

# Importing necessary modules

```
In [1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.patches as Patch
```

## Getting the data from the dataset

```
In [2]: #this data is of the drop out ratio from year 2012 to 2015
dropout = pd.read_csv("dropout-ratio.csv")
```

Printing the information about the dataset

```
In [3]: print(dropout.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   State_UT              110 non-null   object
1   year                  110 non-null   object
2   Primary_Boys          110 non-null   object
3   Primary_Girls         110 non-null   object
4   Primary_Total         110 non-null   object
5   Upper Primary_Boys    110 non-null   object
6   Upper Primary_Girls   110 non-null   object
7   Upper Primary_Total   110 non-null   object
8   Secondary_Boys        110 non-null   object
9   Secondary_Girls       110 non-null   object
10  Secondary_Total       110 non-null   object
11  HrSecondary_Boys      110 non-null   object
12  HrSecondary_Girls     110 non-null   object
13  HrSecondary_Total     110 non-null   object
dtypes: object(14)
memory usage: 12.2+ KB
None
```

In [4]: `print(dropout.describe())`

	State_UT	year	Primary_Boys	Primary_Girls	Primary_Total	\
count	110	110	110	110	110	
unique	39	3	93	87	87	
top	A & N Islands	2013-14	NR	NR	NR	
freq	3	37	15	15	17	

	Upper	Primary_Boys	Upper	Primary_Girls	Upper	Primary_Total	\
count		110		110		110	
unique		91		97		96	
top		NR		NR		NR	
freq		16		8		9	

	Secondary_Boys	Secondary_Girls	Secondary_Total	HrSecondary_Boys	\
count	110	110	110	110	
unique	106	107	106	81	
top	NR	NR	NR	NR	
freq	4	4	4	28	

	HrSecondary_Girls	HrSecondary_Total
count	110	110
unique	76	78
top	NR	NR
freq	34	31

In [5]: `print(dropout.head())`

	State_UT	year	Primary_Boys	Primary_Girls	Primary_Total	\
0	A & N Islands	2012-13	0.83	0.51	0.68	
1	A & N Islands	2013-14	1.35	1.06	1.21	
2	A & N Islands	2014-15	0.47	0.55	0.51	
3	Andhra Pradesh	2012-13	3.3	3.05	3.18	
4	Andhra Pradesh	2013-14	4.31	4.39	4.35	

	Upper	Primary_Boys	Upper	Primary_Girls	Upper	Primary_Total	Secondary_Boys	\
0	Uppe_r_Primary		1.09		1.23		5.57	
1	NR		1.54		0.51		8.36	
2	1.44		1.95		1.69		11.47	
3	3.21		3.51		3.36		12.21	
4	3.46		4.12		3.78		11.95	

	Secondary_Girls	Secondary_Total	HrSecondary_Boys	HrSecondary_Girls	\
0	5.55	5.56	17.66	10.15	
1	5.98	7.2	18.94	12.2	
2	8.16	9.87	21.05	12.21	
3	13.25	12.72	2.66	NR	
4	13.37	12.65	12.65	10.85	

	HrSecondary_Total
0	14.14
1	15.87
2	16.93
3	0.35
4	11.79

## Q1. What is the dropout ratio in Gujarat and India

```
In [6]: dr_in_gj = dropout[dropout['State_UT'] == 'Gujarat'].reset_index(drop=True)
dr_in_india = dropout[dropout['State_UT'] == 'All India'].reset_index(drop=True)
```

```
In [7]: print("Dropout ratio in Gujarat")
display(dr_in_gj)
```

Dropout ratio in Gujarat

	State_UT	year	Primary_Boys	Primary_Girls	Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Pri
0	Gujarat	2012-13	0.21	1.35	0.74	2.75	8.19	
1	Gujarat	2013-14	0.5	1.06	0.76	3.52	8.04	
2	Gujarat	2014-15	0.82	0.98	0.89	4.65	8.54	

```
In [8]: print("Dropout ratio in India")
display(dr_in_india)
```

Dropout ratio in India

	State_UT	year	Primary_Boys	Primary_Girls	Primary_Total	Upper Primary_Boys	Upper Primary_Girls	Pri
0	All India	2012-13	4.68	4.66	4.67	2.3	4.01	
1	All India	2013-14	4.53	4.14	4.34	3.09	4.49	
2	All India	2014-15	4.36	3.88	4.13	3.49	4.6	

## Q2. Calculate total enrollement for each state in year 2014-2015

```
In [9]: total_enrollment_2014_15 = dropout[dropout['year'] == '2014-15'].groupby('State_UT')
```

```
In [10]: display(total_enrollment_2014_15)
```

	Primary_Total	Upper Primary_Total	Secondary_Total	HrSecondary_Total
State_UT				
A & N Islands	0.51	1.69	9.87	16.93
All India	4.13	4.03	17.06	NR
Andhra Pradesh	6.72	5.2	15.71	NR
Arunachal Pradesh	10.82	6.71	17.11	18.42
Assam	15.36	10.51	27.06	NR
Bihar	NR	4.08	25.9	NR
Chandigarh	NR	0.44	NR	10.55
Chhattisgarh	2.91	5.85	21.26	2.76
Dadra & Nagar Haveli	1.47	4.02	16.77	9.47
Daman & Diu	1.11	3.11	32.27	40.48
Delhi	NR	0.76	11.81	17.32
Goa	0.73	0.07	11.15	13.91
Gujarat	0.89	6.41	25.04	7.04
Haryana	5.61	5.81	15.89	5.75
Himachal Pradesh	0.64	0.87	6.07	7.41
Jammu & Kashmir	6.79	5.44	17.28	12.65
Jharkhand	5.48	8.99	24	3.41
Karnataka	2.02	3.85	26.18	1.96
Kerala	NR	NR	12.32	0.47
Lakshadweep	NR	2.78	6.76	3.12
Madhya Pradesh	6.59	9.2	24.77	NR
Maharashtra	1.26	1.79	12.87	1.83
Manipur	9.66	4.2	14.38	NR
Meghalaya	9.46	6.52	20.52	NR
Mizoram	10.1	4.78	21.88	6.91
Nagaland	5.61	7.92	18.23	6.97
Odisha	2.86	3.81	29.56	NR
Puducherry	0.37	0.56	12.19	4.5
Punjab	3.05	3.22	8.86	5.83
Rajasthan	5.02	3.07	13.48	NR
Sikkim	2.27	1.57	15.89	11.76
Tamil Nadu	NR	NR	8.1	3.41
Telangana	2.08	2.3	15.53	0.77
Tripura	1.28	1.99	28.42	8.93
Uttar Pradesh	8.58	2.7	10.22	2.1

	Primary_Total	Upper Primary_Total	Secondary_Total	HrSecondary_Total
State_UT				
Uttarakhand	4.04	1.19	10.4	3.01
West Bengal	1.47	4.3	17.8	8.11

### Q3. Display the state having the highest primary enrollement in 2014-2015

```
In [11]: dr_2014_15 = dropout[dropout['year']=='2014-15']
```

```
In [12]: #Highest primary enrollement
State_with_highest_primaryenrollement = dr_2014_15.loc[pd.to_numeric(dr_2014_15
print(State_with_highest_primaryenrollement)
```

Assam

```
In [13]: #Highest upper primary enrollement
State_with_highest_upperprimaryenrollement = dr_2014_15.loc[pd.to_numeric(dr_20
print(State_with_highest_upperprimaryenrollement)
```

Assam

```
In [14]: #Highest seconary enrollement
State_with_highest_secondaryenrollement = dr_2014_15.loc[pd.to_numeric(dr_2014_
print(State_with_highest_secondaryenrollement)
```

Daman & Diu

### Q4. Display the state with highest primary girls drop out ratio in 2014-2015

```
In [15]: #Highest primary_girls enrollement
State_with_highest_primary_girl = dr_2014_15.loc[pd.to_numeric(dr_2014_15['Prim
print(State_with_highest_primary_girl)
```

Assam

### Q5. Plot the trend of the primary enrollement over the years for a specific state

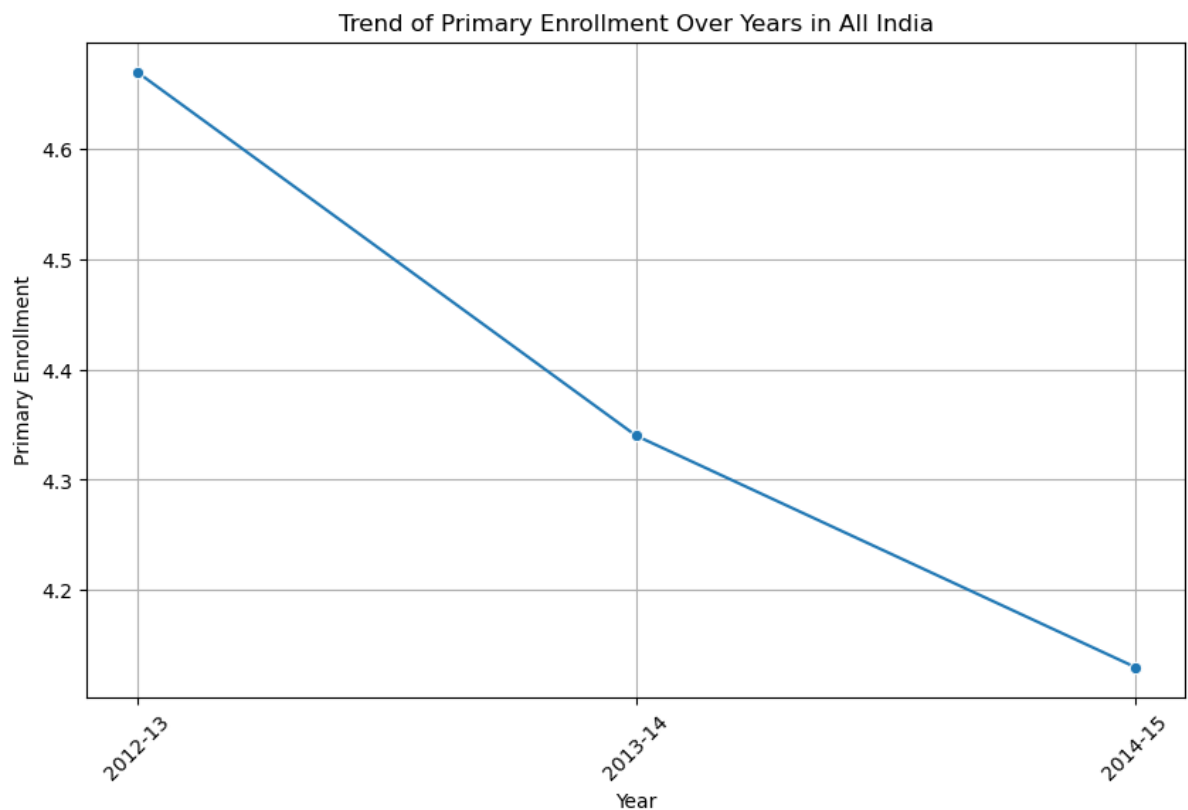
```
In [16]: states = dropout['State_UT'].drop_duplicates().reset_index(drop=True)
display(states)
state = input("Enter the state you want to know about: ")
```

```
0          A & N Islands
1          Andhra Pradesh
2          Arunachal Pradesh
3          Arunachal Pradesh
4              Assam
5              Bihar
6              Chandigarh
7              Chhattisgarh
8      Dadra & Nagar Haveli
9              Daman & Diu
10             Delhi
11             Goa
12             Gujarat
13             Haryana
14      Himachal Pradesh
15      Jammu & Kashmir
16             Jharkhand
17             Karnataka
18             Kerala
19             Lakshadweep
20      Madhya Pradesh
21      Madhya Pradesh
22      Maharashtra
23             Manipur
24             Meghalaya
25             Mizoram
26             Nagaland
27             Odisha
28      Puducherry
29             Punjab
30             Rajasthan
31             Sikkim
32             Tamil Nadu
33             Telangana
34             Tripura
35      Uttar Pradesh
36      Uttarakhand
37      West Bengal
38      All India
Name: State_UT, dtype: object
```

Enter the state you want to know about:

```
In [17]: if state not in dropout['State_UT'].values:
state = "All India"
state_data = dropout[dropout['State_UT'] == state]
state_data.loc[:, 'Primary_Total'] = pd.to_numeric(state_data['Primary_Total'],
state_data = state_data.dropna(subset=['Primary_Total']) #drops data with Nan v
```

```
In [18]: plt.figure(figsize=(10, 6))
sns.lineplot(x='year', y='Primary_Total', data=state_data, marker='o', linestyle='solid')
plt.title(f'Trend of Primary Enrollment Over Years in {state}')
plt.xlabel('Year')
plt.ylabel('Primary Enrollment')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



## Q6. Compare the drop out ratio of girls and boys in upper primary across all states

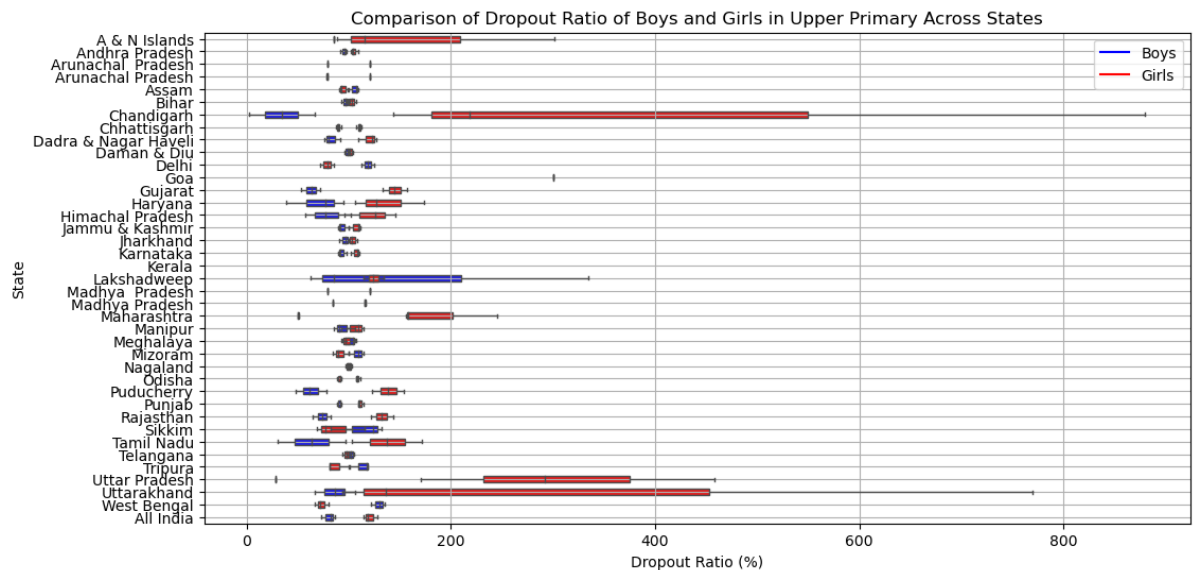
```
In [19]: Upper_primary = dropout.dropna(subset=['Upper Primary_Total', 'Upper Primary_Boys', 'Upper Primary_Girls'])

Upper_primary['Upper Primary_Girls'] = pd.to_numeric(Upper_primary['Upper Primary_Girls'])
Upper_primary['Upper Primary_Total'] = pd.to_numeric(Upper_primary['Upper Primary_Total'])
Upper_primary['Upper Primary_Boys'] = pd.to_numeric(Upper_primary['Upper Primary_Boys'])

Upper_primary['Dropout_Ratio_Boys'] = (Upper_primary['Upper Primary_Boys'] / Upper_primary['Upper Primary_Total'])
Upper_primary['Dropout_Ratio_Girls'] = (Upper_primary['Upper Primary_Girls'] / Upper_primary['Upper Primary_Total'])
```



```
In [20]: plt.figure(figsize=(12, 6))
sns.boxplot(x='Dropout_Ratio_Boys', y='State_UT', data=Upper_primary, color='blue')
sns.boxplot(x='Dropout_Ratio_Girls', y='State_UT', data=Upper_primary, color='red')
plt.title('Comparison of Dropout Ratio of Boys and Girls in Upper Primary Across States')
plt.xlabel('Dropout Ratio (%)')
plt.ylabel('State')
plt.plot([], [], color='blue', label='Boys')
plt.plot([], [], color='red', label='Girls')
plt.legend(loc='upper right')
plt.grid(True)
plt.show()
```



**Q7. Calculate the average enrollement for each education level ( Primary, Upper Primary,Secondary, Higher Secondary) across all states**

```
In [21]: columns = ['Primary_Total', 'Upper Primary_Total', 'Secondary_Total', 'HrSecor']
dropout[columns] = dropout[columns].apply(pd.to_numeric, errors='coerce')
```

```
In [22]: Average_enrollement = dropout.groupby('State_UT')[columns].mean()  
print(Average_enrollement)
```

State_UT	Primary_Total	Upper Primary_Total	Secondary_Total \
A & N Islands	0.800000	1.143333	7.543333
All India	4.380000	3.643333	16.486667
Andhra Pradesh	4.750000	4.113333	13.693333
Arunachal Pradesh	10.890000	5.590000	14.490000
Arunachal Pradesh	12.990000	7.090000	15.020000
Assam	9.680000	8.253333	28.086667
Bihar	2.090000	3.530000	27.123333
Chandigarh	NaN	0.540000	NaN
Chhattisgarh	2.823333	5.023333	19.843333
Dadra & Nagar Haveli	1.520000	3.843333	23.256667
Daman & Diu	0.980000	3.335000	21.216667
Delhi	NaN	1.770000	8.676667
Goa	0.465000	0.070000	9.563333
Gujarat	0.796667	5.720000	20.066667
Haryana	2.436667	2.940000	12.460000
Himachal Pradesh	0.516667	0.720000	7.843333
Jammu & Kashmir	6.183333	5.086667	16.656667
Jharkhand	6.366667	7.293333	21.883333
Karnataka	2.436667	3.803333	31.223333
Kerala	NaN	NaN	12.076667
Lakshadweep	2.500000	1.643333	7.490000
Madhya Pradesh	6.110000	8.530000	13.630000
Madhya Pradesh	8.365000	10.450000	25.620000
Maharashtra	0.926667	1.380000	14.513333
Manipur	12.506667	5.760000	12.646667
Meghalaya	9.980000	7.070000	23.766667
Mizoram	15.723333	10.026667	20.666667
Nagaland	10.696667	11.870000	26.616667
Odisha	3.120000	3.473333	42.966667
Puducherry	0.456667	0.613333	13.880000
Punjab	2.110000	2.950000	9.163333
Rajasthan	7.126667	4.506667	15.300000
Sikkim	3.480000	3.286667	12.670000
Tamil Nadu	2.260000	2.880000	10.166667
Telangana	3.945000	3.505000	16.480000
Tripura	2.360000	2.596667	26.336667
Uttar Pradesh	8.646667	1.430000	8.760000
Uttarakhand	2.756667	1.000000	9.593333
West Bengal	3.560000	4.596667	17.813333

State_UT	HrSecondary_Total
A & N Islands	15.646667
All India	1.540000
Andhra Pradesh	6.070000
Arunachal Pradesh	17.070000
Arunachal Pradesh	11.765000
Assam	5.965000
Bihar	NaN
Chandigarh	11.826667
Chhattisgarh	2.760000
Dadra & Nagar Haveli	8.453333
Daman & Diu	17.200000
Delhi	16.013333
Goa	12.843333

Gujarat	5.110000
Haryana	3.965000
Himachal Pradesh	7.833333
Jammu & Kashmir	10.316667
Jharkhand	3.410000
Karnataka	8.645000
Kerala	3.435000
Lakshadweep	4.176667
Madhya Pradesh	NaN
Madhya Pradesh	1.550000
Maharashtra	2.790000
Manipur	5.565000
Meghalaya	NaN
Mizoram	6.910000
Nagaland	13.150000
Odisha	NaN
Puducherry	6.970000
Punjab	6.640000
Rajasthan	NaN
Sikkim	10.923333
Tamil Nadu	3.240000
Telangana	7.100000
Tripura	8.970000
Uttar Pradesh	2.100000
Uttarakhand	2.300000
West Bengal	8.046667

## Q8. Determine the state with Highest Secondary enrollement in 2013-2014

```
In [23]: dr_2013_14 = dropout[dropout['year']=='2013-14']
```

```
In [24]: #Highest primary enrollement
State_with_highest_secondary_enrollement = dr_2013_14.loc[pd.to_numeric(dr_2013_14['secondary_enrollement']) > 0]
print("State:\t",State_with_highest_secondary_enrollement)
```

State: Odisha

## Q9. Find the state with lowest enrollement in each education level in 2014-2015

```
In [25]: State_with_lowest_primary_dropout_ratio = dr_2014_15.loc[pd.to_numeric(dr_2014_15['primary_dropout_ratio']) < 0]
State_with_lowest_upperprimary_dropout_ratio = dr_2014_15.loc[pd.to_numeric(dr_2014_15['upperprimary_dropout_ratio']) < 0]
State_with_lowest_secondary_dropout_ratio = dr_2014_15.loc[pd.to_numeric(dr_2014_15['secondary_dropout_ratio']) < 0]
State_with_lowest_highersecondary_dropout_ratio = dr_2014_15.loc[pd.to_numeric(dr_2014_15['highersecondary_dropout_ratio']) < 0]
```

```
In [26]: print("State with lowest primary dropout_ratio----\t",State_with_lowest_primary)
print("State with lowest upperprimary dropout_ratio\t",State_with_lowest_upperp
print("State with lowest secondary dropout_ratio---\t",State_with_lowest_second
print("State with lowest highersecondary dropout_ratio\t",State_with_lowest_hig
```

```
State with lowest primary dropout_ratio----      Puducherry
State with lowest upperprimary dropout_ratio      Goa
State with lowest secondary dropout_ratio---      Himachal Pradesh
State with lowest highersecondary dropout_ratio  Kerala
```

## Q10. Identify the state with highest enrollement growth rate from 2012-2013 to 2014-2015

```
In [27]: dr_2012_13 = dropout[dropout['year']=='2012-13']
#data for 2014-15 is already there in dr_2014_15
```

```
In [34]: #columns are also defined earlier
dr_2012_13.loc[:,columns] = dr_2012_13[columns].apply(pd.to_numeric, errors='co
dr_2014_15.loc[:,columns] = dr_2014_15[columns].apply(pd.to_numeric, errors='co
```

```
In [35]: total_gr1213 = dr_2012_13.groupby('State_UT')[columns].sum().sum(axis=1)
total_gr1415 = dr_2014_15.groupby('State_UT')[columns].sum().sum(axis=1)
```

```
In [37]: Enrollement_growth_rate = ((total_gr1415 - total_gr1213)/total_gr1213)*100
max_Enrollment_growth_rate = Enrollement_growth_rate.idxmax()
print(max_Enrollment_growth_rate)
```

Daman & Diu

## Q11. Investigate the correlation between primary enrollement and dropout ratio for each state

```
In [38]: primary_columns = ['Primary_Boys', 'Primary_Girls', 'Primary_Total']
dropout[primary_columns] = dropout[primary_columns].apply(pd.to_numeric, errors
```

```
In [39]: dropout_cleaned = dropout.dropna(subset=primary_columns)
correlation_coefficient = dropout_cleaned.groupby('State_UT')[primary_columns].
display(correlation_coefficient)
```

```
State_UT
A & N Islands      0.885012
All India          0.974527
Andhra Pradesh     0.999008
Arunachal Pradesh  NaN
Arunachal Pradesh  1.000000
Assam              0.999994
Bihar              NaN
Chhattisgarh       0.996292
Dadra & Nagar Haveli -1.000000
Daman & Diu         -1.000000
Goa                1.000000
Gujarat            -0.941060
Haryana            0.988487
Himachal Pradesh   -0.885520
Jammu & Kashmir     0.811652
Jharkhand           0.962555
Karnataka           0.991725
Lakshadweep        NaN
Madhya Pradesh      NaN
Madhya Pradesh     1.000000
Maharashtra         0.969429
Manipur             0.992840
Meghalaya           0.845922
Mizoram             0.997318
Nagaland            0.998102
Odisha              0.902049
Puducherry          0.894702
Punjab              0.999058
Rajasthan           0.990801
Sikkim              0.999819
Tamil Nadu          1.000000
Telangana           1.000000
Tripura             0.999711
Uttar Pradesh       0.999299
Uttarakhand         1.000000
West Bengal         0.998767
dtype: float64
```

**Q12. Compare the enrollement distribution across different education level for specific state**

```
In [40]: display(states)
state_e = input("Enter the state you want to know about: ")
if state_e not in dropout['State_UT'].values:
    state_e = "All India"
```

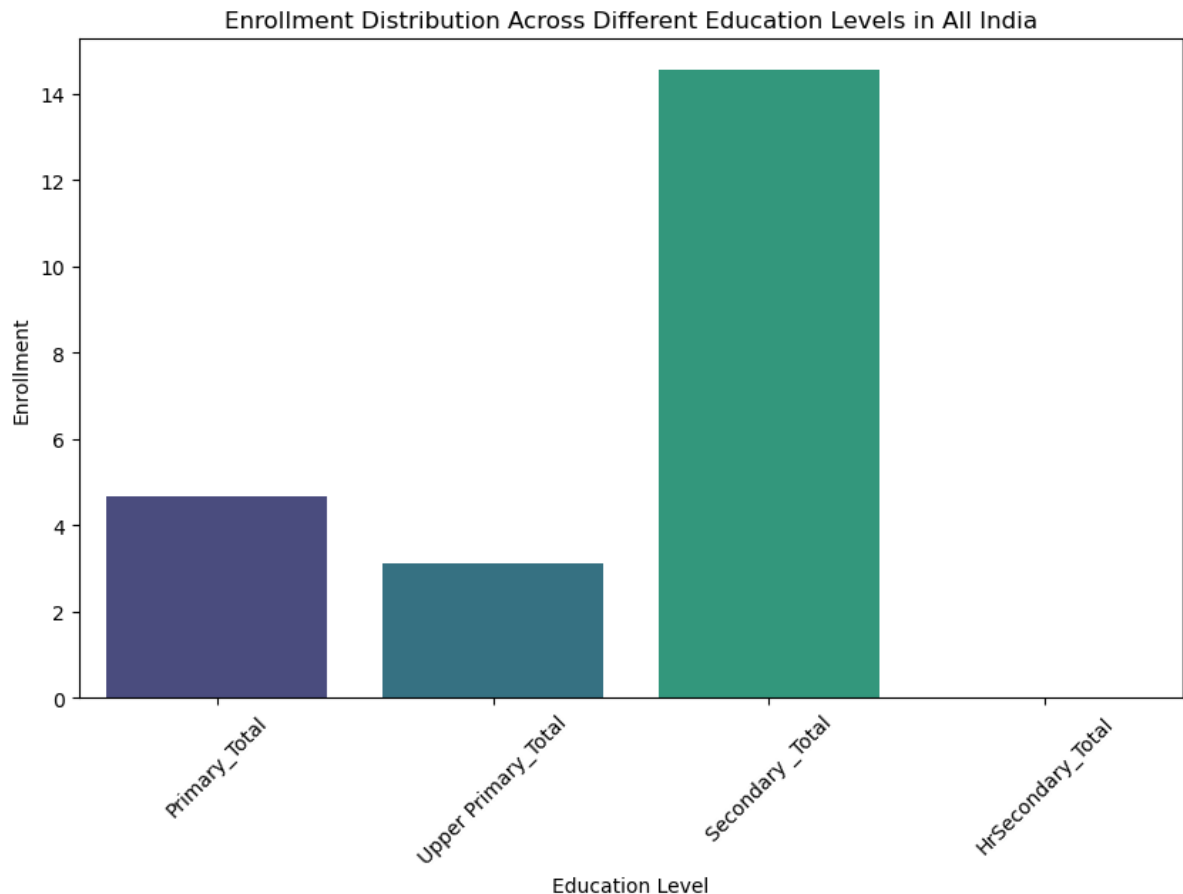
```
0          A & N Islands
1          Andhra Pradesh
2          Arunachal Pradesh
3          Arunachal Pradesh
4              Assam
5              Bihar
6              Chandigarh
7          Chhattisgarh
8      Dadra & Nagar Haveli
9          Daman & Diu
10         Delhi
11         Goa
12         Gujarat
13         Haryana
14     Himachal Pradesh
15     Jammu & Kashmir
16         Jharkhand
17         Karnataka
18         Kerala
19         Lakshadweep
20     Madhya Pradesh
21     Madhya Pradesh
22     Maharashtra
23         Manipur
24         Meghalaya
25         Mizoram
26         Nagaland
27         Odisha
28         Puducherry
29         Punjab
30         Rajasthan
31         Sikkim
32         Tamil Nadu
33         Telangana
34         Tripura
35     Uttar Pradesh
36     Uttarakhand
37         West Bengal
38         All India
Name: State_UT, dtype: object
```

Enter the state you want to know about:

```
In [41]: state_data_e = dropout[dropout['State_UT'] == state_e]
```

```
In [42]: enrollment_data = state_data_e[columns].iloc[0]
```

```
In [43]: plt.figure(figsize=(10, 6))
sns.barplot(x=enrollment_data.index, y=enrollment_data.values, palette='viridis')
plt.title(f'Enrollment Distribution Across Different Education Levels in {state}')
plt.xlabel('Education Level')
plt.ylabel('Enrollment')
plt.xticks(rotation=45)
plt.show()
```



### Q13 Identify the state with highest dropout ratio of primary boys in 2014-2015

```
In [50]: # data for the year 2014-15 is already stored in dr_2014_15
dr_2014_15.loc[:, 'Primary_Boys'] = pd.to_numeric(dr_2014_15['Primary_Boys'], errors='coerce')
```

```
In [51]: max_index = dr_2014_15['Primary_Boys'].idxmax()
state_highest_ratio = dr_2014_15.loc[max_index, 'State_UT']
print("State:\t", state_highest_ratio)
```

State: Assam

### Q14 Calculate the total dropout ratio for each education level across all states.



```
In [52]: Total_dropout = dropout[columns].sum()
print(Total_dropout)
```

```
Primary_Total          456.50
Upper_Primary_Total    422.29
Secondary_Total        1843.75
HrSecondary_Total      655.41
dtype: float64
```

## Q15 Find the state with the highest drop out in each education level in 2012-13.

```
In [53]: dr_2012_13 = dr_2012_13.reset_index()
State_with_highest_primary_dropout_ratio1213 = dr_2012_13.loc[pd.to_numeric(dr_
State_with_highest_upperprimary_dropout_ratio1213 = dr_2012_13.loc[pd.to_numeri
State_with_highest_secondary_dropout_ratio1213 = dr_2012_13.loc[pd.to_numeric(c
State_with_highest_highersecondary_dropout_ratio1213 = dr_2012_13.loc[pd.to_nun
```

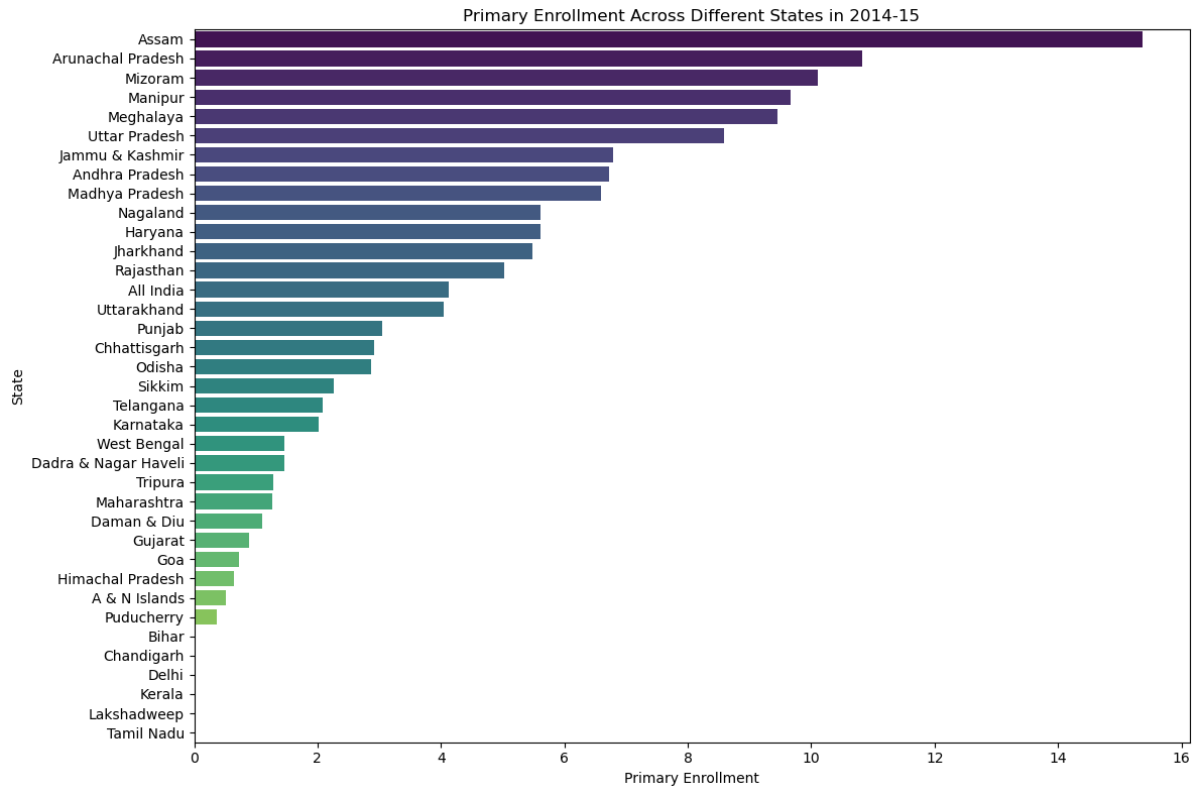
```
In [54]: print("State with highest primary dropout_ratio----\t",State_with_highest_prima
print("State with highest upperprimary dropout_ratio\t",State_with_highest_uppe
print("State with highest secondary dropout_ratio---\t",State_with_highest_secc
print("State with highest highersecondary dropout_ratio\t",State_with_highest_h
```

```
State with highest primary dropout_ratio----      Gujarat
State with highest upperprimary dropout_ratio      Gujarat
State with highest secondary dropout_ratio---      Puducherry
State with highest highersecondary dropout_ratio      Puducherry
```

## Q16 Plot a graph to show Primary dropout across different states in year 2014-15

```
In [55]: dropout_primary = dr_2014_15.sort_values(by= 'Primary_Total',ascending=False)
```

```
In [56]: plt.figure(figsize=(12, 8))
sns.barplot(x='Primary_Total', y='State_UT', data=dropout_primary, palette='vir
plt.title('Primary Enrollment Across Different States in 2014-15')
plt.xlabel('Primary Enrollment')
plt.ylabel('State')
plt.tight_layout()
plt.show()
```



## Q17 Find the average drop out ratio for each education level

```
In [99]: dropout['Primary_Total'] = pd.to_numeric(dropout['Primary_Total'], errors='coer
dropout['Upper Primary_Total'] = pd.to_numeric(dropout['Upper Primary_Total'],
dropout['Secondary_Total'] = pd.to_numeric(dropout['Secondary_Total'], errors
dropout['HrSecondary_Total'] = pd.to_numeric(dropout['HrSecondary_Total'], erro
```

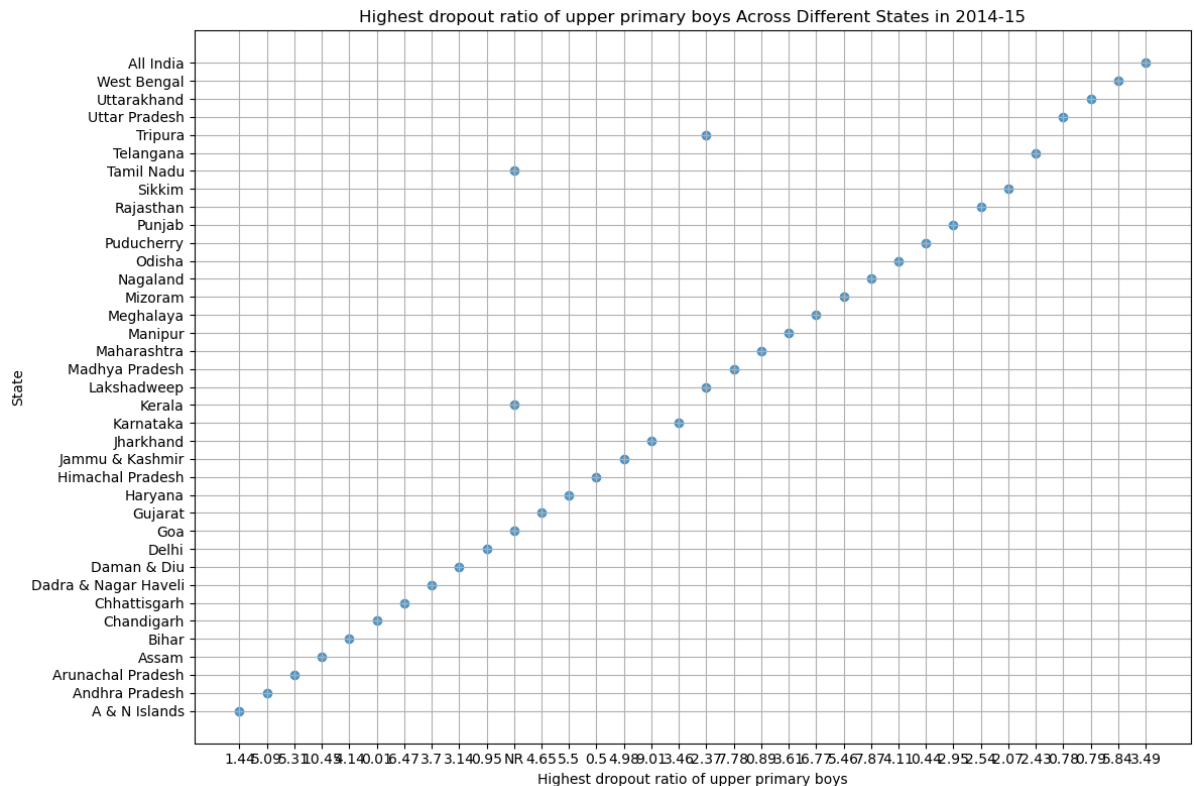
```
In [100]: avg=dropout['Primary_Total'].mean()
print(f"The average dropout for Primary total: {avg:.2f}")
avg=dropout['Upper Primary_Total'].mean()
print(f"The average dropout for Upper Primary total: {avg:.2f}")
avg=dropout['Secondary_Total'].mean()
print(f"The average dropout for Secondary total: {avg:.2f}")
avg=dropout['HrSecondary_Total'].mean()
print(f"The average dropout for Higher secondary total: {avg:.2f}")
```

The average dropout for Primary total: 4.91  
The average dropout for Upper Primary total: 4.18  
The average dropout for Secondary total: 17.39  
The average dropout for Higher secondary total: 8.30

## Q18 Visualize the highest dropout ratio of upper primary boys

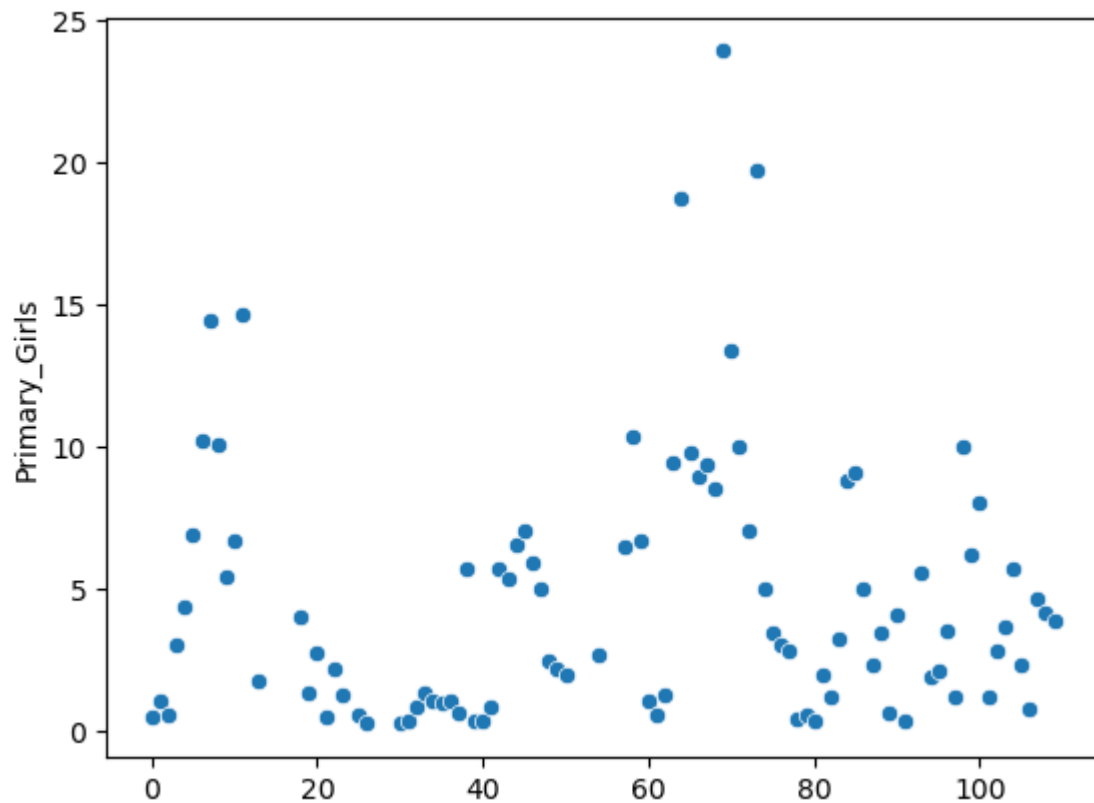
```
In [104]: # Filter data for the year 2014-15
dropout_2014_15 = dropout[dropout['year'] == '2014-15']
```

```
In [105]: # Create a scatter plot
plt.figure(figsize=(12, 8))
plt.scatter(enrollment_2014_15['Upper Primary_Boys'], dropout_2014_15['State_UT
plt.title('Highest dropout ratio of upper primary boys Across Different States
plt.xlabel('Highest dropout ratio of upper primary boys')
plt.ylabel('State')
plt.grid(True)
plt.tight_layout()
plt.show()
```



**Q19 Investigate the relationship of dropout ratio using scatter plots of primary girls.**

```
In [108]: sns.scatterplot(dropout['Primary_Girls'])
plt.show()
```



## Q20 Show Drop Out ratio of higher secondary girls across the state using graphs

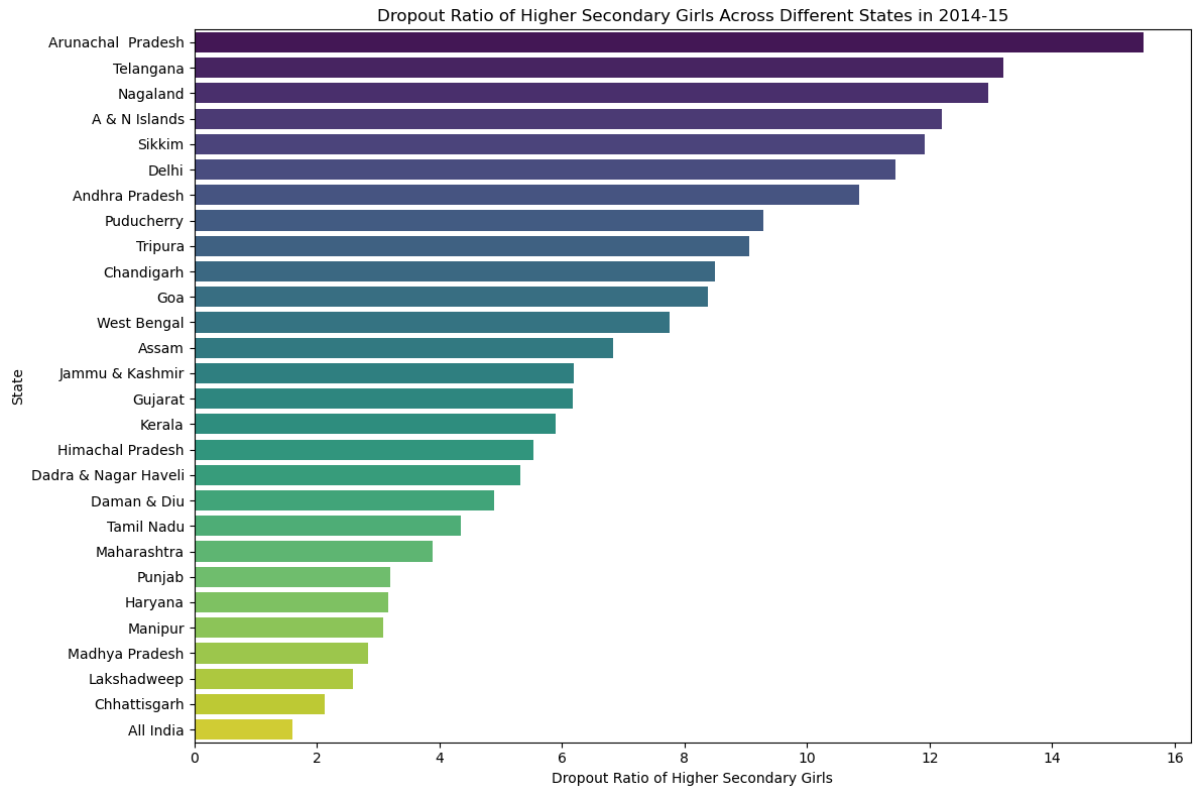
```
In [114]: dr_2013_14 = dropout[dropout['year'] == '2013-14']
```

```
In [115]: dr_upper_primary_boys = dr_2013_14[['State_UT', 'HrSecondary_Girls']]
```

```
In [116]: dr_upper_primary_boys.loc[:, 'HrSecondary_Girls'] = pd.to_numeric(dr_upper_prime
```

```
In [117]: dr_upper_primary_boys = dr_upper_primary_boys.dropna()
dr_upper_primary_boys_sorted = dr_upper_primary_boys.sort_values(by='HrSecondary
```

```
In [118]: plt.figure(figsize=(12, 8))
sns.barplot(x='HrSecondary_Girls', y='State_UT', data=dr_upper_primary_boys_sor
plt.title('Dropout Ratio of Higher Secondary Girls Across Different States in 2
plt.xlabel('Dropout Ratio of Higher Secondary Girls')
plt.ylabel('State')
plt.tight_layout()
plt.show()
```



In [ ]:

In [ ]:

In [ ]: