# Maharashtra rainfall analysis

**Developers**: M R Abhishek
            K Vagdevi
**Guide**: Dr. Mainak Thakur

# Details of data

- **Source of data**: https://www.indiawaterportal.org/met_data/

- **State**: Maharashtra

- **Total number of districts**: 29

- **Data type**: Precipitation

- **Data Range**: Monthly mean precipitation for all the years from 1901 to 2003

- We have 1224 (102 *12) data points for each district in Maharashtra. In total, we have 35,496 data points.

# Format of Maharashtra Data

| | District | Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 26826 | Ahmadnagar | 1901 | 2.51 | 11.489 | 11.325 | 33.931 | 30.401 | 262.971 | 162.397 | 120.152 | 58.708 | 62.633 | 0.596 | 0.0 |
| 26827 | Ahmadnagar | 1902 | 0.603 | 0.0 | 0.003 | 1.34 | 6.338 | 90.293 | 223.483 | 101.852 | 297.212 | 61.453 | 41.992 | 26.715 |
| 26828 | Ahmadnagar | 1903 | 0.227 | 0.0 | 0.0 | 2.602 | 69.076 | 182.652 | 500.983 | 156.236 | 164.433 | 88.419 | 2.022 | 3.999 |
| 26829 | Ahmadnagar | 1904 | 0.937 | 1.126 | 1.051 | 2.378 | 4.365 | 120.658 | 117.058 | 38.47 | 225.726 | 72.389 | 0.271 | 0.0 |
| 26830 | Ahmadnagar | 1905 | 0.0 | 0.622 | 0.0 | 0.452 | 4.115 | 114.385 | 197.696 | 116.061 | 54.338 | 66.508 | 6.581 | 0.0 |
| 26831 | Ahmadnagar | 1906 | 3.061 | 1.005 | 0.124 | 0.006 | 4.612 | 259.556 | 159.938 | 285.418 | 62.414 | 14.387 | 3.823 | 19.299 |
| 26832 | Ahmadnagar | 1907 | 0.0 | 4.317 | 1.187 | 57.017 | 0.282 | 148.949 | 407.846 | 199.486 | 82.819 | 4.017 | 0.544 | 1.968 |
| 26833 | Ahmadnagar | 1908 | 0.228 | 0.0 | 4.065 | 2.63 | 2.059 | 118.277 | 151.933 | 107.993 | 212.96 | 16.046 | 0.678 | 0.0 |
| 26834 | Ahmadnagar | 1909 | 0.745 | 0.345 | 0.301 | 2.323 | 72.218 | 208.955 | 143.086 | 159.47 | 207.009 | 20.276 | 0.271 | 5.98 |
| 26835 | Ahmadnagar | 1910 | 0.0 | 0.0 | 0.008 | 0.006 | 7.74 | 296.432 | 318.644 | 191.564 | 265.571 | 60.167 | 52.152 | 0.0 |
| 26836 | Ahmadnagar | 1911 | 1.584 | 0.0 | 0.521 | 0.351 | 3.158 | 101.505 | 444.751 | 144.037 | 31.325 | 2.432 | 49.787 | 2.649 |
| 26837 | Ahmadnagar | 1912 | 0.0 | 1.166 | 0.0 | 5.83 | 12.037 | 87.747 | 206.988 | 112.025 | 7.127 | 71.342 | 40.658 | 0.0 |
| 26838 | Ahmadnagar | 1913 | 0.0 | 0.0 | 0.0 | 1.719 | 23.627 | 361.091 | 312.628 | 74.991 | 72.612 | 36.054 | 0.33 | 2.47 |
| 26839 | Ahmadnagar | 1914 | 0.0 | 0.168 | 0.07 | 1.169 | 11.297 | 268.759 | 228.972 | 131.181 | 167.434 | 12.747 | 30.745 | 9.453 |
| 26840 | Ahmadnagar | 1915 | 3.081 | 0.315 | 15.25 | 25.554 | 90.597 | 255.959 | 342.74 | 58.797 | 271.783 | 91.273 | 14.348 | 25.365 |
| 26841 | Ahmadnagar | 1916 | 0.0 | 0.0 | 2.368 | 4.886 | 54.521 | 140.368 | 581.759 | 94.054 | 236.88 | 149.621 | 24.658 | 0.0 |
| 26842 | Ahmadnagar | 1917 | 0.126 | 5.627 | 3.366 | 9.791 | 10.82 | 159.173 | 89.7 | 254.389 | 253.643 | 147.429 | 22.548 | 0.0 |
| 26843 | Ahmadnagar | 1918 | 0.569 | 10.947 | 1.81 | 1.685 | 140.872 | 66.522 | 47.636 | 54.943 | 64.184 | 18.455 | 8.772 | 6.489 |
| 26844 | Ahmadnagar | 1919 | 0.295 | 0.255 | 1.448 | 3.082 | 18.446 | 240.662 | 185.198 | 66.495 | 266.7 | 24.18 | 38.946 | 0.954 |
| 26845 | Ahmadnagar | 1920 | 3.554 | 0.522 | 0.0 | 18.65 | 6.27 | 91.08 | 95.867 | 51.391 | 139.689 | 14.646 | 0.272 | 0.0 |

# Monthly mean rainfall of Pune over 100 years

- Total number of data points for Pune: 1224 (102 years * 12 months)
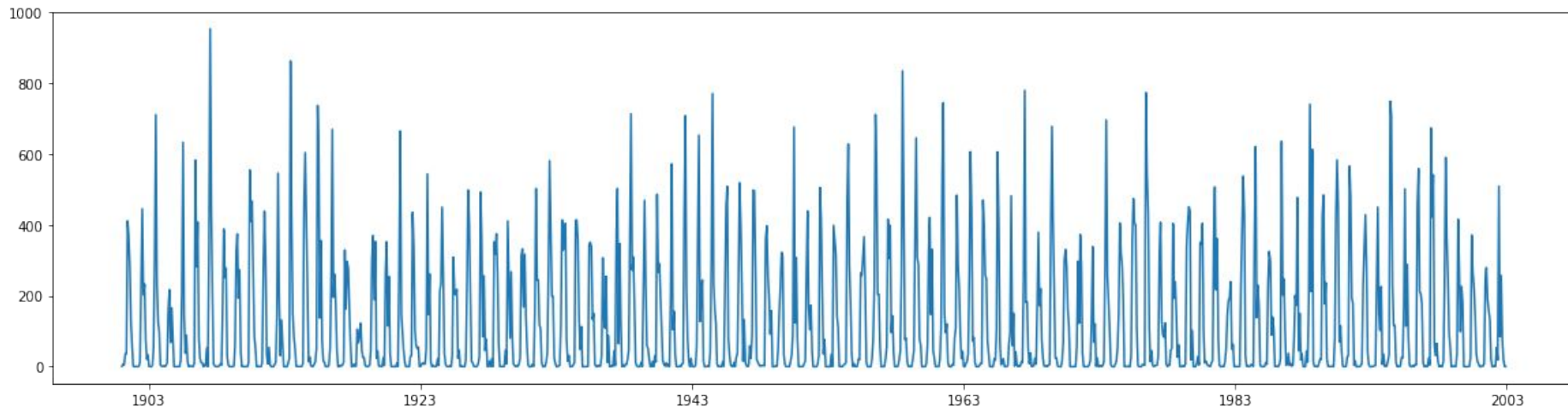
# Correlation of rainfall between districts

- The correlation between districts ranges from 0.63 to 0.992
- The least correlation is 0.63 which says that the variation of rainfall in Maharashtra is almost same in all districts.
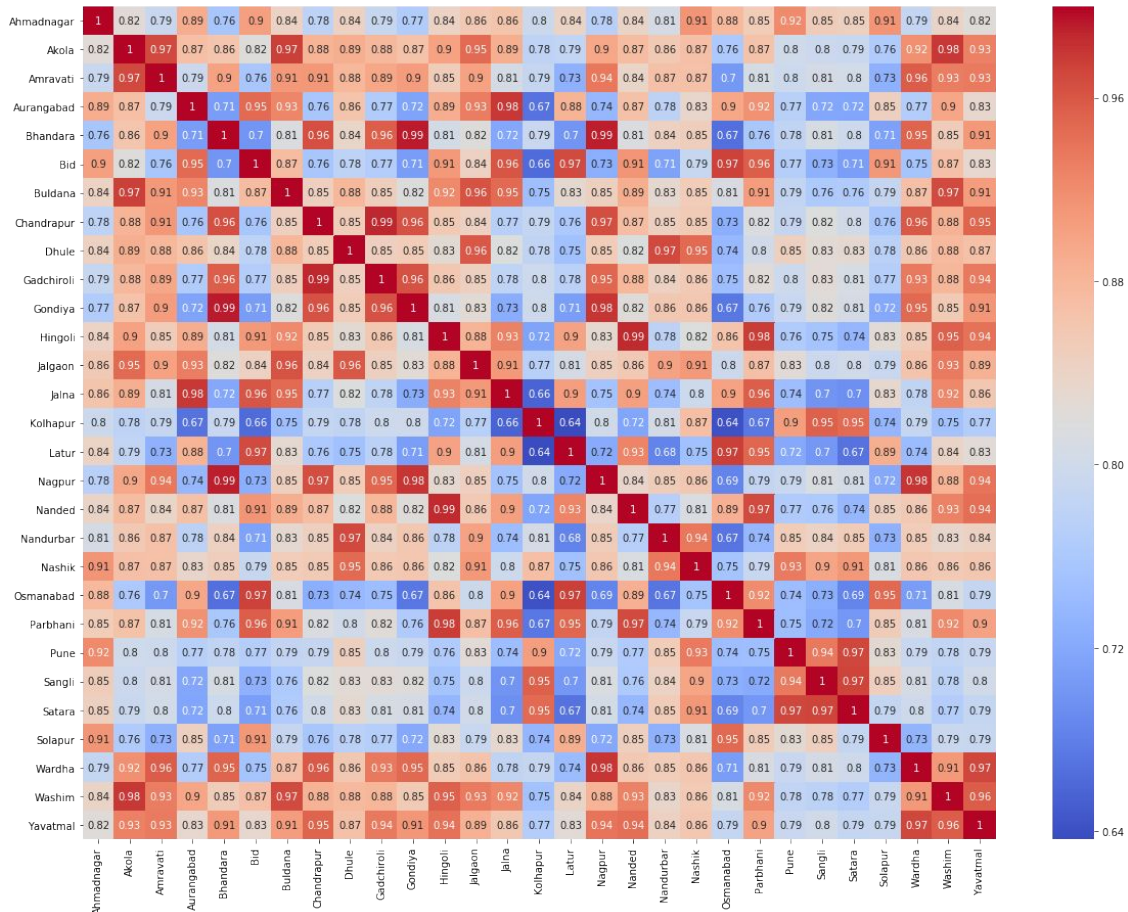
# Highly correlated districts

- The correlation between Bhandara and Gondiya is 0.992(highest).
- The districts that are correlated more than 0.98 are plotted on the map.
- We can see that the districts that are highly correlated are adjacent to each other.



Highly correlated districts
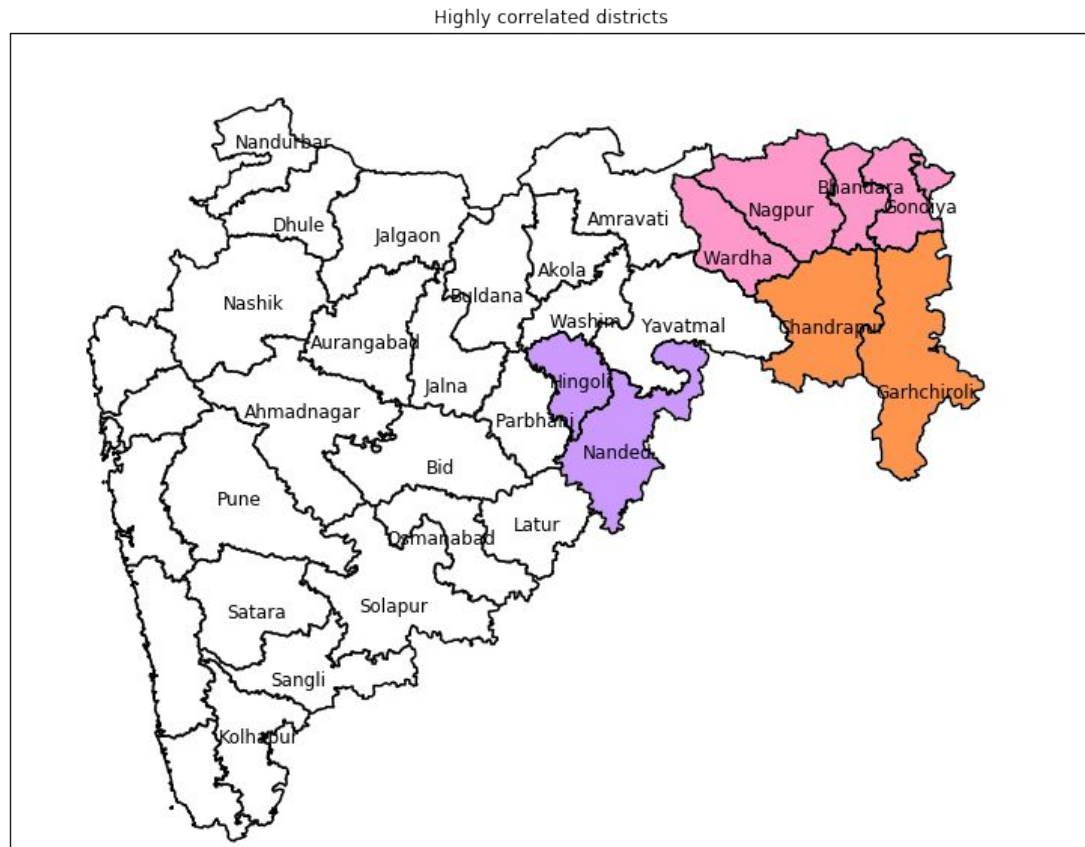
# Box plot: Range of rainfall in all districts

- In Kolhapur district, the rainfall is more than 800mm for about 49 times in the past 100 years(1901 - 2002), 25 times in Satara.
- If we observe, Kolhapur and Satara are very near to each other (in map).



District wise precipitation across all years
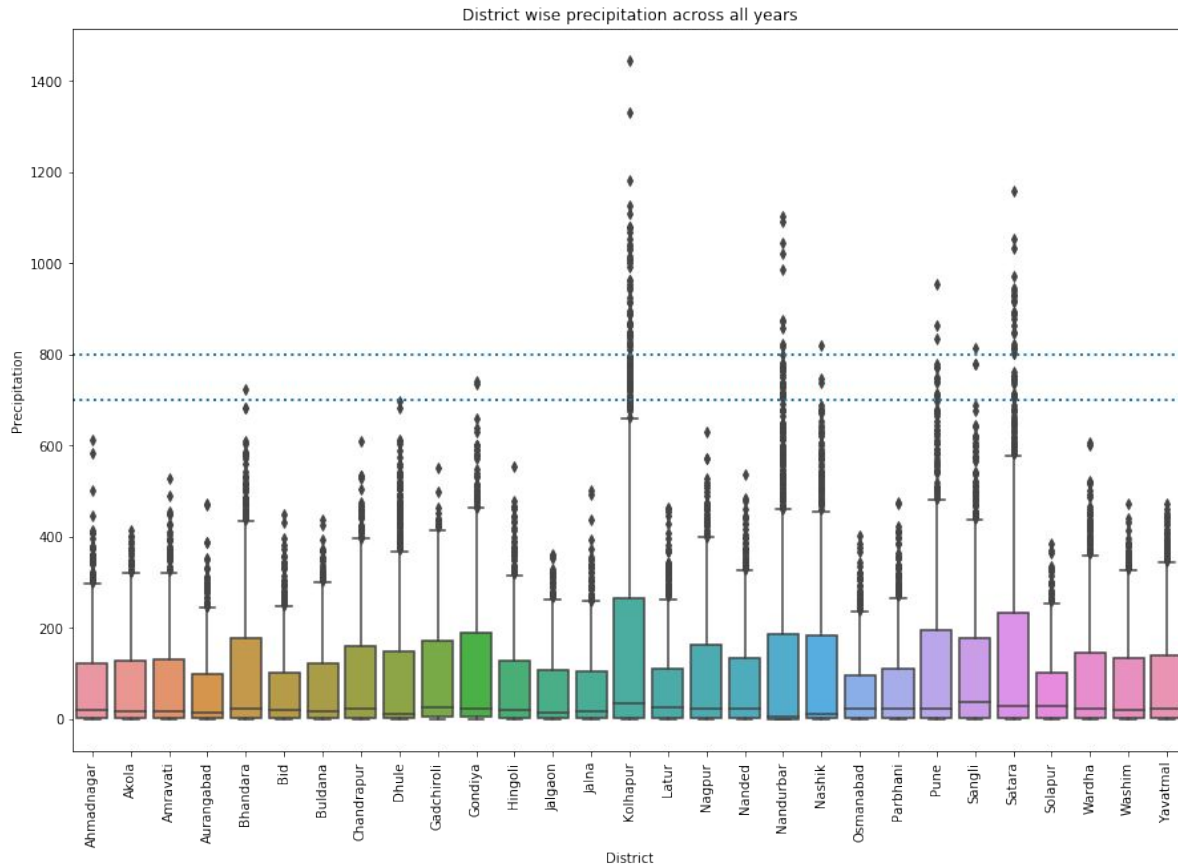
# Heavy rainfall districts

- In Kolhapur district, the rainfall is more than 800mm for about 49 times in the past 100 years(1901 - 2002), 25 times in Satara.
- If we observe, Kolhapur and Satara are very near to each other (in map).
- All the districts lie in a row.



Heavy rainfall districts
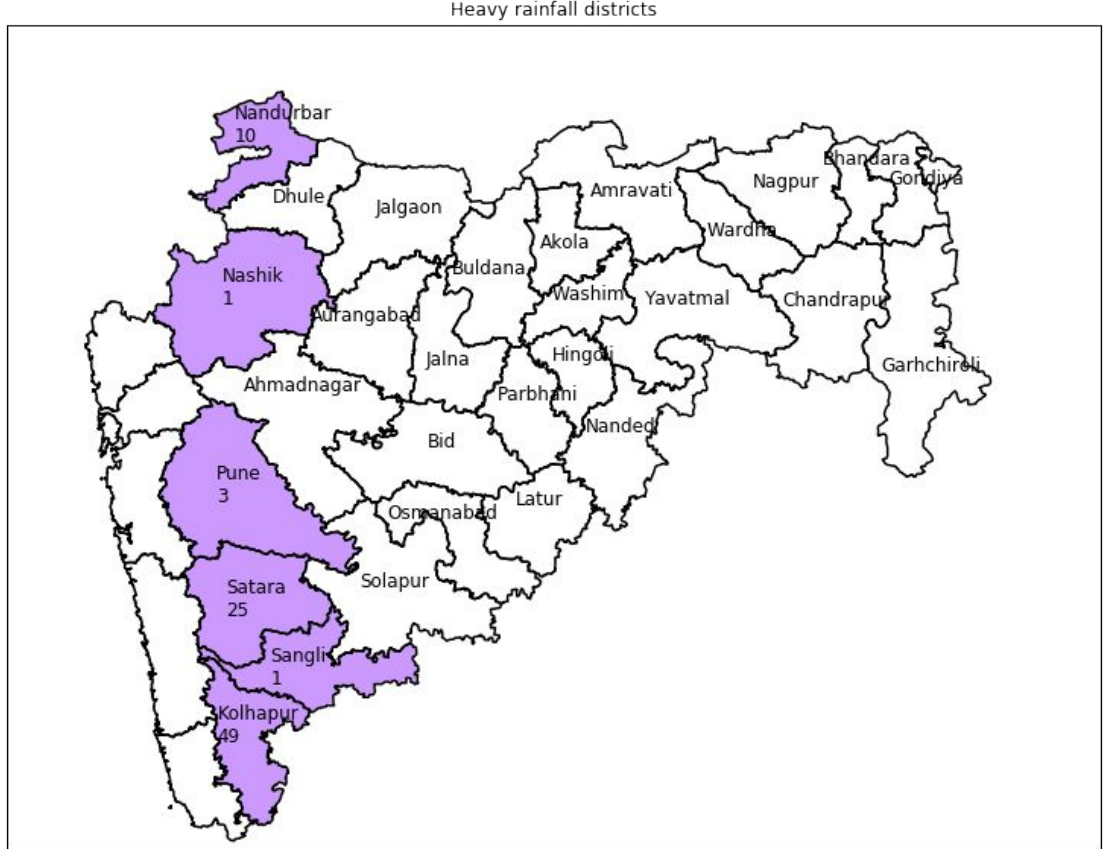
# Rainfall >= 800mm in 100 years

- The rainfall is heavy only in the months of June, July and August.
- Except Kolhapur, the rainfall was not much heavy in June and August.
- In Kolhapur, the average rainfall in July is 676.47mm.

| District | Count | June | July | August |
|----------|-------|------|------|--------|
| Kolhapur | 49 | 17 | **27** | 5 |
| Nandurbar | 10 | - | **10** | - |
| Nashik | 1 | - | **1** | - |
| Pune | 3 | 1 | **2** | - |
| Sangli | 1 | - | **1** | - |
| Satara | 25 | 5 | **20** | - |



Rainfall of Kolhapur district
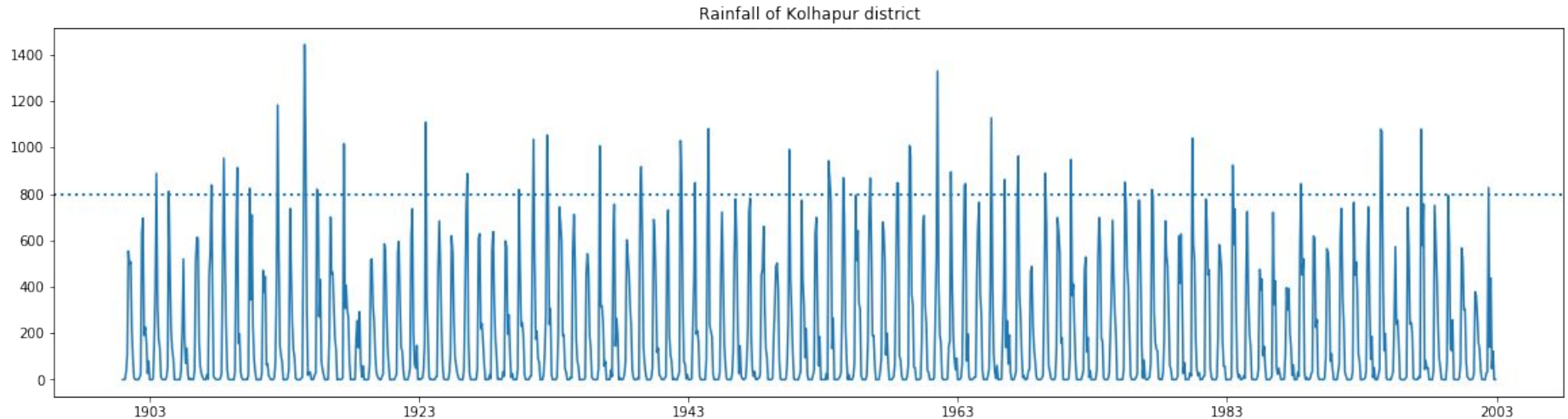
# Monthly rainfall for each year in Maharashtra

- The rainfall is added over all districts of Maharashtra to plot this graph.
- The graph shows that the rainfall is high in July. We can easily interpret(pink colour line).
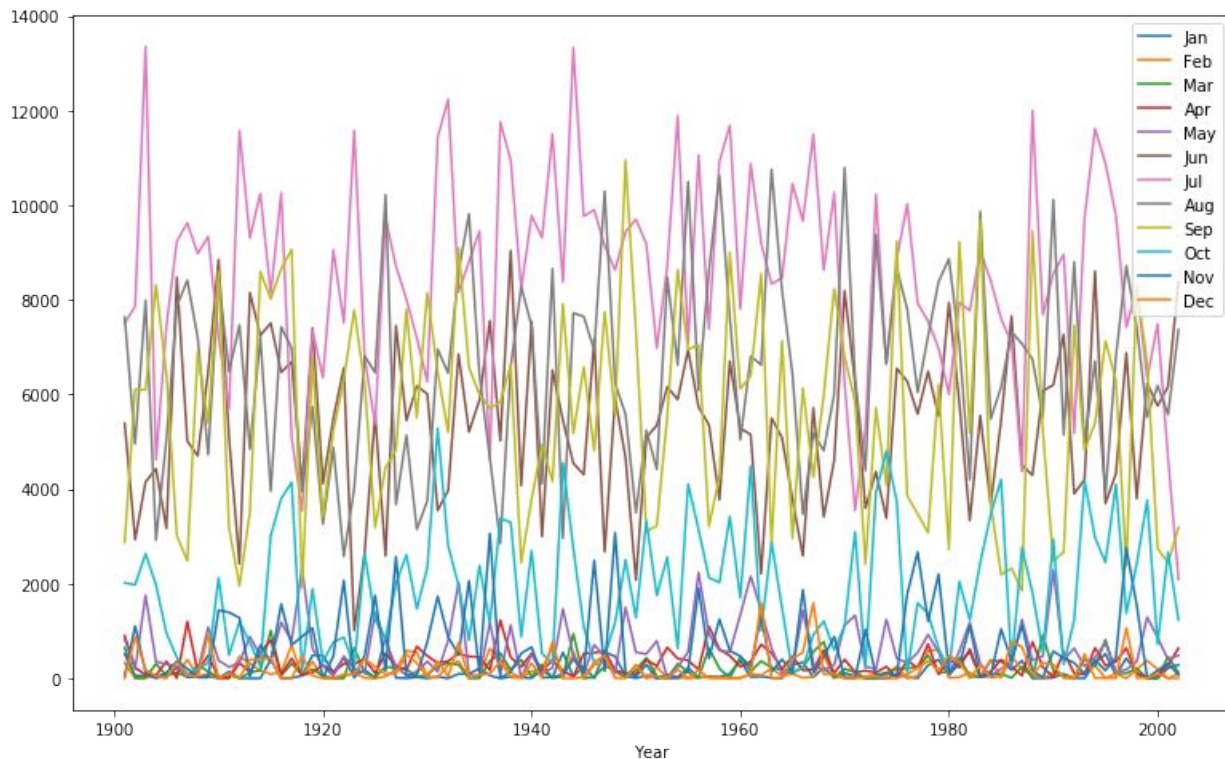
# Seasonal rainfall for each year in Maharashtra

- We divided the months into four seasons to know the season in which the rainfall is high.
- From the graph, we can say that the rainfall is high compared in June September when compared to other seasons.
- The rainfall is very less in the months of January and February.
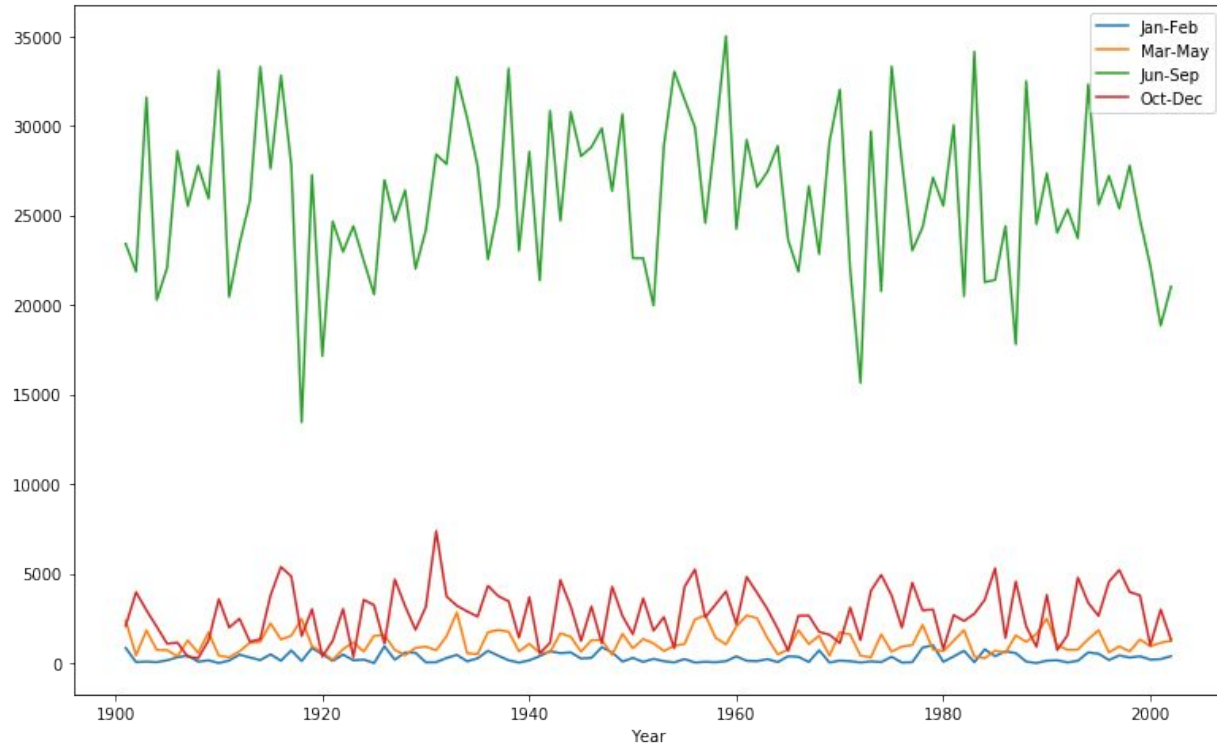
# District wise Seasonal rainfall across all years

- This graph gives intuition about rainfall in all districts.
- The rainfall is high in the season June to September in all districts.
- If we see Kolhapur district, the rainfall is very high compared to other districts(June to September). This is because the rainfall is more than 800mm 43 times(we got this before itself).
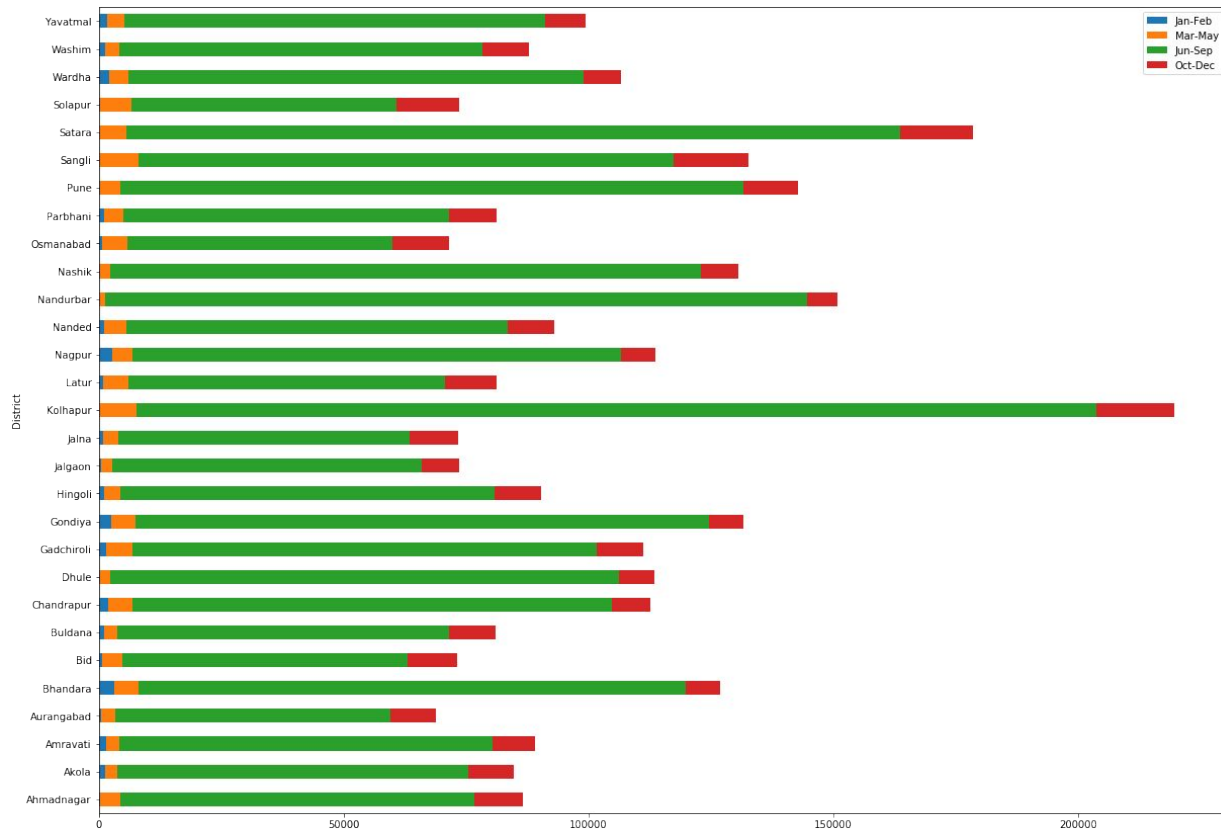
# District wise Monthly mean rainfall across all years

- This is similar to the previous graph. This is for all months instead of seasons.
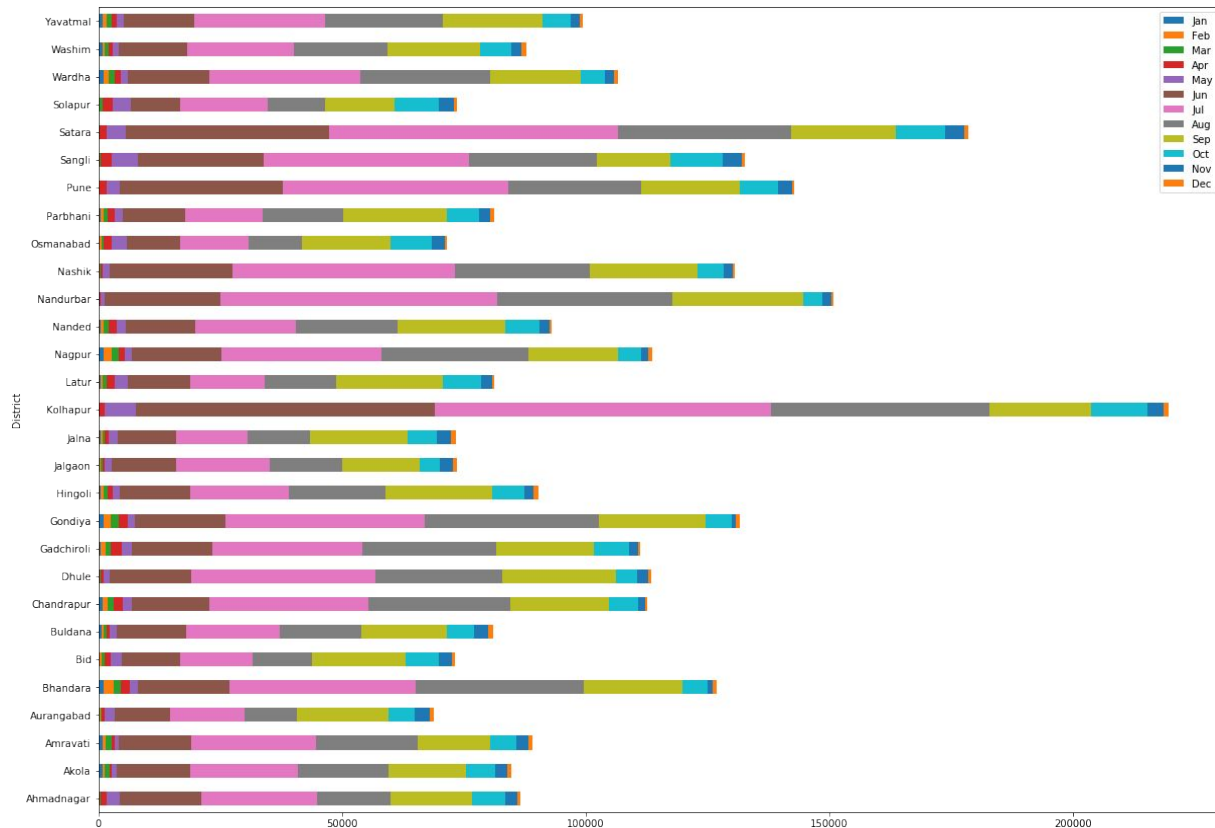
# Annual rainfall in Maharashtra

- If we see, the rainfall is very very less in the years 1918, 1920 and 1972.
- From a quick Google search, we got this:
  - "1970-73 **years** were **years** of droughts over **Maharashtra**. These **years** created an extremely difficult situation for the government and people of **Maharashtra**. 1971 and particularly **1972** were the **years** of deficient rainfall over **Maharashtra**."
  - The worst **drought years** observed over **Maharashtra** State are 1918 and 1920.
- So, the data is saying the truth.
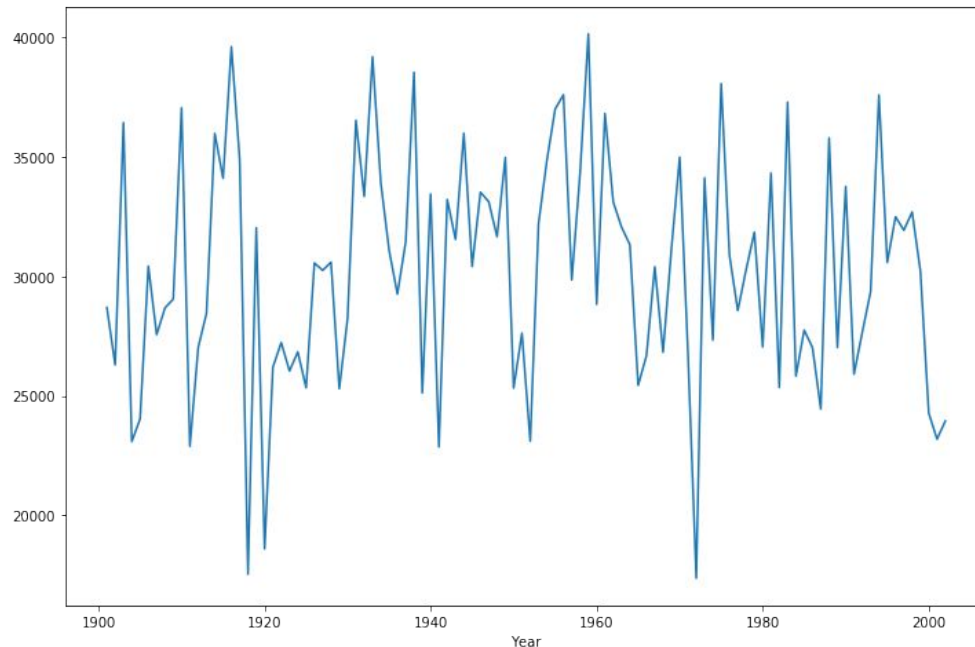
# Neural Networks - Forecasting time series data

- Some of the Neural Network architectures that are widely used for time series data are:
    - Feed forward Neural Network(FNN)
    - Time Lagged Neural Network(TLNN)
    - Seasonal Artificial Neural Network(SANN)
    - Long Short Term Memory(LSTM)
- ARIMA is a statistical method that is widely used for time series data.
- We have compared the above methods for each district of Maharashtra.

# Neural Networks - Feed forward Neural Network

- The most widely used ANNs in forecasting problems are multi-layer perceptrons (MLPs), which use a single hidden layer feed forward network(FNN).
- The model is characterized by a network of three layers, viz. input, hidden and output layer. There may be more than one hidden layer.
- Here, the output($y_t$) depends on the previous 'p' time steps.
- So, the output layer consists of one unit and input layer layer consists of p units.



Outputs

Output Layer

Weights

Middle or Hidden Layer

Weights

Input Layer
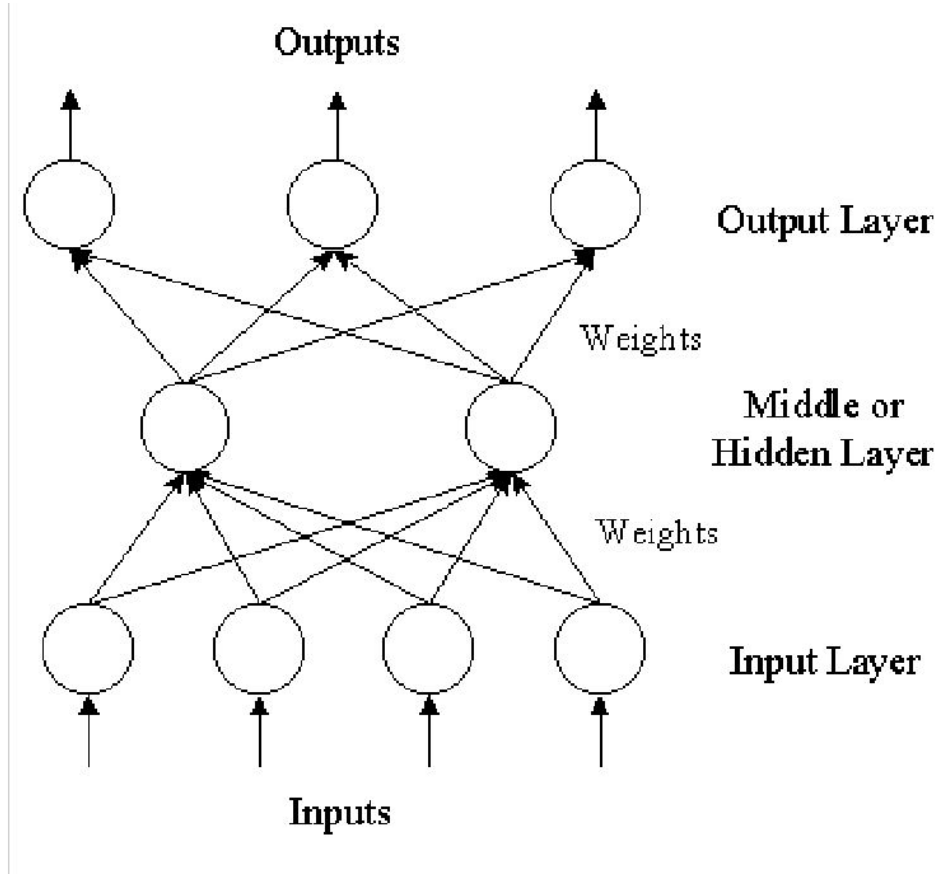
Inputs

# Neural Networks - Time Lagged Neural Network

- TLNN is another variation of FNN and it also widely used.
- In TLNN, the input nodes are the time series values at some particular lags. For example, a typical TLNN for a time series, with seasonal period s = 12 can contain the input nodes as the lagged values at time t − 1 , t − 2 and t − 12 . The value at time t is to be forecasted using the values at lags 1, 2 and 12 as shown in the figure.



**Input Layer**          **Hidden Layer**          **Output Layer**
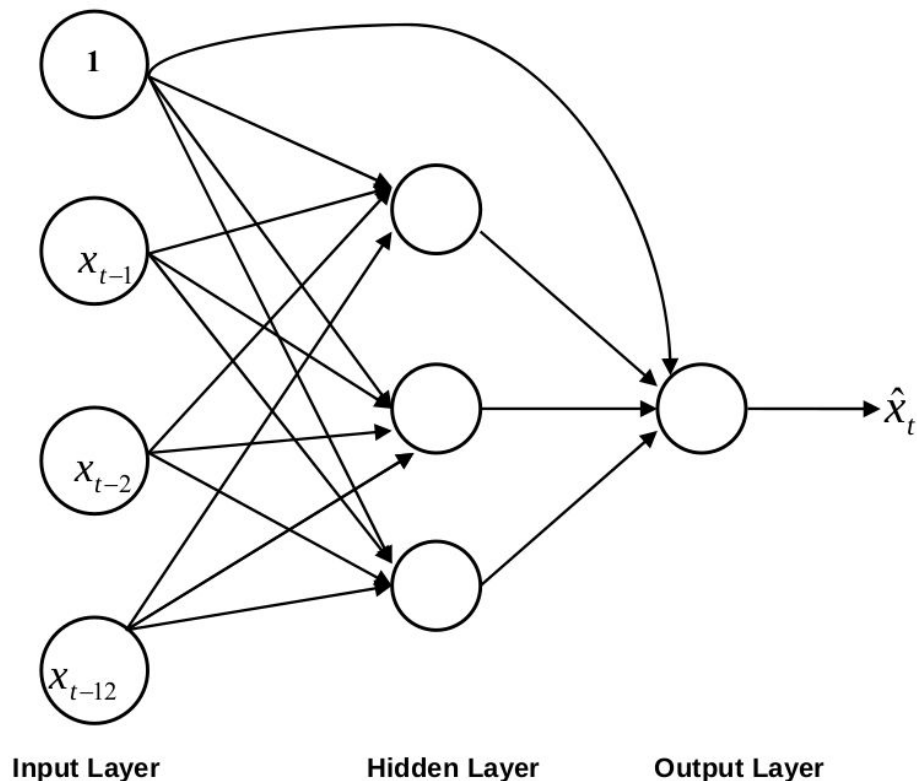
# Neural Networks - Seasonal Artificial Neural Network

- The SANN structure is proposed to improve the forecasting performance of ANNs for seasonal time series data.
- SANN can learn the seasonal pattern in the series, without removing them, contrary to some other traditional approaches, such as SARIMA.
- Thus while forecasting with SANN, the number of input and output neurons should be taken as 12 for monthly and 4 for quarterly time series. The appropriate number of hidden nodes can be determined by performing suitable experiments on the training data.
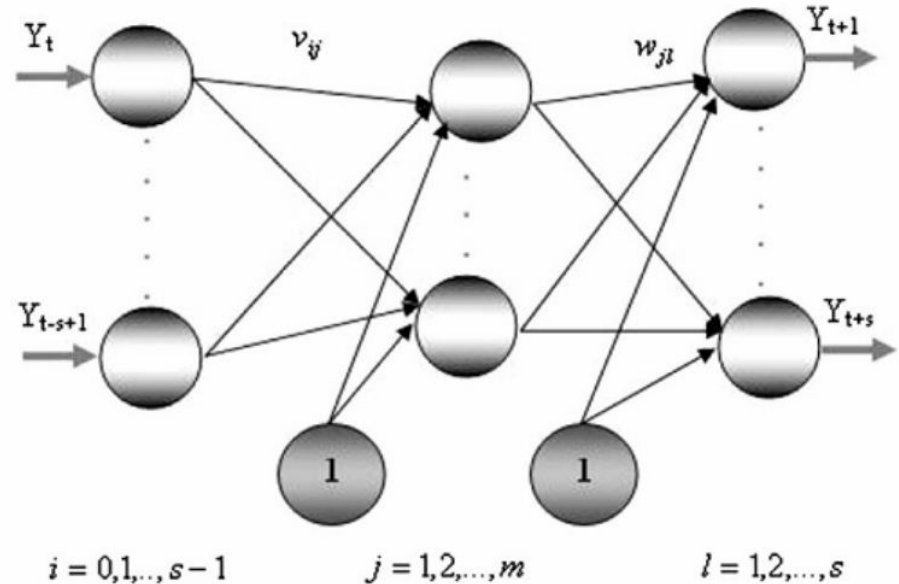
# Results of Kolhapur district

- LSTM
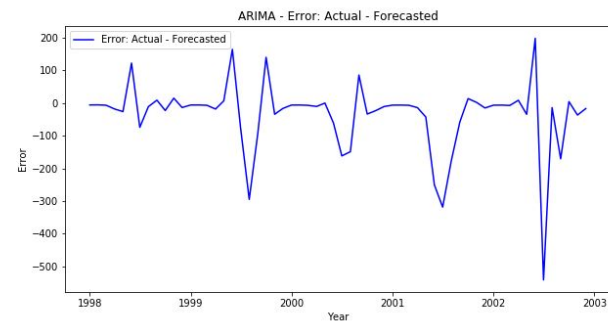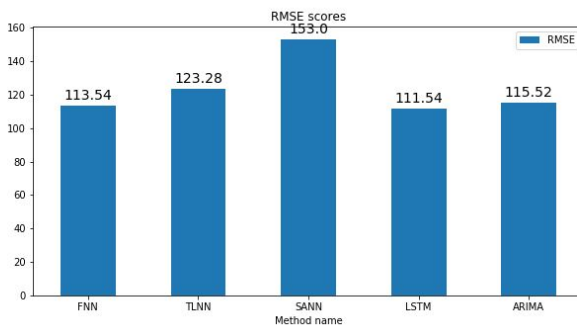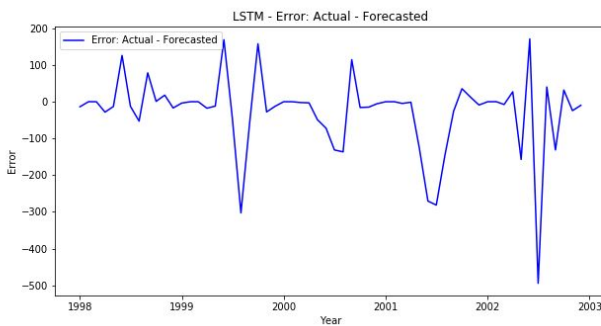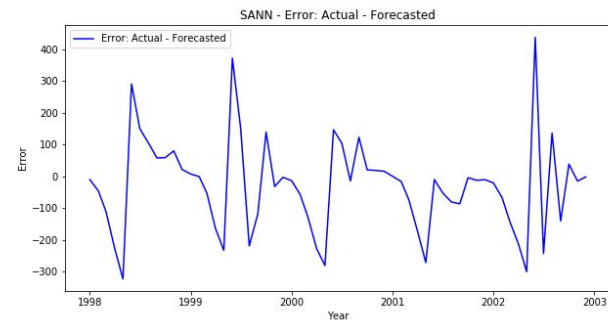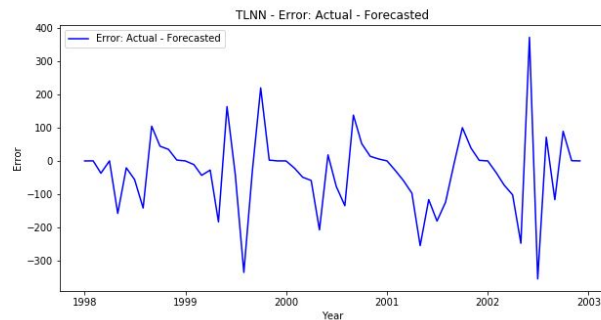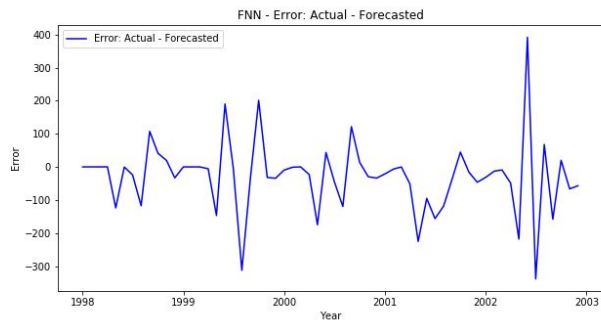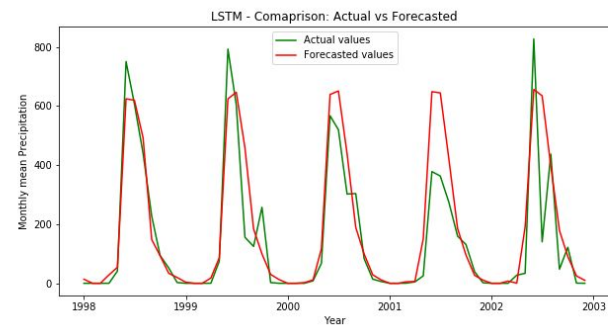
# District wise best method

- Out of 29 districts, ARIMA performed well for 9 districts and LSTM performed well for the other 20 districts.
- TLNN, SANN didn't perform well at least for one district.
- If we group the districts based on the best method of that district, fortunately they are clustered together.
- But, there is no reason for why ARIMA or LSTM is working better.
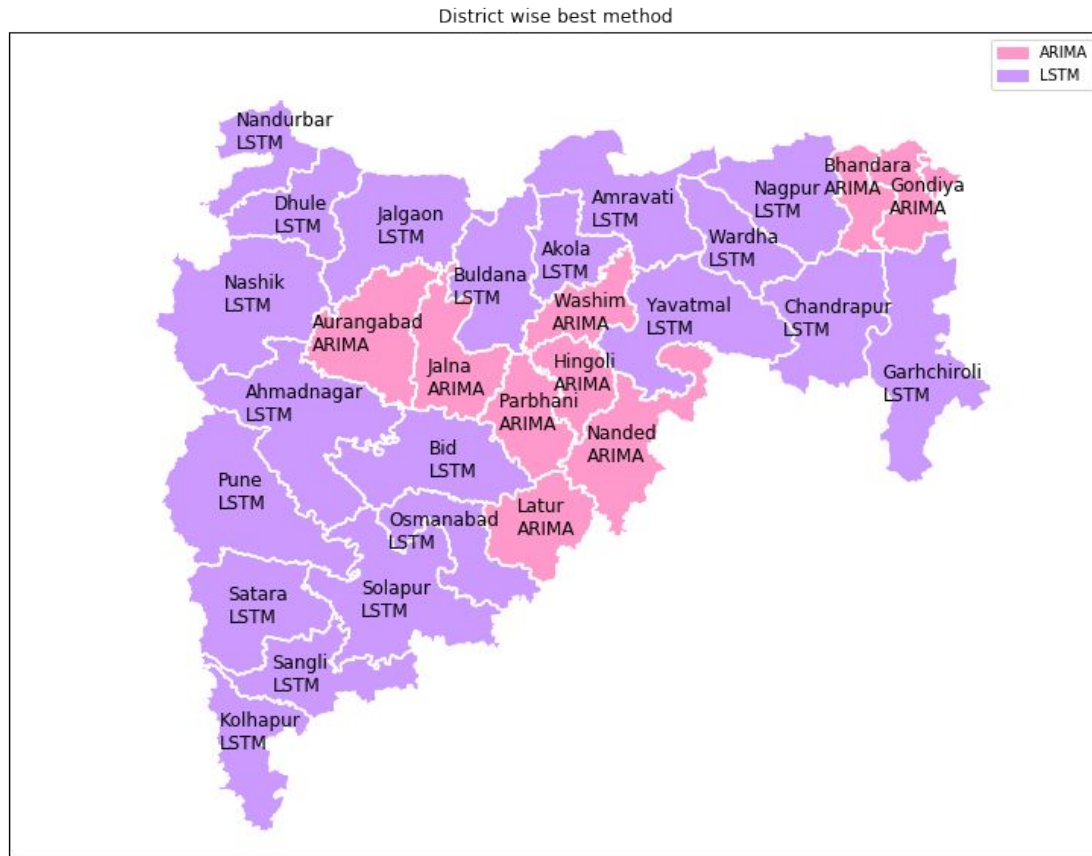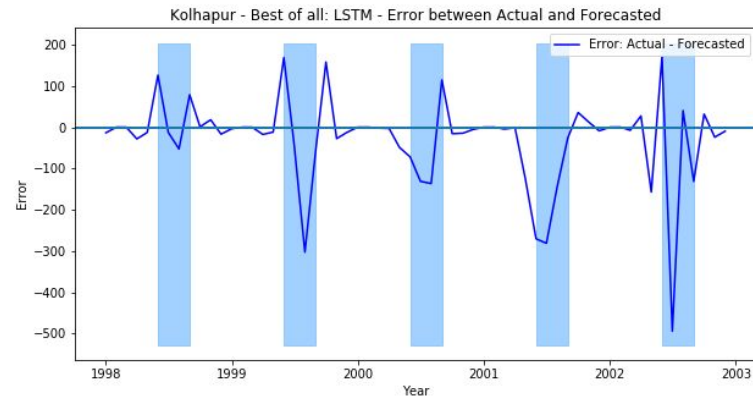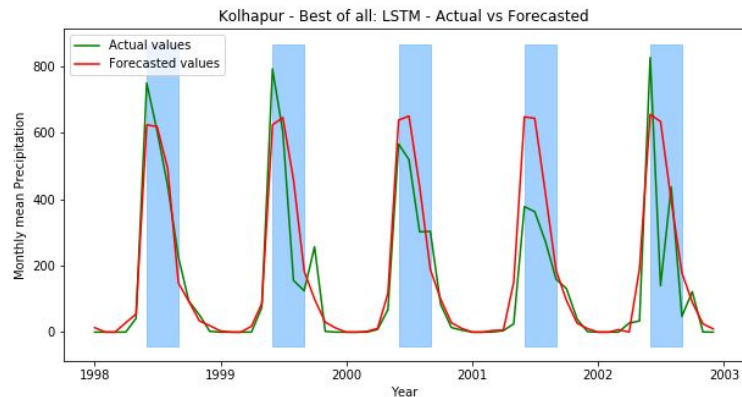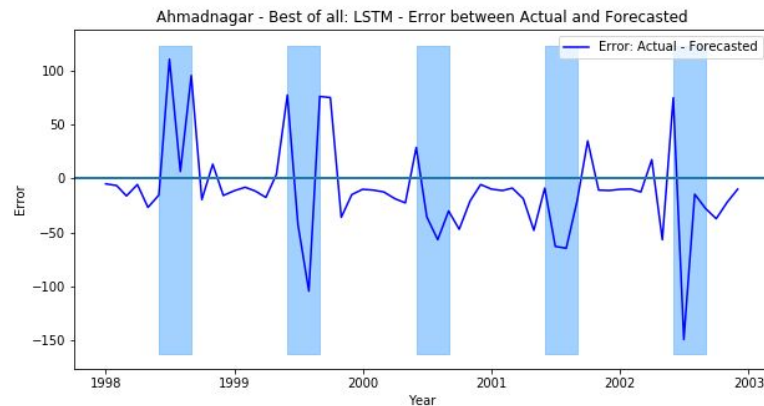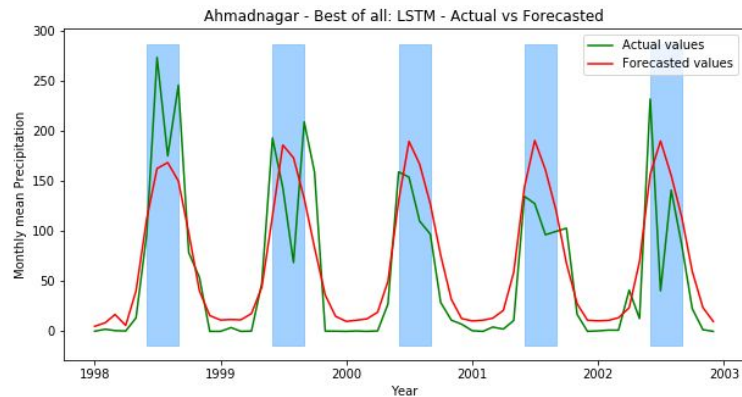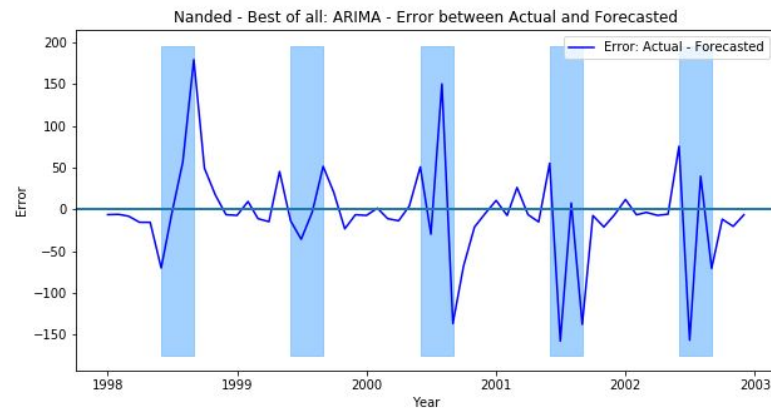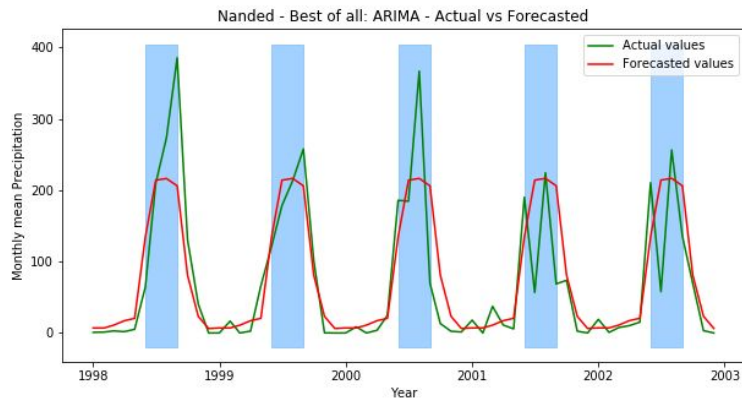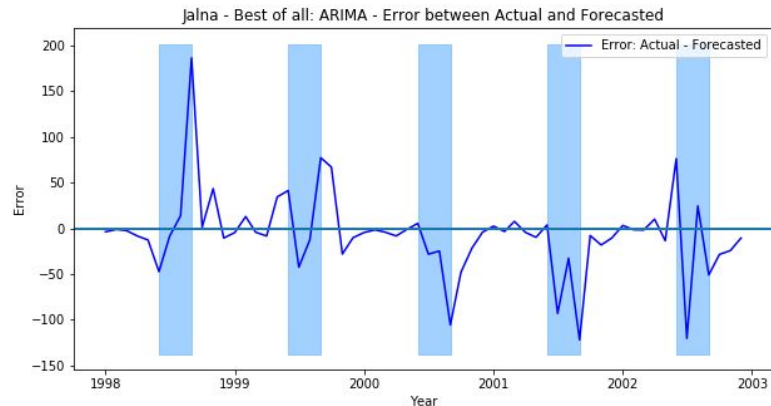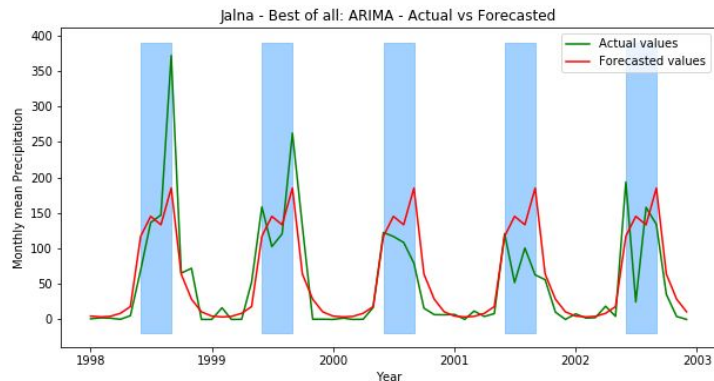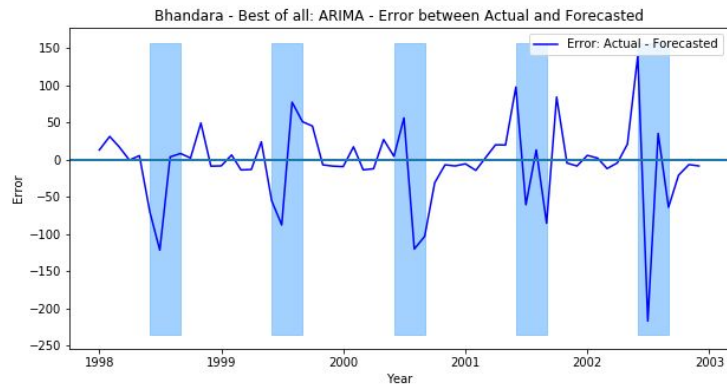
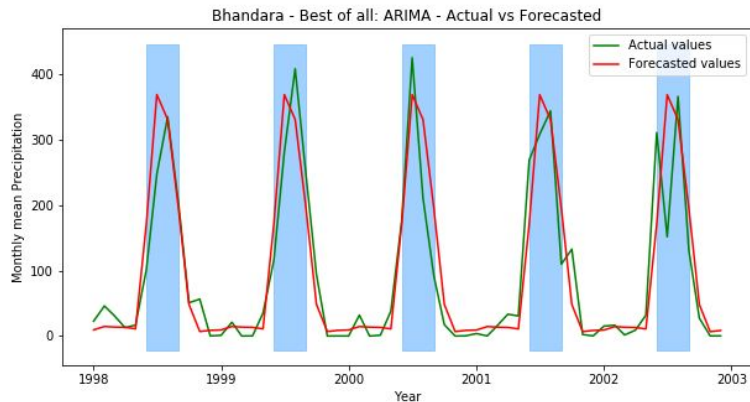

District wise best method

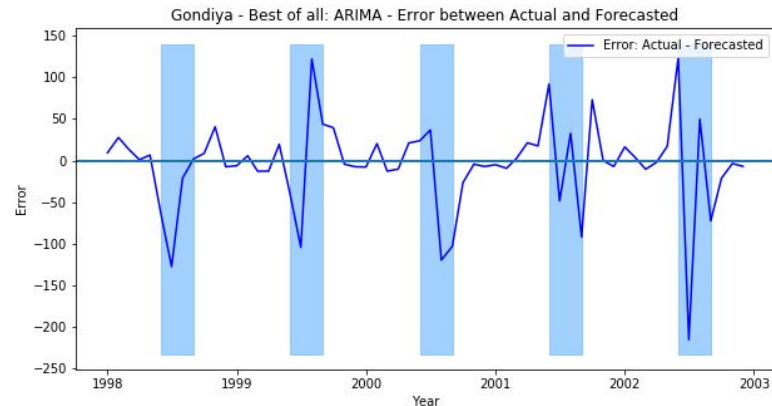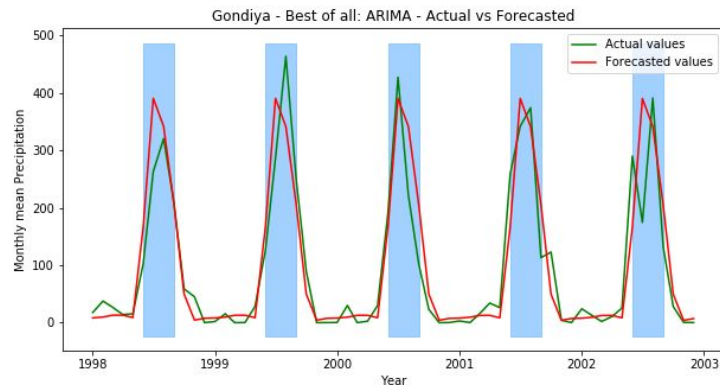# Some of the districts for which LSTM performed well

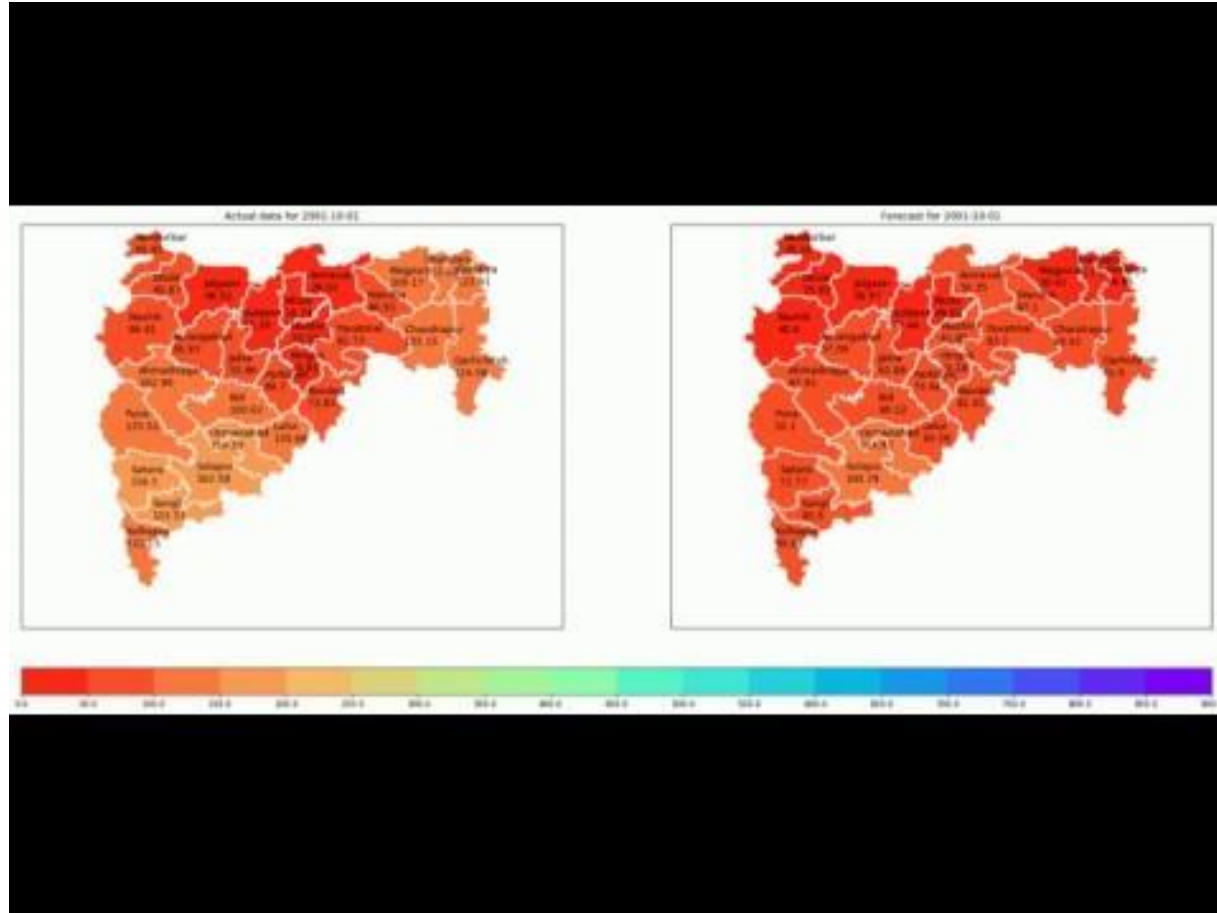# Some of the districts for which LSTM performed well
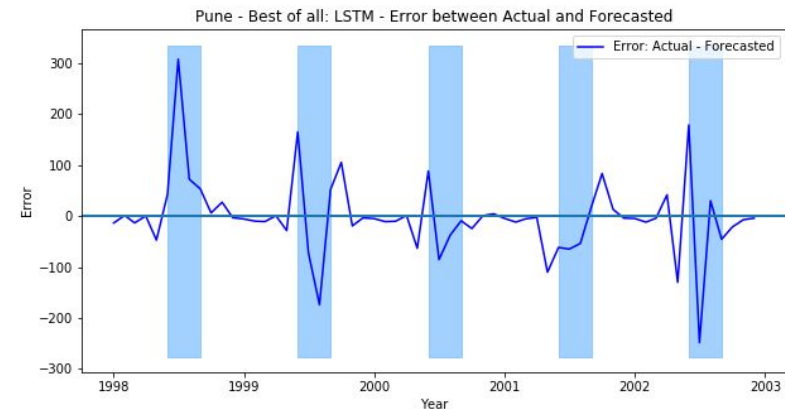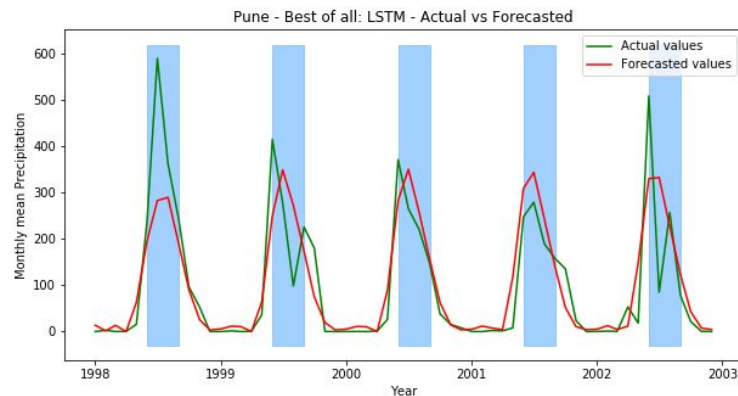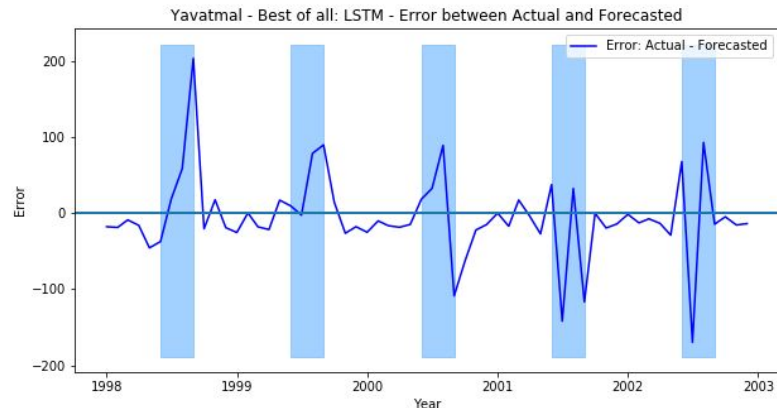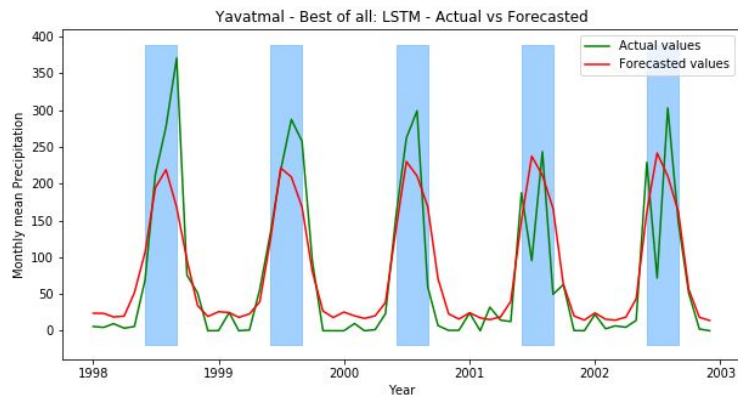
# Some of the districts for which ARIMA performed well

# Some of the districts for which ARIMA performed well

# Visualization of Actual and forecasted values

# Future work

- We will get real daily rainfall data for each station of Maharashtra for the past 40 - 50 years in a week.
- Till now, We have worked on Monthly rainfall. We may get more insights from the new data because that will be more accurate.
- Will work on Spatio-temporal correlation.

Thank you