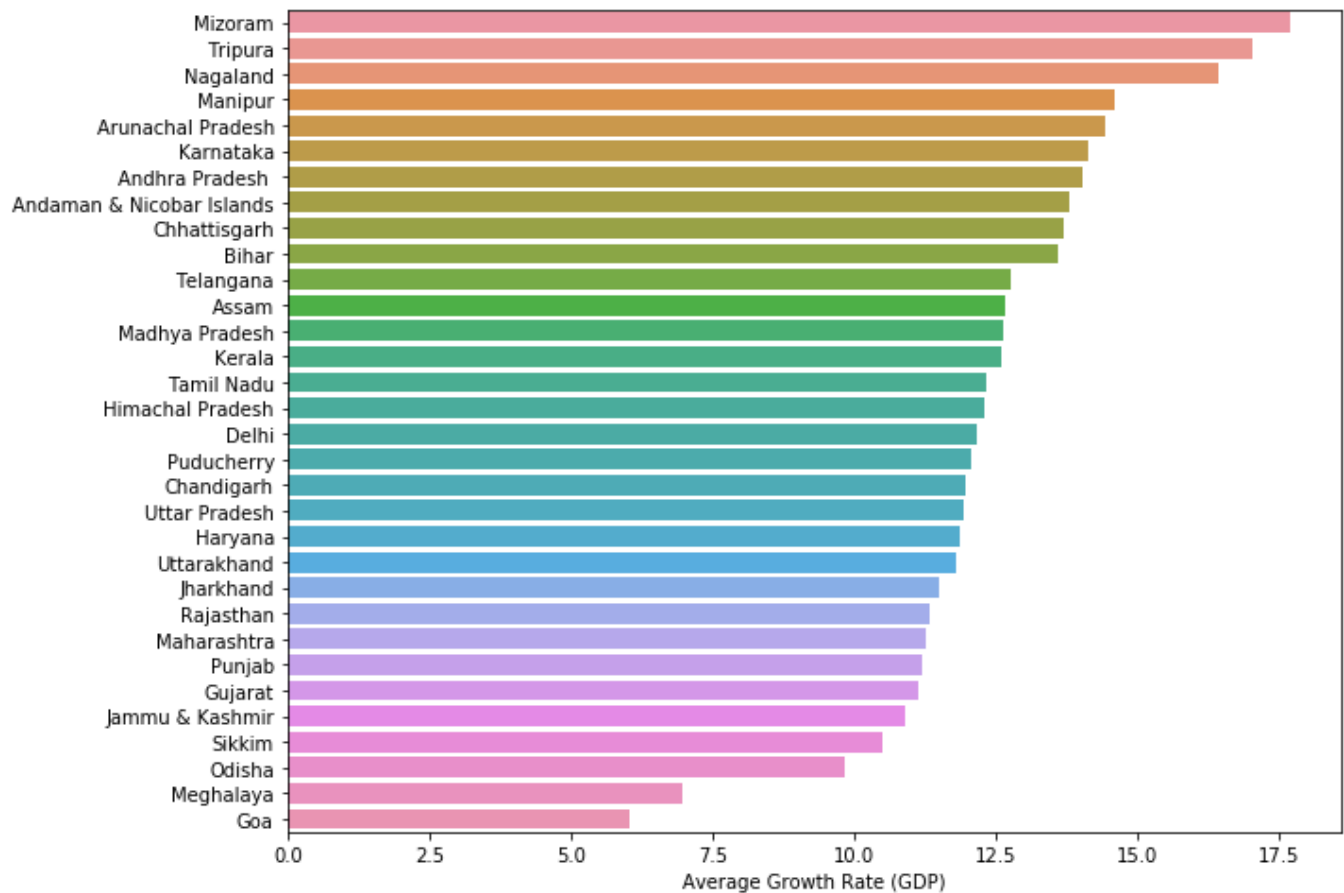


Summary of the Analysis –

EDA of GDP and GDP per capita:

Plot 1 - The average growth rate of the following states (taking the years 2012-2013, 2013-2014, 2014-2015 and 2015-2016 into account) is as follows –

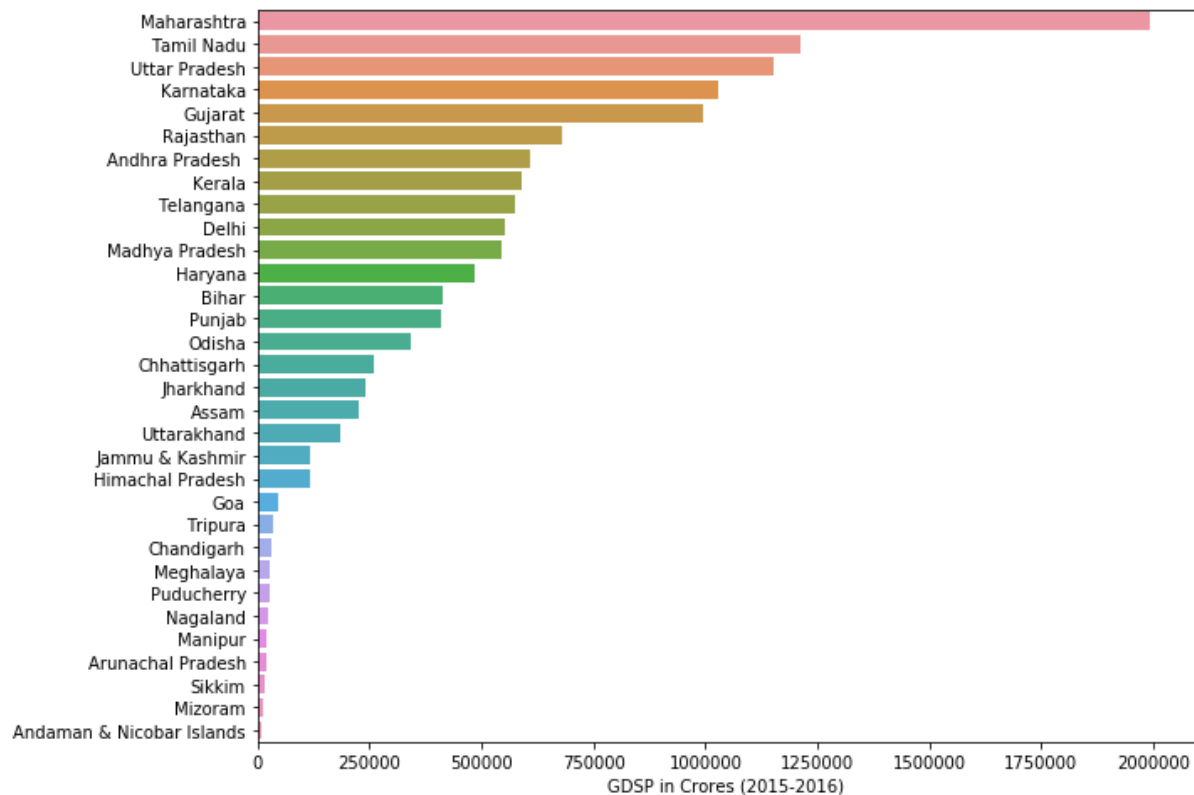


- Looking at the above figure it's apparent that states in the North-east region of India have been consistently growing fast. In fact, the top 5 states with the highest average growth rates over the aforementioned years are all from North-East India, these are – Mizoram, Tripura, Nagaland, Manipur and Arunachal Pradesh.
- It's interesting to note that 4 (Mizoram, Tripura, Nagaland and Manipur) of the top 5 states belong to the so called 7 sister states of India.
- The states with the lowest average GDP growth rate include – Goa, Meghalaya, Odisha, Sikkim and Jammu & Kashmir.

My home-state of Maharashtra has an average GDP growth rate of 11.26% which is slightly below the average All-India GDP of 11.8675% (obtained by taking the average of 'All_India GDP' column) but it's also way below the median GDP growth rate of the states i.e. 12.22%.

It's important to note that the bigger states (the one's which contribute more GDP than others) will have more of an influence on the All India GDP growth rate.

Plot 2 - The total GDP of the states for the year 2015-16 –



The top 5 states based on the total GDP for the year 2015-2016 (in descending order) are –

1. Maharashtra
2. Tamil Nadu
3. Uttar Pradesh
4. Karnataka
5. Gujarat

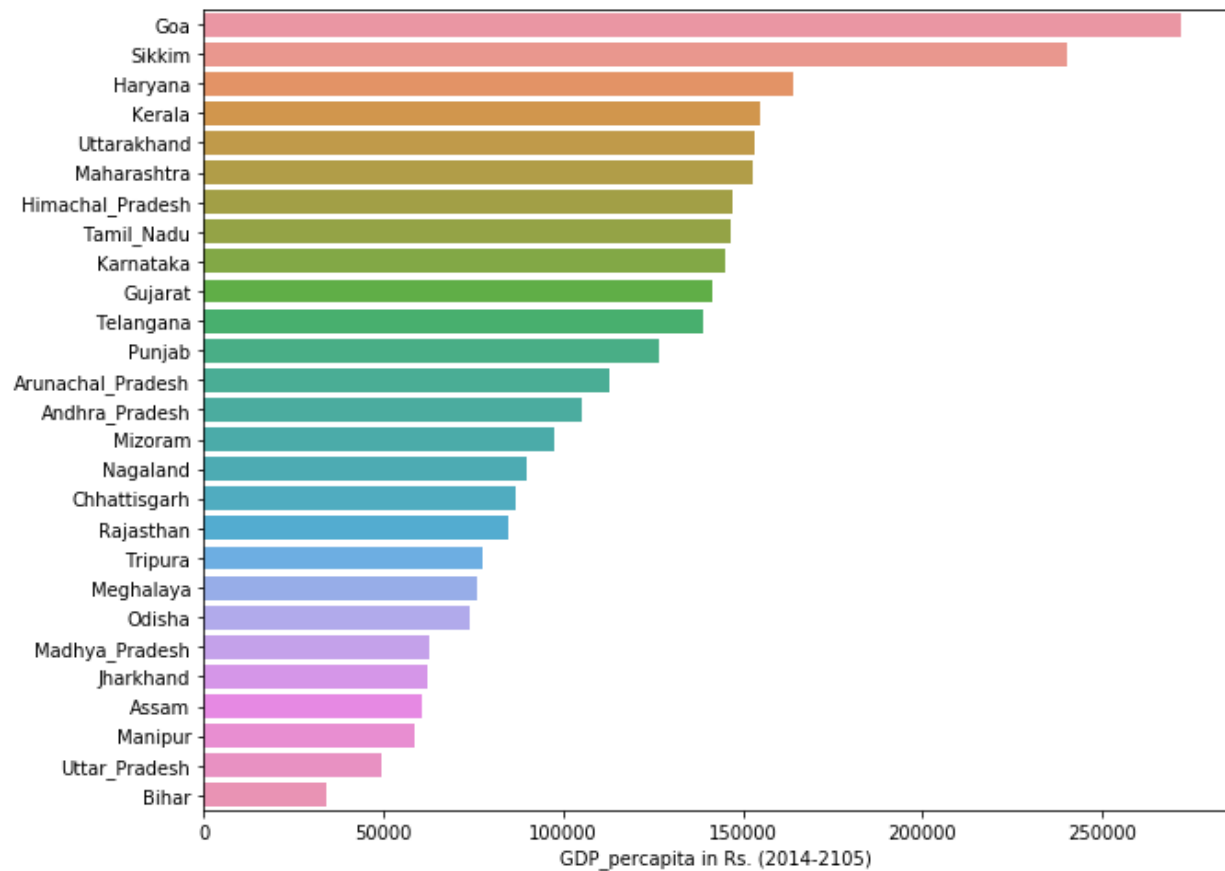
The bottom 5 states based on the total GDP for the year 2015-2016 (in descending order) are –

1. Manipur
2. Arunachal Pradesh
3. Sikkim
4. Mizoram
5. Andaman & Nicobar Islands

It's interesting to see that few of the states with the most Avg. Growth Rate (GDP) (Manipur, Arunachal Pradesh and Mizoram) are actually the ones which have the lowest total GDP.

PART 1-B:

Plot 3 - GDP per capita for all the states



Top 5 states with GDP per capita (in descending order) =

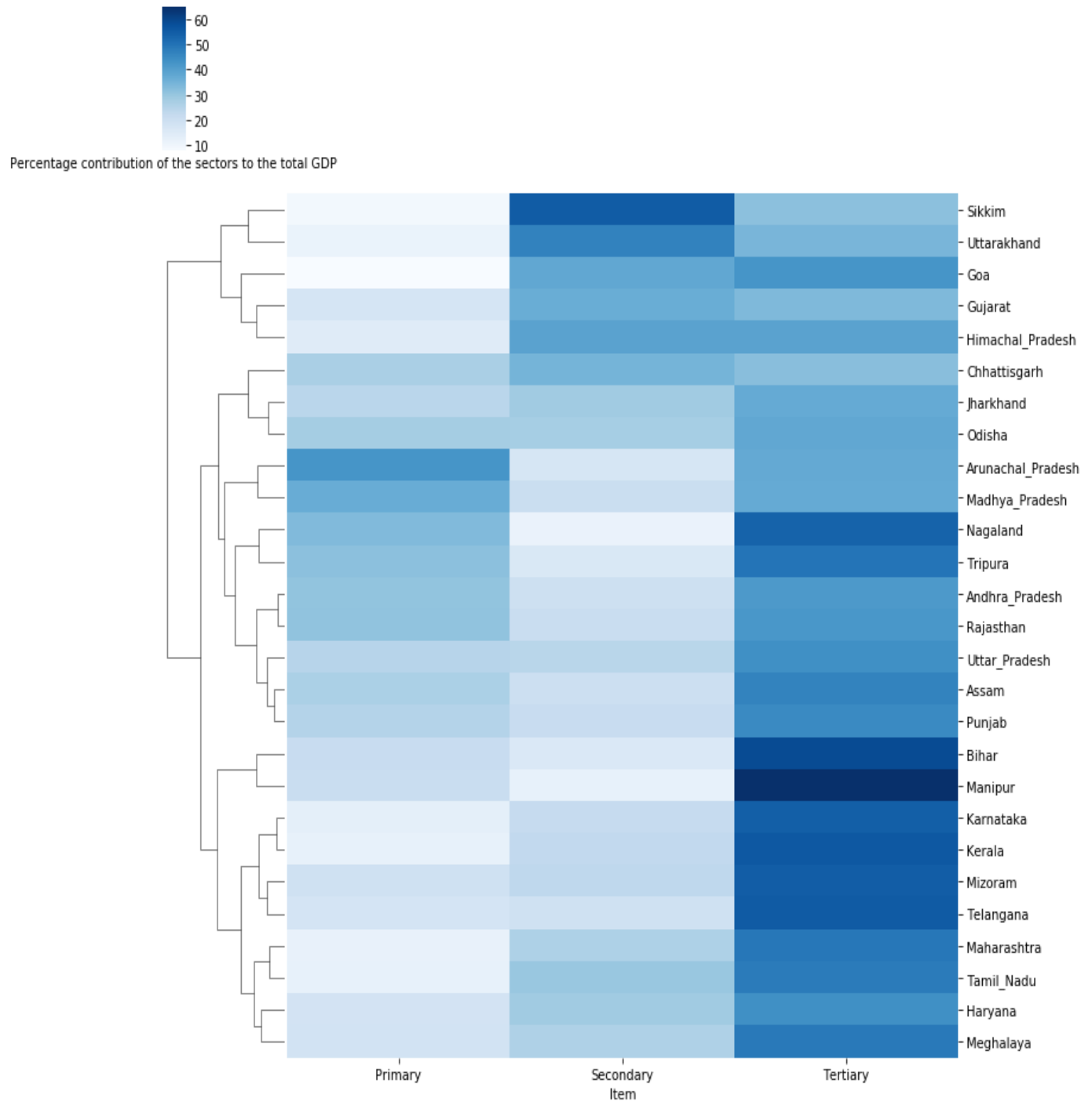
1. Goa
2. Sikkim
3. Haryana
4. Kerala
5. Uttarakhand

Bottom 5 states with GDP per capita (in descending order) =

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

Ratio of the state with the highest per capita GDP to the lowest per capita GDP is = 8.004.

Plot 4 - The percentage contribution of primary, secondary and tertiary sectors as a percentage of total GDP for all the states –



Used a hierarchically clustered heatmap (along the columns) to see if there are any emerging trends in the data.

- It has broadly formed three clusters; the first cluster has states which have more contribution from secondary and tertiary sectors (the first 5 states of the above plot).

- The second cluster has a lot of subdivisions but mostly contains states with high tertiary and primary sectors.
- The third cluster has states whose GDSP is most contributed by the tertiary sector (the last 10 states).

Have created categories C1, C2, C3 and C4 for the states based on their GDP per capita where the GDP per capita has been divided by quantile values 0.20, 0.5, 0.85 and 1.

The sub sectors contributing to ~80% of each of the user created categories (C1, C2, C3 and C4) are as follows:

Sub-sectors of C1 -

1. 'Real estate, ownership of dwelling & professional services'
2. 'Agriculture, forestry and fishing',
3. 'Trade, repair, hotels and restaurants'
4. 'Manufacturing'
5. 'Construction'
6. 'Taxes on Products'

Sub-sectors of C2 -

1. 'Manufacturing'
2. 'Real estate, ownership of dwelling & professional services'
3. 'Agriculture, forestry and fishing'
4. 'Taxes on Products'
5. 'Trade, repair, hotels and restaurants'
6. 'Construction'
7. 'Financial services'

Sub-sectors of C3 -

1. 'Agriculture, forestry and fishing'
2. 'Manufacturing'
3. 'Trade, repair, hotels and restaurants'
4. 'Taxes on Products'
5. 'Real estate, ownership of dwelling & professional services'
6. 'Construction'
7. 'Other services'

Sub-sectors of C4 -

1. 'Agriculture, forestry and fishing'
2. 'Trade, repair, hotels and restaurants'
3. 'Manufacturing'
4. 'Real estate, ownership of dwelling & professional services'
5. 'Construction'
6. 'Taxes on Products'
7. 'Transport, storage, communication & services related to broadcasting'

Plot 5 - The contribution of the sub-sectors as a percentage of the GSDP of each category.



1- The primary sector contributes the least (less than 15%) in top states (C1). The secondary and tertiary sector account for most of the GDP in C1 states.

2- Agriculture, forestry and fishing, Manufacturing and Real estate services are the sub-sectors which seem to be highly correlated with the GDP.

3- It seems logical to focus on the sub-sectors in the tertiary and secondary sectors which are highly correlated with GDP.

The breakdown of GDP within the sub-sectors (i.e. the sub-sub sectors) is also uneven. A lot of these sectors seem to be ignored in favor of others. Investing in particular sub-sub sectors which show promise (i.e. have been contributing to high GDP in other states) can also be a good idea.

Recommendations for each category -

C1 - Manufacturing shows up much higher in the other categories. So it can be further invested upon to improve the GDP per capita. Also, manufacturing more products will bring more revenue from 'Taxes on Products'.

C2- Investing in the sub-sector 'Transport, storage, communication & services related to broadcasting' specifically on the various modes of transport will also improve the overall GDP coming from 'Trade, repair, hotels and restaurants'.

C3 and C4 - It seems that most of the GDP coming from these 2 categories is from 'Agriculture, forestry and fishing'. The overall distribution of GDP is uneven and if the Crops fail due to bad weather for a particular year the overall GDP will be significantly affected. Hence, it's wise to shift some of the other well performing sub-sectors like 'Trade, repair, hotels and restaurants' and 'Manufacturing'.

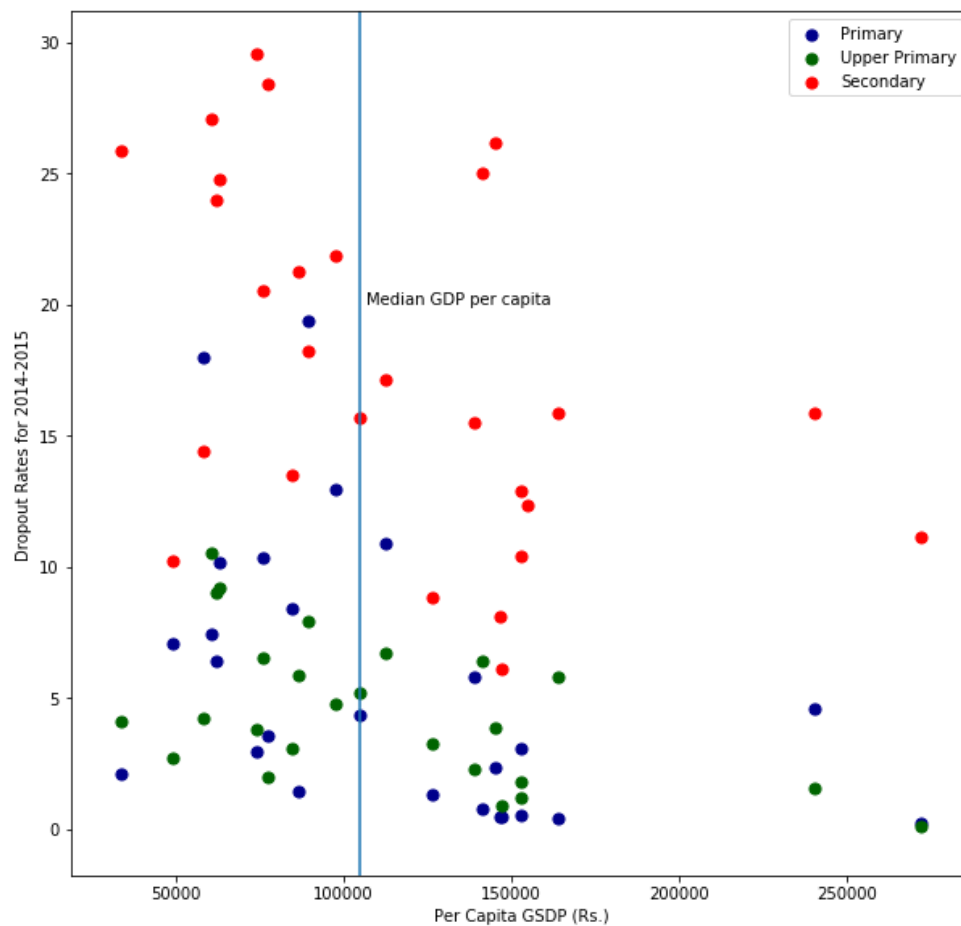
It's also important to note that investing in one sub-sector will definitely have an effect (positive or negative) on another sub-sector. These relationships need to be identified before making any sudden changes.

Checking for any relation between the GDP per capita and Education:

Investigating if there is any correlation between GDP per capita and dropout rates in education (primary, upper primary and secondary) for the year 2014-2015.

Want to see if there is any correlation between the 2 variables, I've made a scatterplot with the median GDP per capita as a vertical line to divide the plot into 2.

Plot 6 –



Observations and hypothesis –

It can be seen that in most of the states the Primary and Upper Primary dropout rates are quite low but they do not seem to affect the GDP per capita. In fact, most of the states with low primary and secondary dropout rates have a low GDP (below median) and a considerable number of states with very low Primary and Upper Primary dropout rates have a GDP per capita above the median.

Now looking at the Secondary dropout rates, it looks like they have a slightly negative correlation with the GDP per capita.

Most of the states with high Secondary dropout rates have their GDP per capita below the median. Most of the states with Secondary dropout rates that are relatively low (below 20%) have their GDP per capita above the median.

Obviously, outliers exist in both these cases but nevertheless a general trend is observed w.r.t the Secondary dropout rates.

Hypothesis – The average dropout rates of primary and Upper primary education are low and it can be assumed the people that are actually completing this level of education will not actively contribute to increasing the GDP of a state since they will not learn a bankable skill at that level. Hence these two do not seem to have any significant correlation with the GDP.

Similarly, we can further assume that the people who have completed secondary education are most likely to complete further education and contribute to increasing GDP. Hence, a high dropout rate in secondary education has a negative correlation with the GDP per capita.