Temporal Analysis and Prediction of Indian Election results

Social Computing Course Term Project (Group 8)

G.Sai Bharath Chandra (14CS10020) R.Sri Charan Reddy (14CS10037) Aakash Verma (14CS30001) B.Sanjith (14CS10010)

Problem Statement:

The objective of our project is to analyze the factors which influence election results and predict the future election outcomes based on these features. We hope to predict the winner of a certain election accurately.

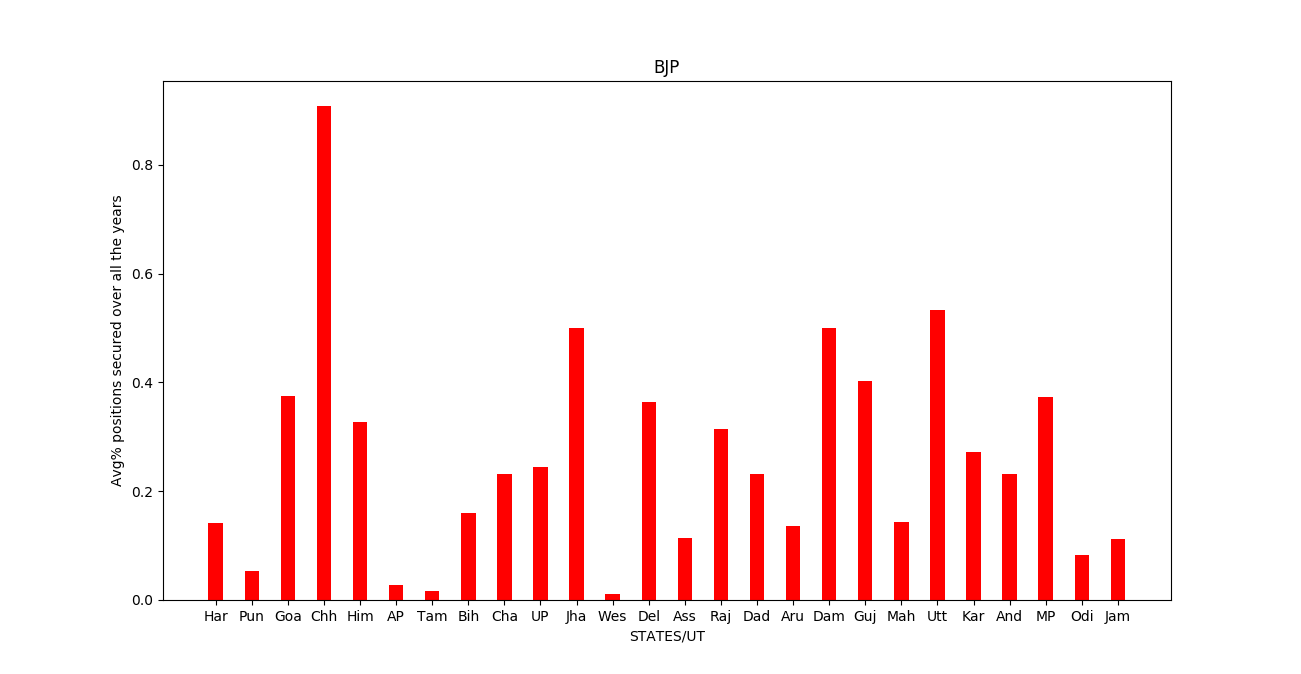
Dataset:

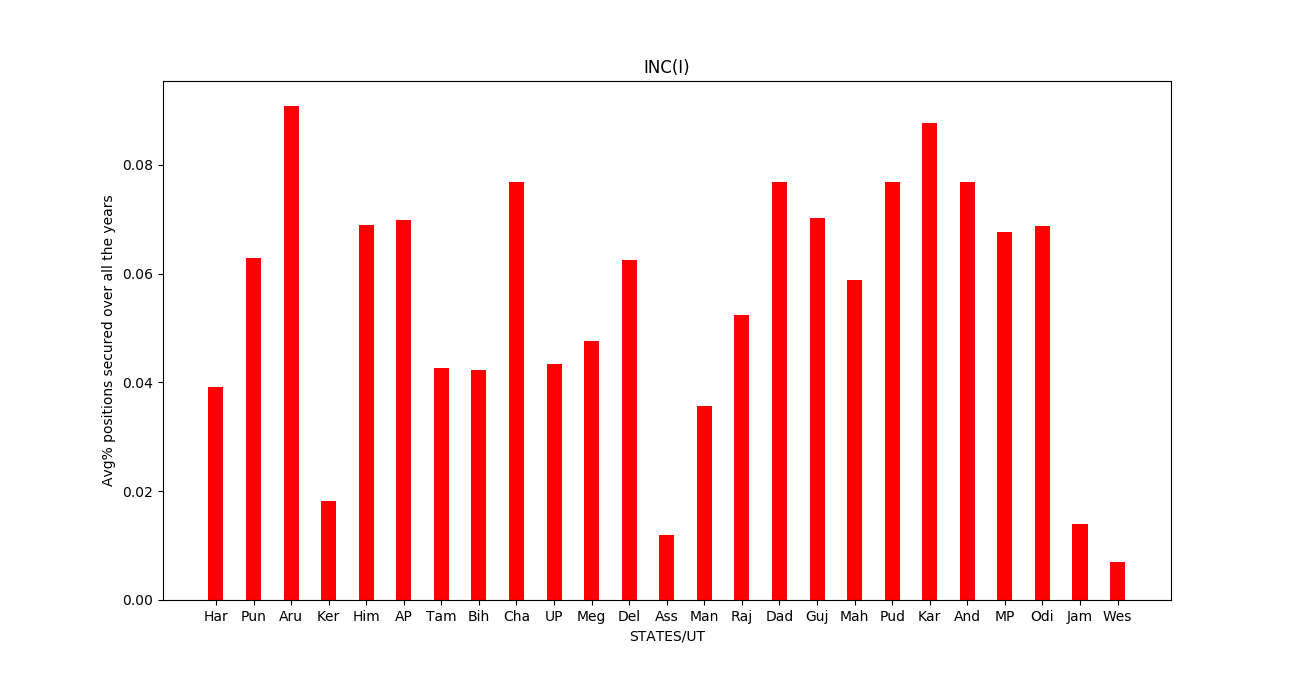
We used a publicly available elections dataset which has the Central Elections data, Assembly Elections results for each state from 1962 to 2014. In both central and assembly election data we have the following attributes.

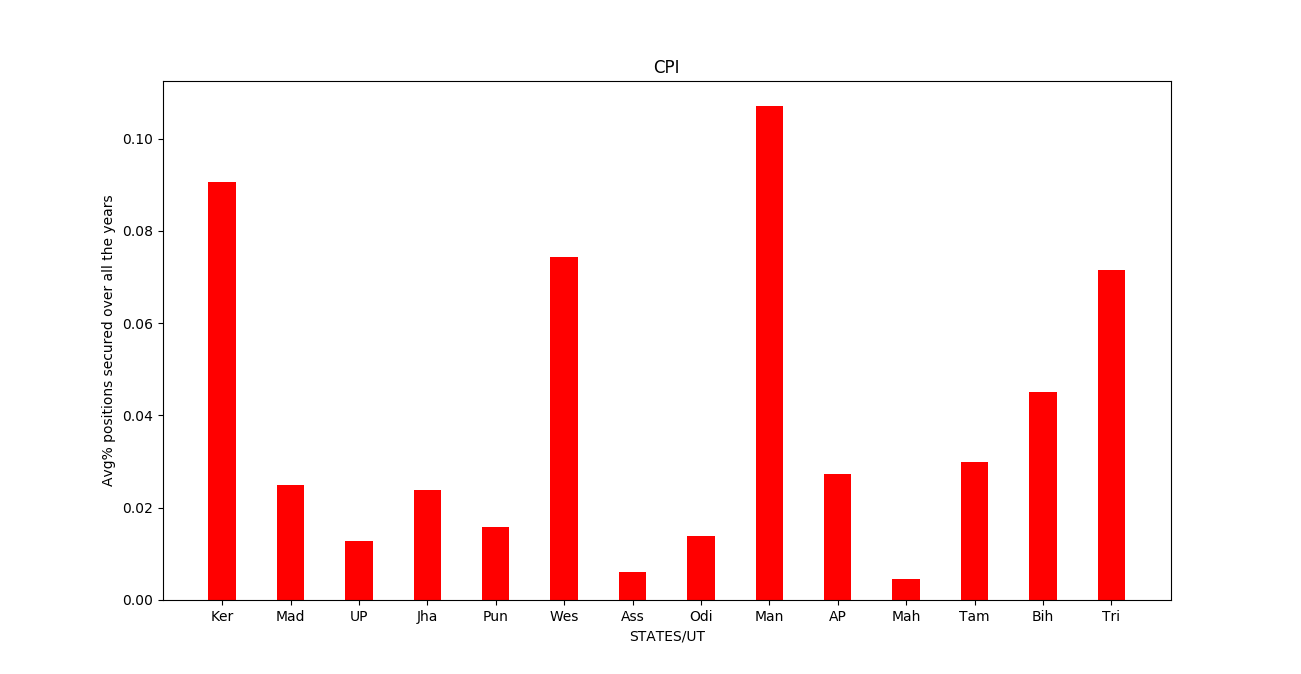
|  |  |
| --- | --- |
| Year | Year of election |
| Votes | Votes obtained by that candidate |
| Vote\_Share\_Percentage | Vote share of a candidate in a given constituency |
| Valid\_Votes | Total valid votes in that constituency |
| Turnout\_Percentage | Turnout in that constituency |
| Sub\_Region | Sub region of that constituency |
| State\_Name | Name of the state |
| Sex | Candidate's gender |
| Position | Candidate's position in election |
| Party | Candidate's party |
| N\_Cand | Number of candidates contested in that constituency |
| Margin\_Percentage | Percentage margin of a candidate in compared to the next position candidate |
| Margin | Difference in votes between a candidate and the next ordered candidate |
| Electors | Total electors in that constituency |
| District\_Name | Name of the district of that constituency |
| Deposit\_Lost | To denote if a candidate lost his/her deposit |
| DelimID | Delimitation ID is a unique id assigned to each delimitation |
| Constituency\_Type | Assembly constituency type |
| Constituency\_No | Assembly constituency number |
| Constituency\_Name | Assembly constituency name |
| Candidate\_Type | Candidate caste (Gen/SC/ST) as specified by the ECI |
| Candidate | Candidate name |
| Assembly\_No | State assembly number |
| Age | Age of the candidate, wherever available |

Observations:

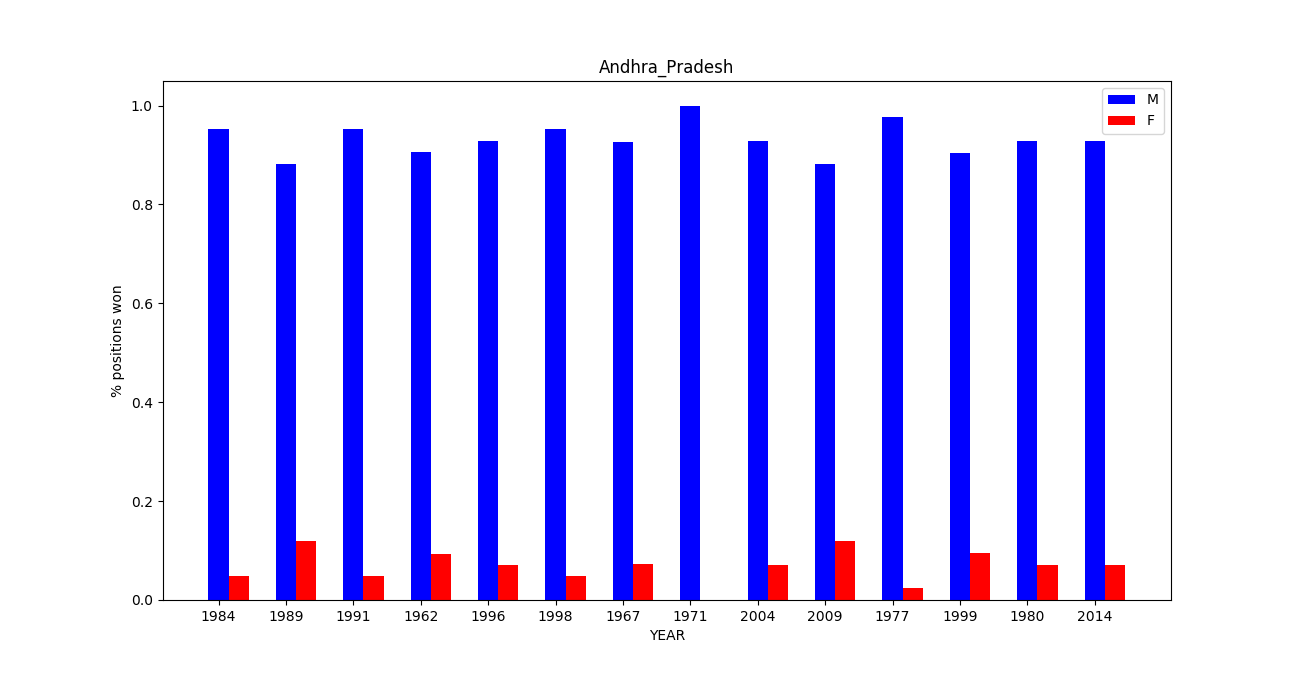
We plotted average number of positions secured by a particular party in all previous central elections for each state & we observed that some parties have an advantage & are more likely to win in certain states/constituencies. We found similar patterns for assembly elections for each constituency within a state.

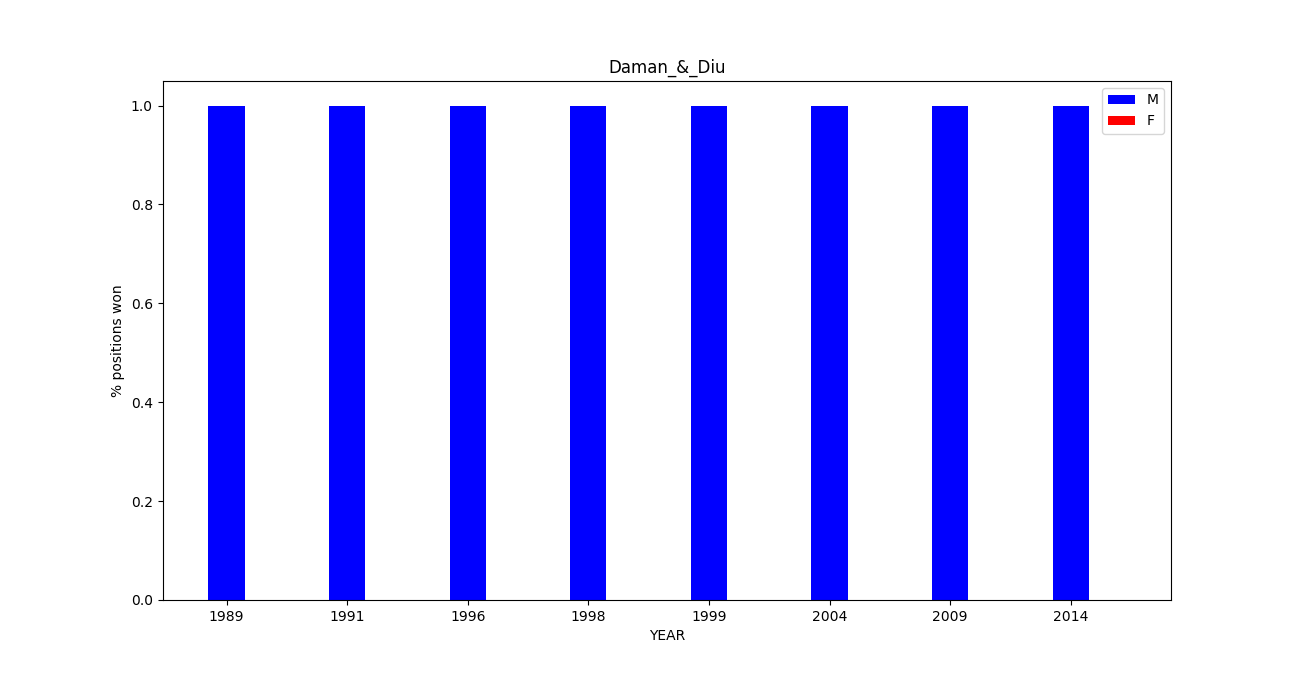


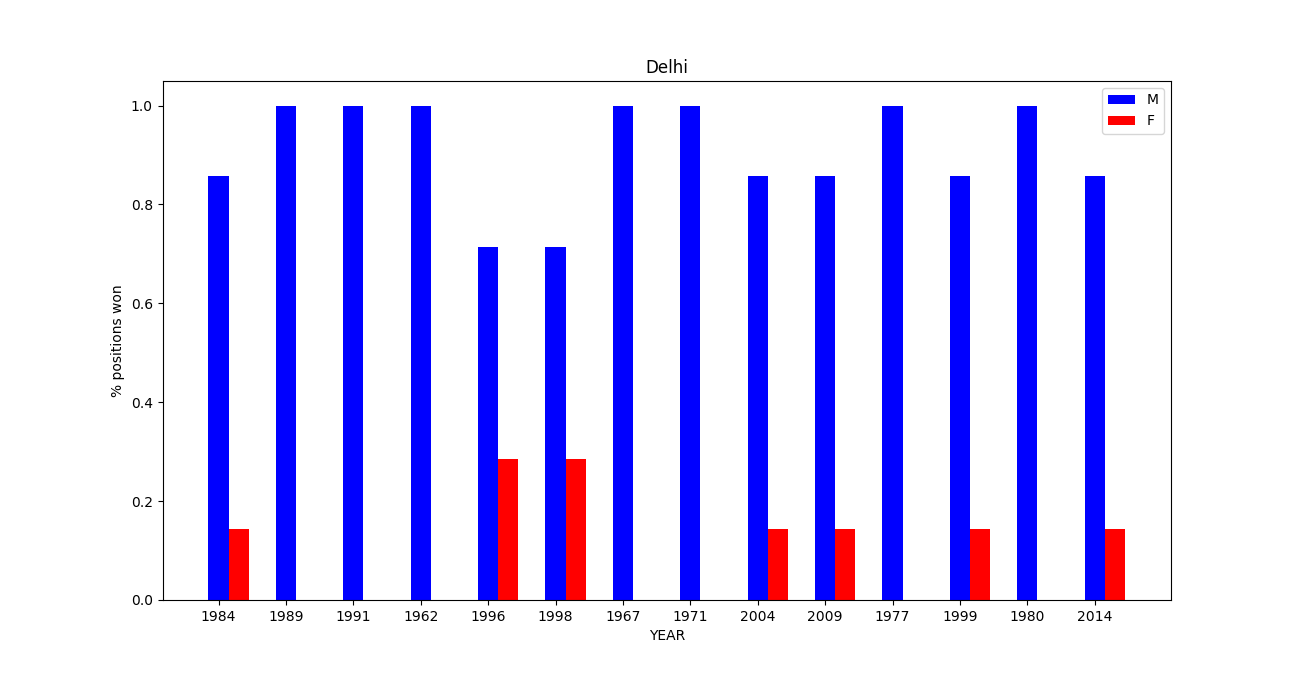




We plotted fraction of positions won by male(female) candidates out of total male(female) candidates for each central election for each state & we observed a certain preference on account of gender of the political candidates & also found similar patterns for assembly elections.







Training Data:

We took all election data before the year '2000' as our training data and the rest as test dataset. Each training example in our training set consists of all the attributes mentioned in the table in the dataset section in which the output/label is the vote share percentage.

Features Extraction:

Factors that influence the outcome of a candidate in an election:

(POPULARITY OF THE PERSON IN THIS CONSTITUENCY, STATE)

(POPULARITY OF THE PARTY IN THIS CONSTITUENCY, STATE)

(POPULARITY OF THE PERSON WITH THIS PARTY, ANY PARTY)

All Features (F1, F2, .... F12, GF1, GF2, ….GF12) have values from [0,1] and are defined as follows

POPULARITY OF THE PARTY IN THIS CONSTITUENCY/STATE:

1) Popularity of this party in this constituency

F1: Fraction of times this party people won from same constituency

F2: Average position this party people won from same constituency normalized by total positions

F3: Average votes this party people secured in past elections in same constituency normalized by total votes

F4: Average Margin loss of previous elections in same constituency

2) Popularity of this party in this state

GF1: Fraction of times this party people won from any constituency

GF2: Average position this party people won from any constituency normalized by total positions

GF3: Average votes he secured in his past elections in any constituency normalized by total votes in each election

GF4: Average Margin loss of previous elections in any constituency

POPULARITY OF THE PERSON WITH THIS PARTY/ANY PARTY:

1) Popularity of this person in this party

F5: Fraction of times he won from this party

F6: Average position he won from same party normalized by total positions

F7: Average votes he secured in his past elections in same party normalized by total votes in each election

F8: Average Margin loss of previous elections in same party

2) Popularity of this person in any party

GF5: Fraction of times he won from any party

GF6: Average position he won from any party normalized by total positions

GF7: Average votes he secured in his past elections in any party normalized by total votes in each election

GF8: Average Margin loss of previous elections in any party

POPULARITY OF THE PERSON IN THIS CONSTITUENCY/STATE:

1) Popularity of this person in this constituency

F9: Fraction of times he won from same constituency

F10: Average position he won from same constituency normalized by total positions

F11: Average votes he secured in past elections in same constituency normalized by total votes in each election

F12: Average Margin loss of previous elections in same constituency

2) Popularity of this person in state

GF9: Fraction of times he won from any constituency

GF10: Average position he won from any constituency normalized by total positions

GF11: Average votes he secured in past elections in any constituency normalized by total votes in each election

GF12: Average Margin loss of previous elections in any constituency

Additionally, we used the gender of a candidate 1-Male/0-Female.

Results:

We train a regression model that takes a feature vector predicts the vote share percentage of that candidate. In a particular election, in a particular constituency, we predict vote share percentages of all candidates and choose the one with highest vote share as the winner of that constituency.

Since predicting a winner of an election heavily depends on the characteristics of a person which we don't have in our dataset, we couldn't achieve an accuracy more than 50% in predicting a winner, however when we considered Top-P predicted candidates as winners we observed a better accuracy.

We've tried with several regression models and compared the accuracy below for P=1,2,3 i.e taking winner to be in Top-P predicted candidates. Our accuracy is much better at linear regression model & also didn't change much at all when gender is omitted from the features.

|  |  |  |
| --- | --- | --- |
| State | Assembly Elections (Accuracy-Linear Regression)  Three-P Two-P One-P | Central Elections (Accuracy-Linear Regression)  Three-P Two-P One-P |
| Andhra Pradesh | 88.48 80.95 37.2 | 82.53 79.36 35.71 |
| Arunachal Pradesh | 94.91 82.48 53.67 | 83.33 66.66 33.33 |
| Assam | 83.73 71.82 45.63 | 85.71 66.66 42.85 |
| Bihar | 72.18 62.65 41.86 | 95.00 75.00 35.00 |
| Delhi | 90.35 69.28 41.78 | 100 100 76.19 |
| Goa | 73.75 53.75 33.75 | 83.33 66.66 50.00 |
| Gujarat | 96.80 93.79 40.22 | 100 100 47.43 |
| Haryana | 85.27 57.50 26.38 | 100 80.00 30.00 |
| Himachal Pradesh | 91.66 81.86 41.66 | 91.66 83.33 41.66 |
| Jammu & Kashmir | 59.38 48.65 29.88 | 77.77 50.00 27.77 |
| Karnataka | 76.63 62.35 38.24 | 94.04 91.66 26.19 |
| Kerala | 96.07 94.46 53.39 | 95.00 91.66 65.00 |
| Madhya Pradesh | 94.92 90.72 31.01 | 98.85 98.85 35.63 |
| Maharashtra | 77.85 64.68 43.93 | 93.05 75.69 37.50 |
| Manipur | 82.66 68.00 47.33 | 83.33 83.33 50.00 |
| Meghalaya | 77.77 69.40 38.80 | 100 100 83.33 |
| Mizoram | 98.33 86.66 43.30 | 100 100 66.66 |
| Nagaland | 84.40 65.00 35.00 | 100 66.66 66.66 |
| Odisha | 90.30 82.65 45.00 | 100 95.23 58.73 |
| Puducherry | 84.16 75.00 45.83 | 66.66 66.66 66.66 |
| Punjab | 92.29 83.29 44.96 | 89.74 84.61 51.28 |
| Rajasthan | 90.61 86.43 31.49 | 98.66 97.33 33.33 |
| Sikkim | 66.30 48.91 39.13 | 100 100 100 |
| Tamilnadu | 63.09 46.24 24.89 | 99.14 95.72 67.52 |
| Tripura | 100 99.44 63.33 | 100 66.66 16.66 |
| Uttar Pradesh | 73.40 51.93 23.65 | 81.25 59.16 32.50 |
| West Bengal | 98.04 81.00 42.33 | 96.82 89.68 38.09 |