



Doing Business in India's Cities
First attempt at rating the top 36 cities

BIBEK DEBROY & INDICUS ANALYTICS
FOR
CONFEDERATION OF INDIAN INDUSTRY



SECTION 1: INTRODUCTION

This is a study done by Indicus Analytics for the Confederation of Indian Industry (CII). Bibek Debroy (Director, Rajiv Gandhi Institute for Contemporary Studies) and Laveesh Bhandari (Indicus Analytics) are the authors of this report.¹

The 23 million plus cities in 1991 have now increased to 35 (Census 2001). These cities are relatively well spread across the country, though the North does have more than its fair share. Apart from the four major traditional metros of Mumbai, Delhi, Kolkata, and Chennai, all available evidence points towards the rise of Bangalore, Pune, Ahmedabad and Hyderabad as the emerging metros.

But there are many other large cities that have remained off our radar. Asansol, Indore, Vijaywada, Ludhiana, and Madurai are only a few such examples. These cities are regional centers of education, trade, industry, or cater to rich agricultural catchments. Some have only recently taken off, others have had a more sustained growth over the last few decades and have only lately crossed the one million population benchmark.

Many factors affect the ease of doing business. Though some of these factors differ across the type of business, the type of individual, etc., we have only chosen those that we consider to be the key ones. These categories were chosen as they reflect the ***lowest common denominator of important factors across all business activities***. In addition we also had good quality data for 2001 for various variables under each of the categories.

Figure 1: List of categories on which cities have been ranked

Categories
1. Communication
2. Private Finance
3. Road Transport
4. Professional Education
5. Hotel Infrastructure
6. Growth of city economy
7. Overall Rating

We have not included any factors that are based on the perception, but only those that reflect hard facts. We have also not included some issues, that we thought would be important, but for which we found that the quality of available information



was too poor. And as will become clear later, we have minimized biases that sometimes creep up because of authors' subjectivities.

Note that we have not included any items on government's efficiency. That is for three reasons. First, city level data is rarely available for such information. Second, and this is related, we wanted to stick to those factors that are easily quantifiable. Third, we have included a variable "growth" that includes factors that reflect growth in the economy – and government's efficiency should be reflected in that category.

The rating on these categories involved the use of about 25 variables. All the data used is from published and highly credible sources, and all of it represents the situation in the cities as of 2001. We chose the 35 cities that the provisional Census 2001 population figures report to be above one million population urban agglomerations. To those we have added Goa and Chandigarh.²

Figure 2: List of cities included in the ranking

No.	City	No.	City
1	Agra	19	Jaipur
2	Ahmedabad	20	Jamshedpur
3	Allahabad	21	Kanpur
4	Amritsar	22	Kochi
5	Asansol	23	Kolkata
6	Bangalore	24	Lucknow
7	Bhopal	25	Ludhiana
8	Chandigarh	26	Madurai
9	Chennai	27	Meerut
10	Coimbatore	28	Nagpur
11	Delhi	29	Patna
12	Dhanbad	30	Pune
13	Faridabad	31	Rajkot
14	Goa	32	Surat
15	Mumbai	33	Vadodara
16	Hyderabad	34	Varanasi

¹ Peeyush Bajpai's brave information gathering efforts are gratefully acknowledged.

² Serious data availability problems required us to remove Nashik from the ratings.



17	Indore	35	Vijayawada
18	Jabalpur	36	Vishakhapatnam

SECTION 2: DATA

The data sources are given in the data tables that follow. We began the process by collecting around 50 variables for the cities. Many of the variables however duplicated what was being captured by other variables. Hence we reduced the set of 50 variables to about 25, the intention being to include three to five variables for each of the categories mentioned. The variables are distributed across the categories or heads in the following way.

Figure 3: Details of Categories and variables used for Rating Cities

Category	Includes Variables	Explanation
Professional Education	Number of students, number of institutes, and Number of students as a share of population	Professional Education includes MBA and Engineering institutes recognized by AICTE,
Road Transport	Fuel (HSD + Petrol) consumption to Population, Four and Two wheelers to Population, All vehicles to population	Data has many estimated figures to make it comparable across cities, and also to make it relevant for 2001
Communication	Telephones, mobiles, and internet connections as share of population	All are as share of urban agglomeration population
Private Finance	Number of Accounts to Population, Credit to Population, and Credit to Deposits (in Scheduled Commercial Banks)	All information from RBI
Tourism	Number of Good and Medium Quality Hotels, No. of Restaurants, Number of International Tourists (Ministry of Tourism)	Hotels and restaurants are as classified by the Restaurant and Hotel Federation of India, Tourists from unpublished figures from Ministry of tourism in 1998
Growth of Economy	Growth in HSD + Petrol Consumption from 1992 to 1999; Credit growth between 1992 and 2001, growth in MBA students	Absolute growth is in terms of absolute changes



Category	Includes Variables	Explanation
	between 1995 and 2001, growth in population of the district between 1991 and 2001	
Overall Ranking	Professional Education, Transport, Communication, Private Finance, Tourism, Growth of Economy	All given equal weights; all variables normalized and made 'unit free' before aggregation

It should be pointed out that information at the city level is extremely hard to come by. As we move away from the top 4 cities the paucity of information becomes even more severe. And as we limit ourselves to only the latest information – we are left with extremely scarce data resources. However, there is still a lot that can be ascertained with the help of imaginative robust statistical and econometric techniques.

In this exercise, however, we have avoided too much estimation. Estimations have generally been done only to make the data comparable across different cities. This generally has been due to differing coverage of the data across time and geography.

A more liberal use of statistical analysis would have allowed us to compare cities across many more categories, but that is an exercise we leave for the future.

SECTION 3: THE METHODOLOGY

Many factors or variables affect the rating of a City. It is necessary to form a composite or aggregate index that incorporates all these diverse variables into a single or summary measure. The problem in developing a composite index is that related to the process of integrating various variables into a single measure. The identification of weights to be assigned to different variables is one such issue in the creation of a composite index. There are different methods available to form a composite index. One way to do this is to use subjective preferences to identify the magnitude of weights to be assigned to each factor or variable.

Another method, which reduces subjectivity, is to use a type of Factor Analytic Model called Principal Components Analysis (henceforth PCA). PCA is one of the better methods of computing composite indices where the analysis involves relatively low levels of subjectivity on the part of the researcher. This well used econometric tool assigns weights to variables based on relationships among them and therefore minimizes subjectivity. This is the main departure with similar exercises conducted to



rank Cities or even countries. There is no subjective or perceptual element in our exercise. The details of this methodology are presented now.

Principal Component Analysis is a part of the Factor Analytic Model. To derive the composite index for Cities, the following steps were followed:

- Step 1: Identification of appropriate categories
- Step 2: Identification of appropriate variables under each category
- Step 3: Collecting information and normalization
- Step 4: Missing data and imputations
- Step 5: Generating combined indices

Since the main objective was the determination of the performance rankings of the Cities, and as mentioned before, the following criteria were judged as important and for which good quality data was available:

Categories

- Communication
- Finance
- Transport
- Professional Education
- Tourism (Leisure & Business)
- Growth
- Overall Rating

Three to five variables were used for each of these categories. The data were collected from diverse sources and appropriately normalized to account for differences across Cities. The data pertain generally to 2001, when data for that year were not available, appropriate computations were conducted to generate our estimates for the same.

For some variables, data were not available for some Cities. For reporting purposes, the missing data are not included. However, this creates a problem for the composite index creation exercise. This is so because it is not possible to aggregate various variables into a single one when some data are missing. If only those variables are used for which data for all Cities are available, then very few variables would be usable. Consequently, the figures where data were missing were imputed. The imputation was done on the following basis. The variables for which most City data were available were used as explanatory variables in an Ordinary Least Squares (OLS) multivariate regression with the variable with the missing data as the dependent variable. On the basis of the relationship so estimated, the missing data were calculated (or imputed). The results so obtained were also cross-checked with



other similar data. Note that the imputation was done by using variables within a category. This was to avoid the impact of other categories showing up. In that sense, care was taken to keep the categories as mutually exclusive as possible.

Thus, data for all the major Cities and all the variables identified were generated. Next principal components factor analysis methods were used to generate the composite indices. This was done in a two step procedure. First, the indices were generated at each category level. That is, for each of the twelve categories mentioned, there is a ranking across Cities, which is reported.

Once all the indices for all the categories were generated, then the same procedure was used to derive the overall performance index for each City. This procedure used the twelve individual indices mentioned earlier.

To recapitulate briefly, PCA undertakes the following steps:

1. First the analysis involves standardization of data in question. This is done for many reasons. One such reason is that standardization (that involves subtraction of the mean value and division by the standard deviation) eliminates unnecessary weights given to some measures on account of their high unit values.
2. Following the standardization, PCA involves finding that relationship between the variables that explains the maximum possible variation in the total data. This is done by generating various factors.
3. Each factor is nothing but a linear weighted combination of the various variables used. The factors are ranked according to their ability to explain the maximum possible variation among all the variables. The factors are ranked according to their ability to explain the total variance. In all the indices calculated, we used the first factor only. The first factor in all the cases, explained more than 60 per cent of the variation.
4. Such analysis sometimes involves giving negative weights to some of variables. However, no negative weights are observed in any of the indices generated by our exercise.
5. Once the weights for each measure are obtained (also sometimes referred to as factor loading), then the composite index was calculated as the equal weighted average. The rankings were done on this basis.

The indices calculated for each of the categories were then used to calculate the overall index. This was done by calculating the equi-weighted average of all the indices. (see methodology appendix for details on factor analysis)

SECTION 4: THE FINDINGS



We first give the rankings for the 36 cities in accordance with the categories.

Figure 4: City Ranks in 6 Categories

City	Professional Education	Private Finance	Communication	Road Transport	Tourism (Business & Leisure)	Relative Growth
Greater Mumbai	4	1	6	35	1	30
Calcutta	11	13	24	36	4	1
Delhi	6	5	1	11	2	10
Chennai	3	4	7	24	3	27
Bangalore	1	9	8	21	5	7
Hyderabad	5	11	12	28	9	25
Ahmedabad	16	12	13	16	11	14
Pune	2	15	10	29	6	12
Surat	27	26	20	20	29	4
Kanpur	17	27	30	30	21	3
Jaipur	9	14	25	13	7	8
Lucknow	10	22	21	18	16	17
Nagpur	8	21	26	33	13	24
Patna	21	31	23	31	23	23
Indore	15	8	29	7	17	9
Vadodara	33	6	9	5	18	18
Bhopal	14	17	32	17	22	22
Coimbatore	7	3	2	12	20	13
Ludhiana	28	7	5	4	28	11
Kochi	24	10	3	22	10	26
Visakhapatnam	13	18	28	25	19	35
Agra	19	32	31	14	12	28
Varanasi	26	33	22	23	15	34
Madhurai	18	19	14	32	30	19
Meerut	12	28	27	27	34	16
Jabalpur	25	23	33	26	31	36
Jamshedpur	34	29	34	19	32	31
Asansol	36	34	35	10	36	32



Dhanbad	30	36	36	34	35	29
Faridabad	23	24	18	15	33	2
Allahabad	22	35	19	6	25	21
Vijayawada	20	20	15	9	26	33
Amritsar	32	25	16	2	24	20
Rajkot	31	30	11	3	27	6
Chandigarh	29	2	4	1	14	5
Goa	35	16	17	8	8	15

Cities sorted by Population size

The next table gives rankings in terms of the overall index, and the values of the index as a test of the robustness of the rankings.

Figure 5: Overall Ranking

City	Overall Index Ranking	Overall Index
Delhi	1	8.74
Greater Mumbai	2	7.71
Chandigarh	3	7.16
Coimbatore	4	5.22
Bangalore	5	4.75
Chennai	6	4.65
Ludhiana	7	2.90
Pune	8	2.49
Hyderabad	9	1.07
Kochi	10	1.04
Kolkata	11	0.88
Jaipur	12	0.78
Vadodara	13	0.61
Indore	14	0.50
Rajkot	15	0.24
Faridabad	16	-0.17
Ahmedabad	17	-0.23
Goa	18	-0.56
Amritsar	19	-0.67
Surat	20	-1.18
Lucknow	21	-1.36
Kanpur	22	-1.46
Vijayawada	23	-1.64
Allahabad	24	-1.89



Nagpur	25	-1.99
Bhopal	26	-2.36
Madhurai	27	-2.42
Visakhapatnam	28	-2.61
Agra	29	-2.66
Meerut	30	-2.82
Varanasi	31	-3.36
Patna	32	-3.62
Asansol	33	-3.86
Jamshedpur	34	-4.16
Jabalpur	35	-4.45
Dhanbad	36	-5.29

Note: Two consequent cities with same shade imply that they have received similar rating, and difference in their rank is coming only from minor difference in their overall index rating

The next table maps strengths and weaknesses of the Cities. Strengths and weaknesses are interpreted in the following way. A City has an overall ranking. If its ranks in the twelve categories are higher than the overall rank, those categories constitute the State's strengths. Conversely, if its ranks in the twelve categories are lower than the overall rank, those categories constitute the State's weaknesses. The table is thus self-explanatory.

Figure 6: Strengths and Weaknesses – Top 16 cities

City	Weaknesses	Strengths
Delhi	Poor transport for a city of its size, slowing growth, relatively poor in number of professional education institutions, relatively lower growth (though on a large base)	Best communications penetration, good hotel infrastructure
Greater Mumbai	Extremely poor in road transport, lower growth (though on a large base), communications poor for the size of its economy	Best in finance in the whole country, and best hotel infrastructure
Chandigarh	Poor in professional education, hotel infrastructure also requires improvements	Private finance and transport its major strengths
Coimbatore	Poor in road transport, hotel infrastructure, and relatively low growth	Very good in finance and communications for a city of its size
Bangalore	Poor in private finance, transport, and communications	Best in professional education, and hotel infrastructure



Chennai	Poor road transport and relatively lower growth in the nineties	Professional, education, finance and hotel infrastructure are its major strengths
Ludhiana	Extremely poor in professional education and poor hotel infrastructure for a city of its characteristics, relatively low growth	Finance, road transport, and communications are its key strengths
Pune	Poor in finance, communications and road transport, relatively low growth	Private education and good Hotel infrastructure
Hyderabad	Poor road transport, and relatively low growth, moderate in other categories	Good professional education, moderately good hotel infrastructure
Kochi	Poor in professional education and road transport	Relatively low growth, few professional education institutions, and poor in road transport
Kolkata	Moderate to moderately poor in most categories, requires overall improvement in all	Good hotel infrastructure and fast growth in nineties
Jaipur	Poor in communications and private finance	Good hotel infrastructure, moderately good in professional education institutions, fast growth in nineties
Vadodara	Few engineering and management institutions for it's the size of its economic activity, lacks adequate hotel infrastructure and low growth	Good in finance, communications and transport
Indore	Moderately poor in number of engineering institutions, and hotel infrastructure, communications has been improving only lately	High growth levels in the nineties, road transport and finance levels are high
Rajkot	Very poor in communications, finance, and hotel infrastructure	High growth levels, good in communications and road transport
Ahmedabad	Poor education and personal transport for a city of its size and character	Moderately good in most categories