NIRF Rank Predictor

The goal of this machine learning problem is to build a predictive model that can accurately estimate the NIRF ranking of colleges and universities based on a set of relevant features and historical ranking data. By doing so, we aim to assist colleges and universities in assessing and enhancing their performance in various NIRF ranking parameters.

It is carried out in following steps:-

- Data Preprocessing and data cleaning
- Transforming raw data into features that can be used to create predictive models
- Exploratory Data Analysis
- Assessing various Machine Learning models
- Training and Testing ML models
- Finalizing the best model suited.

As per the methodology of NIRF, there are five ranking parameters which are as follows:

- Teaching, learning, and resources
- Research and professional practice
- Graduation outcomes
- Outreach and inclusivity
- Peer perception

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Data of year 2016, 2017, 2018, 2019, 2020, 2021 is arranged from kaggle. The dataset contains:-

- Institute ID
- Institute Name
- City where it is located
- State
- Rank of various years
- TLR (Teaching, Learning and Resources)
- RPC (Research And Professional Practice)
- GO (Graduation Outcomes)
- OI (Outreach And Inclusivity)
- Perception

The dataset is then combined into one dataset.

```
data_2016 = pd.read_csv('EngineeringRanking_2016.csv')
```

```
data_2016.head()
```

```
        Institute Id                                Institute Name
City  \
0   NIRF-ENGG-INF-77     Indian Institute Of Technology, Madras
Chennai
1  NIRF-ENGG-INF-312     Indian Institute Of Technology, Bombay
Bombay
2  NIRF-ENGG-INF-300  Indian Institute Of Technology, Kharagpur
Kharagpur
3   NIRF-ENGG-INF-79      Indian Institute Of Technology, Delhi  New
Delhi
4  NIRF-ENGG-INF-228     Indian Institute Of Technology, Kanpur
Kanpur

           State  Score Rank    TLR    RPC     GO     OI  Perception
0     Tamil Nadu  89.41    1  88.26  94.02  81.81  86.11          98
1    Maharashtra  87.66    2  85.93  94.14  84.97  74.84          99
2    West Bengal  83.91    3  76.23  92.68  83.95  78.05          97
3          Delhi  82.02    4  80.27  91.62  74.72  66.17          98
4  Uttar Pradesh  81.07    5  66.08  93.52  85.62  70.59          98
```

```
data_2017 = pd.read_csv('EngineeringRanking_2017.csv')
```

```
data_2017.head()
```

```
        Institute Id                                 Institute Name
City  \
0   IR17-ENGG-1-1-77     Indian Institute of Technology Madras
Chennai
1  IR17-ENGG-2-18633     Indian Institute of Technology Bombay
Mumbai
2  IR17-ENGG-2-18630  Indian Institute of Technology Kharagpur
Kharagpur
3   IR17-ENGG-2-1-79      Indian Institute of Technology Delhi  New
Delhi
4  IR17-ENGG-2-18248     Indian Institute of Technology Kanpur
Kanpur

           State  Score  Rank    TLR    RPC     GO     OI  Perception

0     Tamil Nadu  87.96     1  91.85  92.60  83.78  77.19       81.46

1    Maharashtra  87.87     2  91.15  94.68  83.64  69.70       84.24

2    West Bengal  81.93     3  76.03  89.23  88.02  74.11       73.43

3          Delhi  81.08     4  79.63  89.47  77.45  71.41       77.24

4  Uttar Pradesh  76.83     5  84.28  77.28  74.29  61.35       73.59
```

```
data_2018 = pd.read_csv('EngineeringRanking_2018.csv')

data_2018.head()
```

```
          Institute Id                                Institute Name
City  \
0     IR-2-E-OE-U-0456     Indian Institute of Technology Madras
Chennai
1    IR-3-E-OEM-U-0306     Indian Institute of Technology Bombay
Mumbai
2    IR-3-E-OEM-I-1074      Indian Institute of Technology Delhi  New
Delhi
3  IR-5-E-OEMAL-U-0573  Indian Institute of Technology Kharagpur
Kharagpur
4    IR-3-E-OEM-I-1075     Indian Institute of Technology Kanpur
Kanpur

           State  Score Rank    TLR    RPC     GO     OI  Perception
0     Tamil Nadu  88.95    1  93.83  91.44  84.91  63.88      100.00
1    Maharashtra  84.82    2  89.61  96.04  76.53  44.71       93.48
2          Delhi  82.18    3  80.83  89.35  81.47  59.72       88.60
3    West Bengal  77.78    4  73.73  84.26  85.65  53.99       78.51
4  Uttar Pradesh  75.24    5  78.51  77.15  78.99  41.46       85.89
```

```
data_2019 = pd.read_csv('EngineeringRanking_2019.csv')

data_2019.head()
```

```
  Institute Id                                Institute Name       City  \
0  IR-E-U-0456     Indian Institute of Technology Madras    Chennai
1  IR-E-I-1074      Indian Institute of Technology Delhi  New Delhi
2  IR-E-U-0306     Indian Institute of Technology Bombay     Mumbai
3  IR-E-U-0573  Indian Institute of Technology Kharagpur  Kharagpur
4  IR-E-I-1075     Indian Institute of Technology Kanpur     Kanpur

           State  Score  Rank    TLR    RPC     GO     OI  Perception

0     Tamil Nadu  89.05     1  93.55  92.39  84.36  63.99      100.00

1          Delhi  85.36     2  85.80  96.18  80.32  56.19       90.85

2    Maharashtra  84.40     3  89.12  95.30  76.47  48.17       89.61

3    West Bengal  79.41     4  73.14  88.20  84.12  57.79       84.14

4  Uttar Pradesh  77.57     5  79.07  81.79  82.56  46.61       81.35
```

```
data_2020 = pd.read_csv('EngineeringRanking_2020.csv')

data_2020.head()
```

```
  Institute Id                                Institute Name       City  \
0  IR-E-U-0456     Indian Institute of Technology Madras    Chennai
1  IR-E-I-1074      Indian Institute of Technology Delhi  New Delhi
2  IR-E-U-0306     Indian Institute of Technology Bombay     Mumbai
3  IR-E-I-1075     Indian Institute of Technology Kanpur     Kanpur
4  IR-E-U-0573  Indian Institute of Technology Kharagpur  Kharagpur

           State  Score  Rank    TLR    RPC     GO     OI  Perception

0     Tamil Nadu  89.93     1  95.42  94.64  83.90  61.31      100.00

1          Delhi  88.08     2  90.79  96.15  80.36  64.81       94.46

2    Maharashtra  85.08     3  91.00  93.37  77.60  49.99       92.51

3  Uttar Pradesh  82.18     4  86.22  82.08  88.44  54.21       85.78

4    West Bengal  80.56     5  77.32  87.11  83.21  56.62       89.31
```

```
data_2021 = pd.read_csv('EngineeringRanking_2021.csv')

data_2021.head()
```

```
  Institute Id                                Institute Name       City  \
0  IR-E-U-0456     Indian Institute of Technology Madras    Chennai
1  IR-E-I-1074      Indian Institute of Technology Delhi  New Delhi
2  IR-E-U-0306     Indian Institute of Technology Bombay     Mumbai
3  IR-E-I-1075     Indian Institute of Technology Kanpur     Kanpur
4  IR-E-U-0573  Indian Institute of Technology Kharagpur  Kharagpur

           State  Score  Rank    TLR    RPC     GO     OI  Perception

0     Tamil Nadu  90.19     1  95.47  96.43  81.92  62.44      100.00

1          Delhi  88.96     2  91.76  95.82  80.97  66.39       98.63

2    Maharashtra  85.16     3  89.32  92.56  79.71  53.68       92.88

3  Uttar Pradesh  83.22     4  86.71  83.13  89.79  55.96       87.05

4    West Bengal  82.03     5  80.51  88.59  83.01  58.46       88.50
```

```
data_2016['year'],data_2017['year'],data_2018['year'],data_2019['year'
],data_2020['year'],data_2021['year'] = '2016' , '2017' , '2018' ,
'2019' , '2020' , '2021'

df = [data_2016,data_2017,data_2018,data_2019,data_2020,data_2021]
df_combined = pd.concat(df , axis =0 , ignore_index = 'True')
```

```
# Combined dataset
df_combined.head()
```

```
        Institute Id                                Institute Name
City  \
0   NIRF-ENGG-INF-77     Indian Institute Of Technology, Madras
Chennai
1  NIRF-ENGG-INF-312     Indian Institute Of Technology, Bombay
Bombay
2  NIRF-ENGG-INF-300  Indian Institute Of Technology, Kharagpur
Kharagpur
3   NIRF-ENGG-INF-79      Indian Institute Of Technology, Delhi  New
Delhi
4  NIRF-ENGG-INF-228     Indian Institute Of Technology, Kanpur
Kanpur

            State  Score Rank    TLR    RPC     GO     OI  Perception
year
0     Tamil Nadu  89.41    1  88.26  94.02  81.81  86.11        98.0
2016
1    Maharashtra  87.66    2  85.93  94.14  84.97  74.84        99.0
2016
2    West Bengal  83.91    3  76.23  92.68  83.95  78.05        97.0
2016
3          Delhi  82.02    4  80.27  91.62  74.72  66.17        98.0
2016
4  Uttar Pradesh  81.07    5  66.08  93.52  85.62  70.59        98.0
2016
```

```
df_combined
```

```
          Institute Id                                  Institute Name  \
0     NIRF-ENGG-INF-77      Indian Institute Of Technology, Madras
1    NIRF-ENGG-INF-312      Indian Institute Of Technology, Bombay
2    NIRF-ENGG-INF-300   Indian Institute Of Technology, Kharagpur
3     NIRF-ENGG-INF-79       Indian Institute Of Technology, Delhi
4    NIRF-ENGG-INF-228      Indian Institute Of Technology, Kanpur
..                 ...                                         ...
895        IR-E-C-1438       The National Institute of Engineering
896       IR-E-C-33584        K. J. Somaiya College of Engineering
897       IR-E-C-27400  Kakatiya Institute of Technology & Science
898       IR-E-C-11015            Walchand College of Engineering
899        IR-E-U-0037                Sri Venkateswara University

           City           State  Score Rank    TLR    RPC     GO
OI  \
0       Chennai      Tamil Nadu  89.41    1  88.26  94.02  81.81
86.11
1        Bombay     Maharashtra  87.66    2  85.93  94.14  84.97
74.84
```

```
2     Kharagpur     West Bengal  83.91    3  76.23  92.68  83.95
78.05
3    New Delhi            Delhi  82.02    4  80.27  91.62  74.72
66.17
4        Kanpur   Uttar Pradesh  81.07    5  66.08  93.52  85.62
70.59
..          ...             ...    ...  ...    ...    ...    ...    ..
.
895      Mysore       Karnataka  32.52  196  53.79   2.33  51.03
50.48
896      Mumbai     Maharashtra  32.48  197  52.22   3.33  58.94
38.08
897    Warangal       Telangana  32.48  197  55.80   1.64  49.13
49.25
898      Sangli     Maharashtra  32.46  199  48.25   4.54  56.11
47.93
899    Tirupati  Andhra Pradesh  32.42  200  41.76  25.59  37.58
43.31

     Perception  year
0         98.00  2016
1         99.00  2016
2         97.00  2016
3         98.00  2016
4         98.00  2016
..          ...   ...
895        4.23  2021
896        2.17  2021
897        4.90  2021
898        6.19  2021
899        3.56  2021

[900 rows x 12 columns]

df_combined.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 900 entries, 0 to 899
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Institute Id    900 non-null    object
 1   Institute Name  900 non-null    object
 2   City            900 non-null    object
 3   State           900 non-null    object
 4   Score           900 non-null    float64
 5   Rank            900 non-null    object
 6   TLR             900 non-null    float64
 7   RPC             900 non-null    float64
 8   GO              900 non-null    float64
```

```
 9   OI              900 non-null    float64
 10  Perception      900 non-null    float64
 11  year            900 non-null    object
dtypes: float64(6), object(6)
memory usage: 84.5+ KB
```

There are certain anomalies present in the data like in some fields Rank 21A , 26A is present. The letter is removed using lambda function and the datatype of rank is converted from string to float.

```
df_combined[219:228]
```

```
            Institute Id                                       Institute
Name  \
219     IR-2-E-EM-I-1480     Thapar Institute of Engineering and
Technology
220     IR-2-E-OE-U-0237         National Institute of Technology
Surathkal
221     IR-2-E-OE-U-0584  Indian Institute of Engineering Science and
Te...
222     IR-2-E-OE-U-0378               Indian Institute of Technology
Ropar
223     IR-2-E-OE-U-0255   Indian Institute of Space Science and
Technology
224     IR-2-E-OE-U-0064               Indian Institute of Technology
Patna
225     IR-2-E-OE-U-0025          National Institute of Technology
Warangal
226  IR-5-E-OEMAP-U-0202                     Birla Institute of
Technology
227     IR-2-E-OE-U-0184               Indian Institute of Technology
Mandi

                   City              State  Score Rank    TLR    RPC
GO  \
219             Patiala            Punjab  56.14   20  69.77  45.75
73.11
220           Surathkal         Karnataka  53.16   21  60.34  38.36
71.27
221              Howrah       West Bengal  53.24  21A  67.15  45.47
61.09
222            Rupnagar            Punjab  52.80   22  77.84  29.53
65.29
223  Thiruvananthapuram            Kerala  52.74   23  78.87  20.76
62.12
224               Patna             Bihar  52.37   24  74.43  35.03
65.64
225            Warangal         Telangana  51.82   25  67.25  31.43
72.38
```

```
226              Ranchi         Jharkhand  51.12   26  71.22  36.61
59.88
227               Mandi  Himachal Pradesh  51.28  26A  76.90  30.48
63.07

        OI  Perception  year
219  56.55        12.01  2018
220  51.10        41.93  2018
221  40.27        32.03  2018
222  60.61        14.73  2018
223  60.82        43.47  2018
224  49.30        14.73  2018
225  54.79        22.59  2018
226  51.69        16.21  2018
227  54.64         9.96  2018
```

```
df_combined['Rank'] = df_combined['Rank'].apply(lambda x: x if
str(x).isdigit() else x[:-1])

df_combined['Rank'] = df_combined['Rank'].astype('float64')

df_combined[219:228]
```

```
            Institute Id                                        Institute
Name  \
219     IR-2-E-EM-I-1480     Thapar Institute of Engineering and
Technology
220     IR-2-E-OE-U-0237          National Institute of Technology
Surathkal
221     IR-2-E-OE-U-0584  Indian Institute of Engineering Science and
Te...
222     IR-2-E-OE-U-0378                Indian Institute of Technology
Ropar
223     IR-2-E-OE-U-0255   Indian Institute of Space Science and
Technology
224     IR-2-E-OE-U-0064                Indian Institute of Technology
Patna
225     IR-2-E-OE-U-0025          National Institute of Technology
Warangal
226  IR-5-E-OEMAP-U-0202                       Birla Institute of
Technology
227     IR-2-E-OE-U-0184                Indian Institute of Technology
Mandi

                   City              State  Score  Rank    TLR    RPC
GO  \
219             Patiala            Punjab  56.14  20.0  69.77  45.75
73.11
220           Surathkal         Karnataka  53.16  21.0  60.34  38.36
71.27
```

```
221              Howrah       West Bengal  53.24  21.0  67.15  45.47
61.09
222            Rupnagar            Punjab  52.80  22.0  77.84  29.53
65.29
223  Thiruvananthapuram            Kerala  52.74  23.0  78.87  20.76
62.12
224               Patna             Bihar  52.37  24.0  74.43  35.03
65.64
225            Warangal         Telangana  51.82  25.0  67.25  31.43
72.38
226              Ranchi         Jharkhand  51.12  26.0  71.22  36.61
59.88
227               Mandi  Himachal Pradesh  51.28  26.0  76.90  30.48
63.07

        OI  Perception  year
219  56.55        12.01  2018
220  51.10        41.93  2018
221  40.27        32.03  2018
222  60.61        14.73  2018
223  60.82        43.47  2018
224  49.30        14.73  2018
225  54.79        22.59  2018
226  51.69        16.21  2018
227  54.64         9.96  2018
```

```
df_combined.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 900 entries, 0 to 899
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Institute Id    900 non-null    object
 1   Institute Name  900 non-null    object
 2   City            900 non-null    object
 3   State           900 non-null    object
 4   Score           900 non-null    float64
 5   Rank            900 non-null    float64
 6   TLR             900 non-null    float64
 7   RPC             900 non-null    float64
 8   GO              900 non-null    float64
 9   OI              900 non-null    float64
 10  Perception      900 non-null    float64
 11  year            900 non-null    object
dtypes: float64(7), object(5)
memory usage: 84.5+ KB
```

The Institute name in 2016 is different from other year datasets so a operation is carried out to make it same.

```
data_2016.head(1)
```

```
       Institute Id                             Institute Name
City  \
0  NIRF-ENGG-INF-77  Indian Institute Of Technology, Madras  Chennai


        State  Score Rank    TLR    RPC     GO     OI  Perception
year
0  Tamil Nadu  89.41    1  88.26  94.02  81.81  86.11          98
2016
```

```
data_2017.head(1)
```

```
       Institute Id                             Institute Name     City  \
0  IR17-ENGG-1-1-77  Indian Institute of Technology Madras  Chennai

        State  Score  Rank    TLR   RPC     GO     OI  Perception
year
0  Tamil Nadu  87.96     1  91.85  92.6  83.78  77.19       81.46
2017
```

```
data_2016['Institute Name'] = data_2016['Institute
Name'].str.replace(',', '')
```

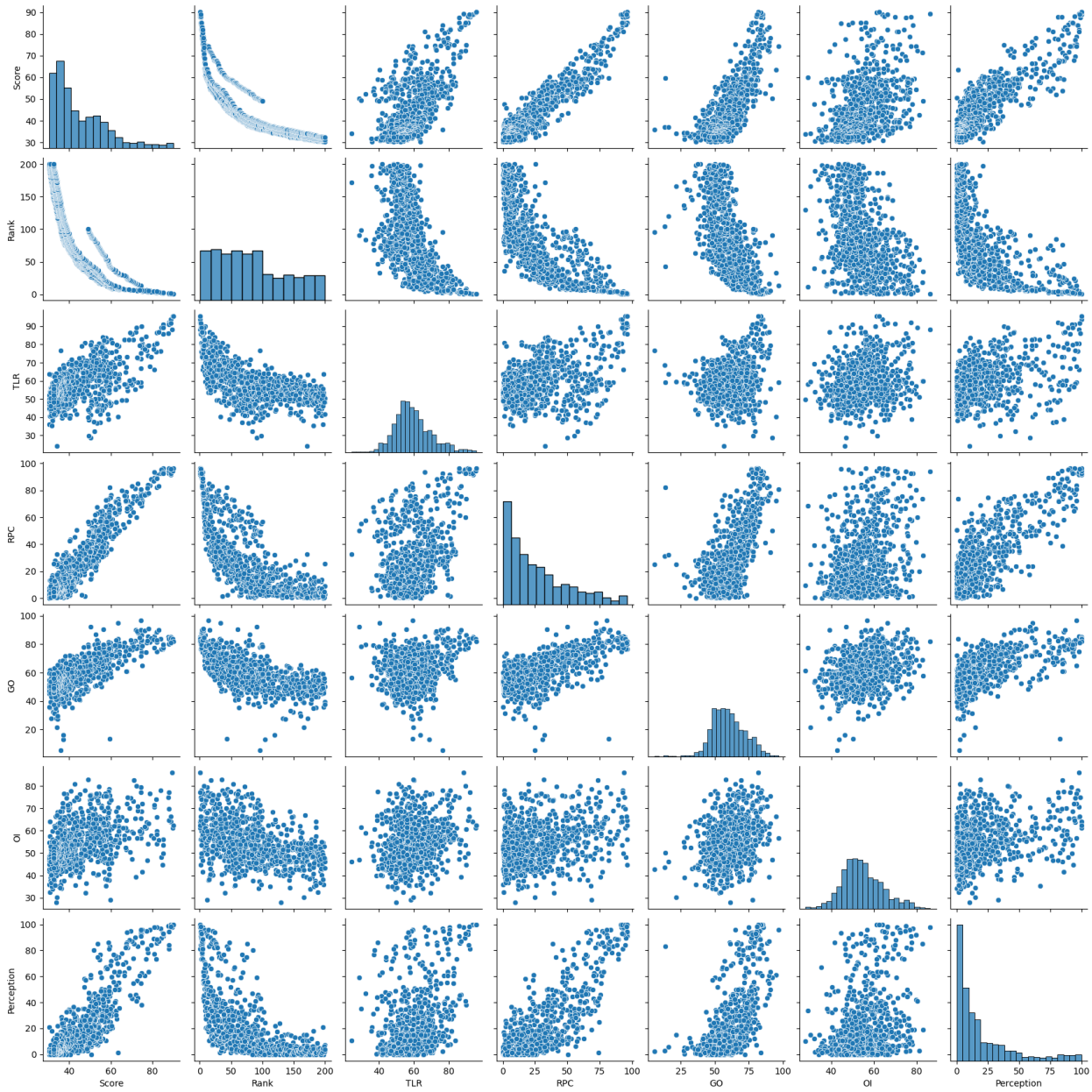
```
data_2016.head(1)
```

```
       Institute Id                             Institute Name     City  \
0  NIRF-ENGG-INF-77  Indian Institute Of Technology Madras  Chennai

        State  Score Rank    TLR    RPC     GO     OI  Perception
year
0  Tamil Nadu  89.41    1  88.26  94.02  81.81  86.11          98
2016
```

EXPLORATORY DATA ANALYSIS

```
sns.pairplot(df_combined)
```

```
C:\Users\ASUS\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)

<seaborn.axisgrid.PairGrid at 0x1bbeab8b350>
```



Mean RPC is low implies research sector is weak in clgs..

```
df_combined.describe()
```

```
             Score        Rank         TLR         RPC          GO
OI  \
count  900.000000  900.000000  900.000000  900.000000  900.000000
900.000000
mean    45.122317   83.700000   59.289333   26.604467   59.949256
54.611733
std     12.890893   55.335713   10.674734   23.832522   12.112018
9.368713
```

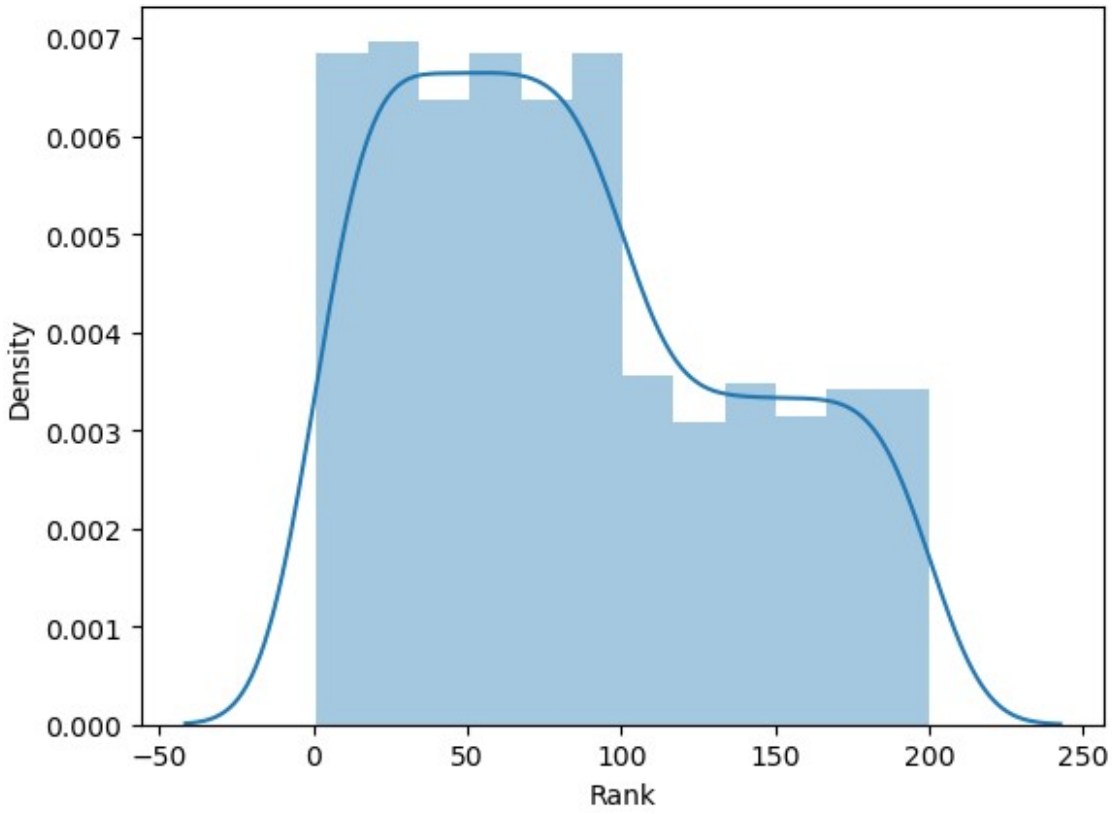
```
min     30.310000    1.000000   24.310000    0.170000    5.460000
28.020000
25%     35.465000   38.000000   52.340000    6.960000   51.617500
48.217500
50%     40.495000   75.000000   57.820000   19.365000   58.995000
53.505000
75%     52.755000  125.250000   64.847500   39.247500   67.667500
60.335000
max     90.190000  200.000000   95.470000   96.430000   96.800000
86.110000
```

```
       Perception
count  900.000000
mean    20.129267
std     24.133590
min      0.000000
25%      3.560000
50%     10.320000
75%     27.402500
max    100.000000
```

```
sns.distplot(df_combined['Rank'])
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_13276\3765976300.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df_combined['Rank'])
```

```
<Axes: xlabel='Rank', ylabel='Density'>
```




```
df_combined.columns
```

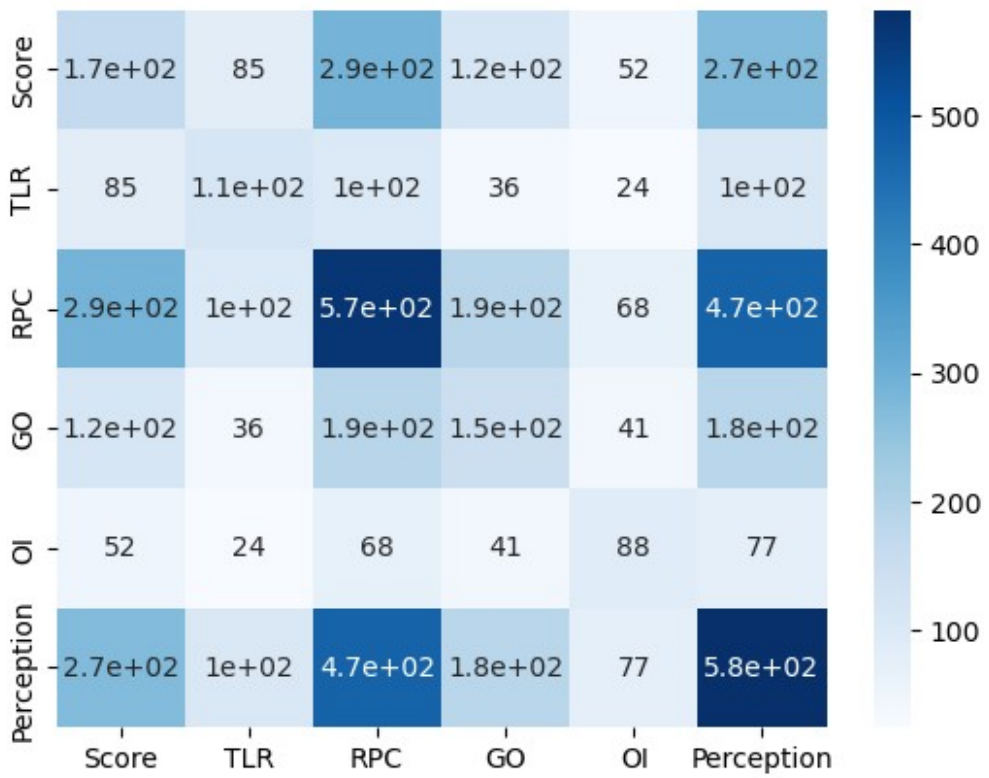
```
Index(['Institute Id', 'Institute Name', 'City', 'State', 'Score',
'Rank',
       'TLR', 'RPC', 'GO', 'OI', 'Perception', 'year'],
      dtype='object')
```

```
df_parameters = df_combined.drop(columns = ['Institute Id', 'Institute
Name', 'City', 'State' , 'year', 'Rank'])
```

```
df_parameters.head()
```

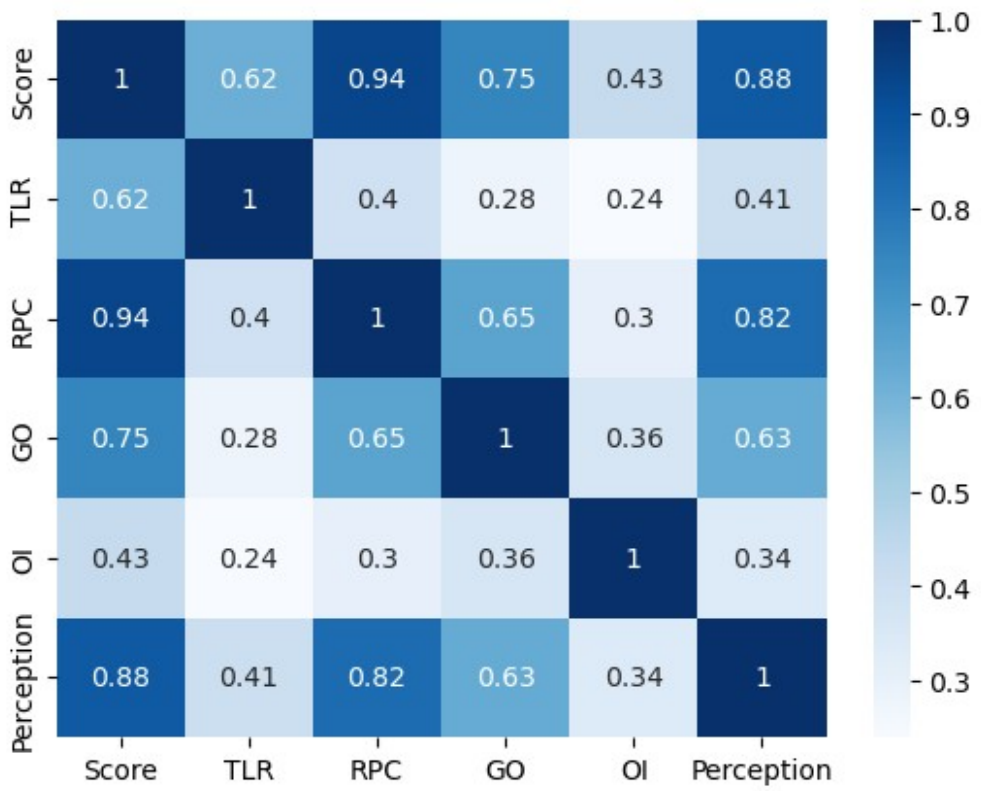
	Score	TLR	RPC	GO	OI	Perception
0	89.41	88.26	94.02	81.81	86.11	98.0
1	87.66	85.93	94.14	84.97	74.84	99.0
2	83.91	76.23	92.68	83.95	78.05	97.0
3	82.02	80.27	91.62	74.72	66.17	98.0
4	81.07	66.08	93.52	85.62	70.59	98.0

```
covmat = df_parameters.cov()
ax = sns.heatmap(covmat, annot = True , cmap = 'Blues')
plt.figure(figsize=(15,15))
plt.show()
```




```
<Figure size 1500x1500 with 0 Axes>
```

```
corrmat = df_parameters.corr()
ax = sns.heatmap(corrmat, annot = True , cmap = 'Blues')
plt.figure(figsize=(15,15))
plt.show()
```




```
<Figure size 1500x1500 with 0 Axes>
```

Here correlation of rpc with score is 0.94 it implies it greatly affects the overall score.

Liner Regressor

```
X = df_combined[['TLR' , 'RPC' , 'GO' , 'OI' , 'Perception']]
y = df_combined['Score']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=101)

from sklearn.linear_model import LinearRegression

lm = LinearRegression()

lm.fit(X_train,y_train)

LinearRegression()

# print the intercept
print(lm.intercept_)
```

```
-0.37743966713058086
```

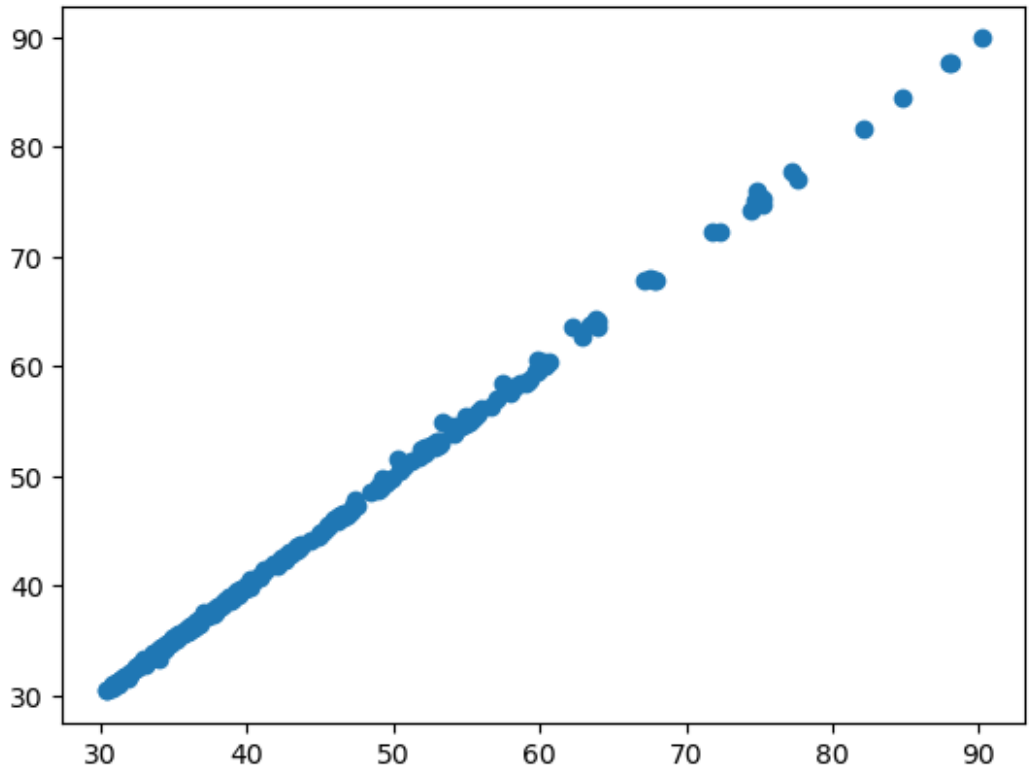
```
coeff_df = pd.DataFrame(lm.coef_,X.columns,columns=['Coefficient'])
coeff_df
```

	Coefficient
TLR	0.314024
RPC	0.294388
GO	0.191477
OI	0.102426
Perception	0.098394

```
predictions = lm.predict(X_test)
```

```
plt.scatter(y_test,predictions)
```

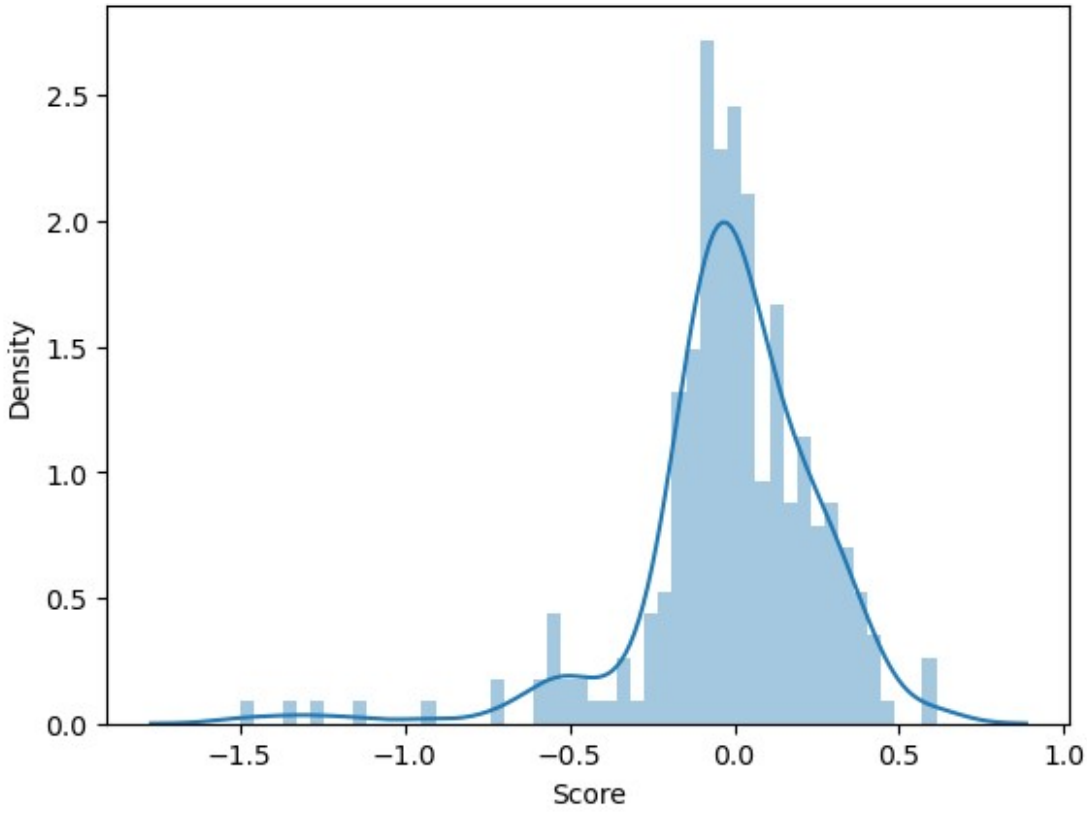
```
<matplotlib.collections.PathCollection at 0x1bbef772990>
```




```
sns.distplot((y_test-predictions),bins=50);
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_13276\1326397652.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.
```

```
Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot((y_test-predictions),bins=50);
```




```
from sklearn import metrics

print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test,
predictions)))

MAE: 0.18383221952155035
MSE: 0.07726423877762947
RMSE: 0.27796445596088265
```

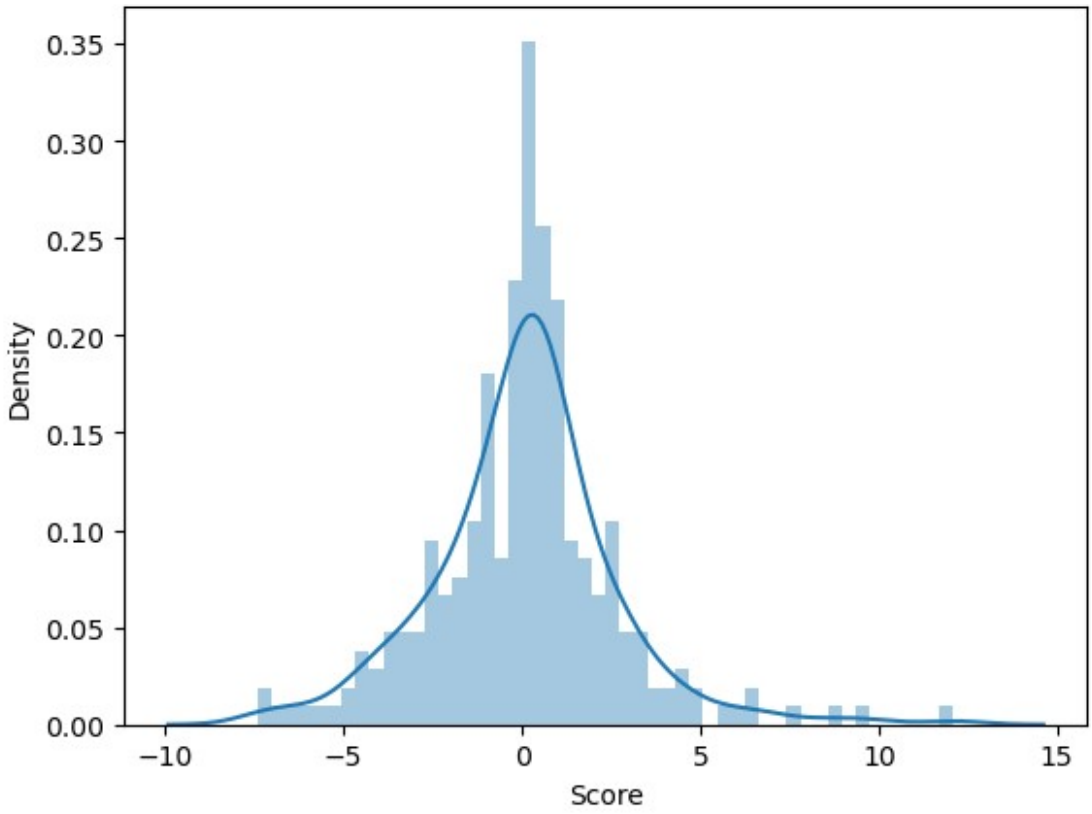
Decision Tree Regressor

```
from sklearn.tree import DecisionTreeRegressor

dtree = DecisionTreeRegressor()

dtree.fit(X_train,y_train)
```

```
DecisionTreeRegressor()
```

```
pred1 = dtree.predict(X_test)
sns.distplot((y_test-pred1),bins=50);
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_13276\3993809026.py:2:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot((y_test-pred1),bins=50);
```




```
print('MAE:', metrics.mean_absolute_error(y_test, pred1))
print('MSE:', metrics.mean_squared_error(y_test, pred1))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, pred1)))
```

```
MAE: 1.7854444444444442
MSE: 6.59323
RMSE: 2.5677285682096542
```

Random Forest Regressor

```
from sklearn.ensemble import RandomForestRegressor
rfc = RandomForestRegressor(n_estimators=100)
rfc.fit(X_train, y_train)
```

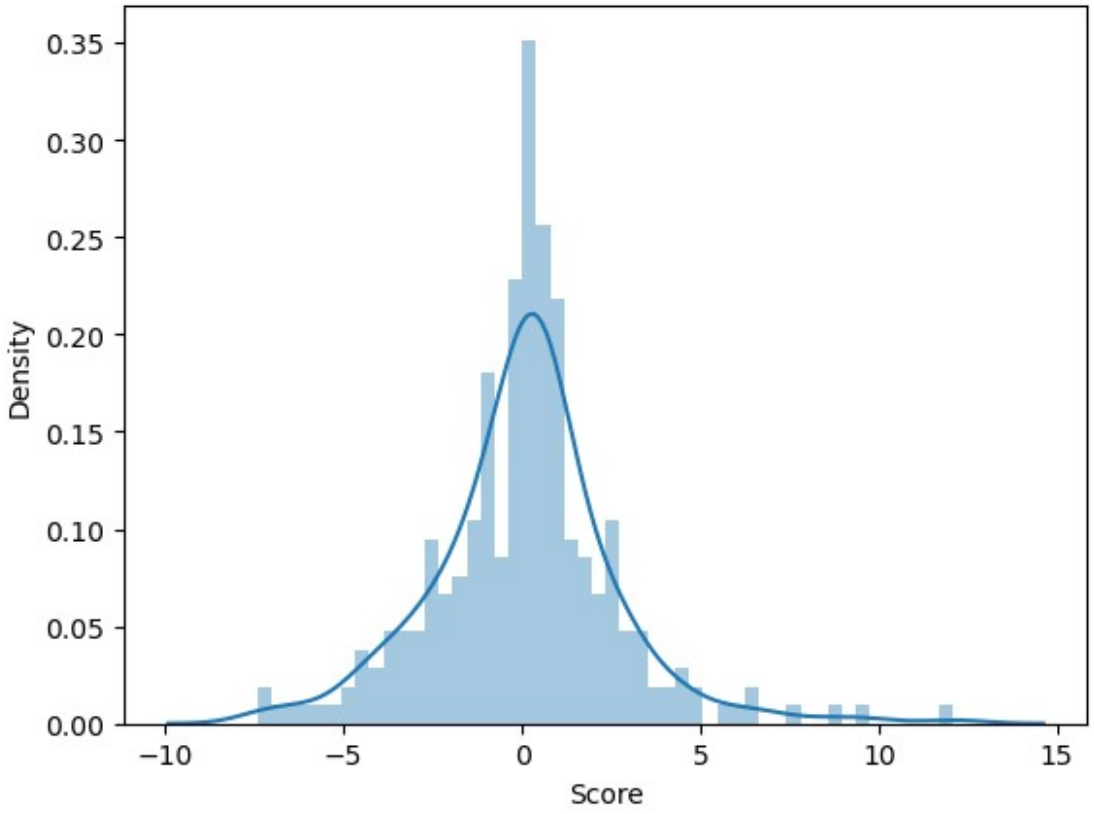
```
RandomForestRegressor()
```

```
rfc_pred = rfc.predict(X_test)
```

```
sns.distplot((y_test-pred1),bins=50);
```

```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_13276\172980259.py:1:
UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot((y_test-pred1),bins=50);
```

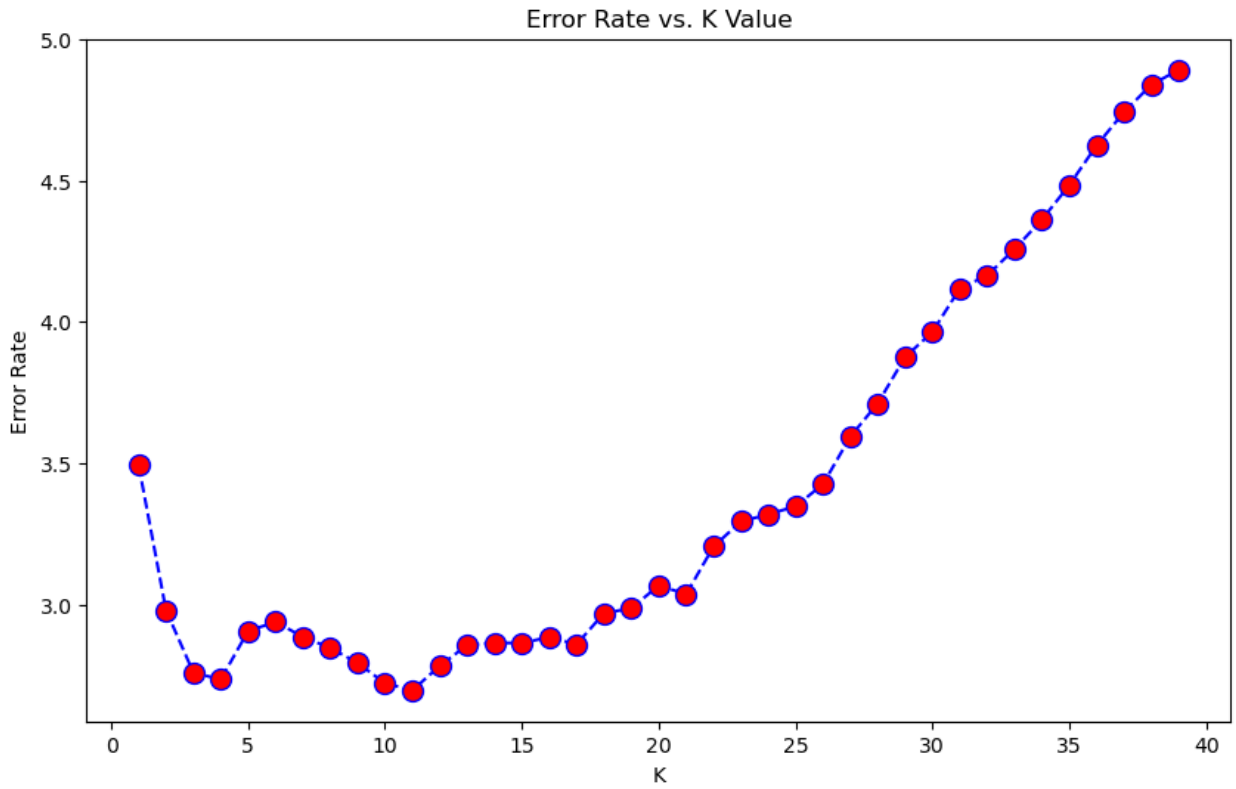



```
print('MAE:', metrics.mean_absolute_error(y_test, rfc_pred))
print('MSE:', metrics.mean_squared_error(y_test, rfc_pred))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, rfc_pred)))

MAE: 1.0658215925925947
MSE: 2.530237056329273
RMSE: 1.5906718883318687
```

KNN Regressor

```
from sklearn.neighbors import KNeighborsRegressor

error_rate = []

for i in range(1,40):

    knn = KNeighborsRegressor(n_neighbors=i)
    knn.fit(X_train,y_train)
    pred_i = knn.predict(X_test)
    exp_i = list(y_test)
    error_rate.append((np.square(np.subtract(pred_i, exp_i))).mean())

plt.figure(figsize=(10,6))
plt.plot(range(1,40),error_rate,color='blue', linestyle='dashed',
marker='o',
         markerfacecolor='red', markersize=10)
plt.title('Error Rate vs. K Value')
plt.xlabel('K')
plt.ylabel('Error Rate')
```

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

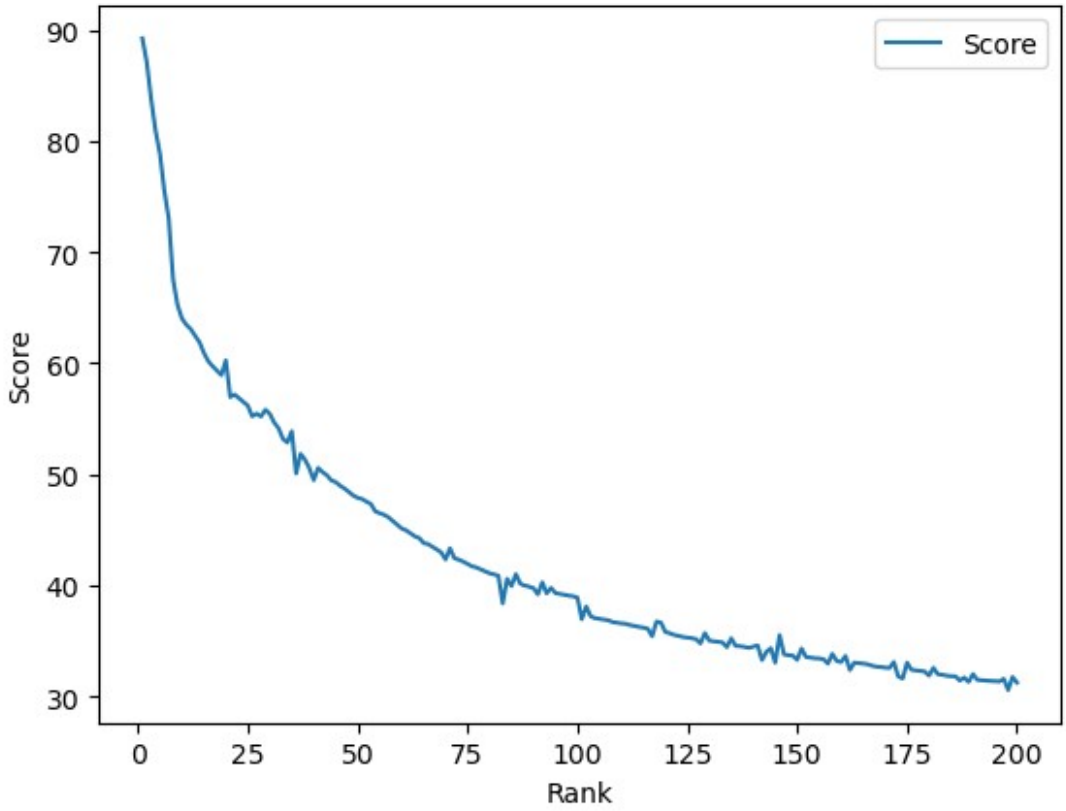


```
# for k =19
knn = KNeighborsRegressor(n_neighbors=19)

knn.fit(X_train,y_train)
pred2 = knn.predict(X_test)

print('WITH K=19')
print('\n')
print('MAE:', metrics.mean_absolute_error(y_test, pred2))
print('MSE:', metrics.mean_squared_error(y_test, pred2))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, pred2)))
```

```
WITH K=19


MAE: 1.1800072124756336
MSE: 2.9860027697240166
RMSE: 1.7280054310458681
```

```
df_score_rank = df_combined[['Score' , 'Rank']]
df_rel_score_rank = df_score_rank.groupby('Rank').mean()
df_rel_score_rank.plot(kind = 'line' , ylabel = 'Score')
```

```
<Axes: xlabel='Rank', ylabel='Score'>
```



Here it is shown the relation between Overall Score and the Rank of the institute.

Comparison of Regression Models

```
test_values = [83.50, 89.80, 78.00, 65.50, 94.0]

arr = np.asarray(test_values)
y_predict = lm.predict(arr.reshape(1,-1))
print('Predicted score: ', y_predict[0])
rank_score = df_rel_score_rank.values     # extract the ranks and
scores from the dataframe

pos, value = min(enumerate(rank_score), key=lambda x:
y_predict[0]<=x[1]) # to predict the rank by interpolation

print('Predicted rank: ',pos+1)
```

```
Predicted score:  83.17274823359435
Predicted rank:  4

C:\Users\ASUS\anaconda3\Lib\site-packages\sklearn\base.py:464:
UserWarning: X does not have valid feature names, but LinearRegression
was fitted with feature names
  warnings.warn(
```

```
arr = np.asarray(test_values)
y_predict = dtree.predict(arr.reshape(1,-1))
print('Predicted score: ', y_predict[0])

pos, value = min(enumerate(rank_score), key=lambda x:
y_predict[0]<=x[1]) # to predict the rank by interpolation

print('Predicted rank: ',pos+1)
```

```
Predicted score:  87.66
Predicted rank:  2

C:\Users\ASUS\anaconda3\Lib\site-packages\sklearn\base.py:464:
UserWarning: X does not have valid feature names, but
DecisionTreeRegressor was fitted with feature names
  warnings.warn(
```

```
arr = np.asarray(test_values)
y_predict = rfc.predict(arr.reshape(1,-1))
print('Predicted score: ', y_predict[0])

pos, value = min(enumerate(rank_score), key=lambda x:
y_predict[0]<=x[1]) # to predict the rank by interpolation

print('Predicted rank: ',pos+1)
```

```
Predicted score:  84.6031
Predicted rank:  3

C:\Users\ASUS\anaconda3\Lib\site-packages\sklearn\base.py:464:
UserWarning: X does not have valid feature names, but
RandomForestRegressor was fitted with feature names
  warnings.warn(
```

```
arr = np.asarray(test_values)
y_predict = knn.predict(arr.reshape(1,-1))
print('Predicted score: ', y_predict[0])

pos, value = min(enumerate(rank_score), key=lambda x:
y_predict[0]<=x[1]) # to predict the rank by interpolation

print('Predicted rank: ',pos+1)
```

```
Predicted score:  84.28684210526318
Predicted rank:  3

C:\Users\ASUS\anaconda3\Lib\site-packages\sklearn\base.py:464:
UserWarning: X does not have valid feature names, but
KNeighborsRegressor was fitted with feature names
  warnings.warn(
```

As from above we can comprehend that the Mean absolute error(MAE) and Root mean squared error(RMSE) of Linear Regressor is least , so it is best suited model for prediction of rank.

Final NIRF Rank predictor model

```
test_values = [83.50, 89.80, 78.00, 65.50, 94.0]
arr = np.asarray(test_values)
y_predict = lm.predict(arr.reshape(1,-1))
print('Predicted score: ', y_predict[0])
rank_score = df_rel_score_rank.values     # extract the ranks and
scores from the dataframe

pos, value = min(enumerate(rank_score), key=lambda x:
y_predict[0]<=x[1]) # to predict the rank by interpolation

print('Predicted rank: ',pos+1)
```

```
Predicted score:  83.17274823359435
Predicted rank:  4

C:\Users\ASUS\anaconda3\Lib\site-packages\sklearn\base.py:464:
UserWarning: X does not have valid feature names, but LinearRegression
was fitted with feature names
  warnings.warn(
```