

# India: A ‘Wiki’ Tale of Twenty Nine States

---

**Sai Venkatesh Balasubramanian**

Sree Sai VidhyaMandhir, Mallasandra, Bengaluru-560109, Karnataka, India.

[saivenkateshbalasubramanian@gmail.com](mailto:saivenkateshbalasubramanian@gmail.com)

## Abstract

After understanding the subtle difference between comprehensive and complete coverage of an entity in an encyclopedia, the present article poses the question: For an encyclopedia service such as Wikipedia, what does it take to ‘comprehensively’ describe a particular state of India, covering in reasonable detail all its diversity? In line with this thought, a simple exercise, consisting for searching for the number of occurrences of each district name in the article about its home state, was carried out. This is done for all state mentions in the article about India, as well as each district of the 29 states in the respective articles, sorted out in seven zones: Himalayas, North, Centre, West, South, East and Sisters. The article in particular focuses on significance of the capital in each state, quantified as the Rajdhani Effect, as well as districts not being mentioned anywhere in the article, quantified as the Zero Effect. With these points in consideration, the article explores the Wikipedia articles of each of the 29 states in India, searching for and studying the number of times each district features therein. The result is a different perspective on the states and on the entire country of India, understanding and highlighting the districts that have made a mark significant enough to feature in the ‘comprehensive’ Wiki tale of twenty nine states.

## 1. Introduction

India is the seventh largest nation in the world by area and the second most populous. With its vast geographical diversity comprising snow-clad mountains, rivers, fertile farmlands, serene coasts and barren deserts, India has a lot to offer for the traveler addicted to variety and nature. Over multiple centuries, such diversity helped not only to secure India’s borders from neighboring countries, but also to shape the rich heritage and cultures of India, that have come to mark the country’s key highlight. Unlike countries such as Australia and the United States, most of India’s state borders, at the end of the colonial period, were drawn respecting this diversity and heritage. A person could travel from say, Tamilnadu to Gujarat, and feel as though he has travelled between two different countries such as France and Norway. Such is the diversity in the Indian states, seen by the differences in language, cuisine, costume, customs, landscape, rites and rituals.

With this background in mind, it is interesting to contemplate upon the following question: For an encyclopedia service such as Wikipedia, what does it take to ‘comprehensively’ describe a particular state of India, covering in reasonable detail all its diversity? It is important to consider that no amount of encyclopedia information or even first-hand experience can give absolute and complete information of any place, anywhere in the universe, and thus the use of the term ‘comprehensive’, rather than ‘complete’.

This article is a consequence of exploring the above question. Specifically, it is noted that the first two levels of administrative subdivisions in India are States/Union Territories and Districts respectively. We then try to understand, in the Wikipedia article about any given state of India, how many times would the names each of its districts feature in the said article.

To answer this question, it is important to understand few key points:

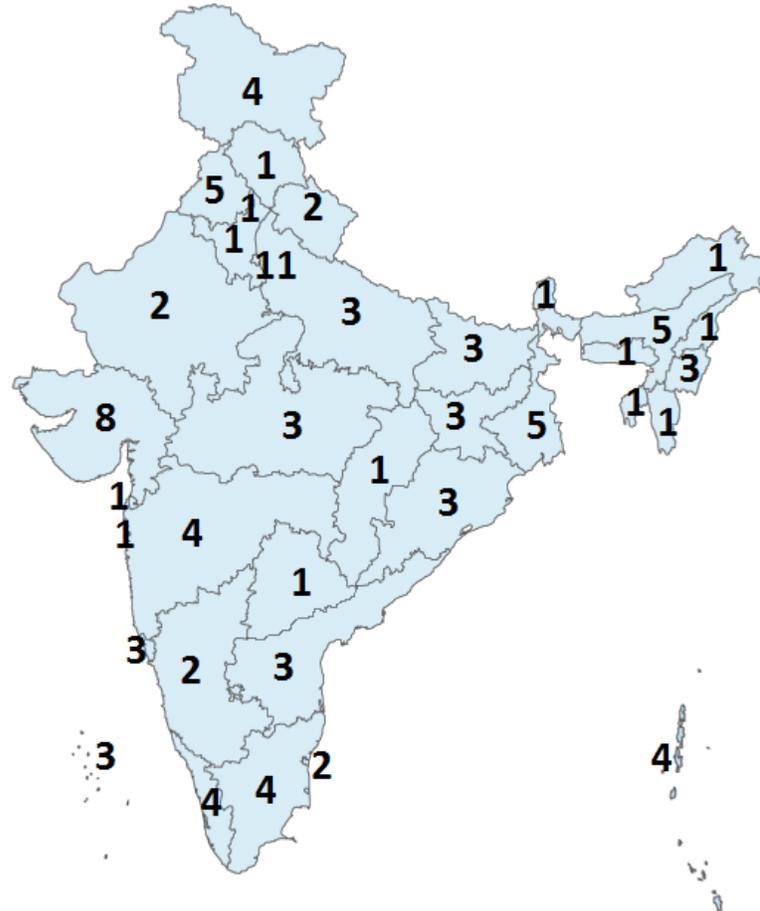
1. In any given state (such as Tamilnadu), the capital of that state (such as Chennai), plays a key role in administration, transport, career, healthcare and other related aspects. The capital is rightfully the heart of a state, and the existence of the whole state revolves around its capital. Thus, an article of a state would certainly feature its capital name multiple times. This behavior shall be termed 'The Rajdhani Effect', with the term Rajdhani translating to capital.
2. Either due to the presence of unusual geographical features, or due to important kingdoms or border issues, certain districts, which are not capitals of a state, find mention multiple times. Examples include Mysore in Karnataka, and Tawang in Arunachal Pradesh. This behavior shall be termed 'Significance Effect'.
3. In any state, it is common to find certain districts not getting mentioned at all. In most cases, this is because, most of the history and geography of the district turns out to be rather average, with nothing unusual to report. This is called the 'Zero Effect'. At this point, it should be noted that the present article only focuses on mentions of district names, focusing on the 'collective' influence of a district on a state. For example, a discussion on architecture and tourism in Tamilnadu may find mentions of Mahabalipuram, without mention of its district Kanchipuram. This is because, the discussion focuses on a particular area of Kanchipuram, and not the whole district. On the other hand, a discussion on industry and economy of Tamilnadu will mention Kanchipuram in the context of silk and weaving industries, and this time, it is the whole district that is being mentioned. This study takes into account the latter, and not the former.

With these points in consideration, the article explores the Wikipedia articles of each of the 29 states in India, searching for and studying the number of times each district features therein. The result is a different perspective on the states and on the entire country of India, understanding and highlighting the districts that have made a mark significant enough to feature in the 'comprehensive' Wiki tale of twenty nine states.

## 2. India - States

The study starts with examining the Wikipedia article of India in the English language, available at <https://en.wikipedia.org/wiki/India>.

In this article, the number of occurrences of each of the 29 states and 7 union territories are recorded, and plotted on the map, as shown below.



One can infer the following from the results:

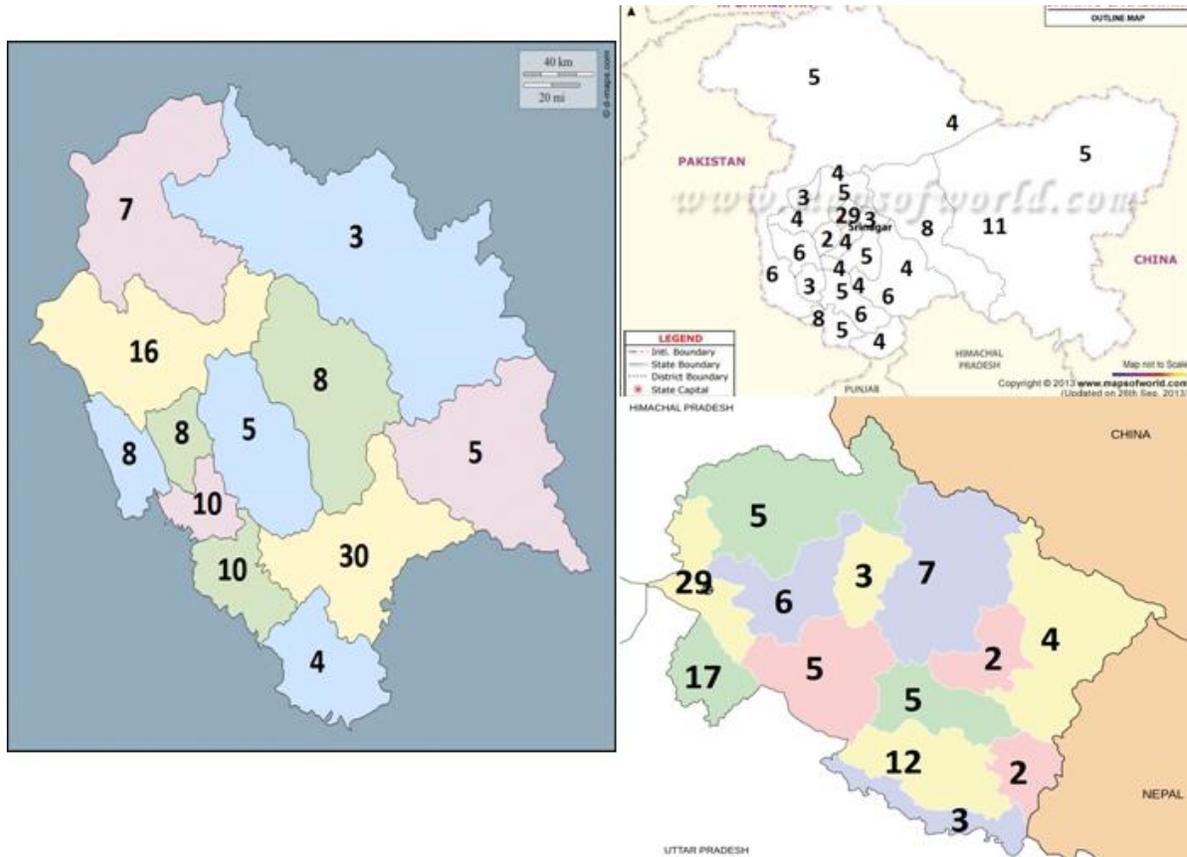
1. The administrative subdivision most frequently mentioned is the National Capital Territory of Delhi at 11, which falls in line with the Rajdhani Effect mentioned earlier.
2. The next highest 'mention frequencies' are Gujarat at 8, and Punjab, West Bengal and Assam, at 5 each. Mentions of these states are found in sections of geography, culture, languages etc.
3. Considering that the wiki page has a listing of states and union territories, entities which are marked with '1' on the map indicate that they occur only in the listing and nowhere else in the article. These include the states of Himachal Pradesh, Haryana, Chhattisgarh, Telangana, Sikkim, Arunachal Pradesh, Nagaland, Mizoram, Tripura and Meghalaya, and the Union Territories of Chandigarh, Daman & Diu, and Dadra & Nagar Haveli. All of these correspond to the Zero Effect described earlier.

### 3. Key Inferences

In this section, analysis of the 29 states is carried out, searching for number of occurrences of each district within the article of its corresponding state. The results and key inferences are presented, organized into seven zones.

#### A. The Himalayas

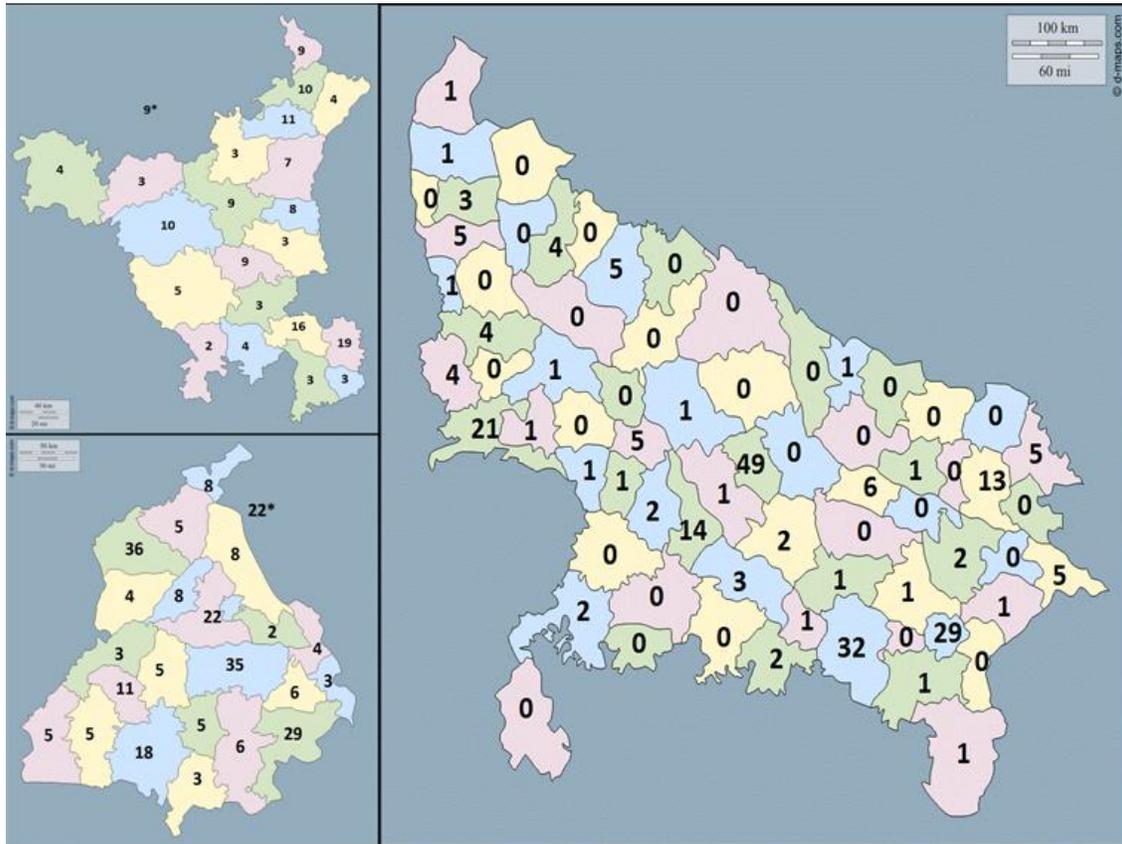
This zone comprises of the states of Jammu & Kashmir (J&K), Himachal Pradesh and Uttarakhand.



- In scenic J&K, apart from the capital Srinagar, Kargil, presumably due to its strategic importance and Leh, due to its military importance, as well as otherworldly experience receives highest mentions.
- In mountainous Himachal Pradesh, second to Shimla is Kangra in significance, owing to the cultural importance, as well as presence of Tibetan Settlement in Dharamshala.
- In ‘Devbhoomi’ Uttarakhand, the capital Dehradun is most significant at 29, followed by Haridwar at 17, owing to its spiritual significance, and Nainital at 12, owing to its tourist appeal.

## B. The North

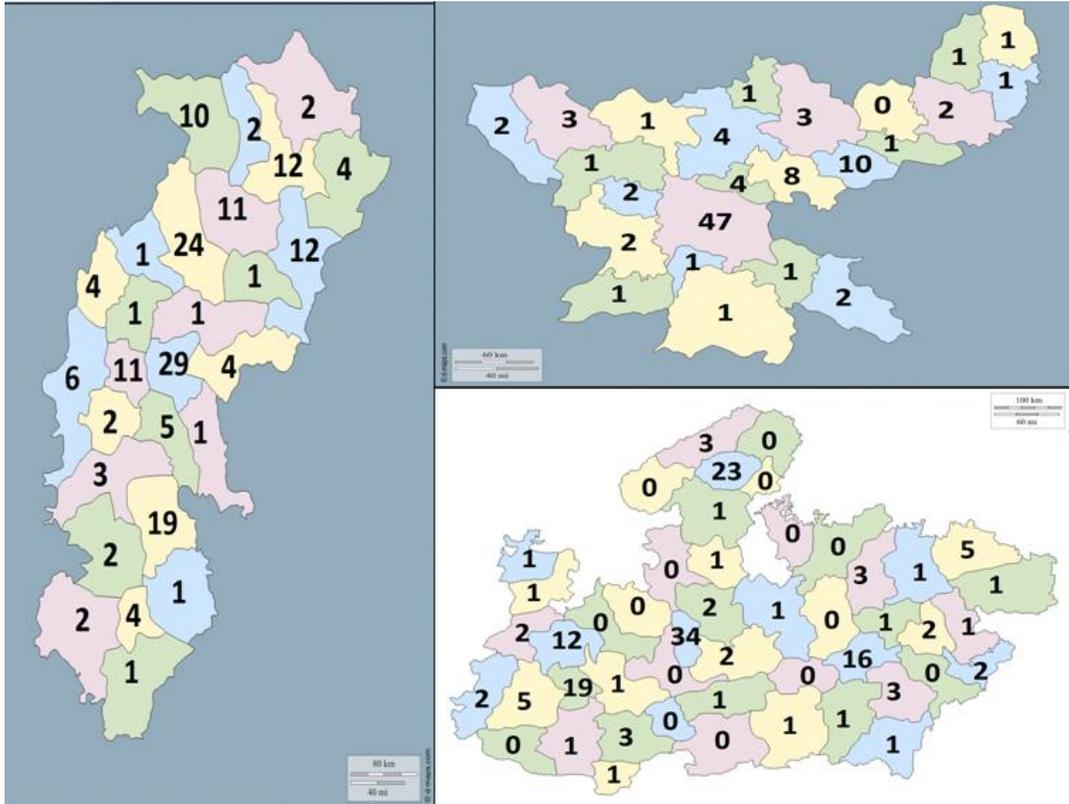
The states in this zone are Punjab, Haryana and Uttar Pradesh.



- Haryana, with its capital outside the state at Chandigarh, has a rare property of highest mentioned districts being the non-capital Gurgaon and Faridabad, both due to industrial significance and proximity to Delhi, and Kurukshetra, due to religious and historical significance.
- The Sikh land of Punjab, which shares its capital Chandigarh with Haryana, has maximum mentions corresponding to the districts of Amritsar, owing to the religious importance, industrially significant Ludhiana and culturally important Patiala.
- In India's most populous state of Uttar Pradesh, the capital Lucknow features most at 49 times. Following it are culturally and spiritually important Allahabad and Varanasi, followed by the historically significant tourist capital of India, Agra.

### C. The Centre

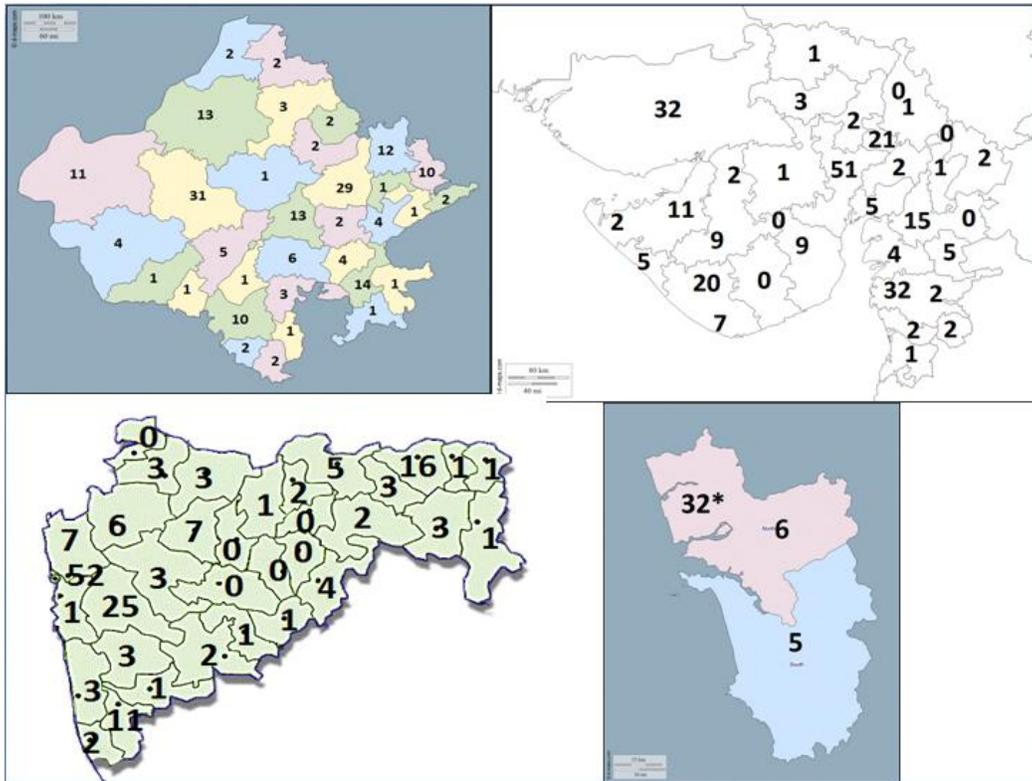
The states in this zone are Madhya Pradesh, Chhattisgarh and Jharkhand.



- In the heart of India at Madhya Pradesh, Bhopal, the capital leads at 34, followed by religiously significant Gwalior and the districts of Indore and Jabalpur.
- In Chhattisgarh, the only districts other than capital Raipur significantly mentioned are Bastar and Bilaspur.
- In nature abound Jharkhand, the Ranchi leads at 49, and at a distant second are the mining stronghold of Dhanbad and industrial Bokaro.

## D. The West

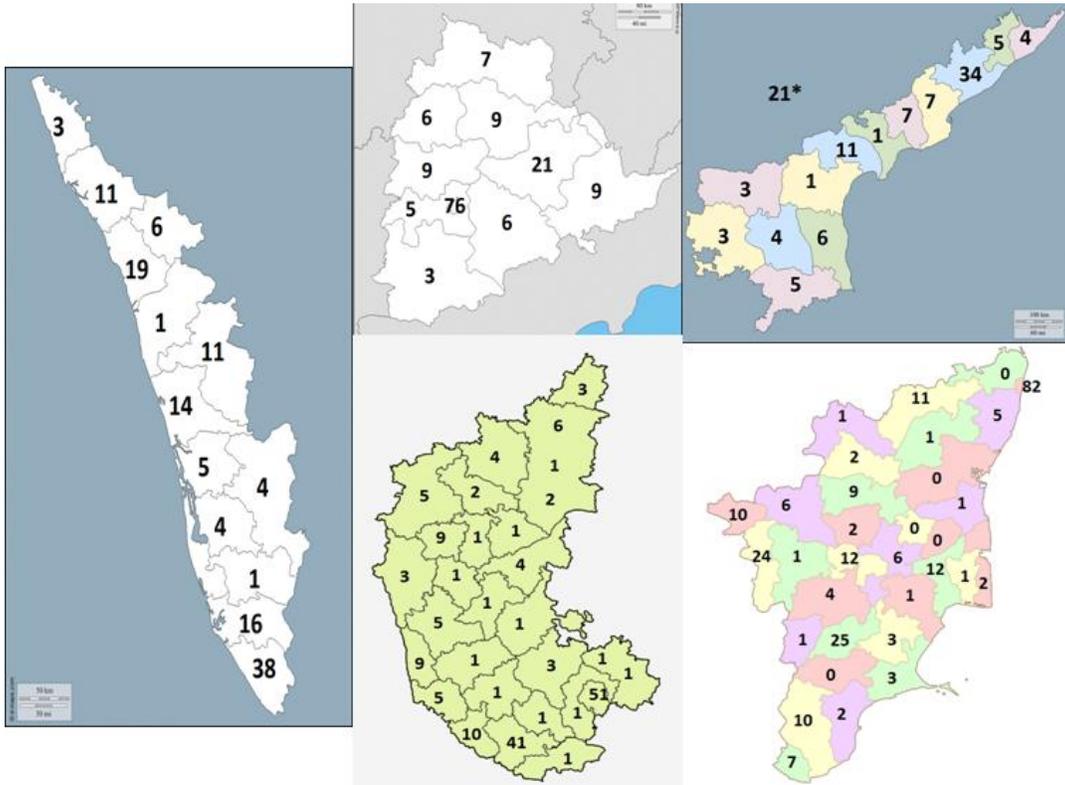
The four states in this zone are Rajasthan, Gujarat, Maharashtra and Goa.



- In arid Rajasthan, princely Jodhpur leads the tally with 31, with the pink city capital Jaipur closely behind at 29, and all other districts a distant behind.
- In Gujarat, one sees a high total for the largest city Ahmedabad, followed by 32 for the unique landscape of Kutch and port city of Surat. Next comes the capital Gandhinagar.
- In Maharashtra, the financial capital of India, Mumbai, leads with 52 mentions, followed by the culturally and industrially significant Pune at 25, and ‘centre of India’ Nagpur at 16.
- In Goa, the capital Panaji occurs 32 times, and north Goa features slightly more than the south.

## E. The South

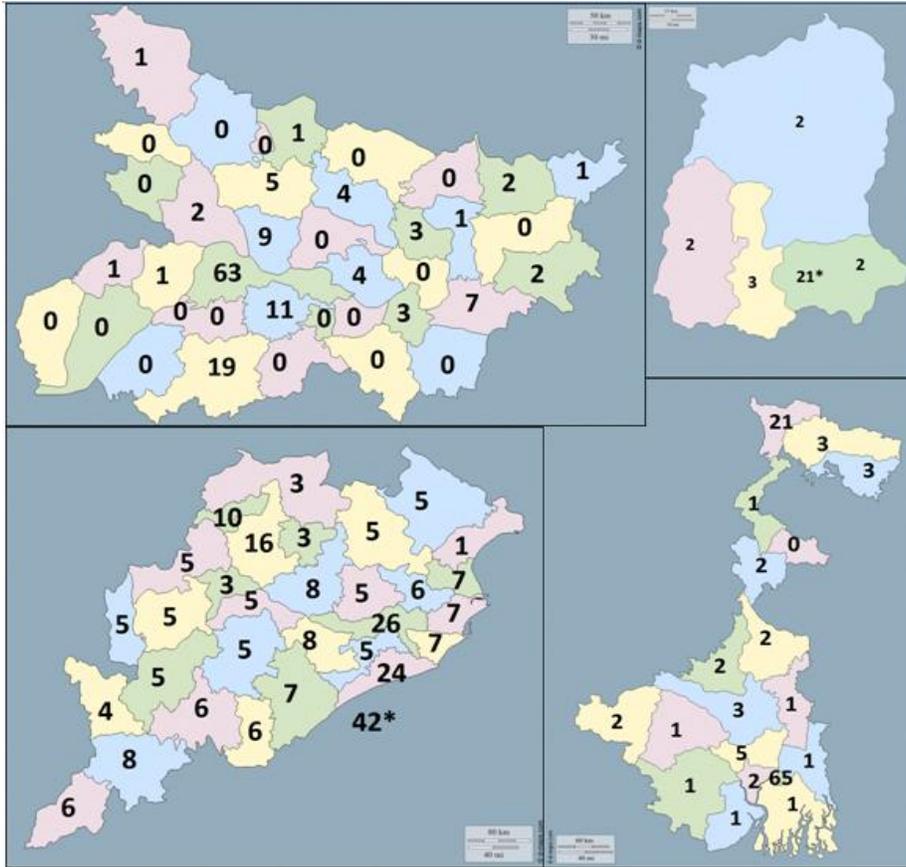
The states here are Karnataka, Kerala, Tamilnadu, Andhra Pradesh and Telangana.



- In ‘God’s Own Country’ Kerala, second to the capital Thiruvananthapuram is the culturally important Kozhikode, followed by Ernakulam and Kollam.
- In India’s newest state Telangana, a distant second to capital Hyderabad (76) is Karimnagar, with 21 occurrences.
- In Andhra Pradesh, the important port city of Visakhapatnam features highest at 34, followed by the capital Hyderabad, and proposed capital district of Guntur.
- In Karnataka, the ‘IT Capital’ of India, Bengaluru reigns at 51, followed by royal city of Mysuru. The ‘Scotland of India’ Kodagu is a distant third.
- In Temple State Tamilnadu, the capital Chennai features 82 times, distantly followed by industrial Coimbatore and culturally significant Madurai.

## F. The East

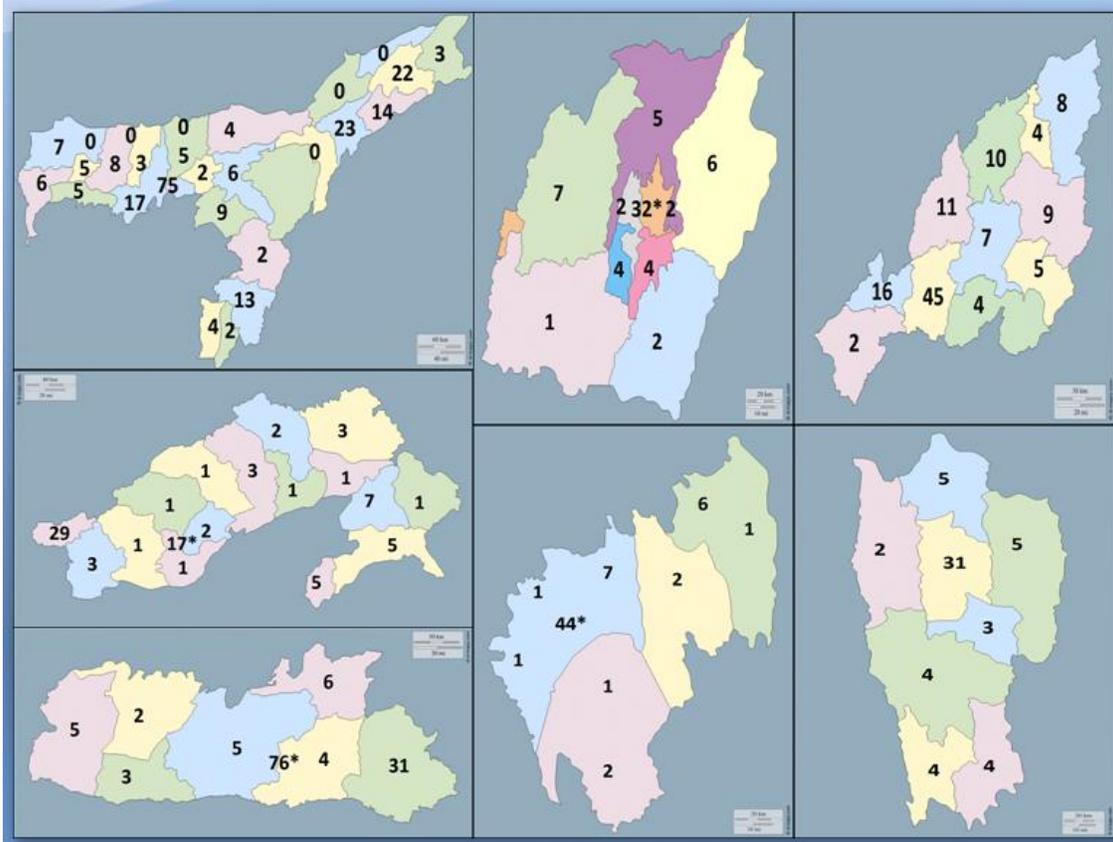
This zone comprises of the quartet of Odisha, West Bengal, Bihar and Sikkim.



- In Bihar, the state with highest population density, Patna leads at 63 mentions, followed distantly by Buddhist strongholds Gaya and Nalanda.
- In coastal Odisha, the capital Bhubaneswar leads the tally at 42, with the spiritually and culturally significant Cuttack and Puri following.
- In mountain jewel Sikkim, the capital Gangtok records 21 mentions, while South Sikkim slightly edges past other districts.
- In West Bengal, metropolitan Kolkata tops at 65, followed by picturesque Darjeeling.

## G. The Sisters

In this zone are the seven sister states: Assam, Meghalaya, Arunachal Pradesh, Nagaland, Manipur and Mizoram.



- In Assam, the capital Guwahati leads at 75 and is distantly followed by Jorhat and Dibrugarh.
- In frontier Arunachal Pradesh, strategically significant Buddhist town Tawang leads at 29 and outperforms the capital Itanagar, which is at 17.
- In Meghalaya, the capital Shillong features 76 times, followed by Jaintia Hills at 31.
- In Manipur, the capital Imphal features 32 times, standing out prominently among all districts.
- In Tripura, similar behavior is seen with the capital Agartala at 44, and all other districts far behind.
- In tribal Nagaland, the capital Kohima leads trends at 45, followed by gateway district Dimapur.
- In hilly Mizoram, the capital Aizawl stands out prominently, featuring 31 times.

## H. Consolidation

In summary of the above illustrated results, the following is a tabulation of all 29 states, illustrating for each state, two key statistical values:

1. The 'Rajdhani Effect', quantified as percentage by dividing the number of mentions of capital district by total number of mentions of all districts.
2. The 'Zero Effect' in percent, computed by dividing number of districts with zero mentions by the total number of districts in a state. In states whose article pages give a listing of constituent districts, those districts featuring only once are taken as 'Zero' districts.

| Name        | Capital | Rajdhani | Zero E | Name              | Capital | Rajdhani | Zero E |
|-------------|---------|----------|--------|-------------------|---------|----------|--------|
| Goa         | 32      | 74.42    | 0.00   | Karnataka         | 51      | 29.14    | 43.33  |
| Sikkim      | 21      | 70.00    | 0.00   | Uttarakhand       | 29      | 29.00    | 0.00   |
| Tripura     | 44      | 67.69    | 50.00  | Kerala            | 38      | 28.57    | 0.00   |
| Meghalaya   | 76      | 57.58    | 0.00   | Himachal Pradesh  | 30      | 26.32    | 0.00   |
| West Bengal | 65      | 55.56    | 42.11  | Madhya Pradesh    | 34      | 22.08    | 29.41  |
| Mizoram     | 31      | 53.45    | 0.00   | Uttar Pradesh     | 49      | 20.85    | 38.67  |
| Manipur     | 34      | 52.31    | 11.11  | Arunachal Pradesh | 17      | 20.48    | 43.75  |
| Telangana   | 76      | 50.33    | 0.00   | Jammu & Kashmir   | 29      | 19.21    | 0.00   |
| Jharkhand   | 47      | 47.00    | 50.00  | Andhra Pradesh    | 21      | 18.75    | 15.38  |
| Bihar       | 63      | 45.00    | 50.00  | Chhattisgarh      | 29      | 16.57    | 25.93  |
| Nagaland    | 45      | 37.19    | 0.00   | Odisha            | 42      | 16.28    | 3.33   |
| Tamilnadu   | 82      | 33.61    | 15.63  | Rajasthan         | 29      | 15.51    | 27.27  |
| Assam       | 75      | 31.91    | 22.22  | Punjab            | 22      | 9.52     | 0.00   |
| Maharashtra | 52      | 30.59    | 16.67  | Gujarat           | 21      | 8.40     | 15.15  |
| Haryana     | 9       | 5.84     | 0.00   | India             | 11      | 10.78    | 0      |

From the table, it can be inferred that a significant proportion of the Rajdhani Effect occurs in the Sister states such as Tripura, Meghalaya, Mizoram and Manipur, as well as small states of Goa and Sikkim. In effect, this can be correlated to unequally dense distribution of population in the capitals of these states, with other regions being sparsely populated, without much activity.

It is interesting to note that two of the lowest Rajdhani Effect values are seen for Punjab and Haryana, both of which share a capital, Chandigarh, that is situated outside both states. The average Rajdhani Effect value is 34%, and the state closest to this value is Tamilnadu. Among states containing metropolitan cities, Kolkata emerges on top, followed by Hyderabad and Chennai.

Among Zero Effect values, the highest proportions are seen at 50% for Tripura, Bihar and Jharkhand, which suggest that in order to have comprehensive information about these states, merely half the districts need to be mentioned, and are thus noteworthy. The rest of the districts might hopefully await their day in the spotlight, in the near future. Quite a lot of states have 0% Zero Effect, suggesting that all district names have been mentioned atleast once, in different parts of the article. The average Zero value is 17.2% and the state closest to this is Maharashtra.

## 4. Conclusion

After understanding the subtle difference between comprehensive and complete coverage of an entity in an encyclopedia, the present article poses the question: For an encyclopedia service such as Wikipedia, what does it take to ‘comprehensively’ describe a particular state of India, covering in reasonable detail all its diversity? In line with this thought, a simple exercise, consisting for searching for the number of occurrences of each district name in the article about its home state, was carried out. This is done for all state mentions in the article about India, as well as each district of the 29 states in the respective articles, sorted out in seven zones: Himalayas, North, Centre, West, South, East and Sisters. The article in particular focuses on significance of the capital in each state, quantified as the Rajdhani Effect, as well as districts not being mentioned anywhere in the article, quantified as the Zero Effect. With these points in consideration, the article explores the Wikipedia articles of each of the 29 states in India, searching for and studying the number of times each district features therein. The result is a different perspective on the states and on the entire country of India, understanding and highlighting the districts that have made a mark significant enough to feature in the ‘comprehensive’ Wiki tale of twenty nine states.

## References

- [1] Khilnani, Sunil. *The idea of India*. Penguin Books India, 1999.
- [2] Wiener, Myron. *State politics in India*. Princeton University Press, 2015.
- [3] Bhattacharjee, Pranabjyoti, and Gajanan Narayan Shastri. *Population in India: A Study of Inter-State Variations*. Vol. 3. International Book Distributors, 1976.
- [4] Dhawan, Nisha, Ira J. Roseman, R. K. Naidu, Komilla Thapa, and S. Ilsa Rettek. "Self-concepts across two cultures India and the United States." *Journal of Cross-Cultural Psychology* 26, no. 6 (1995): 606-621.