# Assignment: GDP Analysis

Part 1:GDP Analysis of Indian States

Part 1 A:

1. This dataset (Data I-A) contains the GSDP (Gross State Domestic Product) data for the states and union territories for the years 2011-12 to 2016-17.
2. As we need to do the analysis only for 2013-14, 2014-15 and 2015-16 years. Remove the unnecessary rows and columns. Remove the NaN values.
3. Calculate the average growth of states over the duration 2013-14, 2014-15 and 2015-16.
   A. Create a new data set which contain only the required data for %Growth over previous year for all the states for the year 2013-14, 2014-15, 2015-16.
   B. Transpose the data set and Calculate the state wise mean of % Growth over previous year.
   C. Sort the data and plot a Bar graph to represent % Growth of each state over 3 years 2013-2016
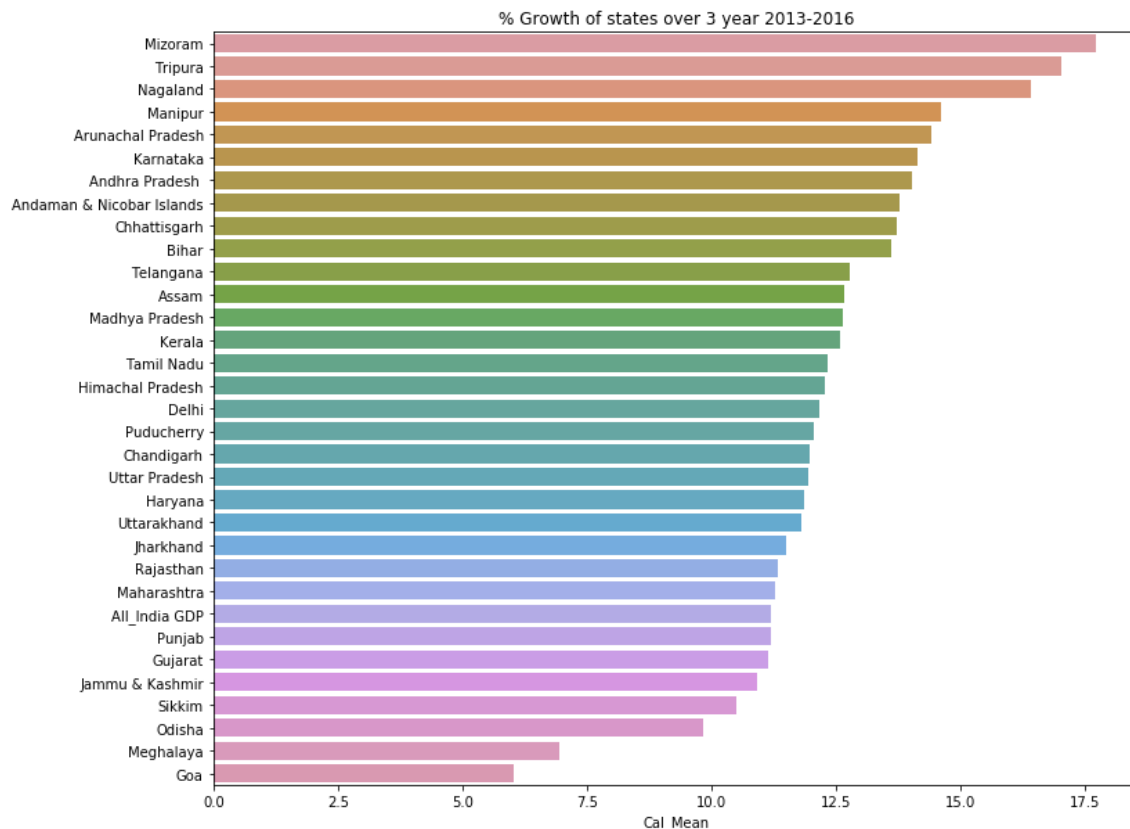

% GROWTH OF EACH STATE OVER 3 YEARS 2013-2016

Observation:

- From the Graph it is clear that, Mizoram is the state growing consistently fast over a period of 3 year and Goa is the struggling state
- My home state is Karnataka and Growth rate of Karnataka is more than the National Average represented by "All_India GDP'.

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

Consider the total GDP of the states data only for 2015-16 and remove rest of the data and remove NaN values. Sort the data and plot a bar graph.
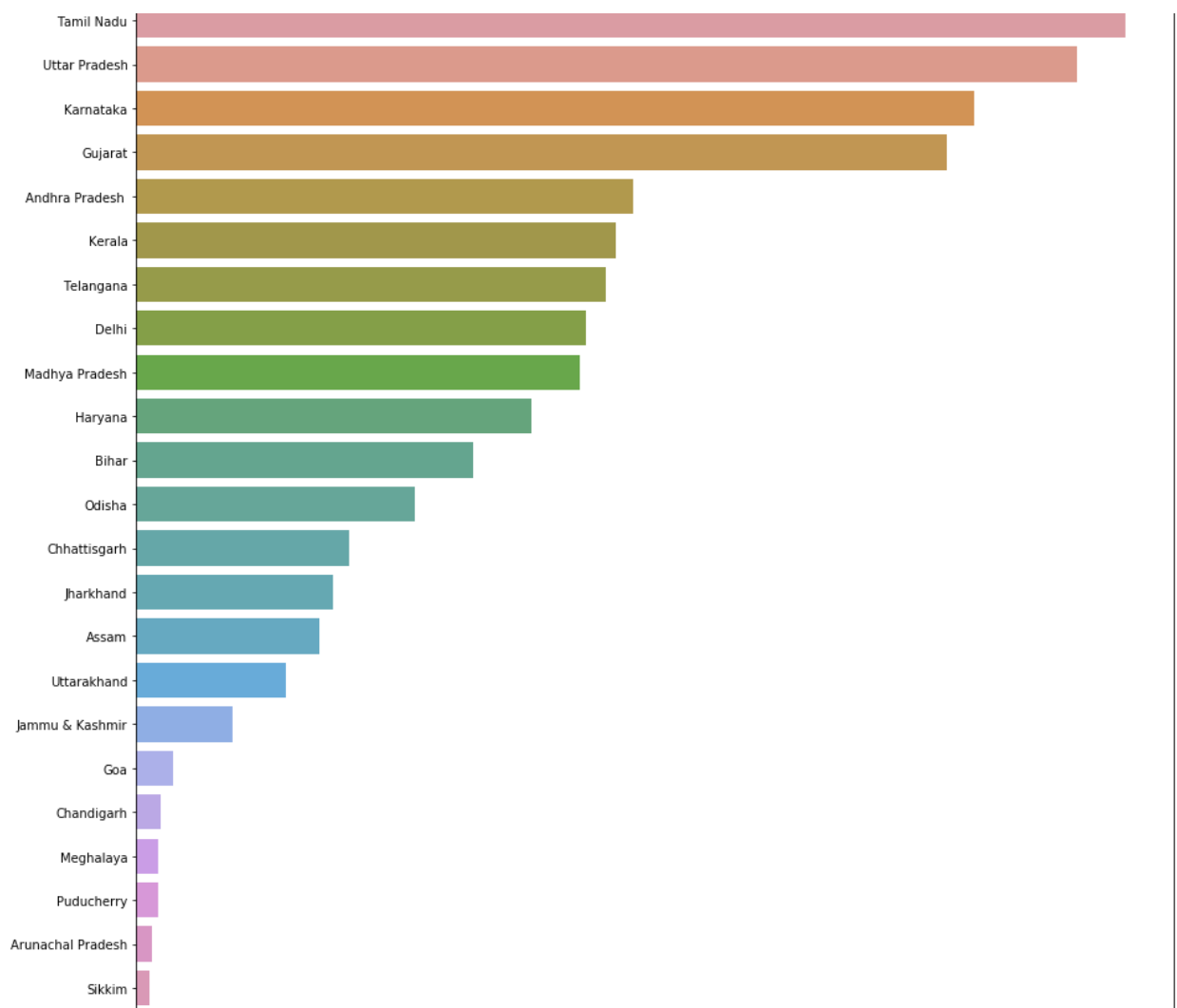
TOTAL GDP OF THE STATES DATA ONLY FOR 2015-16

Observation:

- We can see the top 5 state having higher GDP and and bottom 5 states with lowest GDP for year 2015-16

List of Top 5 states	List of bottom 5 states
Tamil Nadu	Chandigarh
Uttar Pradesh	Meghalaya
Karnataka	Puducherry
Gujarat	Arunachal Pradesh
Andhra Pradesh	Sikkim

Part-IB

1. This dataset(Data I-B) contains the distribution of GSDP among three sectors: the primary sector (agriculture), the secondary sector (industry), and the tertiary sector (services) along with taxes and subsidies. There is separate dataset for each of the states.
2. Read data files for all the state and take out the data only for 2014-15 and merge the data for all the states. Filter out the Union Territories. No need to filter out the NaN value as they are not impacting the Analysis we want to make
3. Plot the GDP per capita for all the states

GDP per capita for all the states for year 2014-15
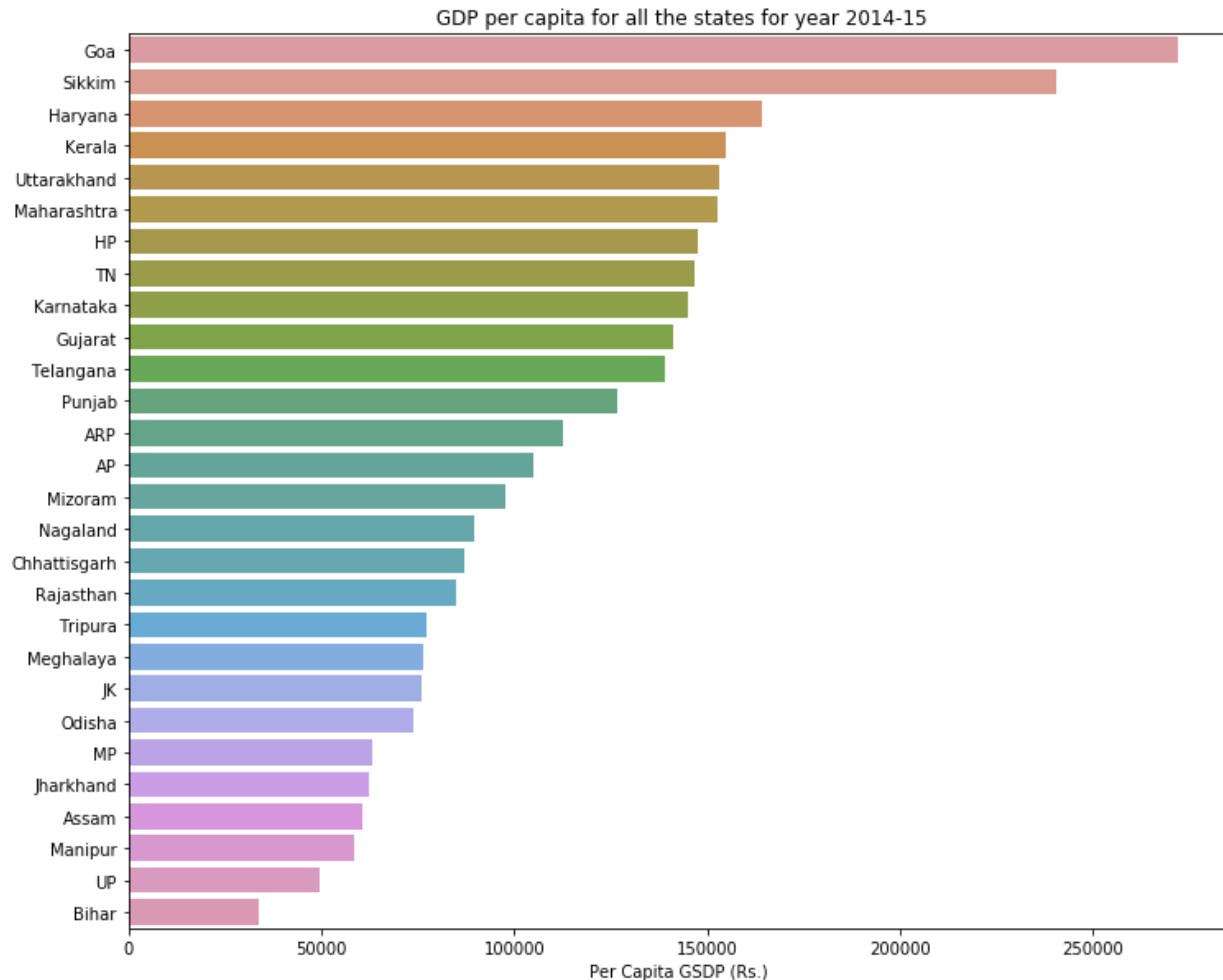
4. Identify the top-5 and the bottom-5 states based on GDP per capita for 2014-15

List of Top 5 states	List of bottom 5 states
Goa	Jharkhand

List of Top 5 states	List of bottom 5 states
Sikkim	Assam
Haryana	Manipur
Kerala	UP
Uttarakhand	Bihar

5. Find the ratio of highest per capita GDP to the lowest per capita GDP. Calculate the ratio by dividing Max value of the column 'per capita GDP' (that is highest per capita GDP is for Goa) and Min value of the column 'per capita GDP' (that is lowest value per capita GDP is for Bihar)

6. Plot the percentage contribution of primary, secondary and tertiary sectors as a percentage of total GDP for all the states. Calculate the % contribution of each sector by dividing the values of sectors with the value of total GDP of state.

Assumption: I have considered the given value of primary, secondary and tertiary sectors and total GDP and not calculated the values by adding all the sectors of different categories.

Observation : Stack plot below shows the % contribution of each sector(primary, secondary and tertiary) in the GDP of each state

% contribution of primary, secondary and tertiary sectors as a % of total GDP for all the states
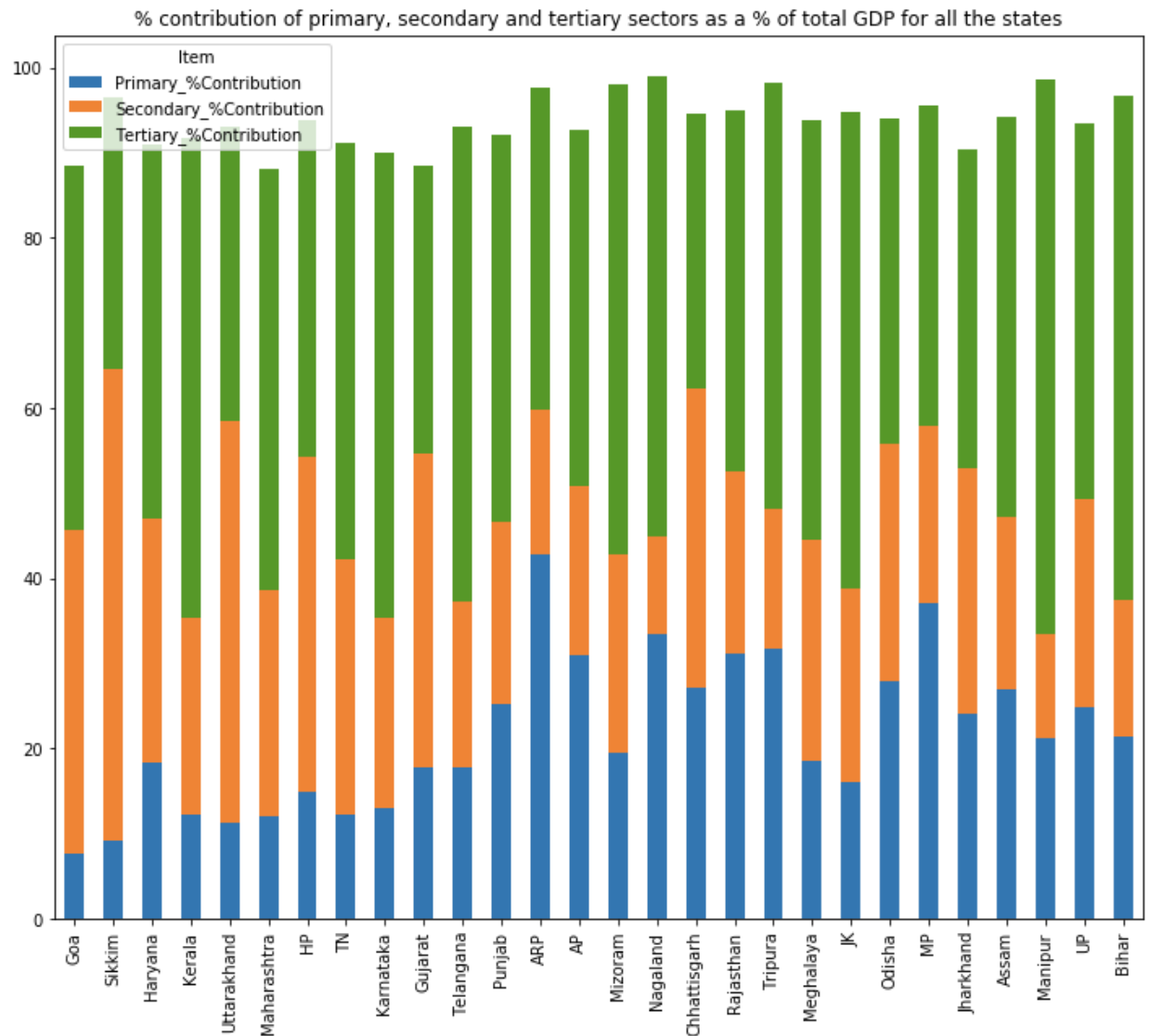
7. Categorize the states into four categories based on GDP per capita (C1, C2, C3, C4) as per the instruction in Assignment (using quantile values) using the qcut function Sort the data frame and group by the category with aggregate function as sum.

8. For each category C1, C2, C3, C4; Find the top 3/4/5 sub-sectors (such as agriculture, forestry and fishing, crops, manufacturing etc.) , which contribute to approx. 80% of the GSDP of each category.
   • As the above data frame is categorical type, New columns cannot be added for further analysis, hence copy this data frame into a new one DS_SECTORS.

- Sort the data frame and Calculate Category wise Cumulative Sum of all the Sectors and put it in a new data frame DS_SECTORS_ALL. now we can Print the list of sectors which contribute to nearly 80% for each Category.

9. Plot the contribution of the sub-sectors as a percentage of theGSDP of each category. Calculate percentage contribution of each category in GSDP for all the sector(100* sector value/GDP) and put the output (float type convert it into Data frame type) in a new data frame DS_ALL for further analysis. plot the 4 bar plot one for each category to show the % contribution of each sector towards GDP.
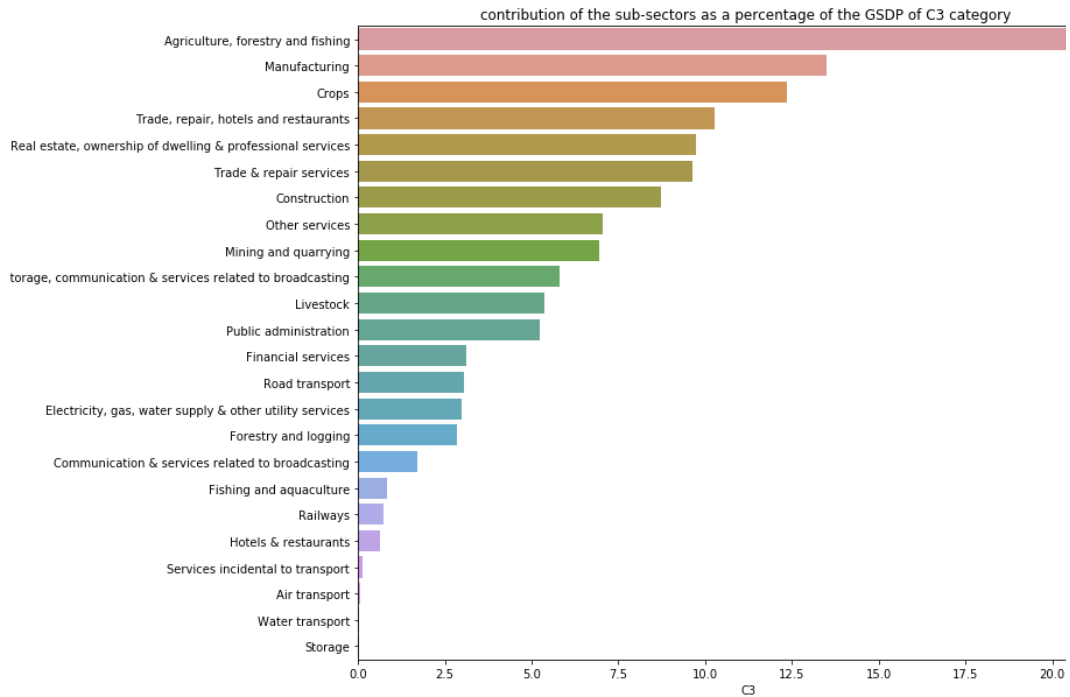

contribution of the sub-sectors as a percentage of the GSDP of C1 category
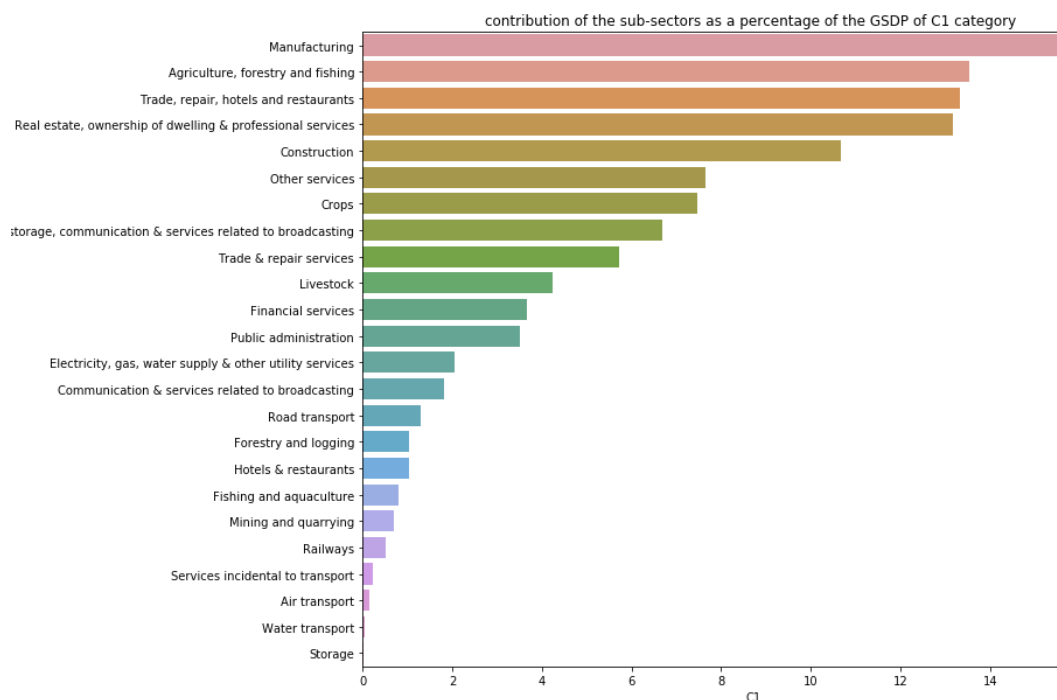
% CONTRIBUTION OF EACH SECTOR TOWARDS GDP FOR C1 STATES

contribution of the sub-sectors as a percentage of the GSDP of C2 category
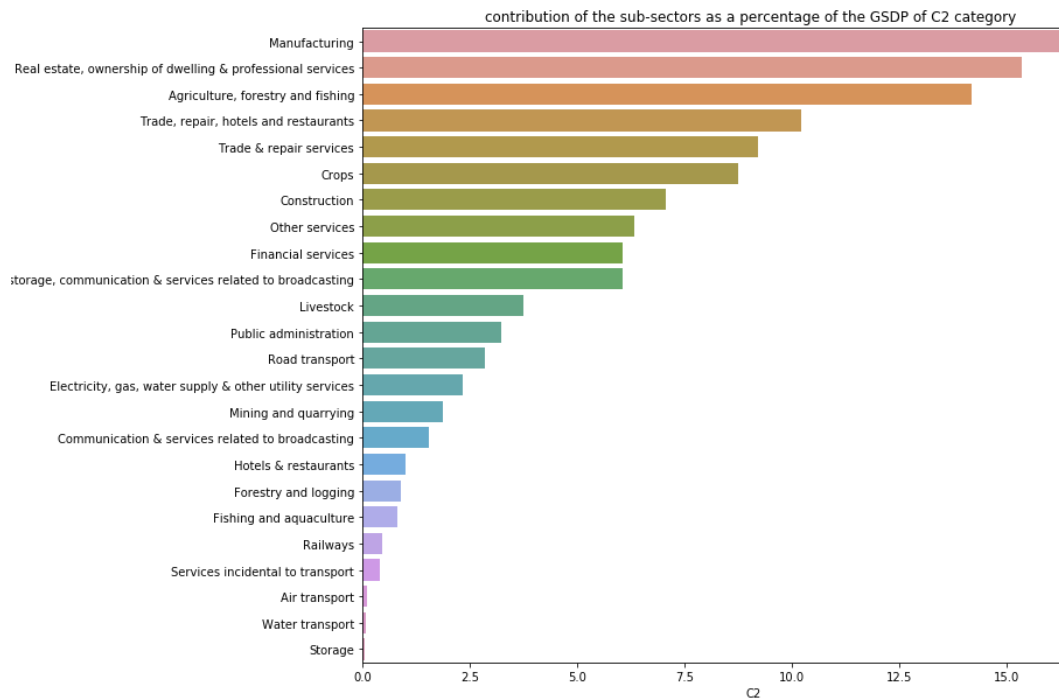
CONTRIBUTION OF EACH SECTOR TOWARDS GDP FOR C2 STATES

contribution of the sub-sectors as a percentage of the GSDP of C3 category

CONTRIBUTION OF EACH SECTOR TOWARDS GDP FOR C3 STATES

contribution of the sub-sectors as a percentage of the GSDP of C4 category
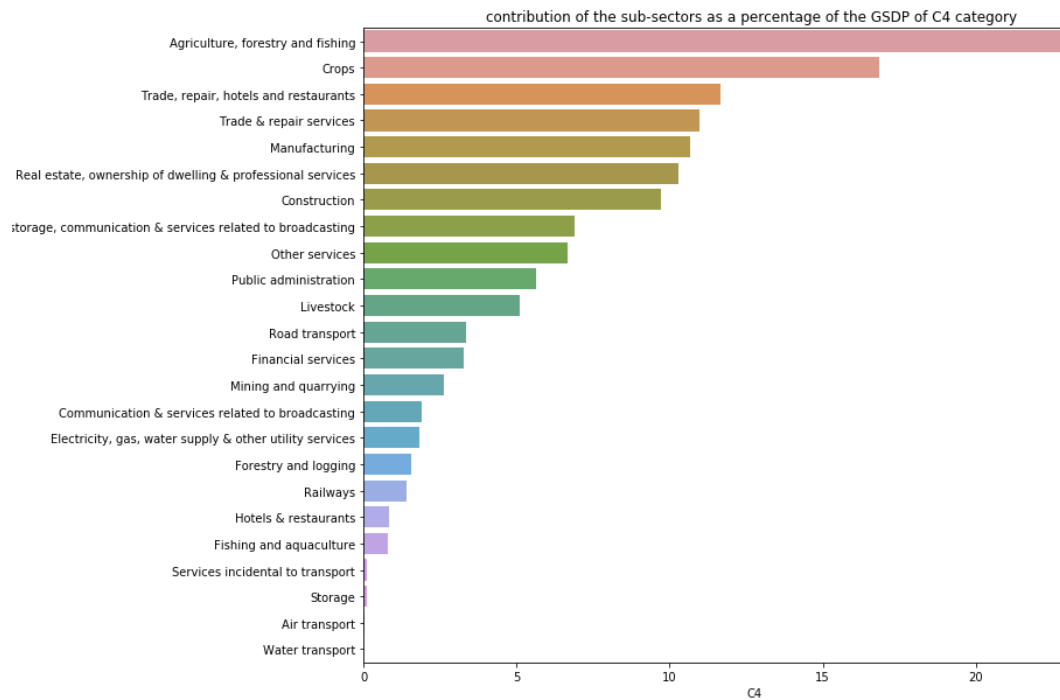
CONTRIBUTION OF EACH SECTOR TOWARDS GDP FOR C4 STATES

10. lets try to observe non-obvious insights from it and answer the below
    Questions from Assignment
  A. How does the GDP distribution of the top states (C1) differ from the
     others
    - Ans : GDP distribution of the top states (C1) are in Secondary and
      tertiary sectors mainly Manufacturing, Services Sector where as GDP
      distribution of C3 , C4 are in primary sectors like Agriculture, Crops,
      forestry and fishing and CROPS
  B. Which sub-sectors seem to be correlated with high GDP?
    - Ans : From the above data it is clear that Secondary and Tertiary
      sectors, like manufacturing and services industries, Real estate, etc
      are contributing to the high GDP
  C. Which sub-sectors do the various categories need to focus on?
   - Ans : shifting the distribution of GDP towards the secondary and
   tertiary sectors may increase per capita GDP
  D. Finally, provide at least two recommendations for each category to
     improve the per capita GDP
    - Ans :
      - Category C1 - As Manufacturing and Service sectors are doing
        good, it can focus more on Real estate,Agriculture
      - Category C2 - As Manufacturing and Real estate are doing good , it
        can focus more on Service sectors,Construction,Agriculture
      - Category C3 - As Agriculture is doing good,it can focus more on
        Manufacturing, Service sectors,Real estate
      - Category C4 - As Agriculture is doing good,it can focus more on
        Manufacturing, Service sectors,Real estate

Part-II: GDP and Education Drop-out Rates

1. Data II: Reading and cleaning of data Read all the files from given source, Th1s file includes dropout rates in education (primary, upper primary and secondary, senior scecondary) for the year 2014-2015 for the states
2. Select only the columns which are required (dropout rates in education (primary, upper primary and secondary, senior secondary) for the year 2014-2015 for all the states)
3. Remove UT data as in Part 1 B also UT data is not used because that are goverend by center not state.
4. Merge the data(DS_DROP) with data(DS_GDP) from part 1B to include the GDP per capita. Merge key is the state column which is common in both dataset.
5. Create the Scatter Plot for All the 4 education level((primary, upper primary and secondary, senior scecondary)) with GSDP per capita on x axis and sector on y axis.
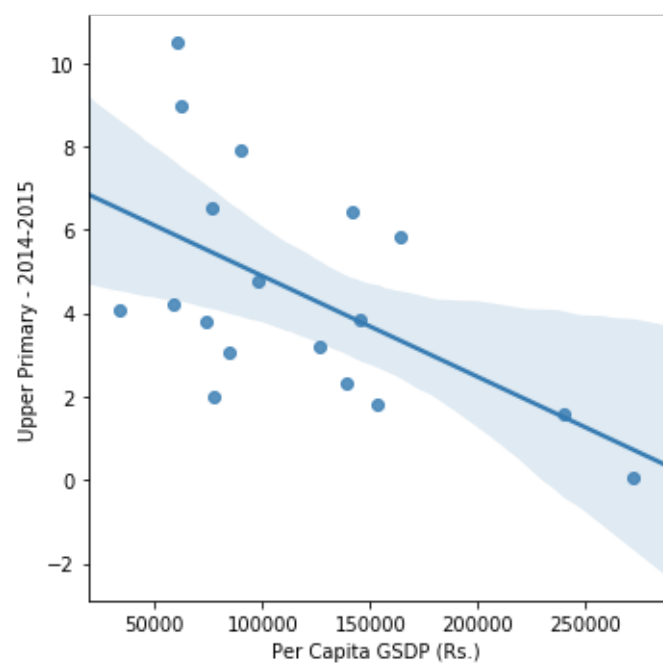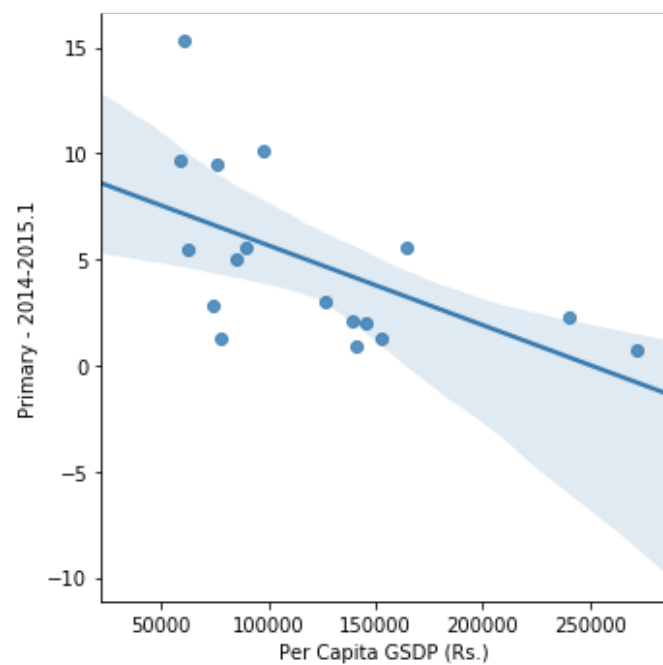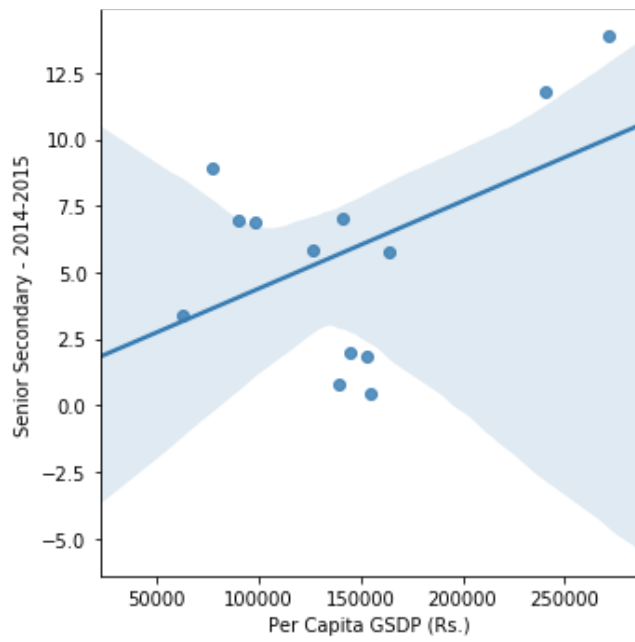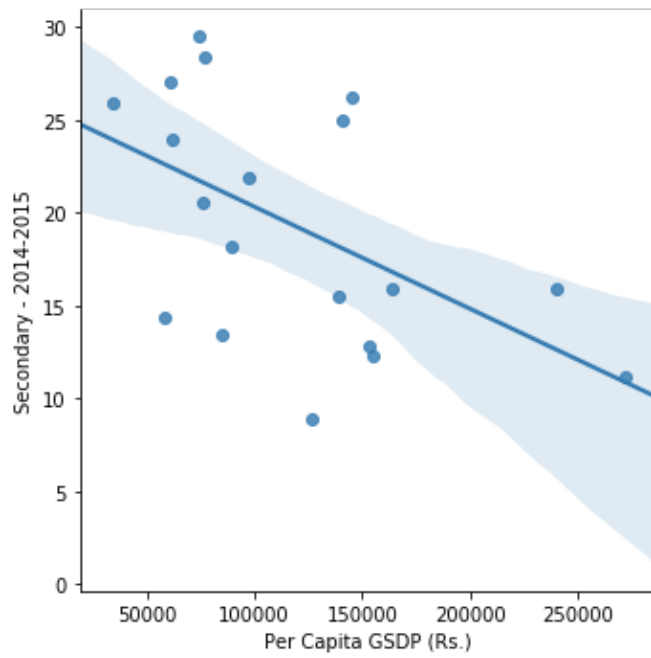
Observation:

From the above plots we can make out the below observation
- primary : As the Drop out rate decrease, per capita GSDP increases
- upper primary : As the Drop out rate decrease, per capita GSDP increases
- secondary: As the Drop out rate decrease, per capita GSDP increases
- Senior Secondary: As the Drop out rate increases, per capita GSDP increases, this looks different from above graph

Correlation of GDP per capita with dropout rates in education:
- As the Drop out rates decrease, GDP per capita of the state increase

Hypothesis:
- As the drop out rates are decreasing , per capita GSDP is increasing up to Secondary level of Education.
- This clearly indicates that Education level directly contribute to the high per capita GSDP of the states.if the population of the state is educated, they will be better able to contribute more to the progress of state.
- Whatever is the Education level(primary, upper primary and secondary), it will definitely improve GSDP even if the population is atleast educated up to primary level, that will also contribute to GSDP.
- There is an exception for the Senior secondary, As the dropout rate increases, GSDP per capita decrease