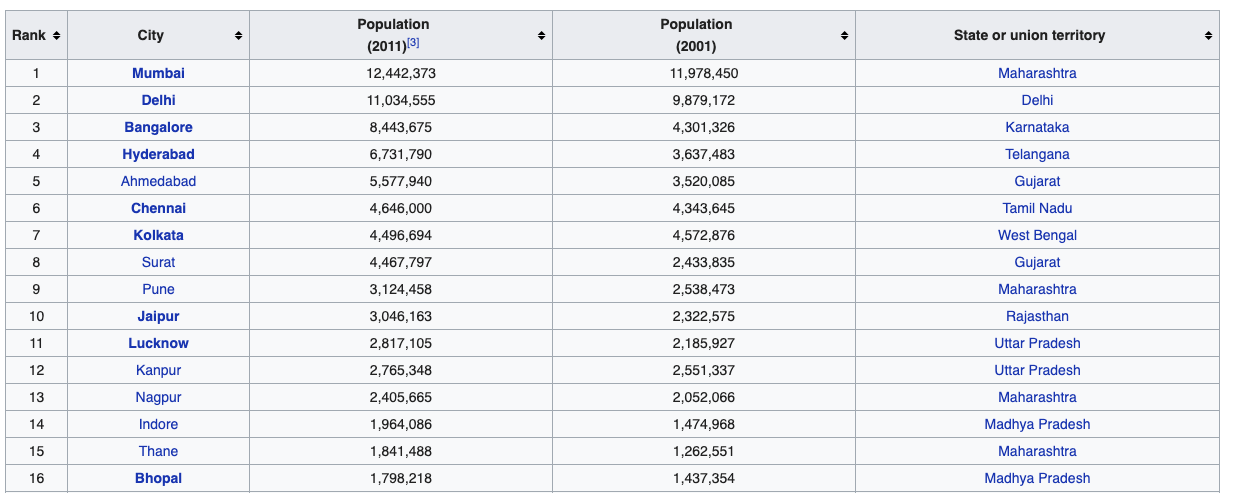
* List of Indian Cities

To start with we will need a list of various Indian cities that we will be classifying under this project. Such list is available on the following Wikipedia page :

<https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population>

This page gives a list of 300 cities of India along with respective populations as per 2011 census and the states these cities belong to.

A sample screen shot is given below :



We will scrap this data from the Wikipedia page using a python code. For the purpose of this project we will limit our analysis to first 50 cities mentioned in the list.

* Latitude & Longitude of the cities

We will use the geocoder package in Python to extract latitude and longitude of each city.

* Foursquare data

Upon entering the details of each city in the Foursquare API we get a list of venues in that city. The venues have different categories such as cafes, diners, schools, parks, metro station, libraries etc. We will use this data and classify the results obtained for each city into following categories. What all venues should be categorised under each of the following categories will be decided after content analysis of all the unique category of venues obtained for all the cities.

* *Recreation*: No of parks, cinema halls, cafes, theatres, restaurants, community centres, zoos etc. for each city.
* *Schools & Education* : No of schools, colleges, universities, learning centres for each city
* *Public Services & Transport* : Number of metro stations, bus stops, taxi stands for each city
* *Medical and health considerations:* No of clinics, hospitals, pharmacies, healing centres, gyms, yoga centres for each city
* *Consumer goods:* Number of shopping centres, supermarkets, shops etc. for each city
* Political and social environment : An important criteria under political and social environment which can be of great concern to people could be the crime rates in the city. To get this data we will mine the Wikipedia page :

<https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_crime_rate>

Since the data given in this page is as state level and not city level, we will map this data to each city depending on the state each city belongs to.

A combination of all the above mentioned data points will then be used in the next stage of analysis to classify cities as per the quality of living offered by them.

**Methodology**

**Web Scraping :**  For data collection from the Wikipedia Page, we have used web scraping techniques using the ‘Beautifulsoup’ package. This package fetches the html code behind any web page and then individual component of the web page are explored to extract the exact data required. Using this technique we were able to extract the details of 25 Indian cities containing the following data for each city :

* Name of the City
* State to which the city belongs
* City population in 2011
* City population in 2001

Similar process was then used to extract the crime data. Data for following fields was downloaded :

* Name of the state
* Crime incidences in years 2014, 2015 & 2016
* Percentage share of state in 2016
* Rank based on Incidence % share (2016)
* Rate of cognizable crimes (2016)
* Rank based on crime rate (2016)

**Geocoder :** In order to extract co-ordinates of each city in terms of its latitude & longitudes, geocoder library was used. The results for ‘Vishakhapatnam’ city were not found by the Geocoder, because of which we had to drop this city from our analysis.

**Foursquare Api:** In order to extract information about other factors such as educational facilities, recreational facilities, public services, Medical & Health care, socio-cultural environment etc, we used Foursquare Api. We fed the co-ordinates of each city to the Api and collected all the resulting values for each city. This gave us a collection of various different venues that were resulted for each city on four square.

**Exploratory Analysis:** A total of 166 unique venues were found for the 25 cities. A detailed manual study of these venue categories revealed that these can be categorized into following three categories .

* Education : Contains venues like University, Book store, School etc
* Public Service : Contains venues like parks, monuments, grounds, Train station etc
* Recreation : Contains venues like Hotels, Deli, Coffee Shops etc

So we recoded the existing categories for each of the venues into one of the above three categories.

**Quality of Living Index:** In order to create quality of living index for each city we first calculated the per capita levels of Education, Public Services and Recreation by dividing the previous scores of each city with their respective population. There are many techniques for creation of Index mainly based on identifying the weights that should be given to each factor. For the purpose of this project we assume that the weight to each of the factors is equal and shall consider a quality of living index as :

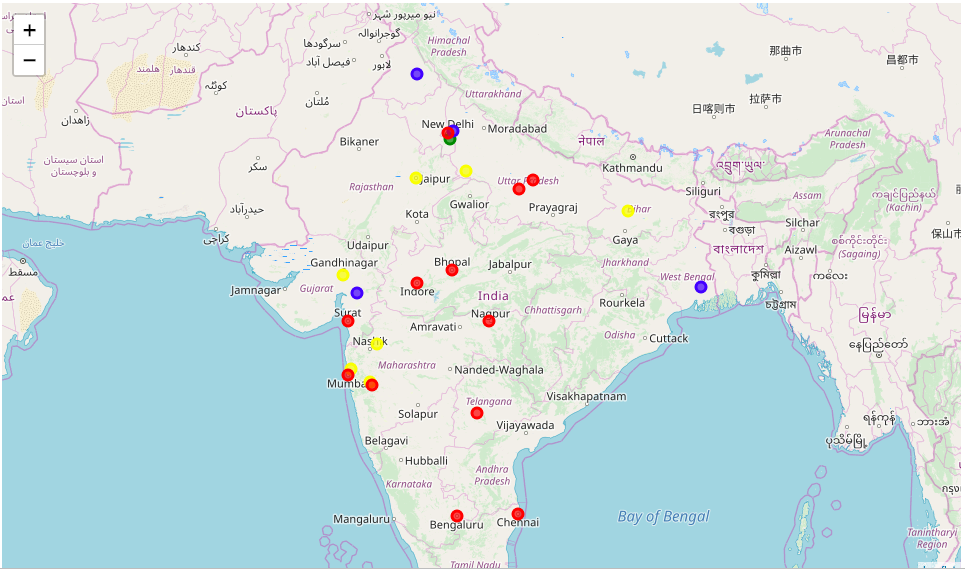
Quality of Living Score = (Education Score+ Recreation Score + Public Service Score)-Crime Score

The units of these individual components are different. So we first standardized all the components and then created the index using above formula.

**Results**

Following are the results of our exercise:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| City | Education | Public Service | Recreation | Population | Crime Percent | Quality\_of\_living\_index | marker\_color |
| Faridabad | 1.89888586 | 1.49777892 | 2.13830487 | -0.7826671 | -1.6749221 | 7.20989177 | green |
| Ghaziabad | 1.53759986 | 1.80603126 | 1.52615338 | -0.7045943 | 1.08127884 | 3.78850567 | blue |
| Ludhiana | -0.6401142 | -0.0046412 | 1.80150426 | -0.7144998 | -2.3957747 | 3.5525235 | blue |
| Vadodara | 1.5087128 | -0.0511725 | 0.7490279 | -0.6972184 | -0.8692634 | 3.07583162 | blue |
| Kolkata | 2.59169105 | -0.5473073 | -0.4943903 | 0.22559484 | -0.4452325 | 1.99522587 | blue |
| Pimpri-Chinchwad | 1.4379605 | -0.4496609 | 1.58315981 | -0.6782867 | 0.78445719 | 1.78700224 | yellow |
| Agra | -0.6401142 | 2.70158278 | 0.39931696 | -0.7255405 | 1.08127884 | 1.3795067 | yellow |
| Thane | -0.6401142 | 1.13057314 | 1.25543205 | -0.6404153 | 0.78445719 | 0.96143379 | yellow |
| Nashik | -0.6401142 | 0.94468692 | 0.61644454 | -0.7587044 | 0.78445719 | 0.13656005 | yellow |
| Ahmedabad | 0.64719725 | -0.7414085 | -0.7115285 | 0.60307925 | -0.8692634 | 0.0635236 | yellow |
| Patna | -0.6401142 | 1.01660062 | -1.0057713 | -0.6927536 | -0.6148448 | -0.0144401 | yellow |
| Jaipur | -0.6401142 | 0.68628811 | -0.4370061 | -0.2394984 | -0.3604263 | -0.030406 | yellow |
| Chennai | -0.6401142 | -0.1976077 | -0.5327976 | 0.29292879 | -0.4028294 | -0.9676901 | red |
| Hyderabad | -0.6401142 | -0.8717476 | -0.8576369 | 0.98708157 | -1.3781005 | -0.9913983 | red |
| Pune | 0.50897246 | -0.9200266 | 0.13096553 | -0.2134418 | 0.78445719 | -1.0645458 | red |
| Bangalore | -0.2145588 | -1.0708978 | -1.020847 | 1.55446883 | -0.8268603 | -1.4794433 | red |
| Surat | -0.6401142 | -0.9593345 | -0.9456059 | 0.23362267 | -0.8692634 | -1.6757912 | red |
| Bhopal | -0.6401142 | -0.1539584 | -0.462385 | -0.6548156 | 0.82686029 | -2.0833179 | red |
| Nagpur | -0.6401142 | -0.2424243 | -0.5231096 | -0.4526567 | 0.78445719 | -2.1901053 | red |
| Indore | -0.6401142 | -0.5762985 | -0.4151441 | -0.5996146 | 0.82686029 | -2.4584171 | red |
| Kanpur | -0.6401142 | -0.6253529 | -0.1431507 | -0.3329539 | 1.08127884 | -2.4898966 | red |
| Delhi | -0.3147478 | -1.0076138 | -1.1884625 | 2.41904563 | 0.02120155 | -2.5320257 | red |
| Lucknow | -0.6401142 | -0.6414559 | -0.1918683 | -0.3157291 | 1.08127884 | -2.5547173 | red |
| Mumbai | -0.6401142 | -0.7226335 | -1.2706054 | 2.88756872 | 0.78445719 | -3.4178103 | red |



**Discussion**

The resulting Quality of Living Indexed has been bucketed into four levels :

- Green : Highest score

- Blue : Less than green

- Yellow : Second Lowest

- Red: Lowest score

As is visible from the resulting Map, only one city 'Faridabad' is green with highest Quality of Living score as 7.2 This is owing to positive and high scores of Faridabad on all criteria of Education, Public Service & Recreation while having a low score for Crime percentage.

Other cities Specially the ones in Red have attained a low score owing to low scores specially for Education & Public Services combined with higher population

**Conclusion**

Based on analysis of our results, we can conclude that for a city to improve its quality of living score it needs to improve upon facilities like Education, Recreation & Public Service in Proportion to its population. Also, focus needs to be given to specially improve upon Education Services & Public Services while trying to reduce Crime Rate

Data provided by Foursquare API for Indian Cities was very limited. This may be the reason for the cities scoring so low specially for factors like Education, Public Services etc. To improve upon the insights from analysis, it may be better to collect data for these criteria from other sources such as government records or United Nations reports. Also advanced index creation techniques like Principal Component Analysis or Structural Equation Methods may be used for better results.