

GDP Assignment

(Submitted By - Manvinder Singh)

(Roll Number - DDS1960097)

Folder Structure:

- gdp_assignment.ipynb is the jupyter notebook which contains all the code
- All the data files are present in the dataset folder
 - Data-1-A-GSDP.csv is the file used "Part I-A" Problem
 - Data-1-B folder contains all the files required for "Part I-B" Problem
 - "Dropout_data.csv" is the file for "Part-II" problem

Data I-A:

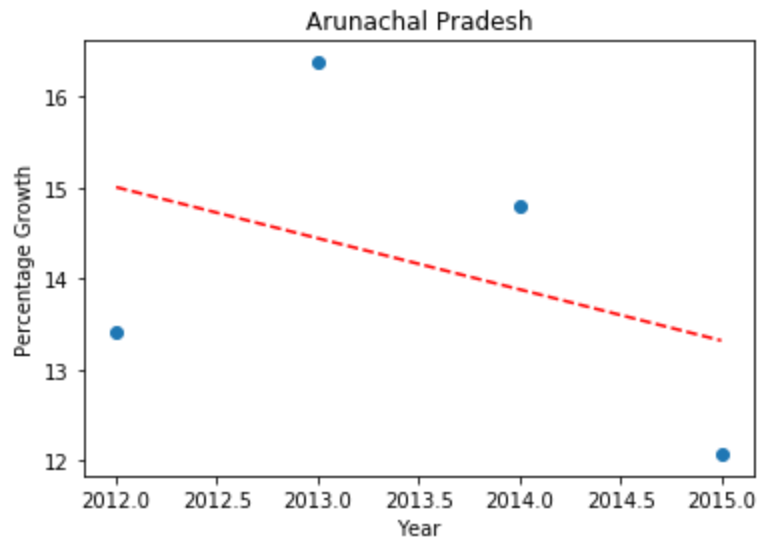
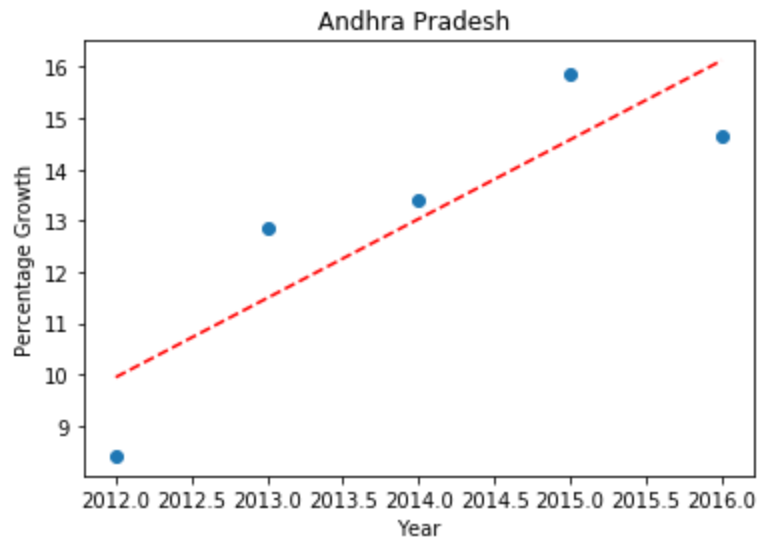
Steps Followed:

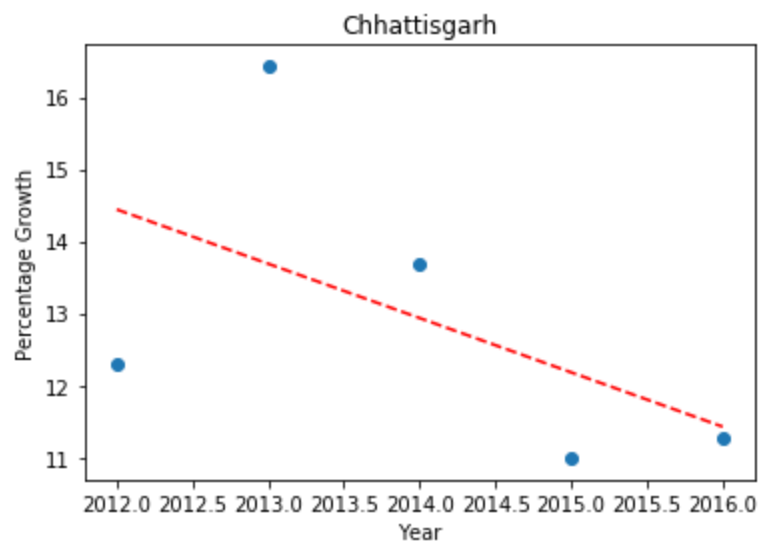
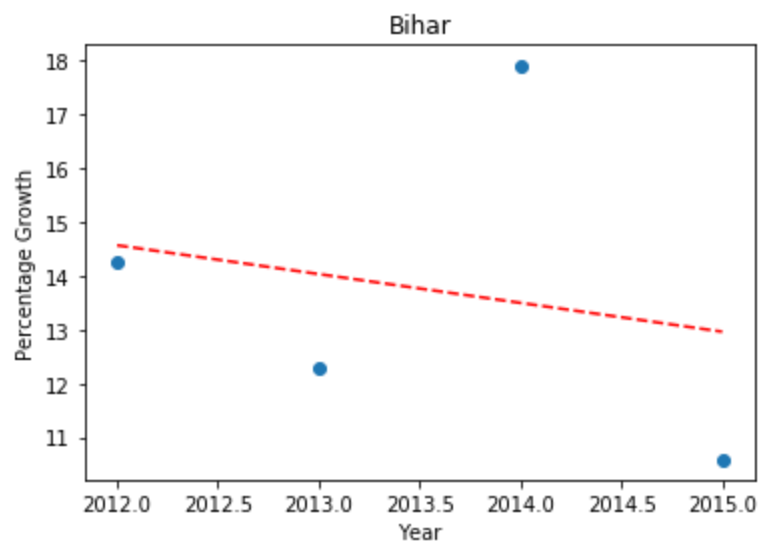
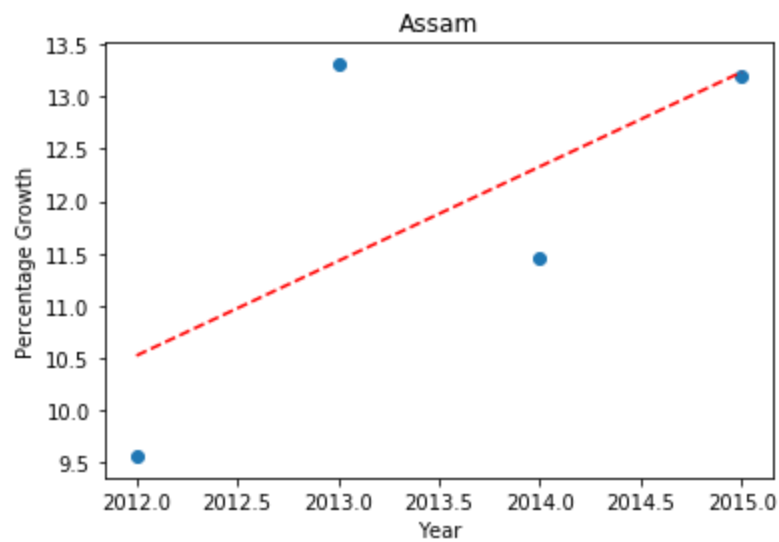
- Firstly the data file is loaded
- Then set the duration column to just the first part of duration for ease of calculation and visualization. i.e Convert "2014-15" to "2014"
- Print the percentage of "NA" values in each column.
 - Result:

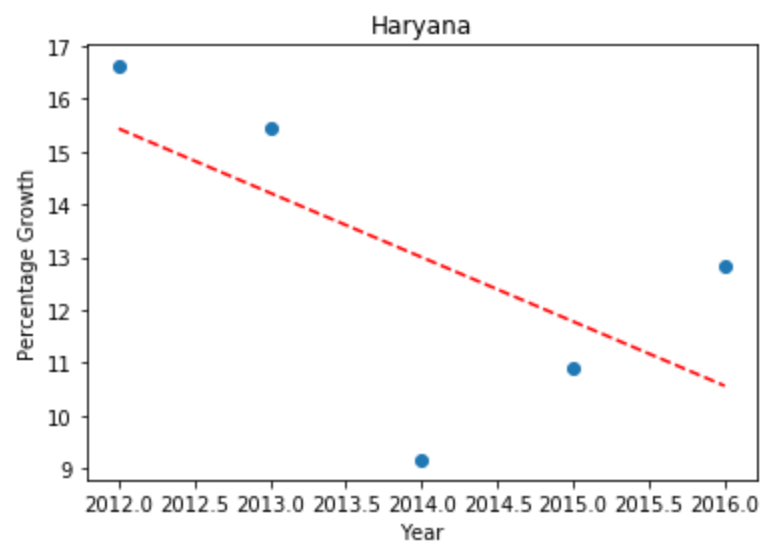
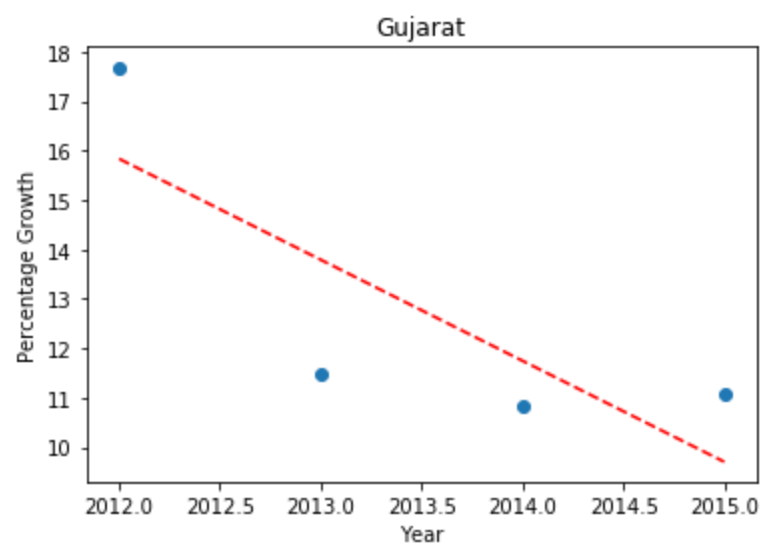
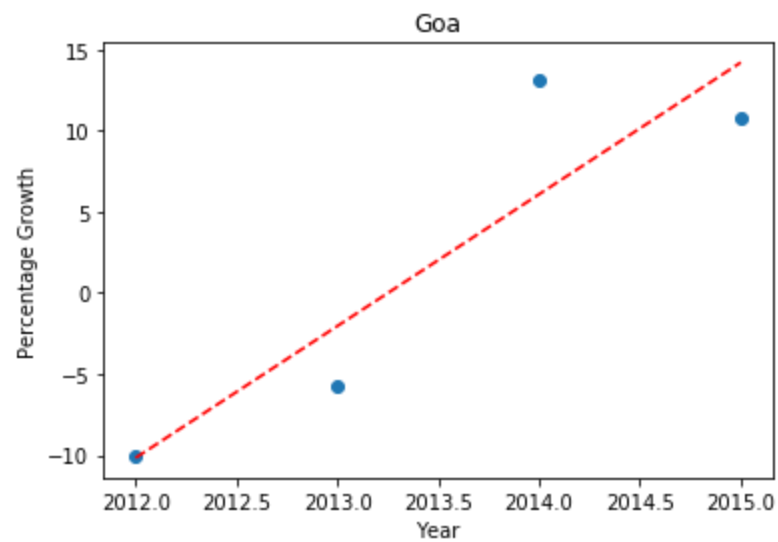
```
Items Description      0.000000
Duration              0.000000
Andhra Pradesh        0.000000
Arunachal Pradesh    18.181818
Assam                 18.181818
Bihar                 18.181818
Chhattisgarh         0.000000
Goa                   18.181818
Gujarat               18.181818
Haryana               0.000000
Himachal Pradesh     36.363636
Jammu & Kashmir       18.181818
Jharkhand             18.181818
Karnataka             18.181818
Kerala                18.181818
Madhya Pradesh        0.000000
Maharashtra           36.363636
Manipur               36.363636
Meghalaya             0.000000
Mizoram               36.363636
Nagaland              36.363636
Odisha                0.000000
Punjab                36.363636
Rajasthan             36.363636
Sikkim                18.181818
Tamil Nadu            0.000000
Telangana             0.000000
Tripura               36.363636
Uttar Pradesh         18.181818
Uttarakhand           18.181818
West Bengal1         100.000000
Andaman & Nicobar Islands 36.363636
Chandigarh            18.181818
Delhi                 0.000000
Puducherry            0.000000
All_India GDP         0.000000
dtype: float64
```

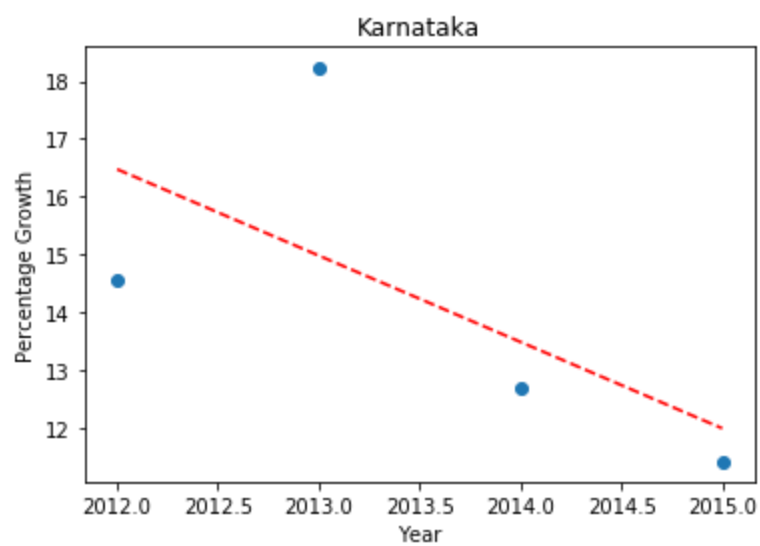
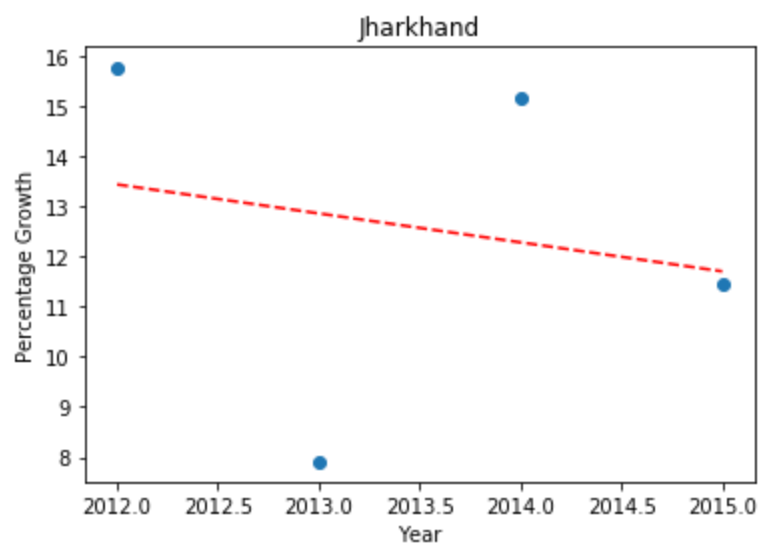
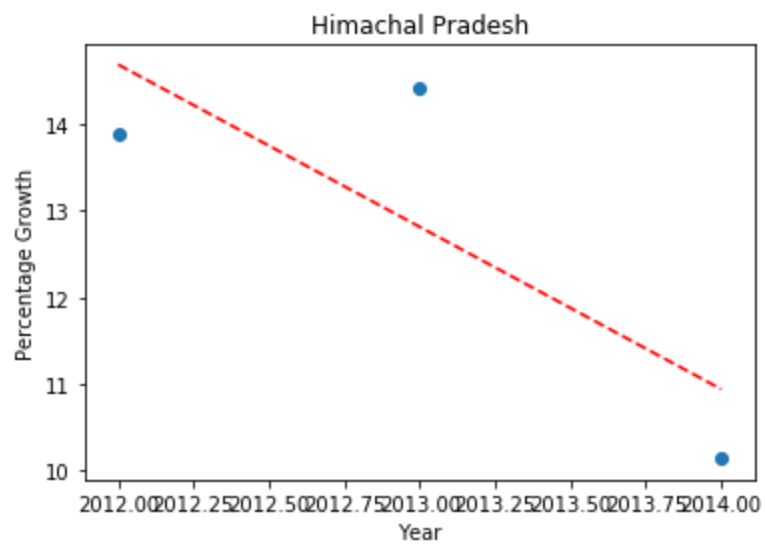
- Here you can see the for WestBengal1 100% data is not present so remove WestBengal

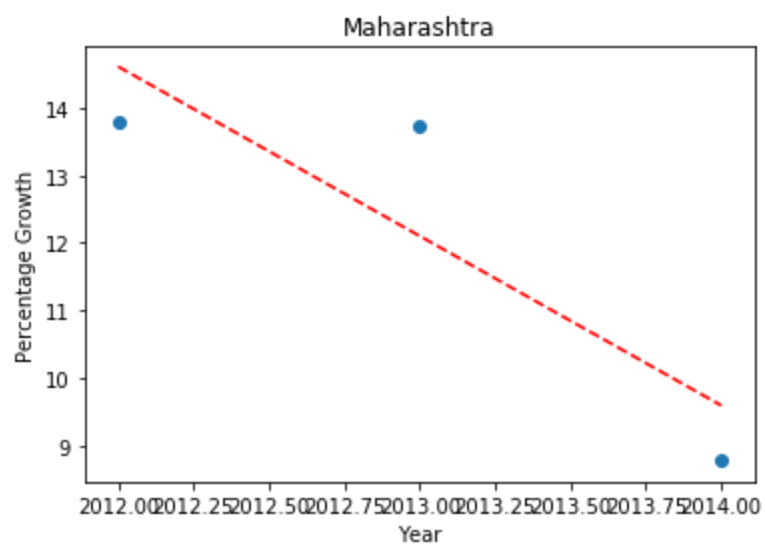
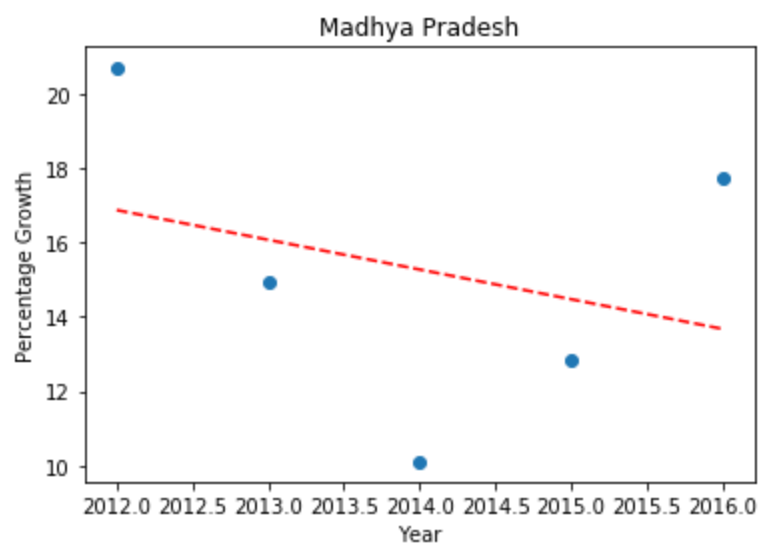
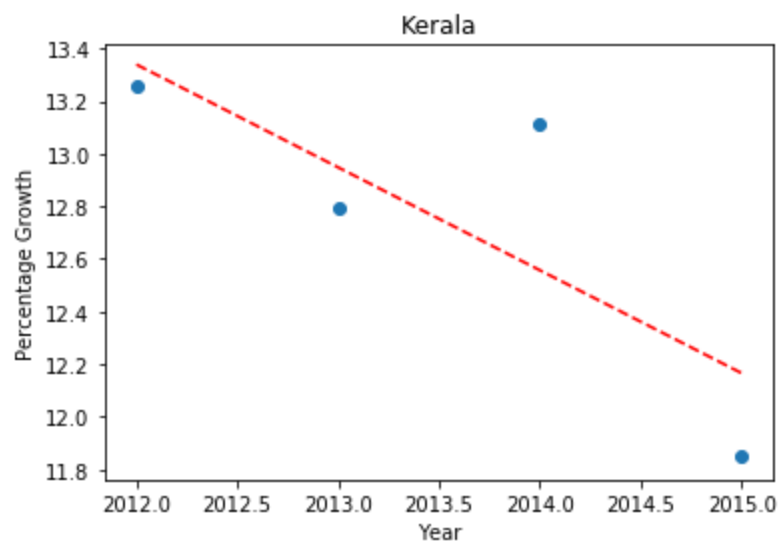
- Then Plot the graph of percentage growth over previous each for each state(not UT)

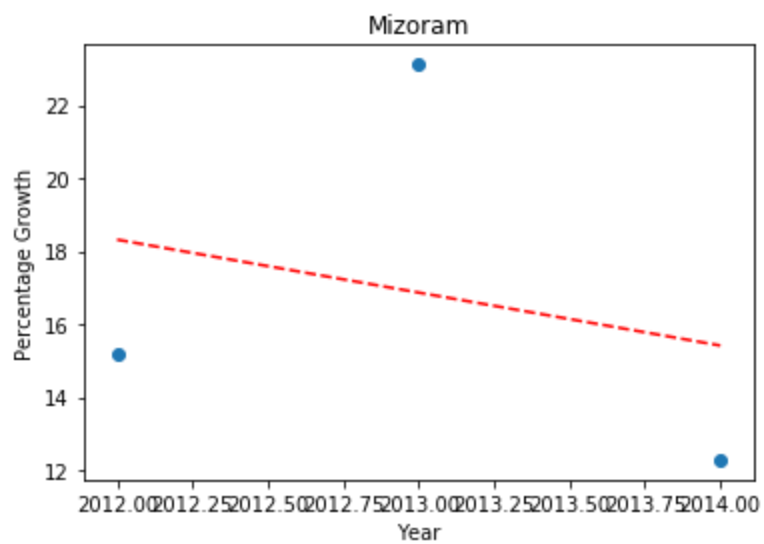
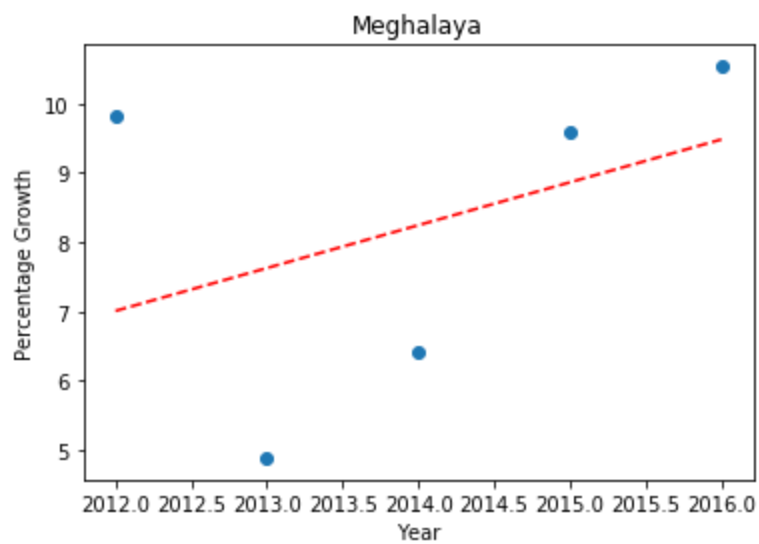
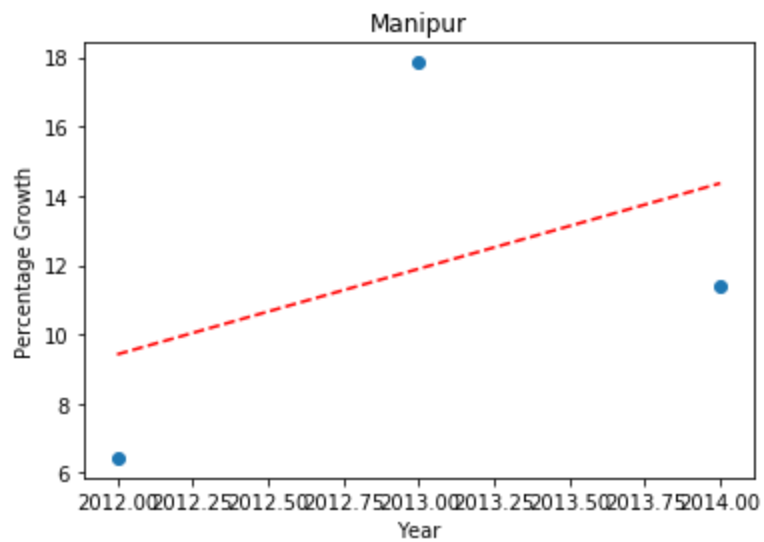


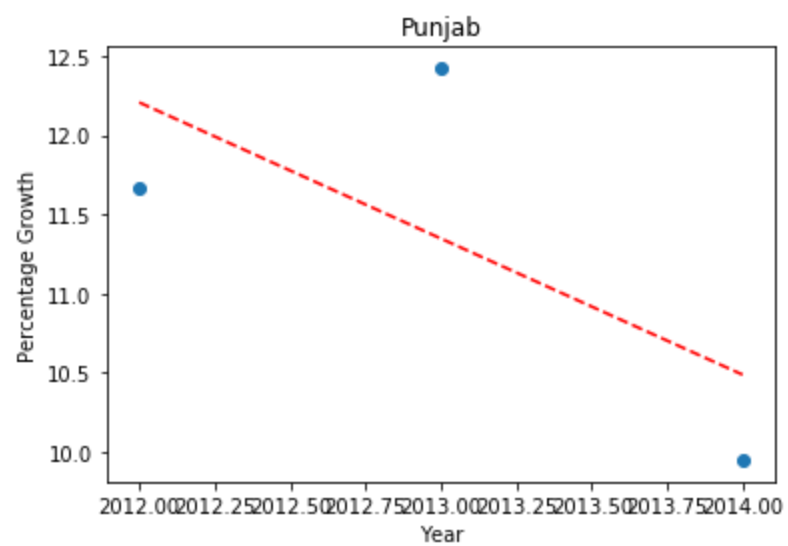
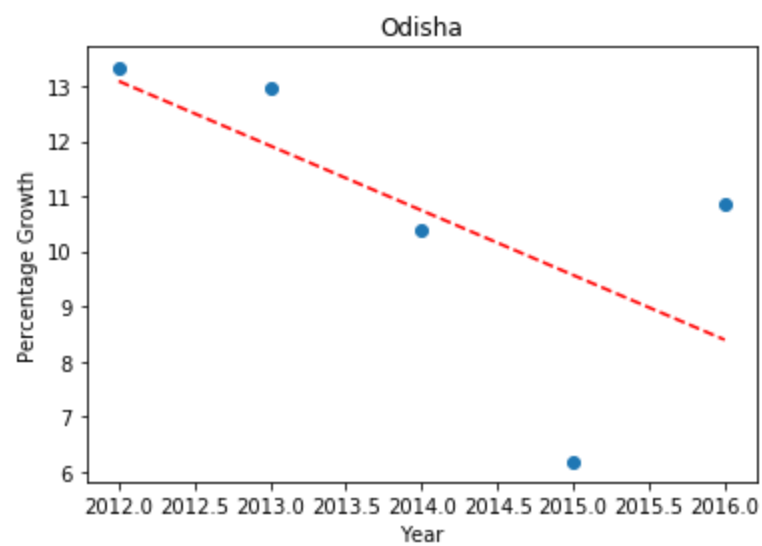
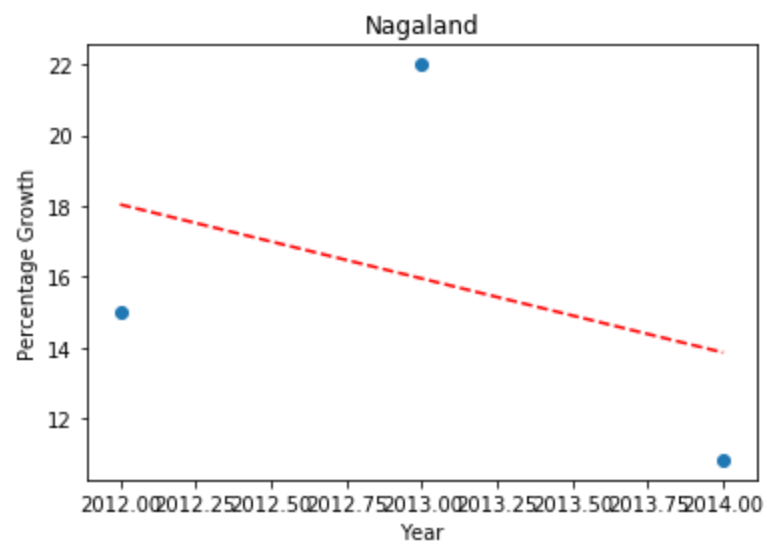


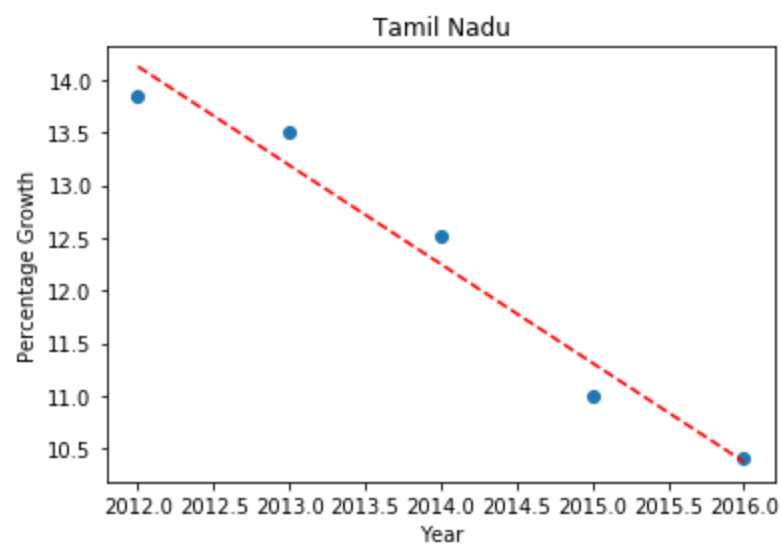
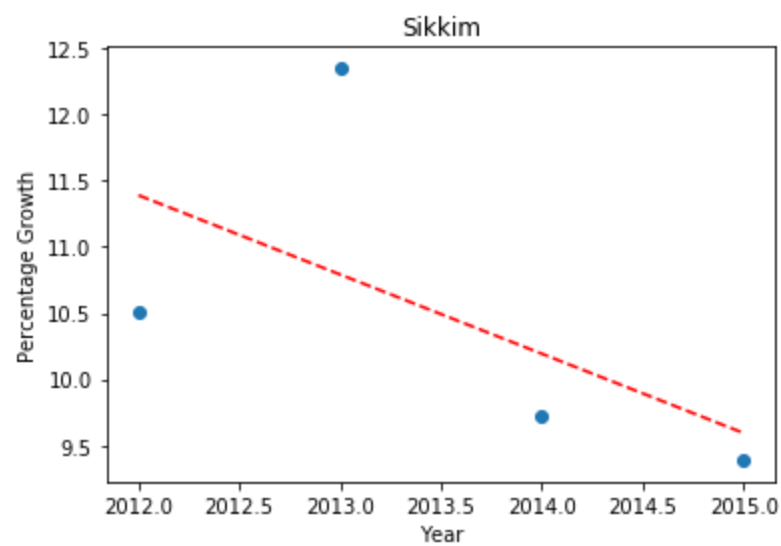
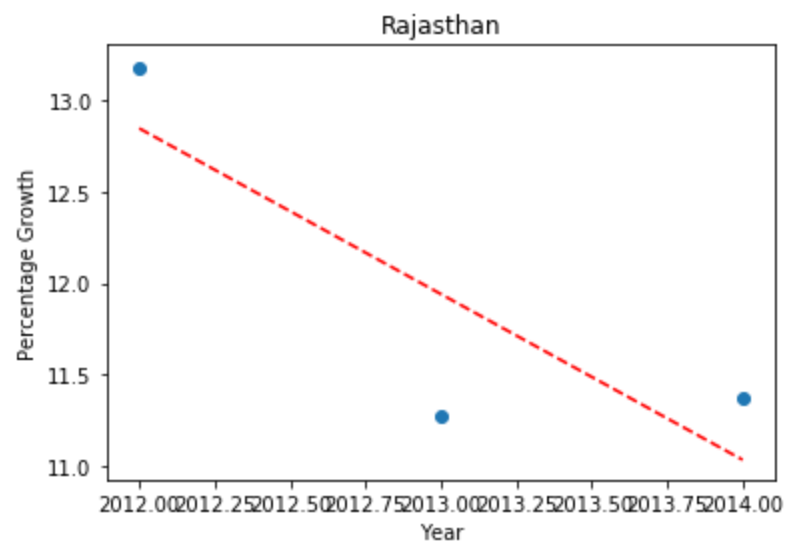


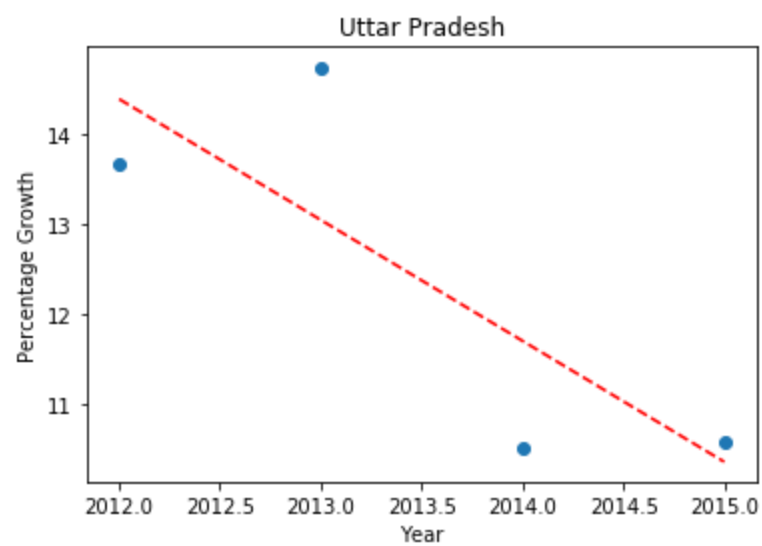
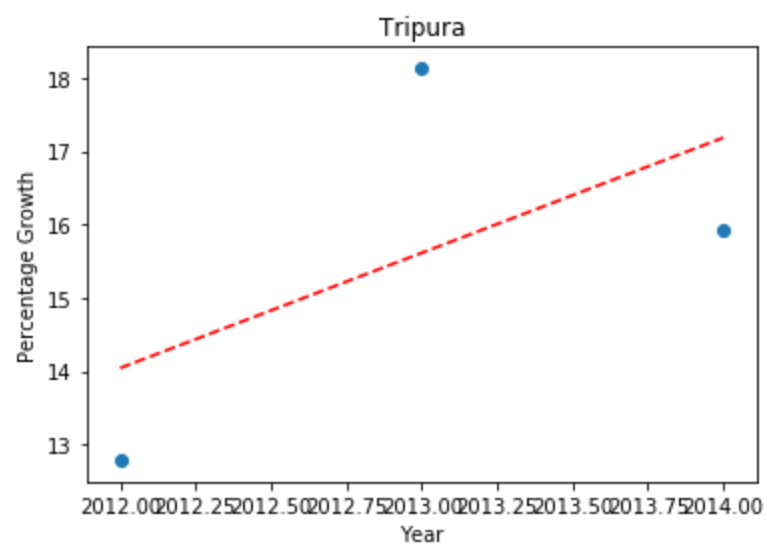
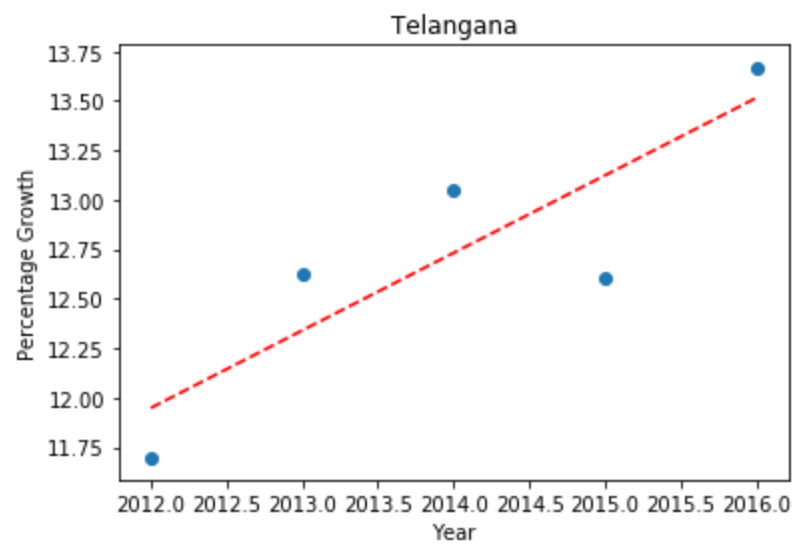


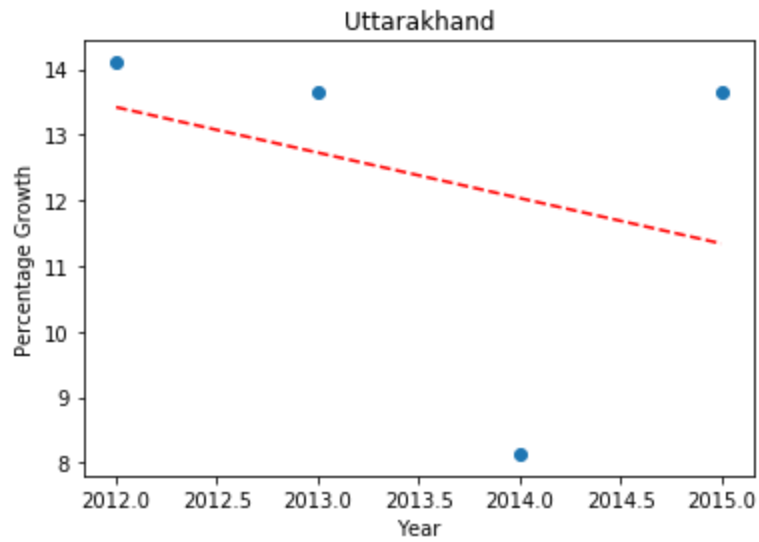








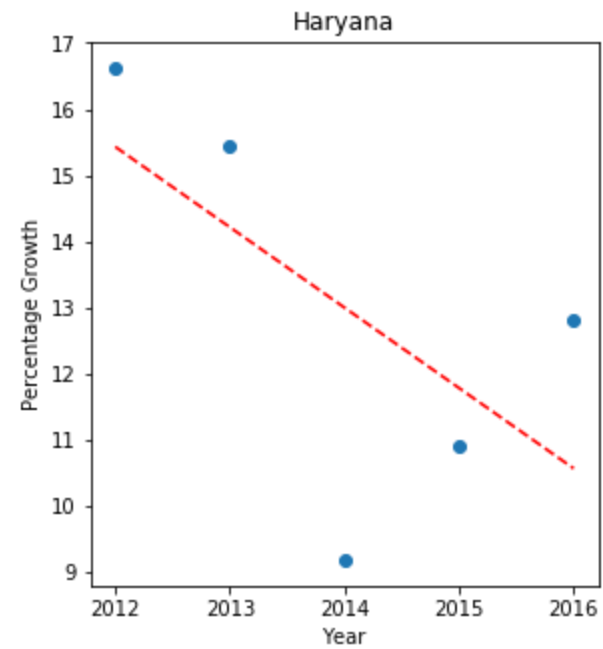
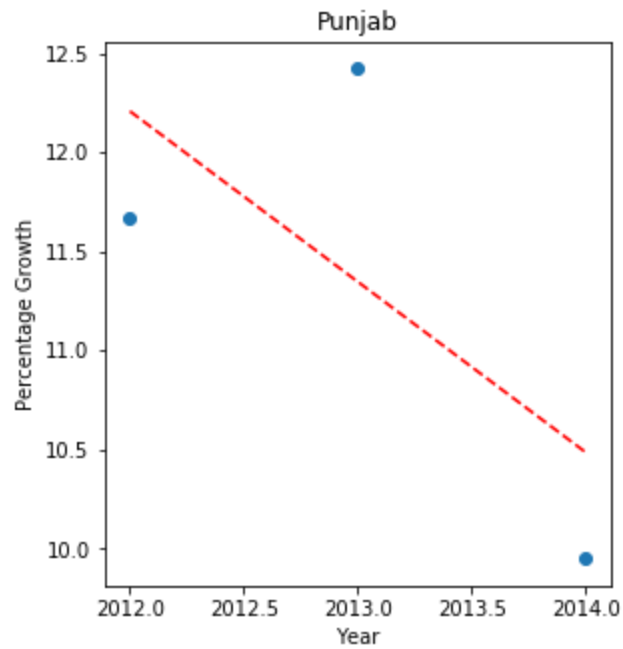




-
- Print the graph of All India Percentage growth
-



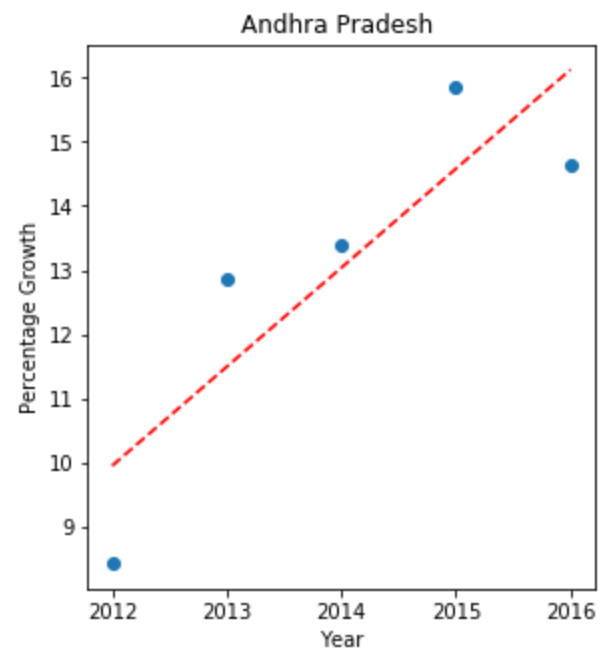
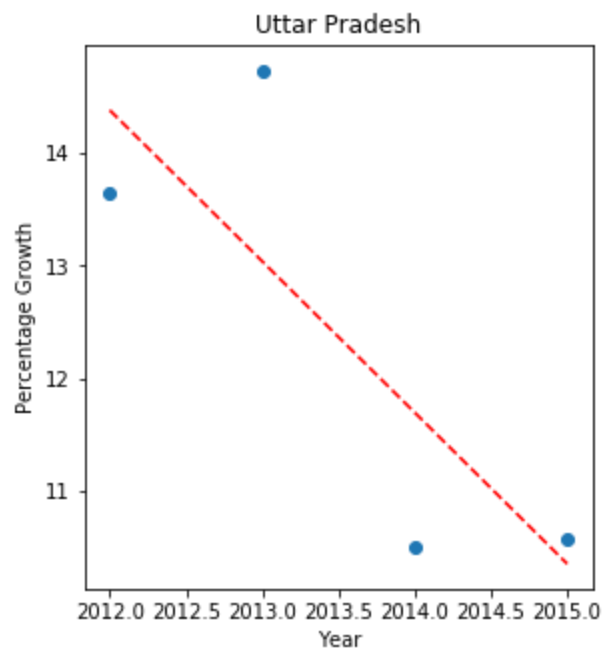
- **Comparing States:**
We can compare the states by looking at the slop:



Punjab: -0.86

Haryana: -1.22

Growth of Punjab is more than Haryana

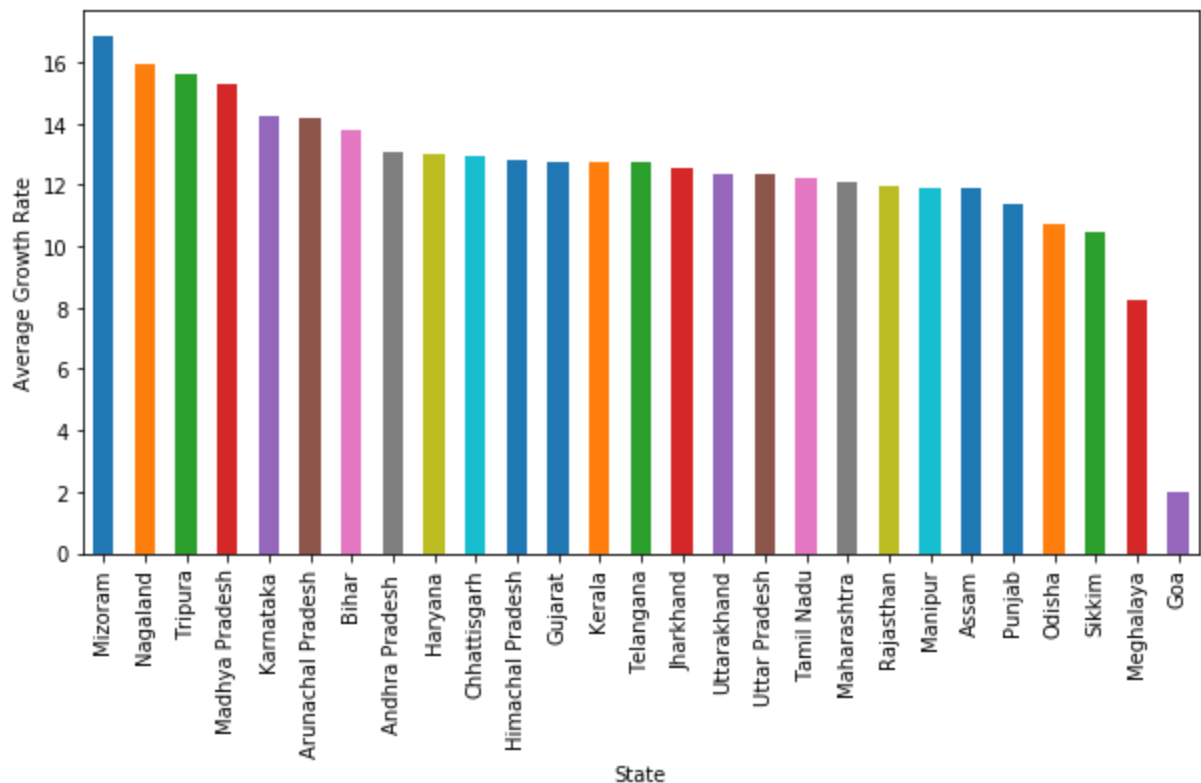


Uttar Pradesh: -1.34

Andhra Pradesh : 1.54

Growth of Andhra Pradesh is more than Uttar Pradesh

- Rank top 3 fastest and 3 slowest-growing states.
 - Average Growth Rate Chart



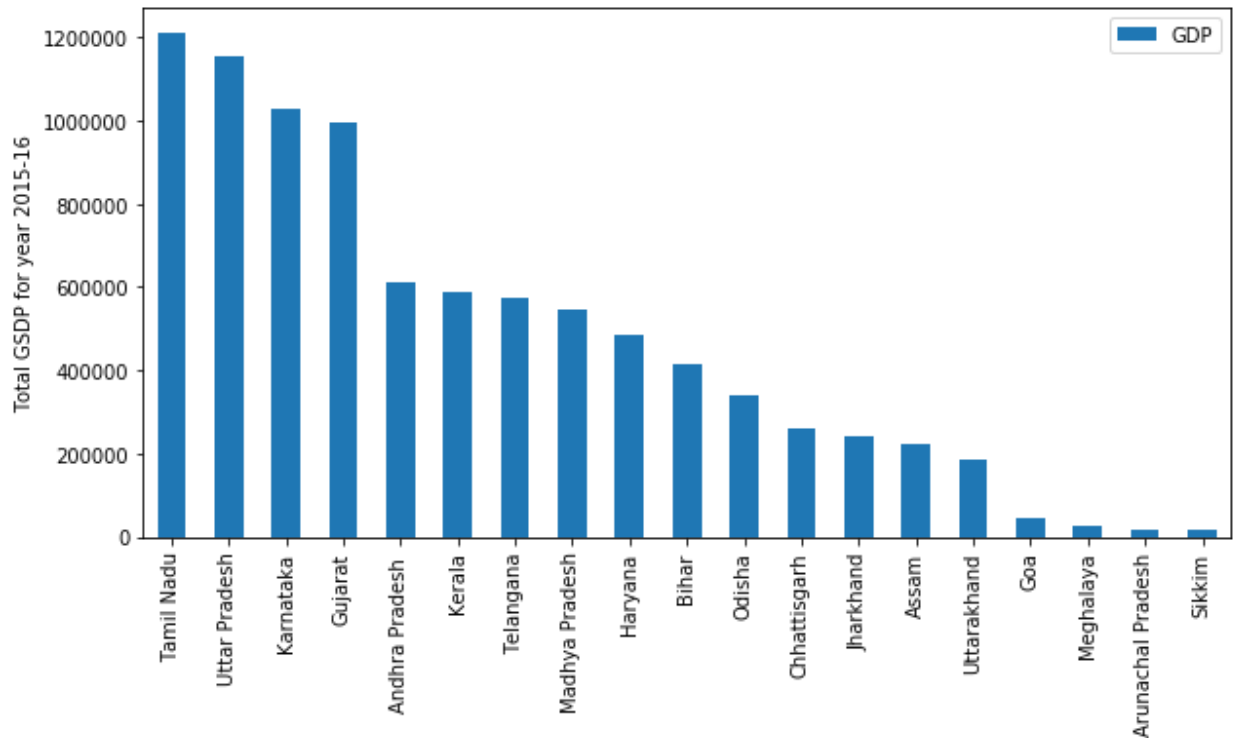
Fastest Growing states:

- Mizoram
- Nagaland
- Tripura

Slowest Growing states:

- Goa
- Meghalaya
- Sikkim

- Nation's growth rate is: 11.8% (Average Growth Rate)
- (Home State)Haryana's growth rate is: 13.0 %
 - Haryana's Average Growth is more than National Average Growth
- Plot the total GDP of the states for the year 2015-16:
 - I used the bar char for this with states shown in decreasing order of Total GDP(2015-16) because number of states are high and with a bar chart, we can easily look and compare which have higher value and which have lower



- Identify the top 5 and the bottom 5 states based on total GDP

Top 5 States:

- Tamil Nadu
- Uttar Pradesh
- Karnataka
- Gujarat
- Andhra Pradesh

Bottom 5 States:

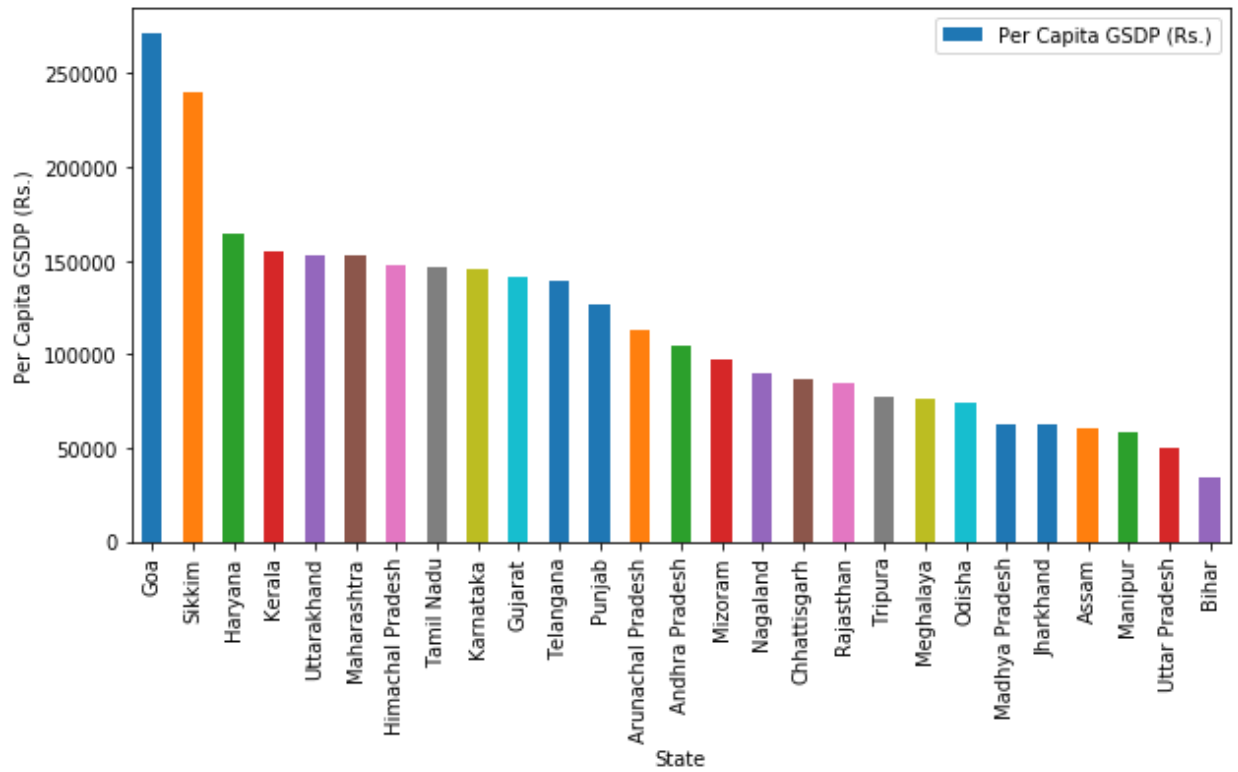
- Uttarakhand
- Goa
- Meghalaya
- Arunachal Pradesh
- Sikkim

- Insights
 - All the states with low GDP are smaller states

Data I-B:

- Import All the files using the glob module
- Include the state name in the data frame which is present in the file name
- Remove the trailing “*” present in the sector names in some files
- Keep only 2014-15 data because only that is required for further analysis

- Filter out the union territories
- Plot the GDP per capita for all the states.



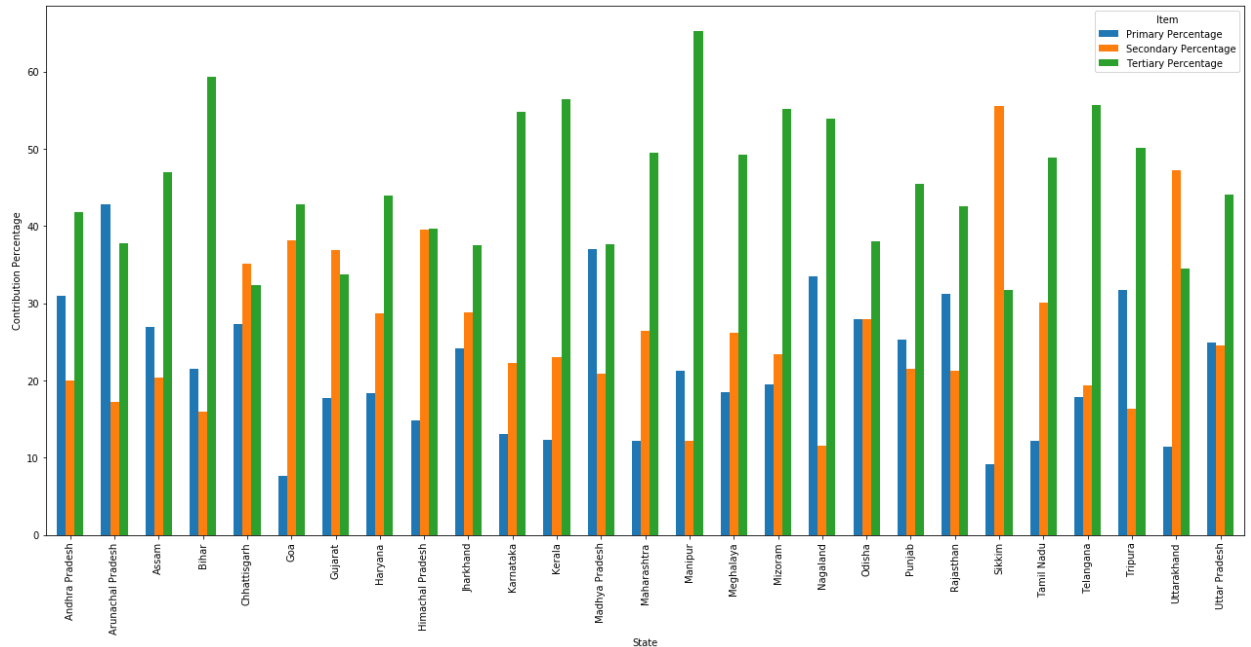
- Identify the top 5 and the bottom 5 states based on the GDP per capita
- Top 5 states based on the GDP per capita:

1. Goa
2. Sikkim
3. Haryana
4. Kerala
5. Uttarakhand

Bottom 5 states based on the GDP per capita

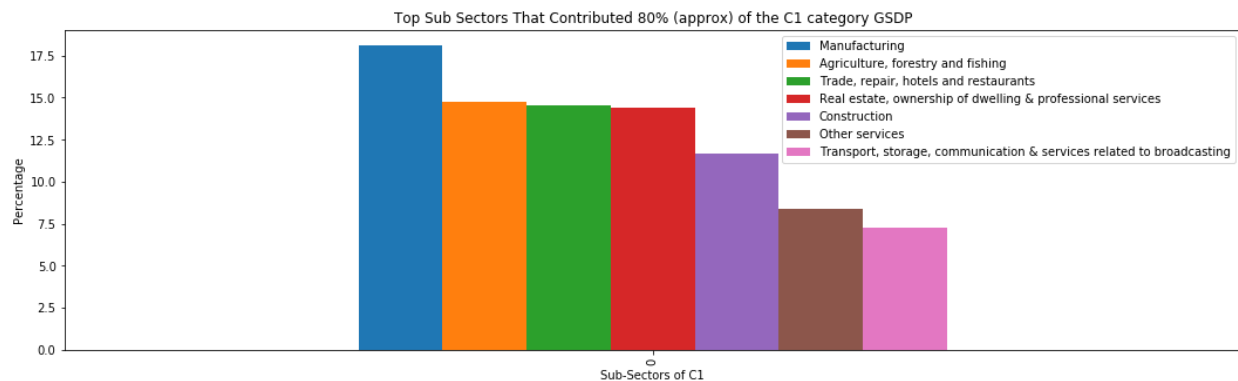
1. Bihar
2. Uttar Pradesh
3. Manipur
4. Assam
5. Jharkhand

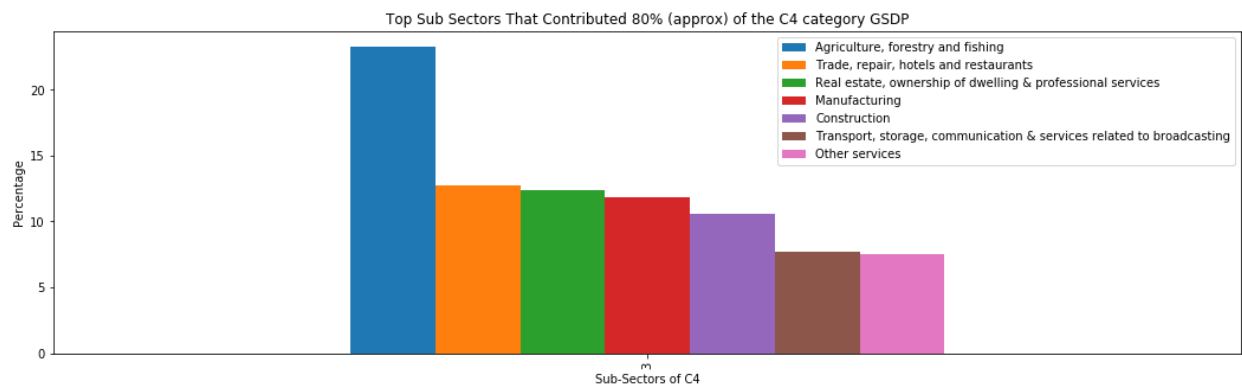
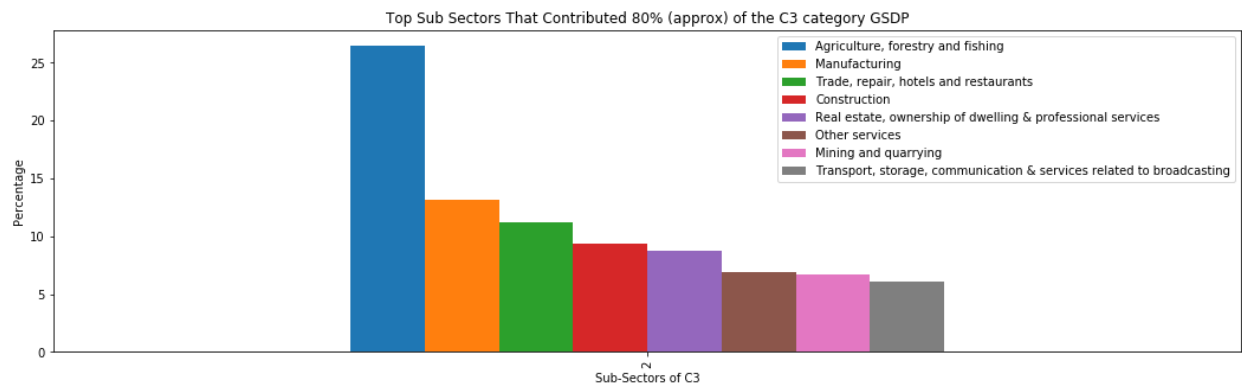
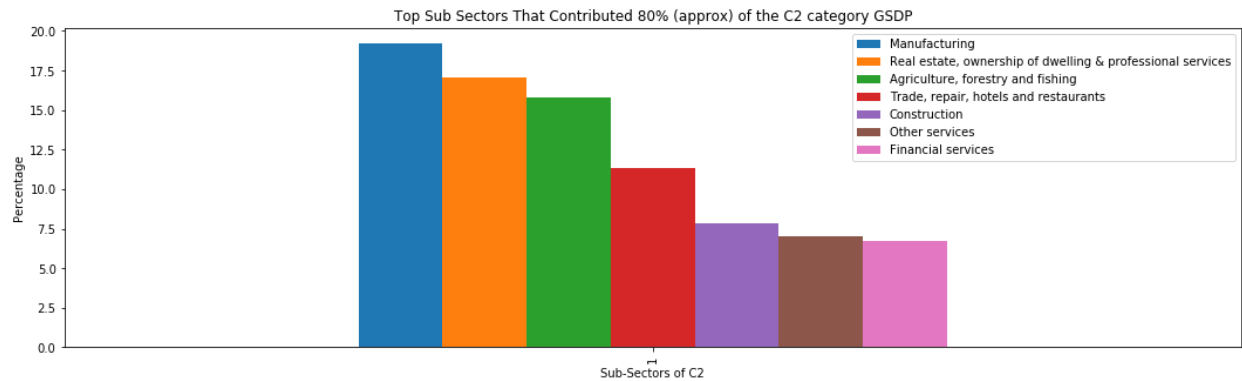
- The ratio of the highest per capita GDP to the lowest per capita GDP is 8.005.
- The percentage contribution of the primary, secondary and tertiary sectors as a percentage



- I used the Grouped Bar chart here because with this we can easily compare sectors within a state and can also compare easily compare the same sector of different states
- Primary + Secondary + Tertiary is not equal to total GDP because the total GDP also includes the Taxes on Products and excludes the Subsidies on products.
 - Total GDP = Primary + Secondary + Tertiary + Taxes on Products - Subsidies on products
- Insights:
 - For mostly all of the states Tertiary sector the major contributor
- Correlation of percentile of the state (% of states with lower per capita GDP) and %contribution of Primary sector to total GDP is: -0.62002
- Categorise the states into four groups based on the GDP per capita (C1, C2, C3, C4)
- Top 3/4/5 sub-sectors that contribute to approximately 80% of the GSDP of each category.

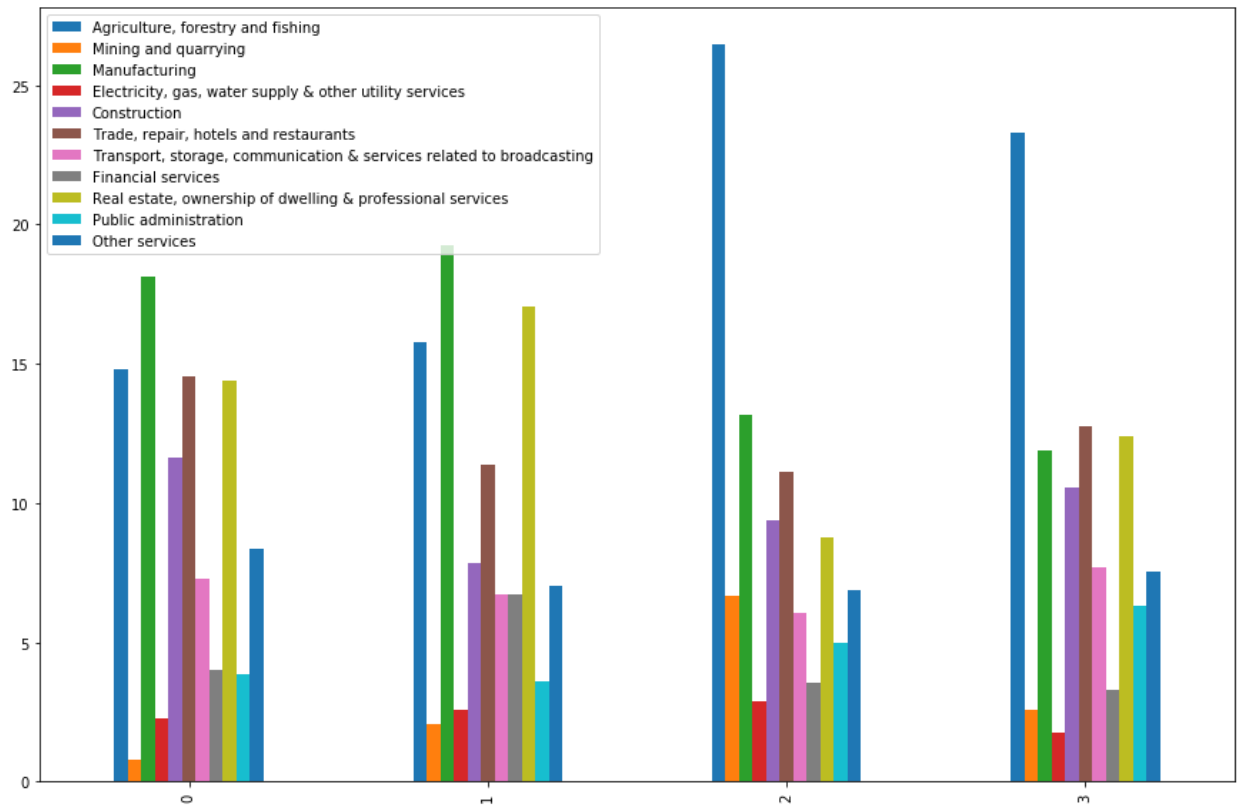
Figure size 1008x432 with 0 Axes>





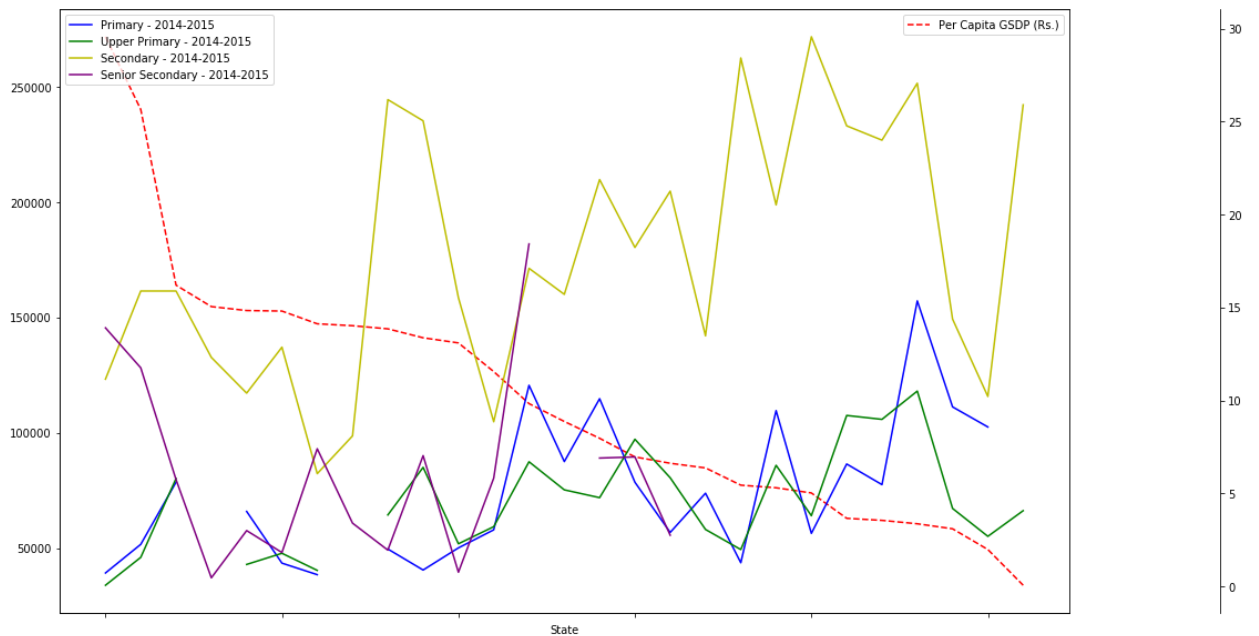
- For Top States Manufacturing is the main contributor whereas for bottom states agriculture is the main contributor
- Manufacturing seem to be correlated with high GDP
- Subsector Focus Area'
 - C1 Transport
 - C2 Financial Service
 - C3 Transport
 - C4 Other Services
- Recommendation:
 - C1 : Improve Transport and other service
 - C2: Improve Other and Final services

- C3: Sector difference is very high so try to neutralize that and improve transport
- C4: Sector difference is very high so try to neutralize that and improve Other services
- Contribution of the sub-sectors as a percentage of the GSDP of each category.



Part-II: GDP and Education Dropout Rates

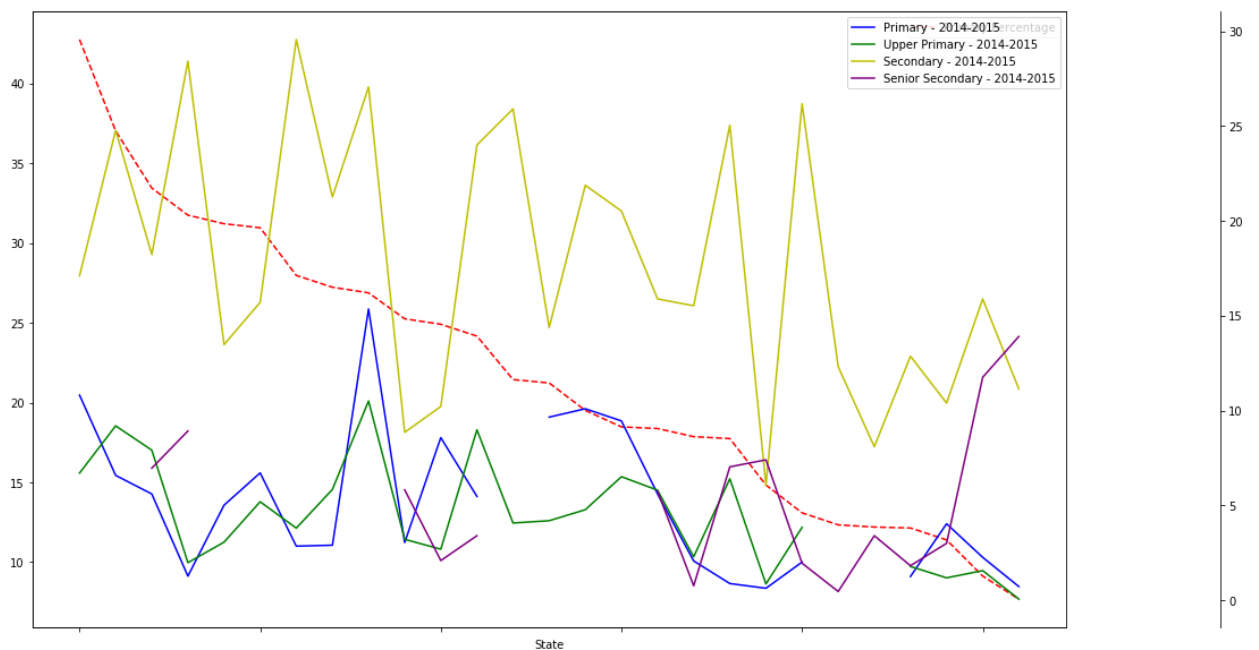
- Load Dropout data
- Fix name of Chhattisgarh and Uttarakhand in Dropouts dataframe because that will cause issue in merge
- GDP per capita with dropout rates in education (primary, upper primary and secondary) for the year 2014-2015 for each state



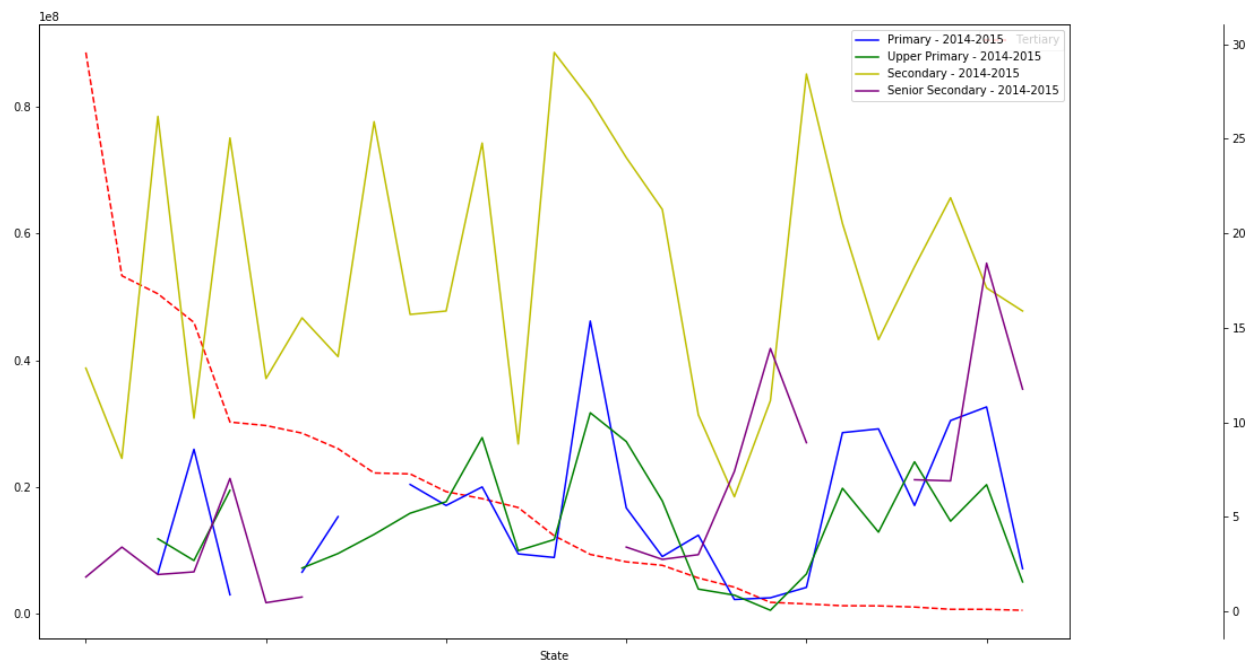
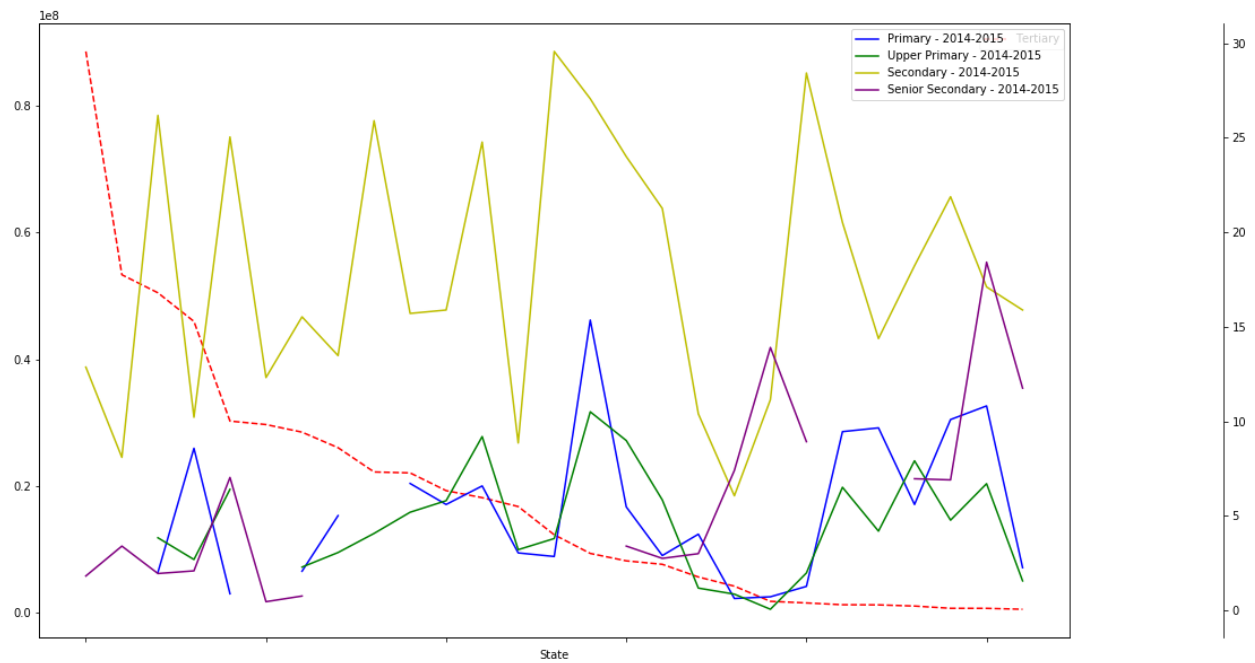
You can see as the Per Capita GDP decreases

- The dropout rate in secondary education level increases very much (High Correlation)
- For senior secondary level there is slight decreasing trend but there seems very low correlation
- For Primary and Upper Primary level there is increases in dropout by small amount

- dropout rate and %contribution of each sector (Primary, Secondary and Tertiary) to the total GDP

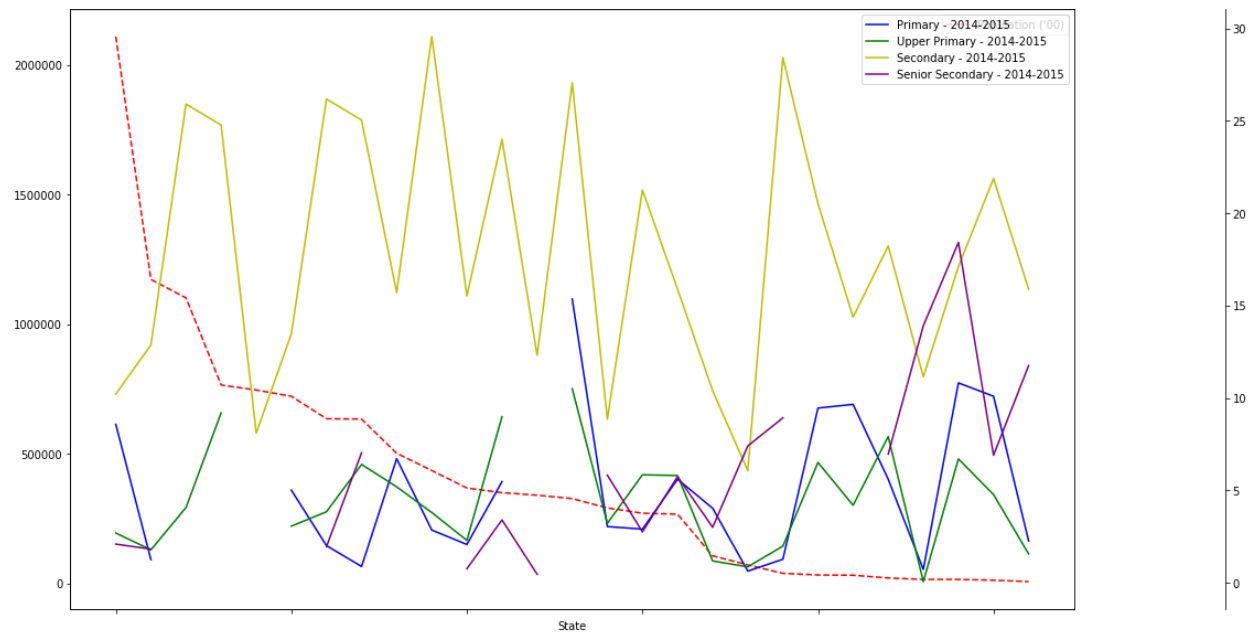


- As Primary sector percentage increases, Senior secondary dropout increases



As Tertiary sector percentage decreases, Senior secondary dropout increases
 Insight: Secondary and Tertiary required educated people

- Population graph with dropout rate



Expected High Correlation between them
 Observed: No/little correlation

Hypothesis:

- There is a steady increase in GSDP by the decrease in the dropout rate
- The states where agriculture is major contributor have high senior secondary drop rate and for manufacturing reverse is true