Project Title :

Detailed Analysis of Nirf Ranking(2016-2021) of Different Institute's Across India.

Creator :-
Abhishek kumar,13547
Ambedkar,13545
Chestha Sharma,13502
Shubham kumar,13525

Inspiration -

I was always curious about how institutes have performed over the year, which institutes got improved or which institutes got worsened, how many of them were able to retain their top position ? And the nirf ranking seemed to offer reliable and comprehensive data on institutes since 2016.

Description of Dataset :

Content

The dataset contains available Rankings and Different Scores of the Institutes over 2016 to 2021. The dataset contains folders for each year containing the scores of the institutes and ranking for that year. *The files outside the folder contains the Combined data of institute from 2016 to 2021 which can help in analyzing the performance of the institutes over the years.*

Parameters for Ranking :-

A) Teaching, Learning & Resources (TLR)

1.Faculty-Student Ratio with Emphasis on Permanent Faculty (FSR) 2.Combine Metric for Faculty with PhD and Experience (FQE) 3.Metric for Library and Laboratory Facilities (LL) 4.Metric for Sports and Extra-Curricular Facilities, Activities (SEC)

B) Research and Professional Practice (RP)

1.Combined Metric for Publications (PU) 2.Combined Metric for Citations (CI) 3.IPR and Patents: Granted, Filed, Licensed (IPR) 4.Percentage of Collaborative Publications and Patents (CP) 5.Footprints of Projects and Professional Practice (FPPP)

C) Graduation Outcome (GO)

1.Combined Performance in Public and University Examinations (PUE) 2.Combined Percentage for Placement, Higher Studies and Entrepreneurship (PHE) 3.Mean Salary for Employment (MS)

D) Outreach and Inclusivity (OI)

1.Outreach Footprint (Continuing Education, Service) (CES) 2.Percentage of Students from Other States/Countries(Region Diversity – RD) 3.Percentage of Women Students and Faculty (WS) 4.Percentage of Economically and Socially 5.Disadvantaged Students (ESDS)

E) Peer Perception (PR)

Process for Peer Rating in Category (PR)Scrutinizing by these two parameters itself, things don't seem to be falling in a rigorous order. Interestingly, the parameter on perception at number (f) is somehow interspersed between those of Publication Details and IPR Summary. This is a significant departure from the established practice of rigorous data processing whereby the consistency of the framework has to be ensured by all means.
Note:-
Incidentally, Datasets for some of the parameters are missing altogether.

Analysis of Dataset

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
pd.set_option('mode.chained_assignment',None)
pd.options.plotting.backend = "plotly"
base_dir = "C:/Users/Shubham's/Desktop/NirfRanking_2016-2021/"
df = pd.read_csv(f'{base_dir}/OverallRanking.csv')
df.head(2)
```

Out[2]:

	Institute Id	Institute Name	City	State	Score_21	Rank_21	TLR_21	RPC_21	GO_21	OI
0	IR-O-U-0456	Indian Institute of Technology Madras	Chennai	Tamil Nadu	86.76	1.0	85.61	92.51	86.32	63
1	IR-O-U-0220	Indian Institute of Science	Bengaluru	Karnataka	82.67	2.0	79.13	91.48	78.23	58

2 rows × 39 columns

Dataset Summary :

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 39 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Institute Id    149 non-null    object 
 1   Institute Name  149 non-null    object 
 2   City            149 non-null    object 
 3   State           149 non-null    object 
 4   Score_21        100 non-null    float64
 5   Rank_21         100 non-null    float64
 6   TLR_21          100 non-null    float64
 7   RPC_21          100 non-null    float64
 8   GO_21           100 non-null    float64
 9   OI_21           100 non-null    float64
 10  Perception_21   100 non-null    float64
 11  Score_20        101 non-null    float64
 12  Rank_20         101 non-null    float64
 13  TLR_20          101 non-null    float64
```

Data Pre-processing:

Now, For Data Analysis and Visualization of a dataset.There are many some asymmetry present in the Dataset which further,give us problem during the time of Analysis. So,For Escape this we remove this redundancy and kind of Asymmetry by data-pre-processing.

Firstly, we replace null cell or missing data in dataset with NaN value.

In [4]:

```
df = df.replace(r'^\s*$', np.NaN, regex=True)
print("After Replacing, There are ",len(df.isnull()==True),"NaN cells Values in the dataset
```

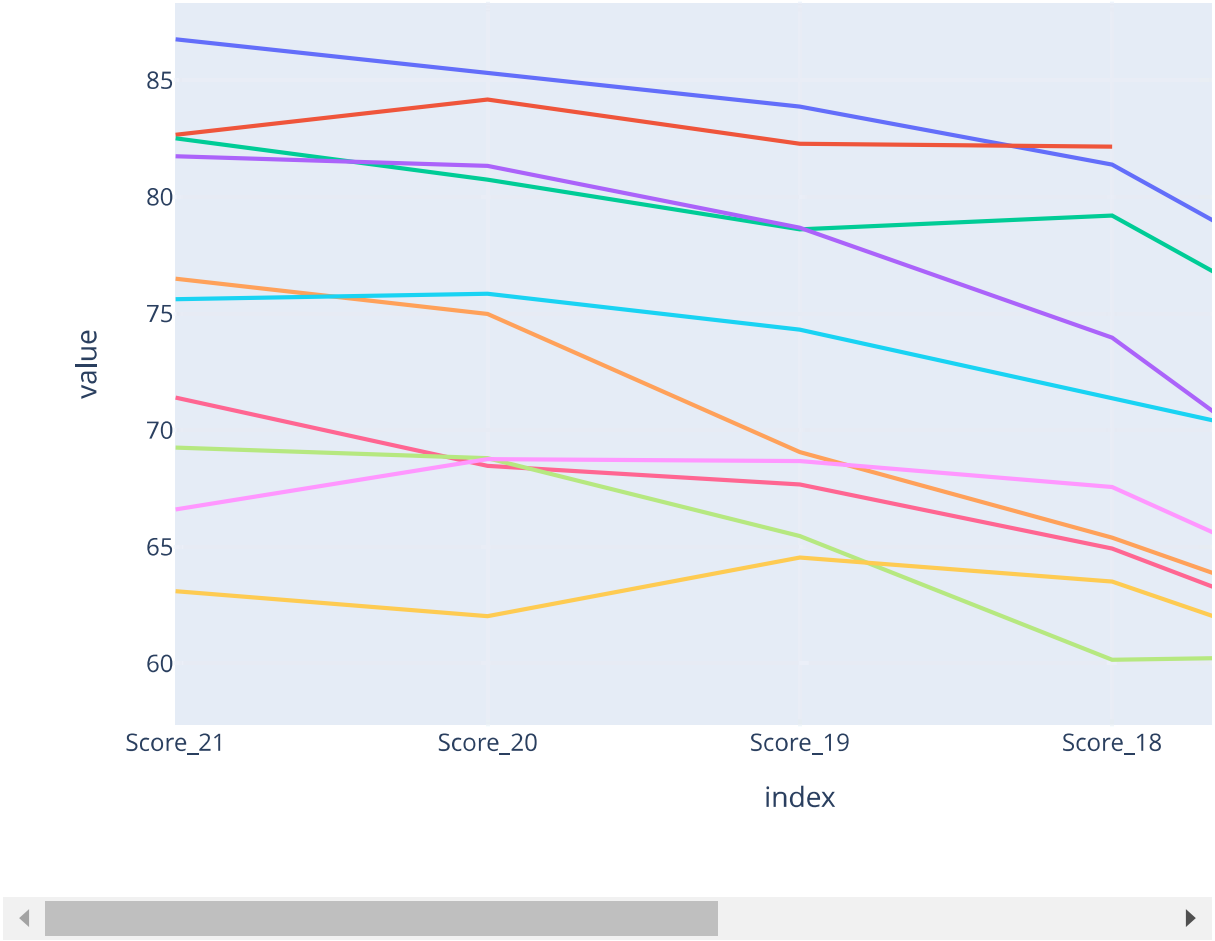
```
After Replacing, There are  149 NaN cells Values in the dataset.
```

Data Analysics & Visualization on Dataset

Query1: what is the performance of Top 10 institute's during 2016-2021?

In [5]:

```
df = pd.read_csv(f'{base_dir}/OverallRanking.csv')
df_scores = df[['Institute Name', 'Score_21', 'Score_20', 'Score_19', 'Score_18', 'Score_17
fig = df_scores.head(10).set_index('Institute Name').T.plot()
fig.update_layout(width=1000, height=500)
fig.show()
```



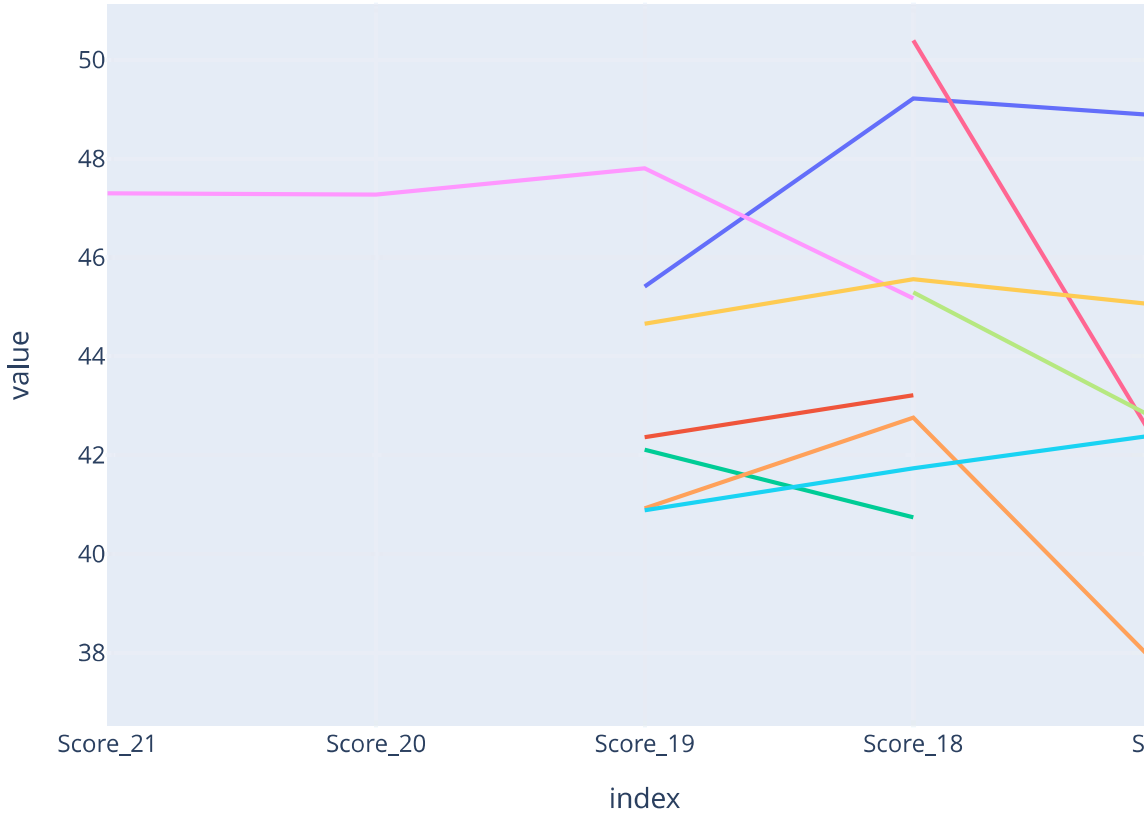
conclusion:

1. These top institutes are constantly working on there weak area and improve there rank continously.

Query2: what is the performance of Bottom 10 institute's during 2016-2021?

In [6]:

```
df_scores.loc[:,'Nan'] = df_scores.isnull().sum(axis=1)
df_scores = df_scores[df_scores['Nan'] < 4]
df_scores.reset_index(inplace=True, drop=True)
df_scores = df_scores.drop(columns='Nan', axis=1)
fig = df_scores.tail(10).set_index('Institute Name').T.plot()
fig.update_layout(width=1000, height=500)
fig.show()
```



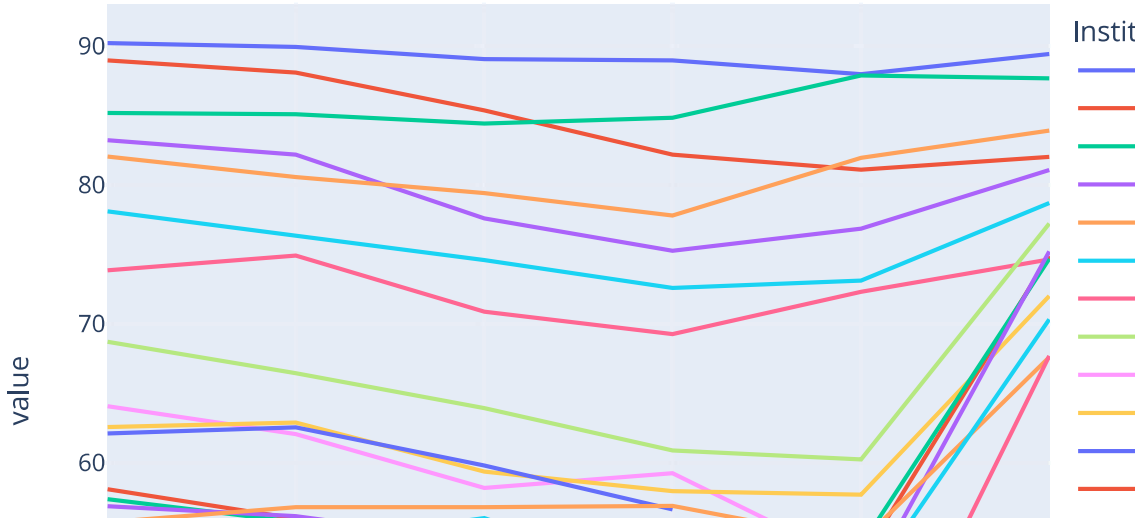
Conclusion :

Most of the Collages are not participate after 2019 because of there bad previous Score.

Query3: what is the performance of IIT's from 2016-2021 ?

In [7]:

```
df = pd.read_csv(f'{base_dir}/EngineeringRanking.csv')
df_scores = df[['Institute Name', 'Score_21', 'Score_20', 'Score_19', 'Score_18', 'Score_17
df_scores.loc[:,'Nan'] = df_scores.isnull().sum(axis=1)
df_scores = df_scores[df_scores['Nan'] < 4]
df_scores.reset_index(inplace=True, drop=True)
df_scores = df_scores.drop(columns='Nan', axis=1)
sp_df = df_scores[df_scores['Institute Name'].str.contains('Indian Institute of Technology'
fig = sp_df.set_index('Institute Name').T.plot()
# plt.suptitle("IIT's Performance (2016-2021)")
fig.show()
```



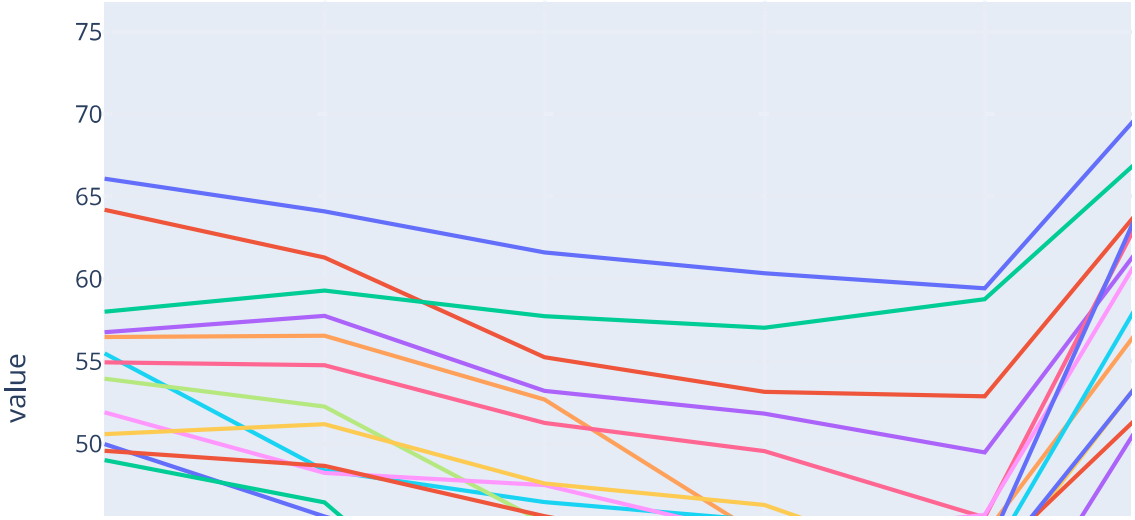
Conclusion :

1. There is sudden fallen of performance is coming in 2017.
2. But some Top IIT's Maintain there performance in 2017 like :- IIT Bombay , IIT madras etc.
3. After 2017,There are most of the IIT's are come back and there performance Increase.

Query4: what is the performance of NIT's from 2016-2021 ?

In [8]:

```
sp_df = df_scores[df_scores['Institute Name'].str.contains('National Institute of Technolog
fig = sp_df.set_index('Institute Name').T.plot()
# plt.suptitle("NIT's Performance (2016-2021)")
fig.show()
```



Conclusion :

1. There is sudden fallen of performance is coming in 2017.
2. After 2017,There are most of the IIT's are come back and there Nirf Score Increase.
3. some New NIT's are also added in between a year of 2018-2019

Query5: Changes in Educational Growth of Different States on basis of there Institue's NIRF Score Rank improves or vice versa ?

In [9]:

```
df = pd.read_csv(f'{base_dir}/OverallRanking.csv')
df.loc[:,'Nan'] = df.isnull().sum(axis=1)
df = df[df['Nan'] < 3]
df.reset_index(inplace=True, drop=True)
df = df.drop(columns='Nan', axis=1)
df_Score = pd.DataFrame()
df_Score['2017'] = df['Score_17'].groupby(df['State']).sum()
df_Score['2018'] = df['Score_18'].groupby(df['State']).sum()
df_Score['2019'] = df['Score_19'].groupby(df['State']).sum()
df_Score['2020'] = df['Score_20'].groupby(df['State']).sum()
df_Score['2021'] = df['Score_21'].groupby(df['State']).sum()
df_Score.head()
```

Out[9]:

State	2017	2018	2019	2020	2021
Andhra Pradesh	41.48	43.44	44.88	46.14	42.78
Assam	148.57	153.60	159.51	161.12	160.35
Bihar	39.87	44.25	46.28	48.09	47.67
Chandigarh	43.13	50.94	51.25	50.24	50.31
Delhi	277.67	306.72	314.76	319.93	310.39

Query6: What is the Mean/Average values of TLR, RPC, GO, OI, Perception for each year?

In [12]:

```
data = pd.read_csv(f'{base_dir}/Final_Dataset.csv')
data['year']=pd.DatetimeIndex(data['Accreditation valid up to']).year
data_all_year=data[['year','TLR','RPC','GO','OI','Perception']]
data_all_parameter=data_all_year.groupby('year').mean()
data_all_parameter.plot(figsize=(20,8),fontsize=20,linewidth=10)
```

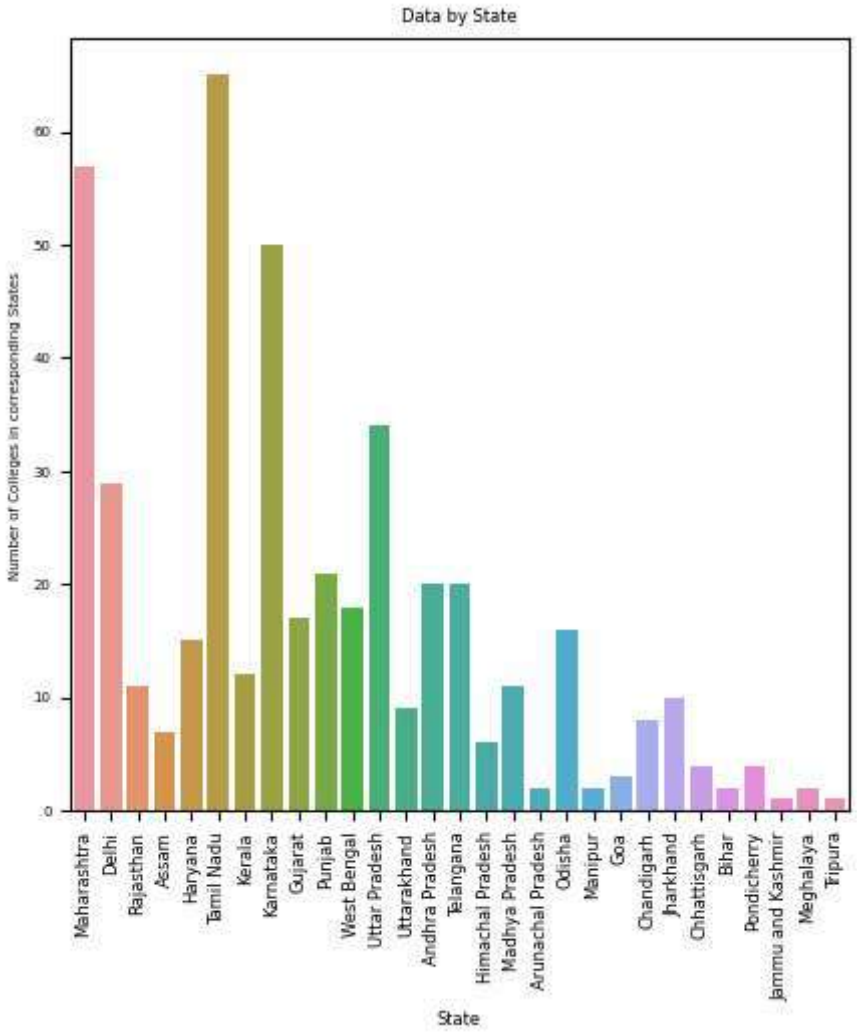
...

Conclusion : it basically gives an an broad idea about the indian Education System.

Query7: What is the distribution of Collages between the states Across India ?

In [83]:

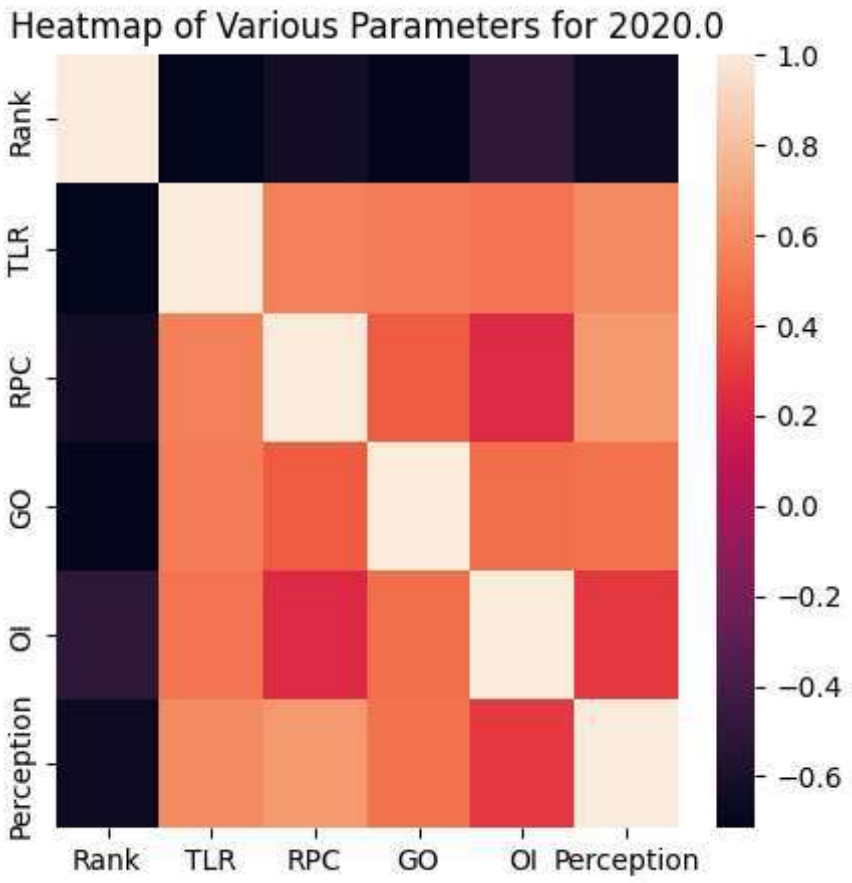
```
plt.figure(figsize=(5,5))
ax=sns.countplot(data['State'])
plt.title(f'Data by State',fontsize=6)
plt.xticks(rotation='vertical',fontsize=6)
plt.yticks(fontsize=5)
plt.xlabel('State',fontsize=6)
plt.ylabel('Number of Colleges in corresponding States',fontsize=5)
plt.show()
```

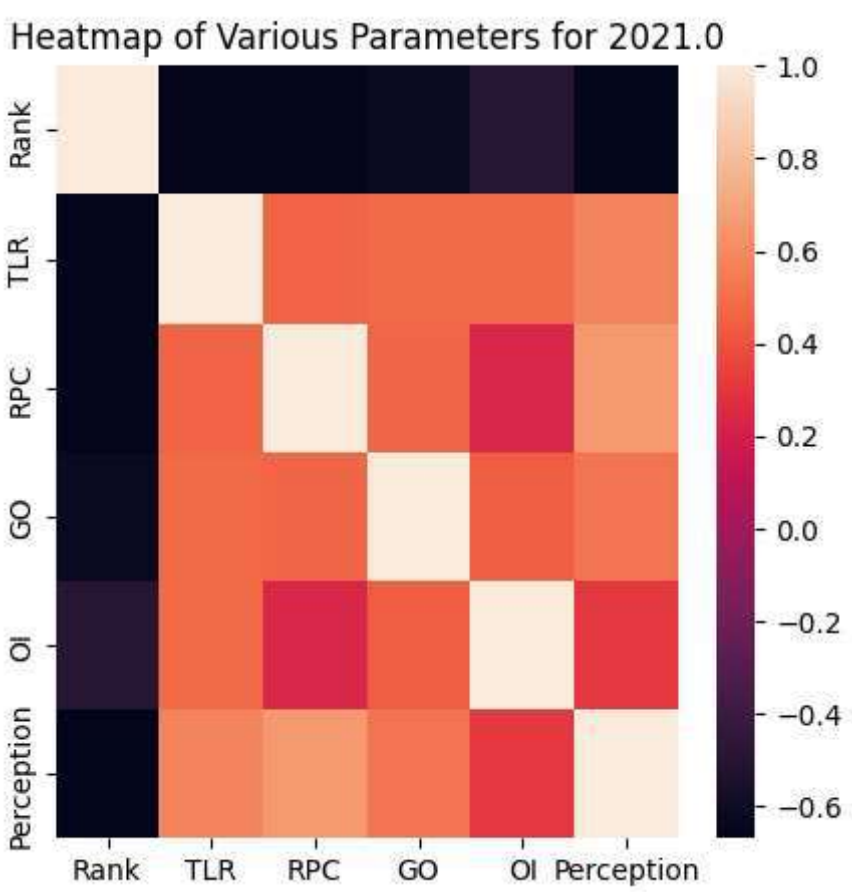


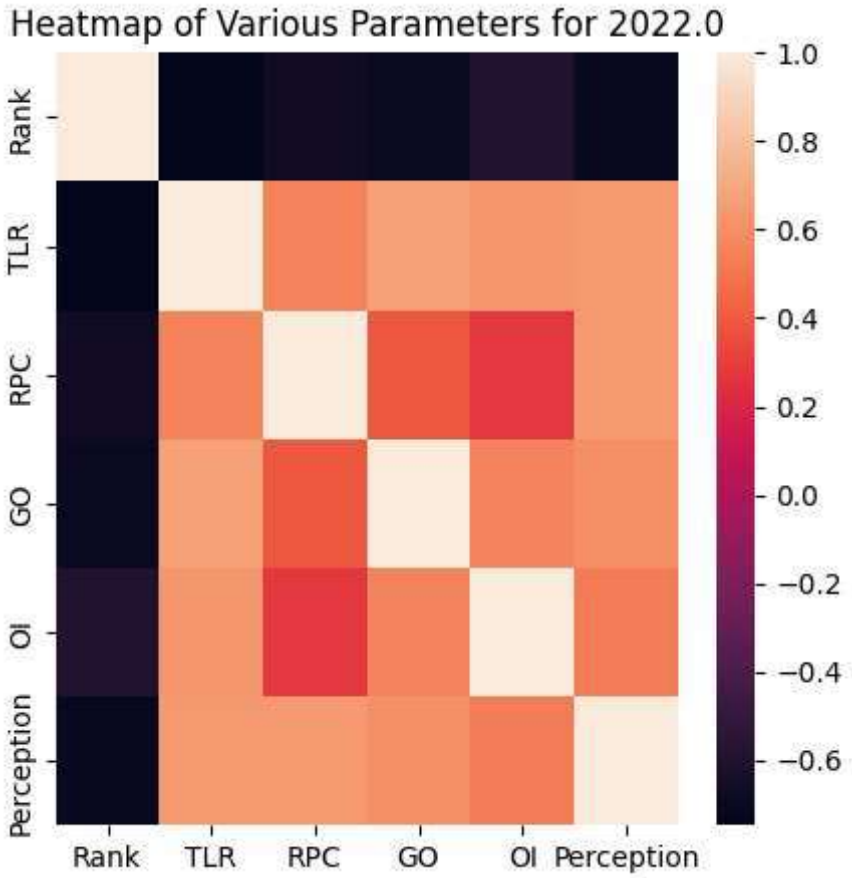
Query8: what is the various Co-relation between the various factors of the ranking ?

In [72]:

```
plt.rcParams['figure.figsize'] = [5,5]
k=['Rank','TLR','RPC','GO','OI','Perception']
for i,j in data.groupby('year'):
    plt.title(f"Heatmap of Various Parameters for {i}")
    sns.heatmap(j[k].corr())
    plt.show()
```




Heatmap of Various Parameters for 2021.0
Rank
TLR
RPC
GO
OI
Perception
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

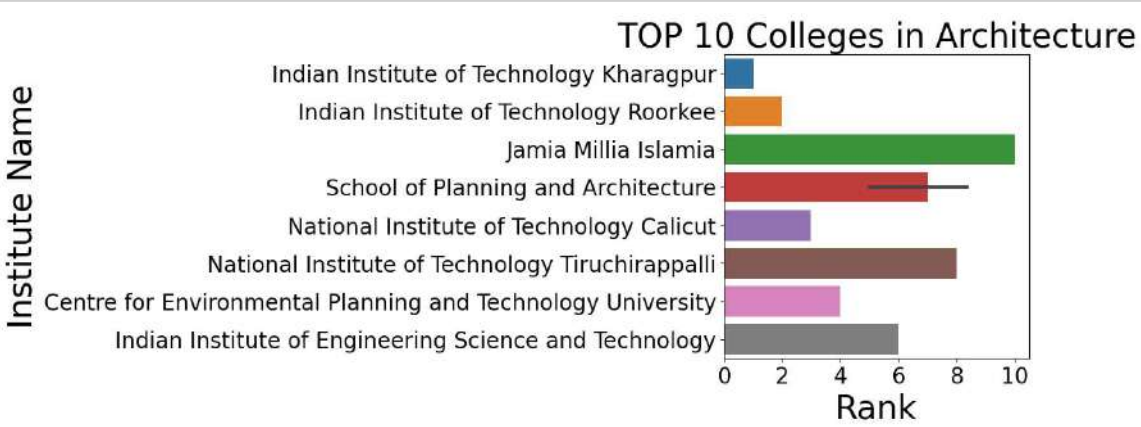

Heatmap of Various Parameters for 2022.0
Rank
TLR
RPC
GO
OI
Perception
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

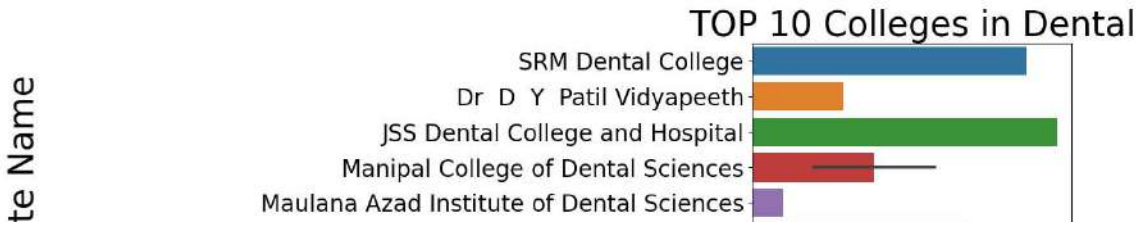
Conclusion: This Shows that how different factors are varying and how to corelated with each other.

Query9: what is the Top Colleges in each domain of study on the basis of rank ?

In [76]:

```
top10Colleges = data.query("Rank <= 10")
for i,j in top10Colleges.groupby('Domain'):
    sns.barplot(j['Rank'],j['Name'],label=f"{i}")
    plt.title(f"TOP 10 Colleges in {i}",fontsize=30)
    plt.ylabel('Institute Name',fontsize=30)
    plt.xlabel("Rank",fontsize=30)
    plt.xticks(fontsize = 20)
    plt.yticks(fontsize = 20)
    plt.show()
```





Query10: What is Top 10 Public and Private Universities Other than IIT's and NIT's ?

In [126]:

```
df = pd.read_csv(f'{base_dir}/UniversityRanking.csv')
df.head()
```

Out[126]:

	Institute Id	Institute Name	City	State	Score_21	Rank_21	TLR_21	RPC_21	GO_21
0	IR-O-U-0220	Indian Institute of Science	Bengaluru	Karnataka	82.67	1.0	79.13	91.48	78.23
1	IR-O-U-0109	Jawaharlal Nehru University	New Delhi	Delhi	67.99	2.0	71.19	44.96	95.07
2	IR-O-U-0500	Banaras Hindu University	Varanasi	Uttar Pradesh	64.02	3.0	64.50	45.00	100.00
3	IR-O-U-0570	Calcutta University	Kolkata	West Bengal	62.06	4.0	66.20	43.92	91.72
4	IR-O-U-0436	Amrita Vishwa Vidyapeetham	Coimbatore	Tamil Nadu	61.23	5.0	64.23	54.33	71.35

5 rows × 46 columns

Conclusion: there are many University in India whose Score Very well and compete Iit's and Nit's.

Query11: Graduation Outcomes on the basis of different Domain ?

In [38]:

```
GO_Domain = data.groupby('Domain')['GO'].sum()
GO_Domain
```

Out[38]:

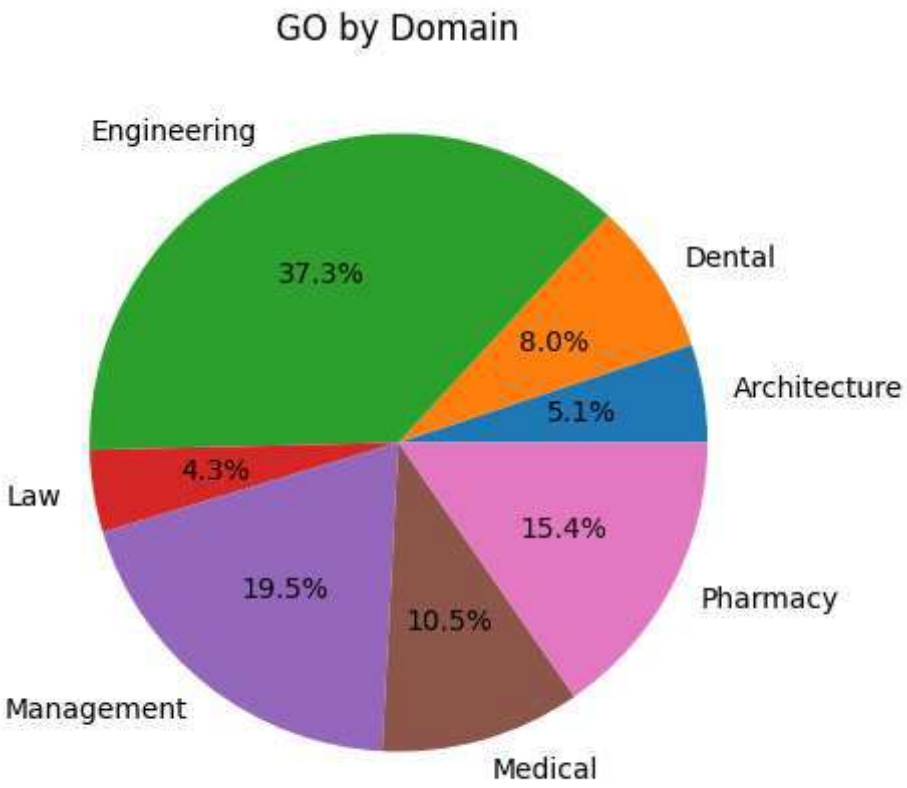
```
Domain
Architecture     1559.63
Dental           2436.89
Engineering     11340.72
Law              1310.91
Management       5928.82
Medical          3192.65
Pharmacy         4674.33
Name: GO, dtype: float64
```

In [39]:

```
GO_Domain.plot.pie(autopct = '%1.1f%%')
plt.title('GO by Domain')
plt.ylabel('')
```

Out[39]:

```
Text(0, 0.5, '')
```



Conclusion :So there are maximum colleges in Engineering domain are having highest Share on GO and minimum is Law and Architecture.

Query12: Analysis of participation of colleges in ranking process

In [24]:

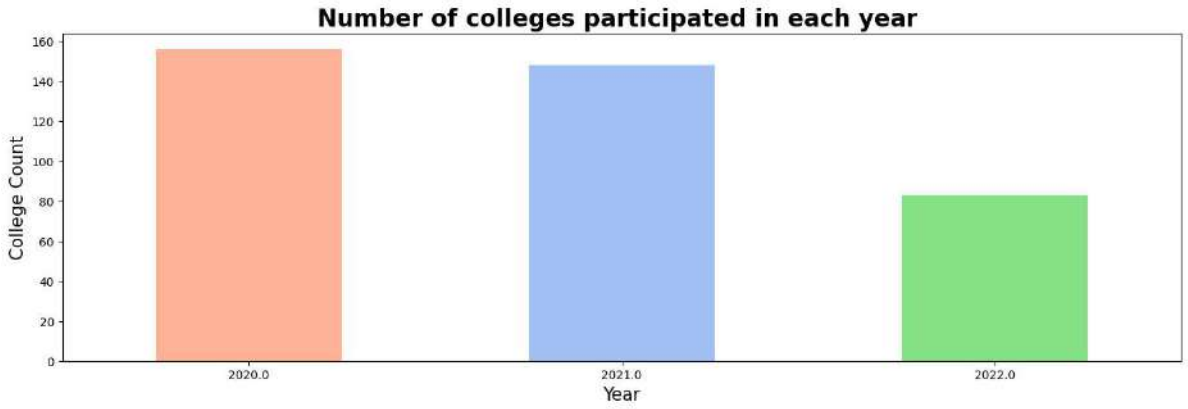
```
# count of Colleges participated in each year
count_college= data.groupby('year')['Institute_ID'].nunique()
print(count_college)

# Bar plot for the No. of Colleges participated in each Season
count_college.plot.bar(color=[ 'coral', 'cornflowerblue', 'limegreen'], figsize=(17,5), alp
plt.title("Number of colleges participated in each year ",fontsize=20,fontweight="bold")
plt.ylabel("College Count", size = 15)
plt.xlabel("Year", size = 15)
plt.xticks(size = 10, rotation='horizontal')
plt.yticks(size = 10)
```

```
year
2020.0    156
2021.0    148
2022.0     83
Name: Institute_ID, dtype: int64
```

Out[24]:

```
(array([  0.,  20.,  40.,  60.,  80., 100., 120., 140., 160., 180.]),
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```

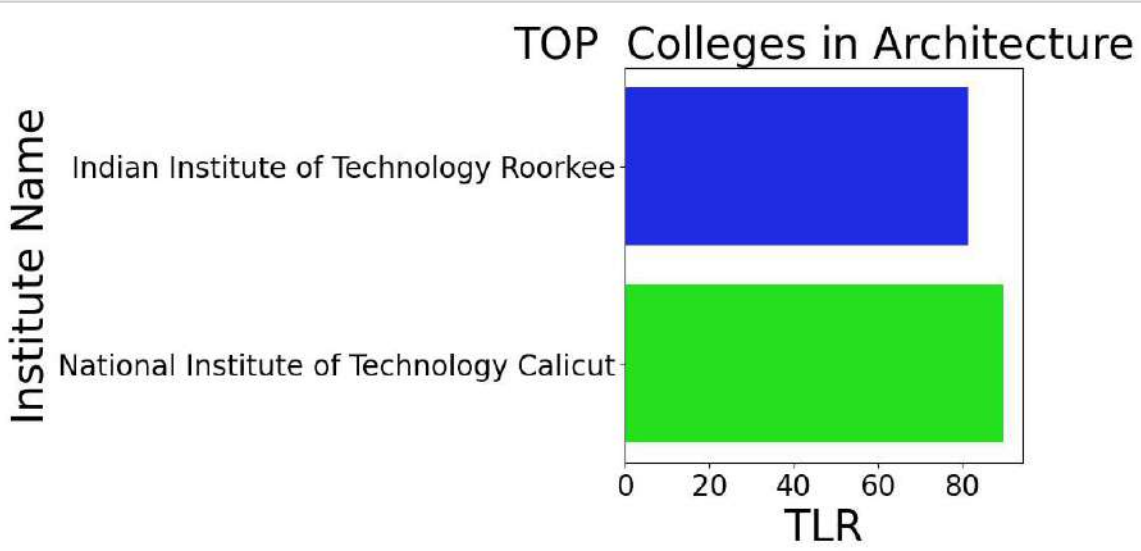


Conclusion : So the maximum colleges participated in year 2020.

Query13 : Top Colleges in each domain of study on the basis of Teacher and Learning Resource

In [130]:

```
# data=pd.read_csv("C:/Users/HP/Downloads/Final_Dataset.csv")
data = pd.read_csv(f'{base_dir}/Final_Dataset.csv')
top10data = data.query("TLR >= 80")
for i,j in top10data.groupby('Domain'):
    sns.barplot(j['TLR'],j['Name'],label=f"{i}",palette='hsv_r')
    plt.title(f"TOP  Colleges in {i}",fontsize=30)
    plt.ylabel('Institute Name',fontsize=30)
    plt.xlabel("TLR",fontsize=30)
    plt.xticks(fontsize = 20)
    plt.yticks(fontsize = 20)
    plt.show()
```



Query14 : Colleges whoose perception is zero

In [34]:

```
zero_perc=data.loc[data['Perception']==0]
zero_perc[['Institute_ID','Name']]
```

Out[34]:

	Institute_ID	Name
7	IR-E-U-0332	The Rashtrasant Tukadoji Maharaj Nagpur Univer...
35	IR-E-C-19650	Vardhaman College of Engineering
50	IR-E-C-18254	Yeshwantrao Chavan College of Engineering
68	IR-E-C-1345	P E S College of Engineering
117	IR-M-U-0379	Lovely Professional University
189	IR-P-U-0190	Shoolini University of Biotechnology and Manag...
229	IR-E-U-0604	Amity University Gwalior
237	IR-E-C-34167	Ramrao Adik Institute of Technology
241	IR-E-C-16537	St Josephs College of Engineering
261	IR-E-U-0163	The Northcap University
290	IR-M-U-0373	Chitkara University
304	IR-N-U-0461	M G R Educational and Research Institute
310	IR-E-C-1398	BNM Institute of Technology
319	IR-P-U-0491	Vels Institute of Science Technology Advanc...
325	IR-E-U-0190	Shoolini University of Biotechnology and Manag...
366	IR-E-U-0373	Chitkara University
386	IR-P-U-0085	Guru Ghasidas Vishwavidyalaya
389	IR-N-N-116	Pacific Dental College
406	IR-E-U-0535	Rajiv Gandhi Institute of Petroleum Technology
422	IR-E-C-26929	Sree Vidyanikethan Engineering College

Conclusion :These colleges has 0 extra curricular activities

Query15 : Comparing percentage of colleges in different zones

In [35]:

```
north=[ 'Himachal Pradesh', 'Punjab', 'Uttarakhand' ,'Delhi', 'Uttar Pradesh' , 'Haryana']
east=[ 'Bihar', 'Orissa', 'Jharkhand',  'West Bengal']
west= ['Rajasthan' , 'Gujarat', 'Goa' , 'Maharashtra']
south=['Andhra Pradesh', 'Karnataka', 'Kerala' , 'Tamil Nadu']
central=['Madhya Pradesh', 'Chattisgarh']
northeast=['Assam', 'Sikkim', 'Nagaland', 'Meghalaya', 'Manipur', 'Mizoram', 'Tripura' , 'A
for i in data['State']:
    print(i,"in for")
    if  i in north:
        print(i,data['State'])
        data['zone']='north'
    elif i in east:
        data['zone']='east'
    elif i in west:
        data['zone']='west'
    elif i in south:
        data['zone']='south'
    else:
        data['zone']='undetermined'
# elif data['State'] in central:
#     data['zone']='central'
# elif data['State'] in northeast:
#     data['zone']='northeast'
data[['zone','State']]
```

```
Maharashtra in for
Maharashtra in for
Maharashtra in for
Delhi in for
Delhi 0         Maharashtra
1         Maharashtra
2         Maharashtra
3               Delhi
4         Maharashtra
            ...
455     Uttar Pradesh
456    Andhra Pradesh
457     Uttar Pradesh
458         Karnataka
459               Goa
Name: State, Length: 460, dtype: object
Maharashtra in for
Rajasthan in for
Assam in for
```

Query-16 Top states with most colleges in a domain

In [36]:

```
data=pd.read_csv(f'{base_dir}/Final_Dataset.csv')
data['year']=pd.DatetimeIndex(data['Accreditation valid up to']).year
Eng_Clg = data[data['Domain']=='Engineering']
Mng_Clg = data[data['Domain']=='Management']
```

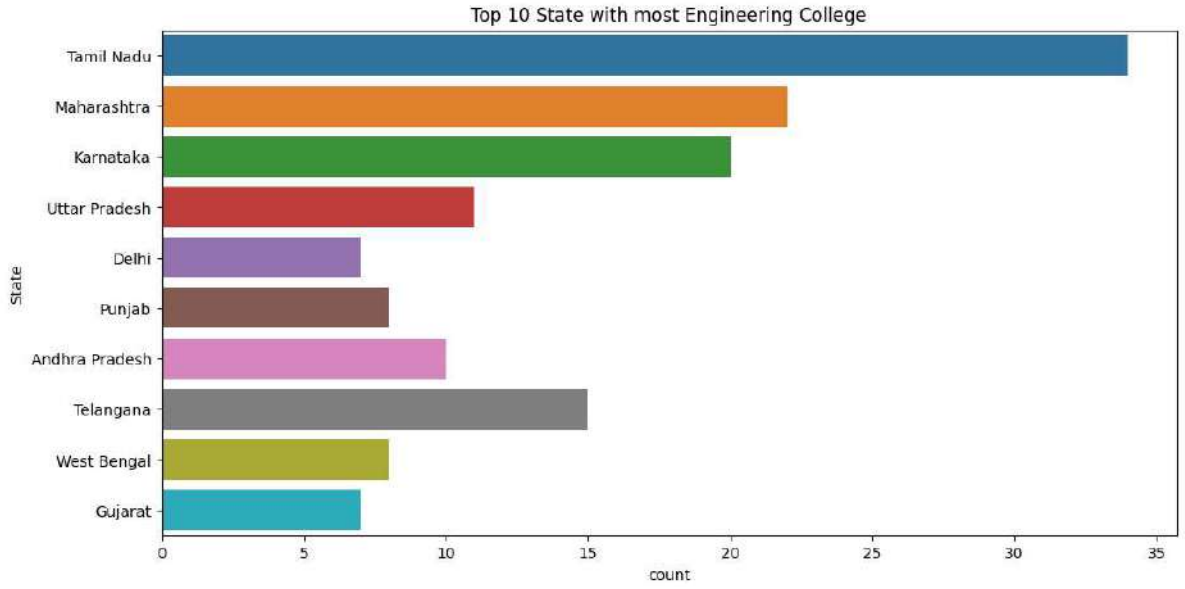
In [37]:

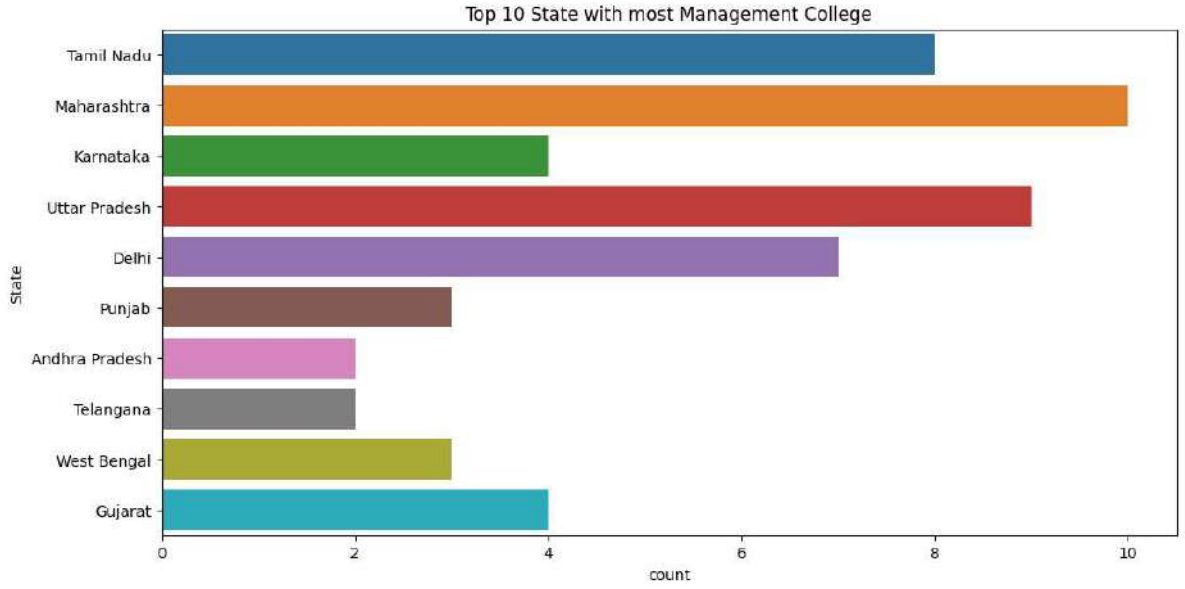
```
plt.figure(figsize=(12,6))
sns.countplot(y='State',order = data['State'].value_counts().index[0:10],data=Eng_Clg)
plt.title('Top 10 State with most Engineering College')

plt.figure(figsize=(12,6))
sns.countplot(y='State',order = data['State'].value_counts().index[0:10],data=Mng_Clg)
plt.title('Top 10 State with most Management College')
```

Out[37]:

```
Text(0.5, 1.0, 'Top 10 State with most Management College')
```





Conclusion:

Tamil Nadu is the state with highest no. of top ranked Engineering and Management colleges.and Top 10 states for both remains the same.

Query -17: Colleges appearing max times in ranking

In [38]:

```
data.Name.value_counts().head(10)
```

Out[38]:

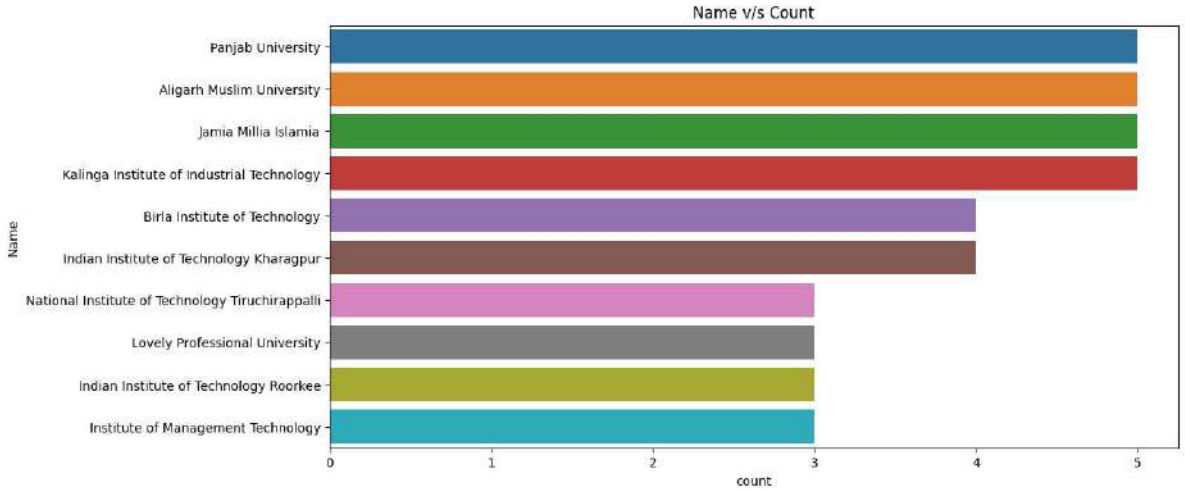
```
Panjab University                                5
Aligarh Muslim University                        5
Jamia Millia Islamia                             5
Kalinga Institute of Industrial Technology       5
Birla Institute of Technology                    4
Indian Institute of Technology Kharagpur         4
National Institute of Technology Tiruchirappalli 3
Lovely Professional University                   3
Indian Institute of Technology Roorkee           3
Institute of Management Technology               3
Name: Name, dtype: int64
```

In [39]:

```
plt.figure(figsize=(12,6))
sns.countplot(y='Name',order = data['Name'].value_counts().index[0:10],data=data)
plt.title('Name v/s Count')
```

Out[39]:

```
Text(0.5, 1.0, 'Name v/s Count')
```



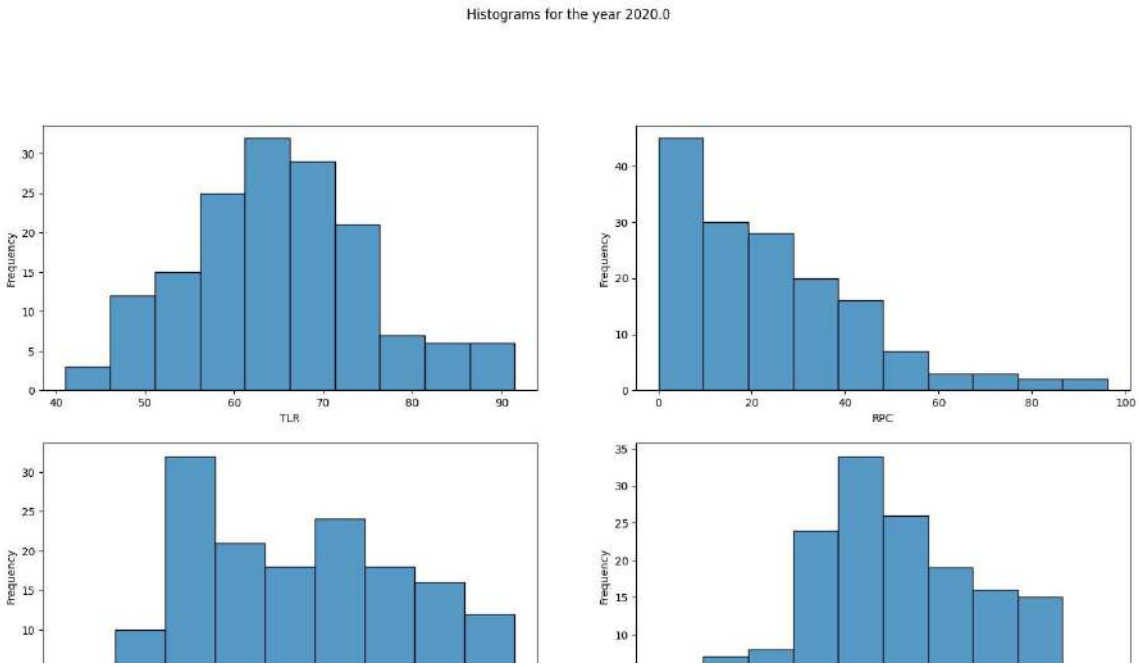
Conclusion:

1.Panjab University,AMU,JMI,KIIT appered most times i.e. 5 in ranking .This means they were among top ranked in different domains 5 times.

Query-18 : Distribution of various parameters over the years

In [40]:

```
plt.rcParams['figure.figsize'] = [18,15]
for i,j in data.groupby('year'):
  cols=['Rank','TLR','RPC','GO','OI','Perception']
  stat_data = j[cols]
  cols =['TLR','RPC','GO','OI','Perception']
  plt.figure()
  plt.suptitle(f"Histograms for the year {i}")
  for k in range(len(cols)):
    plt.subplot(3,2,k+1)
    sns.histplot(stat_data[cols[k]], bins='auto')
    plt.xlabel(f"{cols[k]}")
    plt.ylabel(f"Frequency")
  plt.show()
```



Conclusion : This Shows that how different factors are varying over the years.

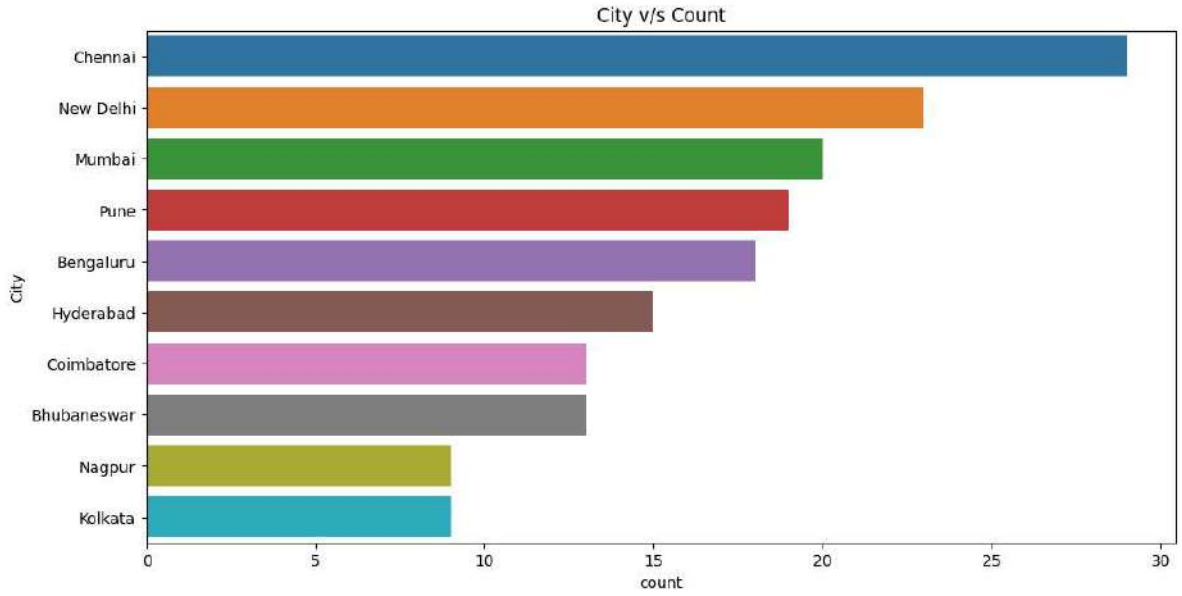
Query-19 : Distribution of colleges across cities in India

In [54]:

```
data['City'].value_counts()
plt.figure(figsize=(12,6))
sns.countplot(y='City',order = data['City'].value_counts().index[0:10],data=data)
plt.title('City v/s Count')
```

Out[54]:

```
Text(0.5, 1.0, 'City v/s Count')
```



Conclusion :

1.Chennai is the city with most colleges. 2.There are more colleges in metropolitan cites compared to other cities.

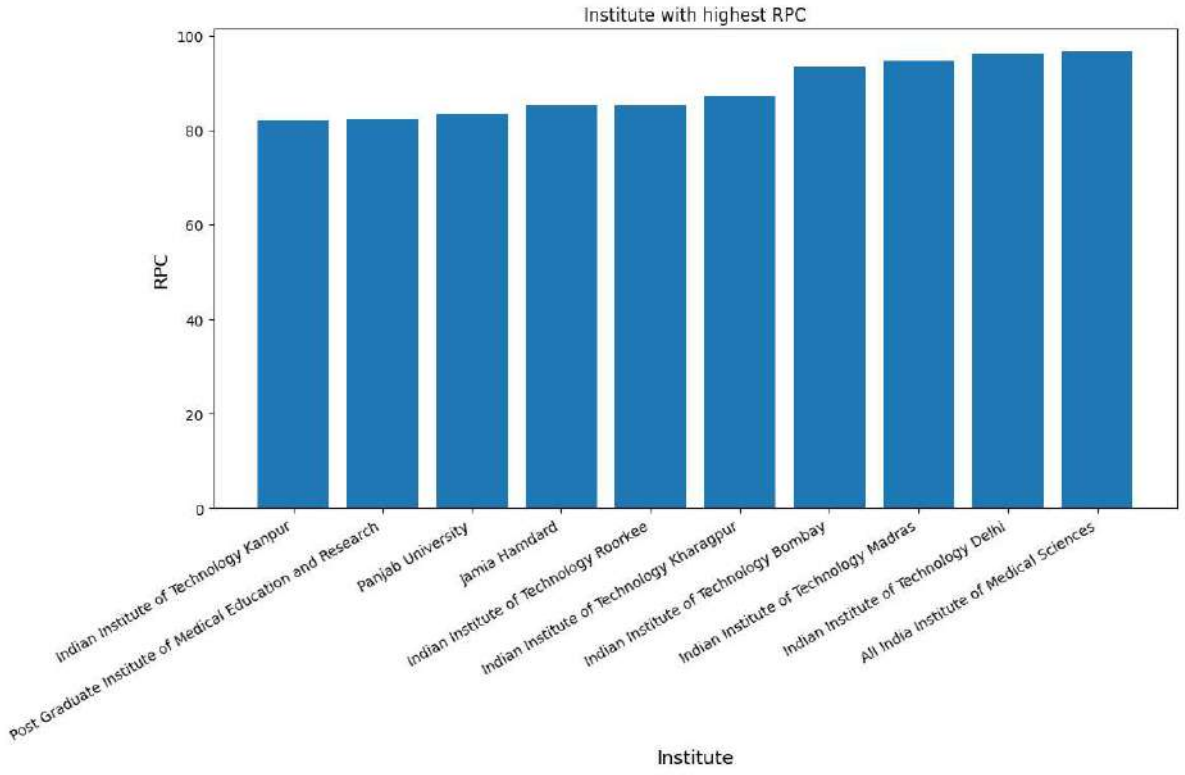
Query-20 : Top Colleges on the basis of Research and Professional Practice (RP)

In [41]:

```
states = data.sort_values(by='RPC').tail(10)
plt.rcParams['figure.figsize']=(12,6)
plt.title('Institute with highest RPC')
plt.xlabel('Institute', fontsize=13)
plt.ylabel('RPC', fontsize=13)
plt.xticks(rotation=30,ha='right')
plt.bar(states.Name,states.RPC)
```

Out[41]:

```
<BarContainer object of 10 artists>
```



Conclusion :

IIT Kanpur is the best college on the basis of RPC.

Final Outcomes of our Analysis :

From the analysis, it is observed that NIRF ranking has attracted more participation of institution of higher education from across india in all categories and subject domains.The number of particiapants has increased every year.It may also be noted that highest number of institutions are located in southern region of India.Major number of institute are in Tamil Nadu.And among these institutes most are in chennai.

In [ ]: