Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mour
```

```
ls
```

```
'Computer vision End To End.ipynb'            New_Data.csv
'Copy of LibraryCode.ipynb'                   t10k-images-idx3-ubyte/
 Data_Set_all.csv                             t10k-labels-idx1-ubyte/
 Deepa_End_To_End_Code_Implementation.ipynb   train-images-idx3-ubyte/
 Img/                                         train-labels-idx1-ubyte/
 LibraryCode.ipynb                           'With Sigmoid.ipynb'
'Mnist Analysis.ipynb'
```

1.which lilbray was used more -- UG or Knowledge lib 2.WHich class is being used most of the libray 3.Group of MCA classes and then plot respect to 3i_time spend vs class 3ii_class vs venue 3iii_histo

4.which user had used mosty the library-- Regno with timespend 5.month wise

Refer note : http://www.dontpad.com/dee123

```
cd drive/My\ Drive
```

```
[Errno 2] No such file or directory: 'drive/My Drive'
/content/drive/My Drive/Colab Notebooks
```

```
cd Colab\ Notebooks
```

```
[Errno 2] No such file or directory: 'Colab Notebooks'
/content/drive/My Drive/Colab Notebooks
```

Main Page

```
import pandas as pd
import matplotlib as plot
import seaborn as sns
```

```
#reading the file and print
csvfile = 'Data_Set_all.csv'
phd_record_data = pd.read_csv(csvfile)
phd_record_data.head()
```

	SL.NO	Venue	Register No	First Name	Class	Department
0	1	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
1	2	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
2	4	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
3	5	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
4	7	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE

```
phd_record_data.info
```

```
<bound method DataFrame.info of        SL.NO                          Venue  ...     Ou
0          1        UG Library - Main campus  ...  11:58:32 AM    0:18:00
1          2        UG Library - Main campus  ...   3:57:01 PM    0:37:46
2          4        UG Library - Main campus  ...   4:40:58 PM    0:02:35
3          5        UG Library - Main campus  ...   4:01:28 PM    0:25:42
4          7        UG Library - Main campus  ...   4:00:43 PM    0:00:52
...      ...                             ...  ...          ...        ...
12679  17996  Knowledge Center - Main campus  ...   1:57:20 PM    0:18:09
12680  17998  Knowledge Center - Main campus  ...   1:41:47 PM    0:07:23
12681  17999  Knowledge Center - Main campus  ...   2:14:45 PM    0:32:02
12682  18000  Knowledge Center - Main campus  ...   3:17:38 PM    0:46:41
12683  18001  Knowledge Center - Main campus  ...   4:09:41 PM    0:50:10

[12684 rows x 10 columns]>
```

```
#drop na records in the dataset
phd_record_data= phd_record_data.dropna()
```

```
phd_record_data.dtypes
```

```
SL.NO           int64
Venue          object
Register No     int64
First Name     object
Class          object
Department     object
Date           object
In Time        object
Out Time       object
Time Spend     object
dtype: object
```

```
phd_record_data['Time Spend'] = pd.to_datetime(phd_record_data['Time Spend'])
#phd_record_data['Time Spend'] = phd_record_data['Time Spend'] .astype('datetime64[ns]')
```

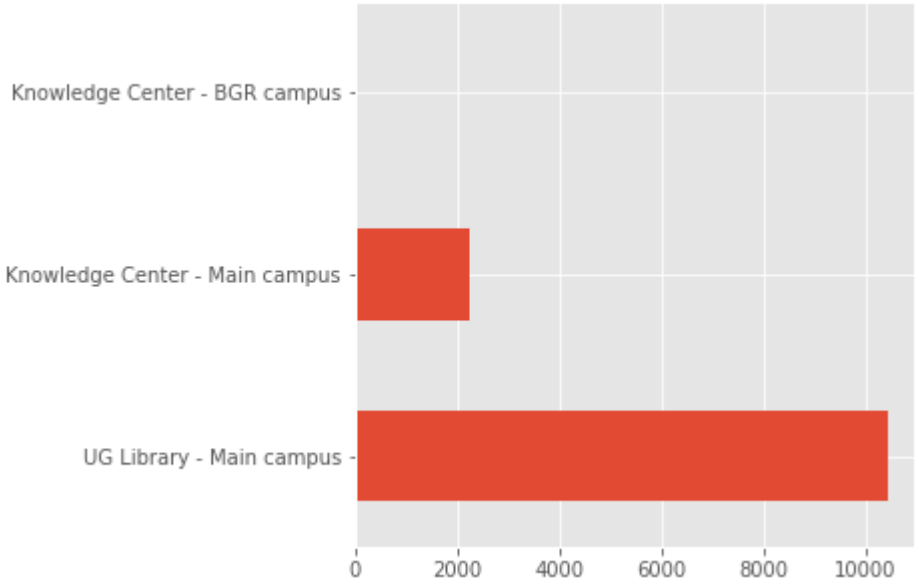
```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
```

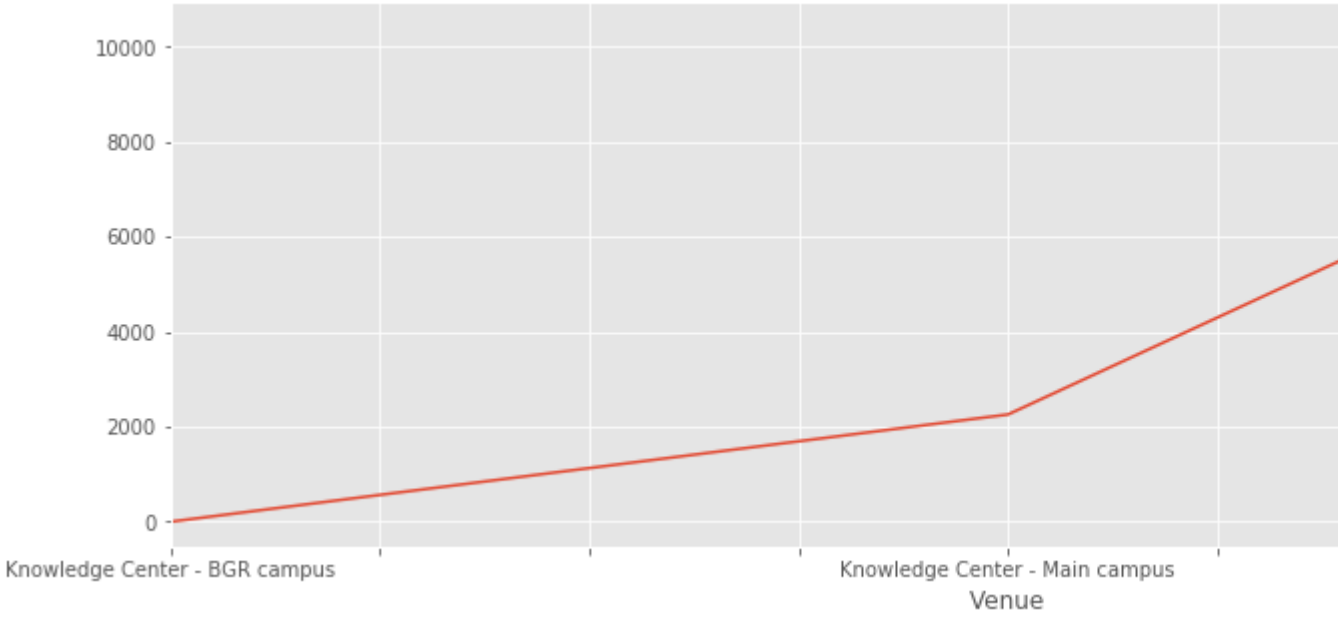
```
phd_record_data.dtypes
```

```
SL.NO                  int64
Venue                 object
Register No            int64
First Name            object
Class                 object
Department            object
Date                  object
In Time               object
Out Time              object
Time Spend     datetime64[ns]
dtype: object
```

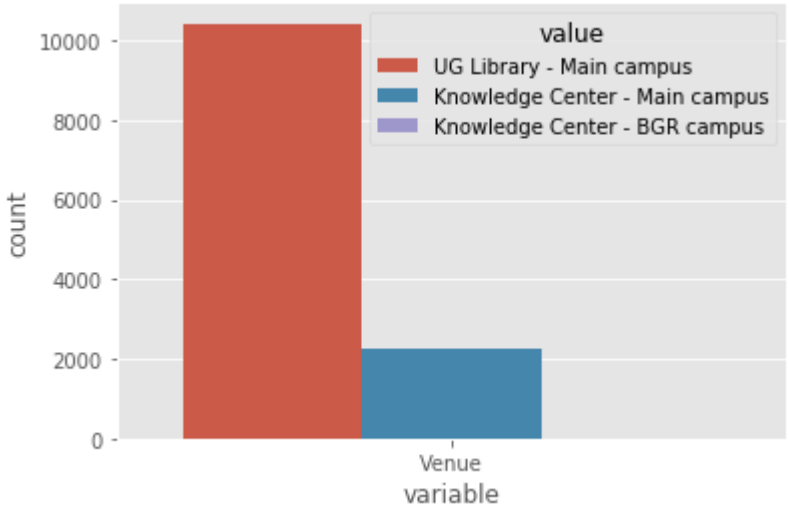
```
#which lilbray was used more -- UG or Knowledge lib
import matplotlib.pyplot as plt
phd_record_data['Venue'].value_counts().head(30).plot(kind='barh', figsize=(5,5))

fig, ax = plt.subplots(figsize=(15,5))
phd_record_data.groupby(['Venue']).count()['Class'].plot(ax=ax)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa62b15780>
```



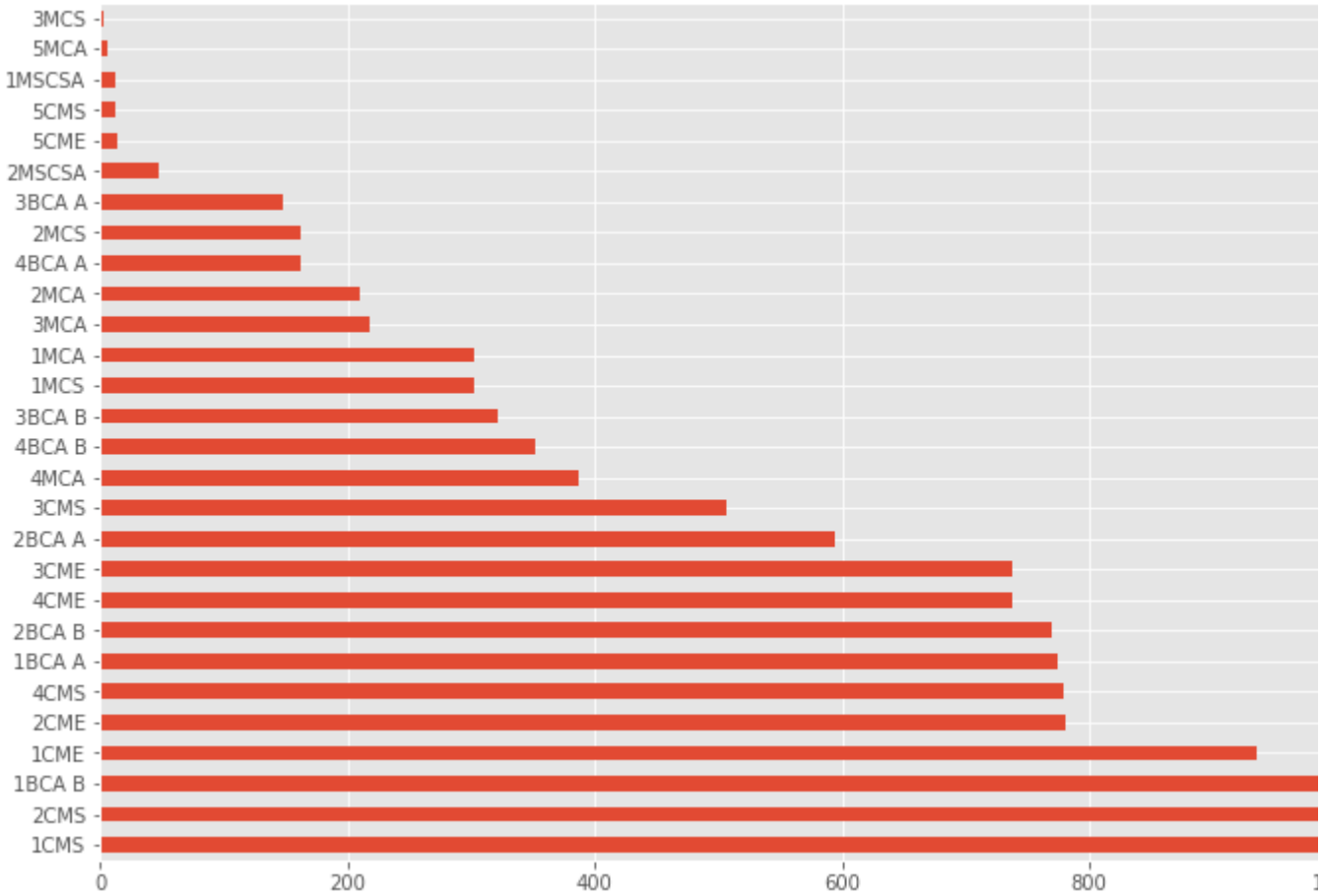



```
dataset_categ = phd_record_data.loc[:,['Venue']]
sns.countplot(x="variable", hue="value",data= pd.melt(dataset_categ));
```



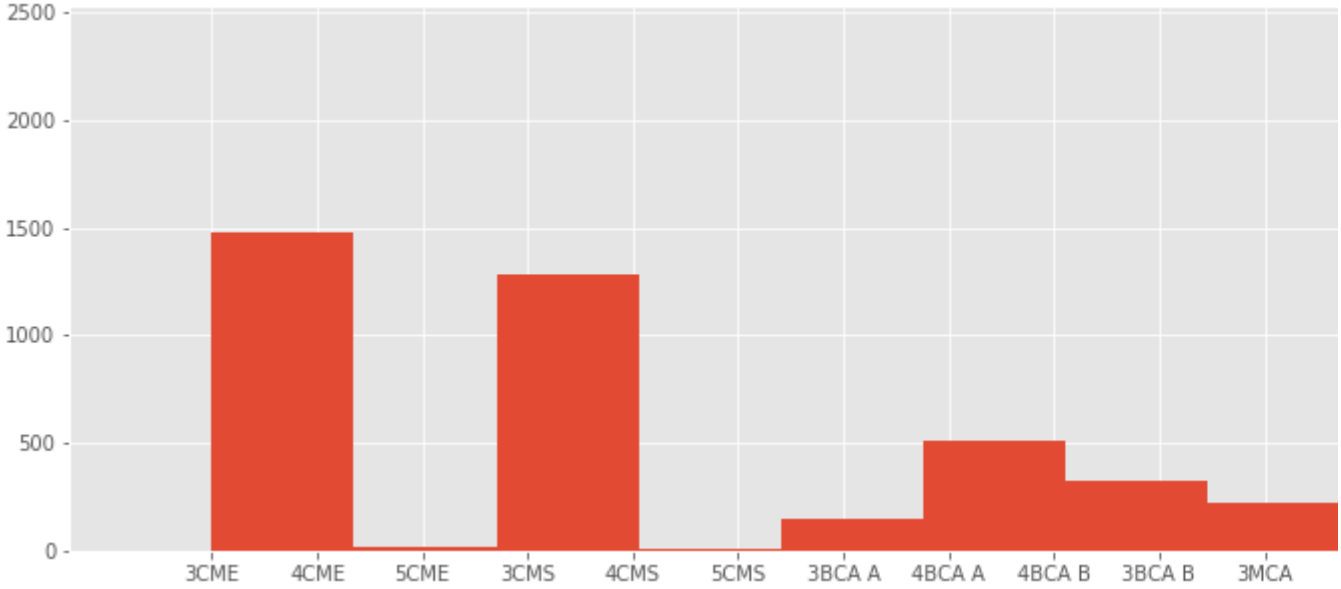

```
#WHich class is being used most of the libray
phd_record_data['Class'].value_counts().head(30).plot(kind='barh', figsize=(15,8))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa6325fdd8>
```



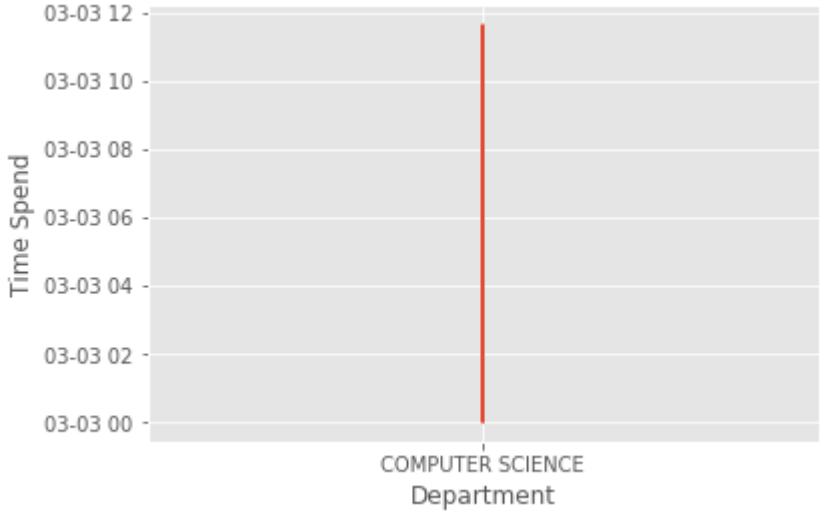

```
plt.figure(figsize=(28,5))
plt.style.use('ggplot')
plt.hist(phd_record_data['Class'], bins=20)
```

```
(array([1476.,   14., 1286.,   12.,  147.,  513.,  321.,  218.,  392.,
         936.,  780., 2408.,  774.,  594., 1773.,   12.,   47.,  465.,
           3.,  512.]),
 array([ 0.  ,  1.35,  2.7 ,  4.05,  5.4 ,  6.75,  8.1 ,  9.45, 10.8 ,
        12.15, 13.5 , 14.85, 16.2 , 17.55, 18.9 , 20.25, 21.6 , 22.95,
        24.3 , 25.65, 27.  ]),
 <a list of 20 Patch objects>)
```




```
#time spend vs Deparment
plt.plot(phd_record_data['Department'], phd_record_data['Time Spend'])
plt.xlabel('Department')
plt.ylabel('Time Spend')
```

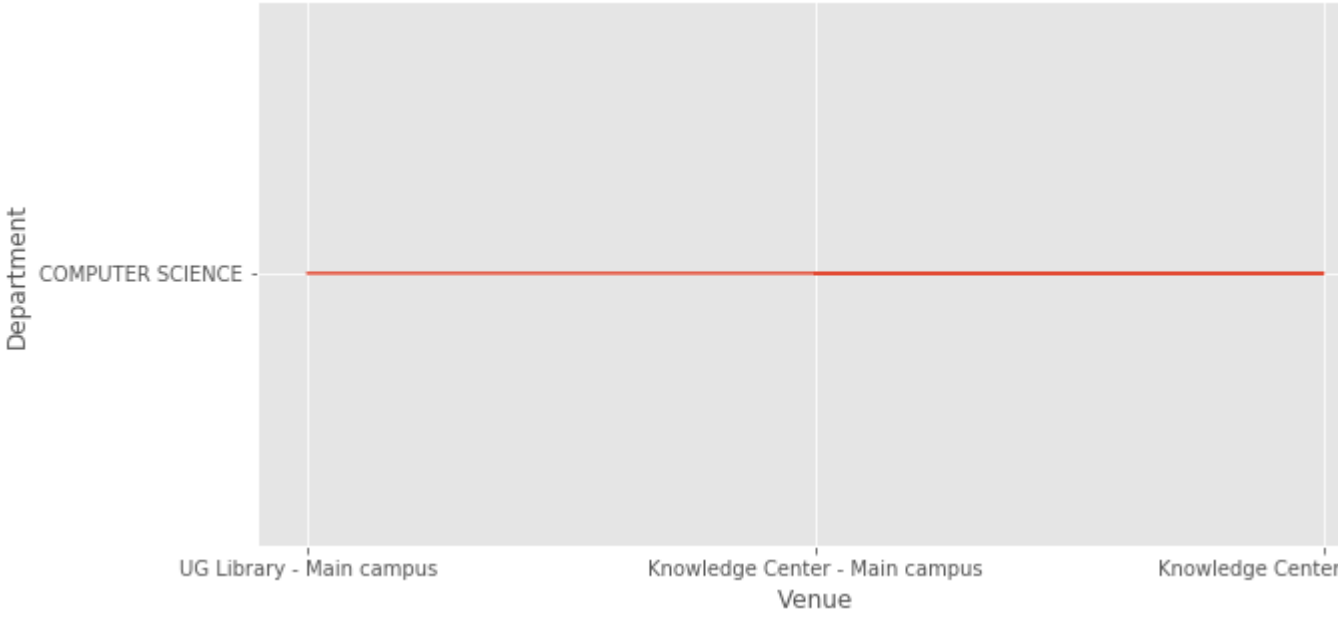
```
Text(0, 0.5, 'Time Spend')
```




```
#Venue vs Deparment
plt.figure(figsize=(10,5))
plt.plot(phd_record_data['Venue'], phd_record_data['Department'])
plt.xlabel('Venue')
```

```
plt.ylabel('Department')
```

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




```
##check any null value in data
phd_record_data.isnull().values.any()
```

```
False
```

```
phd_record_data.describe()
```

	SL.NO	Register No
count	12683.000000	1.268300e+04
mean	9045.619964	1.807312e+06
std	5248.402971	4.751938e+04
min	1.000000	1.740101e+06
25%	4396.500000	1.741115e+06
50%	9102.000000	1.840209e+06
75%	13610.500000	1.841050e+06
max	18001.000000	1.847267e+06

```
phd_record_data['singleclass']= phd_record_data['Class'].map(lambda x:x.lstrip('12345'))
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
phd_record_data['singlesem']= phd_record_data['Class'].map(lambda x:str(x)[:1])
phd_record_data['singlesem']
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
0        3
1        3
2        3
3        3
4        3
        ..
12679    3
12680    3
12681    3
12682    3
12683    3
Name: singlesem, Length: 12683, dtype: object
```

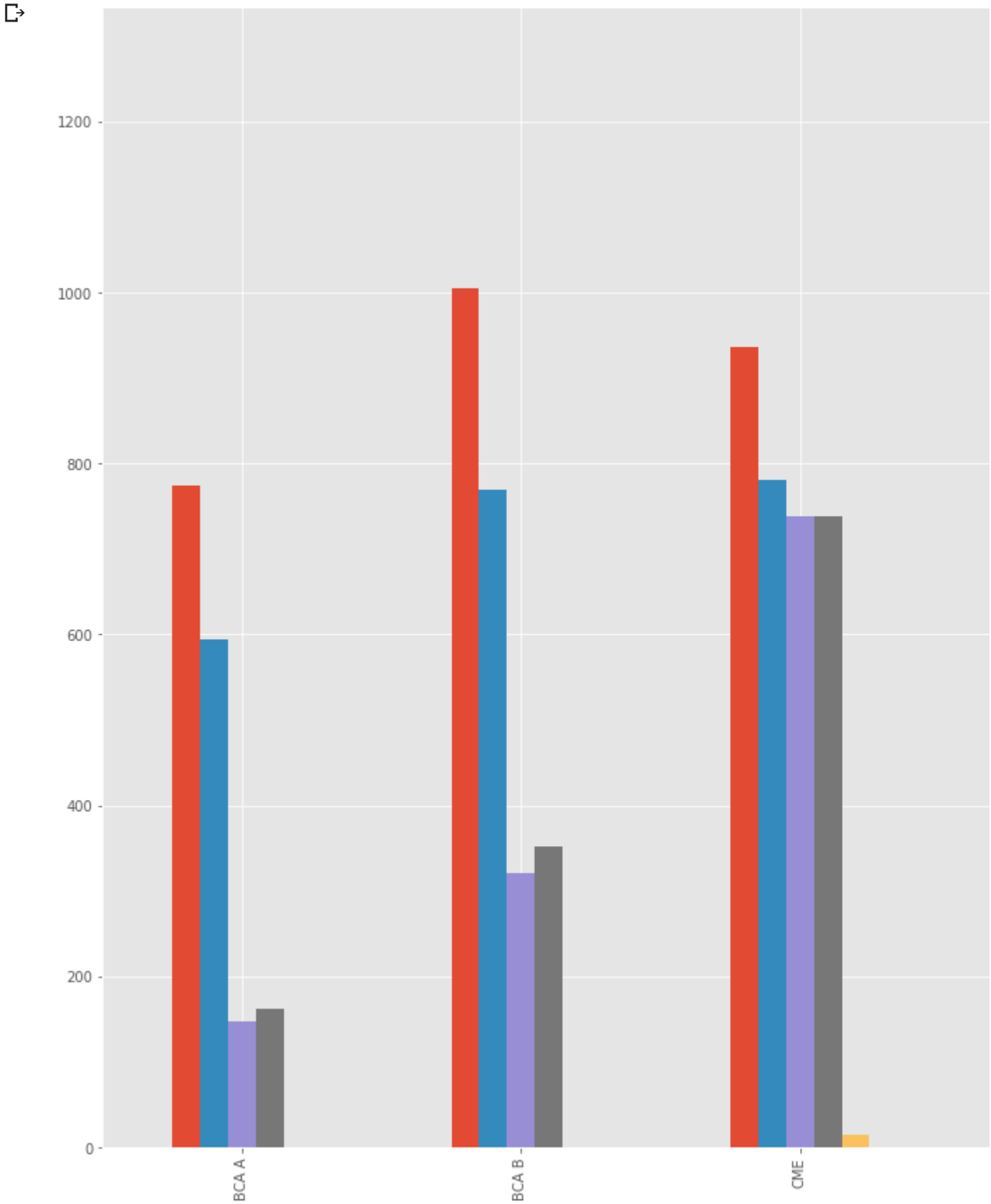
```
phd_record_data
```

	SL.NO	Venue	Register No	First Name	Class
0	1	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
1	2	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
2	4	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
3	5	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
4	7	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
...	...	...	...	...	...
12679	17996	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12680	17998	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12681	17999	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12682	18000	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12683	18001	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA

12683 rows × 12 columns

```
#Coursewise analysis by grouping similar groups
```

```
pd.crosstab(phd_record_data.singleclass,phd_record_data.singlesem).plot(kind="bar",figsize=(4
plt.show()
```



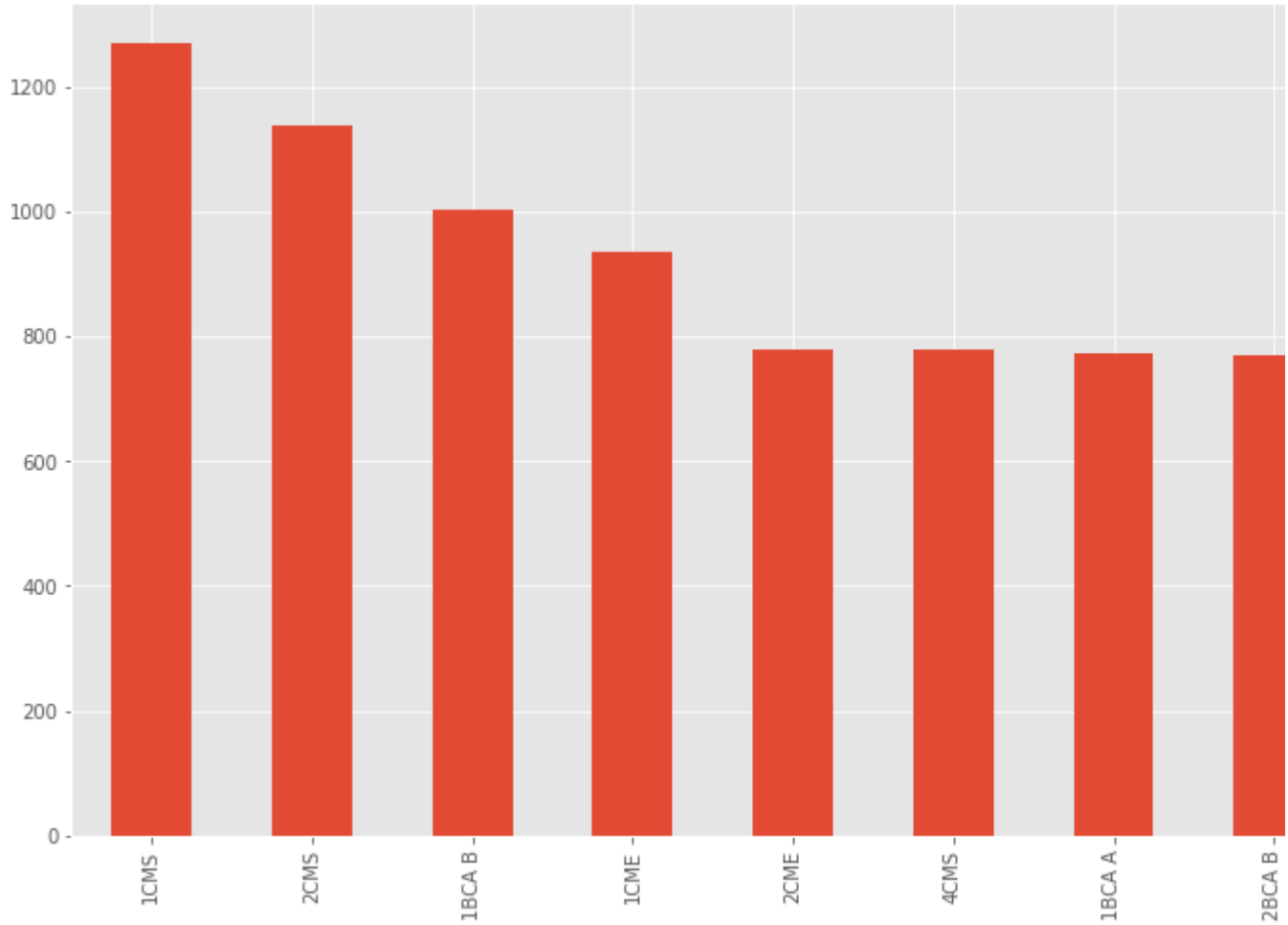

```
#phd record data['sort']= phd record data[['Class']].groupby([phd record data['singleyear'],p
```

```
phd_record_data['s'] = phd_record_data['Class'].value_counts()
```

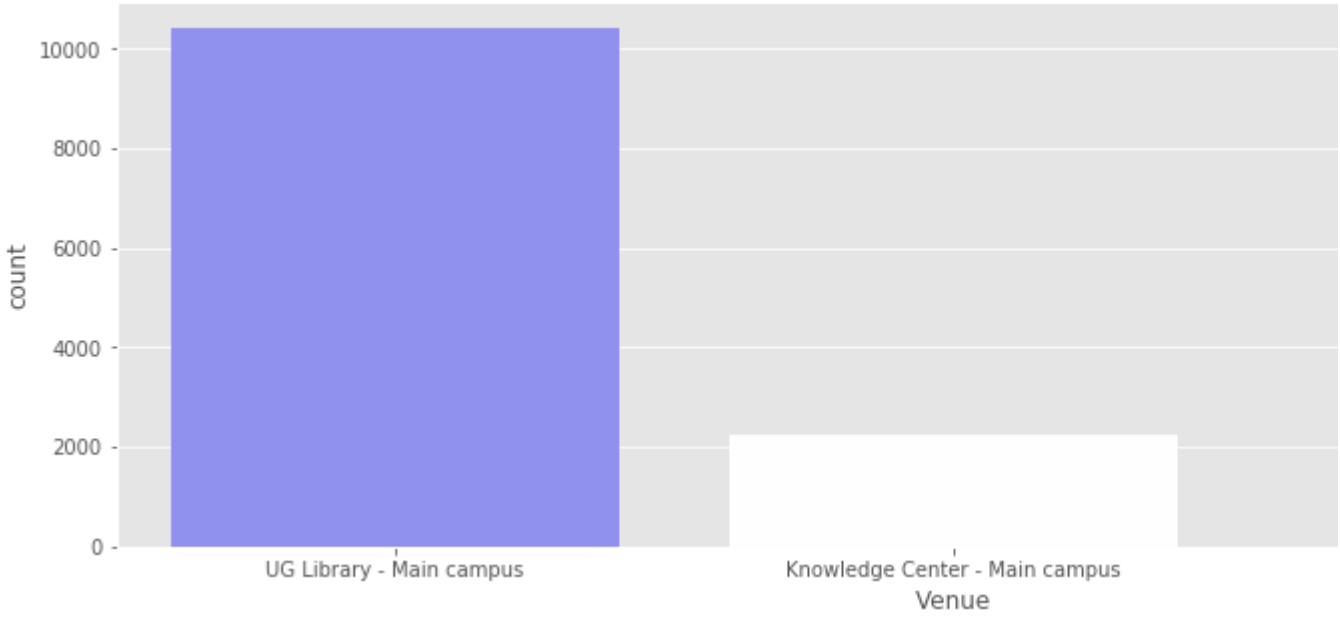
```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
```

```
#Sorting the classes based on highest number of usage
phd_record_data['Class'].value_counts().head(10).plot(kind='bar', figsize=(15,8))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa63177710>
```

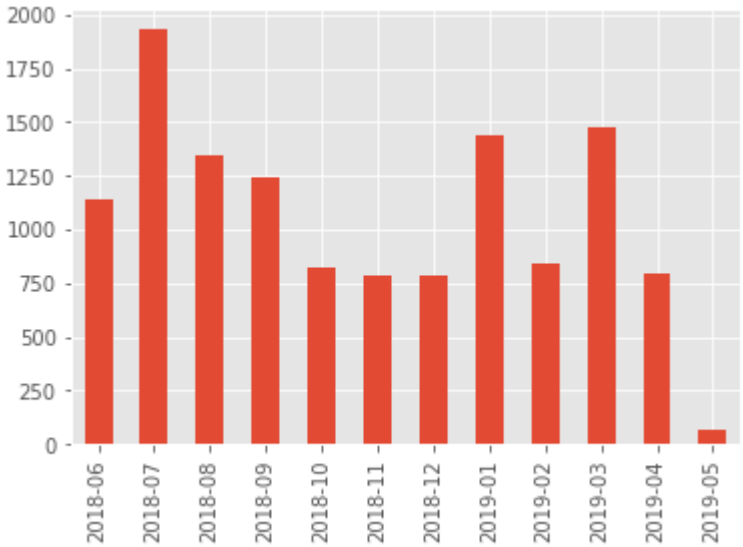



```
#Which library used more all over students
fig, ax = plt.subplots(figsize=(15,5))
sns.countplot(x="Venue",data=phd_record_data,palette = "bwr")
plt.show()
```




```
#month wise most used library using plot
import pandas as pd
s = pd.to_datetime(pd.Series(phd_record_data['Date']), format='%d/%m/%Y')
s.index = s.dt.to_period('m')
s = s.groupby(level=0).size()
s = s.reindex(pd.period_range(s.index.min(), s.index.max(), freq='m'), fill_value=0)
print (s)
s.sort_values()
s.plot.bar()
```

```
2018-06    1143
2018-07    1932
2018-08    1349
2018-09    1241
2018-10     823
2018-11     788
2018-12     785
2019-01    1436
2019-02     843
2019-03    1476
2019-04     799
2019-05      68
Freq: M, Name: Date, dtype: int64
<matplotlib.axes._subplots.AxesSubplot at 0x7faa627b1d30>
```




```
phd_record_data['month'] = pd.DatetimeIndex(phd_record_data['Date']).month
phd_record_data.groupby(['month']).count()
grouped = phd_record_data.groupby(['month'])
size = grouped.size()
size
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
month
1     1003
2      834
3     1692
4      711
5      428
6     1348
7     1612
8      760
9     1334
10     671
11    1350
12     940
dtype: int64
```

```
s = pd.to_datetime(pd.Series(phd_record_data['Date']), format='%d/%m/%Y')
```

```
phd_record_data['Date']=pd.to_datetime(phd_record_data['Date'], format='%d/%m/%Y').dt.strftim
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
```

```
phd_record_data['month'] = pd.DatetimeIndex(phd_record_data['Date']).month
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """Entry point for launching an IPython kernel.
```

```
phd_record_data.tail()
phd_record_data.head()
```

	SL.NO	Venue	Register No	First Name	Class	Department
0	1	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
1	2	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
2	4	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
3	5	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE
4	7	UG Library - Main campus	1740101	ABHAY TOMAR	3CME	COMPUTER SCIENCE

```
for i in phd_record_data.index:
  if(phd_record_data['month'][i] in [6,7,8]):
     phd_record_data['month'][i]= "Q1"
  elif(phd_record_data['month'][i] in [9,10,11]):
     phd_record_data['month'][i]= "Q2"
  elif(phd_record_data['month'][i] in [12,1,2]):
     phd_record_data['month'][i]= "Q3"
  else:
      phd_record_data['month'][i]= "Q4"
phd_record_data
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  This is separate from the ipykernel package so we can avoid doing imports until
/usr/local/lib/python3.6/dist-packages/pandas/core/indexing.py:205: SettingWithCopyWarni
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  self._setitem_with_indexer(indexer, value)
/usr/local/lib/python3.6/dist-packages/pandas/core/series.py:1221: SettingWithCopyWarnir
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  self.loc[key] = value
/usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2882: SettingWit
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  exec(code_obj, self.user_global_ns, self.user_ns)
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  """
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  import sys
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_g
  if __name__ == '__main__':
```

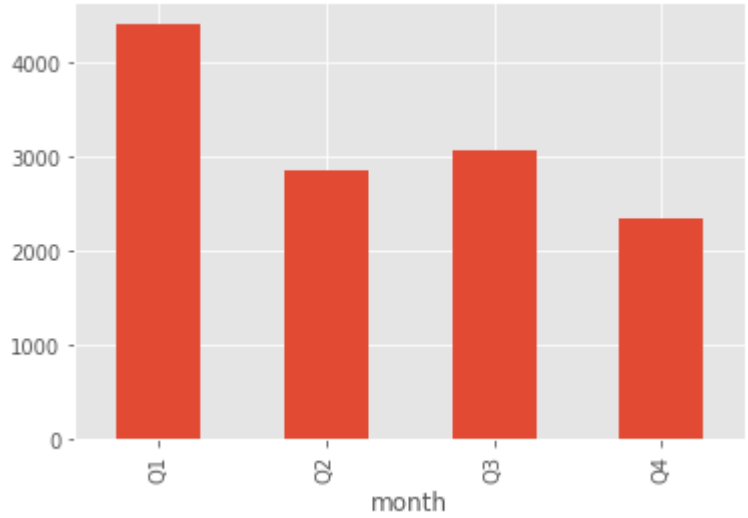
	SL.NO	Venue	Register No	First Name	Class
0	1	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
1	2	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
2	4	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
3	5	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
4	7	UG Library - Main campus	1740101	ABHAY TOMAR	3CME
...	...	...	...	...	...
12679	17996	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA

12680	17998	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12681	17999	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12682	18000	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA
12683	18001	Knowledge Center - Main campus	1847267	KUMAR NAVIN BARNWAL	3MCA

12683 rows × 14 columns

```
#Quarterly based analysis
#Q1 Jun Jul Aug Q2 Sep Oct Nov Q3 Dec Jan Feb Q4 Mar Apr May
phd_record_data.groupby(['month']).count()
grouped = phd_record_data.groupby(['month'])
size = grouped.size()
size
size.plot.bar()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa62763f28>
```




```
print(phd_record_data.columns)
```

```
Index(['SL.NO', 'Venue', 'Register No', 'First Name', 'Class', 'Department',
       'Date', 'In Time', 'Out Time', 'Time Spend', 'singleclass', 'singlesem',
       's', 'month'],
      dtype='object')
```

unwanted

15 cells hidden

END