

GDP Assignment

Ishar Mohapatra

Steps carried out:

1. Downloading of the data as CSV files from mentioned site
2. Understanding of Datasets.
3. Understanding the requirements out of the data.
4. Cleaning of Data
 - Filter out the union territories (Delhi, Chandigarh, Andaman and Nicobar Islands, etc.) for further analysis, as they are governed directly by the central, not state governments.
 - Cleaning of data with no values like West Bengal
 - In order to merge two data frame we need to format the state names to one format.
 - Remove unnecessary columns from the data set that are not relevant
5. Reading the csv from the downloaded file path

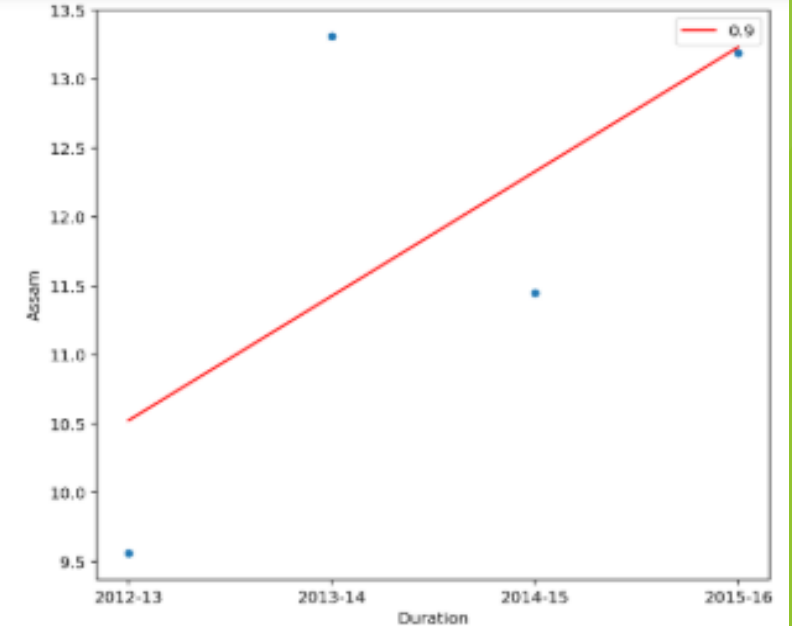
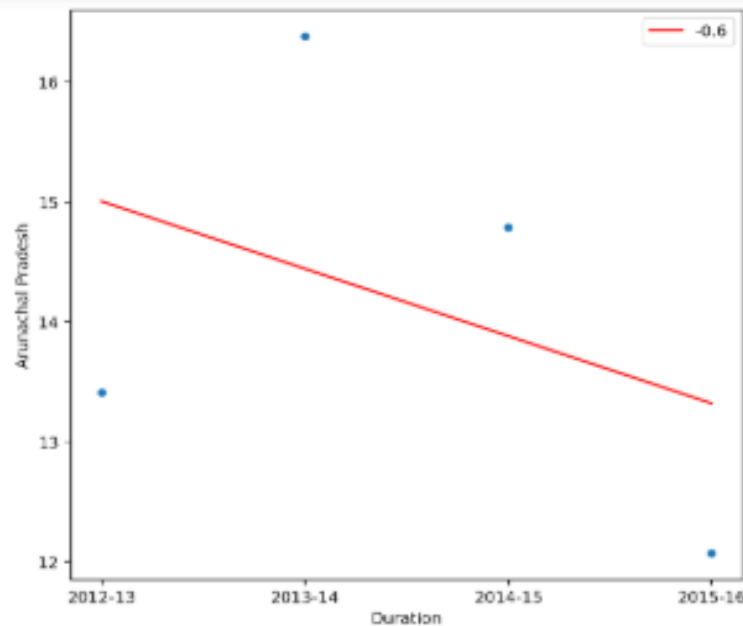
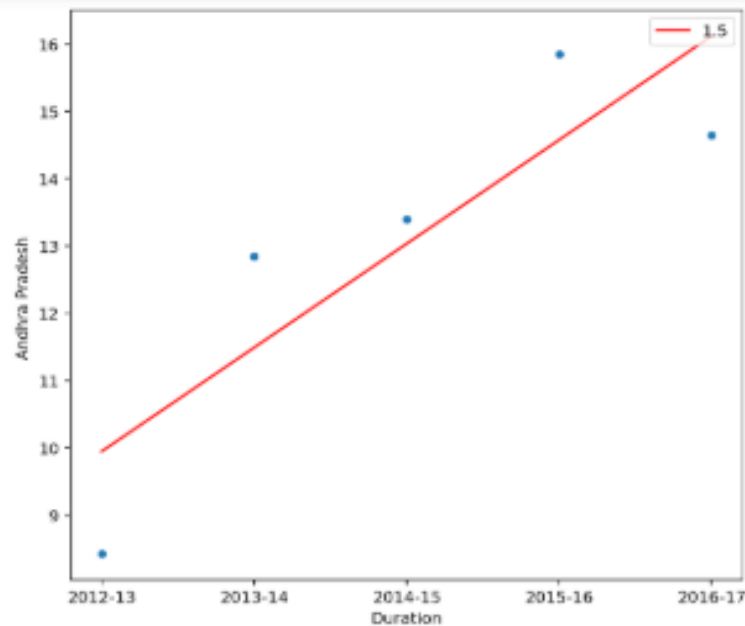
Part-I: GDP Analysis of the Indian States

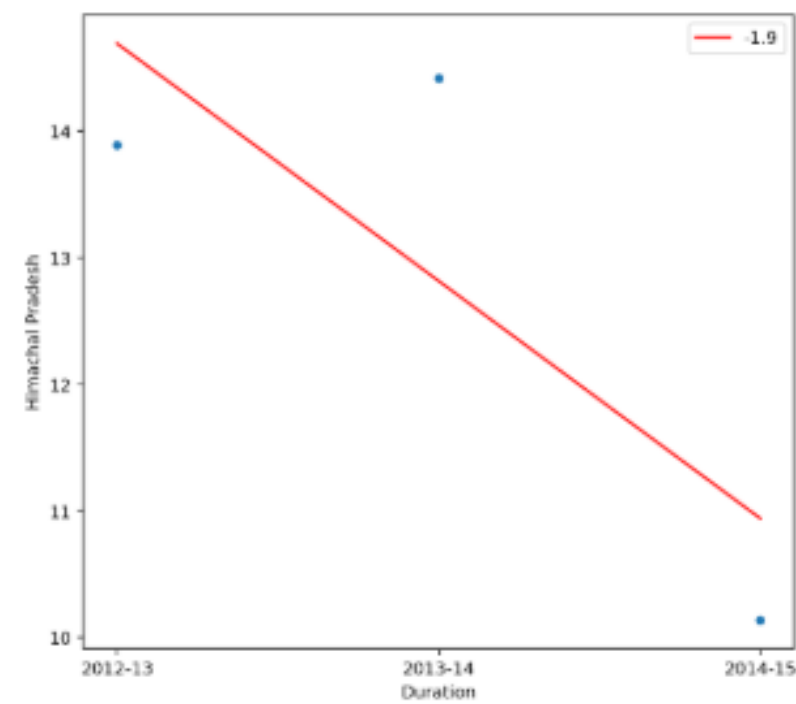
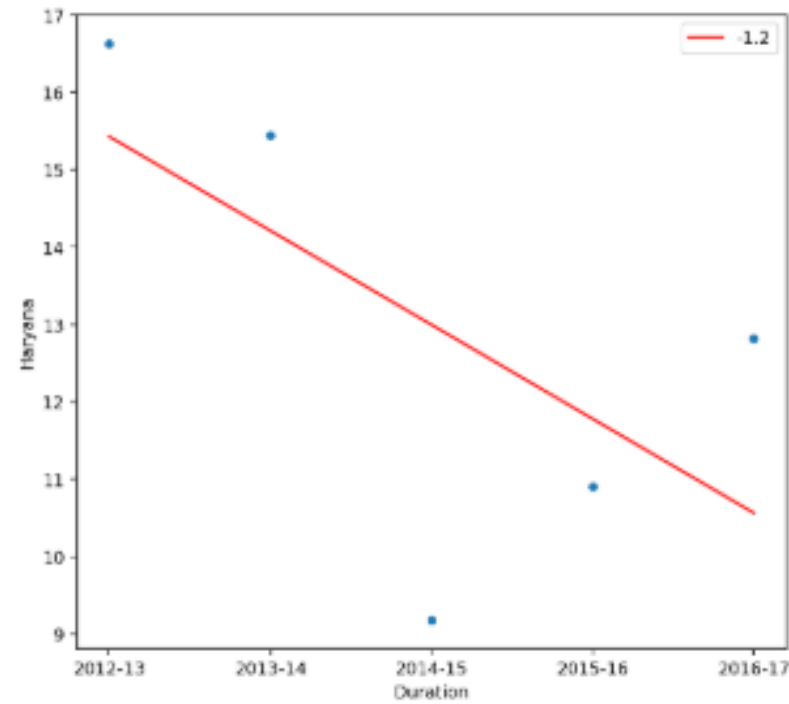
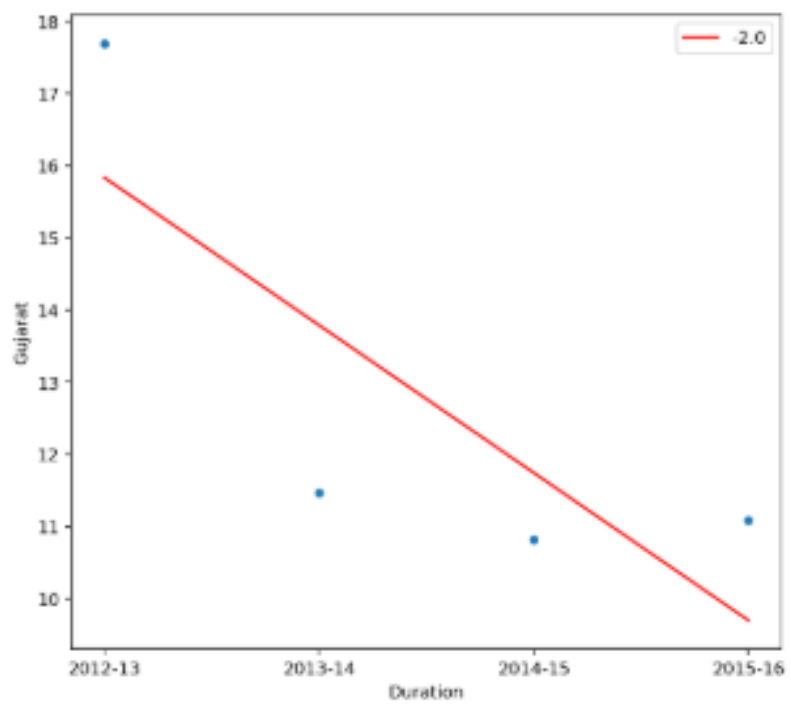
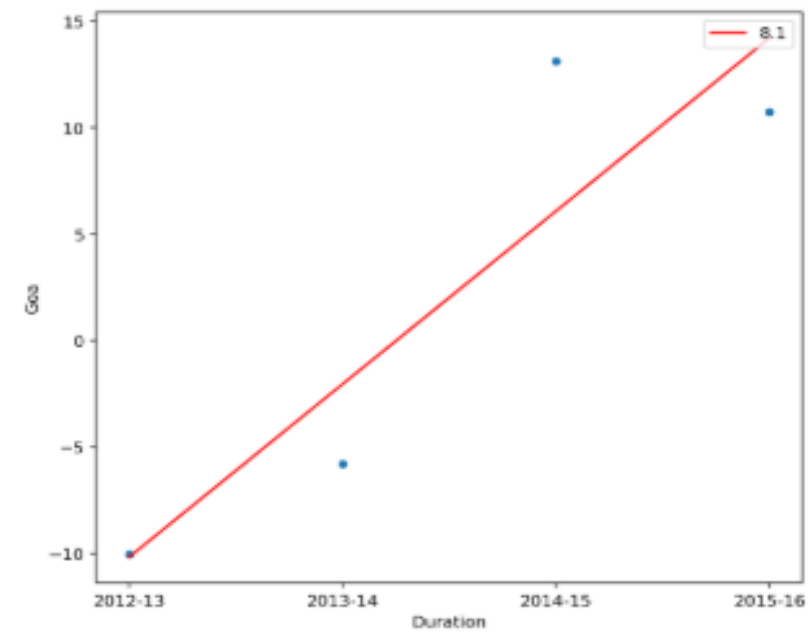
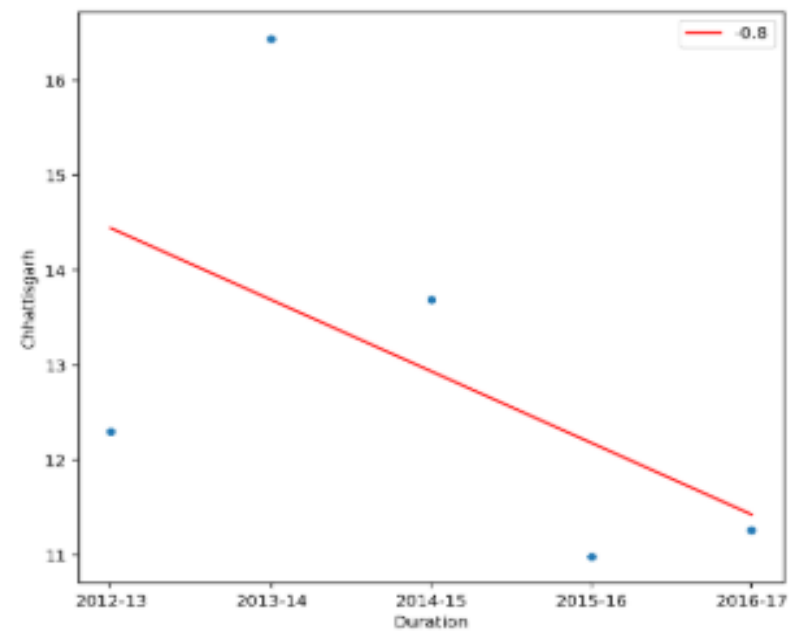
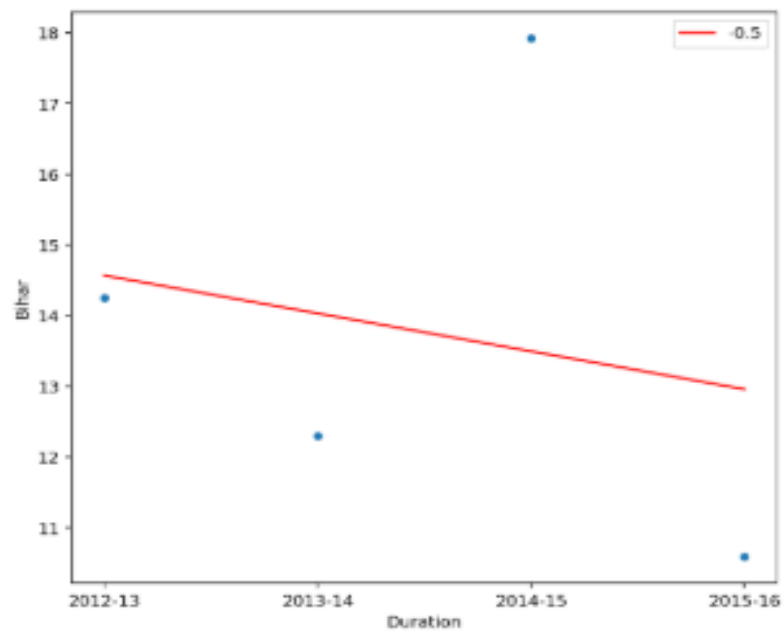
Part I-A:

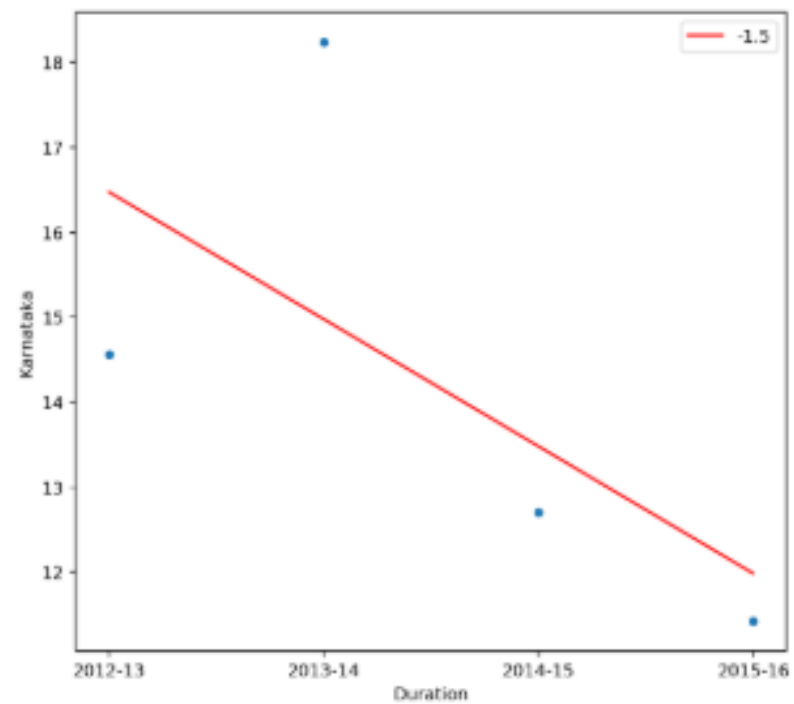
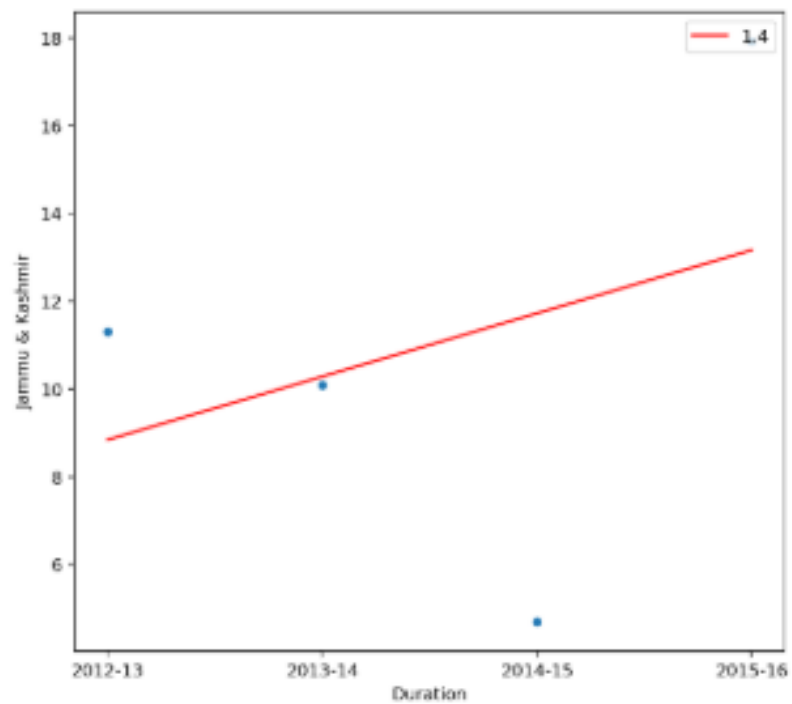
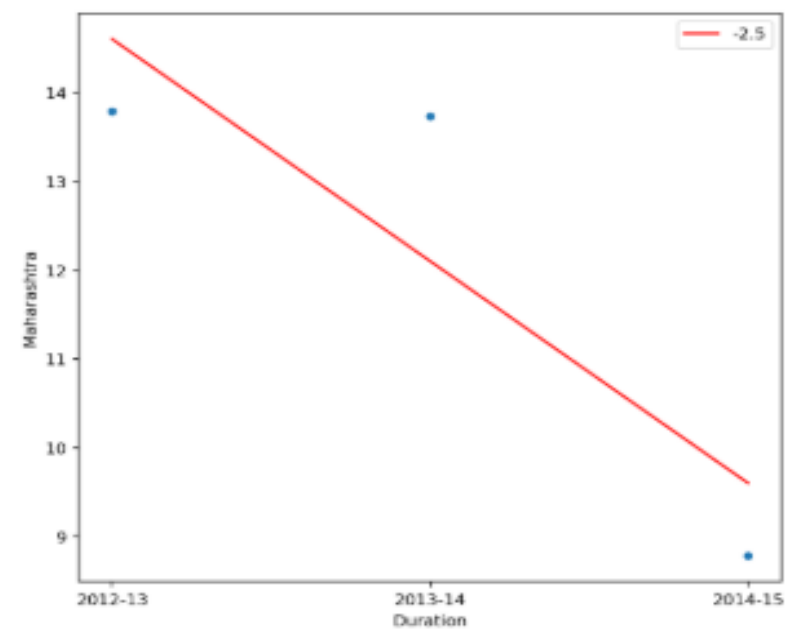
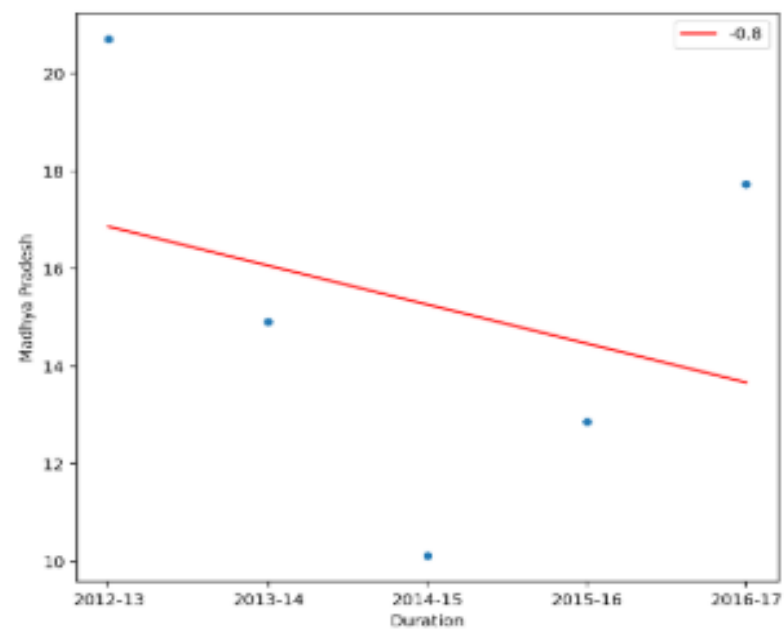
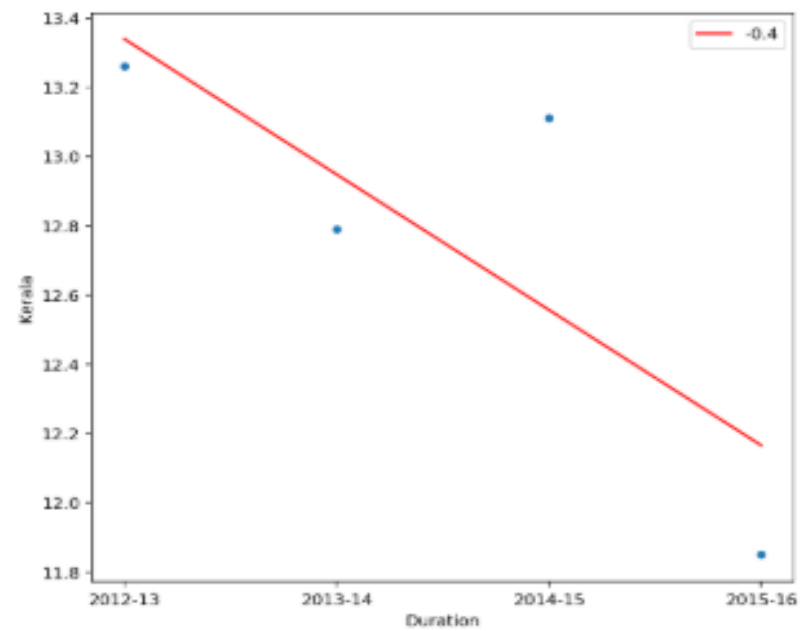
For the analysis below, **Data I-A** is used.

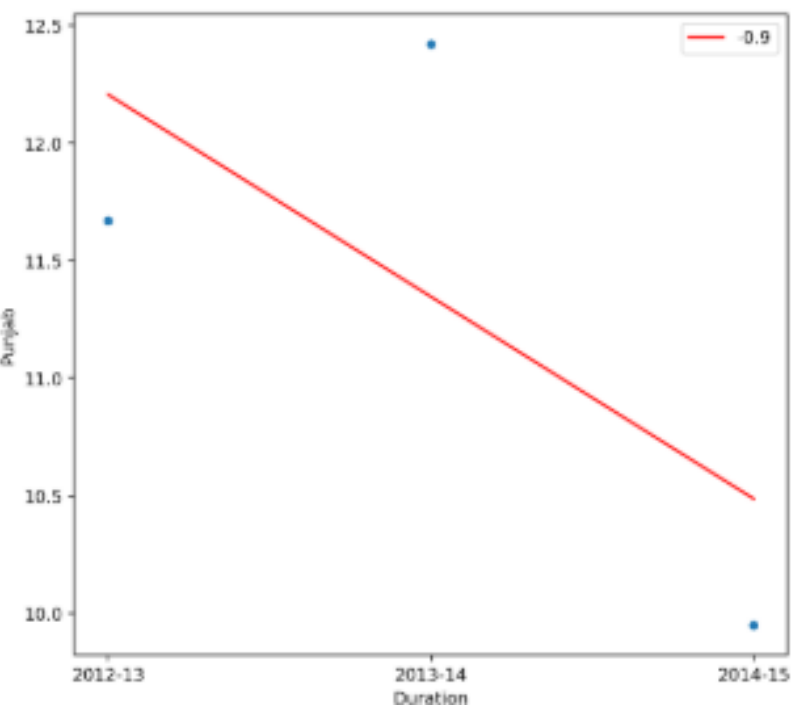
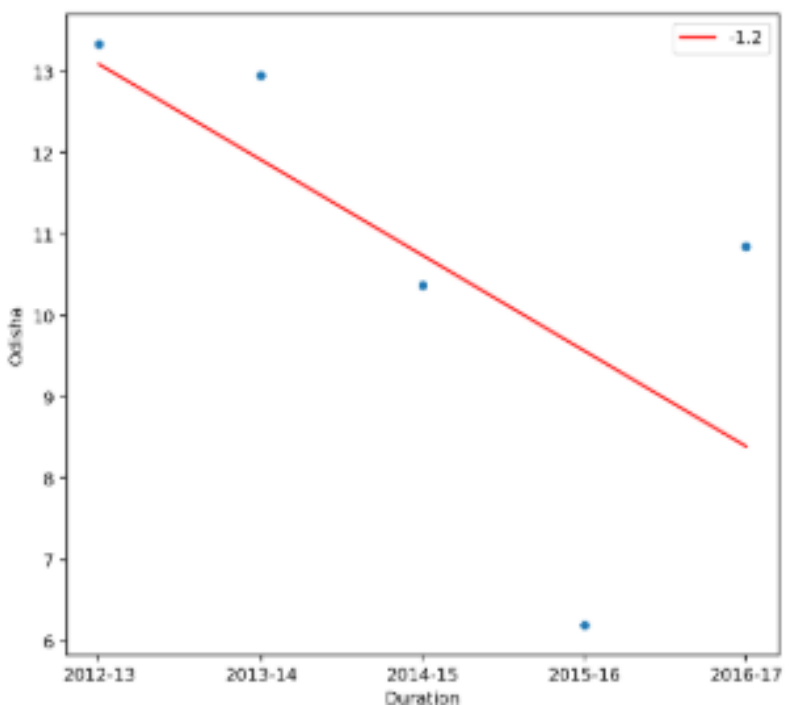
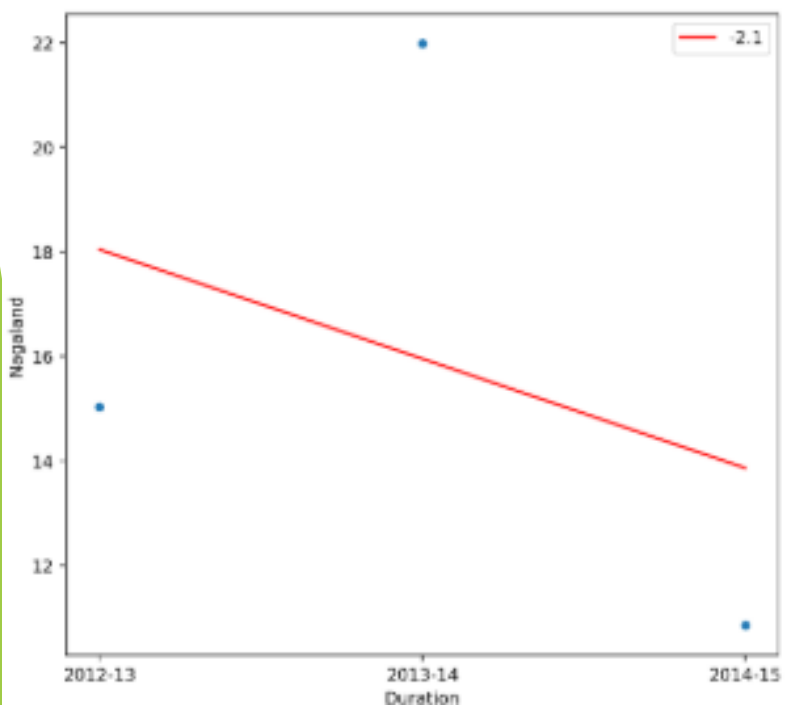
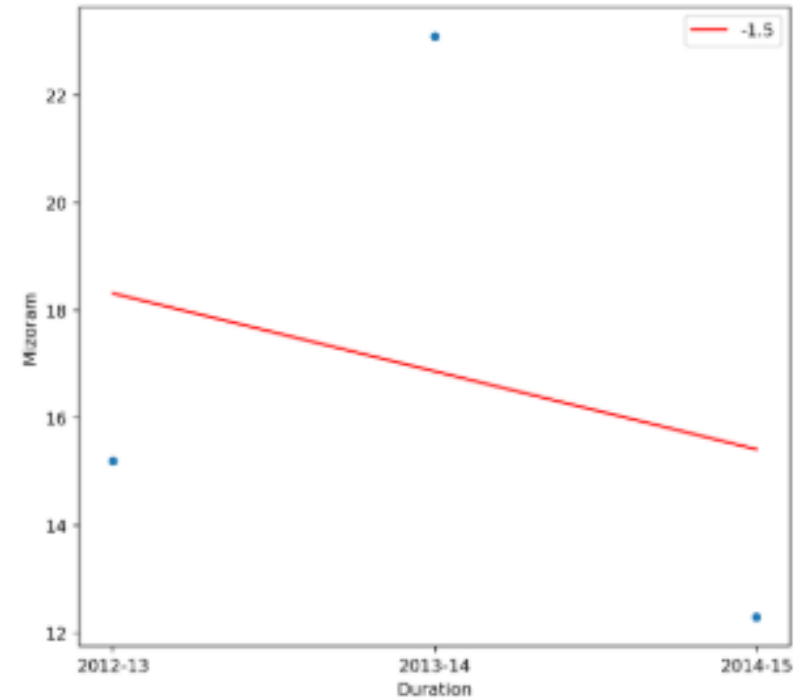
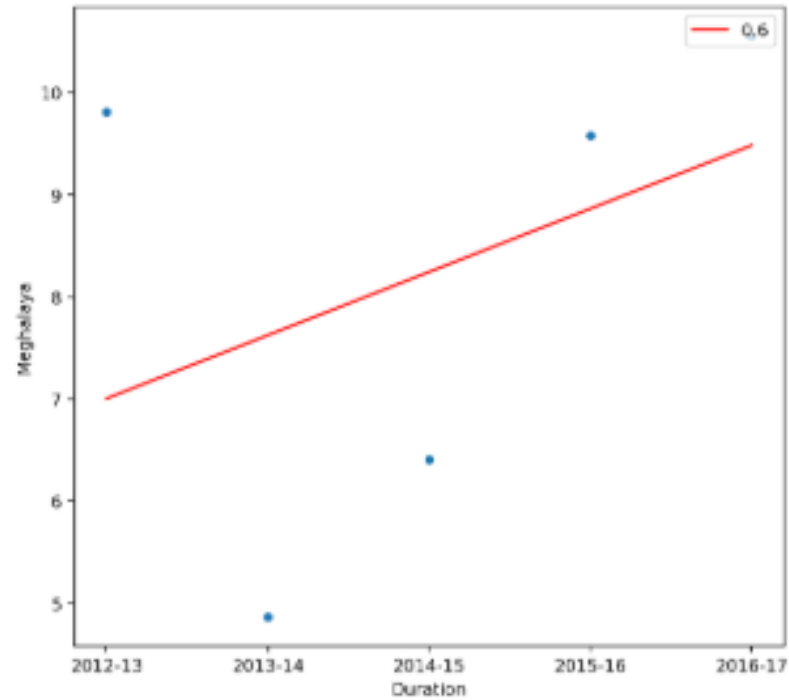
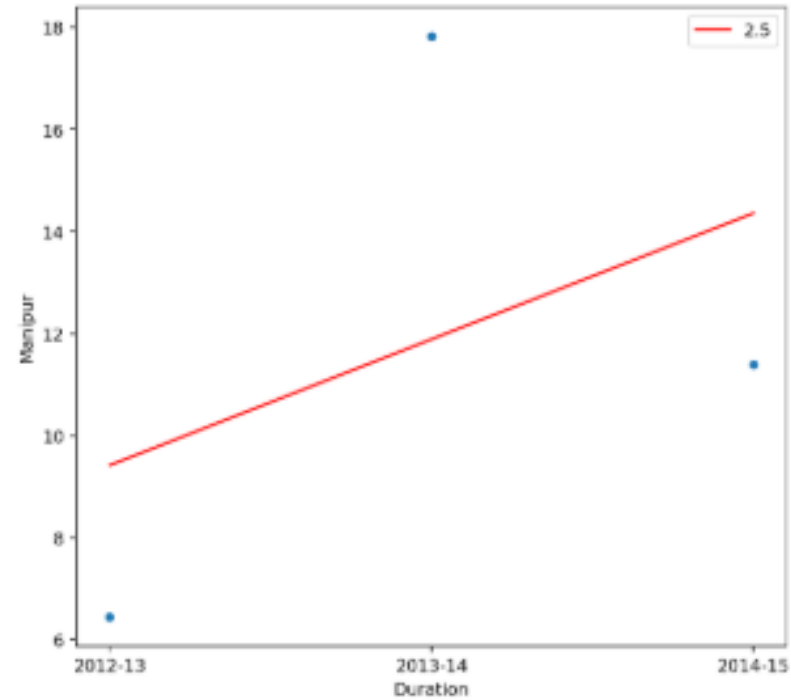
Plot a graph for rows "% Growth over previous year" for all the states (not union territories) whose data is available, use as much data as possible for this exercise. Use the best fit line to represent the growth for each state. Draw a similar line graph for the nation as well.

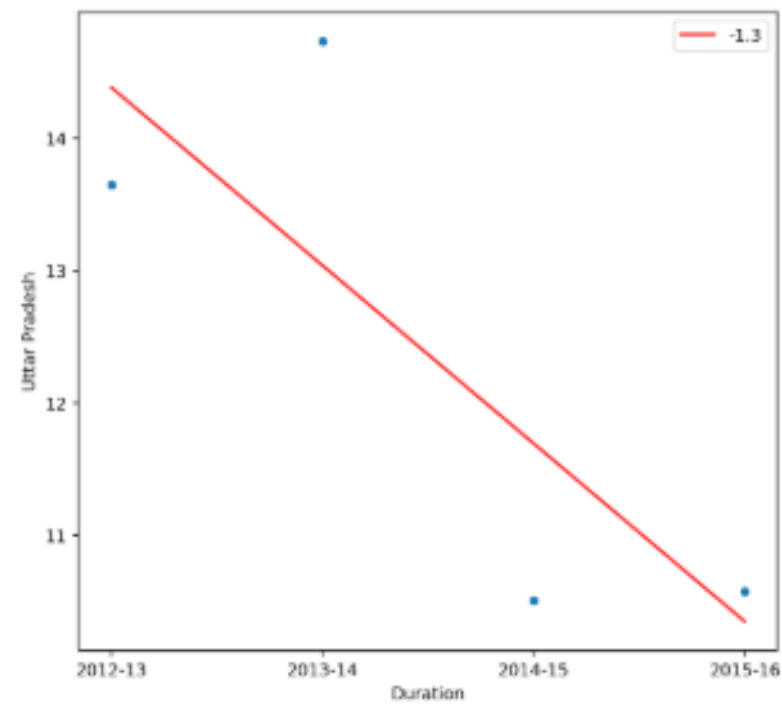
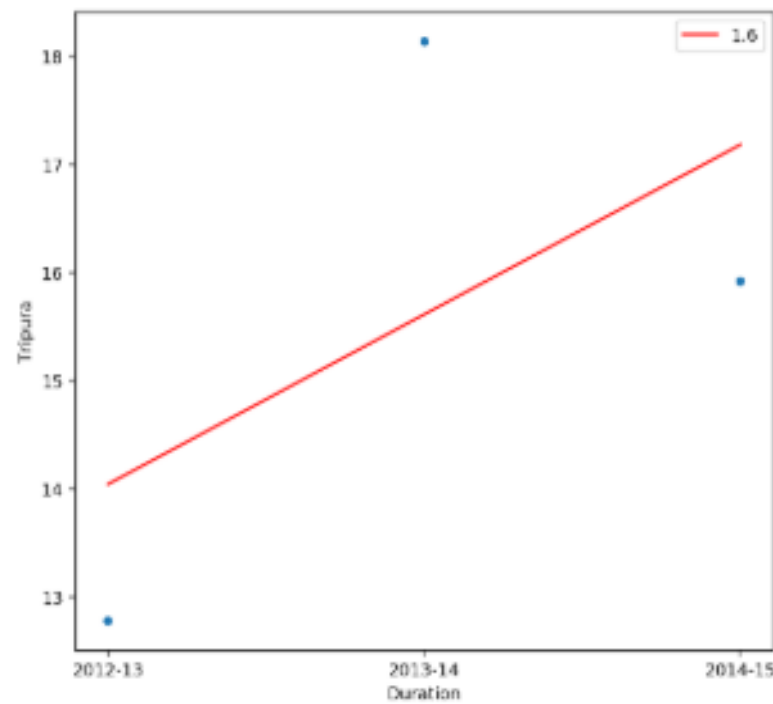
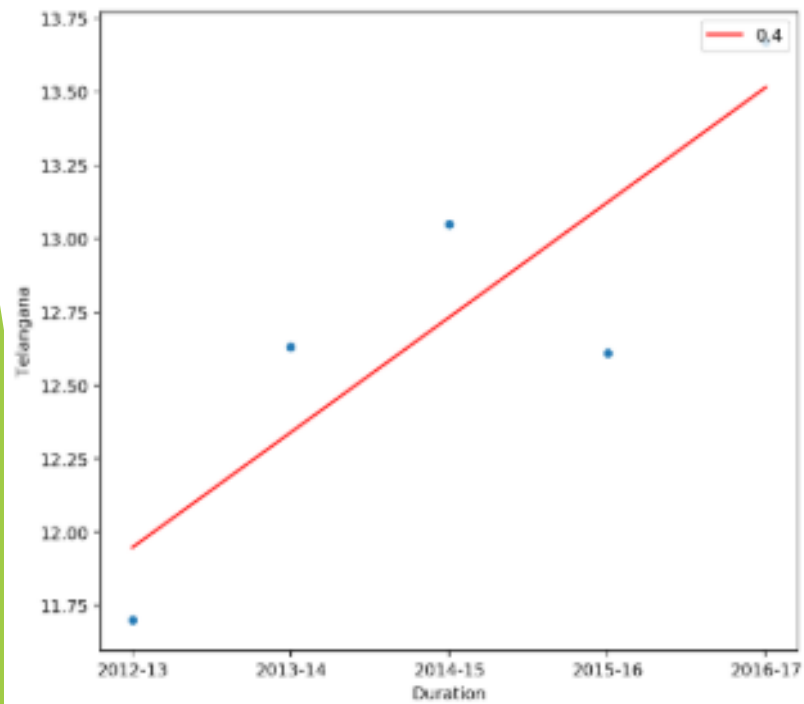
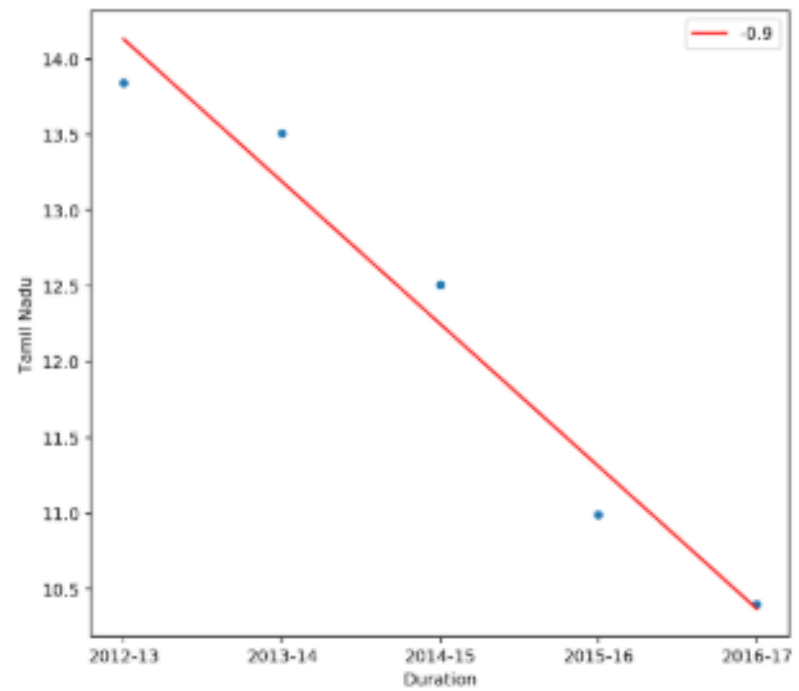
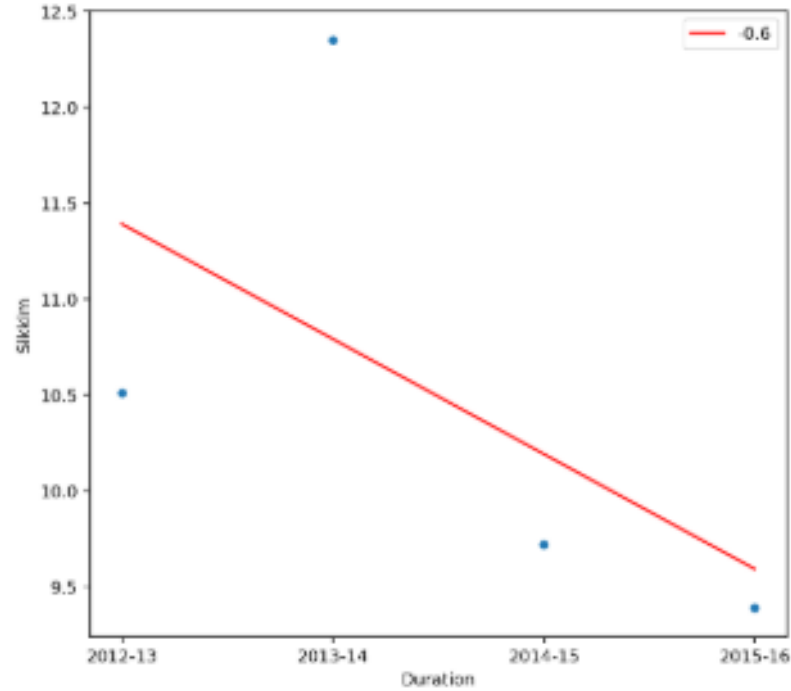
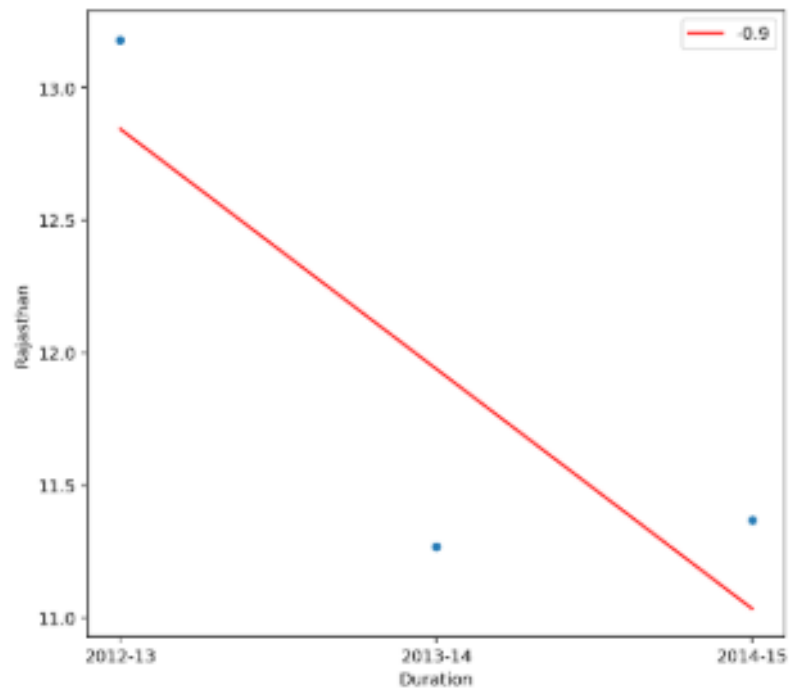
Here are the plots for different states based on the data of % Growth over previous year.

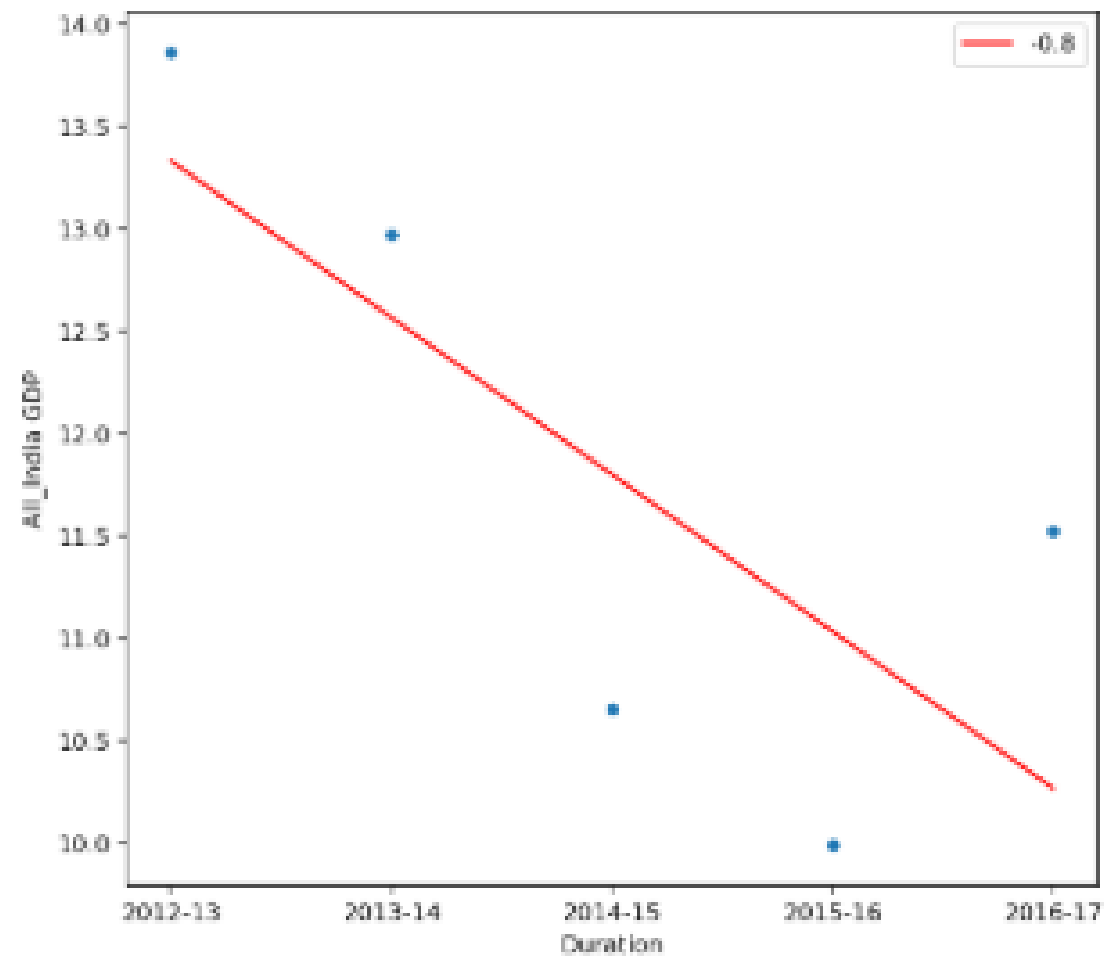
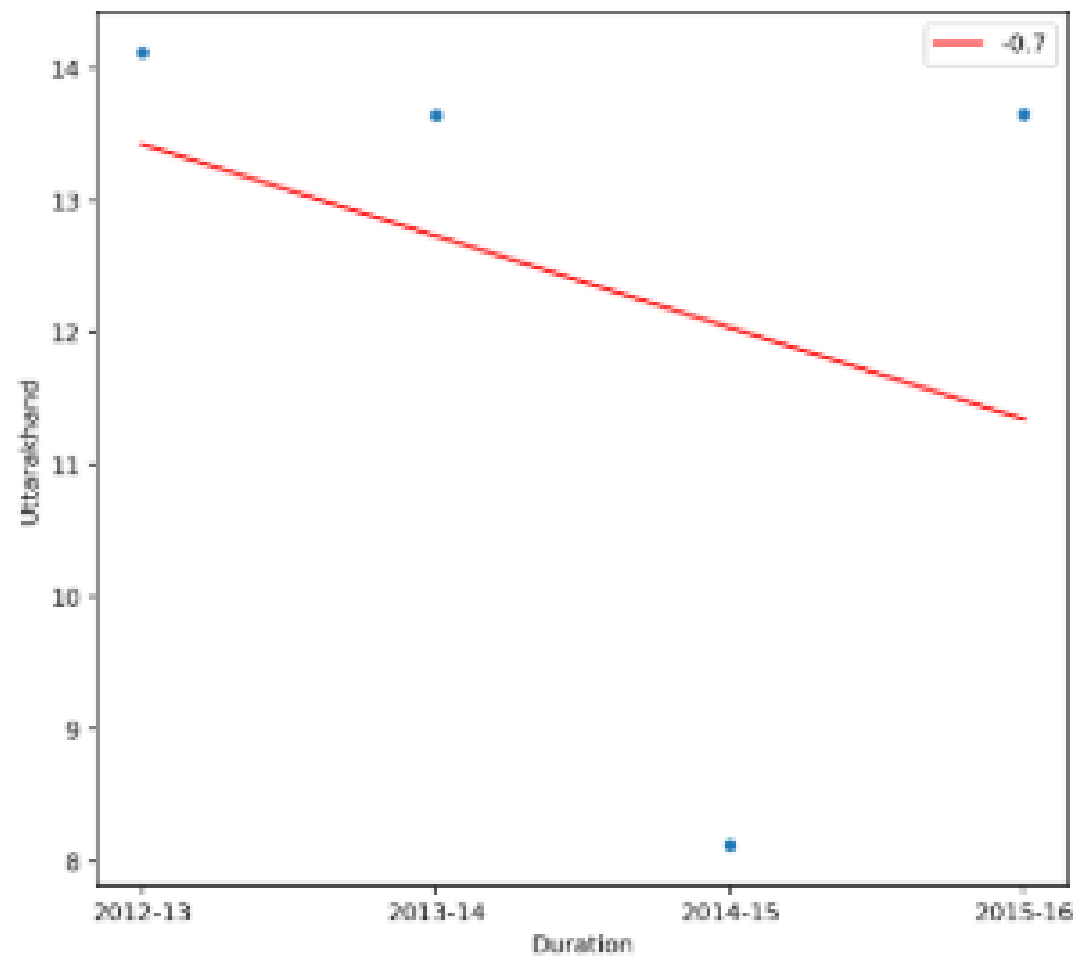






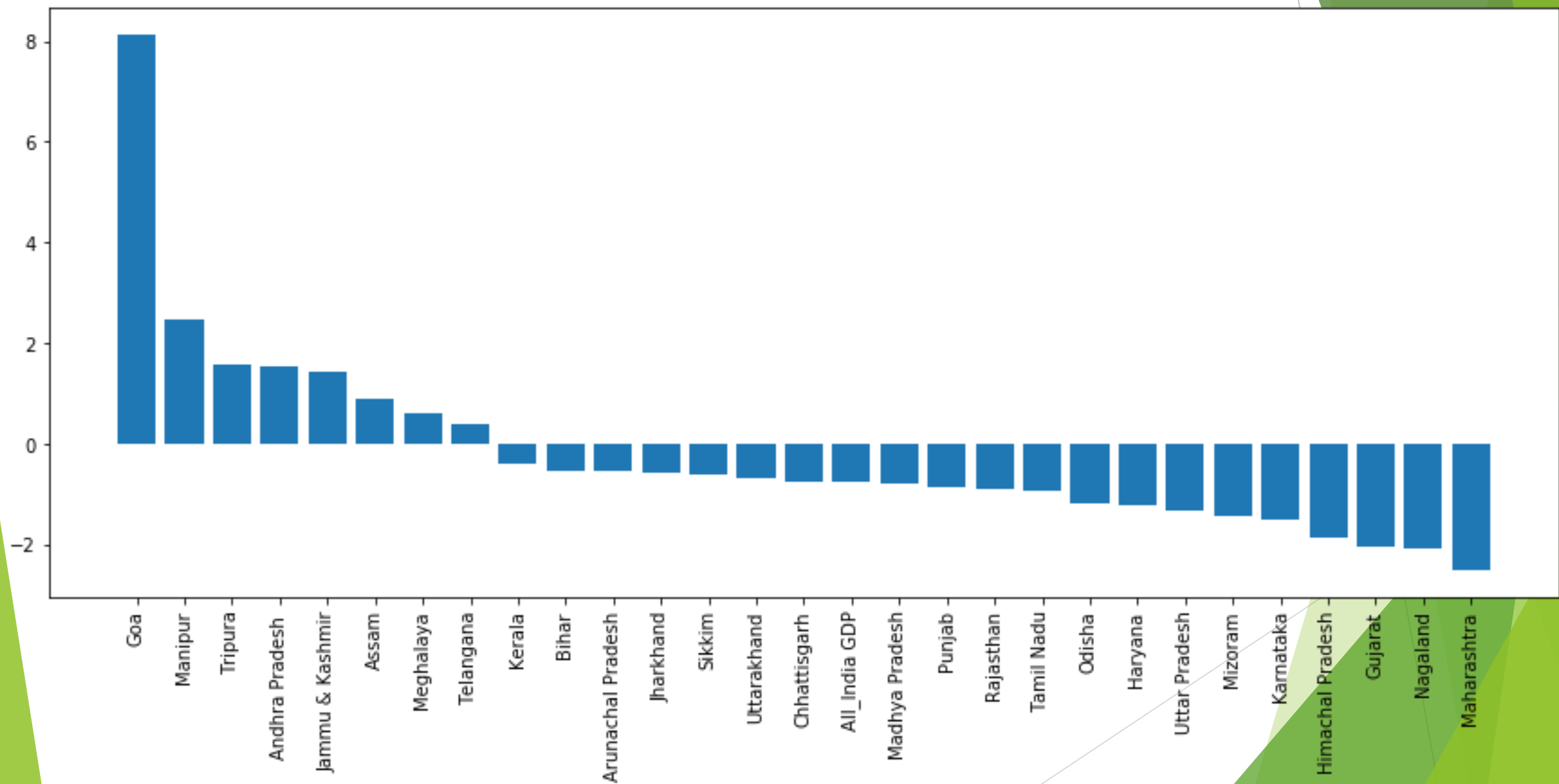






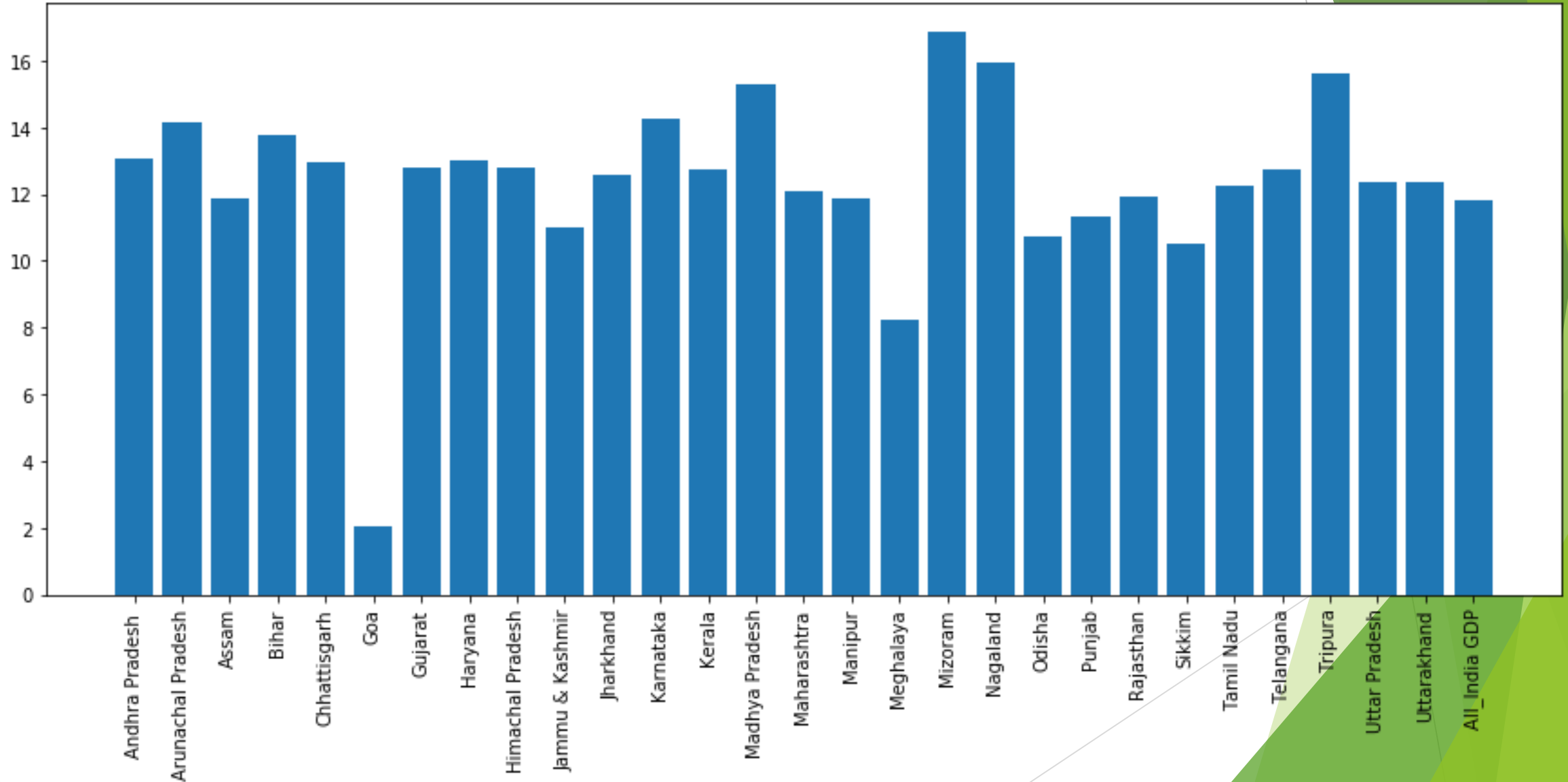
Plot the slope of GDP growth % of the states for the year 2015-16:

This graph represents growth rates of different states.



Plot the average growth rate of GDP growth % of the states for the year 2015-16:

This graph represents consistent growth rates of different states.



How will you compare the growth rates of any two states?

We can compare the growth rate of any two states by plotting a best fit line on the scatter plot using the % Growth over previous year. Now once we have the best fit line, we can calculate the slope to identify the growth rate.

Which states have been growing consistently fast, and which ones have been struggling? Rank top 3 fastest and 3 slowest-growing states.

Top 3 consistently Growing States	Slowest growing State
<ul style="list-style-type: none">• Mizoram• Nagaland• Tripura	<ul style="list-style-type: none">• Goa• Meghalaya• Sikkim

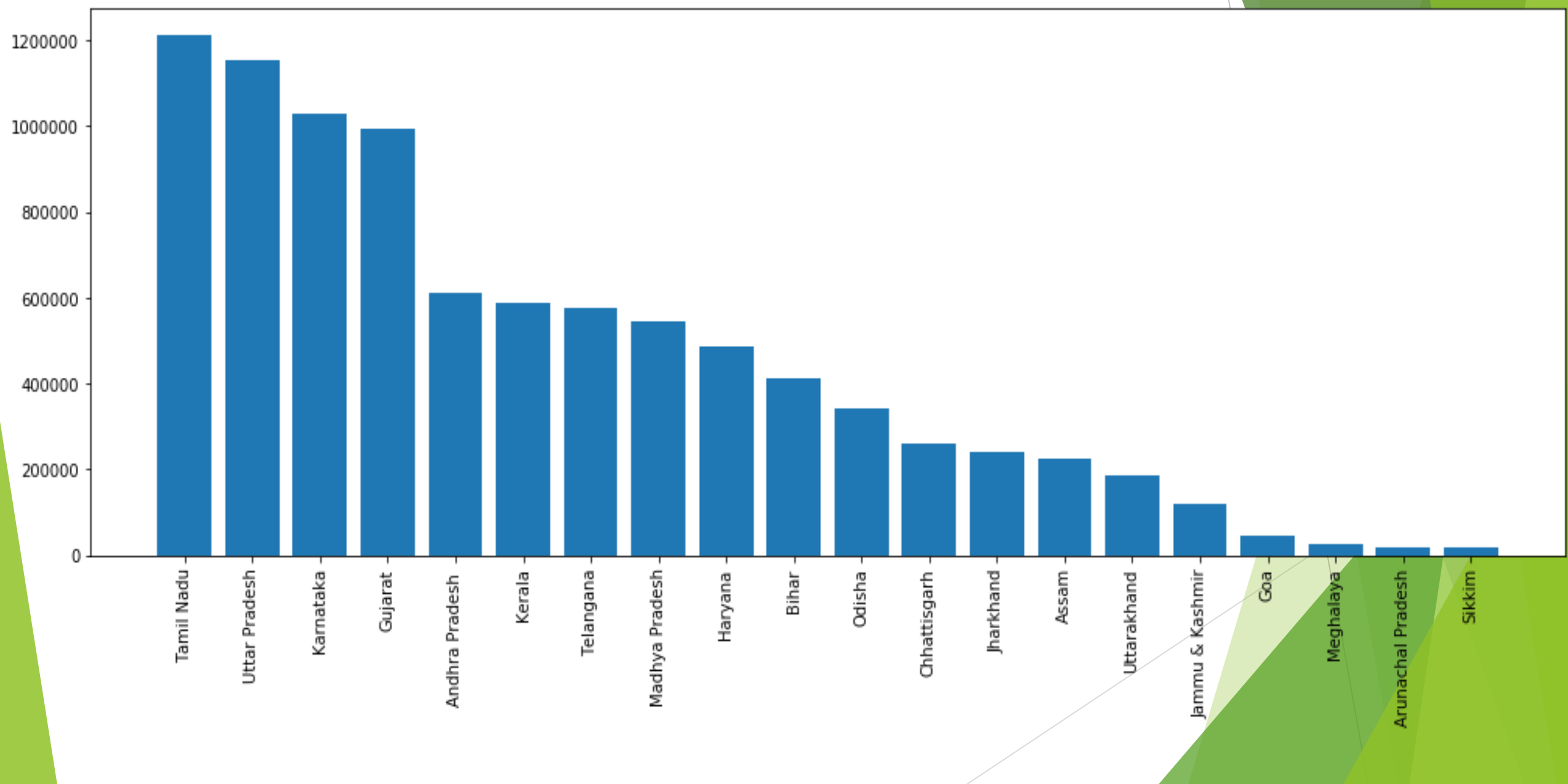
What is the Nation's growth rate?

Nation Growth rate is negative All India GDP is -0.766 where as Average growth rate is around 11.78

What has been the growth rate of your home state, and how does it compare to the national growth rate?

My Home State is Odisha, Growth rate is -1.174 and Growth rate of All India GDP is -0.766 Both are having negative growth rate but average growth rate is similar 10.74 to that of National growth rate.

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



Which Plot will you use for this? Why?

I have used a Bar chart plot for analyzing GDP of the states, bar chart are used to compare values between a categorical value.

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

In order to identify the best and worst performing states I have sorted the data frame. Here we can clearly see following results based on the GDP value.

Top 5 states	Bottom 5 states
<ul style="list-style-type: none">• Tamil Nadu• Uttar Pradesh• Karnataka• Gujarat• Andhra Pradesh	<ul style="list-style-type: none">• Sikkim• Arunanchal Pradesh• Meghalaya• Goa• Jammu & Kashmir

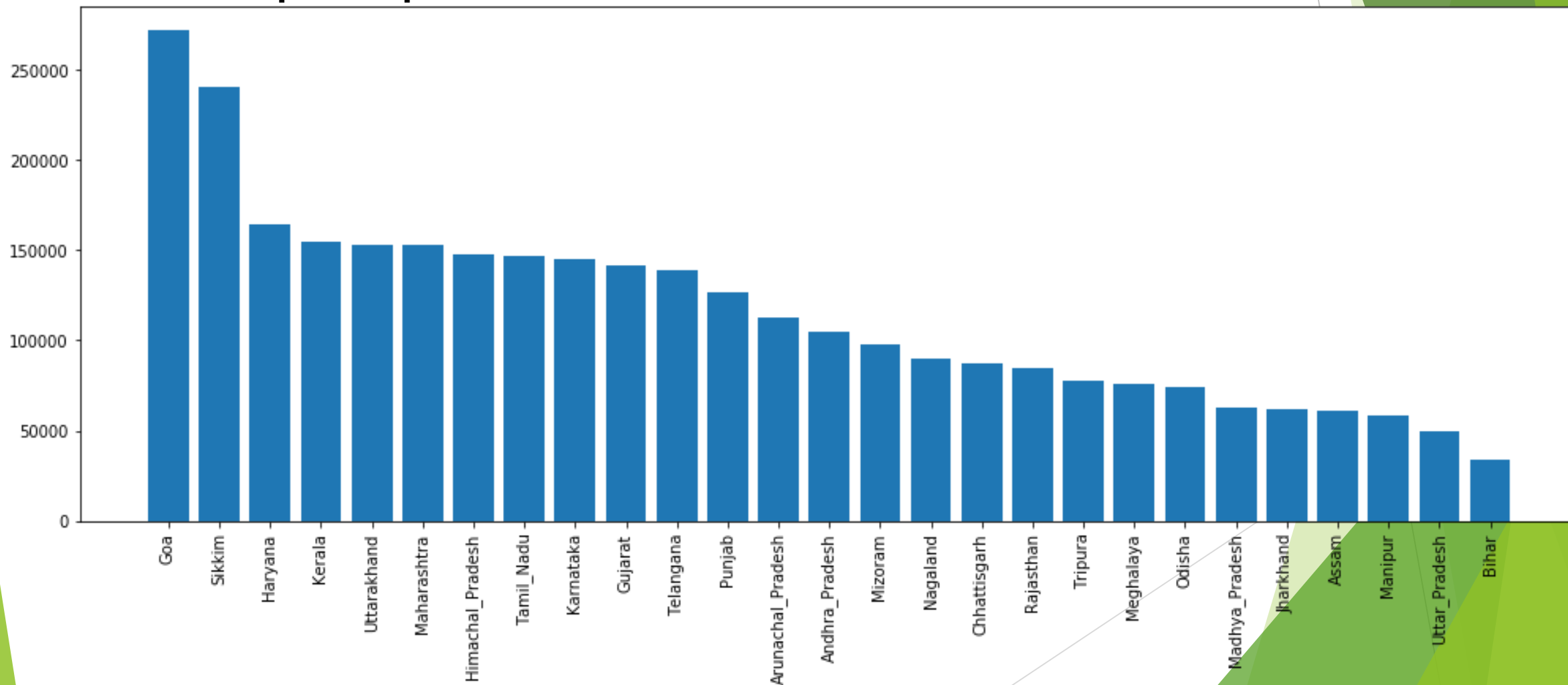
What insights can you draw from this graph? What states are performing poorly? (Remember: this will not solely based on total GDP)

- Goa has best speed of growth but not performing consistently across the years and is the poorest average performer.
- North Eastern states like Sikkim, Arunanchal Pradesh, Meghalaya are in bottom GSDP and also are performing negatively in growth rate.
- States like Tamil Nadu, Uttar Pradesh, Karnataka, Gujarat has been **topping the list of GDSP in 2015-16** as well as performing above average consistently.

Part I-B:

For the analysis below, Data I-B is used.

Plot the GDP per capita for all the states.



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

Top 5 states	Bottom 5 states
<ul style="list-style-type: none">• Goa• Sikkim• Haryana• Kerala• Uttarakhand	<ul style="list-style-type: none">• Bihar• Uttar Pradesh• Manipur• Assam• Jharkhand

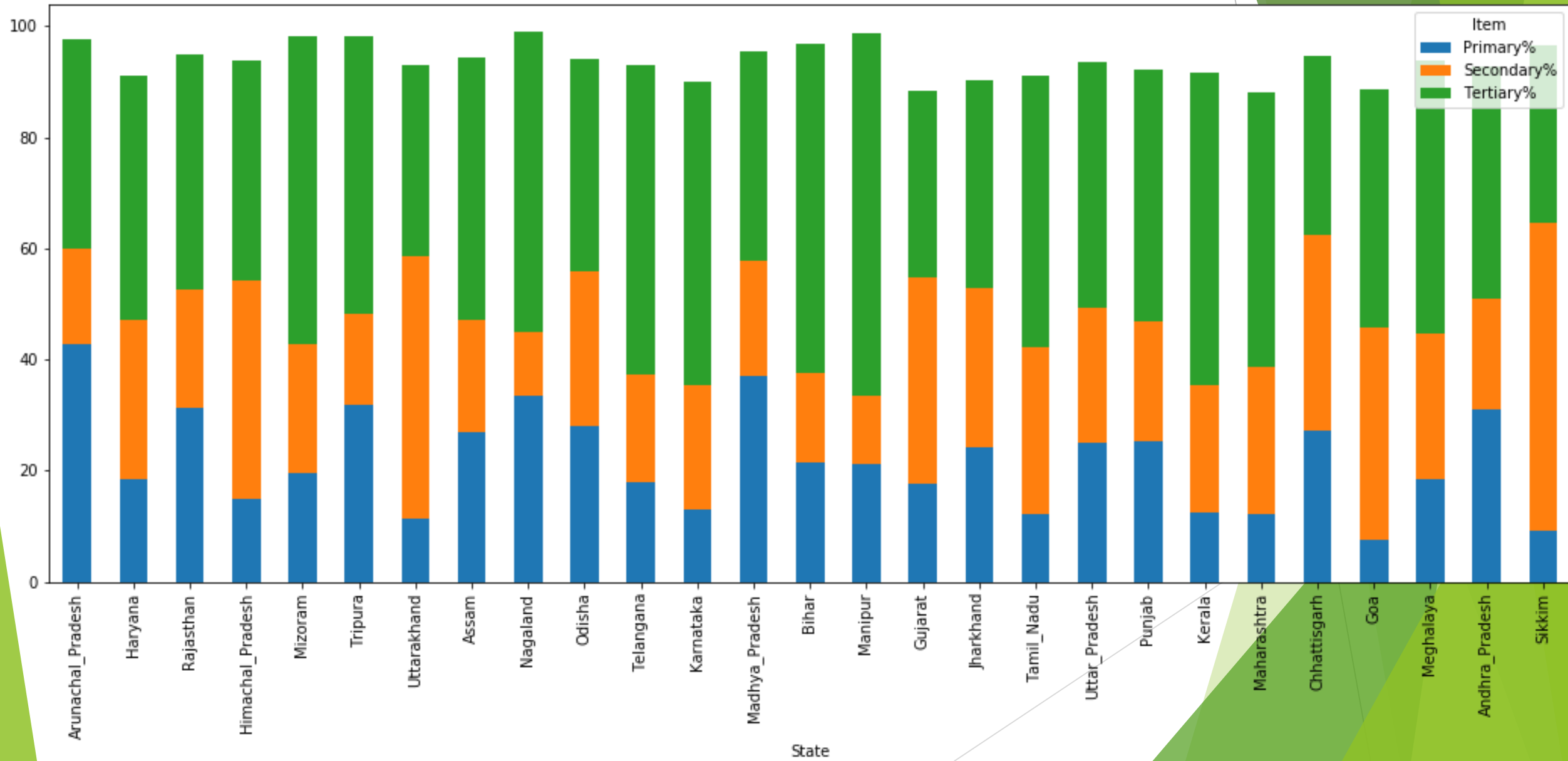
Find the ratio of the highest per capita GDP to the lowest per capita GDP.

Command used:

```
print(round(GDPPER_Capita['2014-15'].max() / GDPPER_Capita['2014-15'].min(),2))
```

Result : 8.0

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



Which plot will you use here? Why?

I have used stacked bar plot for showing 3 different quantity based values on each categorical value.

Why is (Primary + Secondary + Tertiary) not equal to total GDP?

Total GDP consists of "Subsidies on products" and "Taxes on Products" to be calculated also.

Calculation is as follows

1. Primary + Secondary + Tertiary = TOTAL GSVA at basic prices
2. TOTAL GSVA at basic prices + Taxes - Subsidies = Total GDP

Can you draw any insight from this? Find correlation of percentile of the state (% of states with lower per capita GDP) and %contribution of Primary sector to total GDP.

	Primary%	PerCapitaGDP	Rank_Pct
Primary	1.000000	-0.663396	-0.620022
PerCapitaGDP	-0.663396	1.000000	0.932707
Rank_Pct	-0.620022	0.932707	1.000000

We can clearly see a negative correlation of -0.62 between %contribution of primary sector to Total GDP percentile rank. We also noticed %contribution of primary sector is also quite significantly PerCapitaGDP with negative correlation of -0.66

Categorize the states 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), i.e., the states lying between the 85th and the 100th percentile are in C1; those between the 50th and the 85th percentiles are in C2, and so on.

Quartile	Value
0.20	65187.0
0.5	104977.0
0.85	153246.2
1	271793.0

C1 Category	C2 Category	C3 Category	C4 Category
Goa	Arunachal_Pradesh	Chhattisgarh	Assam
Haryana	Gujarat	Meghalaya	Bihar
Kerala	Himachal_Pradesh	Mizoram	Jharkhand
Sikkim	Karnataka	Nagaland	Madhya_Pradesh
	Maharashtra	Odisha	Manipur
	Punjab	Rajasthan	Uttar_Pradesh
	Tamil_Nadu	Tripura	
	Telangana	Andhra_Pradesh	
	Uttarakhand		

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) that contribute to approximately 80% of the GSDP of each category.

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

Category C1 States

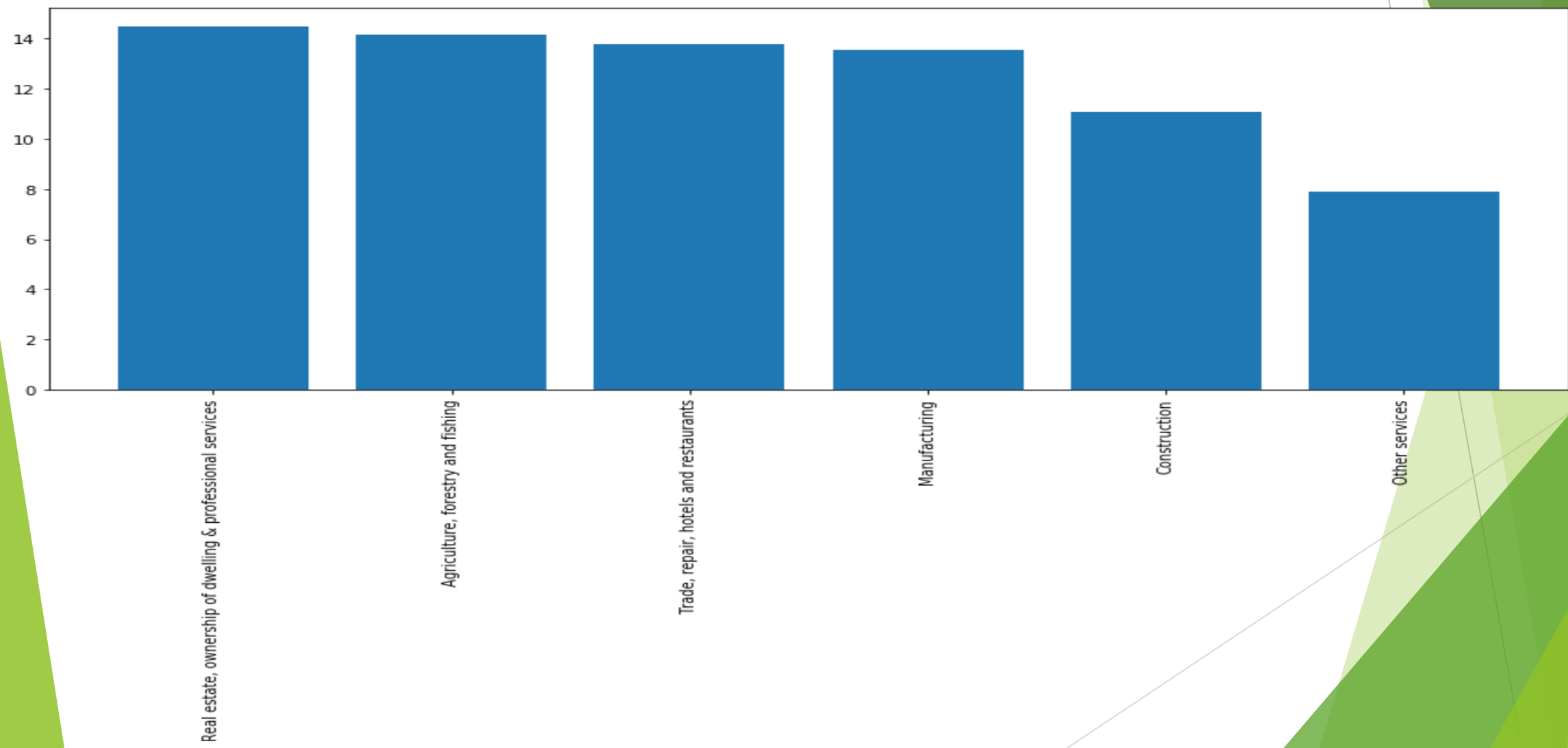
Table:

```
C1_DF = final_df[(final_df.Category== 'C1')]  
C1_DF = C1_DF.sort_values(['%Contribution'],ascending = False)  
C1_DF['CumSum'] = C1_DF['%Contribution'].cumsum()  
C1_DF[:7]
```

	Category	Item	SubContribution	GDSP	%Contribution	Cum Sum
8	C1	Real estate, ownership of dwelling & professio...	14740245.0	101930677.0	14.46	14.46
0	C1	Agriculture, forestry and fishing	14391809.0	101930677.0	14.12	28.58
9	C1	Trade, repair, hotels and restaurants	13995159.0	101930677.0	13.73	42.31
4	C1	Manufacturing	13758793.0	101930677.0	13.50	55.81
1	C1	Construction	11264451.0	101930677.0	11.05	66.86
6	C1	Other services	8059922.0	101930677.0	7.91	74.77
10	C1	Transport, storage, communication & services r...	6818560.0	101930677.0	6.69	81.46

Category C1 States

Graph:



Category C2 States

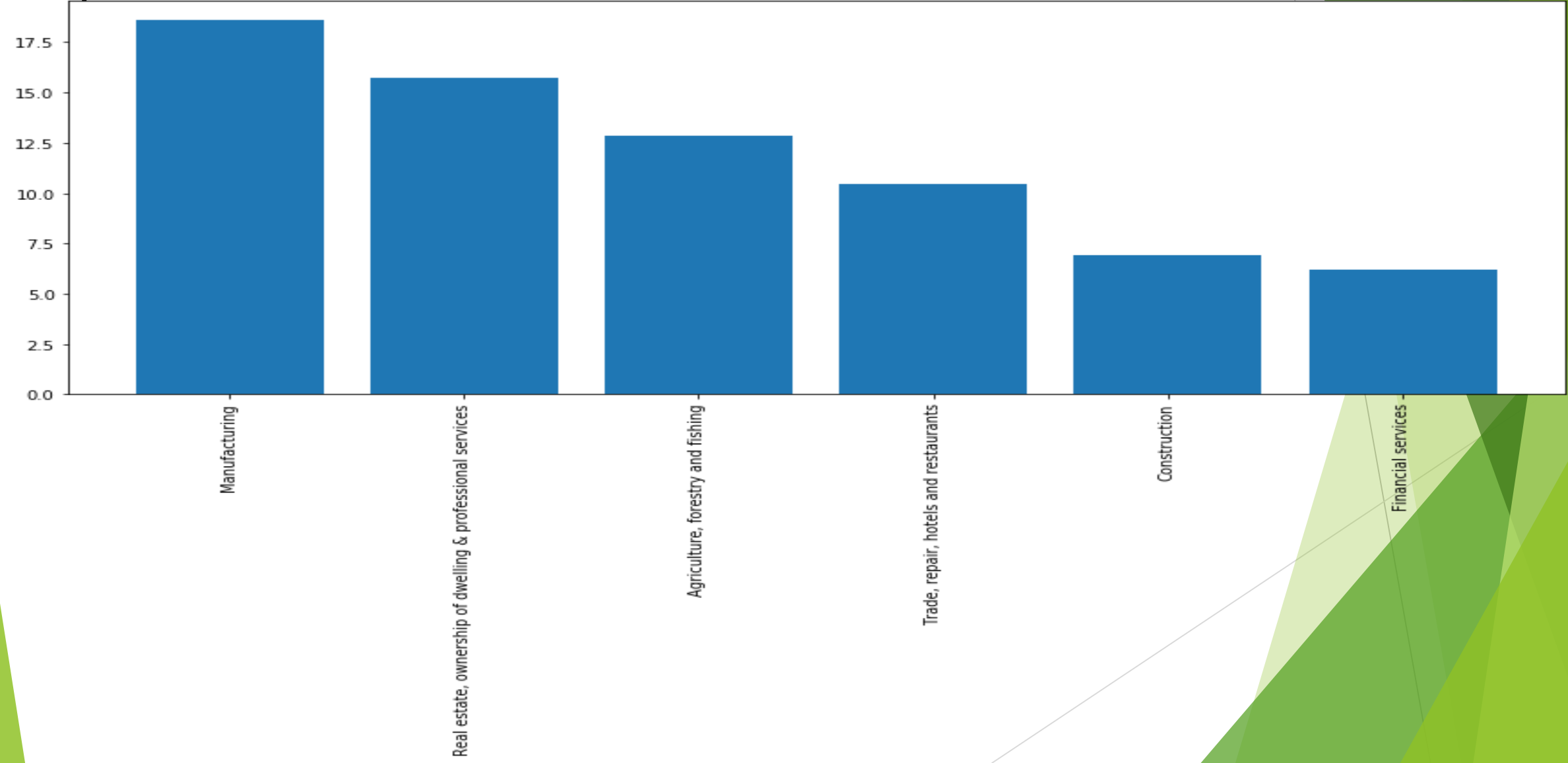
Table:

```
C2_DF = final_df[(final_df.Category == 'C2')]  
C2_DF = C2_DF.sort_values(['%Contribution'],ascending = False)  
C2_DF['CumSum'] = C2_DF['%Contribution'].cumsum()  
C2_DF[:7]
```

	Category	Item	SubContribution	GDSP	%Contribution	CumSum
15	C2	Manufacturing	109196530.0	586380452.0	18.62	18.62
19	C2	Real estate, ownership of dwelling & professio...	92121446.0	586380452.0	15.71	34.33
11	C2	Agriculture, forestry and fishing	75209022.0	586380452.0	12.83	47.16
20	C2	Trade, repair, hotels and restaurants	61238862.0	586380452.0	10.44	57.60
12	C2	Construction	40653562.0	586380452.0	6.93	64.53
14	C2	Financial services	36296642.0	586380452.0	6.19	70.72
17	C2	Other services	36253657.0	586380452.0	6.18	76.90

Category C2 States

Graph:



Category C3 States

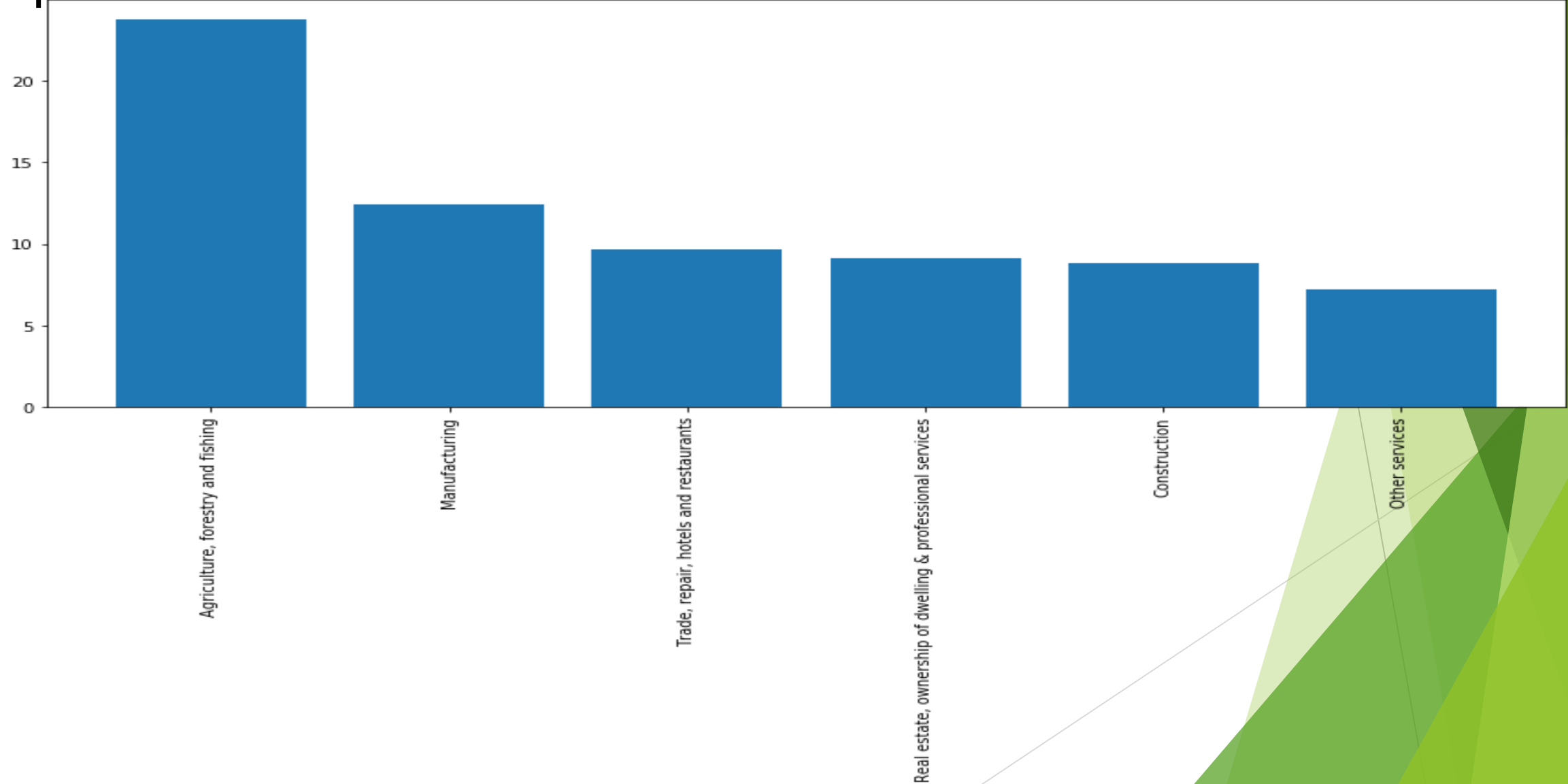
Table:

```
C3_DF = final_df[(final_df.Category == 'C3')]
C3_DF = C3_DF.sort_values(['%Contribution'],ascending = False)
C3_DF['CumSum'] = C3_DF['%Contribution'].cumsum()
C3_DF[:7]
```

	Category	Item	SubContribution	GDSP	%Contribution	CumSum
22	C3	Agriculture, forestry and fishing	42226888.0	177966387.0	23.73	23.73
26	C3	Manufacturing	22038331.0	177966387.0	12.38	36.11
31	C3	Trade, repair, hotels and restaurants	17245309.0	177966387.0	9.69	45.80
30	C3	Real estate, ownership of dwelling & professio...	16224118.0	177966387.0	9.12	54.92
23	C3	Construction	15707921.0	177966387.0	8.83	63.75
28	C3	Other services	12815417.0	177966387.0	7.20	70.95
32	C3	Transport, storage, communication & services r...	12231484.0	177966387.0	6.87	77.82

Category C3 States

Graph:



Category C4 States

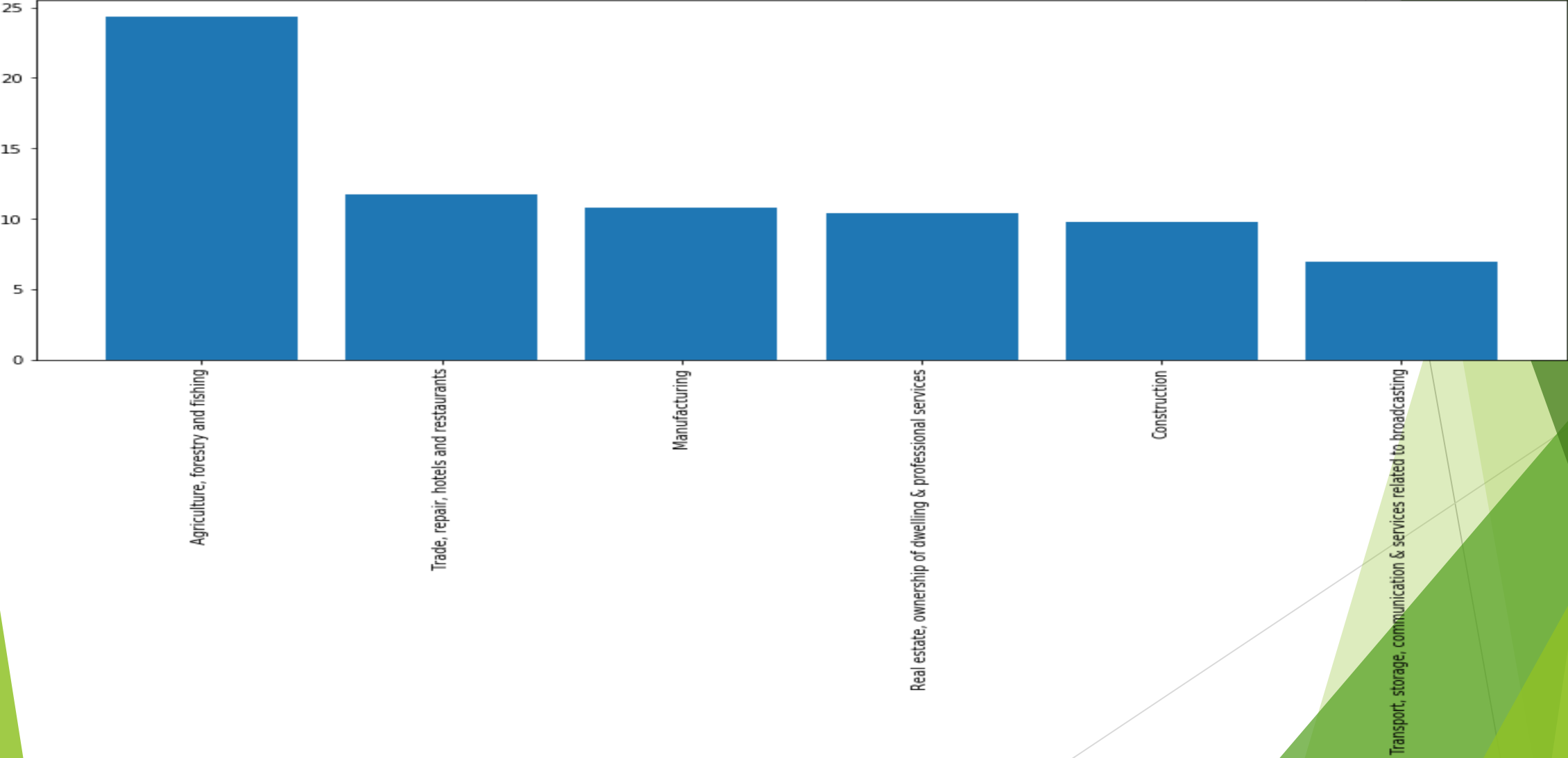
Table:

```
C4_DF = final_df[(final_df.Category == 'C4')]
C4_DF = C4_DF.sort_values(['%Contribution'],ascending = False)
C4_DF['CumSum'] = C4_DF['%Contribution'].cumsum()
C4_DF[:7]
```

	Category	Item	SubContribution	GDSP	%Contribution	CumSum
33	C4	Agriculture, forestry and fishing	56735044.0	233252066.0	24.32	24.32
42	C4	Trade, repair, hotels and restaurants	27221645.0	231447790.0	11.76	36.08
37	C4	Manufacturing	24937531.0	231447790.0	10.77	46.85
41	C4	Real estate, ownership of dwelling & professio...	24018199.0	231447790.0	10.38	57.23
34	C4	Construction	22632766.0	231447790.0	9.78	67.01
43	C4	Transport, storage, communication & services r...	16082074.0	231447790.0	6.95	73.96
39	C4	Other services	15541085.0	231447790.0	6.71	80.67

Category C4 States

Table:





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

Since we know the per capita GDP of C1 states is obviously better than category so we see majority of the GDP contribution is distributed among tertiary, primary sectors from Real estate, ownership of dwelling & professional services , Agriculture, forestry and fishing & Trade, repair, hotels and restaurants subsectors.

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

Sub Sectors like below are having high correlation with GDP

- Trade, repair, hotels and restaurants
- Transport, storage, communication & services related to broadcasting
- Other services & Manufacturing

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

Various categories should focus on more Tertiary sectors and Secondary sectors, i.e., the manufacturing and services industries.

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

1. Government should consider ways to increase per capita GDP, by involving more secondary and tertiary sectors, i.e., the manufacturing and services industries.
2. Government should also promote skilled labor for working in these areas.
3. We also see most categories are already having good contribution from Agriculture, forestry and fishing need to increase assist in other sectors as well.

Key Insights

- ▶ Its clearly observed from the correlation of subcategories that every subcategory has positive correlation with Total GDP

Analyzing category wise:

- ▶ In C1 & C2 category, more significance should be given to the following sub categories improvements
 - Financial services
 - Public administration
 - Electricity, gas, water supply & other utility
 - Mining and quarrying
- ▶ In C3 & C4 category, more significance should be given to the following sub categories improvements
 - Electricity, gas, water supply & other utility
 - Financial services
 - Public administration
 - Transport, storage, communication & services
- ▶ Hence on a high level Tertiary sectors and Second Sectors needs to improve.

Part-II: GDP and Education Dropout Rates

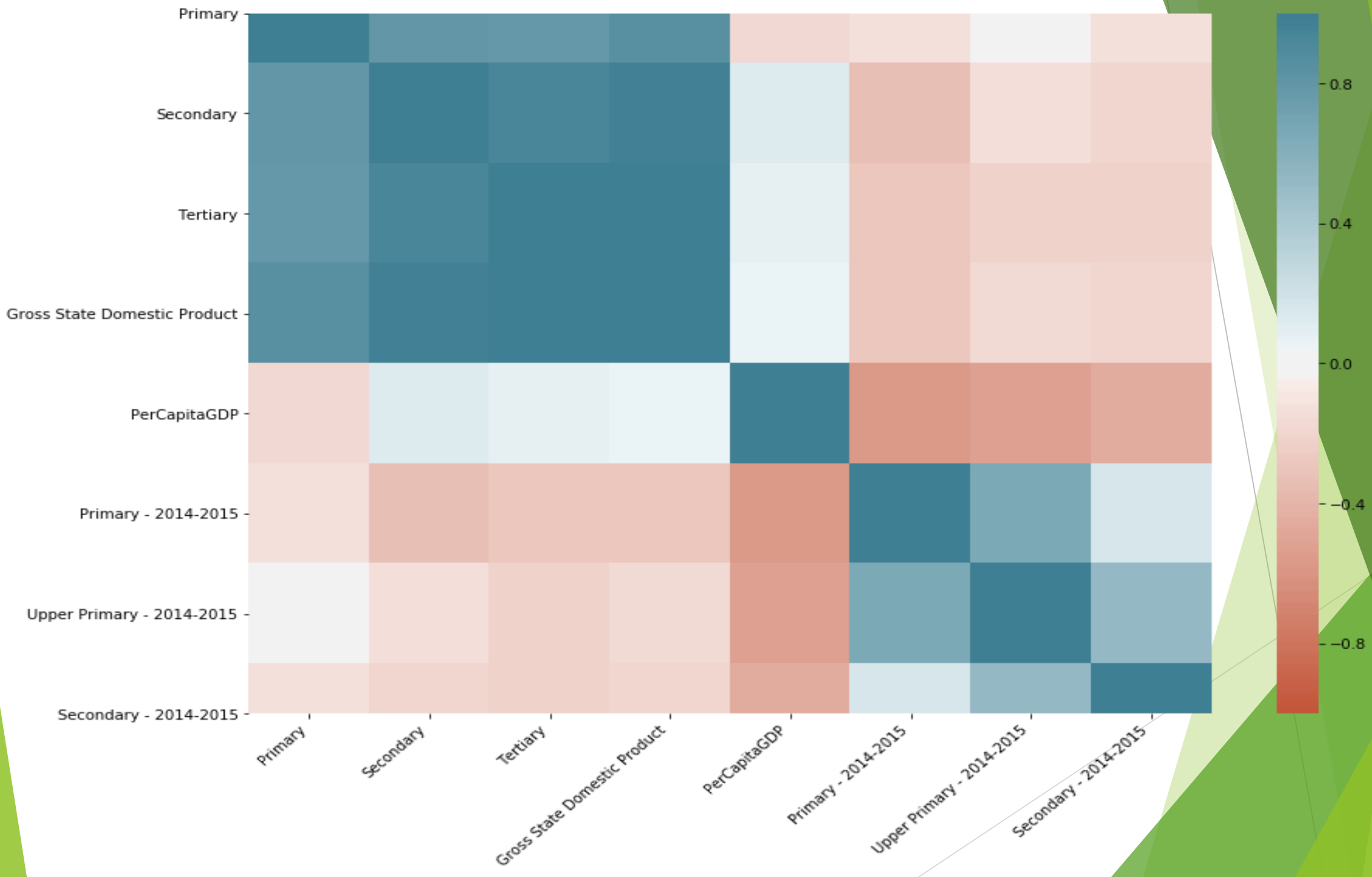
Part-II: GDP and Education

Analyze 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. Choose an appropriate plot to conduct this analysis.

Is there any correlation between dropout rate and %contribution of each sector (Primary, Secondary and Tertiary) to the total GDP?

Here are the correlation matrix and plot

	Primary	Secondary	Tertiary	Gross State Domestic Product	PerCapitaGDP	Primary - 2014-2015	Upper Primary - 2014-2015	Secondary - 2014-2015
Primary	1.000000	0.790880	0.788877	0.851747	-0.181934	-0.125473	-0.026308	-0.121918
Secondary	0.790880	1.000000	0.936539	0.974081	0.121730	-0.331897	-0.141517	-0.192913
Tertiary	0.788877	0.936539	1.000000	0.984848	0.081409	-0.286294	-0.211630	-0.214815
Gross State Domestic Product	0.851747	0.974081	0.984848	1.000000	0.061279	-0.287916	-0.163672	-0.197693
PerCapitaGDP	-0.181934	0.121730	0.081409	0.061279	1.000000	-0.574049	-0.524320	-0.446081
Primary - 2014-2015	-0.125473	-0.331897	-0.286294	-0.287916	-0.574049	1.000000	0.648542	0.167977
Upper Primary - 2014-2015	-0.026308	-0.141517	-0.211630	-0.163672	-0.524320	0.648542	1.000000	0.515896
Secondary - 2014-2015	-0.121918	-0.192913	-0.214815	-0.197693	-0.446081	0.167977	0.515896	1.000000



Yes there is a correlation between dropout rate and %contribution of each sector (Primary, Secondary and Tertiary) to the total GDP

	Gross State Domestic Product	Primary%	Secondary%	Tertiary%	PerCapitaGDP	Primary - 2014-2015	Upper Primary - 2014-2015	Secondary - 2014-2015
Gross State Domestic Product	1.000000	-0.267906	0.030890	0.009358	0.061279	-0.287916	-0.163672	-0.197693
Primary%	-0.267906	1.000000	-0.611285	-0.167047	-0.652104	0.454678	0.535019	0.335541
Secondary%	0.030890	-0.611285	1.000000	-0.635294	0.693038	-0.447592	-0.357448	-0.231190
Tertiary%	0.009358	-0.167047	-0.635294	1.000000	-0.268230	0.221519	-0.088830	-0.017582
PerCapitaGDP	0.061279	-0.652104	0.693038	-0.268230	1.000000	-0.574049	-0.524320	-0.446081
Primary - 2014-2015	-0.287916	0.454678	-0.447592	0.221519	-0.574049	1.000000	0.648542	0.167977
Upper Primary - 2014-2015	-0.163672	0.535019	-0.357448	-0.088830	-0.524320	0.648542	1.000000	0.515896
Secondary - 2014-2015	-0.197693	0.335541	-0.231190	-0.017582	-0.446081	0.167977	0.515896	1.000000

Key Insights

- ▶ It's clearly observed that there is negative correlation between GDP per capita with dropout rates in education (primary, upper primary and secondary) for the year 2014-2015 for each state.
- ▶ We have around correlation value of -0.57 between GDP per capita and primary school drop outs clearly it means more the drop outs lesser the GDP.
- ▶ Similarly we have correlation value of -0.52 between GDP per capita and Upper Primary - 2014-2015 and -0.44 between GDP per capita and Secondary - 2014-2015.
- ▶ Interestingly we can see a positive correlation of primary drop outs and %contribution of each Primary sector and negative correlation towards Secondary sector
- ▶ Similar correlation can be seen for Upper Primary and Secondary drop outs.
- ▶ Assumption is school dropouts generally earn through the following subsectors

Agriculture, forestry and fishing

Crops

Livestock

Forestry and logging

Fishing and aquaculture

Mining and quarrying

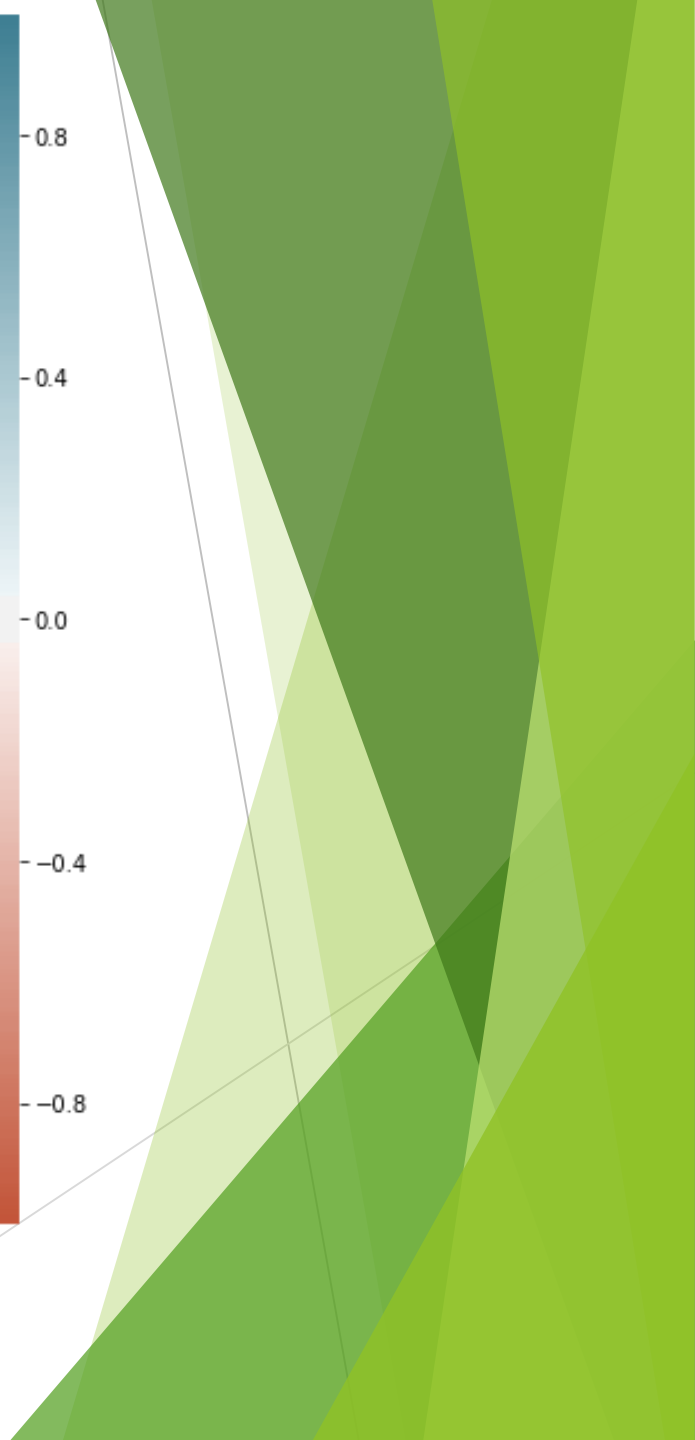
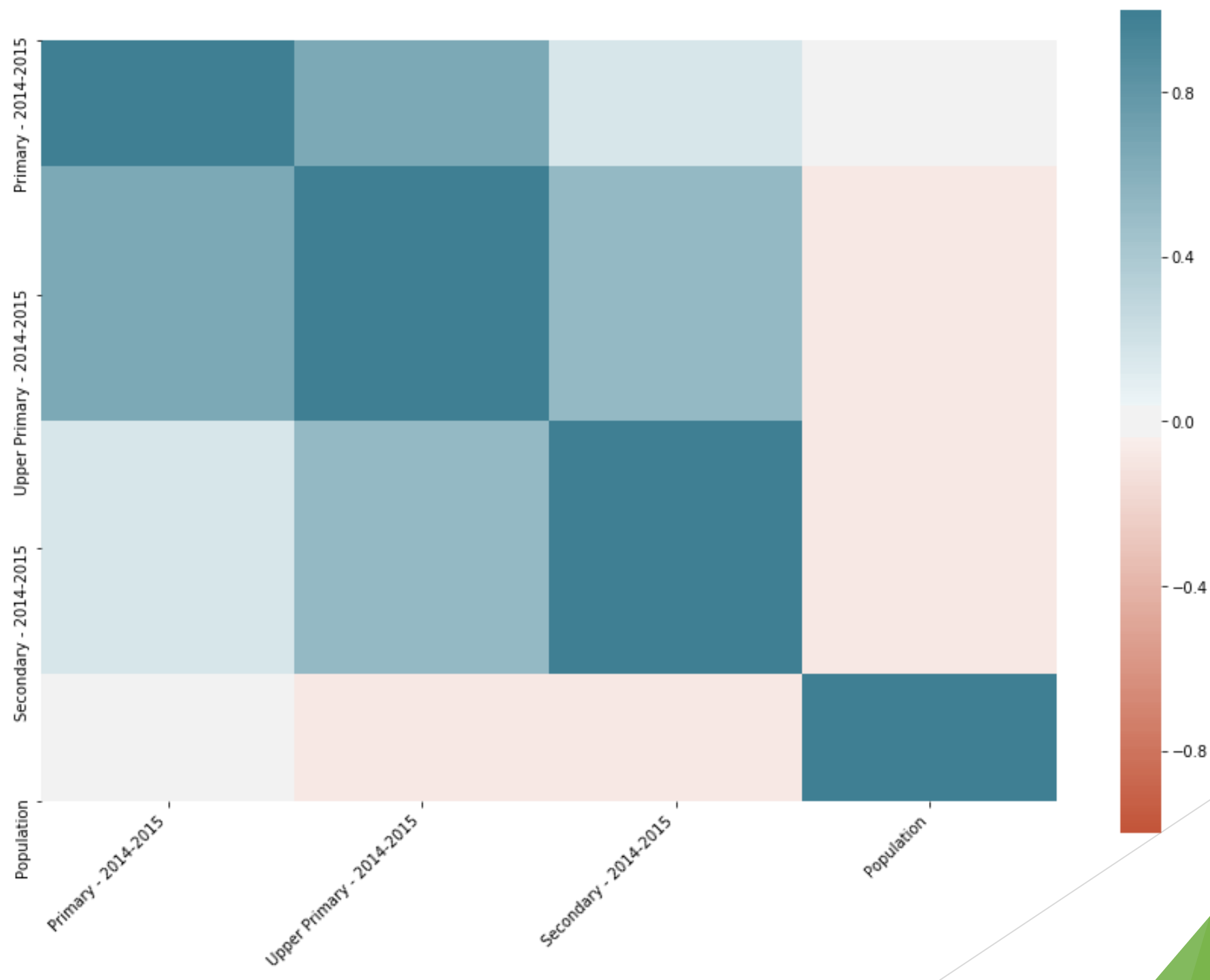
You have the total population of each state from the data in part I. Is there any correlation between dropout rates and population? What is the expected trend and what is the observation?

Write down the key insights you draw from this data:

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

Here are the correlation matrix and plot

	Primary - 2014-2015	Upper Primary - 2014-2015	Secondary - 2014-2015	Population
Primary - 2014-2015	1.000000	0.648542	0.167977	-0.014683
Upper Primary - 2014-2015	0.648542	1.000000	0.515896	-0.079808
Secondary - 2014-2015	0.167977	0.515896	1.000000	-0.083989
Population	-0.014683	-0.079808	-0.083989	1.000000



Key Insights

- ▶ Very slight negative correlation is observed between population and drop out rates.
- ▶ Expectation was more the population more the drop out rates should occur but its slightly negative is observed.
- ▶ One Hypothesis is population doesn't have to do anything with drop out rates.

END