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

from sklearn.preprocessing import MinMaxScaler

from sklearn.model_selection import train_test_split

from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error

from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression

%matplotlib inline

pd.set_option('display.max_columns', None)
```

In [2]:
```
scores=pd.read_csv("2013-2017_School_Math_Results_-_All.csv",na_values="s")
```

In [3]:
```
scores
```

Out[3]:

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
0	01M015	PS 015 ROBERTO CLEMENTE	3	2013	All Students	27	278.0	16.0	59.3
1	01M015	PS 015 ROBERTO CLEMENTE	3	2014	All Students	18	286.0	6.0	33.3
2	01M015	PS 015 ROBERTO CLEMENTE	3	2015	All Students	17	280.0	10.0	58.8
3	01M015	PS 015 ROBERTO CLEMENTE	3	2016	All Students	21	275.0	13.0	61.9
4	01M015	PS 015 ROBERTO CLEMENTE	3	2017	All Students	29	302.0	8.0	27.6
5	01M015	PS 015 ROBERTO CLEMENTE	4	2013	All Students	20	277.0	12.0	60.0
6	01M015	PS 015 ROBERTO CLEMENTE	4	2014	All Students	24	282.0	9.0	37.5
7	01M015	PS 015 ROBERTO CLEMENTE	4	2015	All Students	17	281.0	8.0	47.1
8	01M015	PS 015 ROBERTO CLEMENTE	4	2016	All Students	15	304.0	3.0	20.0
9	01M015	PS 015 ROBERTO CLEMENTE	4	2017	All Students	23	301.0	6.0	26.1
10	01M015	PS 015 ROBERTO CLEMENTE	5	2013	All Students	24	274.0	17.0	70.8
11	01M015	PS 015 ROBERTO CLEMENTE	5	2014	All Students	21	266.0	17.0	81.0

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
12	01M015	PS 015 ROBERTO CLEMENTE	5	2015	All Students	24	275.0	16.0	66.7
13	01M015	PS 015 ROBERTO CLEMENTE	5	2016	All Students	16	283.0	7.0	43.8
14	01M015	PS 015 ROBERTO CLEMENTE	5	2017	All Students	17	322.0	2.0	11.8
15	01M015	PS 015 ROBERTO CLEMENTE	All Grades	2013	All Students	71	276.0	45.0	63.4
16	01M015	PS 015 ROBERTO CLEMENTE	All Grades	2014	All Students	63	278.0	32.0	50.8
17	01M015	PS 015 ROBERTO CLEMENTE	All Grades	2015	All Students	58	278.0	34.0	58.6
18	01M015	PS 015 ROBERTO CLEMENTE	All Grades	2016	All Students	52	286.0	23.0	44.2
19	01M015	PS 015 ROBERTO CLEMENTE	All Grades	2017	All Students	69	307.0	16.0	23.2
20	01M019	PS 019 ASHER LEVY	3	2013	All Students	36	311.0	8.0	22.2
21	01M019	PS 019 ASHER LEVY	3	2014	All Students	35	311.0	8.0	22.9
22	01M019	PS 019 ASHER LEVY	3	2015	All Students	28	303.0	6.0	21.4
23	01M019	PS 019 ASHER LEVY	3	2016	All Students	33	316.0	5.0	15.2
24	01M019	PS 019 ASHER LEVY	3	2017	All Students	28	313.0	7.0	25.0
25	01M019	PS 019 ASHER LEVY	4	2013	All Students	42	289.0	18.0	42.9
26	01M019	PS 019 ASHER LEVY	4	2014	All Students	34	316.0	7.0	20.6

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
27	01M019	PS 019 ASHER LEVY	4	2015	All Students	32	322.0	4.0	12.5
28	01M019	PS 019 ASHER LEVY	4	2016	All Students	28	315.0	4.0	14.3
29	01M019	PS 019 ASHER LEVY	4	2017	All Students	29	324.0	3.0	10.3
...	...	...	...	...	...	...	...	...	...
23866	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	6	2016	All Students	60	341.0	1.0	1.7
23867	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	6	2017	All Students	58	336.0	1.0	1.7
23868	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	7	2013	All Students	59	330.0	2.0	3.4
23869	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	7	2014	All Students	60	328.0	1.0	1.7
23870	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	7	2015	All Students	59	334.0	4.0	6.8
23871	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	7	2016	All Students	60	338.0	1.0	1.7
23872	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	7	2017	All Students	59	346.0	0.0	0.0
23873	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	8	2013	All Students	35	322.0	2.0	5.7

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
23874	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	All Grades	2013	All Students	155	325.0	8.0	5.2
23875	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	All Grades	2014	All Students	118	330.0	2.0	1.7
23876	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	All Grades	2015	All Students	119	337.0	4.0	3.4
23877	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	All Grades	2016	All Students	120	339.0	2.0	1.7
23878	32K554	ALL CITY LEADERSHIP SECONDARY SCHOOL	All Grades	2017	All Students	117	341.0	1.0	0.9
23879	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	6	2013	All Students	145	266.0	107.0	73.8
23880	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	6	2014	All Students	107	264.0	78.0	72.9
23881	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	6	2015	All Students	98	262.0	76.0	77.6
23882	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	6	2016	All Students	89	269.0	54.0	60.7

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
23883	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	6	2017	All Students	112	265.0	78.0	69.6
23884	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	7	2014	All Students	124	274.0	84.0	67.7
23885	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	7	2015	All Students	98	274.0	76.0	77.6
23886	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	7	2016	All Students	89	268.0	67.0	75.3
23887	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	7	2017	All Students	94	275.0	69.0	73.4
23888	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	8	2015	All Students	128	268.0	97.0	75.8
23889	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	8	2016	All Students	91	274.0	55.0	60.4
23890	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	8	2017	All Students	88	268.0	59.0	67.0

	DBN	School Name	Grade	Year	Category	Number Tested	Mean Scale Score	Level1_N	Level
23891	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	All Grades	2013	All Students	145	266.0	107.0	73.8
23892	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	All Grades	2014	All Students	231	269.0	162.0	70.1
23893	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	All Grades	2015	All Students	324	268.0	249.0	76.9
23894	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	All Grades	2016	All Students	269	270.0	176.0	65.4
23895	32K562	EVERGREEN MIDDLE SCHOOL FOR URBAN EXPLORATION	All Grades	2017	All Students	294	269.0	206.0	70.1

23896 rows × 17 columns

In [4]:
```
scores.dropna(inplace=True)
```

In [5]:
```
all_grades_filter=scores["Grade"]=="All Grades"
```

In [6]:
```
scores2=scores[~all_grades_filter]
```

In [7]:
```
x=scores2[["Grade"]]
y=scores2[["Mean Scale Score"]]
```

In [8]:
```
x.head()
```

Out[8]:

	Grade
0	3
1	3
2	3
3	3
4	3

In [9]:
```
y.head()
```

Out[9]:

	Mean Scale Score
0	278.0
1	286.0
2	280.0
3	275.0
4	302.0

In [10]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

In [11]:
```
knn = KNeighborsRegressor(n_neighbors = 10)
knn.fit(x_train, y_train)
```

Out[11]:
```
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
          metric_params=None, n_jobs=None, n_neighbors=10, p=2,
          weights='uniform')
```

In [12]:
```
y_test_preds=knn.predict(x_test)
```

In [13]:
```
mean_squared_error(y_test_preds,y_test)
```

Out[13]:
```
463.72891197375617
```

In [14]:
```
y_train_preds=knn.predict(x_train)
```

In [15]:
```
mean_squared_error(y_train_preds, y_train)
```

Out[15]:
```
467.3111098961181
```

In [16]:
```
x2=scores2[["Level1_N","Level1_%"]]
y2=scores2[["Level2_N"]]
```

In [17]:
```
x2.head()
```

Out[17]:

	Level1_N	Level1_%
0	16.0	59.3
1	6.0	33.3
2	10.0	58.8
3	13.0	61.9
4	8.0	27.6

In [18]:
```
y2.head()
```

Out[18]:

	Level2_N
0	11.0
1	9.0
2	4.0
3	4.0
4	9.0

In [19]:
```
x2_train,x2_test,y2_train,y2_test=train_test_split(x2,y2,test_size=0.2)
```

In [20]:
```
knn2 = KNeighborsRegressor(n_neighbors = 10)
knn2.fit(x2_train, y2_train)
```

Out[20]:
```
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
          metric_params=None, n_jobs=None, n_neighbors=10, p=2,
          weights='uniform')
```

In [21]:
```
y_test_preds2=knn2.predict(x2_test)
```

In [22]:
```
mean_squared_error(y_test_preds2,y2_test)
```

Out[22]:
```
110.07492072170585
```

In [23]:
```
y_train_preds2=knn2.predict(x2_train)
```

In [24]:
```
mean_squared_error(y_train_preds2, y2_train)
```

Out[24]:
```
81.72234691088026
```

In [25]:
```
polynomial_features= PolynomialFeatures(degree=2)
x_poly = polynomial_features.fit_transform(scores2[["Level3_N"]])
```

In [26]:
```
model = LinearRegression()
```

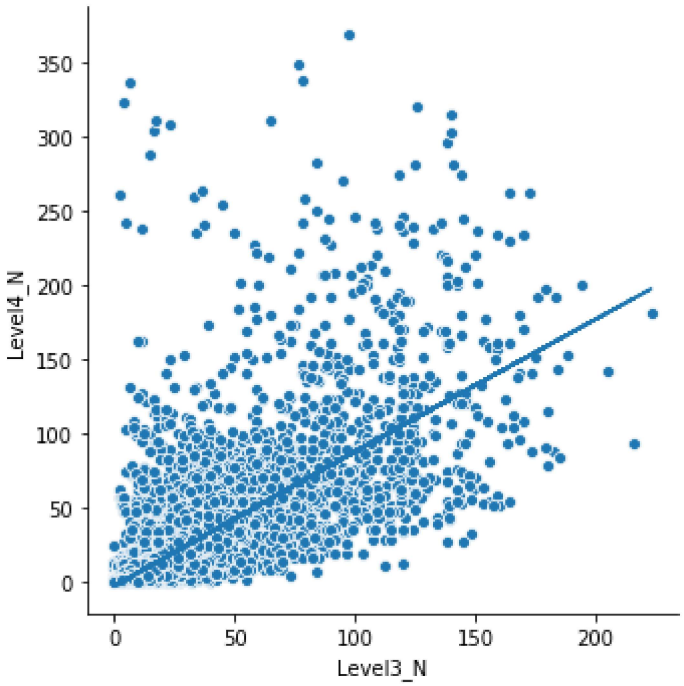
In [27]:
```
model.fit(x_poly, scores2["Level4_N"])
```

Out[27]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [28]:
```
y_pred = model.predict(x_poly)
```

In [29]:
```
sns.relplot(x="Level3_N",y="Level4_N",data=scores2)
plt.plot(scores2["Level3_N"],y_pred)
```

Out[29]:
```
[<matplotlib.lines.Line2D at 0x7ff6327af588>]
```



In [30]:
```
polynomial_features2= PolynomialFeatures(degree=2)
x_poly2 = polynomial_features2.fit_transform(scores2[["Level3+4_N"]])
```

In [31]:
```
model2 = LinearRegression()
```

In [32]:
```
model2.fit(x_poly2, scores2["Level3+4_%"])
```

Out[32]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
```

In [33]:
```
y_pred2 = model.predict(x_poly2)
```

In [34]:
```
sns.relplot(x="Level3+4_N",y="Level3+4_%",data=scores2)
plt.plot(scores2["Level3+4_N"],y_pred2)
```

Out[34]: [<matplotlib.lines.Line2D at 0x7ff633d68748>]