Name: Dibyalok Bhutiarai

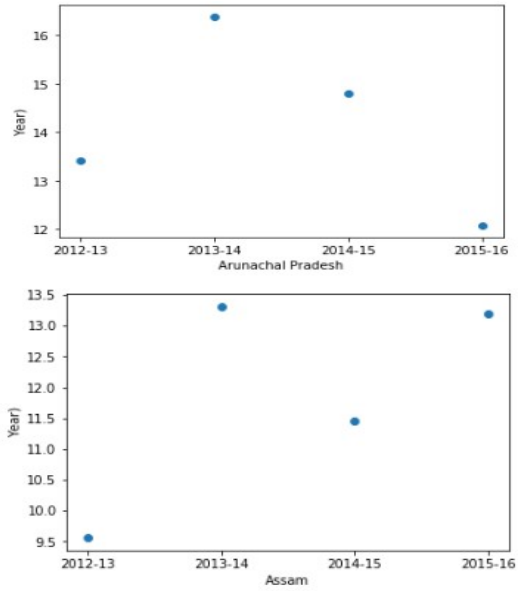
Part I-A:

- For the analysis below, use the **Data I-A**.
- First, you need to load the data in Python properly and then clean it. This also involves the treatment of missing values, you can choose to drop the row or column as well. Remember this will affect your next analysis and results drastically.
- Plot a graph for rows " % Growth over previous year" for all the states (not union territories) whose data is available, use as much data as possible for this exercise. Use the best fit line to represent the growth for each state. Draw a similar line graph for the nation as well.
    - How will you compare the growth rates of any two states?
    - Which states have been growing consistently fast, and which ones have been struggling? Rank top 3 fastest and 3 slowest-growing states.
    - What is the Nation's growth rate?
    - What has been the growth rate of your home state, and how does it compare to the national growth rate?
    -
- Plot the total GDP of the states for the year 2015-16:

    - Which Plot will you use for this? Why? (Remeber to plot the graph in a way such as it is easier to read and compare)
    - Identify the top 5 and the bottom 5 states based on **total** GDP.
    - What insights can you draw from this graph? What states are performing poorly? (Remember: this will not be solely based on total GDP)

ANALYSIS:

- **How will you compare the growth rates of any two states?**
    - ✓ Lets us compare two states: Arunachal Pradesh and Assam



The growth rate for Assam was almost 1.4 less than Arunachal Pradesh in 2012-2013. But Assam saw a significant gain in 2013-2014 and both of the state gained in the % growth in 2013-2014 and declined in 2014-2015. However the growth for Assam has again picked up in 2015-2016 whereas the growth for Arunachal Pradesh has further declined.

- **Which states have been growing consistently fast, and which ones have been struggling? Rank top 3 fastest and 3 slowest-growing states.**

Assumption: We can compare the growth rate from 2012-2013 to 2016-2017. But for 2016-2017 for around 20 States the Growth Rate is missing. So we can also compare the growth rate from 2012-2013 to 2015-2016.

2012-2013 to 2016-2017:

Top 3 Growing: 1. Andhra Pradesh, 2. Telengana, 3. Meghalaya

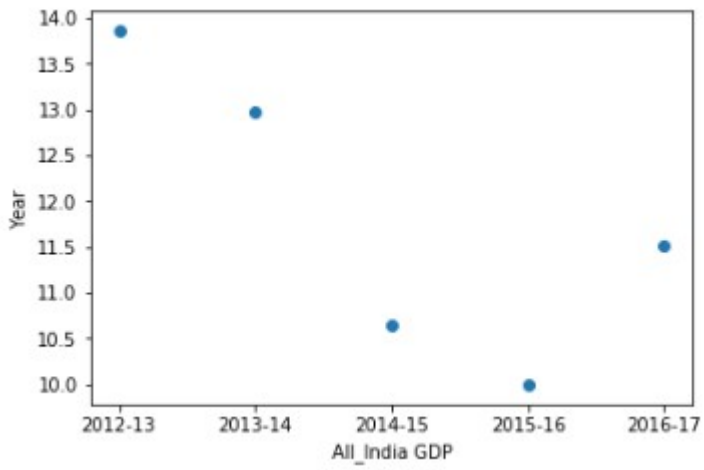
Bottom 3 Growing: 1.Tamil Nadu, 2. Haryana, 3. Odisha

2012-2013 to 2015-2016:

Top 3 Growing: 1. Goa, 2. Andhra Pradesh, 3. Assam

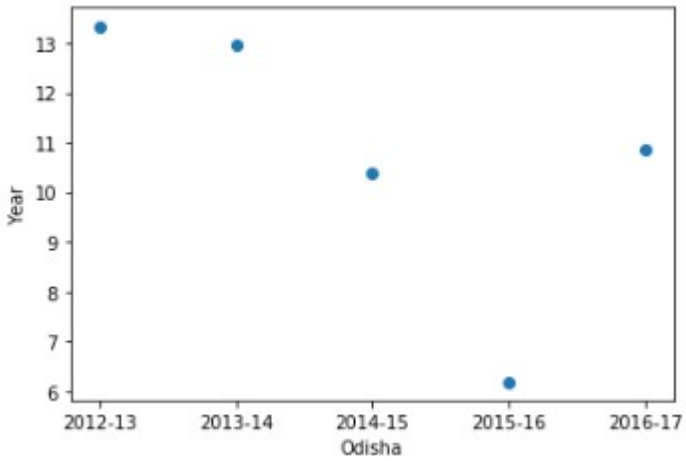
Bottom 3 Growing: 1. Odisha, 2. Madhya Pradesh, 3. Gujarat

- **What is the Nation's growth rate?**



The All-India GDP has shown a decline from 2012-2013 to 2015-2016. This is a decline of 28%. From 2015-2016 it has increased by 15% to 2016-17. This increase is inconclusive as data for 20 states are missing.

- **What has been the growth rate of your home state, and how does it compare to the national growth rate?**



My Home State is Odisha. If we compare the growth rate from 2012-2013 to 2016-2017 there has been a dip of 19%. Every Year from 2012-2013 till 2015-2016 there has been a dip in the growth rate as compared to the previous year.

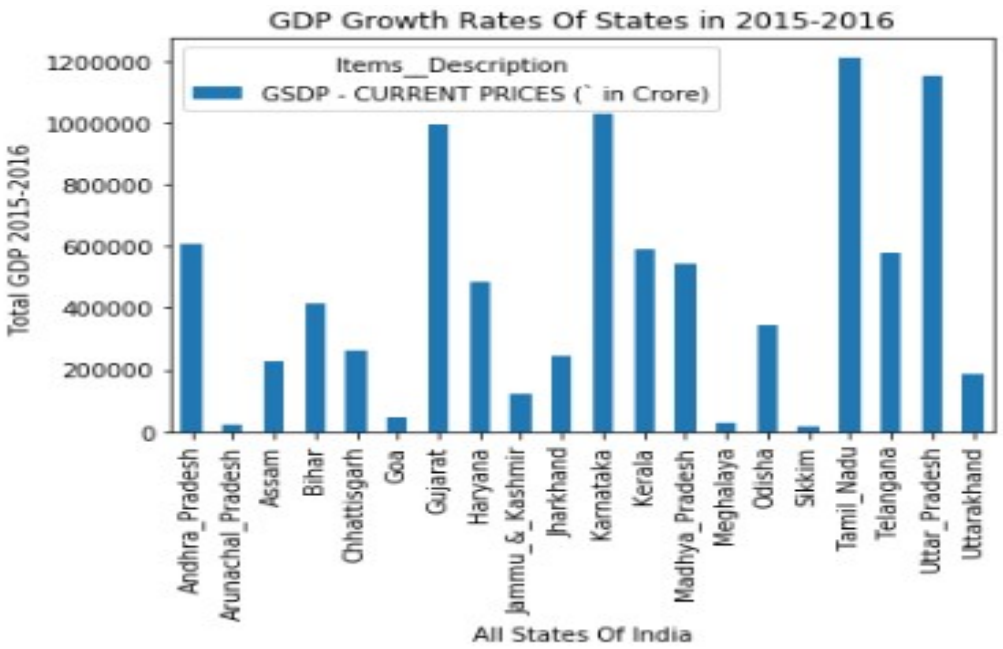
In the last year i.e. from 2015-2016 to 2016-2017, the growth rate has increased by almost 75%.

- **Plot the total GDP of the states for the year 2015-16:**

    - **Which Plot will you use for this? Why? (Remember to plot the graph in a way such as it is easier to read and compare)**
    - **Identify the top 5 and the bottom 5 states based on** total **GDP.**
    - **What insights can you draw from this graph? What states are performing poorly? (Remember: this will not be solely based on total GDP)**

✓ We can use a Bar Chart to compare the GDP of the year 2015-16.

✓



- ✓ **The top 5 States: 1. Tamil Nadu, 2. Uttar Pradesh, 3. Karnataka, 4. Gujarat, 5. Andhra Pradesh.**

- ✓ **Bottom 5 States: 1. Sikkim, 2. Arunachal Pradesh 3. Meghalaya, 4. Goa, 5. Jammu & Kashmir**

- ✓ **Insights we draw:**

    - **Bigger States have bigger GSDP i.e. Total GDP is directly proportional to the size of the State.**

Part I-B:

- For the analysis below, use **Data I-B**. You can also use **Data I-B** along with **Data I-A** if required. Also, perform the analysis only for the duration 2014-15.
- Filter out the union territories (Delhi, Chandigarh, Andaman and Nicobar Islands, etc.) for further analysis, as they are governed directly by the central, not state governments.
- Plot the GDP per capita for all the states.
    - Identify the top 5 and the bottom 5 states based on the GDP per capita.
    - Find the ratio of the highest per capita GDP to the lowest per capita GDP.

- Plot the percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states.
    - Which plot will you use here? Why?
    - Why is (Primary + Secondary + Tertiary) not equal to total GDP?
    - Can you draw any insight from this? Find correlation of percentile of the state (% of states with lower per capita GDP) and %contribution of Primary sector to total GDP.
- Categorise the states into four groups based on the GDP per capita (C1, C2, C3, C4, where C1 would have the highest per capita GDP and C4, the lowest). The quantile values are (0.20,0.5, 0.85, 1), i.e., the states lying between the 85th and the 100th percentile are in C1; those between the 50th and the 85th percentiles are in C2, and so on.
    - **Note:** Categorisation into four groups will simplify the subsequent analysis, as otherwise, comparing the data of all the states would become quite exhaustive.
- For each category (C1, C2, C3, C4):
    - Find the top 3/4/5 **sub-sectors (**such as agriculture, forestry and fishing, crops, manufacturing etc., not primary, secondary and tertiary) that contribute to approximately 80% of the GSDP of each category.
        - **Note-I:** The nomenclature for this project is as follows: primary, secondary and tertiary are named 'sectors', while agriculture, manufacturing etc. are named 'sub-sectors'.
        - **Note-II**: If the top 3 sub-sectors contribute to, say, 79% of the GDP of some category, you can report "These top 3 sub-sectors contribute to approximately 80% of the GDP". This is to simplify the analysis and make the results consumable. (Remember, the CEO has to present the report to the CMs, and CMs have limited time; so, the analysis needs to be sharp and concise.)
    - Plot the contribution of the sub-sectors as a percentage of the GSDP of each category.

Now that you have summarised the data in the form of plots, tables, etc., try to draw non-obvious insights from it. Think about questions such as:

- How does the GDP distribution of the top states (C1) differ from the others?
- Which sub-sectors seem to be correlated with high GDP?
- Which sub-sectors do the various categories need to focus on?

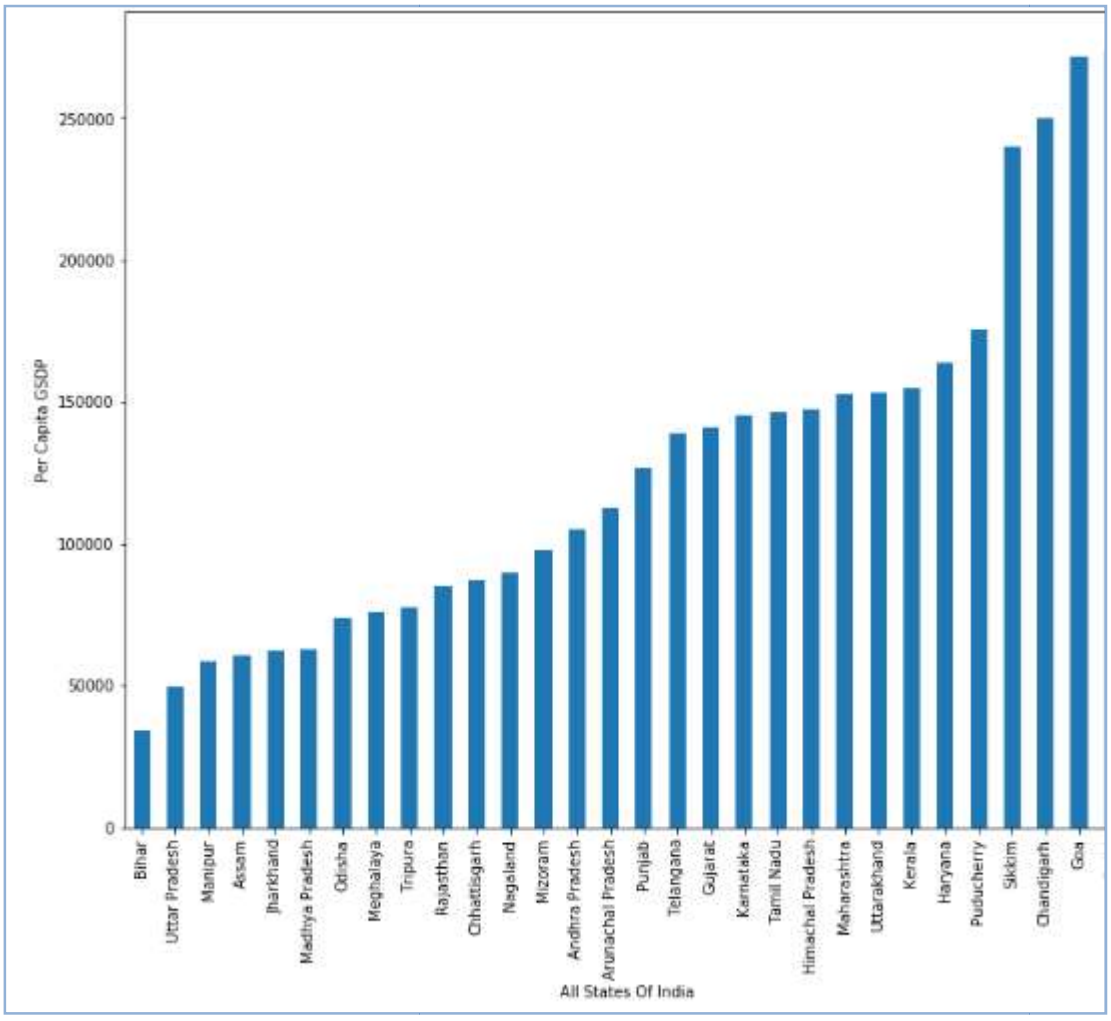
Ask other such relevant questions, which you think are important, and note your insights for category separately. More insights are welcome and will be awarded accordingly.

- Finally, provide at least two recommendations for each category to improve the per capita GDP.

Analysis:

- Plot the GDP per capita for all the states.
    - Identify the top 5 and the bottom 5 states based on the GDP per capita.

Find the ratio of the highest per capita GDP to the lowest per capita GDP



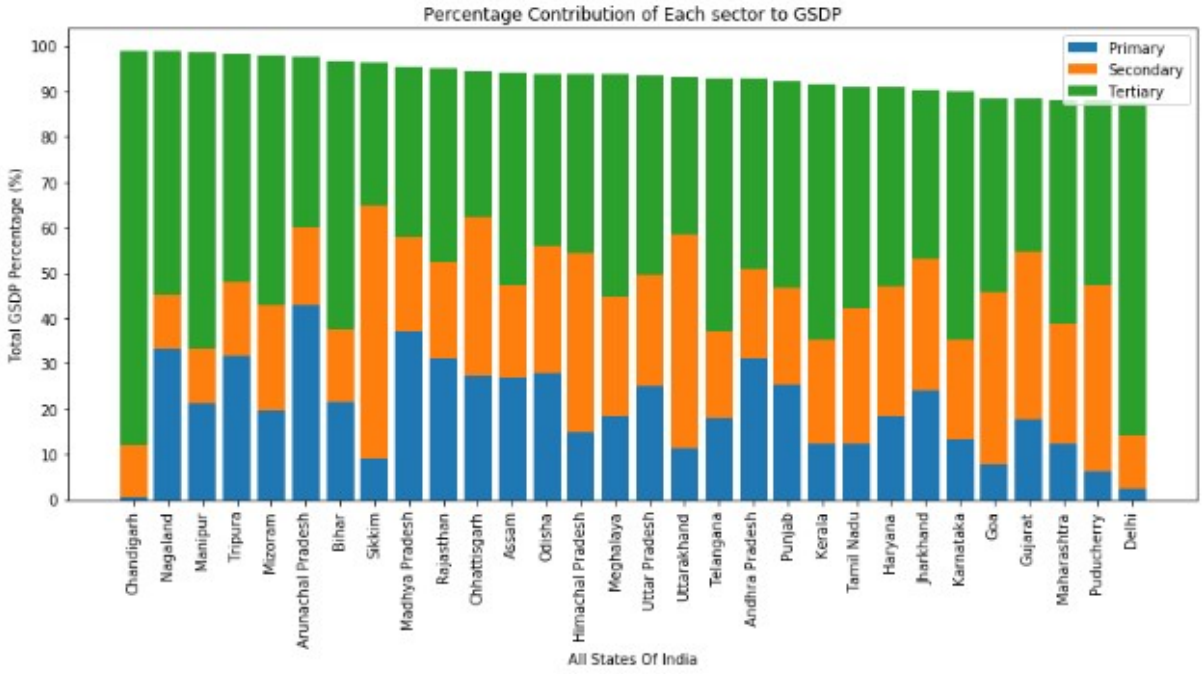
Identify the top 5 States: 1. Goa, 2. Chhattisgarh,3.Sikkim,4. Haryana, 5. Kerala

Identify bottom 5 states: 1. Bihar, 2.Uttar Pradesh,3. Manipur, 4. Assam, 5.Jharkhand

Find the ratio of the highest per capita GDP to the lowest per capita GDP : This compares the highest per Capita GDP of Goa Vs the lowest per Capita GDP of Bihar and the answer is 8.004741709371503.

- Plot the percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states.
    - Which plot will you use here? Why?
    - Why is (Primary + Secondary + Tertiary) not equal to total GDP?

Can you draw any insight from this? Find correlation of percentile of the state (% of states with lower per capita GDP) and %contribution of Primary sector to total GDP



We will use a BAR plot. This is because we can compare the Primary, Secondary and Tertiary sectors to the Total GSDP %. This way we can get a look&feel of the three sectors and compare them.

The Tertiary Sector is the maximum contribution to the GDP. This is because the Services sector is the largest sector of India and this contributes to the Tertiary Sector. It contributes to almost 45-50% of the GDP for most of the states i.e. the Primary and Secondary sector contribute to 50% of the GDP combined. The Secondary Sector or the Industrial Sector contributes 25%-30% of the GDP for most of the states.

Analysis for correlation of percentile of the state (% of states with lower per capita GDP) and %contribution of Primary sector to total GDP :

Lets us consider the below states **1. Bihar, 2.Uttar Pradesh,3. Manipur, 4. Assam, 5.Jharkhand which are the bottom 5 States in terms of Per Capita GDP.**

If we consider states with lower Per capita GDP, contribution from the Primary Sector is approximately **25%.**

Part-II: GDP and Education Dropout Rates

In Part-I, you would have noticed that (one) way to increase per capita GDP is by shifting the distribution of GDP towards the secondary and tertiary sectors, i.e., the manufacturing and services industries. But these industries can thrive only when there is an availability of educated and skilled labour.

In this part of the analysis, you will investigate whether there is any relationship between per capita GDP with dropout rates in education.

Data

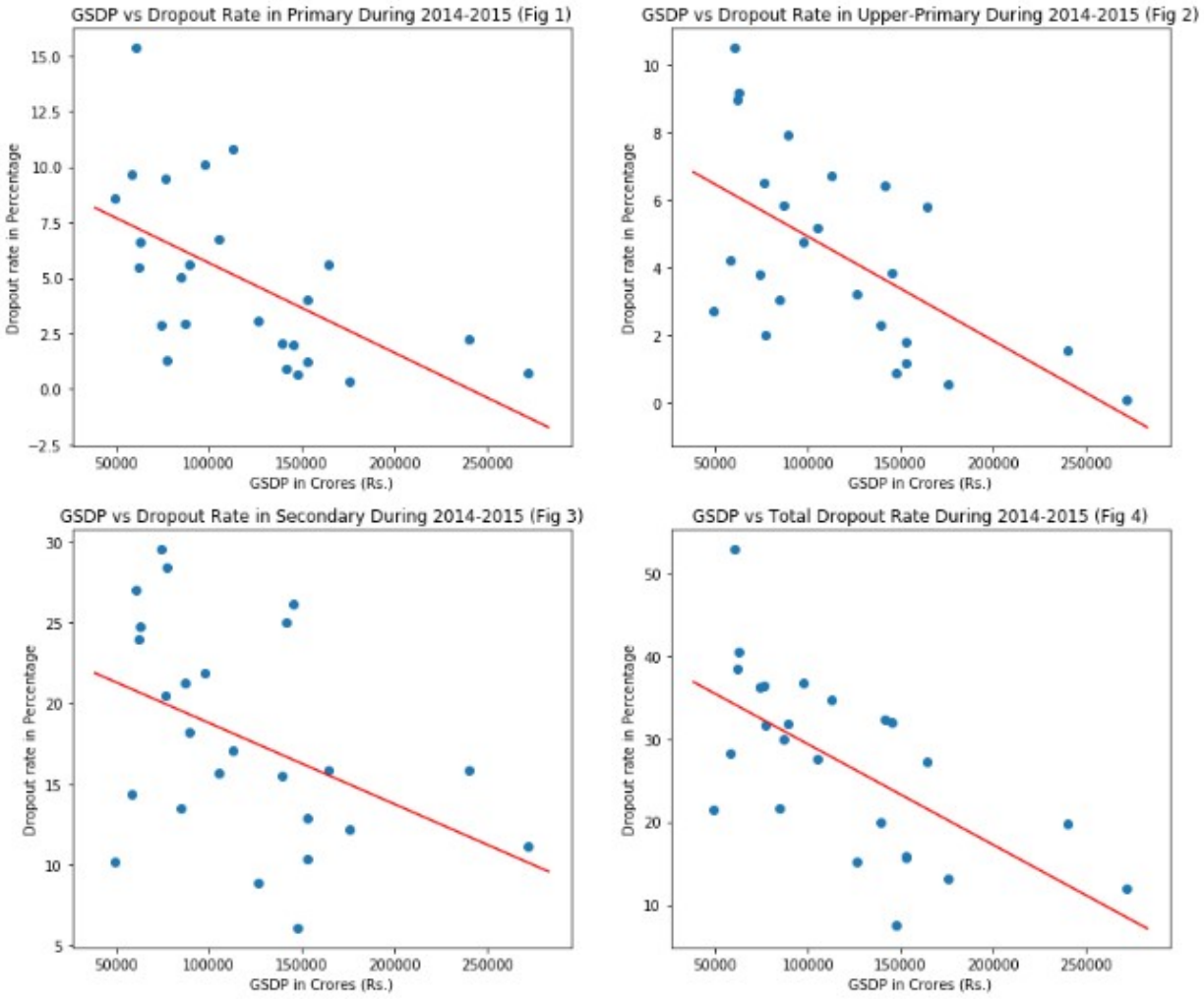
Data II: This section will require the dropout rate dataset apart from the dataset that you used in Part-1 of the case study. Download instructions are provided in the next segment.

Part-II: GDP and Education

- Analyse if there is any correlation of GDP per capita with dropout rates in education (primary, upper primary and secondary) for the year 2014-2015 for each state. Choose an appropriate plot to conduct this analysis.
    - Is there any correlation between dropout rate and %contribution of each sector (Primary, Secondary and Tertiary) to the total GDP?
- You have the total population of each state from the data in part I. Is there any correlation between dropout rates and population? What is the expected trend and what is the observation?
- Write down the key insights you draw from this data:
    - Form at least one reasonable hypothesis for the observations from the data

Analysis:

Let us consider the below plots:



Analysis:

- Looking at all the above plots, it is clear in all the four cases i.e. GSDP Vs DropOut rate in Primary, GSDP Vs Dropout rate in Upper-Primary and GSDP Vs Drpout Rate in Secondary, increase steadily by the decrease in dropout rate.
- Comparing at the plots, GSDP Vs DropOut rate in Primary, GSDP Vs Dropout rate in Upper-Primary, the increase in GSDP is approximately the same for the decrease in dropout rate, if you check the regression line and the plotted points
- The GSDP Vs DropOut rate in Primary and Upper-Primary as compared to GSDP Vs Drpout Rate in Secondary states that increase in GSDP is 50% lesser to the decrease in dropout rate of Secondary education students, comparing to the primary and upper-primary students.