Problem Description - I

[NITI Aayog](https://en.wikipedia.org/wiki/NITI_Aayog) (National Institution for Transforming India) is a policy think tank of the Government of India which provides strategic inputs to the central and the state governments to achieve various development goals. In the past, NITI Aayog has played an important role in initiatives such as Digital India, Atal Innovation Mission, various agricultural reforms etc. and have designed various policies in education, skill development, water management, healthcare, etc.

NITI Aayog was established to replace the [Planning Commission of India](https://en.wikipedia.org/wiki/Planning_Commission_(India)) which used to follow a top-down model for policy making, i.e. it typically designed policies at the central level (such as the various [5-year plans](https://en.wikipedia.org/wiki/Five-Year_Plans_of_India)). On the other hand, NITI Aayog designs policies specific to the different states or segments of the economy.

Finance Minister [Arun Jaitley](https://en.wikipedia.org/wiki/Arun_Jaitley) made the following observation on the necessity of creating NITI Aayog, "The 65-year-old Planning Commission had become a redundant organisation. It was relevant in a command economy structure, but not any longer. India is a diversified country and its states are in various phases of economic development along with their own strengths and weaknesses. In this context, a ‘one size fits all’ approach to economic planning is obsolete..."

Project Brief

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 states as well.

The overall goal of this project is to help the CMs focus on areas which will foster economic development for their respective states. Since the most common measure of economic development is the [GDP](https://en.wikipedia.org/wiki/Gross_domestic_product), you will analyse the GDP of the various states of India and suggest ways to improve it.

Understanding GDP

[Gross domestic product (GDP)](https://en.wikipedia.org/wiki/Gross_domestic_product) at current prices is the GDP at the market value of goods and services produced in a country during a year. In other words, GDP measures the 'monetary value of final goods and services produced by a country/state in a given period of time'.

GDP can be broadly divided into goods and services produced by three sectors: the primary sector (agriculture), the secondary sector (industry), and the tertiary sector (services).

It is also known as nominal GDP. More technically, (real) GDP takes into account the price change that may have occurred due to inflation. This means that the real GDP is nominal GDP adjusted for inflation. We will use the nominal GDP for this exercise. Also, we will consider the financial year 2015-16 as the base year since most of the data required for this exercise is available for the above period.

Per Capita GDP and Income

Total GDP divided by the population gives the per capita GDP (which roughly measures the average value of goods and services produced per person). The per capita income is closely related to per capita GDP (though they are not the same). In general, the per capita income increases when per capita GDP increases and vice-versa. For instance, in the financial year 2015-16, the per capita income of India was Rs 93,293, whereas the per capita GDP of India was $1717 which roughly translates to Rs 1,11,605.

India ranks 11th in the world in terms of total GDP though it lies at the 139th position in terms of per capita GDP.

Problem Description - II

**Data**

The data is sourced from <https://data.gov.in/> - an Open Government Data (OGD) platform of India. The download instructions are provided in the next segment. The data for GDP Analysis of Indian States is divided into two parts:

**Data I-A:** This dataset contains the GSDP (Gross State Domestic Product) data for the states and union territories.

**Data I-B:** This dataset 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. You are expected to read the dataset for the available states and join these (in Python) if needed.

There are two parts to this project. In the first part, you will analyse and compare the GDPs of various Indian states (both total and per capita). The GDP of a state is referred to as**the GSDP (Gross State Domestic Product)**. Then, you will categorise the states into four categories based on GDP per capita and, for each of these four categories, you will analyse the sectors which contribute most to the GDP (such as agriculture, real estate, manufacturing, etc.).

In the second part, you will analyse whether GDP per capita is related to drop out rates in schools and colleges.

**Part-I: GDP Analysis of Indian States**

For each of the following steps of analysis, choose an appropriate type of plot for comparing the data. Also, ensure that the plots are in increasing or decreasing order for better comparison. For e.g., if you make a bar plot to compare the GDPs of various states, ensure that the bars are in either increasing or decreasing order of GDP.

**Part I-A:**

For the analysis below, use the **Data I-A**.

Remove the rows: '(% Growth over the previous year)' and 'GSDP - CURRENT PRICES (` in Crore)' for the year 2016-17.

Calculate the average growth of states over the duration 2013-14, 2014-15 and 2015-16 by taking the mean of the row '(% Growth over previous year)'. Compare the calculated value and plot it for the states. Make appropriate transformations if necessary to plot the data. Report the average growth rates of the various states:

Which states have been growing consistently fast, and which ones have been struggling?

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?

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

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

**Part I-B:**

For the analysis below, use **Data I-B**. You can also use **Data I-B** along with **Data I-A** if required. Also, perform the analysis only for the duration : 2014-15.

Filter out the Union Territories (Delhi, Chandigarh, Andaman and Nicobar Islands etc.) for further analysis since they are governed directly by the centre, not state governments.

Plot the GDP per capita for all the states.

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

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

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

Categorise the states into four categories based on GDP per capita (C1, C2, C3, C4 - C1 would have the highest per capita GDP, 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 50th and 85th percentile are in C2 and so on.

**Note:**Categorisation into four categories will simplify the subsequent analysis, since comparing data of all the states would become quite exhaustive

For each category C1, C2, C3, C4:

Find the top 3/4/5 **sub-sectors (**such as agriculture, forestry and fishing, crops, manufacturing etc.) [not primary, secondary and tertiary] which contribute to approx. 80% of the GSDP of each category

**Note-I:**The nomenclature for this project is as follows: primary, secondary and tertiary are named 'sectors', while agriculture, manufacturing etc. are named 'sub-sectors'

**Note-II**: If the top-3 sub-sectors contribute to say 79% of the GDP of some category, you can report 'these top-3 sub-sectors contribute to approx. 80% of the GDP'. This is to simplify the analysis and make the results consumable (remember, the CEO has to present the report to the CMs, and CMs have limited time, so the analysis needs to be very sharp and concise)

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

Now that you have summarised the data in the form of plots, tables etc., try to observe non-obvious insights from it. Think about questions such as:

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

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

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

Etc.

Ask other such relevant questions which you think are important and write your insights for category separately.

Finally, provide at least two recommendations for each category to improve the per capita GDP.

**Part-II: GDP and Education Drop-out Rates**

In Part-I, you would have noticed that (one) way to increase per capita GDP is by shifting the distribution of GDP towards the secondary and tertiary sectors, i.e. the manufacturing and services industries. But these industries can thrive only when there is availability of educated, skilled labour.

In this part of the analysis, you will investigate whether there is any relationship between per capita GDP with drop-out rates in education.

**Data**

**Data II:** This section will require the drop-out rate dataset apart from the dataset that you have used in Part-1 of the case study. Download instructions are given in the next segment.

**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 the states. Choose an appropriate plot to conduct this analysis.

Write the key insights you observe from this data:

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