GDP ANALYSIS

-Archit Srivastava

PART – IA

Data Exploration

- The data consists of Items, Duration , and states' & UT's names
- Within Items, 2 values (GSDP at current prices and % growth over previous year) are given for years 2011-12 to 2015-16

Data Cleaning and Manipulation

- Finding for years 2016-17, majority are nulls, so dropping rows.
- Getting West Bengal has all NA values, so dropping that column.
- Since we need only % growth, so filtering other data out
- Also we need plotting statewise % growth, so we need to transpose the dataframe
- There are some NA values in 2015-16 column for some states, but we can't impute them with mean/median as it'll not be right way, so leaving them NA.

% Growth over previous year for states

From the adjoining plot, we can see that mainly NorthEastern states have high growth rates

Whereas Goa, Meghalaya, Odisha, Sikkim have relatively lower growth rates.

Curiosity Exercise: My state UP is fairly performing well above national avg growth


Avg Growth of States' GSDP
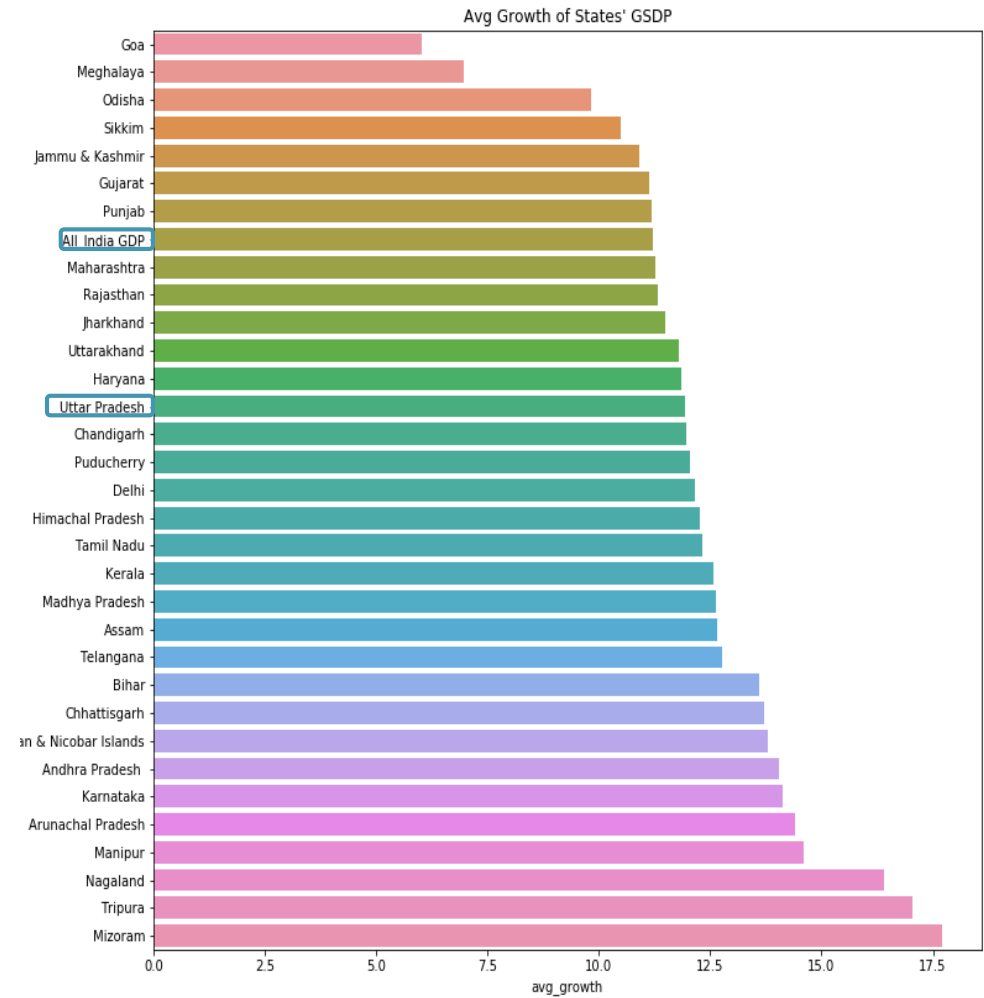
Total GDP of states for year 2015-16

- We shall look at the size of GDP through adjacent plot for all the states.
- Though some values were not available, we could have imputed them based on some methods, but it would rather be dropped from analysis.
- Since values differ on large scale, we are rather taking log of values to compare states
- We can see that 5 highest GDP states are as follows:
    1. Tamil Nadu
    2. UP
    3. Karnataka
    4. Gujarat
    5. Andhra Pradesh
- Bottom-5 consist 2 Union Territories
    1. Sikkim
    2. Arunachal Pradesh
    3. Puducherry
    4. Meghalaya
    5. Chandigarh

ANALYSIS:

- High GDP can be attributed to large size of states and better administration in some others
- While Low GDP is mainly due to small size and less accessibility to other parts of country
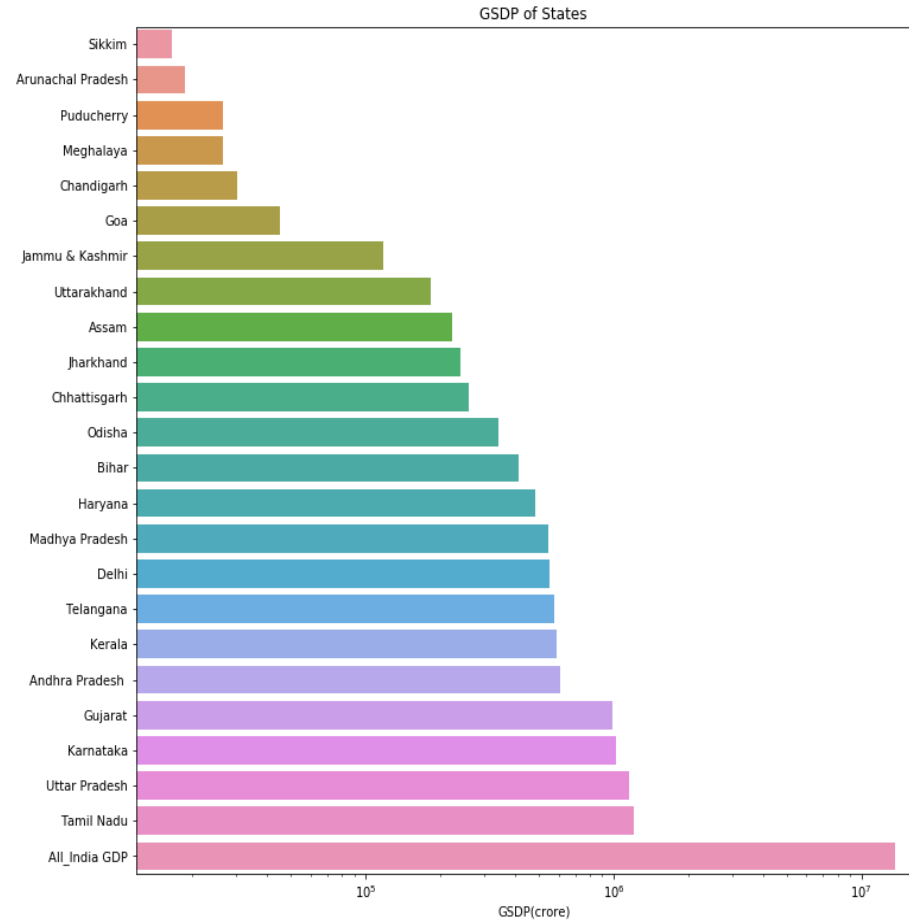
PART – I-B

Data Exploration

- The data consists of multiple files of each state having sector/subsector wise GSVP contribution for years 2011-12 to 2015-16

Data Cleaning and Manipulation

- For each state data, we extract 'Item' and '2014-15' columns only
- Some values in 'Item' have * appended at end, we must remove them for not having conflict during merging
- Have to add state column and transpose the data
- Combine all dataframes of states.
- Clear all data pertaining to Union territories after merging for all states.
- There aren't any NA values in the combined dataframe

Plotting GDP per capita for all the states

Analyzing adjacent figure.

❑ We can find the top performing states as
1. Goa
2. Sikkim
3. Haryana
4. Kerala
5. Uttarakhand

❑ While lowest states
1. Bihar
2. Uttar Pradesh
3. Manipur
4. Assam
5. Jharkhand

Analysis:
- Moreover the difference is large among the states,
- This can be seen as the ratio of max and min comes out to be approx. **8**
- This creates a concern on uneven distribution of talent, administration and money across the people of different states in India
- High per capita GDP maybe due to less population with high production
- While low maybe attributed to large unskilled population and lack of industries etc.

Per capita GDP of States
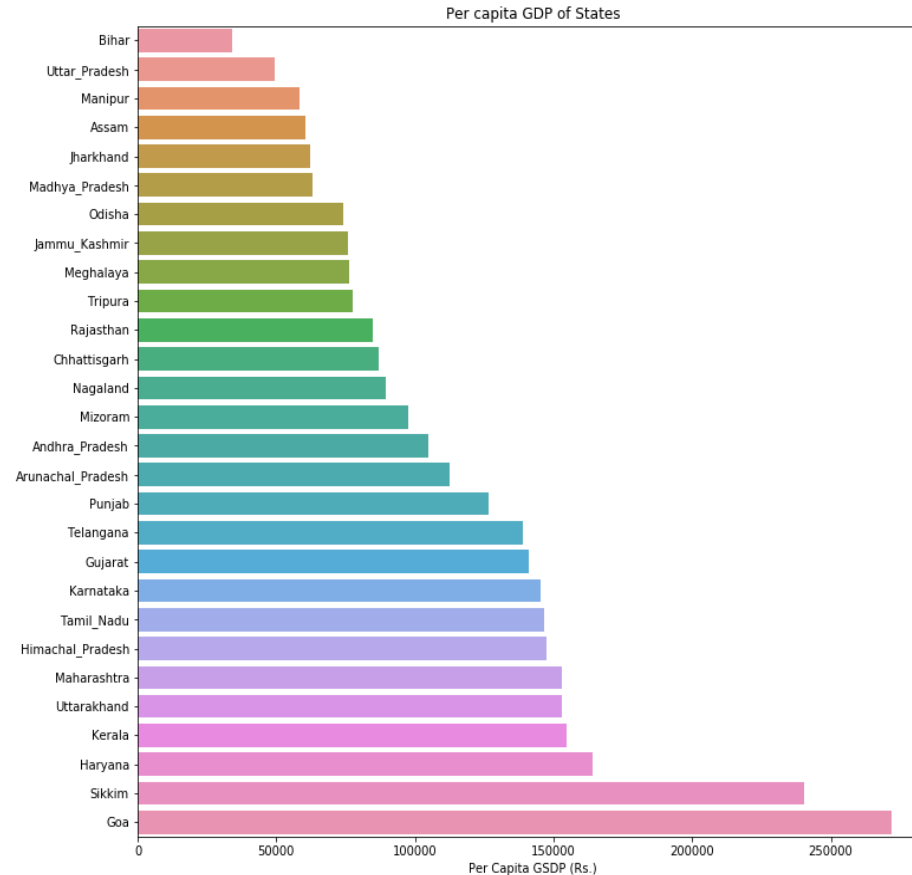
Percentage contribution of the primary, secondary and tertiary sectors
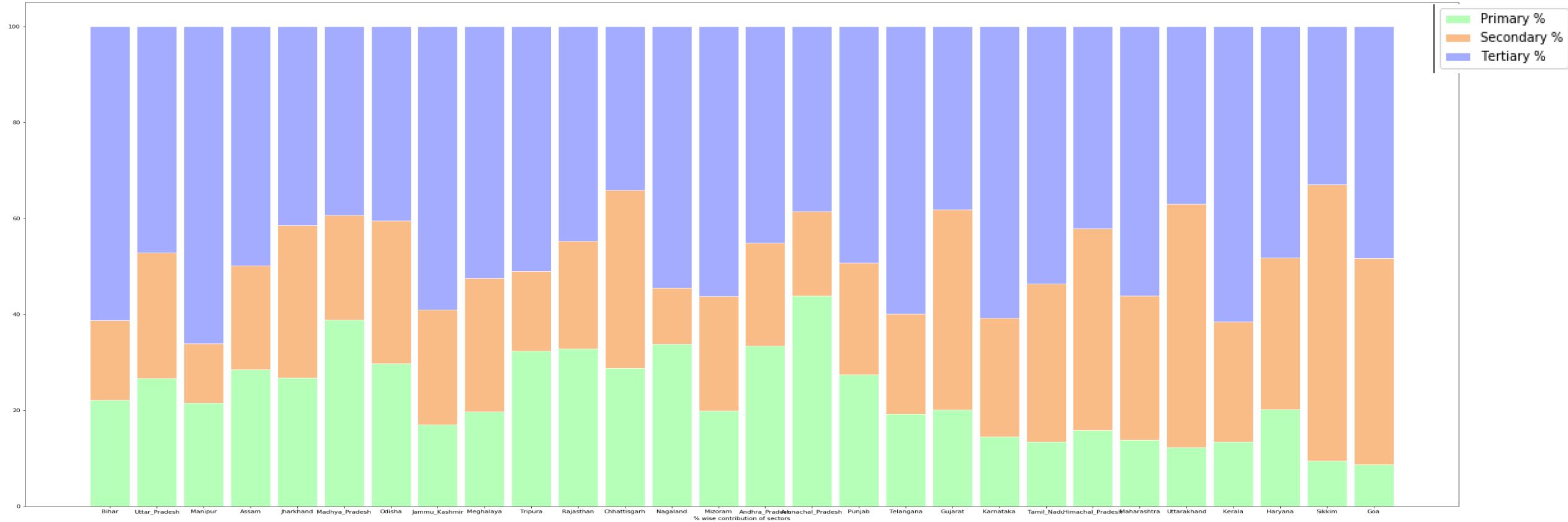
Percentage contribution of the primary, secondary and tertiary sectors (CONTD.)

► From Previous graph, we can see that states having high per capita GDP states, majority of sector is Secondary and Tertiary.

Primary is least contributing sector for these

► Whereas, for lower per capita GDP states, more percentage is of primary sector giving an impression on over dependence on the primary sector and not developing the secondary and tertiary sectors
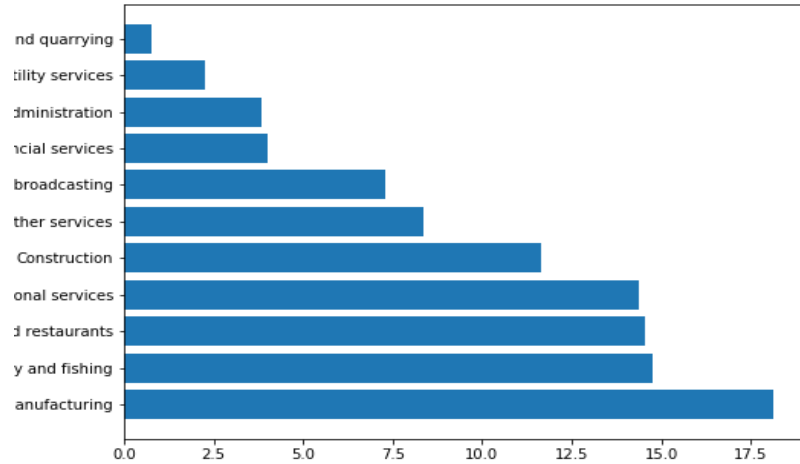
Categorise the states into four groups based on the GDP per capita

▶ Let's now categorise states on basis of which quantile they fall into:
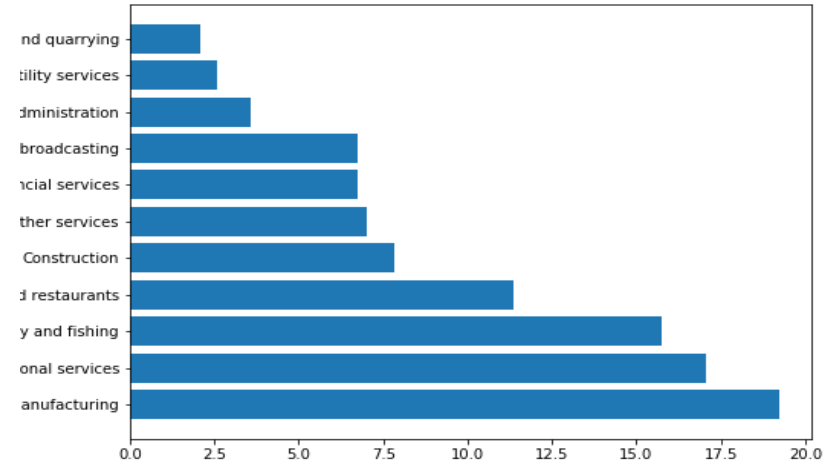
Class	Criteria for GDP
C1	Quantile(0.85,1)
C2	Quantile(0.5,0.85)
C3	Quantile(0.2,0.5)
C4	Quantile(0,0.2)

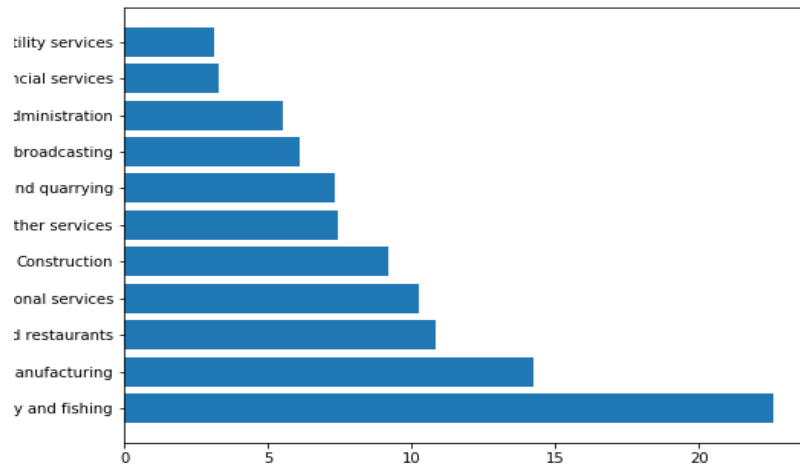# Find the top subsectors contributing towards the GDP of different categories of states
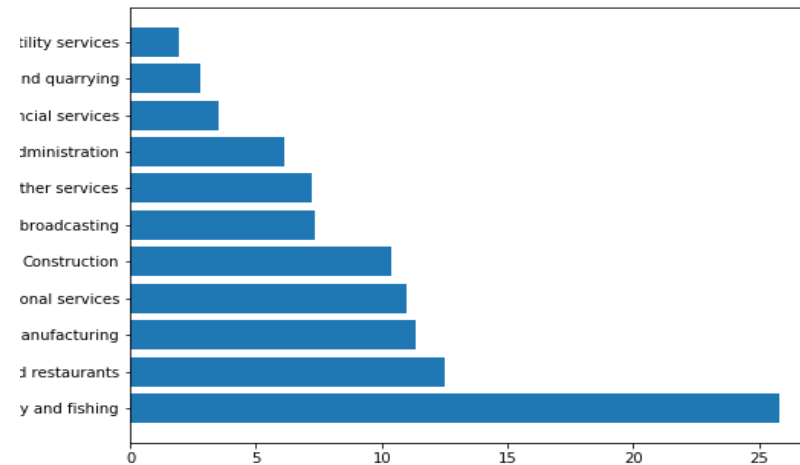
Find the top subsectors contributing towards the GDP of different categories of states (CONTD.)

▶ Top categories contributing to approx. 80% of GSDP in categories

Category	Top sub sectors(contr. 80%)
C1	'Manufacturing', 'Agriculture, forestry and fishing', 'Trade, repair, hotels and restaurants', 'Real estate, ownership of dwelling & professional services', 'Construction', 'Other services'
C2	'Manufacturing', 'Real estate, ownership of dwelling & professional services', 'Agriculture, forestry and fishing', 'Trade, repair, hotels and restaurants', 'Construction', 'Other services', 'Financial services'
C3	'Agriculture, forestry and fishing', 'Manufacturing', 'Trade, repair, hotels and restaurants', 'Real estate, ownership of dwelling & professional services', 'Construction', 'Other services', 'Mining and quarrying'
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'

▶ It can be easily seen that C1,C2 states have mix % of Manufacturing, Agriculture, Trade, Real State and Construction as top contributors.

▶ This is lacking in C3 and C4 states where main categories are Agriculture only. The next big subsectors in C3/C4 are contributing approx. half of what agriculture is contributing

Analysis

- We can clearly see that states in high bracket of GDP have more manufacturing, are not wholly dependent on primary sector activities.
- Thus it can be concluded that more emphasis on Secondary and Tertiary activities is required for betterment of GDP of states, ie, secondary sectors are needed to be paid attention to if we want to increase per capita GDP.

Recommendations for each category

Category	Recommendations
C1	1.Doing well in manufacturing but can improve in construction services 2.Tertiary services can be improved 3.Agricultural sector can also be improved a little
C2	1.Primary sector services can be improved a little. 2.Construction sector must be improved 3.Tertiary sector like transport etc are lacking behind and can be upgraded
C3	1.Need to improve manufacturing and construction sectors 2.Also need to improve promising tertiary sectors like real estate and trade. 3.Decouple with overdependence in primary sector
C4	1.Need more focus on manufacturing and other secondary sectors like construction leading to more infrastructure 2.Also need more attention on tertiary sectors like trade(ease of trade policies) 3.Shift more people towards these sectors through upskilling.

PART II

Data Exploration:

▶ The data consists of dropout rates for different years for different levels of education for each state/UT.

Data Cleaning and Manipulation

➢ Primary dropout for year 2013-14 was wrongly named 2014-15, had to rename 2014-15.1 to 2014-15

➢ Mismatch in names of states with data in section I-B

➢ Have to merge data with left join with data in I-B

➢ Drew scatter plot with regression line to view correlation between different dropout rates and per capita GDP

Correlation of GDP per capita with dropout rates in education

Let's have correlation plots of different levels of education with GDP per capita

GDP/capita Correlation with	Primary	Upper Primary	Secondary
Value	-0.5599	-0.5457	-0.4594
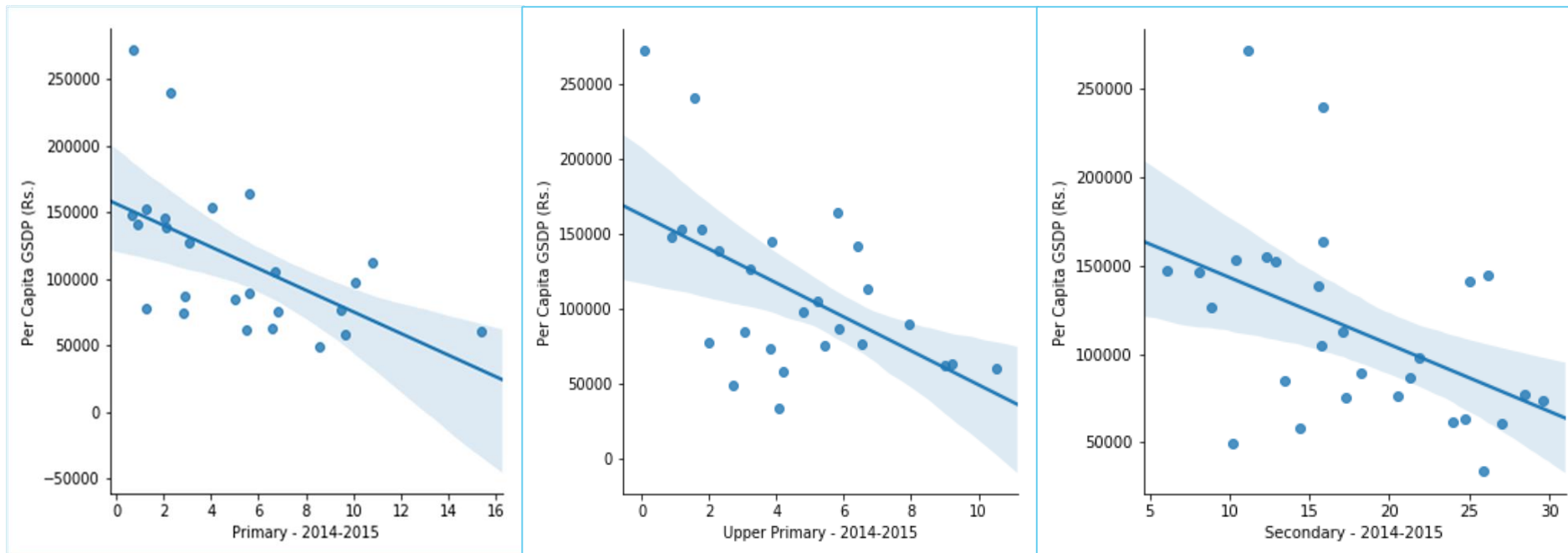
Hypothesis for negative correlation of Dropout rates with GDP/Capita

- ▶ Due to high dropout rates, skilled and educated workforce is not available

- ▶ Thus this creates shortage to contribution made by Tertiary and Secondary sectors.

- ▶ Also since people remain uneducated, there is a chance of disguised unemployment in primary sector, leading to overdependence on agriculture

- ▶ This also is responsible for poor health and innovation indices of a state if the masses are less educated.

Thank You ☺