GDP Analysis Assignment

- Problem Statement:

- You are working as the chief data scientist at NITI Aayog, reporting to the CEO. The CEO has initiated a project wherein the NITI Aayog will provide top-level recommendations to the Chief Ministers (CMs) of various states, which will help them prioritise areas of development for their respective states. Since different states are in different phases of development, the recommendations should be specific to the states.
- The overall goal of this project is to help the CMs focus on areas that will foster economic development for their respective states. Since the most common measure of economic development is the GDP, you will analyse the GDP of the various states of India and suggest ways to improve it.

Data Exploration

- Need to download three sets of data namely:

1. Data I-A: 'State-wise Gross Domestic Product (GDP) at current price on yearly basis

2. Data I-B: Contains data on the Gross State Value Added (GSVA) and the Gross State Domestic Product (GSDP) for all states.

3. Data II: Dropout rates data

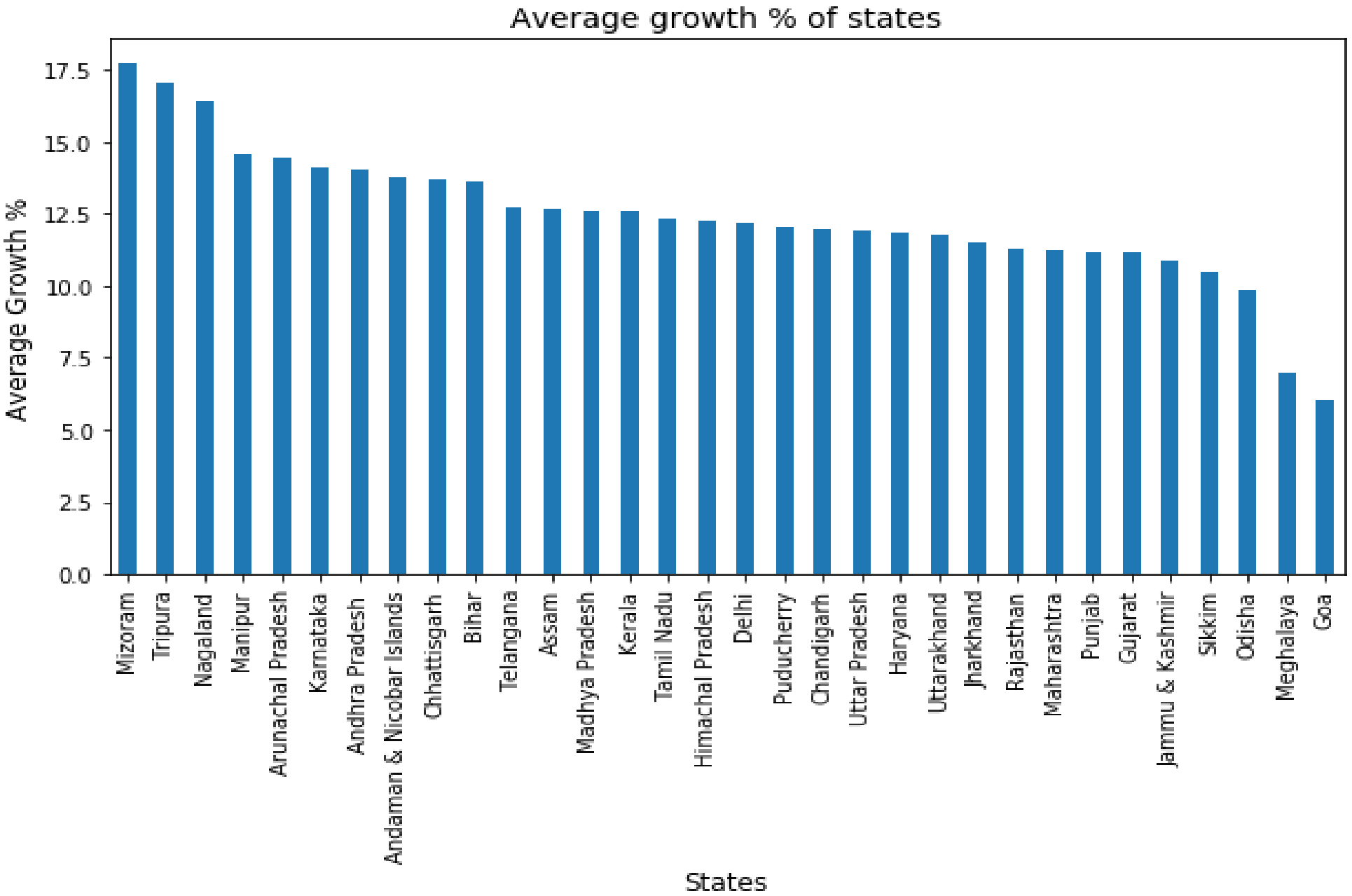
Data Cleaning and Manipulation

- Duplicate rows check has been performed.
- Column-wise null values percentage was taken, and row deleted which had 100% null values (West Bengal).
- The null values in other columns were ignored assuming it is better to avoid it rather than exaggerating the information and distorting the data accuracy.
- Few column names were renamed in Part II to ensure data consistency throughout the dataframe.

PART I-A

Average growth of states for the duration 2013-14, 2014-15 and 2015-16

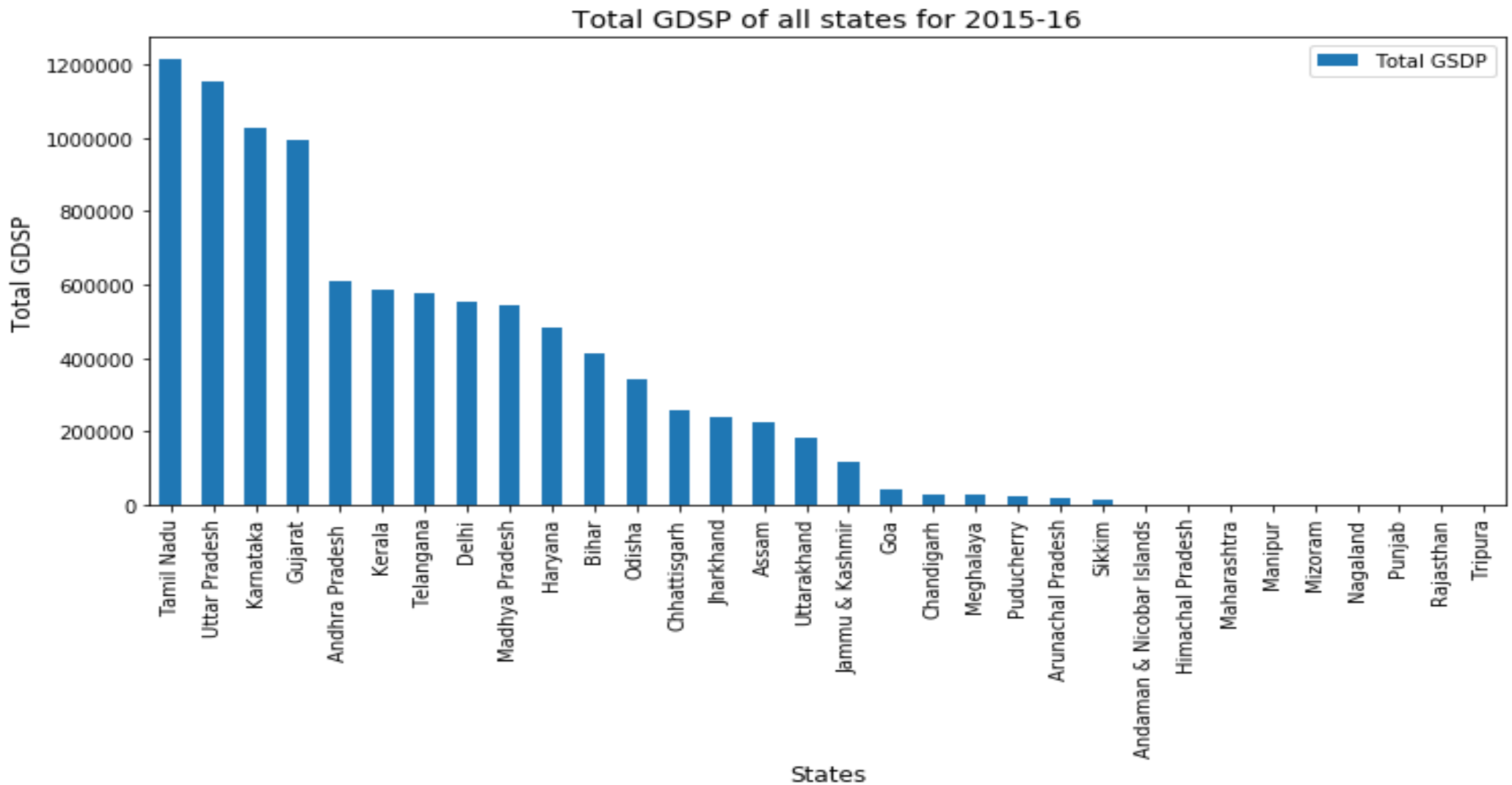
- The average growth of states was calculated by taking mean of the '(% Growth over previous year)' rows for duration 2013-14, 2014-15 and 2015-16.
- Same has been plotted using a bar graph as shown.


Average growth % of states
Average Growth %
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Mizoram
Tripura
Nagaland
Manipur
Arunachal Pradesh
Karnataka
Andhra Pradesh
Andaman & Nicobar Islands
Chhattisgarh
Bihar
Telangana
Assam
Madhya Pradesh
Kerala
Tamil Nadu
Himachal Pradesh
Delhi
Puducherry
Chandigarh
Uttar Pradesh
Haryana
Uttarakhand
Jharkhand
Rajasthan
Maharashtra
Punjab
Gujarat
Jammu & Kashmir
Sikkim
Odisha
Meghalaya
Goa
States

- We can clearly see the states that have been growing consistently fast are: Mizoram (17.70), Tripura (17.03), Nagaland (16.41), Manipur (14.61) etc.
- The states that have been struggling are Goa (6.03), Meghalaya (6.95), Odisha (9.83), Sikkim (10.48), Jammu&Kashmir(10.90) etc.
- Curiosity exercise: What has been the average growth rate of your home state, and how does it compare to the national average over this duration? My home state is Kerala and it seems to be at 14th position from top (including Andaman & Nicobar Islands) with an average growth rate of 12.58% which is more than the All_India GDP of 11.20%.

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

- After understanding which states have been consistently growing, we need to understand the Total GSDP of the states for the year 2015-16.
- Plotting a bar graph for the same.


Total GDSP of all states for 2015-16
Total GDSP
1200000
1000000
800000
600000
400000
200000
0
Total GDSP
Tamil Nadu
Uttar Pradesh
Karnataka
Gujarat
Andhra Pradesh
Kerala
Telangana
Delhi
Madhya Pradesh
Haryana
Bihar
Odisha
Chhattisgarh
Jharkhand
Assam
Uttarakhand
Jammu & Kashmir
Goa
Chandigarh
Meghalaya
Puducherry
Arunachal Pradesh
Sikkim
Andaman & Nicobar Islands
Himachal Pradesh
Maharashtra
Manipur
Mizoram
Nagaland
Punjab
Rajasthan
Tripura
States

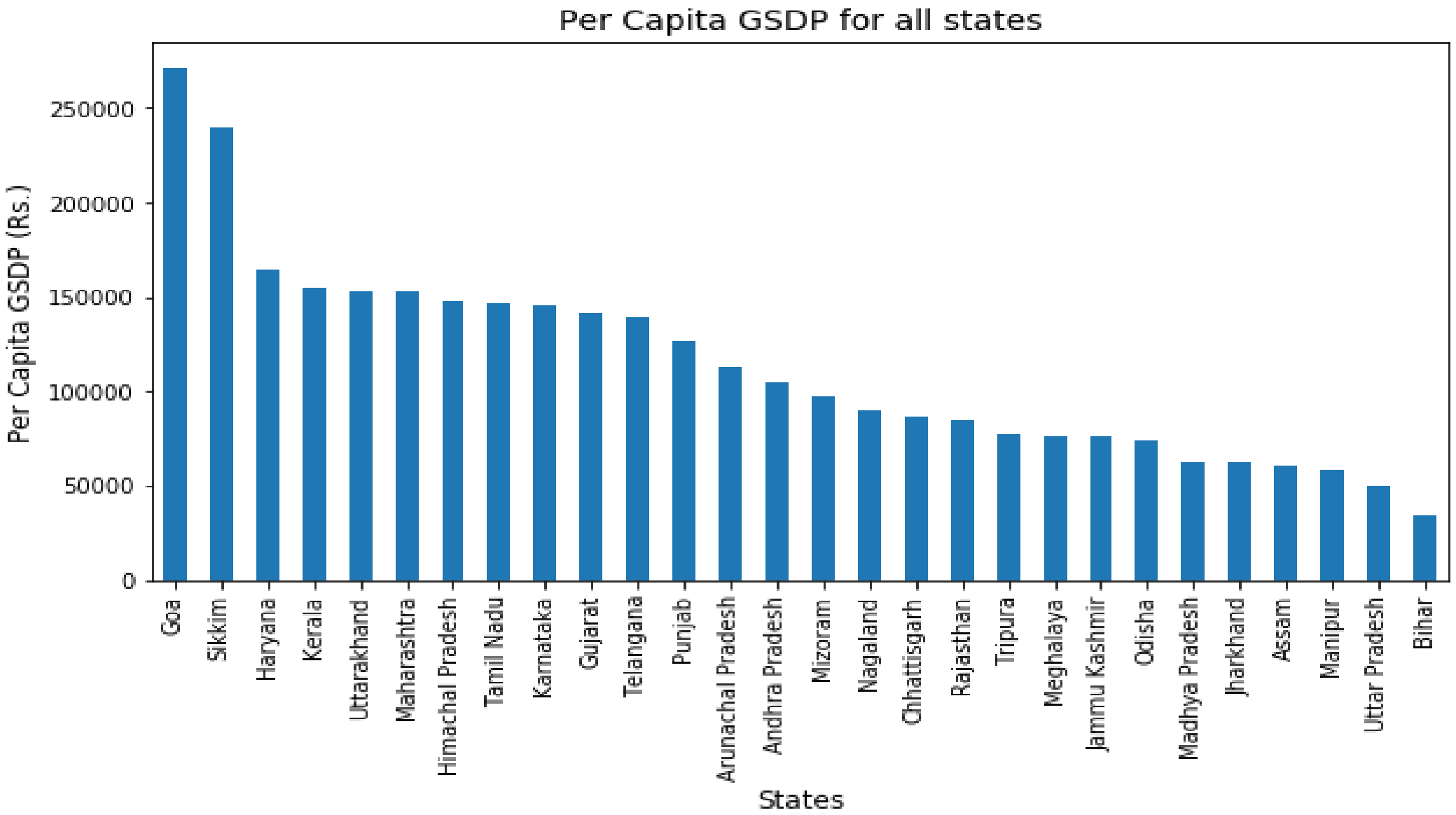
Identify the top 5 and the bottom 5 states based on total GDP.

- Top 5 states are: Tamil Nadu, Uttar Pradesh, Karnataka, Gujarat & Andhra Pradesh

- Bottom 5 states are (including UT): Sikkim, Arunachal Pradesh, Puducherry, Meghalaya, Chandigarh ,Goa.

PART I-B

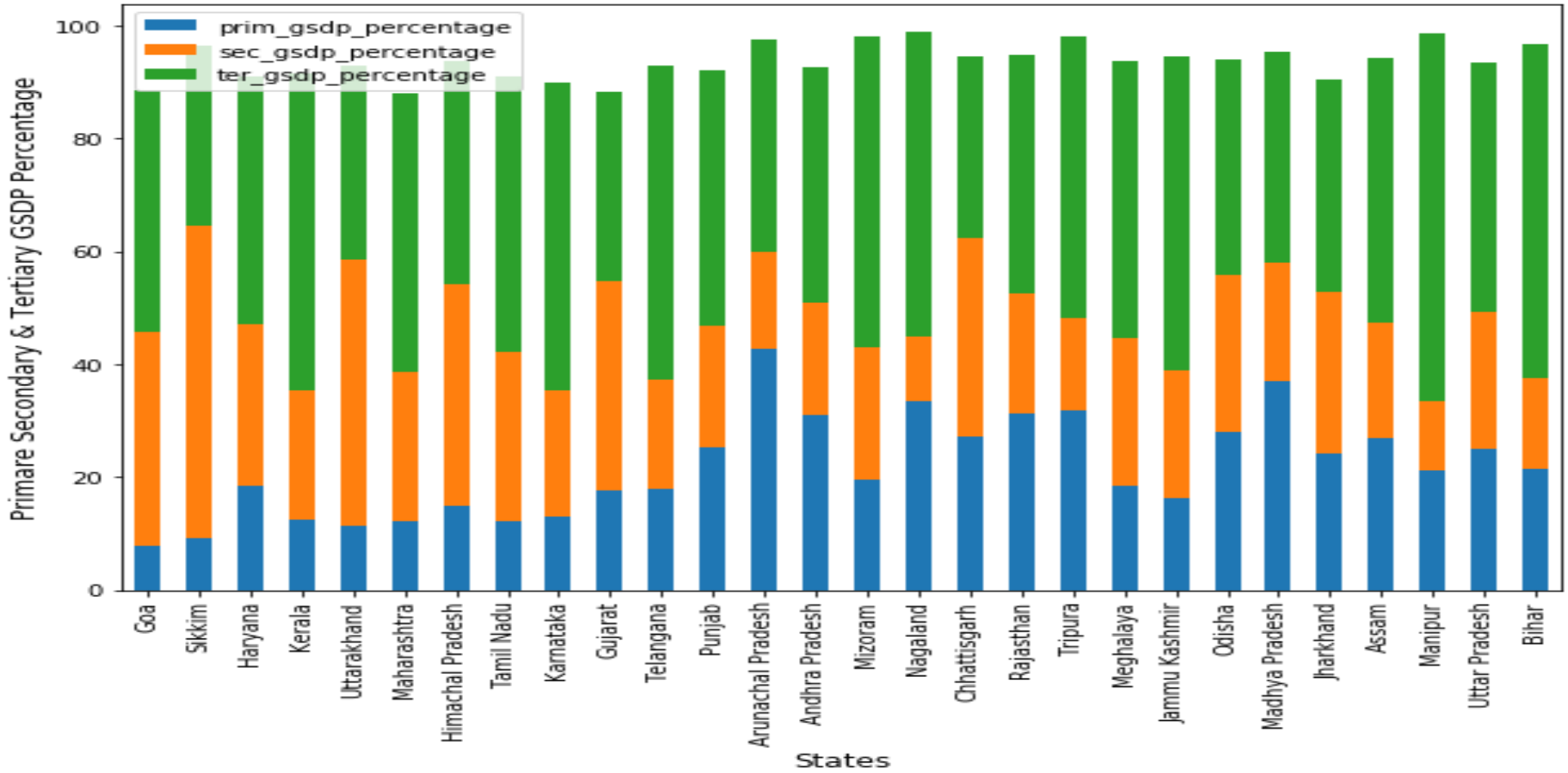
- After merging all the state and union territory datasets into a single dataframe, select data for only 2014-15 duration.
- The Union territories have been filtered out.
- To understand the per capita GDSP of all states, plotting a bar graph for the same.


Per Capita GSDP for all states
Per Capita GSDP (Rs.)
250000
200000
150000
100000
50000
0
Goa
Sikkim
Haryana
Kerala
Uttarakhand
Maharashtra
Himachal Pradesh
Tamil Nadu
Karnataka
Gujarat
Telangana
Punjab
Arunachal Pradesh
Andhra Pradesh
Mizoram
Nagaland
Chhattisgarh
Rajasthan
Tripura
Meghalaya
Jammu Kashmir
Odisha
Madhya Pradesh
Jharkhand
Assam
Manipur
Uttar Pradesh
Bihar
States

Identify the top 5 and the bottom 5 states based on the GDP per capita.

- Top 5 states : Goa, Sikkim, Haryana, Kerala, Uttarakhand
- Bottom 5 states : Bihar,Uttar Pradesh,Manipur,Assam,Jharkhand
- Find the ratio of the highest per capita GDP to the lowest per capita GDP.
- Ratio = 8.00474

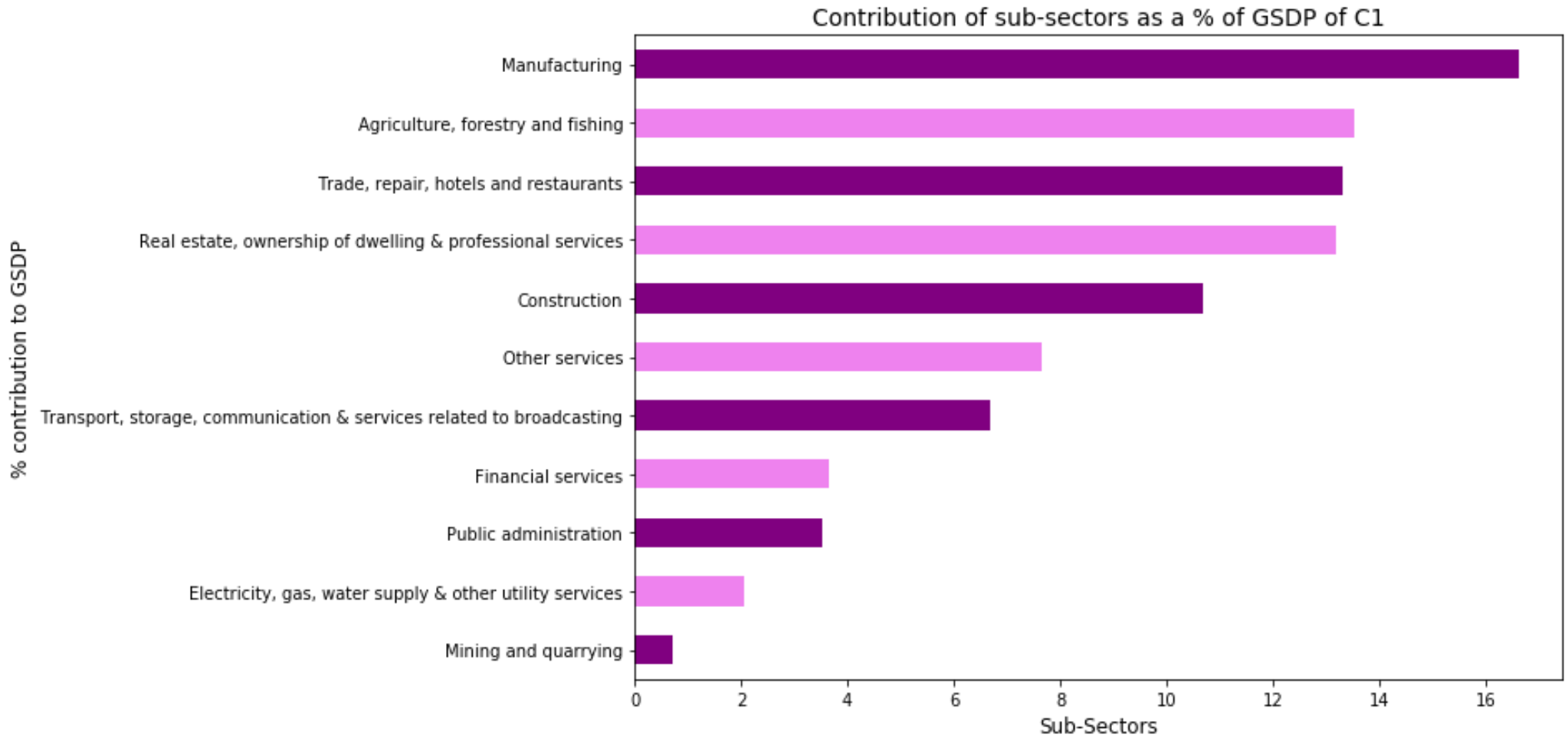
Plot the percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states.



- As seen in the above graph, the primary sector's contribution is minimum in states like Goa, Sikkim, Kerala, Uttarakhand etc. and maximum in Arunachal Pradesh, Rajasthan, Tripura etc. The secondary sector has major contribution in states like Sikkim, Himachal Pradesh, Gujarat etc. The tertiary sector is dominant in Telangana, Bihar, Kerala etc.

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

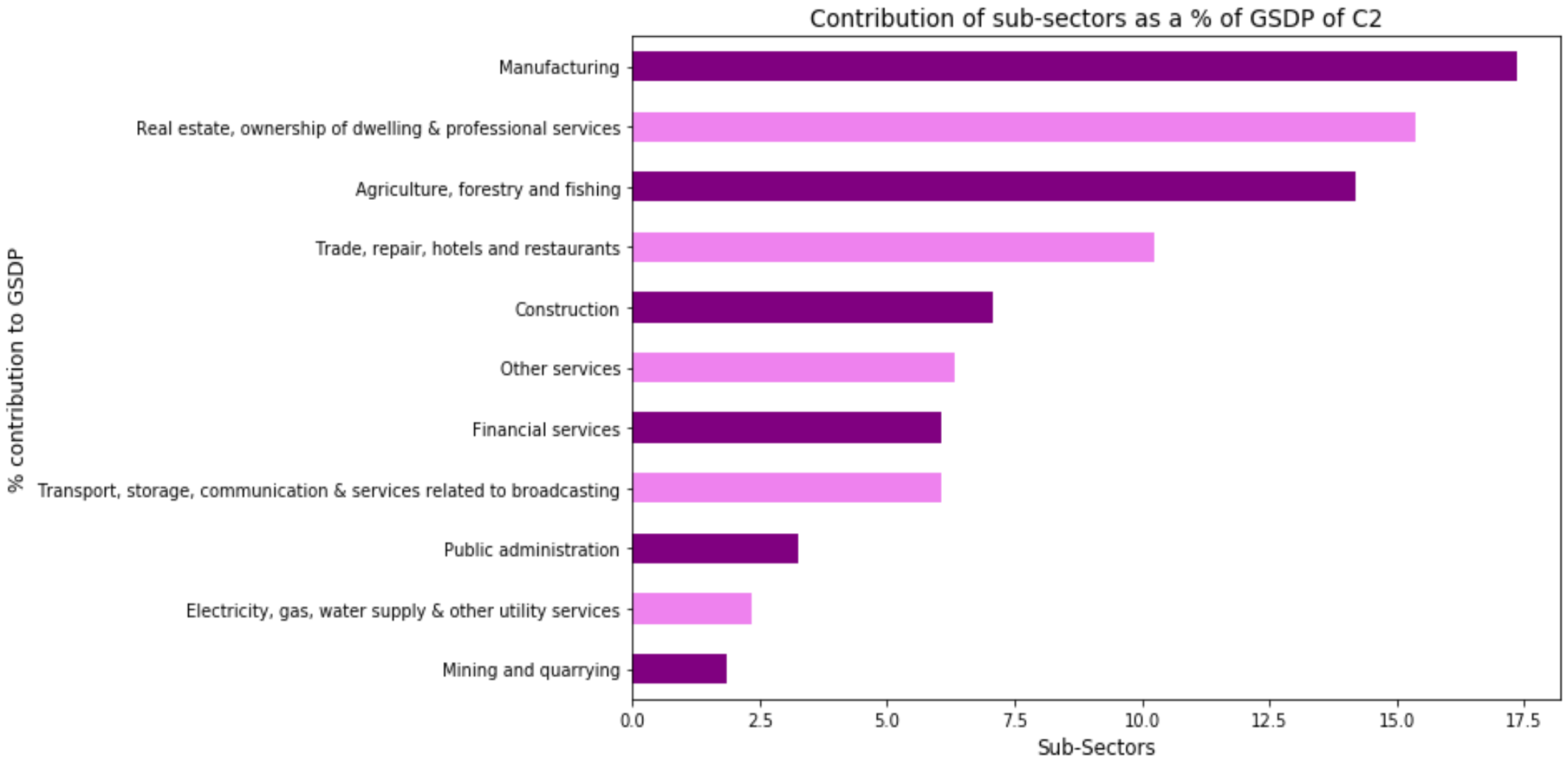
C1 Category



Inference from the Bar graph plotted for Category-1(C1).

- The major contribution to the overall GSDP is coming from the Manufacturing sector followed by ‘Agriculture, forestry &fishing’, ‘Trade, repair, hotels and restaurants’, ‘Real estate, ownership of dwelling & professional services’ etc. Minimum contribution is from ‘Mining & quarrying’ followed by ‘Electricity, gas, water supply &utility services’.
- Almost all the states in this category are famous tourist destinations. Hence, government should focus more on the manufacturing sector as well as ‘Agriculture, forestry &fishing’ and ‘Trade, repair, hotels and restaurants’ sectors so, that we can more and more tourists and increase the GSDP.
- ‘Other services’ and services related to ‘Transport, storage, communication and broadcasting’ have a moderate contribution to the GSDP and hence a little more focus on that side may improve the gain from those sectors as well.

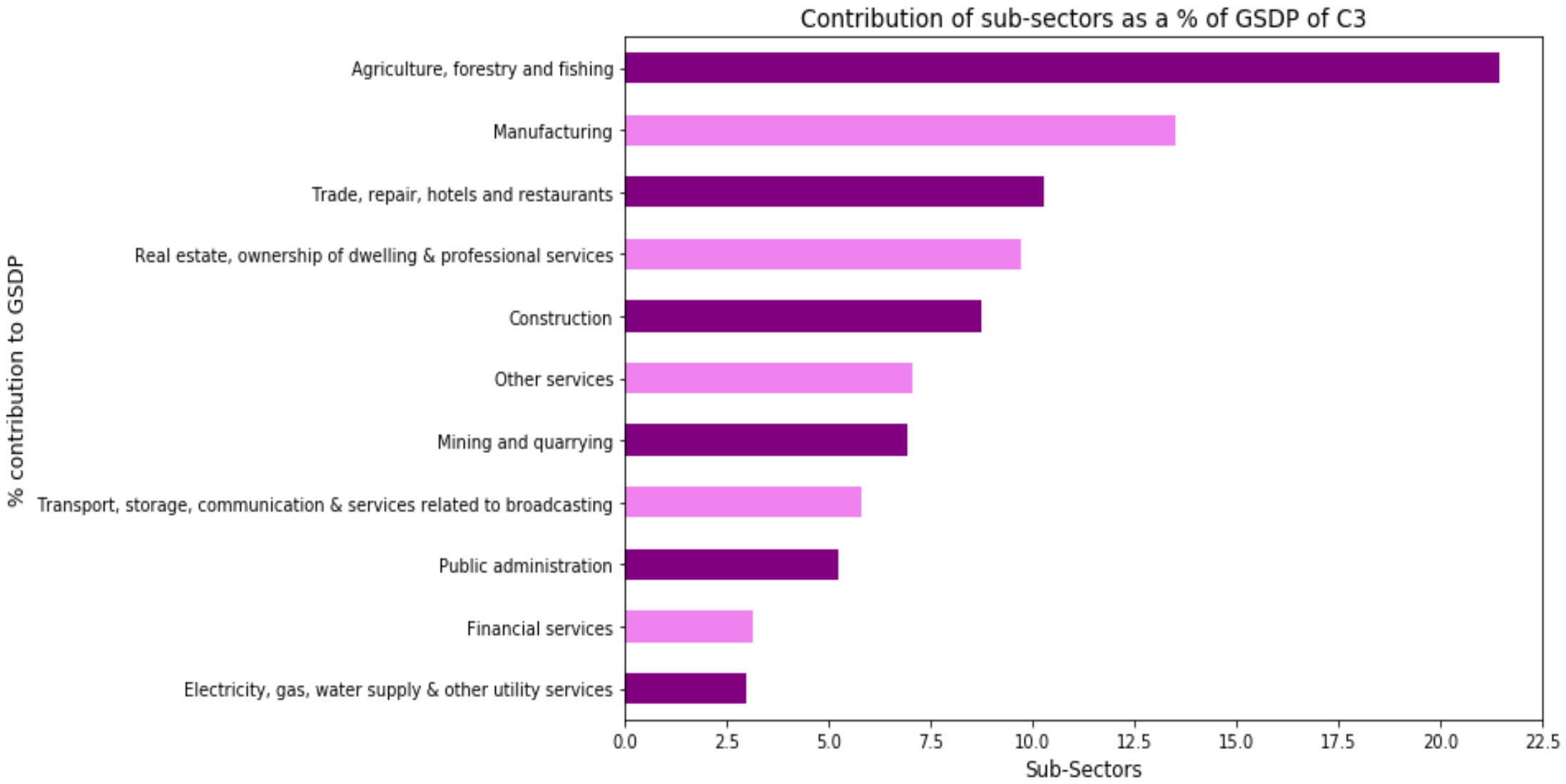
C2 Category


Contribution of sub-sectors as a % of GSDP of C2
Manufacturing
Real estate, ownership of dwelling & professional services
Agriculture, forestry and fishing
Trade, repair, hotels and restaurants
Construction
Other services
Financial services
Transport, storage, communication & services related to broadcasting
Public administration
Electricity, gas, water supply & other utility services
Mining and quarrying
% contribution to GSDP
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Sub-Sectors

Inference from the bar graph plotted for Catgeory-3(C3).

- The sub-sector with highest contribution to GSDP is ‘Agriculture, forestry and fishing’, also the contribution of agriculture sector is very high compared to Category-1 & 2.
- We can infer from the states in this category that number of industries or companies are less compared to other states and hence more policies or schemes should be implemented to increase the revenue coming from this sector.
- There is a sudden drop from the agriculture sector to manufacturing sector. Government should focus more on manufacturing sector as once it improves, automatically there would be increase in the trade and real estate sector.
- One more thing that we notice here is that the contribution of ‘Mining &quarrying’ sector to GSDP is more in this category compared to remaining 3 categories, so they may have potential to bring in more revenue.

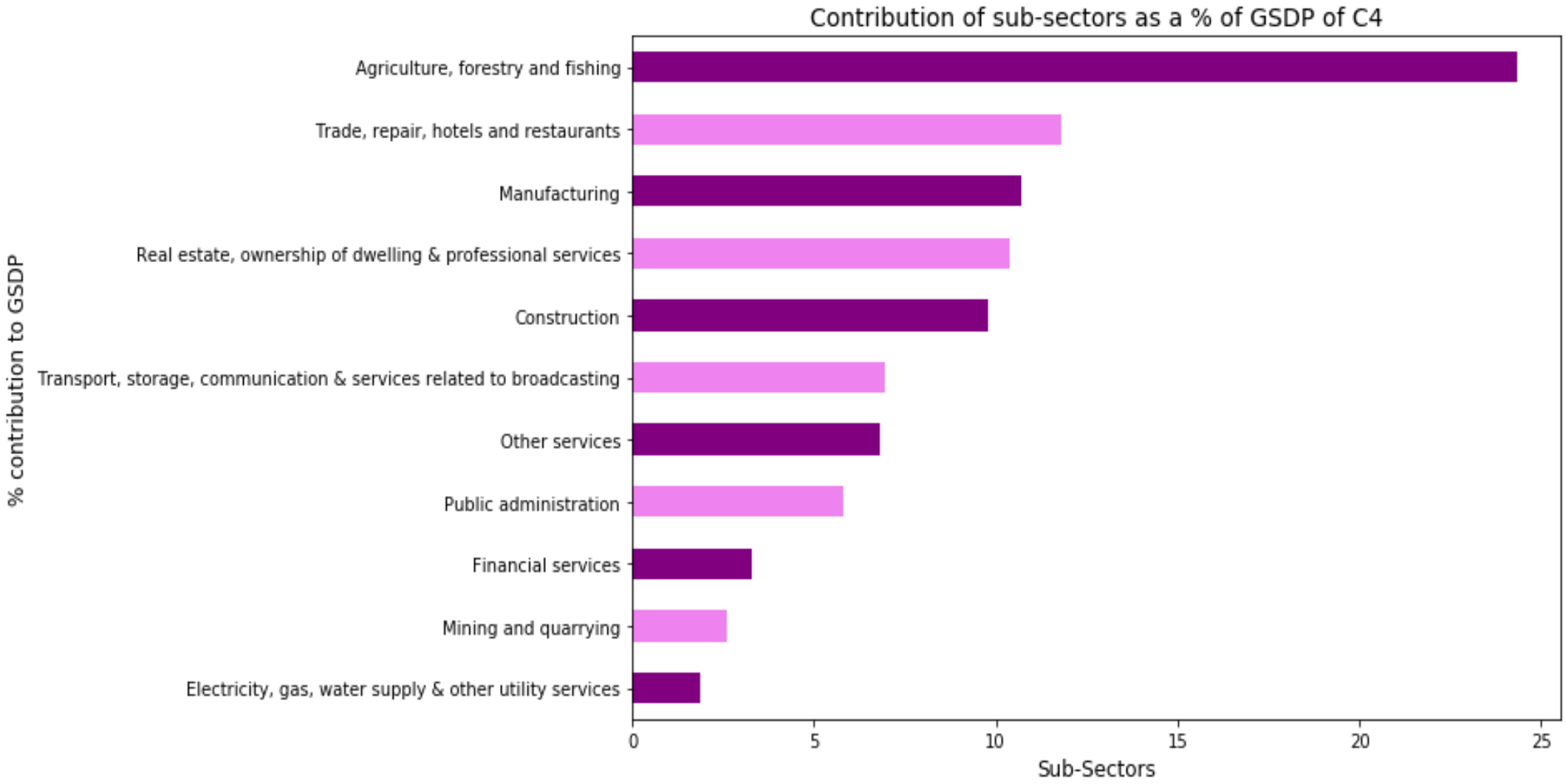
C3 Category


Contribution of sub-sectors as a % of GSDP of C3
Agriculture, forestry and fishing
Manufacturing
Trade, repair, hotels and restaurants
Real estate, ownership of dwelling & professional services
Construction
Other services
Mining and quarrying
Transport, storage, communication & services related to broadcasting
Public administration
Financial services
Electricity, gas, water supply & other utility services
% contribution to GSDP
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
Sub-Sectors

Inference from the bar graph plotted for Catgeory-3(C3).

- The sub-sector with highest contribution to GSDP is ‘Agriculture, forestry and fishing’, also the contribution of agriculture sector is very high compared to Category-1 & 2.
- We can infer from the states in this category that number of industries or companies are less compared to other states and hence more policies or schemes should be implemented to increase the revenue coming from this sector.
- There is a sudden drop from the ‘Agriculture, forestry and fishing’ sector to ‘Manufacturing’ sector. Government should focus more on manufacturing sector as once it improves, automatically there would be increase in the trade and real estate sector.
- One more thing that we notice here is that the contribution of ‘Mining & quarrying’ sector to GSDP is more in this category compared to remaining 3 categories, so they may have potential to bring in more revenue.

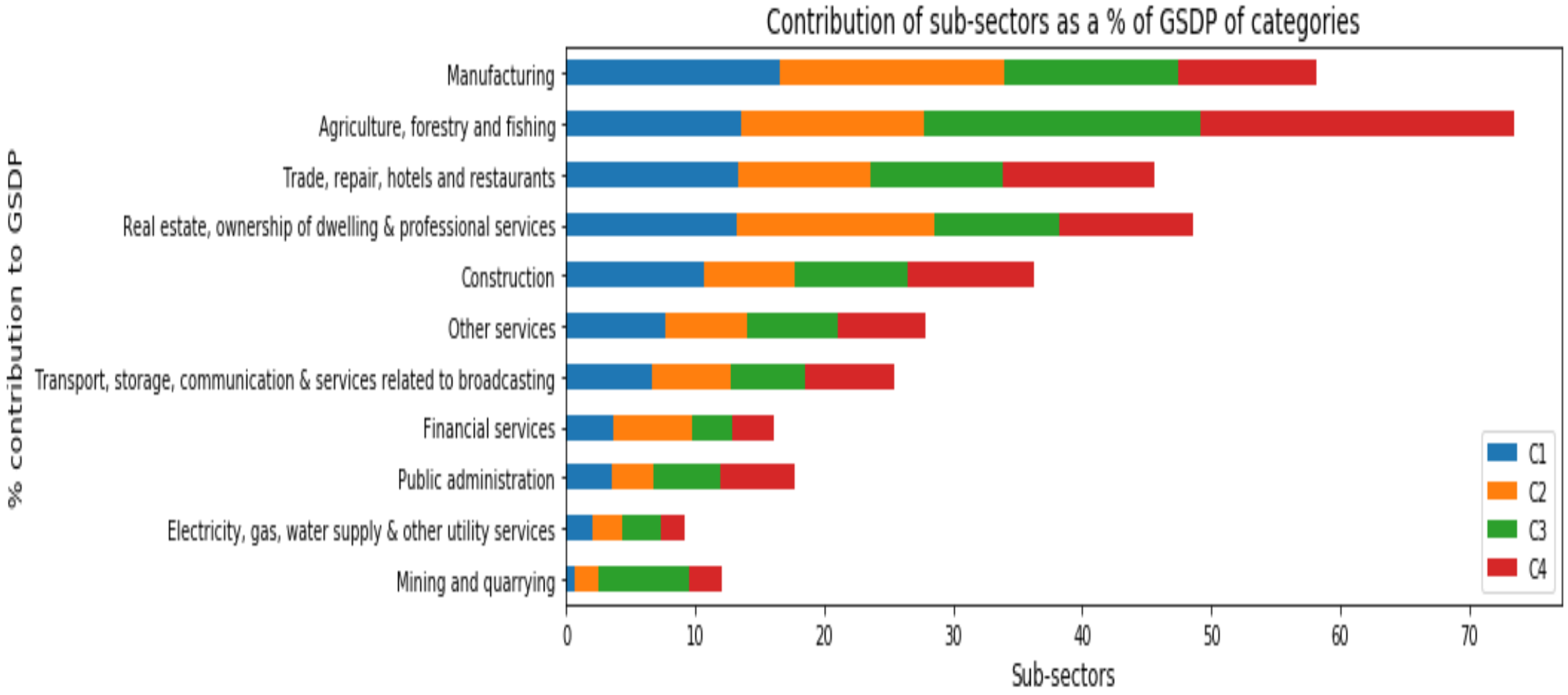
C4 Category


Contribution of sub-sectors as a % of GSDP of C4
Agriculture, forestry and fishing
Trade, repair, hotels and restaurants
Manufacturing
Real estate, ownership of dwelling & professional services
Construction
Transport, storage, communication & services related to broadcasting
Other services
Public administration
Financial services
Mining and quarrying
Electricity, gas, water supply & other utility services
% contribution to GSDP
0
5
10
15
20
25
Sub-Sectors

Inference from the bar graph plotted for Category-4(C4).

- This looks like Category-3. Agriculture sector has the highest contribution to GSDP exactly like C3.
- There is a sudden drop from agriculture to trade, repair, hotels and restaurants sector. Government should try to maintain this by implementing more schemes and policies.
- Even here, the manufacturing and real estate sector seem like having the potential to bring in more revenue, so some focus there would improve the GSDP.

Plot a stacked bar graph to see the contribution of sub-sectors in each category.



Inference from the stacked bar graph showing % contribution to GSDP across all categories.

- One thing that we can see from the above stacked bar chart is the top states in Category C1 and C2 have the Manufacturing sub-sector contributing more to GSDP percentage whereas the C3 and C4 category states have Agriculture, Forestry and fishing leading way.
- 'Real estate, ownership of dwelling & professional services' sector seem to have prominence in Category-2. Similarly, the financial services sector is comparatively more in category-2.
- Electricity, gas, water supply & other utility services seem to be consistently low across all categories.

How does the GDP distribution of the top states (C1) differ from the others?

- One thing that was noticed is that almost all the states falling in C1 category are famous tourist destinations and hence the sector with highest % contribution to GSDP is Manufacturing followed by agriculture and trade, repair &hotel etc.
- The Agriculture sector has the lowest contribution to GDSP in C1 category compared to other categories.
- The Mining and Quarrying sector is comparatively lowest in category C1 compared to other categories.

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

- Manufacturing
- Agriculture, forestry & fishing
- Real estate, ownership of dwelling & professional services
- Trade, repair, hotels & restaurants

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

- Category-1: Manufacturing, Agriculture, forestry & fishing, Trade, repair, hotels & restaurants, Real estate, ownership of dwelling &professional services
- Category-2: Manufacturing, Real estate, ownership of dwelling &professional services, Agriculture, forestry & fishing, Trade, repair, hotels & restaurants
- Category-3: Agriculture, forestry & fishing, Manufacturing, Trade, repair, hotels & restaurants, Real estate, ownership of dwelling &professional services
- Category-4: Agriculture, forestry & fishing, Trade, repair, hotels & restaurants, Manufacturing, Real estate, ownership of dwelling &professional services.

Provide at least two recommendations for each category to improve the per capita GDP.

Category-1

Given below are two recommendations for C1 category

- Almost all the states in this category are famous tourist destinations. Hence, government should focus more on the ‘Manufacturing’ sector as well as ‘Agriculture’ and ‘Trade, repair, hotel & restaurant’ sectors so, that we can more and more tourists and increase the GSDP.
- Other services and services related to ‘Transport, storage, communication and broadcasting’ have a moderate contribution to the GSDP and hence a little more focus on that side may improve the gain from those sectors as well.

Category-2

Given below are two recommendations for C2 category

- Looking at the states falling in this category, the major industries including IT are located here and we may assume this is the reason behind increase in the GSDP from real estate sector as the population would be more in these states. The government should focus more on ‘Manufacturing’, ‘Real estate’, ‘Agriculture’ and ‘Trade, repair, hotels and restaurants’ sector.
- Same as in category-1, a little more attention towards the ‘other services’, ‘financial services’ and ‘Transport, storage, communication and broadcasting.’ can improve the GSDP coming from that side.

Category-3

Given below are two recommendations for C3 category:

- We can infer from the states in this category that number of industries or companies are less compared to other states and hence more policies or schemes should be implemented to increase the revenue from that side.
- There is a sudden drop from the ‘Agriculture’ sector to ‘Manufacturing ‘sector. Government should focus more on manufacturing sector as once it improves, automatically there would be increase in the trade and real estate sector.
- One more thing that we notice here is that the contribution of ‘Mining &quarrying’ sector to GSDP is more in this category compared to remaining 3 categories, so they may have potential to bring in more revenue.

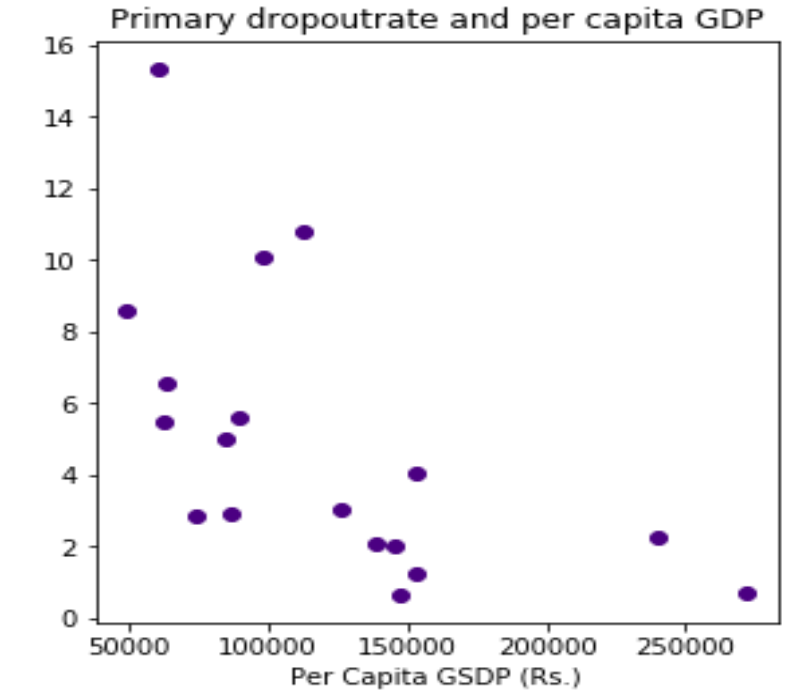
Category-4

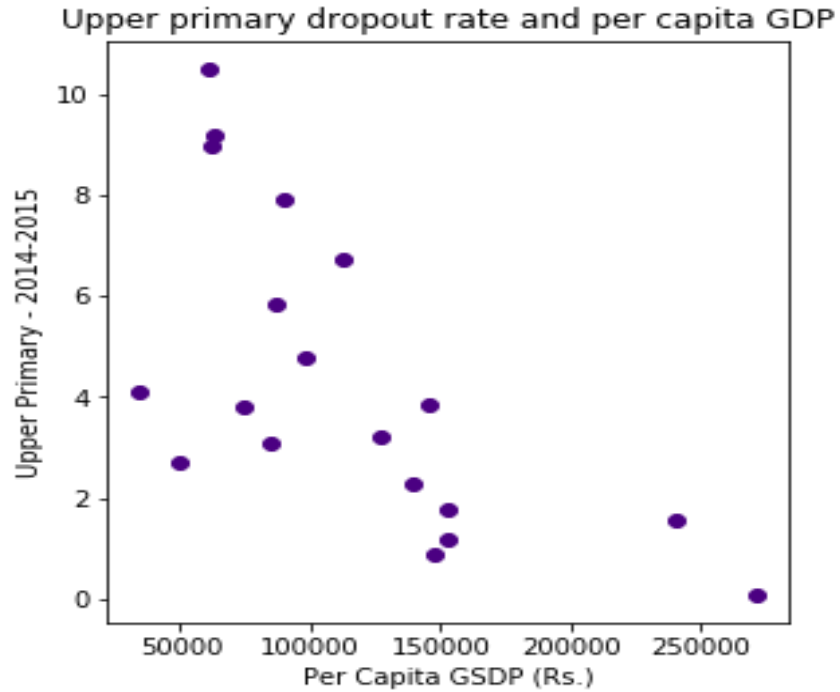
Given below are two recommendations for C4 category:

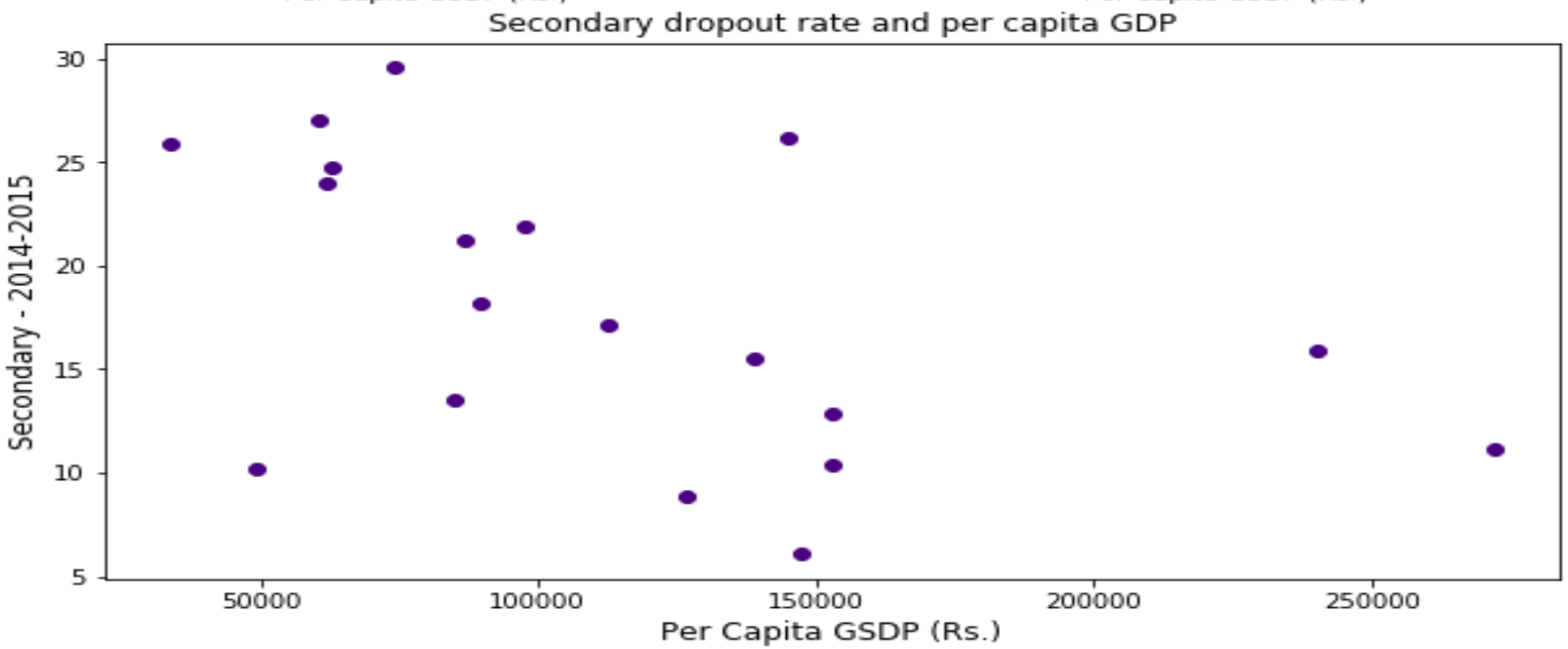
- There is a sudden drop from ‘Agriculture’ to ‘Trade, repair, hotels and restaurants’ sector. Government should try to maintain this by implementing more schemes and policies.
- Even here, the ‘Manufacturing’ and ‘Real estate’ sector seem like having the potential to bring in more revenue, so some focus there would improve the GSDP.

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


Primary dropoutrate and per capita GDP
16
14
12
10
8
6
4
2
0
50000
100000
150000
200000
250000
Per Capita GSDP (Rs.)


Upper primary dropout rate and per capita GDP
Upper Primary - 2014-2015
10
8
6
4
2
0
50000
100000
150000
200000
250000
Per Capita GSDP (Rs.)


Secondary dropout rate and per capita GDP
Secondary - 2014-2015
30
25
20
15
10
5
50000
100000
150000
200000
250000
Per Capita GSDP (Rs.)

- The above shown scatter plots show the correlation between Primary, Upper Primary and Secondary dropout rates with Per Capita GDP. It is clearly seen
- that all of them have a negative linear relationship as high GDSP value has lowest dropout rate and vice versa.
- In all the above plots, if we imagine a line joining the middle points, it will be easier to understand that most of the points do not lie too close to the line, hence we can also say that the relation is not very strong.
- In nutshell, we can say that in order to increase the per capita GDSP, the dropout rates should come down in all the three levels of education.

Form at least one reasonable hypothesis for the observations from the data.

- The tertiary sector has the maximum contribution to GSDP for the states followed by Secondary and Primary being the lowest. Now, if we delve deeper and check which sub-sectors are contributing more to GSDP for states, it was found that following sub-sectors are correlated with high GDP:
- Manufacturing
- Agriculture, forestry & fishing
- Real estate, ownership of dwelling & professional services
- Trade, repair, hotels & restaurants

- The categories 1 & 2 which have the states falling into high GSDP are getting more revenue from ‘Manufacturing’ sectors.
- So, government should focus more on improving these sectors. These require skilled people and as we have already seen that the dropout rates of primary, upper primary and secondary and negatively correlated with the GDP. So, we should come up with new policies which will help in decreasing the dropout rates and thereby increasing the GDP.
- Some investment in the education field, like affordable fee or free education at least till secondary will motivate more people to complete their education. Also, improving the quality of the education and infrastructure in the government run schools will help in achieving the same.