```python
# -*- coding: utf-8 -*-
"""

Created on Thu Jun 28 16:51:21 2018


@author: shubham b thorat
"""

# importing the libraries
"""

university ranking prediction
"""

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np


# importing the data set
dataset=pd.read_csv('university_ranking.csv')
X = dataset.iloc[:,1:13].values
Y = dataset.iloc[:,0].values


# taking care of missing data
from sklearn.preprocessing import Imputer
imputer = Imputer(missing_values="NaN", strategy="mean",axis=0)
imputer = imputer.fit(X[:,2:13], y=None)
X[:,2:14] = imputer.transform(X[:,2:14])


# handling categorical (encoding) data
from sklearn.preprocessing import LabelEncoder
```

```python
label_encoder_X = LabelEncoder()

label_encoder_Y = LabelEncoder()

X[:,0] = label_encoder_X.fit_transform(X[:,0])

X[:,1] = label_encoder_X.fit_transform(X[:,1])

#onehotencoder = OneHotEncoder(categorical_features= [0])

#onehotencoder = OneHotEncoder(categorical_features= [1])

#X = onehotencoder.fit_transform(X).toarray()


# dividing dataset into test and training dataset

from sklearn.model_selection import train_test_split

X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = .3, random_state = 0)


# feature scaling

from sklearn.preprocessing import StandardScaler

sc_X = StandardScaler()

X_train = sc_X.fit_transform(X_train)

X_test = sc_X.fit_transform(X_test)


from sklearn.linear_model import LinearRegression

regressor = LinearRegression()

regressor.fit(X_train,Y_train)


#predicting the test set result

y_pred = regressor.predict(X_test)


""""

Y_test =
```
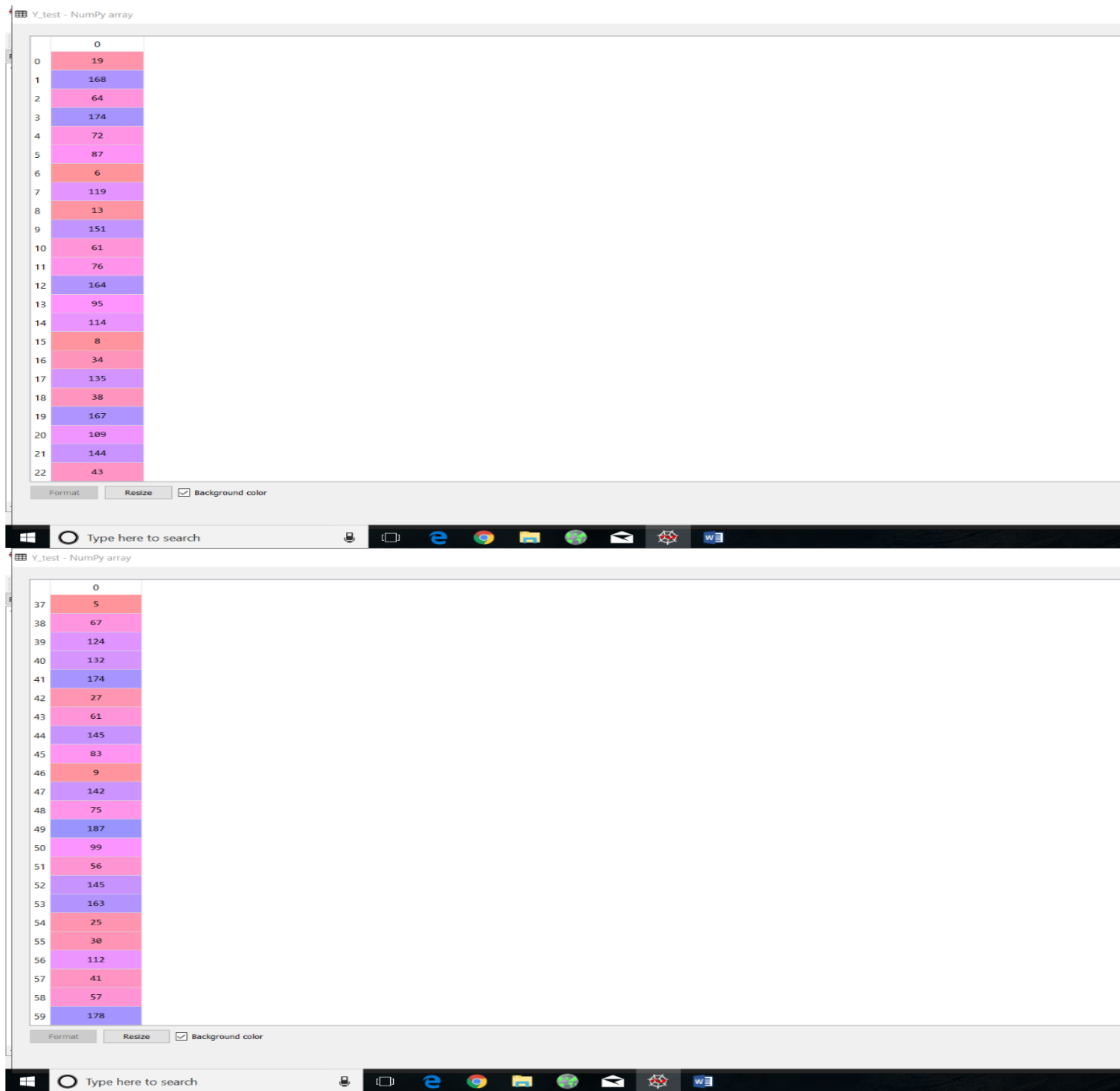
	0
0	19
1	168
2	64
3	174
4	72
5	87
6	6
7	119
8	13
9	151
10	61
11	76
12	164
13	95
14	114
15	8
16	34
17	135
18	38
19	167
20	109
21	144
22	43

Format | Resize | ☑ Background color

	0
37	5
38	67
39	124
40	132
41	174
42	27
43	61
44	145
45	83
46	9
47	142
48	75
49	187
50	99
51	56
52	145
53	163
54	25
55	30
56	112
57	41
58	57
59	178

Format | Resize | ☑ Background color

Y_pred =

y_pred - NumPy array

	0
0	29.353
1	148.13
2	88.6812
3	145.84
4	96.3923
5	116.832
6	-3.41295
7	157.993
8	1.19892
9	177.296
10	83.7012
11	81.7355
12	171.992
13	92.4178
14	133.634
15	-18.767
16	55.8571
17	108.298
18	39.0305
19	146.646
20	137.232
21	153.787
22	61.3178

Format   Resize   ☑ Background color

y_pred - NumPy array

	0
37	-32.8702
38	78.5632
39	144.956
40	157.021
41	154.862
42	44.2373
43	93.0462
44	142.932
45	105.502
46	2.55602
47	97.9081
48	69.4454
49	137.992
50	101.206
51	95.47
52	103.994
53	155.778
54	21.5536
55	51.8203
56	156.661
57	41.7829
58	129.915
59	168.514

Format   Resize   ☑ Background color

"""