

# Dataset of various Colleges according to AISHE data

## Importing Libraries

```
In [1]: import numpy as np
import pandas as pd
import random
import matplotlib.pyplot as plt
```

## Data Preprocessing

```
In [2]: df1=pd.read_csv('out turn.csv', encoding='unicode_escape')
df2=pd.read_csv('teacher student ratio.csv', encoding='unicode_escape')
```

In [3]: df1.head()

Out[3]:

	t4	State Name	Univesity Name / Body Type	College/Stand Alone Institution	Appeared - Male	Appeared - Female	Passed - Male	Passed - Female	Passed with 60% or more - Male	Passed wit more
0	1	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Pondicherry University Campus II	0.0	0.0	22	26	0.0	
1	2	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Dr.B.R. Ambedkar Institute of Technology	124.0	58.0	67	28	46.0	
2	3	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Zoological Survey of India	5.0	0.0	0	0	0.0	
3	4	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Regional Medical Research Institute (I.C.M.R.)	4.0	1.0	4	1	4.0	
4	5	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Tagore Government College of Education	17.0	111.0	17	111	11.0	

```
In [4]: df2.head()
```

```
Out[4]:
```

	Sr. No.	State Name	University District	Univesity Name / Body Type	Institution District	College/Stand Alone Institution	Student Enrolment - Male	Student Enrolment - Female	Student Enrolment - Total	Teaching Staff (Excluding - Male)	Teaching Staff (Excluding - Female)	Teaching Staff (Excluding - Total)
0	1	Andaman and Nicobar Islands	Puducherry	College Institution	Nicobars	Zoological Survey of India	8.0	1.0	9.0	2	0	2
1	2	Andaman and Nicobar Islands	Puducherry	College Institution	South Andaman	Dr.B.R. Ambedkar Institute of Technology	1038.0	454.0	1492.0	38	25	63
2	3	Andaman and Nicobar Islands	Puducherry	College Institution	Nicobars	Regional Medical Research Institute (I.C.M.R.)	17.0	4.0	21.0	8	1	9
		Andaman										

### Merge the two datasets

```
In [5]: df=pd.merge(df1, df2, on="College/Stand Alone Institution", how="right")
```

### Removing rows with 0 values

```
In [6]: df = df.replace(0, np.nan)
df = df.dropna(how='any', axis=0)
```

```
In [7]: df.head()
```

```
Out[7]:
```

	t4	State Name_x	Univesity Name / Body Type_x	College/Stand Alone Institution	Appeared - Male	Appeared - Female	Passed - Male	Passed - Female	Passed with 60% or more - Male	Passed with 60% or more - Female	...	University District	Univesity Name / Body Type_y	Institution District
1	2	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Dr.B.R. Ambedkar Institute of Technology	124.0	58.0	67.0	28.0	46.0	14.0	...	Puducherry	College Institution	Andaman and Nicobar Islands
2	4	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Regional Medical Research Institute (I.C.M.R.)	4.0	1.0	4.0	1.0	4.0	1.0	...	Puducherry	College Institution	Nicobar Islands
4	5	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Tagore Government College of Education	17.0	111.0	17.0	111.0	11.0	110.0	...	Puducherry	College Institution	Nicobar Islands

### Drop unwanted columns

```
In [8]: df.drop(['University District','Univesity Name / Body Type_y','Institution District','Sr. No.','State Name_y','S
```

```
In [9]: df.shape
```

```
Out[9]: (16211, 9)
```

In [10]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16211 entries, 1 to 129622
Data columns (total 9 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   t4                                           16211 non-null  object
1   State Name_x                               16211 non-null  object
2   Univesity Name / Body Type_x              16211 non-null  object
3   College/Stand Alone Institution           16211 non-null  object
4   Appeared - Male                           16211 non-null  float64
5   Appeared - Female                         16211 non-null  float64
6   Passed - Male                             16211 non-null  float64
7   Passed - Female                           16211 non-null  float64
8   PTR EQUAL TO (9/12)                       16211 non-null  float64
dtypes: float64(5), object(4)
memory usage: 1013.2+ KB
```

In [11]: Students\_appeared = df['Appeared - Male'] + df['Appeared - Female']  
Students\_passed=df['Passed - Male'] + df['Passed - Female']

In [12]: df['Students\_appeared']=Students\_appeared  
df['Students\_passed']=Students\_passed

In [13]: df['Passing\_percentage']=(df['Students\_passed']/df['Students\_appeared'])\*100

In [14]: df.drop(['t4',"Passed - Male","Passed - Female",'Appeared - Male','Appeared - Female','Students\_appeared','Stude



In [15]: df=df.rename(columns={'PTR EQUAL TO (9/12)': 'S-T Ratio', 'State Name\_x': 'State Name', 'Univesity Name / Body Type\_



In [16]: df.shape

Out[16]: (16211, 5)

```
In [17]: x=list(range(1,16212))
```

```
In [18]: df.insert(loc=0, column='S.no', value=x)
```

```
In [19]: df.head()
```

Out[19]:

	S.no	State Name	Univesity Name	College/Stand Alone Institution	S-T Ratio	Passing_pe
1	1	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Dr.B.R. Ambedkar Institute of Technology	23.0	5
2	2	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Regional Medical Research Institute (I.C.M.R.)	2.0	10
4	3	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Tagore Government College of Education	55.0	10
768	4	Andhra Pradesh	Jawaharlal Nehru Technological University, Kak...	WEST GODAVARI INSTITUTE OF SCIENCE AND TECHNOL...	23.0	7
770	5	Andhra Pradesh	Acharya Nagarjuna University, Guntur	SIMS College of Education	16.0	7

```
In [20]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16211 entries, 1 to 129622
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   S.no                                  16211 non-null  int64
1   State Name                           16211 non-null  object
2   Univesity Name                       16211 non-null  object
3   College/Stand Alone Institution      16211 non-null  object
4   S-T Ratio                            16211 non-null  float64
5   Passing_percentage                   16211 non-null  float64
dtypes: float64(2), int64(1), object(3)
memory usage: 696.6+ KB
```

**Adding 3 new columns to df**

```
In [21]: rating=[]
infra=[]
place=[]
for i in range(16211):
    rating.append(round(random.uniform(2.5,4.8),1))
    infra.append(round(random.uniform(4.2,9.8),1))
    place.append(round(random.uniform(60,98),1))
```

```
In [22]: df['Rating']=rating
df['Infrastructure']=infra
df['Placement_Record']=place
```

```
In [23]: df.head()
```

Out[23]:

	S.no	State Name	Univesity Name	College/Stand Alone Institution	S-T Ratio	Passing_percentage	Rating	Infrastructure	Placemen
1	1	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Dr.B.R. Ambedkar Institute of Technology	23.0	52.197802	3.6	5.4	
2	2	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Regional Medical Research Institute (I.C.M.R.)	2.0	100.000000	4.1	6.9	
4	3	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Tagore Government College of Education	55.0	100.000000	3.6	8.1	
768	4	Andhra Pradesh	Jawaharlal Nehru Technological University, Kak...	WEST GODAVARI INSTITUTE OF SCIENCE AND TECHNOL...	23.0	77.591973	3.0	5.1	
770	5	Andhra Pradesh	Acharya Nagarjuna University, Guntur	SIMS College of Education	16.0	70.103093	3.5	4.8	

**Corelation between various columns**

In [24]: df.corr()

Out[24]:

	S.no	S-T Ratio	Passing_percentage	Rating	Infrastructure	Placement_Record
S.no	1.000000	0.110598	0.090639	-0.010615	-0.005933	-0.002262
S-T Ratio	0.110598	1.000000	0.011818	-0.002270	-0.002890	-0.011493
Passing_percentage	0.090639	0.011818	1.000000	0.010896	-0.001179	0.002508
Rating	-0.010615	-0.002270	0.010896	1.000000	0.003659	-0.004464
Infrastructure	-0.005933	-0.002890	-0.001179	0.003659	1.000000	0.009602
Placement_Record	-0.002262	-0.011493	0.002508	-0.004464	0.009602	1.000000

## Final Dataset

In [25]: df.head()

Out[25]:

	S.no	State Name	Univesity Name	College/Stand Alone Institution	S-T Ratio	Passing_percentage	Rating	Infrastructure	Placemen
1	1	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Dr.B.R. Ambedkar Institute of Technology	23.0	52.197802	3.6	5.4	
2	2	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Regional Medical Research Institute (I.C.M.R.)	2.0	100.000000	4.1	6.9	
4	3	Andaman and Nicobar Islands	Pondicherry Univeristy, Puducherry	Tagore Government College of Education	55.0	100.000000	3.6	8.1	
768	4	Andhra Pradesh	Jawaharlal Nehru Technological University, Kak...	WEST GODAVARI INSTITUTE OF SCIENCE AND TECHNOLOG...	23.0	77.591973	3.0	5.1	
770	5	Andhra Pradesh	Acharya Nagarjuna University, Guntur	SIMS College of Education	16.0	70.103093	3.5	4.8	



```
In [26]: df.describe()
```

```
Out[26]:
```

	S.no	S-T Ratio	Passing_percentage	Rating	Infrastructure	Placement_Record
count	16211.000000	16211.000000	16211.000000	16211.000000	16211.000000	16211.000000
mean	8106.000000	39.008513	80.435267	3.644902	6.994769	78.986972
std	4679.856942	58.793577	20.235223	0.665392	1.617474	10.962946
min	1.000000	1.000000	0.274876	2.500000	4.200000	60.000000
25%	4053.500000	14.000000	70.765971	3.100000	5.600000	69.500000
50%	8106.000000	21.000000	85.650224	3.600000	7.000000	79.100000
75%	12158.500000	41.000000	95.483581	4.200000	8.400000	88.500000
max	16211.000000	3300.000000	896.000000	4.800000	9.800000	98.000000

## Top 5 institutes according to various parameters

```
In [27]: ['Placement_Record', 'Infrastructure', 'Rating', 'Passing_percentage'],axis=0,ascending=[False,False,False,False],inp
```

◀

```
In [28]: out=result.head(5)
```

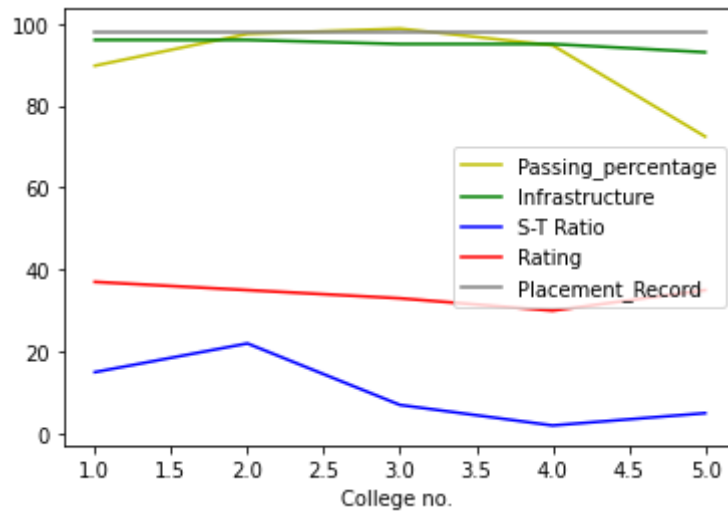
```
In [29]: out[['College/Stand Alone Institution', 'Univesity Name', 'State Name']][:10]
```

```
Out[29]:
```

	College/Stand Alone Institution	Univesity Name	State Name
0	130021-RSSPM'S NANASAHEB YASHAVANTRAO N.CHAVAN...	North Maharashtra University, Jalgaon	Maharashtra
1	Gandhi Shatabdi Smarak Snatkottar Mahavidyalay...	Veer Bahadur Singh Purvanchal University, Jaunpur	Uttar Pradesh
2	jagdish Prasad pandey Smarak Degree College, D...	Deen Dayal Upadhyay Gorakhpur University, Gora...	Uttar Pradesh
3	KSS College of Education and PG centre, Davangere	Davangere University, Davangere	Karnataka
4	Govt. College, Banjar, Distt Kullu	Himachal Pradesh University , Shimla	Himachal Pradesh

```
In [30]: x=[1,2,3,4,5]
plt.plot(x, out['Passing_percentage'], color='y', label='Passing_percentage')
plt.plot(x, out['Infrastructure']*10, color='g', label='Infrastructure')
plt.plot(x, out['S-T Ratio']//10, color='b', label='S-T Ratio')
plt.plot(x, out['Rating']*10, color='r', label='Rating')
plt.plot(x, out['Placement_Record'], color='grey', label='Placement_Record')

plt.xlabel('College no. ')
plt.legend()
plt.show()
```



```
In [31]: x=[1,2,3,4,5]
plt.plot(x, out['Passing_percentage'], color='b', label='Passing_percentage')
plt.plot(x, out['S-T Ratio'], color='g', label='S-T Ratio')
plt.xlabel('College no. ')
plt.legend()
plt.show()
```

