Abstract

GDP Analysis

NITI Aayog (National Institution for Transforming India) is a policy think tank of the Government of India; it provides strategic inputs to the central and the state governments to achieve various development goals. This is 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 the most common measure of economic development is the GDP, in this project, we will analyse the GDP of the various states of India and suggest ways to improve it. Also, as a part of the analysis, we will investigate whether there is any relationship between per capita GDP with dropout rates in education. First dataset consists of the GSDP (Gross State Domestic Product) data for the states and union territories. Second 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. And Finally, another dataset contains dropout rates for different levels of education.

Tools / Skills Used

1. Python Programming
2. Jupyter Notebook
3. Pandas
4. NumPy
5. Matplotlib
6. Seaborn
7. Exploratory Data Analysis
8. Data Visualization

Problem Statement - Introduction to the project

GDP Analysis:

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. 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.

Implementation

Workflow:
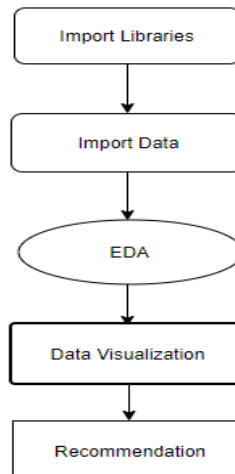
Analysis:

1. **Data Cleaning & EDA:** Cleaning Data and handling Null Values. Also making necessary transmissions to the data to make it visualize-friendly. Required univariate - bivariate analysis is done as a part of Exploratory Data Analysis

2. **GDP and Growth %:** Analysing all the states GDP and Growth % to get the states performing good in terms of GDP and growth rates. Visualizing these using graphs.

3. **Per Capita GDP:** Loading all the states data and joining them to analyse per Capita GDP of all the states with information. Considering all the factors, Identifying Top performing and Poorly Performing States. Analysing theses top and bottom performing states to compare along with visualization graphs.

4. **Sector- Sub Sector:** Considering sectors and Sub sectors contribution for these states, identifying each sector contribution. And finally recommending the sectors which needs to be improved for higher GSDP in poorly performing states.

5. **Drop-out Rates:** Analysing drop-out rates in different levels of education and its effect on Per Capita GDP. Visualizing sector contribution with drop-out rates to find how important education is, in each sector.

Code Snippets:

Importing necessary libraries

```
In [1]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        %matplotlib inline
        import warnings
        warnings.filterwarnings('ignore')
```

```
In [2]: # Reading file
        main = pd.read_csv('Gross Domestic Product (GDP) at Current Price.csv',encoding='utf8')
```

```
In [190]: andhra = pd.read_csv('NAD-Andhra_Pradesh-GSVA_cur_2016-17.csv',encoding='utf8')
          arunachal = pd.read_csv('NAD-Arunachal_Pradesh-GSVA_cur_2015-16.csv',encoding='utf8')
          assam=pd.read_csv('NAD-Assam-GSVA_cur_2015-16.csv',encoding='utf8')
          bihar = pd.read_csv('NAD-Bihar-GSVA_cur_2015-16.csv',encoding='utf8')
          chat=pd.read_csv('NAD-Chandigarh-GSVA_cur_2015-16.csv',encoding='utf8')
          goa = pd.read_csv('NAD-Goa-GSVA_cur_2015-16.csv',encoding='utf8')
          gujrat = pd.read_csv('NAD-Gujarat-GSVA_cur_2015-16.csv',encoding='utf8')
          haryana = pd.read_csv('NAD-Haryana-GSVA_cur_2016-17.csv',encoding='utf8')
          jharkand = pd.read_csv('NAD-Jharkhand-GSVA_cur_2015-16.csv',encoding='utf8')
          karnataka = pd.read_csv('NAD-Karnataka-GSVA_cur_2015-16.csv',encoding='utf8')
          kerala = pd.read_csv('NAD-Kerala-GSVA_cur_2015-16.csv',encoding='utf8')
          mp = pd.read_csv('NAD-Madhya_Pradesh-GSVA_cur_2016-17.csv',encoding='utf8')
          mahara = pd.read_csv('NAD-Maharashtra-GSVA_cur_2014-15.csv',encoding='utf8')
          manipur = pd.read_csv('NAD-Manipur-GSVA_cur_2014-15.csv',encoding='unicode_escape')
          megha = pd.read_csv('NAD-Meghalaya-GSVA_cur_2016-17.csv',encoding='utf8')
          naga = pd.read_csv('NAD-Nagaland-GSVA_cur_2014-15.csv',encoding='utf8')
          odisha = pd.read_csv('NAD-Odisha-GSVA_cur_2016-17.csv',encoding='utf8')
          punjab = pd.read_csv('NAD-Punjab-GSVA_cur_2014-15.csv',encoding='utf8')
          raja = pd.read_csv('NAD-Rajasthan-GSVA_cur_2014-15.csv',encoding='utf8')
          sikkim = pd.read_csv('NAD-Sikkim-GSVA_cur_2015-16.csv',encoding='utf8')
          tamil = pd.read_csv('NAD-Tamil_Nadu-GSVA_cur_2016-17.csv',encoding='utf8')
          tela = pd.read_csv('NAD-Telangana-GSVA_cur_2016-17.csv',encoding='utf8')
          tripura = pd.read_csv('NAD-Tripura-GSVA_cur_2014-15.csv',encoding='utf8')
          up = pd.read_csv('NAD-Uttar_Pradesh-GSVA_cur_2015-16.csv',encoding='utf8')
          uttar = pd.read_csv('NAD-Uttarakhand-GSVA_cur_2015-16.csv',encoding='utf8')
```

Subsector Contribution for Top-Bottom States

```
]: # Identifying the major sub-sectors contributing more to the GSDP  by finding the cumulative sum
   top_contributor = topg_dft[['Sub-sector','Percentage of Total GDP']][:-2].sort_values(by='Percentage of Total GDP', ascending=Fa
   top_contributor.reset_index(drop=True, inplace=True)
   top_contributor['Cumulative sum'] = top_contributor['Percentage of Total GDP'].cumsum()
   top_contributor
```

```
2]: c1_top=top_contributor[1:11]
    c1_top
```

2]:

| | Sub-sector | Percentage of Total GDP | Cumulative sum |
|---|---|---|---|
| 1 | Manufacturing | 18.0354 | 107.621 |
| 2 | Real estate, ownership of dwelling & professio... | 15.6657 | 123.286 |
| 3 | Agriculture, forestry and fishing | 13.2649 | 136.551 |
| 4 | Trade, repair, hotels and restaurants | 9.97303 | 146.524 |
| 5 | Trade & repair services | 8.98206 | 155.506 |
| 6 | Crops | 8.31197 | 163.818 |
| 7 | Construction | 7.26145 | 171.08 |
| 8 | Financial services | 6.18663 | 177.266 |
| 9 | Transport, storage, communication & services r... | 6.04749 | 183.314 |
| 10 | Other services | 5.90735 | 189.221 |

```
: # Identifying the major sub-sectors contributing more to the GSDP  by finding the cumulative sum
  bottom_contributor = bottom_dft[['Sub-sector','Percentage of Total GDP']][:-2].sort_values(by='Percentage of Total GDP', ascendi
  bottom_contributor.reset_index(drop=True, inplace=True)
  bottom_contributor['Cumulative sum'] = bottom_contributor['Percentage of Total GDP'].cumsum()
  bottom_contributor
```

```
: bottom_top=bottom_contributor[1:11]
  bottom_top
```

:

	Sub-sector	Percentage of Total GDP	Cumulative sum
1	Agriculture, forestry and fishing	23.6505	118.276
2	Crops	17.119	135.395
3	Trade, repair, hotels and restaurants	13.1406	148.536
4	Trade & repair services	12.39	160.926
5	Manufacturing	11.5689	172.495
6	Construction	8.64912	181.144
7	Other services	8.06431	189.208
8	Real estate, ownership of dwelling & professio...	7.4188	196.627
9	Transport, storage, communication & services r...	6.72383	203.351
10	Public administration	5.24497	208.596

Visualization Snippets:
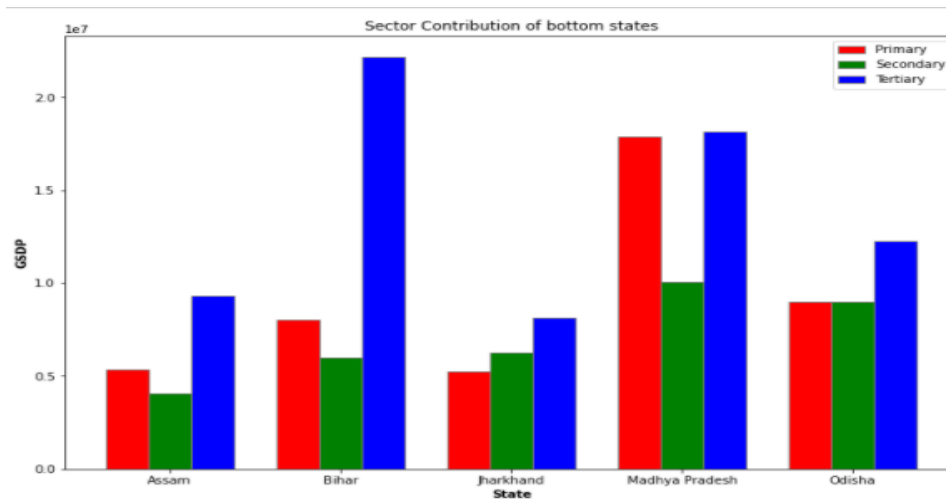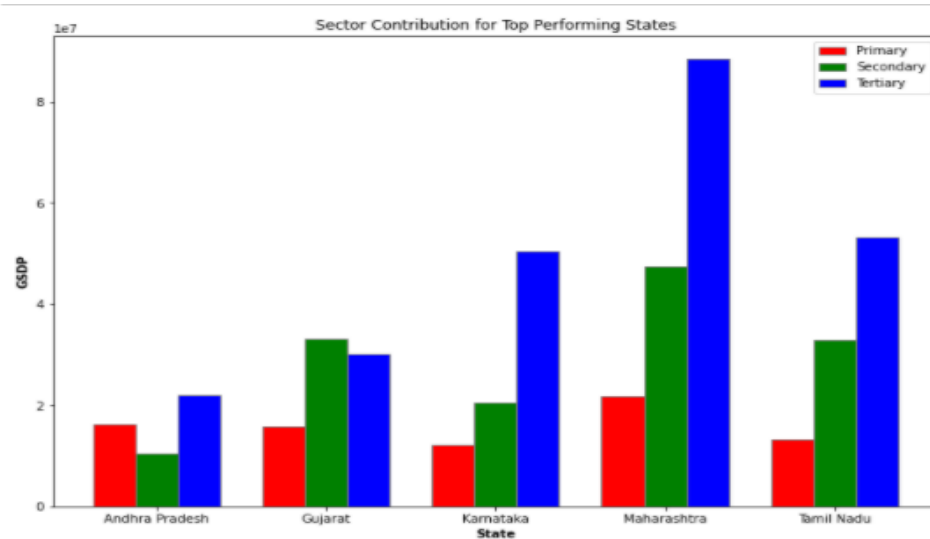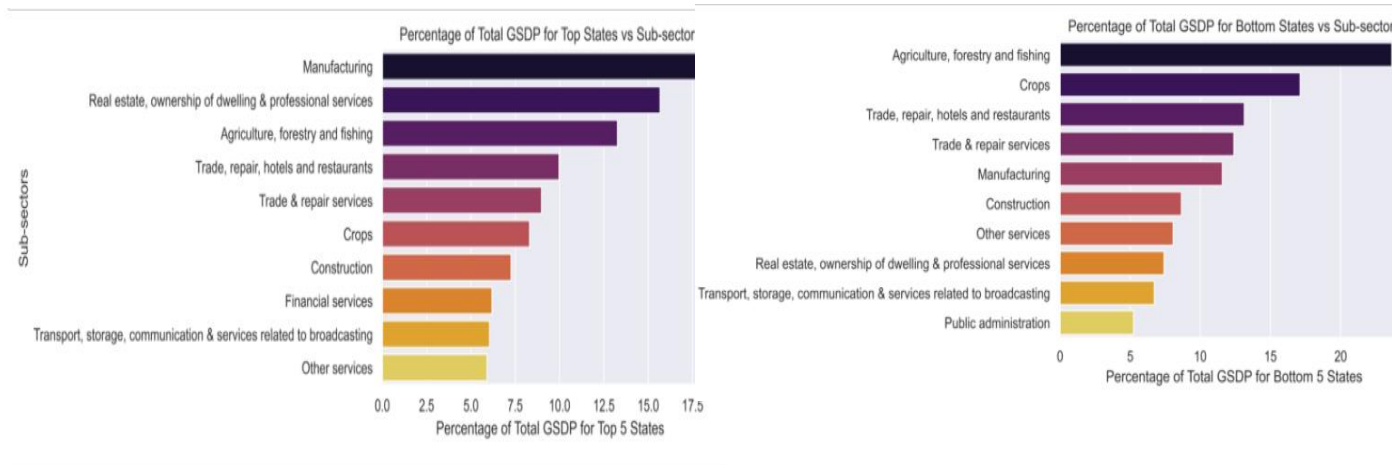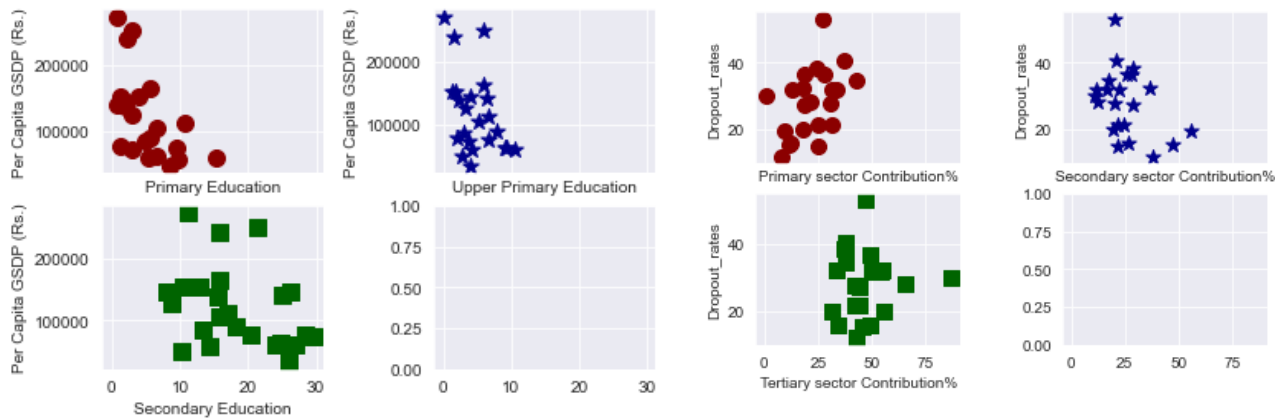
Conclusion/ Recommendation

- Poorly performing states need to concentrate on sectors like Manufacturing and Real estate, ownership of dwelling & professional services since they are the common sectors that drive GSDP in top performing states
- For these sectors to improve it is important to consider education dropout rates since these sectors require professionals
- Financial Services should be promoted further in poorly performing states
- Top performing states are already doing well. Sub sectors like Railways and Forestry can be promoted further for higher increase in GSDP good.
- It is necessary to decrease the drop-out rates to contribute to higher per Capita GSDP of states
- Education holds prior importance when it comes to secondary sector. And hence, States performing poorly in secondary sector should be given education as an important factor for GSDP Growth
- Population is very slightly effecting drop-out rates and hence even smaller cities with less population can consider the above recommendations to increase GSDP growth.

Future Scope

In future, when the GDP of different states fluctuate, using more data informative measures can be taken out and can also use machine learning models to predict GDP of states in the upcoming years.

And from this required measure can be taken if the predicted GDP rates are very low which going to be beneficial in increasing overall GDP of country.