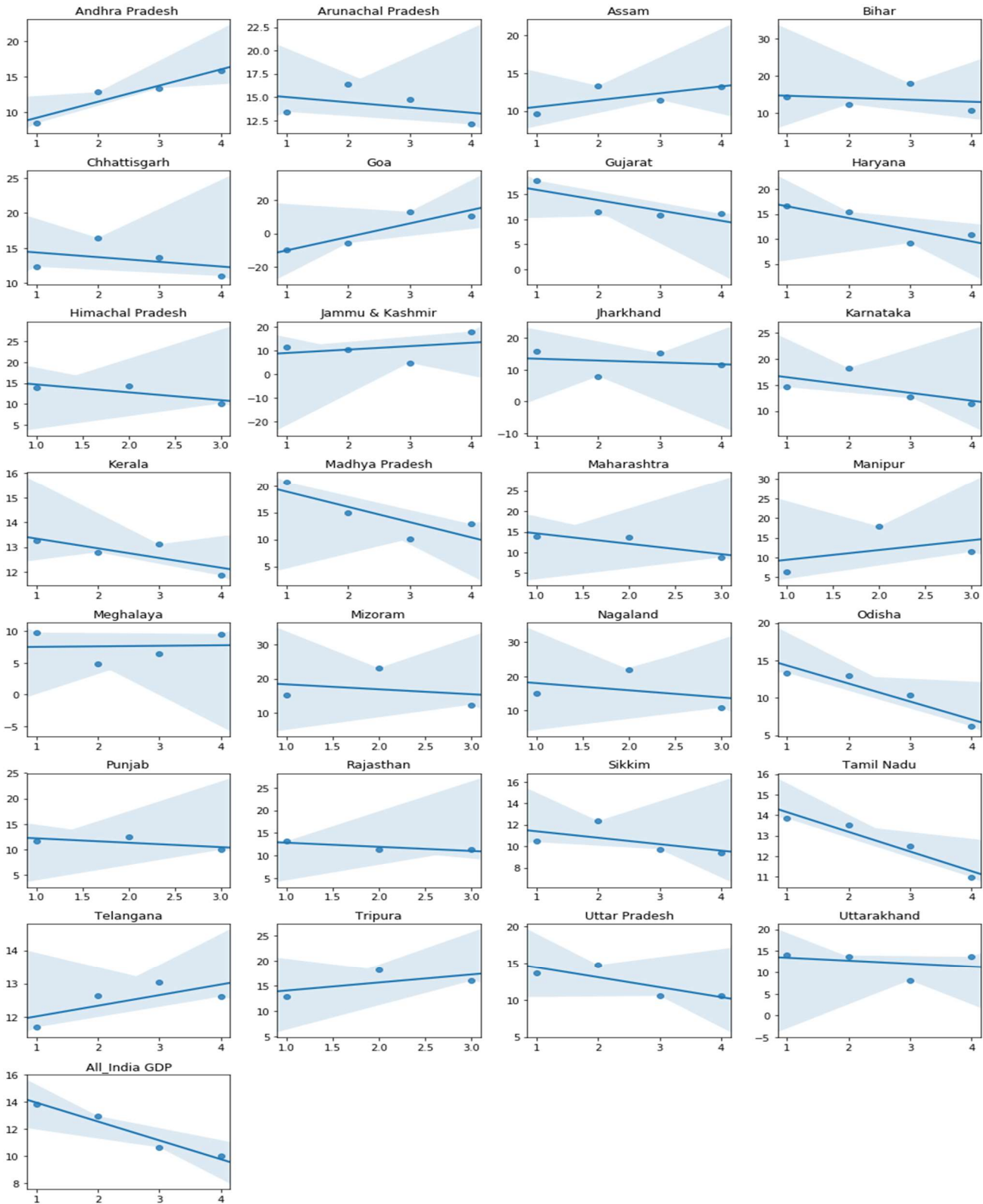# Part I-A

- **Best Fit line to represent the growth for each state:**

**Q. How will you compare the growth rates of any two states?**

**A.** We can compare Growth rate of any two states by looking at the slope of the best fit line of respective states. if the slope of one state is higher than other, its growth rate is higher.

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

**A.** Andhra Pradesh, Goa and Telangana are fastest growing states and Tamil Nadu, Odisha and Madhya Pradesh are slowest.

**Q. What is the Nation's growth rate?**

**A.** As per the plot, we can see that there is a negative slop which indicates a declining National growth rate over the years.

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

**A.** Uttar Pradesh is my home state. Overall its growth rate is negative. However, but as compared to National growth rate, it's better.

- **Plot the total GDP of the states for the year 2015-16:**
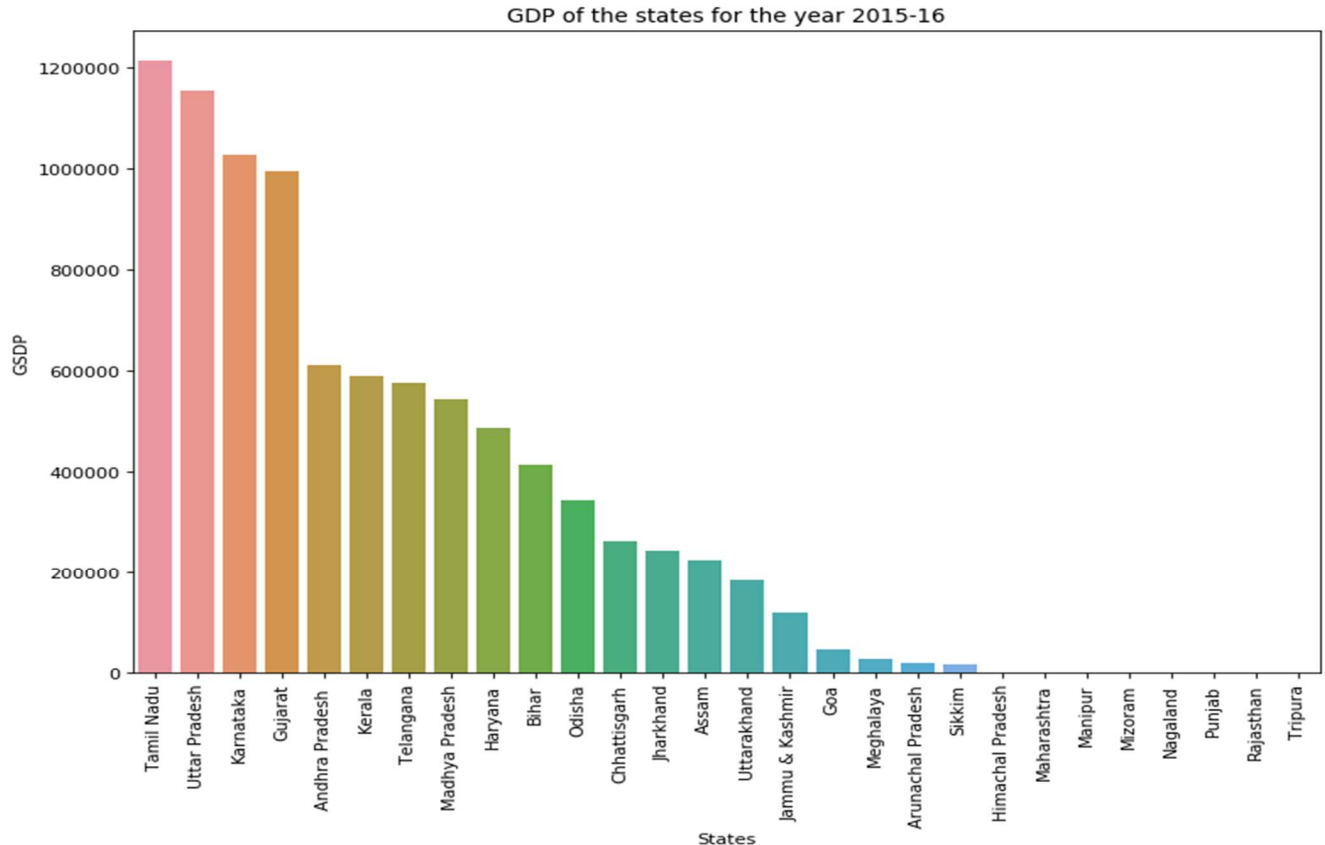
**Q. Which Plot will you use for this? Why? (Remember to plot the graph in a way such as it is easier to read and compare).**

**A. For this** I have used Bar Plot because we have 2 variables; one is categorical and other is numerical. So, to plot this, bar plot is right and simple choice for good visualization.

**Q. Identify the top 5 and the bottom 5 states based on total GDP.**

**A. Top 5 States:** Tamil Nadu, Uttar Pradesh, Karnataka, Gujarat, Andhra Pradesh

   **Bottom 5 States:** Sikkim, Arunachal Pradesh, Meghalaya, Goa, Jammu & Kashmir

**Q. What insights can you draw from this graph? What states are performing poorly?**

**A.** GDP depends on the population, so leaving the states which have missing values; states with less population tend to have smaller GDP and ones with more population tend to have larger GDP. Sikkim, Arunachal Pradesh, Medhalaya etc. are poorly growing states.

# Part I-B

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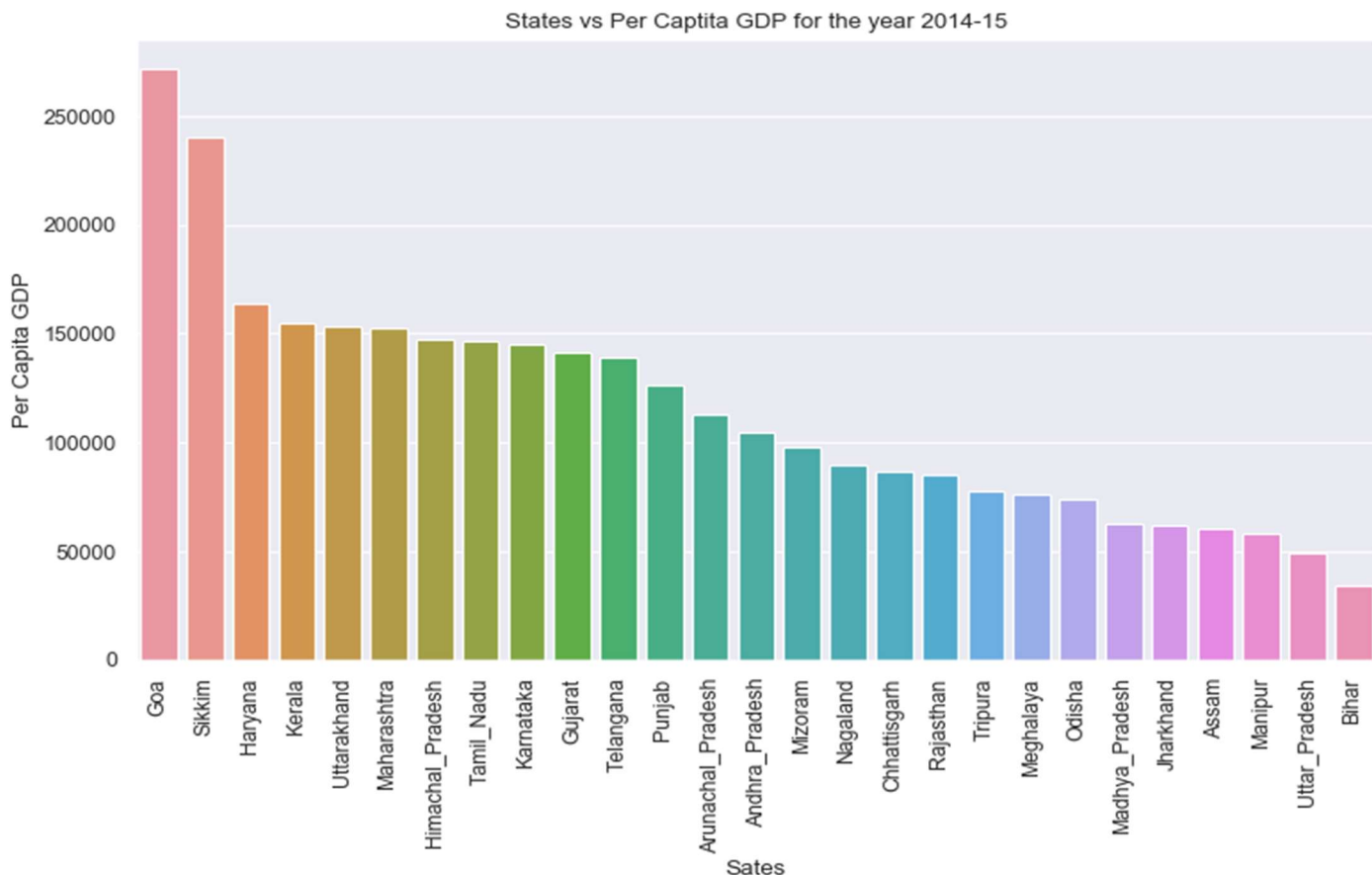
**Q. Identify the top 5 and the bottom 5 states based on the GDP per capita.**

**A. Top 5 states:** Goa, Sikkim, Haryana, Kerala, Uttarakhand

 **Bottom 5 States:** Bihar, Uttar Pradesh, Manipur, Assam, Jharkhand

**Q. Find the ratio of the highest per capita GDP to the lowest per capita GDP.**

**A.** ratio = all_states['Per Capita GSDP (Rs.)'].max()/all_states['Per Capita GSDP (Rs.)'].min()

 print(ratio)

 **Output:** 8.004741709371503

- ## Plot the percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states.
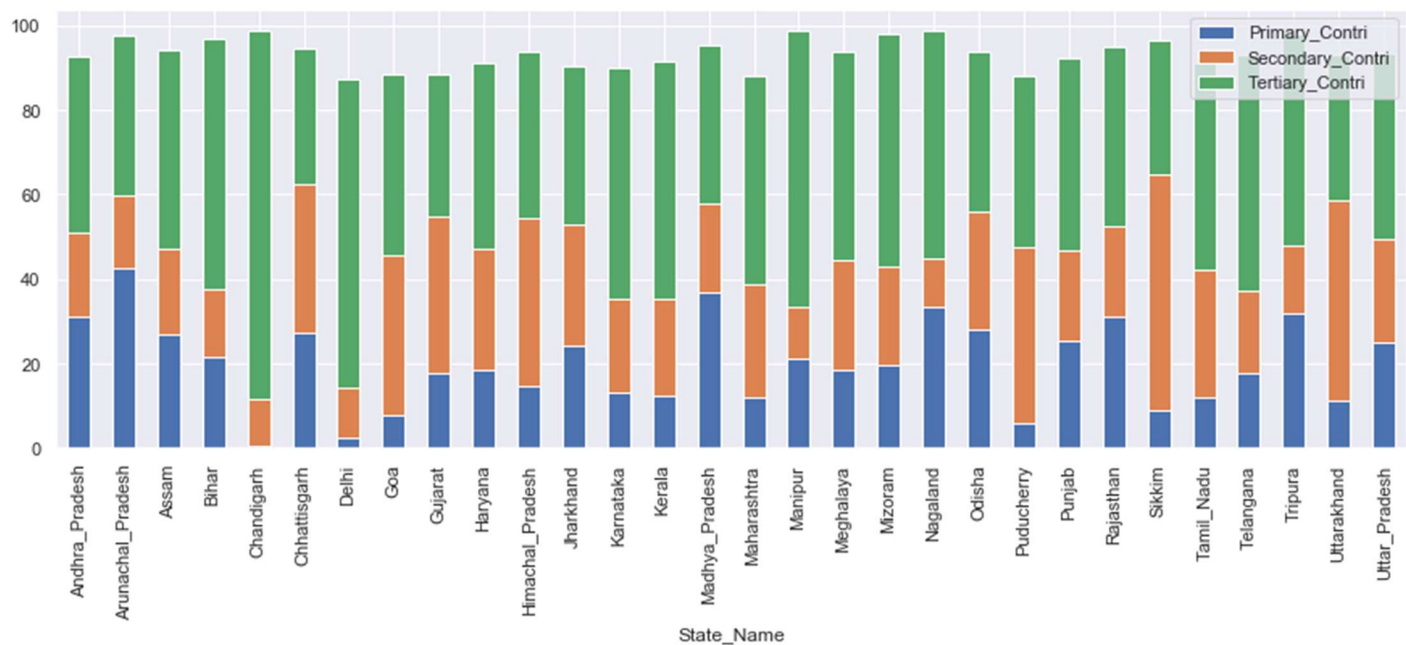


**Q. Which plot will you use here? Why?**

**A.** I have used Stacked Bar Plot. There are t categorical values here, so this plot gives nice idea of distribution of 3 sectors over the states.

**Q. Why is (Primary + Secondary + Tertiary) not equal to total GDP?**

**A.** Gross GDP is calculated as (Total GDP + Taxes - Subsidies). Since, there is no tax and subsidy in Primary + Secondary + Tertiary, this value is not equal to total GDP.

**Q. 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.**

**A.** An observation from the plot it is clear that States with low per capita GDP are getting major contribution from Tertiary sectors. Also, there is not a high correlation between state per capita GDP percentile and % contribution of primary sector. As per the analysis, states with high per capita GDP tend to have a lower contribution from Primary sector compared to states with lower per capita GDP.
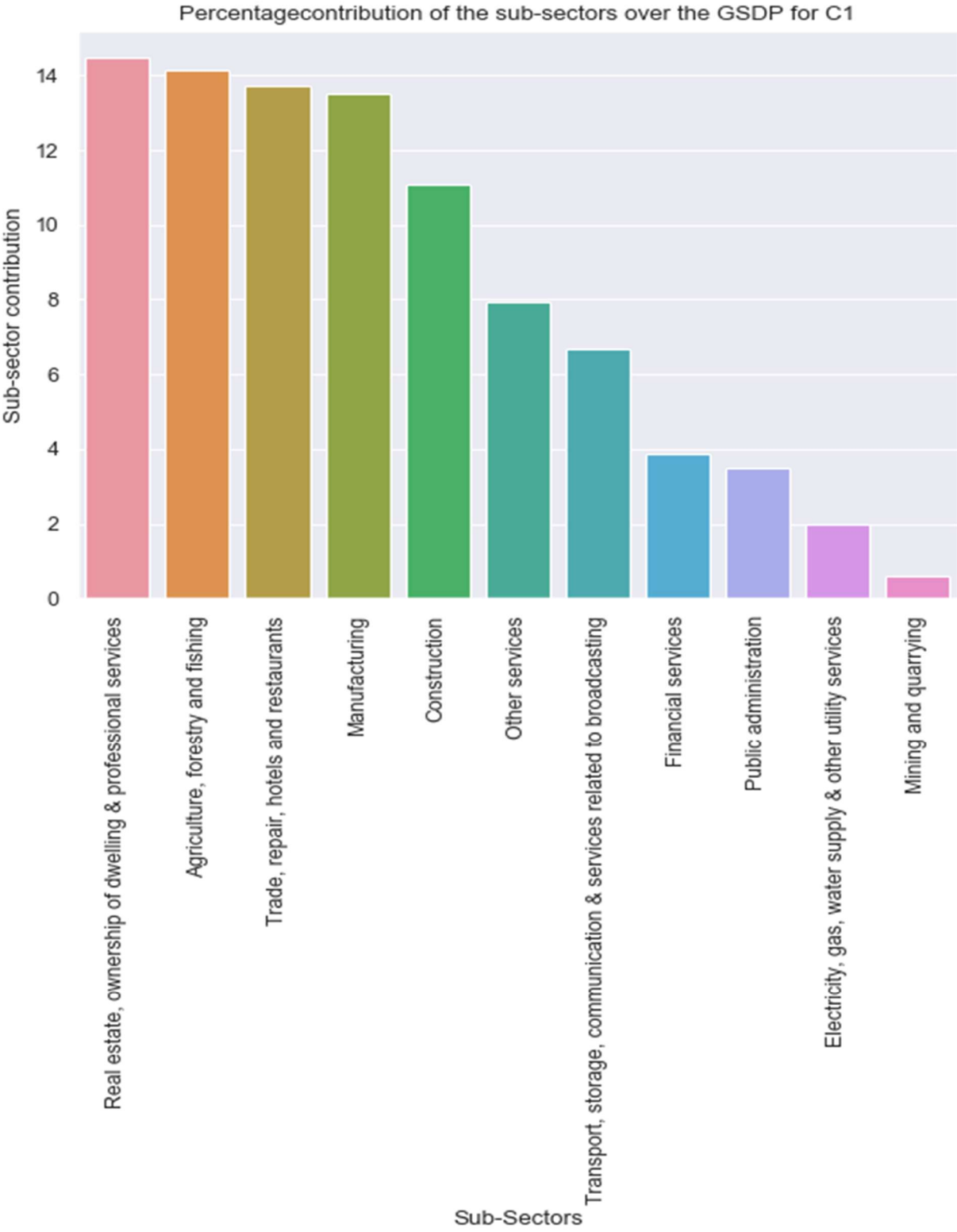
**Q. Categorize the states into four groups based on the GDP per capita (C1, C2, C3, C4, where C1 would have the highest per capita GDP and C4, the lowest). The quantile values are (0.20,0.5, 0.85, 1), i.e., the states lying between the 85th and the 100th percentile are in C1; those between the 50th and the 85th percentiles are in C2, and so on.**

**A.** Please refer to the Jupyter notebook for the code. Below are the results:

- **States in C1 Category:**
    - Goa
    - Sikkim
    - Haryana
    - Kerala
- **States in C2 Category:**
    - Uttarakhand
    - Maharashtra
    - Himachal Pradesh
    - Tamil Nadu
    - Karnataka
    - Gujarat
    - Telangana
    - Punjab
    - Arunachal Pradesh
- **States in C3 Category:**
    - Mizoram
    - Nagaland
    - Chhattisgarh
    - Rajasthan
    - Tripura
    - Meghalaya
    - Odisha
- **States in C4 Category:**
    - Madhya_Pradesh
    - Jharkhand
    - Assam
    - Manipur
    - Uttar_Pradesh
    - Bihar

- **Find the top 3/4/5 sub-sectors:**

  - **Top 3 sub-sectors for C1:**
    1. Real estate, ownership of dwelling & professional services
    2. Agriculture, forestry and fishing
    3. Trade, repair, hotels and restaurants

  - **Top 3 sub-sectors for C2:**
    1. Manufacturing
    2. Real estate, ownership of dwelling & professional services
    3. Agriculture, forestry and fishing

  - **Top 3 sub-sectors for C3:**
    1. Agriculture, forestry and fishing
    2. Manufacturing
    3. Trade, repair, hotels and restaurants

  - **Top 3 sub-sectors for C4:**
    1. Agriculture, forestry and fishing
    2. Trade, repair, hotels and restaurants
    3. Manufacturing

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

**For C1:**



Percentagecontribution of the sub-sectors over the GSDP for C1

**For C2:**



Percentagecontribution of the sub-sectors over the GSDP for C2

**For C3:**



Percentagecontribution of the sub-sectors over the GSDP for C3

**For C4:**
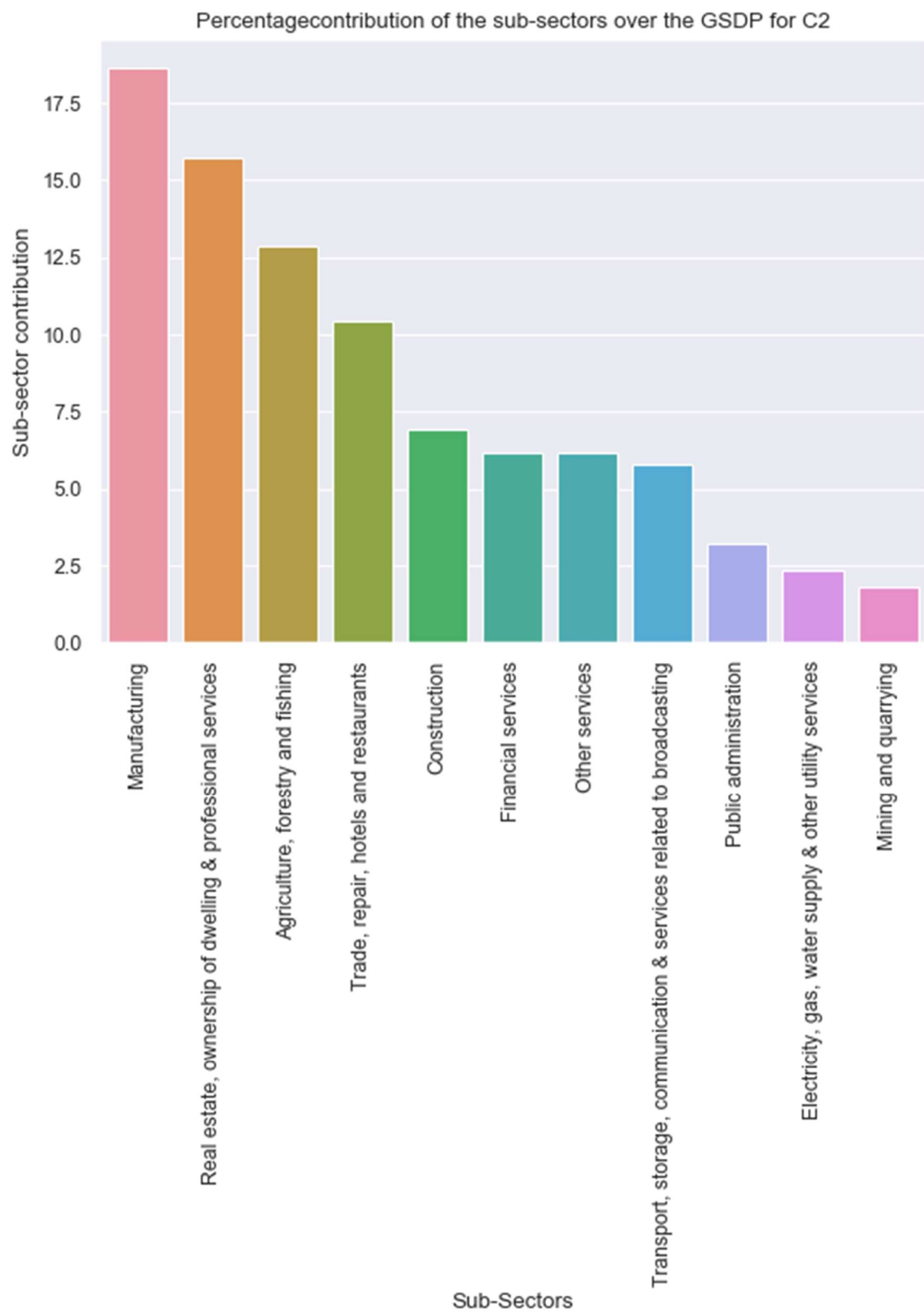


Percentagecontribution of the sub-sectors over the GSDP for C4
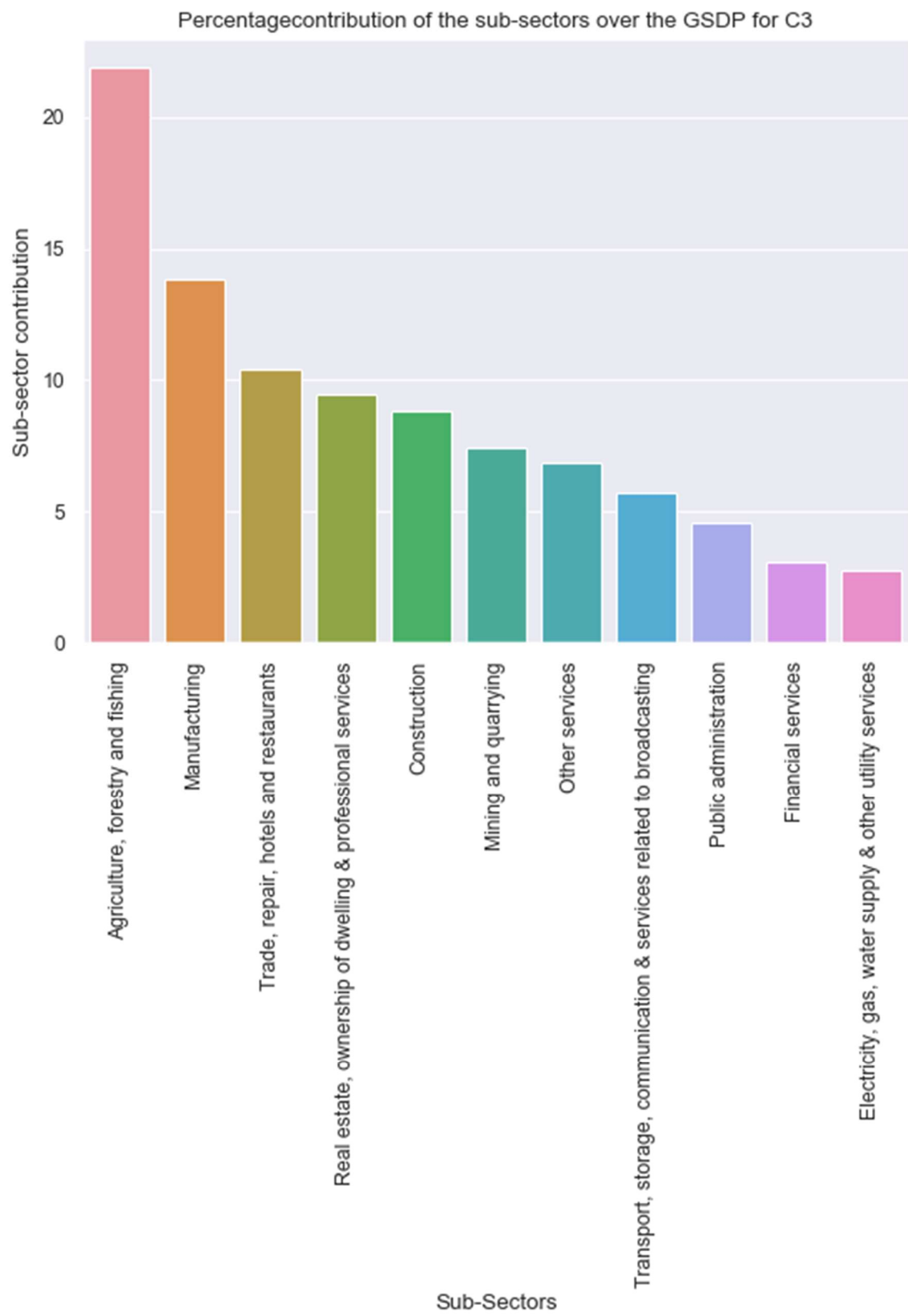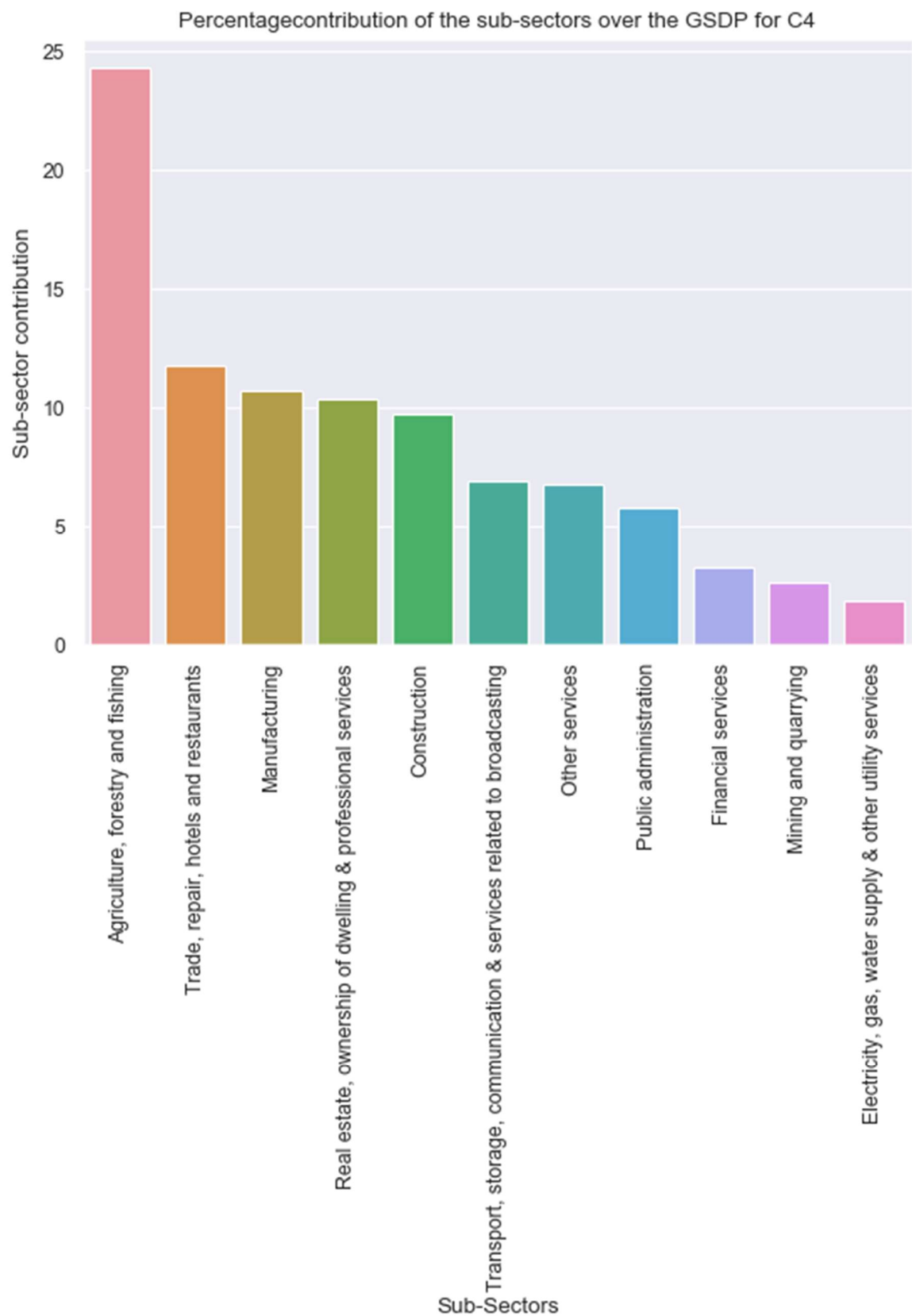
**Q. How does the GDP distribution of the top states (C1) differ from the others?**

**A.** States in the C1 category are majorly getting contribution from Secondary and Tertiary Sectors. As compared to states in other categories, C1 states are less in population, hence, high with per capita GDP. In C1 states main contribution is happening from "Real-estate" sub-sector, which is not on the top of other categories subsectors.

**Q. Which sub-sectors seem to be correlated with high GDP?**

**A.** After comparing graphs, Agriculture, Manufacturing and Trade are highly correlated with GDP.

**Q. Which sub-sectors do the various categories need to focus on?**

**A.** Mining, Electricity and Gas etc.

# Part-II: GDP and Education Dropout Rates

**Q. Is there any correlation between dropout rate and %contribution of each sector (Primary, Secondary and Tertiary) to the total GDP?**

**A.** Below is the graph and please refer to the Jupyter Notebook for the codes:



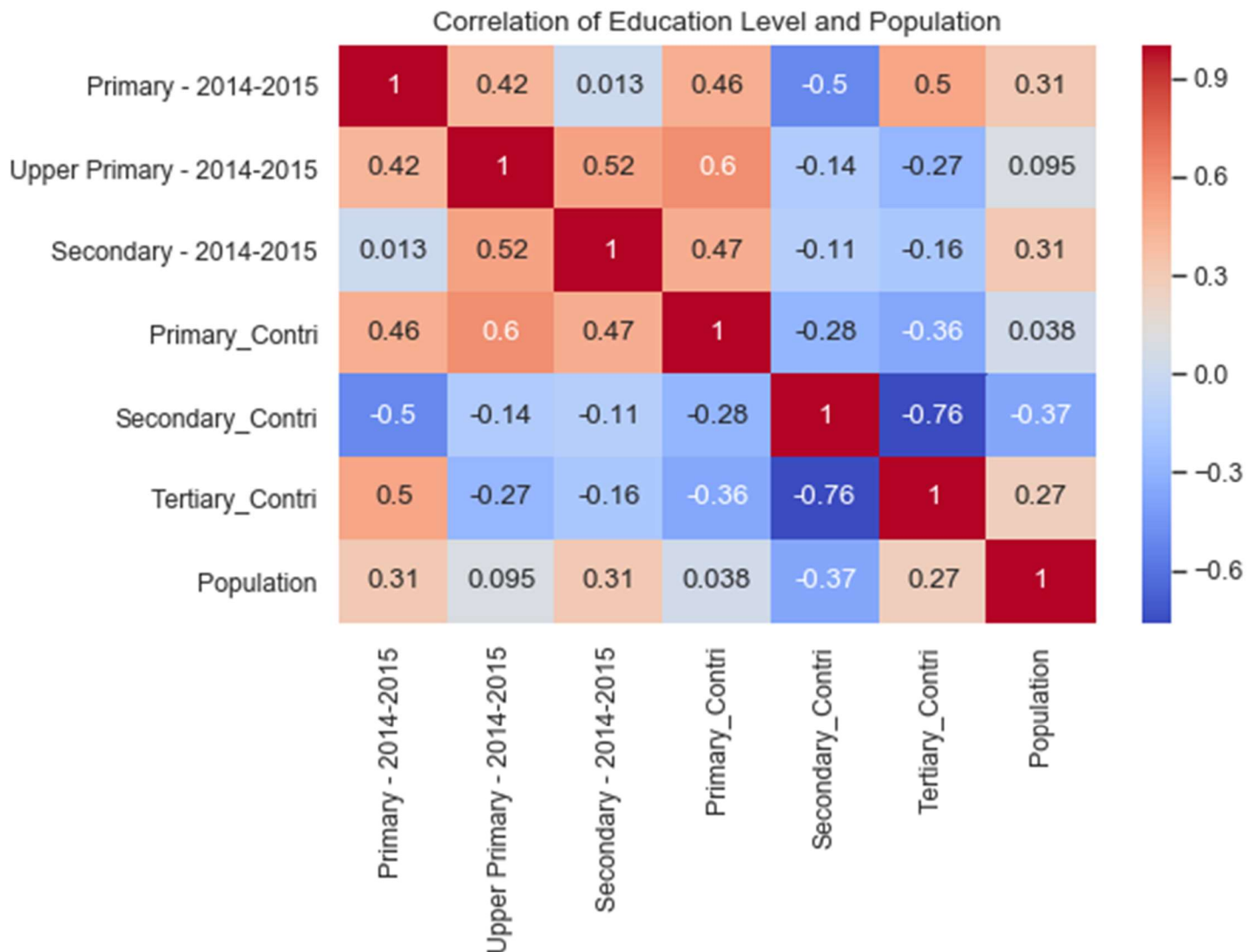Correlation between dropout rate and % contribution of each sector
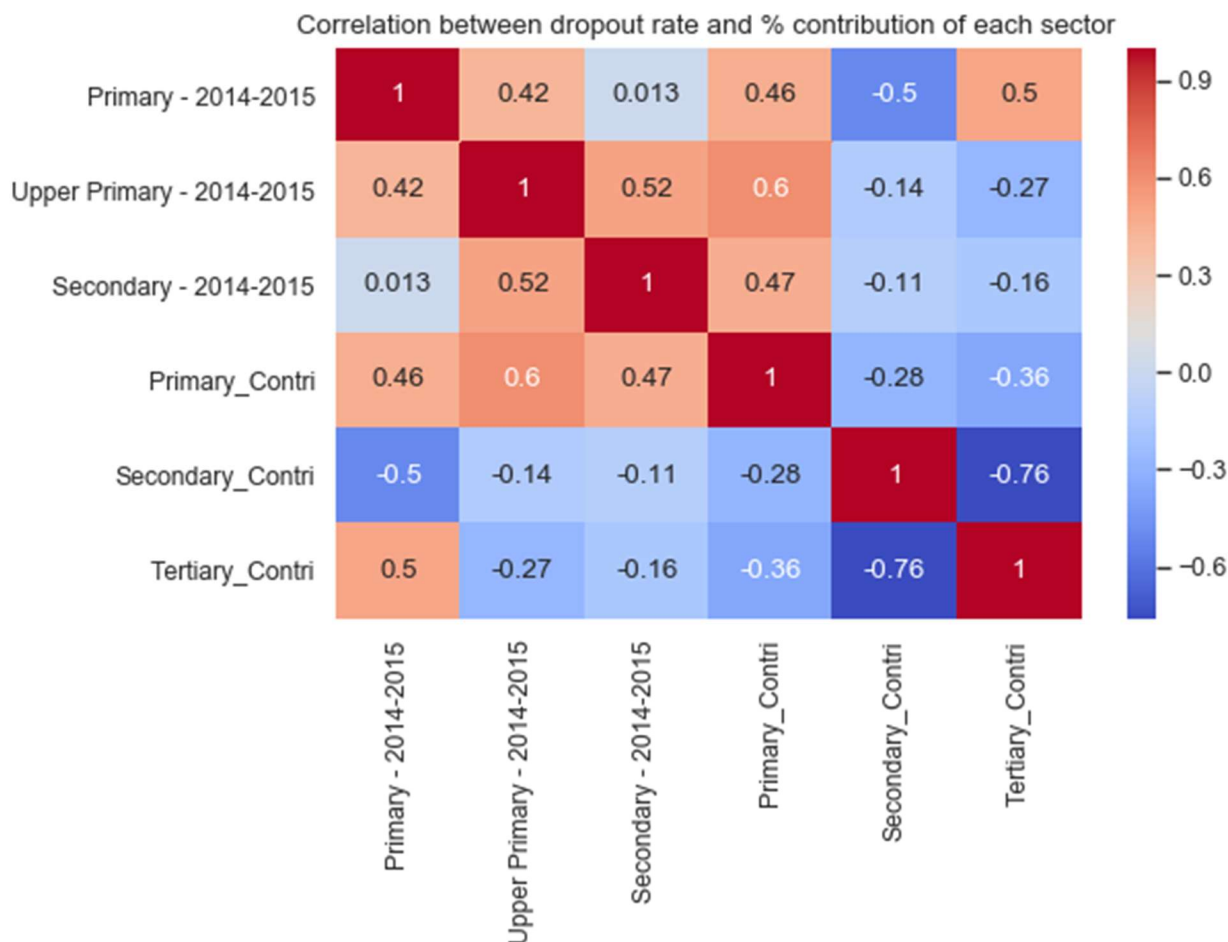
As per the plot, we can infer below:

1. Dropout rate for primary education is positively correlated with Primary and Tertiary contributions, however, it is negatively correlated with Secondary Contribution.
2. Dropout rate for upper primary education is positively correlated with only Primary Contribution, however, it is negatively correlated with Secondary and Tertiary Contributions.
3. Dropout rate for secondary education is also positively correlated with only Primary Contribution, however, it is negatively correlated with Secondary and Tertiary Contributions.

**Q. 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?**

**A.** Below is the graph and please refer to the Jupyter Notebook for the codes:



Correlation of Education Level and Population

| | Primary - 2014-2015 | Upper Primary - 2014-2015 | Secondary - 2014-2015 | Primary_Contri | Secondary_Contri | Tertiary_Contri | Population |
|---|---|---|---|---|---|---|---|
| Primary - 2014-2015 | 1 | 0.42 | 0.013 | 0.46 | -0.5 | 0.5 | 0.31 |
| Upper Primary - 2014-2015 | 0.42 | 1 | 0.52 | 0.6 | -0.14 | -0.27 | 0.095 |
| Secondary - 2014-2015 | 0.013 | 0.52 | 1 | 0.47 | -0.11 | -0.16 | 0.31 |
| Primary_Contri | 0.46 | 0.6 | 0.47 | 1 | -0.28 | -0.36 | 0.038 |
| Secondary_Contri | -0.5 | -0.14 | -0.11 | -0.28 | 1 | -0.76 | -0.37 |
| Tertiary_Contri | 0.5 | -0.27 | -0.16 | -0.36 | -0.76 | 1 | 0.27 |
| Population | 0.31 | 0.095 | 0.31 | 0.038 | -0.37 | 0.27 | 1 |

As per the plot, we can infer below:

Dropout rate for all the educations (Primary. Upper Primary and Secondary) is positively correlated with population, where Primary and Secondary levels have same correlation; however, Upper Secondary has vey less.