# ИУ5-62Б Ковалев Сергей РК2

# Импорт библиотек

CAPITAL OUTLAY EXPENDITURE

GRADES PK G

```
In [1]:
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import warnings
warnings.filterwarnings('ignore')
sns.set(style="ticks")
%matplotlib inline
from sklearn.model selection import train test split
from sklearn.preprocessing import LabelEncoder
In [2]:
data = pd.read csv('/Users/set27/Downloads/states all.csv')
In [3]:
data.head()
Out[3]:
                      STATE YEAR ENROLL TOTAL_REVENUE FEDERAL_REVENUE STATE_REVENUE LOCAL_REVEN
     PRIMARY_KEY
    1992_ALABAMA
                    ALABAMA
                              1992
                                      NaN
                                                 2678885.0
                                                                                                71568
0
                                                                  304177.0
                                                                                1659028.0
1
      1992_ALASKA
                     ALASKA
                              1992
                                      NaN
                                                 1049591.0
                                                                  106780.0
                                                                                 720711.0
                                                                                                22210
2
     1992_ARIZONA
                    ARIZONA
                              1992
                                                 3258079.0
                                                                  297888.0
                                                                                1369815.0
                                                                                               159037
                                      NaN
   1992 ARKANSAS
                              1992
                                      NaN
                                                 1711959.0
                                                                  178571.0
                                                                                 958785.0
                                                                                                57460
                   ARKANSAS
  1992_CALIFORNIA CALIFORNIA
                                                                  2072470.0
                                                                               16546514.0
                              1992
                                      NaN
                                                26260025.0
                                                                                               764104
5 rows × 25 columns
In [4]:
data = data.fillna(1)
In [5]:
data.dtypes
Out[5]:
PRIMARY KEY
                                    object
STATE
                                    object
YEAR
                                     int64
                                   float64
ENROLL
TOTAL REVENUE
                                   float64
FEDERAL REVENUE
                                   float64
STATE REVENUE
                                   float64
LOCAL REVENUE
                                   float64
TOTAL EXPENDITURE
                                   float64
INSTRUCTION EXPENDITURE
                                   float64
SUPPORT SERVICES EXPENDITURE
                                   float64
OTHER EXPENDITURE
                                   float64
```

float64

float64

```
GRADES KG G
                                 float64
GRADES 4 G
                                float64
GRADES 8 G
                                float64
GRADES 12 G
                                float64
GRADES 1 8 G
                                float64
GRADES 9 12 G
                                float64
GRADES ALL G
                               float64
AVG MATH 4 SCORE
                               float64
AVG MATH 8 SCORE
                                float64
                                float64
AVG READING 4 SCORE
AVG READING 8 SCORE
                                float64
dtype: object
In [6]:
data.isnull().sum()
# проверим есть ли пропущенные значения
Out[6]:
PRIMARY KEY
                                 0
                                 0
STATE
                                 0
YEAR
ENROLL
                                 0
                                 0
TOTAL REVENUE
                                 0
FEDERAL REVENUE
STATE REVENUE
                                 0
LOCAL_REVENUE
                                 0
TOTAL EXPENDITURE
                                 0
INSTRUCTION EXPENDITURE
                                0
                                0
SUPPORT SERVICES EXPENDITURE
OTHER EXPENDITURE
                                0
CAPITAL OUTLAY EXPENDITURE
                                0
GRADES PK G
                                 0
GRADES KG G
                                 0
                                 0
GRADES 4 G
GRADES 8 G
                                 0
GRADES 12_G
                                 0
GRADES 1 8 G
                                 0
GRADES 9 12 G
                                 0
GRADES ALL G
                                 0
AVG MATH 4 SCORE
                                 0
AVG_MATH_8_SCORE
                                0
AVG_READING_4_SCORE
                                 0
AVG_READING_8_SCORE
dtype: int64
In [7]:
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1715 entries, 0 to 1714
Data columns (total 25 columns):
                                    Non-Null Count Dtype
 # Column
                                    1715 non-null object
 0 PRIMARY_KEY
   STATE
                                   1715 non-null object
 1
                                   1715 non-null int64
   YEAR
                                   1715 non-null float64
 3
   ENROLL
   TOTAL REVENUE
                                   1715 non-null
                                                    float64
    FEDERAL_REVENUE
 5
                                   1715 non-null
                                                  float64
    STATE REVENUE
 6
                                  1715 non-null float64
 7
    LOCAL_REVENUE
                                   1715 non-null float64
 8 TOTAL_EXPENDITURE 1715 non-null float64
9 INSTRUCTION_EXPENDITURE 1715 non-null float64
 10 SUPPORT_SERVICES_EXPENDITURE 1715 non-null float64
                         1715 non-null float64
```

1715 non-null float64

1715 non-null float64

1715 non-null float64

11 OTHER EXPENDITURE

13 GRADES PK G

14 GRADES KG G

15 GRADES 4 G

12 CAPITAL OUTLAY EXPENDITURE 1715 non-null float64

```
18 GRADES 1 8 G
                                                        float64
                                       1715 non-null
 19 GRADES 9 12 G
                                       1715 non-null
                                                        float64
 20 GRADES_ALL_G
                                       1715 non-null
                                                        float64
 21 AVG MATH_4_SCORE
                                       1715 non-null
                                                        float64
 22 AVG_MATH_8_SCORE
                                      1715 non-null
                                                        float64
 23 AVG_READING_4_SCORE
                                      1715 non-null
                                                        float64
 24 AVG_READING_8_SCORE
                                      1715 non-null
                                                         float64
dtypes: float64(22), int64(1), object(2)
memory usage: 335.1+ KB
In [8]:
data.head()
Out[8]:
                       STATE YEAR ENROLL TOTAL_REVENUE FEDERAL_REVENUE STATE_REVENUE LOCAL_REVEN
     PRIMARY_KEY
    1992_ALABAMA
                    ALABAMA
                              1992
                                                 2678885.0
                                                                                                 71568
0
                                       1.0
                                                                   304177.0
                                                                                 1659028.0
1
      1992 ALASKA
                     ALASKA
                              1992
                                       1.0
                                                 1049591.0
                                                                   106780.0
                                                                                  720711.0
                                                                                                 22210
2
     1992_ARIZONA
                              1992
                                                 3258079.0
                    ARIZONA
                                       1.0
                                                                   297888.0
                                                                                 1369815.0
                                                                                                159037
   1992_ARKANSAS ARKANSAS
                              1992
                                       1.0
                                                 1711959.0
                                                                   178571.0
                                                                                  958785.0
                                                                                                 57460
  1992_CALIFORNIA CALIFORNIA
                              1992
                                       1.0
                                                26260025.0
                                                                  2072470.0
                                                                                16546514.0
                                                                                                764104
5 rows × 25 columns
In [9]:
parts = np.split(data, [1,17,18], axis=1)
X = parts[0]
Y = parts[1]
G = parts[2]
print('Входные данные:\n\n', X.head(), '\n\nВыходные данные:\n\n', G.head())
Входные данные:
        PRIMARY KEY
      1992 ALABAMA
0
       1992 ALASKA
1
2
      1992 ARIZONA
3
     1992 ARKANSAS
  1992 CALIFORNIA
Выходные данные:
   GRADES 12 G
0
       41167
1
         6714
2
        37410
3
        27651
       270675
In [10]:
#Построим корреляционную матрицу
fig, ax = plt.subplots(figsize=(15,7))
sns.heatmap(data.corr(method='pearson'), ax=ax, annot=True, fmt='.2f')
Out[10]:
<AxesSubplot:>
                  YEAR - 1.00 0.19 0.27 0.28 0.26 0.26 0.26 0.26 0.28 0.27 0.21 0.21 0.01 0.00 0.02 0.06 0.27 0.32 0.01
```

1.00 0.92 0.90 0.92 0.86 0.92 0.89 0.92 0.96 0.92

0.92 1.00 0.94 0.98 0.97 1.00 0.99 1.00 0.95 0.93

TOTAL\_REVENUE - 0.27

0.73 0.83 0.84 0.86 0.85 0.91 0.87 0.85

0.68 0.77 0.79 0.81 0.83 0.90 0.88 0.80 0.08 0.05 0.07 0.09

1715 non-null

1715 non-null

float64

float64

16 GRADES 8 G

17 GRADES 12 G

- 0.8

0.01 -0.02 0.03 0.05

```
LUCAL REVENUE
              TOTAL_EXPENDITURE - 0.26
                                        0.92 1.00 0.94 0.97 0.97 1.00 0.99 0.99 0.95 0.94 0.69 0.77 0.79 0.81 0.83 0.90 0.88 0.80 0.08 0.05 0.07 0.09
       INSTRUCTION_EXPENDITURE - 0.26
                                        0.89 0.99 0.91 0.96 0.97 0.99 1.00 0.98 0.92 0.91 0.65 0.75
                                                                                                0.76 0.78 0.80 0.87 0.85 0.77
                                                                                                                            0.08 0.05 0.07 0.09
                                        0.92 1.00 0.94 0.97 0.96 0.99 0.98 1.00 0.95 0.93
                                                                                       0.68 0.77 0.79 0.81 0.83 0.90 0.88 0.80
                                                                                                                            0.08
                                                                                                                                0.05 0.07 0.09
 SUPPORT SERVICES EXPENDITURE -
                                                                                                                                                         - 0.6
                                                                                       0.70 0.80 0.81 0.83 0.84 0.92 0.89 0.82 0.06 0.03 0.05 0.09
             OTHER_EXPENDITURE - 0.27 0.96 0.95 0.95 0.95 0.88 0.95 0.92 0.95 1.00 0.93
    CAPITAL_OUTLAY_EXPENDITURE = 0.21
GRADES_PK_G = 0.21
                                        0.92 0.93 0.92 0.92 0.88 0.94 0.91 0.93
                                                                             0.93 1.00
                                                                                           0.78 0.80 0.82 0.82 0.92 0.89 0.81
                                                                                   0.74 1.00 0.73
                                                                                                                            0.08 0.06 0.07 0.08
                                                                                                               0.74 0.77 1.00 -0.03 -0.02 -0.03 -0.03
                   GRADES_KG_G - -0.01
                                        0.83 0.77 0.77 0.78
                                                           0.72 0.77
                                                                    0.75 0.77 0.80 0.78
                                                                                           1.00 1.00 0.99 0.98
                     GRADES_4_G - -0.00
                                                                                                                                                         - 0.4
                                        0.84 0.79 0.78 0.79
                                                               0.79
                                                                        0.79 0.81 0.80
                                                                                           1.00 1.00 1.00 0.98
                                                                                                                   0.79 1.00
                    GRADES_8_G - 0.02
                                                               0.81 0.78 0.81 0.83 0.82
                                                                                           0.99 1.00 1.00 0.98 0.78 0.81 1.00 -0.02 -0.02 -0.01 -0.01
                    GRADES_12_G - 0.06
                                        0.85 0.83 0.83 0.84 0.77 0.83 0.80 0.83 0.84 0.82
                                                                                           0.98 0.98 0.98 1.00 0.80 0.84 0.98
                                                                                                                            0.00 0.00 -0.00 0.01
                   GRADES_1_8_G -
                                        0.91 0.90 0.91 0.90 0.84 0.90 0.87 0.90 0.92 0.92
                                                                                                     0.78 0.80 1.00 0.96 0.77
                                                                                                                            0.08 0.05 0.07 0.13
                                        0.87 0.88 0.89 0.88 0.82 0.88 0.85 0.88 0.89 0.89
                                                                                       0.72 0.77 0.79 0.81 0.84 0.96 1.00 0.80 0.13 0.10 0.11 0.17
                  GRADES_9_12_G -
                                                                                                                                                         - 0.2
                   GRADES_ALL_G - 0.01 0.85
                                                                                                1.00 1.00 0.98
                                                                                                                            -0.02 -0.02 -0.02 -0.02
                                        0.01 0.08 0.09 0.07
                                                                   0.08 0.08 0.06 0.06 0.08 -0.03 -0.03 -0.02 0.00 0.08 0.13 -0.02
              AVG MATH 4 SCORE -
                                   0.43 -0.02 0.05 0.07 0.05 0.05 0.05 0.05 0.05 0.03 0.04 0.06 -0.02 -0.02 -0.02 0.00 0.05 0.10 -0.02
              AVG_MATH_8_SCORE -
                                   0.42 0.03 0.07 0.08 0.07 0.06 0.07 0.07 0.07 0.05 0.06 0.07 -0.03 -0.02 -0.01 -0.00 0.07 0.11 -0.02
                                                                                                                                      1.00 0.90
           AVG_READING_4_SCORE -
           AVG_READING_8_SCORE - 0.53
                                                                   0.09 0.09
                                                                             0.09 0.08 0.08 0.03 0.03 0.01 0.01 0.13 0.17
                                             TOTAL REVENUE
                                                  FEDERAL REVENUE
                                                       STATE_REVENUE
                                                           LOCAL_REVENUE
                                                                         SUPPORT SERVICES EXPENDITURE
                                                                                   CAPITAL OUTLAY EXPENDITURE
                                                                                       GRADES_PK_G
                                                                                            GRADES KG G
                                                                                                          GRADES_12_G
                                                                                                                        GRADES ALL G
                                                                TOTAL_EXPENDITURE
                                                                     INSTRUCTION_EXPENDITURE
                                                                              OTHER_EXPENDITURE
In [23]:
X = data.drop(['PRIMARY KEY', 'ENROLL', 'TOTAL REVENUE', 'FEDERAL REVENUE', 'AVG MATH 4 SCORE
','AVG MATH 8 SCORE','AVG READING 4 SCORE','AVG READING 8 SCORE','STATE','STATE REVENUE',
'LOCAL REVENUE', 'TOTAL EXPENDITURE', 'INSTRUCTION EXPENDITURE', 'SUPPORT SERVICES EXPENDI
TURE', 'OTHER_EXPENDITURE', 'CAPITAL_OUTLAY_EXPENDITURE', 'GRADES_PK_G', 'GRADES_KG_G', 'GRADE
S_4_G','GRADES_8_G','GRADES_12_G','GRADES_1_8_G','GRADES_9_12_G','GRADES_ALL_G'], axis =
1)
Y = data.YEAR
```

print('Входные данные:\n\n', X.head(), '\n\nВыходные данные:\n\n', Y.head())

# Входные данные:

1992

### Выходные данные:

0 1992 1 1992 2 1992 3 1992 4 1992

Name: YEAR, dtype: int64

# In [24]:

# Входные параметры обучающей выборки:

```
Входные параметры тестовой выборки:
       YEAR
1101 2013
      1992
6
746
      2006
1320
     1989
473
      2001
Выходные параметры обучающей выборки:
 82
        1993
1579
        1989
1544
        1989
1323
        2017
249
        1996
Name: YEAR, dtype: int64
Выходные параметры тестовой выборки:
1101
         2013
        1992
746
        2006
1320
        1989
        2001
473
Name: YEAR, dtype: int64
In [25]:
from sklearn.svm import SVC , LinearSVC
from sklearn.datasets.samples generator import make blobs
from matplotlib import pyplot as plt
In [26]:
svc = SVC(kernel='linear')
svc.fit(X train, Y train)
Out[26]:
SVC(kernel='linear')
In [27]:
pred y = svc.predict(X test)
In [28]:
plt.scatter(X_test.YEAR, Y_test, marker = 's', label = 'Тестовая выборка')
plt.scatter(X_test.YEAR, pred_y, marker = '.', label = 'Предсказанные данные')
plt.legend (loc = 'lower right')
plt.xlabel ('YEAR')
plt.ylabel ('YEAR')
plt.show()
       35
  2020
  2015
  2010
  2005
  2000
  1995
  1990
                              Предсказанные данные
  1985 -
                      2000
                           2005
                                 2010
                                      2015
                                            2020
     1985
           1990
                 1995
```

YEAR

```
In [29]:
from sklearn.ensemble import RandomForestRegressor

In [30]:
forest_1 = RandomForestRegressor(n_estimators=5, oob_score=True, random_state=10)
forest_1.fit(X, Y)
```

```
Out[30]:
```

RandomForestRegressor(n estimators=5, oob score=True, random state=10)

#### In [31]:

```
Y_predict = forest_1.predict(X_test)

print('Средняя абсолютная ошибка:', mean_absolute_error(Y_test, Y_predict))

print('Средняя квадратичная ошибка:', mean_squared_error(Y_test, Y_predict))

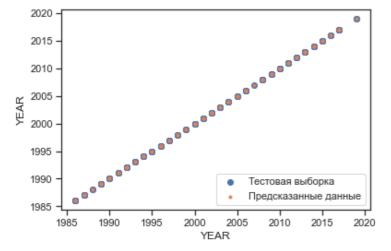
print('Median absolute error:', median_absolute_error(Y_test, Y_predict))

print('Коэффициент детерминации:', r2_score(Y_test, Y_predict))
```

Средняя абсолютная ошибка: 0.0 Средняя квадратичная ошибка: 0.0 Median absolute error: 0.0 Коэффициент детерминации: 1.0

## In [32]:

```
plt.scatter(X_test.YEAR, Y_test, marker = 'o', label = 'Тестовая выборка')
plt.scatter(X_test.YEAR, Y_predict, marker = '.', label = 'Предсказанные данные')
plt.legend(loc = 'lower right')
plt.xlabel('YEAR')
plt.ylabel('YEAR')
plt.show()
```



#### In [ ]: