In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import plotly.express as px

warnings.filterwarnings('ignore')
```

```
C:\Users\Dell\anaconda3\lib\site-packages\pandas\core\computation\expressions.py:21: UserWarning: Pandas requires version '2.8.4' or newer of 'numexpr' (version '2.8.3' currently installed).
  from pandas.core.computation.check import NUMEXPR_INSTALLED
C:\Users\Dell\anaconda3\lib\site-packages\pandas\core\arrays\masked.py:60: UserWarning: Pandas requires version '1.3.6' or newer of 'bottleneck' (version '1.3.5' currently installed).
  from pandas.core import (
```

In [2]:
```
# Read the datasets into pandas DataFrame objects
inter_colleges= pd.read_csv('All the Intermediate Colleges in Pakistan.csv')
```

In [3]:
```
# Exploring Data With pandas method
inter_colleges.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1630 entries, 0 to 1629
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   Name           1630 non-null   object
 1   Location       1626 non-null   object
 2   Study Program  1630 non-null   object
 3   Sector         1101 non-null   object
 4   Affiliation    155 non-null    object
 5   Rating         1630 non-null   object
dtypes: object(6)
memory usage: 76.5+ KB
```

In [4]:
```
inter_colleges
```

Out[4]:

	Name	Location	Study Program	Sector	Affiliation	Rating
0	Campus.pk Lahore	Office 206, 2nd Floor Siddique Trade Center Ma...	FSc Pre Engineering Fsc Pre Medical	Private	ads	0*
1	Allama Iqbal Open University Aiou Islamabad	Sector H-8, Islamabad-44000	FSc Pre Engineering ICs Computer Science FSc P...	Public	NaN	5*
2	Government College University Gcu Lahore	GC University, Katchery Road Lahore 54000	FSc Pre Engineering ICS Computer Science FSc P...	Public	HEC	4*
3	Govt College University GCU Faisalabad	Kotwali Rd, Faisalabad 38000	FSc Pre-Engineering ICS FSc Pre-Medical Icom F...	Public	HEC	3*
4	Lahore College For Women University (Lcwu) Lahore	Near Wapda Flats, Jail Road, Lahore	FA Arts FSc Pre Engineering ICS Physics ICS St...	Public	NaN	5*
...	...	...	...	...	...	...
1625	Riphah International College Lodhran	Riphah International College Lodhran	Fsc Pre Engineering FSc Pre Medical ICS ICOM FA	Private	NaN	0*
1626	Govt College of Education	Bagh Azad Kashmir	FA ICS Computer Science	Public	NaN	0*

	Bagh		FSc Pre Engineering IC...			
1627	Riphah International College Swat	Riphah International College Swat	Fsc Pre Engineering FSc Pre Medical ICS ICOM FA	Private	NaN	0*
1628	Riphah International College Dina	Riphah International College Dina	Fsc Pre Engineering FSc Pre Medical ICS ICOM FA	Private	NaN	0*
1629	Riphah International College Dargai	Riphah International College Dargai	Fsc Pre Engineering FSc Pre Medical ICS ICOM FA	Private	NaN	0*

1630 rows × 6 columns

In [5]:
```
inter_colleges.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1630 entries, 0 to 1629
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   Name           1630 non-null   object
 1   Location       1626 non-null   object
 2   Study Program  1630 non-null   object
 3   Sector         1101 non-null   object
 4   Affiliation    155 non-null    object
 5   Rating         1630 non-null   object
dtypes: object(6)
memory usage: 76.5+ KB
```

In [6]:
```
inter_colleges.isnull().sum()
```

Out[6]:
```
Name              0
Location          4
Study Program     0
Sector          529
Affiliation    1475
Rating            0
dtype: int64
```

Handling the Missing values

In [7]:
```
# Dropping column "Affiliation"
inter_colleges.drop(['Affiliation'], axis ="columns", inplace =True)
```

In [8]:
```
# Filling Semi to null values
inter_colleges['Sector']= inter_colleges['Sector'].fillna(value= 'semi')

#Some Sector have null values. So we fill It with "Semi".
```

In [9]:
```
#Dropping the rows with null values from Location column
inter_colleges.dropna(subset = ["Location"] , inplace= True)
```

In [10]:
```
# Checking any null value remaining
inter_colleges.isnull().sum()
```

Out[10]:
```
Name            0
Location        0
Study Program   0
Sector          0
Rating          0
dtype: int64
```

In [11]:
```
inter_colleges.dtypes
```

Out[11]:
```
Name              object
Location          object
Study Program     object
Sector            object
Rating            object
dtype: object
```

Converting and cleaning the columns

In [12]:
```
inter_colleges["Rating"]
```

Out[12]:
```
0       0*
1       5*
2       4*
3       3*
4       5*
       ..
1625    0*
1626    0*
1627    0*
1628    0*
1629    0*
Name: Rating, Length: 1626, dtype: object
```

In [13]:
```
inter_colleges["Rating"] = inter_colleges["Rating"].str.replace('*','').astype(float)
```

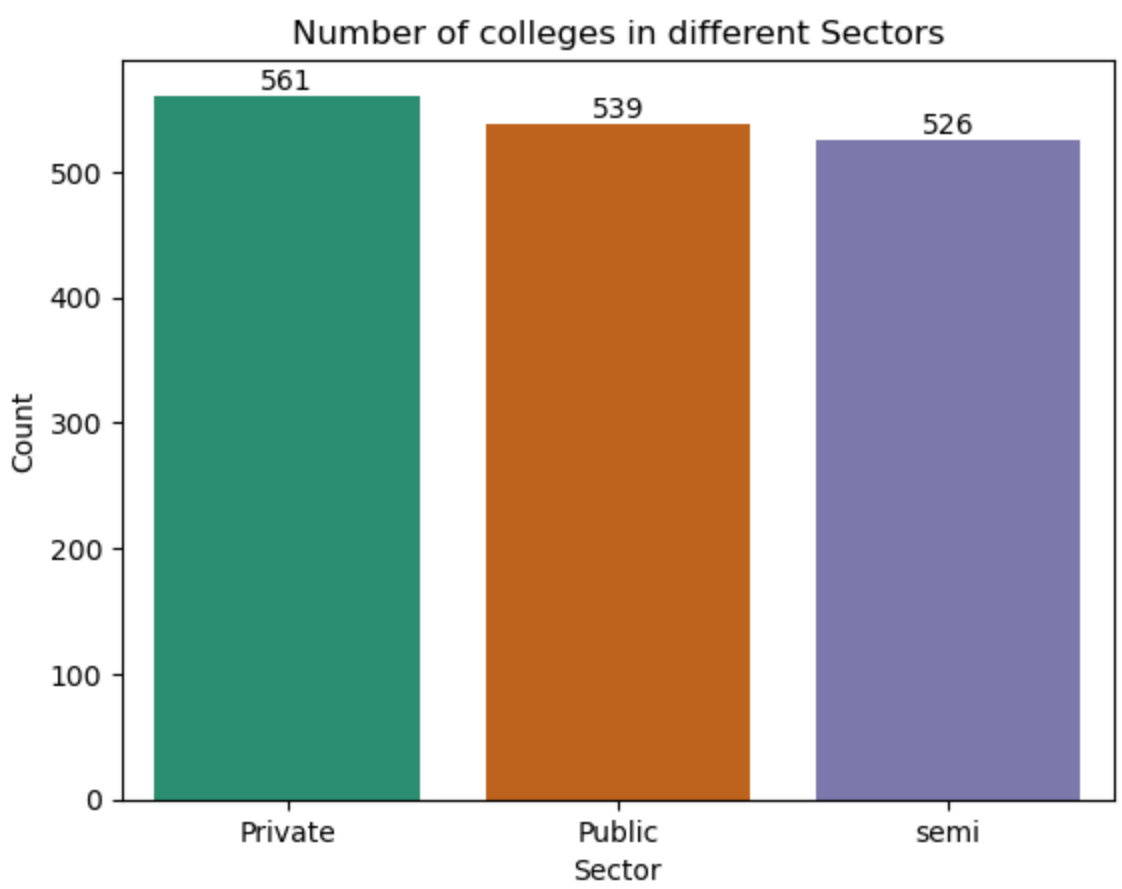
In [14]:
```
inter_colleges["Rating"]= inter_colleges["Rating"].round().astype(int)
inter_colleges["Rating"]
```

Out[14]:
```
0       0
1       5
2       4
3       3
4       5
      ..
1625    0
1626    0
1627    0
1628    0
1629    0
Name: Rating, Length: 1626, dtype: int32
```

In [15]:
```
ax = sns.countplot(x="Sector", data=inter_colleges, palette='Dark2')
for p in ax.patches:
    ax.annotate(format(p.get_height(), '.0f'), (p.get_x() + p.get_width() / 2., p.get_he
                ha='center', va='center', xytext=(0, 5), textcoords='offset points')

# Adjusting plot aesthetics
plt.xlabel("Sector")
plt.ylabel("Count")
plt.title("Number of colleges in different Sectors")

# Displaying the plot
plt.show()
```

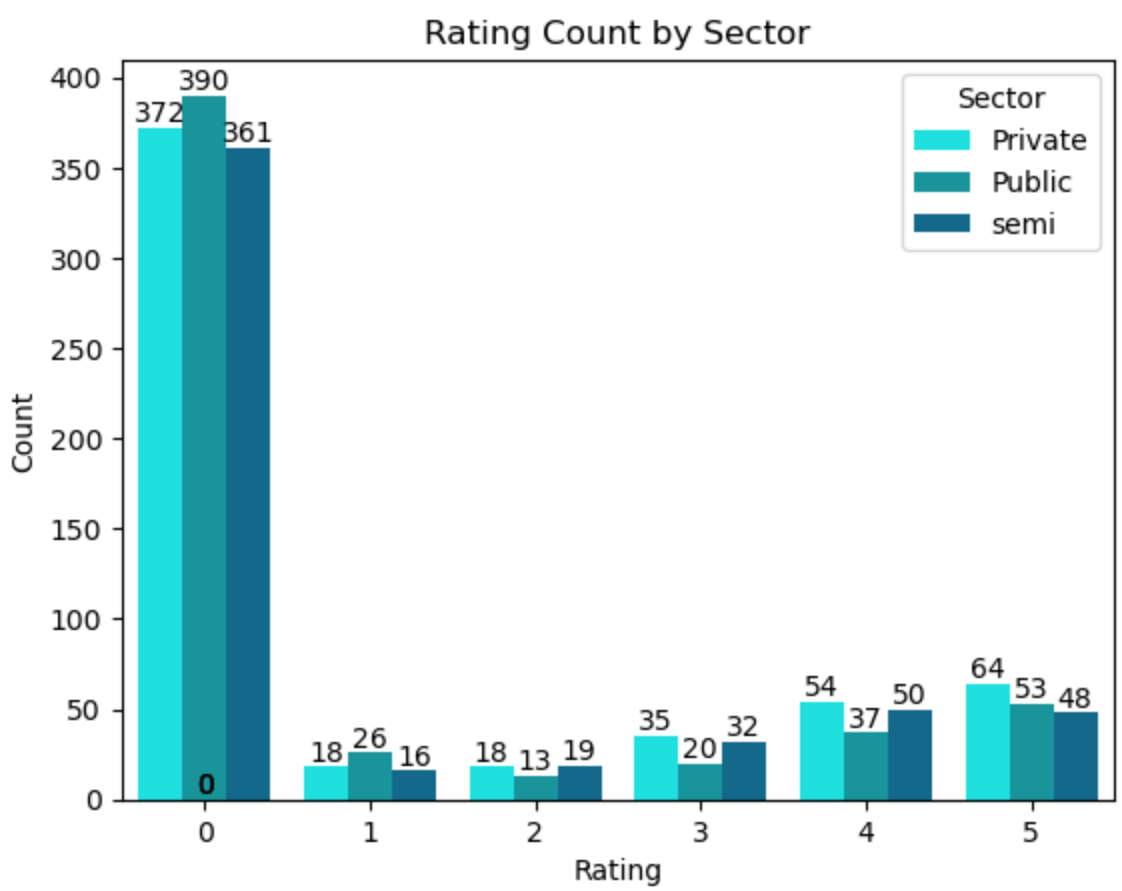


The plot shows that Private **`Colleges`** are more than **`Public Colleges`** and than comes colleges with semi-government system, we can't say anything about them because we're not sure(if they are private or public), because they were Nan values that we named ourselve as **`Semi`**.

In [16]:
```
colors = ['#00FFFF','#03A7B1','#00719F']
ax = sns.countplot(x="Rating", data=inter_colleges, hue="Sector", palette=colors)
for p in ax.patches:
    ax.annotate(format(p.get_height(), '.0f'), (p.get_x() + p.get_width() / 2., p.get_he
                ha = 'center', va = 'center', xytext = (0, 5), textcoords = 'offset poin

# Adjusting plot aesthetics
plt.xlabel("Rating")
plt.ylabel("Count")
plt.title("Rating Count by Sector")

# Displaying the plot
plt.show()
```



The plot above shows the **`Ratings`** of different colleges categorized by **`Sector`**.Somehow, it's showing that in **`higher`** rated colleges(3, 4, 5) we've more **`private colleges`**.

In [17]:
```
# Exracting the City name form the Name column

inter_colleges['City']= inter_colleges['Name'].apply(lambda x: x.split()[-1])
```

In [18]:
```
inter_colleges['City']
```

Out[18]:
```
0           Lahore
1        Islamabad
2           Lahore
3       Faisalabad
4           Lahore
           ...
1625       Lodhran
1626          Bagh
1627          Swat
1628          Dina
1629        Dargai
Name: City, Length: 1626, dtype: object
```

In [19]:
```
inter_colleges['City'].nunique()
```

Out[19]:
```
176
```

Total **`176`** cities data is available.

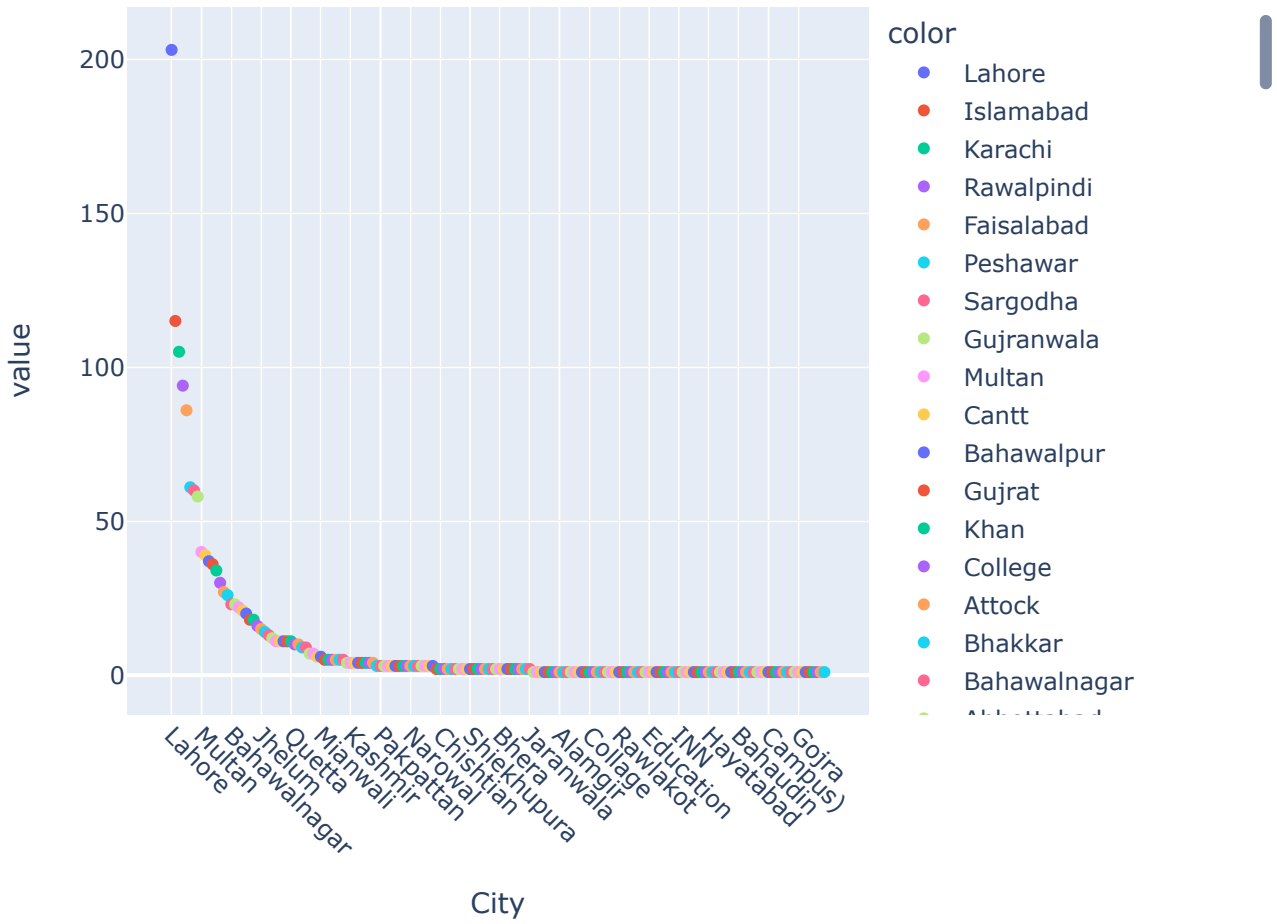
In [20]:
```
inter_colleges['City'].value_counts().head(15)
```

Out[20]:
```
City
Lahore        203
Islamabad     115
Karachi       105
```

```
Rawalpindi      94
Faisalabad      86
Peshawar        61
Sargodha        60
Gujranwala      58
Multan          40
Cantt           39
Bahawalpur      37
Gujrat          36
Khan            34
College         30
Attock          27
Name: count, dtype: int64
```

So, we've gained a glimpse of number of cities in different cities. `Lahore` has highest a number of colleges and then `Islamabad` , `Karachi` , `Rawalpindi` , `Faislabad` and so on...

In [21]:
```
# Number of Colleges city wise
cities = px.scatter(inter_colleges['City'].value_counts() ,color = inter_colleges['City'
cities.update_xaxes(tickangle=45)
```

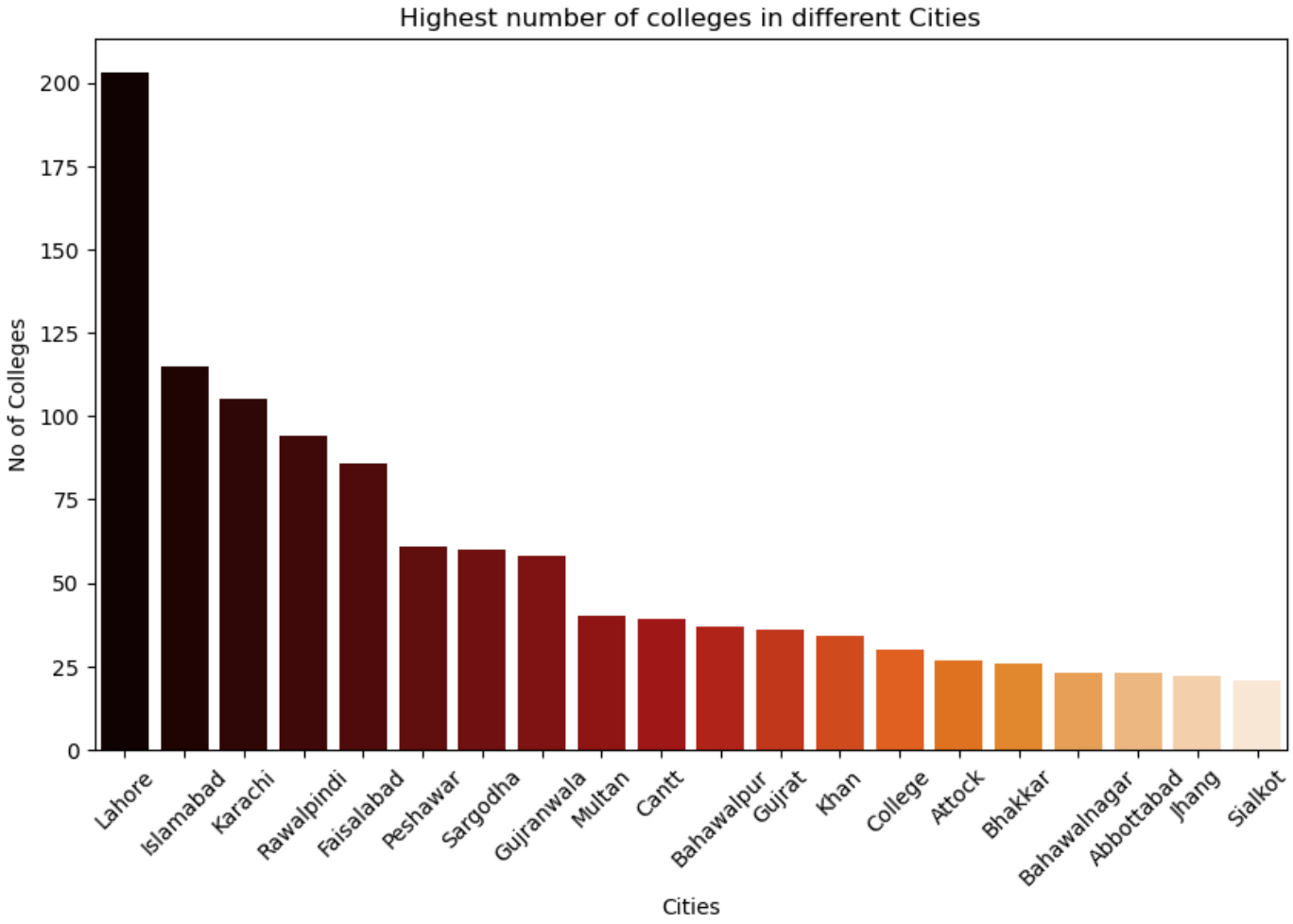


Through this plot we can have an overview of how much colleges are present in a specific city and in which cities we have a highest number of colleges.

`Lahore lies at the top in this regard with highest number of colleges(203) present there.`

In [22]:
```
# Highest number of colleges city wise bar plot
cities_count = inter_colleges['City'].value_counts().head(20)
```

```
plt.figure(figsize=(10,6))
sns.barplot(x=cities_count.index, y=cities_count.values, palette= 'gist_heat')
plt.title("Highest number of colleges in different Cities")
plt.xlabel('Cities')
plt.ylabel('No of Colleges')
plt.xticks(rotation=45)
plt.show()
```



In [23]:

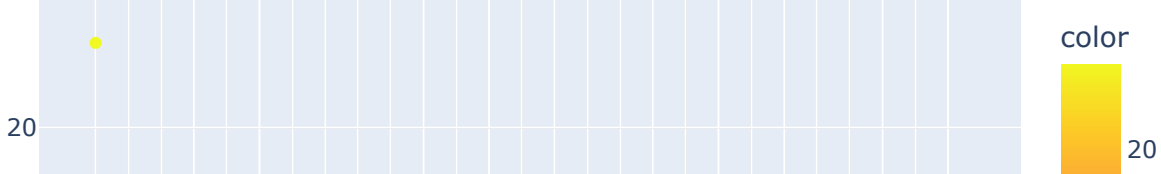
```
# Explor the Colleges with Rating 5.
rating = inter_colleges[inter_colleges['Rating'] == 5]
print('Number of Colleges with Rating 5:',rating.value_counts().sum())

fig =px.scatter(rating['Name'].str.split().str[-1].value_counts(),
           color=rating['Name'].str.split().str[-1].value_counts())


fig.update_layout(
    xaxis_title="City" ,
    yaxis_title='No of college with Rating 5'
)
fig.update_xaxes(tickangle=45)

fig.show()
```

```
Number of Colleges with Rating 5: 165
```



Lahore has the most number of Colleges with **Rating 5**

In [24]:
```
inter_colleges[inter_colleges['Rating'] ==5]
```

Out[24]:

	Name	Location	Study Program	Sector	Rating	City
1	Allama Iqbal Open University Aiou Islamabad	Sector H-8, Islamabad-44000	FSc Pre Engineering ICs Computer Science FSc P...	Public	5	Islamabad
4	Lahore College For Women University (Lcwu) Lahore	Near Wapda Flats, Jail Road, Lahore	FA Arts FSc Pre Engineering ICS Physics ICS St...	Public	5	Lahore
5	University Of Agriculture Faisalabad	University of Agriculture, Agriculture Univers...	FSc Pre Agriculture	Public	5	Faisalabad
6	Pir Mehr Ali Shah Arid Agriculture University ...	PMAS-Arid Agriculture University Rawalpindi, S...	FSc Pre Agriculture	Public	5	Rawalpindi
7	International Islamic University Iiu Islamabad	International Islamic University Sector H-10, ...	FSc Pre-Engineering ICS FSc Pre-Medical Icom F...	Public	5	Islamabad
...	...	...	...	...	...	...
1600	Fazaia Intermediate College Lahore	2-Durrand Road Garhi Shahu, Lahore	FA Icom ICS FSc Pre Engineering FSc Pre Medical	Private	5	Lahore
1604	Punjab Group of Colleges Nowshera	Khushab Road Near Shell Petrol Pump, Dhaka Mor...	FA ICS Computer Science FSc Pre Medical FSc Pr...	Private	5	Nowshera
1610	Govt College Boys Gublerg Lahore	Govt College Boys Gublerg Lahore	FA FSc Pre Engineering FSc Pre Medical ICS ICom	Public	5	Lahore
1616	Army Public School & Degree College For Boys P...	Army Public School & Degree College For Boys P...	FA FSc Pre Engineering FSc Pre Medical ICS Com...	Private	5	Cantt
1618	Govt Degree College for Boys Kahna Nou Lahore	Govt Degree College for Boys Kahna Nou Lahore	ICS Icom FA FSc Pre Engineering FSc Pre Medical	Public	5	Lahore

165 rows × 6 columns

In [25]:
```
# Highest number of colleges city wise bar plot
cities_count = inter_colleges['City'].value_counts().head(20)
plt.figure(figsize=(10,6))
sns.barplot(x=cities_count.index, y=cities_count.values, palette= 'gist_heat')
plt.title("Highest number of colleges in different Cities")
plt.xlabel('Cities')
plt.ylabel('No of Colleges')
plt.xticks(rotation=45)
plt.show()
```