GDP Analysis Assignment- Upgrad

Data Exploration

Data set 1 - contains GSDP and % Growth over previous year for all Indian states for the year 2012-2013, 2013-2014, 2014-2015,2015-2016, 2016-2017

- GSDP - CURRENT PRICES (` in Crore)
- (% Growth over previous year)

Data set 2 - contains the distribution of GSDP among three sectors: the primary sector , the secondary sector and the tertiary sector along with taxes and subsidies for all the states and union territories

- Manufacturing',
- 'Agriculture,
- forestry and fishing', 'Real estate, ownership of dwelling & professional services',
- 'Trade & repair services', 'Construction',
- 'Other services',
- 'Transport, storage, communication & services related to broadcasting',
- 'Financial services',
- 'Public administration', 'Electricity, gas, water supply & other utility services',
- 'Mining and quarrying

Data set 3 – contains Drop out rate for education (primary, upper primary and secondary) for the year 2014-2015 for each state

- Primary
- Upper Primary
- Secondary
- Senior Secondary

Problem Statement

Problem Statement:

1. Calculate the average growth of states for the duration 2013-14, 2014-15 and 2015-16
2. The percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states.
3. 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.
4. 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.

Based on the problem statement,

we should focus on GSDP for each state for year 2012-2016 (ignoring 2016-2017) for all states and Union territoies , GDP per capita for all states , sector and sub sector for all states for the year 2014-2015 (ignoring Union Territoies) and drop out rate for all education categories for year 2014-2015 (ignoring senior secondary)

Data Cleaning and Manipulation

Data inconsistencies:

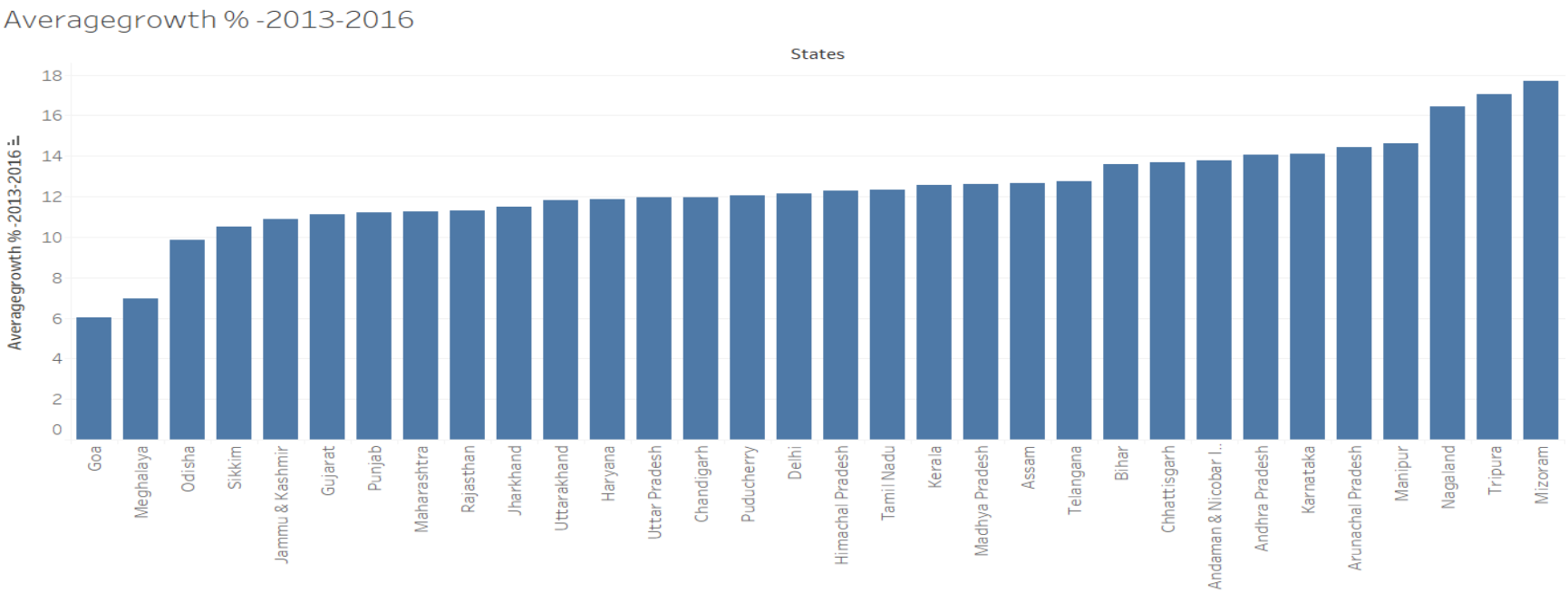
- NA values in GSDP data for the year 2016-2017
- NA values in the West Bengal column

Other Issues:

- State names are named in a different way in different data sets, eg- ***Uttarkhand*** in data set-1 is named as ***Uttrakhand*** in data set-3 which may lead to mismatch of data while merging/joining data from these two data sets.
- Duplicate column names/ wrong column names on drop out rate data set which needs to be corrected
- Data need appropriately transformed, if necessary, for easy analysis and plotting the data

Part-I: GDP Analysis of the Indian States

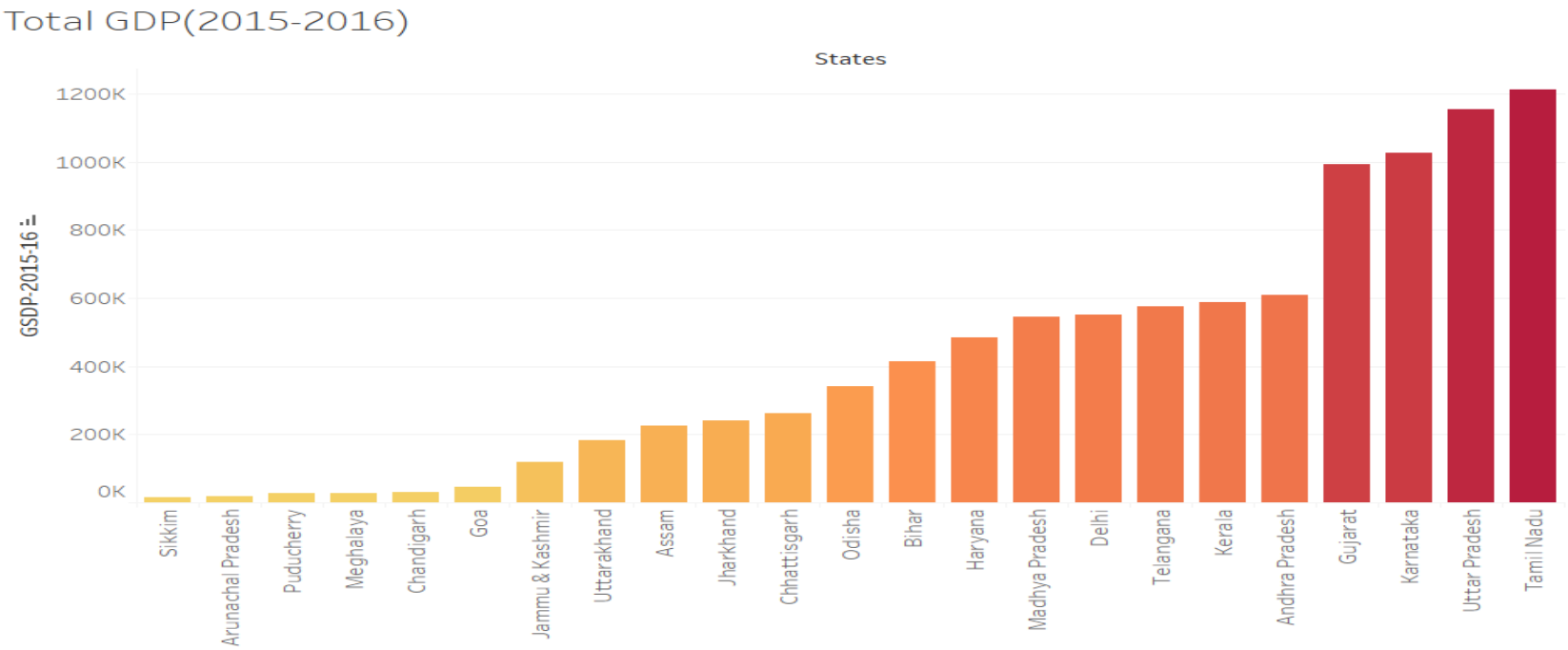
Average growth of states for the duration 2013-14, 2014-15 and 2015-16



From the previous plot, it is evident that **Mizoram, Tripura, Nagaland** are consistently growing states while **Goa, Meghalaya and Odisha** are struggling over the years 2013 to 2016

Part-I: GDP Analysis of the Indian States

Top 5 and the bottom 5 states based on total GDP (2015-2016).



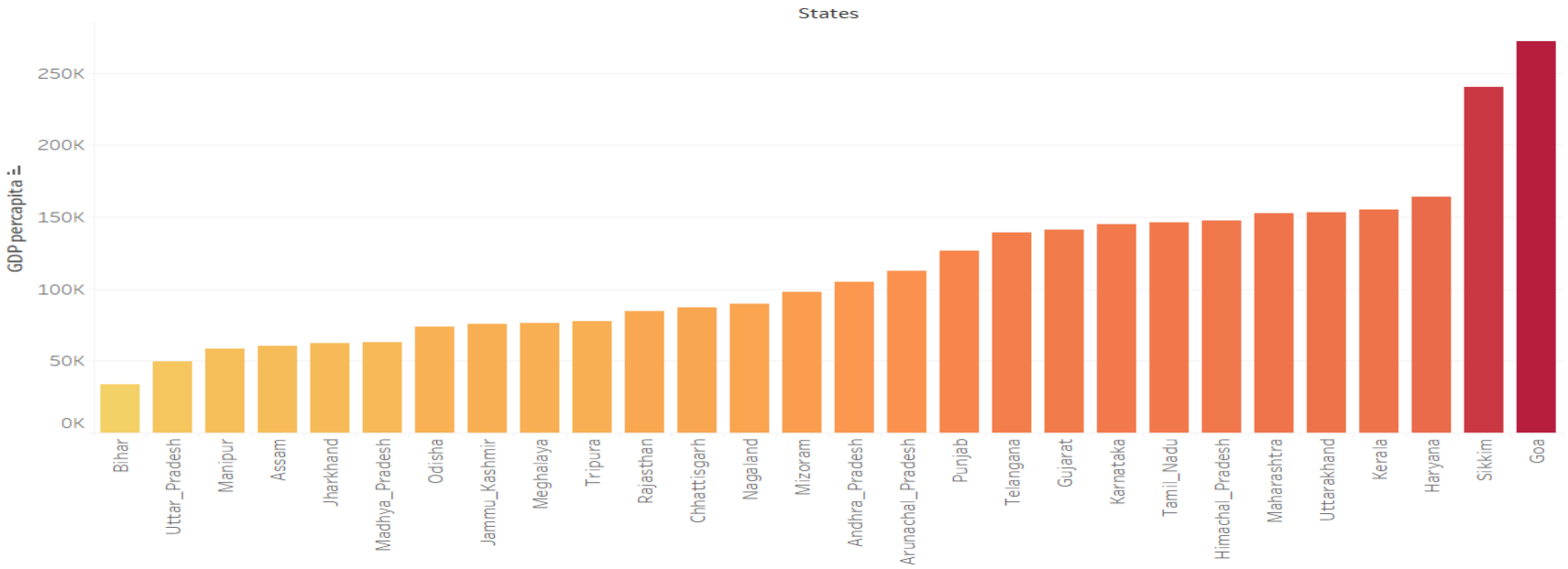
Bottom 5 states with the lowest GDP

Sikkim
Arunachal Pradesh
Puducherry
Meghalaya
Chandigarh

Top 5 states with highest GDP

Tamil Nadu
Uttar Pradesh
Karnataka
Gujarat
Andhra Pradesh

Part I-B – GDP per capita for all the states(2014-2015)



The bottom 5 states with lowest GDP per capita :

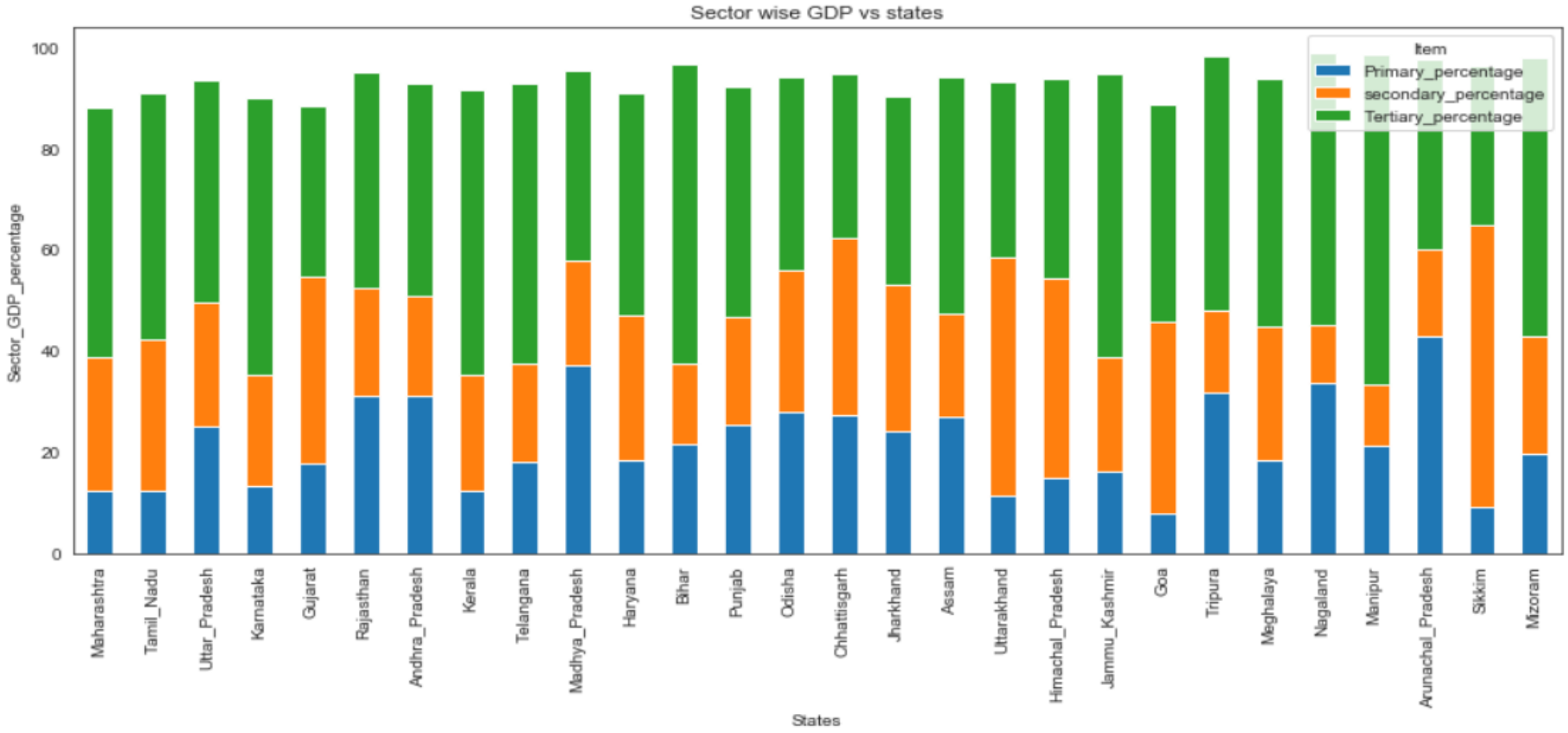
Bihar
Uttar_Pradesh
Manipur
Assam
Jharkhand

The top 5 states with highest GDP per capita :

Goa
Sikkim
Haryana
Kerala
Uttarakhand

The ratio of the highest per capita GDP to the lowest per capita GDP : 8

Part I-B - Percentage contribution of the primary, secondary and tertiary sectors as a percentage of the total GDP for all the states



Inference:

As we could see from the stacked bar plot, it is evident that secondary and tertiary sectors contribute more towards the total GSDP for all the states

Part-1B -Binning states into 4 categories

The states are divided 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)

States under category C1 (quantile between 0.85 and 1) :

Goa ,Haryana ,Kerala ,Sikkim, Uttarakhand

States in category C2 (quantile between 0.5 and 0.85):

Andhra Pradesh, Arunachal Pradesh ,Gujarat ,Himachal Pradesh ,Karnataka, Maharashtra, Punjab ,Tamil Nadu ,Telangana

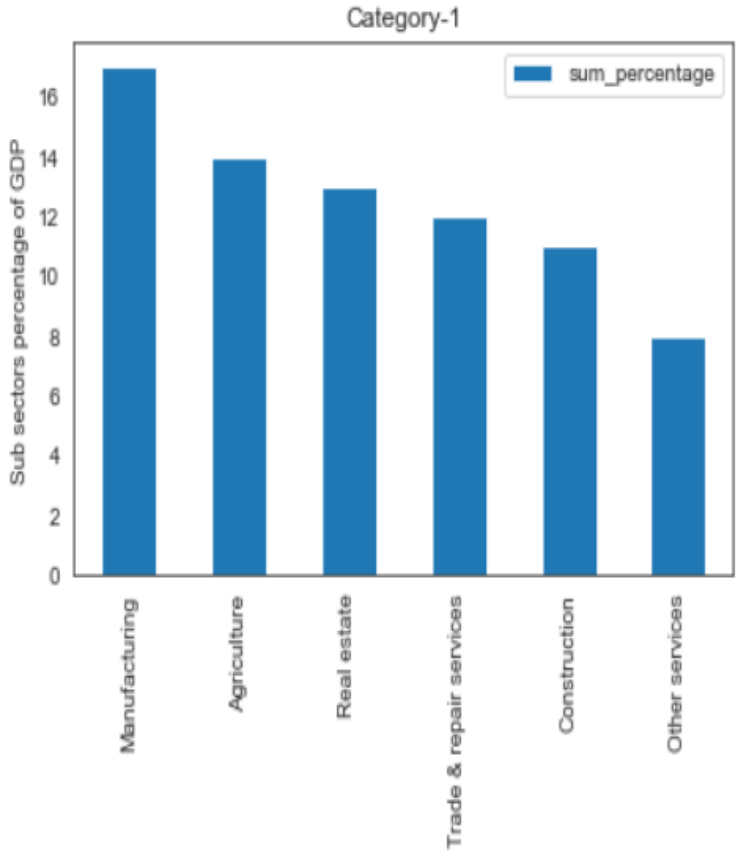
States in category C3 (quantile between 0.2 and 0.5):

Chhattisgarh, Jammu Kashmir, Meghalaya ,Mizoram, Nagaland, Odisha, Rajasthan, Tripura

States in category C4 (quantile less than 0.2):

Assam, Bihar, Jharkhand ,Madhya Pradesh, Manipur, Uttar Pradesh

Part-IB -The Top **sub-sectors** that contribute to approximately 80% of the GSDP of each category -1.

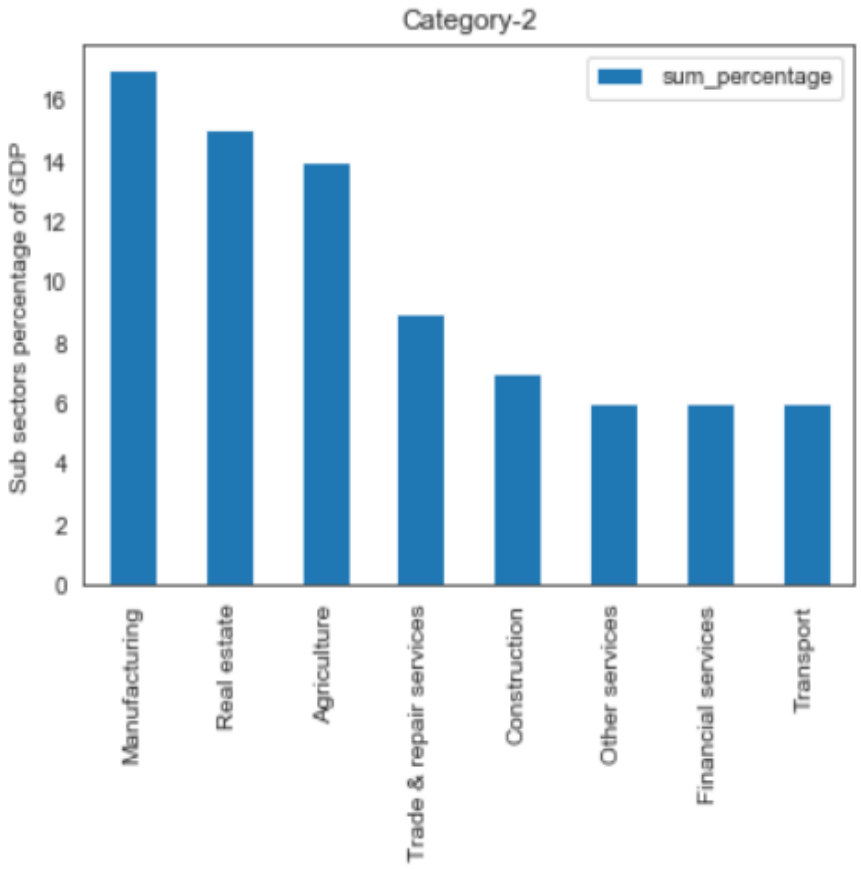


Category 1:

1.Category 1 states has to focus on categories like Construction, Trade, Hotels and restaurant as improving these will help increase the per capita GDP

2.Also focus on development goals on Manufacturing and agriculture as it contributes to the maximum GDP per capita

Part-IB -The Top **sub-sectors** that contribute to approximately 80% of the GSDP of each category-2

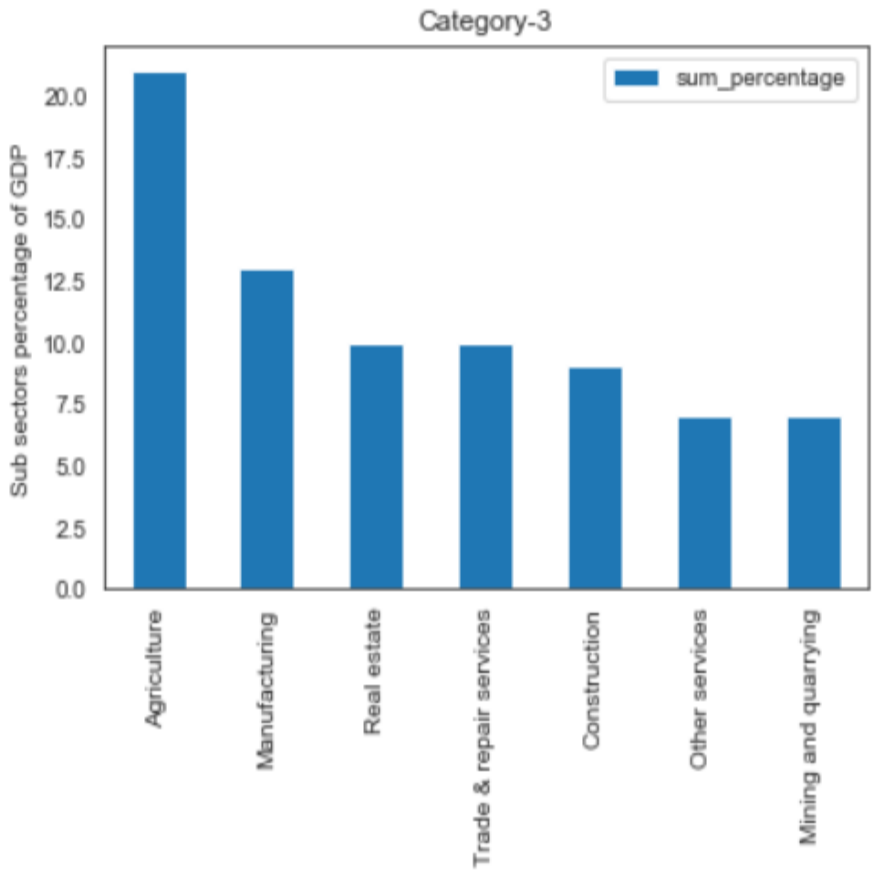


Category 2:

1.Category 2 states has to focus on categories like Financial services and transport and devise possible goals to improve these subsectors

2.Real estate being a greatest contributor in GDP per capita for category 2, they need to involve more strategic plans on maintaining the same.

Part-IB -The Top **sub-sectors** that contribute to approximately 80% of the GSDP of each category-3.

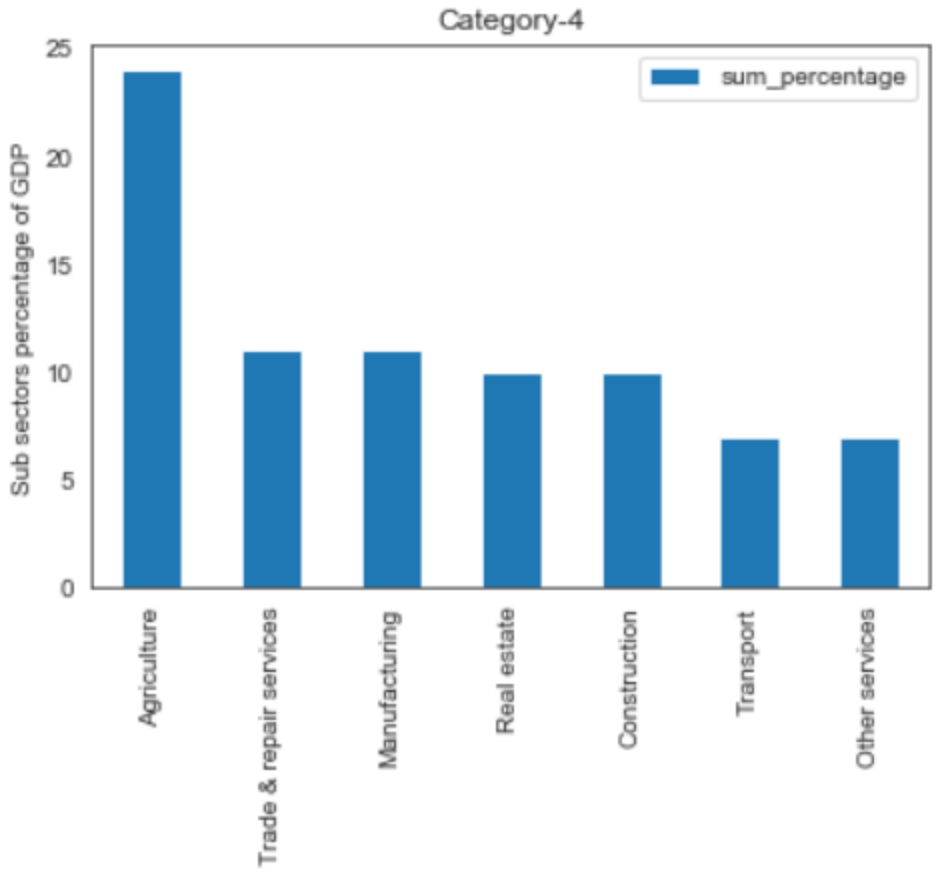


Category 3:

1.Agriculture being the greatest contributed subsector in category 3,the increase productivity in agriculture is needed through appropriate policies and strategies to affect a convergence of labour productivity across agriculture to maintain high GDP per capita

2.Category 3 states has to concentrate on construction, mining and quarrying as these sub sectors seem to be a potential way to increase GDP per capita

Part-IB -The Top **sub-sectors** that contribute to approximately 80% of the GSDP of each category-4.



Category 4 :

1.Agriculture seem to be the major contribution, while providing Education will help improve the GDP per capita of other sub sectors as well.

2.Category 4 states should concentrate more on all possible sectors like Transport, Construction and Real estate to improve their GDP per capita.

Part-1B

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

GDP distribution for category 1 and 2 are highly concentrated in secondary sector like Manufacturing, while GDP distribution for category 3 and 4 (bottom states) are concentrated on Primary sector like Agriculture.

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

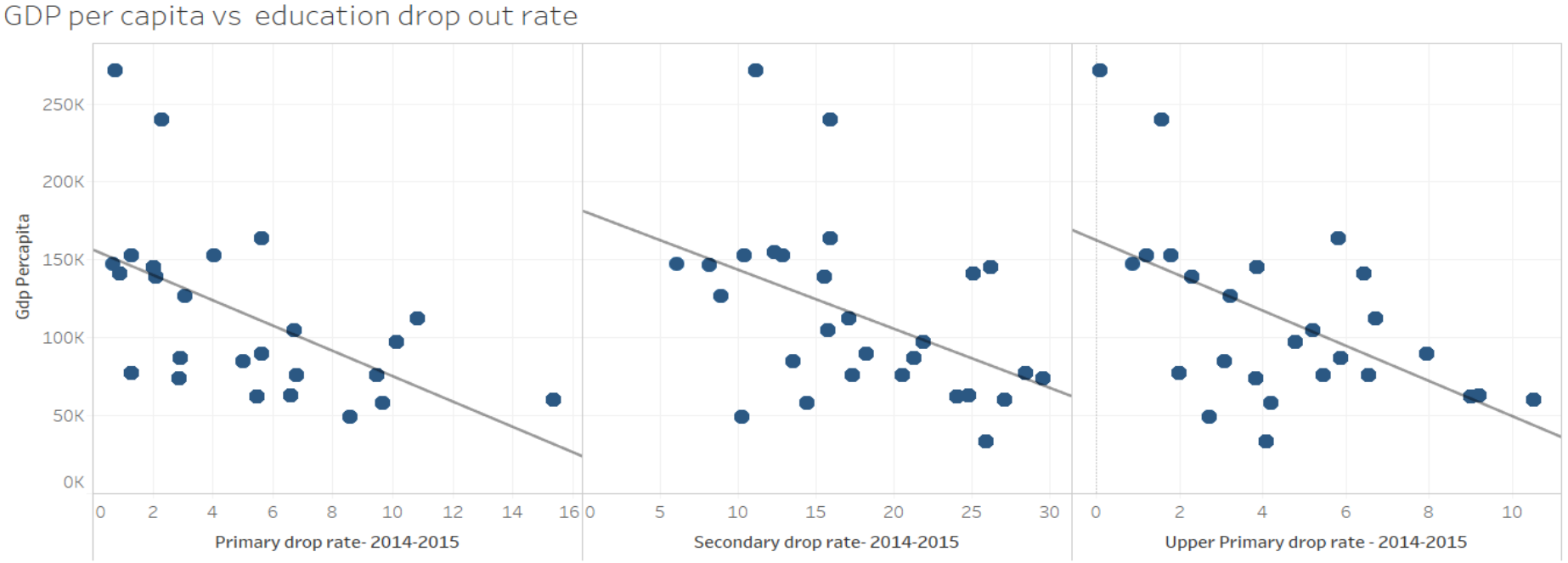
1.Agriculture, forestry and fishing from Primary sector.
2.Manufacturing from Secondary sector.
3.Trade,Hotels,Restaurants and Real estate from Tertiary sectors
are mostly correlated with high GDP

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

Category 1 - Construction, Trade, Repair,Hotels and Restaurants
Category 2 - Financial services , Transport
Category 3 - Construction, Mining and Quarrying
Category 4 - Construction, Transport

Part-II: GDP and Education Dropout Rates

Correlation of GDP per capita with dropout rates in education (primary, upper primary and secondary) for the year 2014-2015



key insights :

As we can see from the previous scatter plot, it is clearly obvious that there is relationship between GDP per capita and drop out rate, when dropout percentage increases the gdp per capita for the state decreases. This is similar in all the three plots which means **GDP per capita of state decreases when drop out rate increases"**. Hence, **GDP per capita is negatively correlated with dropout rate**