РК1

Мартынова П.В. ИУ5-61Б вариант 13 задача 2, датасет 5

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import seaborn as sns

df = pd.read_csv('data/states_all_extended.csv')
df.head(10)
```

```
                 PRIMARY_KEY                 STATE  YEAR  ENROLL  \
0              1992_ALABAMA               ALABAMA  1992     NaN
1               1992_ALASKA                ALASKA  1992     NaN
2              1992_ARIZONA               ARIZONA  1992     NaN
3             1992_ARKANSAS              ARKANSAS  1992     NaN
4            1992_CALIFORNIA            CALIFORNIA  1992     NaN
5              1992_COLORADO              COLORADO  1992     NaN
6           1992_CONNECTICUT           CONNECTICUT  1992     NaN
7              1992_DELAWARE              DELAWARE  1992     NaN
8  1992_DISTRICT_OF_COLUMBIA  DISTRICT_OF_COLUMBIA  1992     NaN
9              1992_FLORIDA               FLORIDA  1992     NaN

   TOTAL_REVENUE  FEDERAL_REVENUE  STATE_REVENUE  LOCAL_REVENUE  \
0      2678885.0         304177.0      1659028.0       715680.0
1      1049591.0         106780.0       720711.0       222100.0
2      3258079.0         297888.0      1369815.0      1590376.0
3      1711959.0         178571.0       958785.0       574603.0
4     26260025.0        2072470.0     16546514.0      7641041.0
5      3185173.0         163253.0      1307986.0      1713934.0
6      3834302.0         143542.0      1342539.0      2348221.0
7       645233.0          45945.0       420942.0       178346.0
8       709480.0          64749.0            0.0       644731.0
9     11506299.0         788420.0      5683949.0      5033930.0

   TOTAL_EXPENDITURE  INSTRUCTION_EXPENDITURE  ...
G08_HI_A_READING  \
0          2653798.0                1481703.0  ...                NaN

1           972488.0                 498362.0  ...                NaN

2          3401580.0                1435908.0  ...                NaN

3          1743022.0                 964323.0  ...                NaN

4         27138832.0               14358922.0  ...                NaN
```

```
5          3264826.0                1642466.0  ...                NaN

6          3721338.0                2148041.0  ...                NaN

7           638784.0                 372722.0  ...                NaN

8           742893.0                 329160.0  ...                NaN

9         11305642.0                5166374.0  ...                NaN


   G08_HI_A_MATHEMATICS  G08_AS_A_READING  G08_AS_A_MATHEMATICS  \
0                   NaN               NaN                   NaN
1                   NaN               NaN                   NaN
2                   NaN               NaN                   NaN
3                   NaN               NaN                   NaN
4                   NaN               NaN                   NaN
5                   NaN               NaN                   NaN
6                   NaN               NaN                   NaN
7                   NaN               NaN                   NaN
8                   NaN               NaN                   NaN
9                   NaN               NaN                   NaN

   G08_AM_A_READING  G08_AM_A_MATHEMATICS  G08_HP_A_READING  \
0               NaN                   NaN               NaN
1               NaN                   NaN               NaN
2               NaN                   NaN               NaN
3               NaN                   NaN               NaN
4               NaN                   NaN               NaN
5               NaN                   NaN               NaN
6               NaN                   NaN               NaN
7               NaN                   NaN               NaN
8               NaN                   NaN               NaN
9               NaN                   NaN               NaN

   G08_HP_A_MATHEMATICS  G08_TR_A_READING  G08_TR_A_MATHEMATICS
0                   NaN               NaN                   NaN
1                   NaN               NaN                   NaN
2                   NaN               NaN                   NaN
3                   NaN               NaN                   NaN
4                   NaN               NaN                   NaN
5                   NaN               NaN                   NaN
6                   NaN               NaN                   NaN
7                   NaN               NaN                   NaN
8                   NaN               NaN                   NaN
9                   NaN               NaN                   NaN

[10 rows x 266 columns]
```

```
df.describe()
```

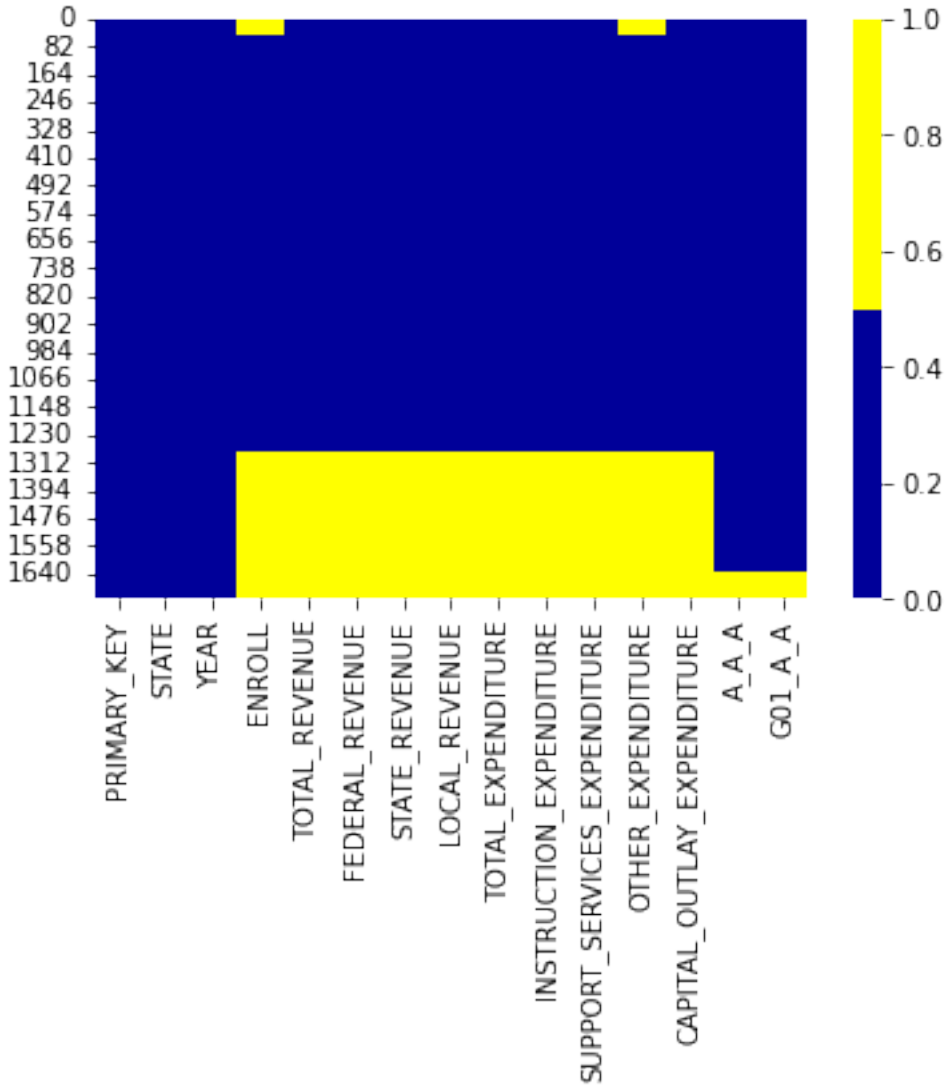
```
              YEAR        ENROLL  TOTAL_REVENUE  FEDERAL_REVENUE  \
count  1715.000000  1.224000e+03   1.275000e+03     1.275000e+03
mean   2002.075219  9.175416e+05   9.102045e+06     7.677799e+05
std       9.568621  1.066514e+06   1.175962e+07     1.146992e+06
min    1986.000000  4.386600e+04   4.656500e+05     3.102000e+04
25%    1994.000000  2.645145e+05   2.189504e+06     1.899575e+05
50%    2002.000000  6.499335e+05   5.085826e+06     4.035480e+05
75%    2010.000000  1.010532e+06   1.084516e+07     8.279320e+05
max    2019.000000  6.307022e+06   8.921726e+07     9.990221e+06

       STATE_REVENUE  LOCAL_REVENUE  TOTAL_EXPENDITURE  \
count   1.275000e+03   1.275000e+03       1.275000e+03
mean    4.223743e+06   4.110522e+06       9.206242e+06
std     5.549735e+06   5.489562e+06       1.199279e+07
min     0.000000e+00   2.209300e+04       4.816650e+05
25%     1.165776e+06   7.151210e+05       2.170404e+06
50%     2.537754e+06   2.058996e+06       5.242672e+06
75%     5.055548e+06   4.755293e+06       1.074420e+07
max     5.090457e+07   3.610526e+07       8.532013e+07

       INSTRUCTION_EXPENDITURE  SUPPORT_SERVICES_EXPENDITURE  \
count             1.275000e+03                  1.275000e+03
mean              4.768010e+06                  2.682587e+06
std               6.300569e+06                  3.357214e+06
min               2.655490e+05                  1.399630e+05
25%               1.171336e+06                  6.380760e+05
50%               2.658253e+06                  1.525471e+06
75%               5.561959e+06                  3.222924e+06
max               4.396452e+07                  2.605802e+07

       OTHER_EXPENDITURE  ...  G08_HI_A_READING  G08_HI_A_MATHEMATICS
\
count       1.224000e+03  ...        246.000000            248.000000

mean        4.299509e+05  ...        254.845528            269.995968

std         5.347893e+05  ...          5.617077              5.992909

min         1.154100e+04  ...        239.000000            255.000000

25%         1.034492e+05  ...        251.000000            266.000000

50%         2.717040e+05  ...        255.000000            270.000000

75%         5.172222e+05  ...        258.000000            274.000000

max         3.995951e+06  ...        277.000000            287.000000
```

```
       G08_AS_A_READING  G08_AS_A_MATHEMATICS  G08_AM_A_READING  \
count        153.000000            157.000000         61.000000
mean         279.803922            306.898089        246.688525
std            9.017570             11.034436          7.630074
min          257.000000            277.000000        229.000000
25%          274.000000            301.000000        242.000000
50%          281.000000            309.000000        247.000000
75%          286.000000            314.000000        252.000000
max          298.000000            332.000000        261.000000

       G08_AM_A_MATHEMATICS  G08_HP_A_READING  G08_HP_A_MATHEMATICS  \
count             60.000000         14.000000             13.000000
mean             261.016667        248.285714            264.846154
std                6.652365          6.568322              7.861787
min              240.000000        234.000000            250.000000
25%              257.750000        244.250000            258.000000
50%              261.000000        248.500000            266.000000
75%              265.000000        253.500000            269.000000
max              275.000000        259.000000            276.000000

       G08_TR_A_READING  G08_TR_A_MATHEMATICS
count        141.000000            145.000000
mean         268.085106            284.082759
std            7.218517              7.684313
min          249.000000            266.000000
25%          263.000000            279.000000
50%          267.000000            284.000000
75%          272.000000            289.000000
max          291.000000            305.000000

[8 rows x 264 columns]
```

Обработка пропусков в данных

```
cols = df.columns[:15]
# желтый - пропущенные данные, синий - не пропущенные
colours = ['#000099', '#ffff00']
sns.heatmap(df[cols].isnull(), cmap=sns.color_palette(colours))
```

```
<AxesSubplot:>
```




```
#Количество пустых ячеек в колонках:
df.isnull().sum()
```

```
PRIMARY_KEY                0
STATE                      0
YEAR                       0
ENROLL                   491
TOTAL_REVENUE            440
                        ...
G08_AM_A_MATHEMATICS    1655
G08_HP_A_READING        1701
G08_HP_A_MATHEMATICS    1702
G08_TR_A_READING        1574
G08_TR_A_MATHEMATICS    1570
Length: 266, dtype: int64
```

```
#Типы данных в колонках:
df.dtypes
```

```
PRIMARY_KEY              object
STATE                    object
YEAR                      int64
ENROLL                  float64
TOTAL_REVENUE           float64
                         ...
G08_AM_A_MATHEMATICS    float64
G08_HP_A_READING        float64
G08_HP_A_MATHEMATICS    float64
G08_TR_A_READING        float64
G08_TR_A_MATHEMATICS    float64
Length: 266, dtype: object
```

```
#Количество пустых числовых значений
num_cols = []
total_count = df.shape[0]
for col in df.columns:
    # Количество пустых значений
    temp_null_count = df[df[col].isnull()].shape[0]
    dt = str(df[col].dtype)
    if temp_null_count>0 and (dt=='float64' or dt=='int64'):
        num_cols.append(col)
        temp_perc = round((temp_null_count / total_count) * 100.0, 2)
        print('Колонка {}. Тип данных {}. Количество пустых значений
{}, {}%.'.format(col, dt, temp_null_count, temp_perc))
```

```
Колонка ENROLL. Тип данных float64. Количество пустых значений 491,
28.63%.
Колонка TOTAL_REVENUE. Тип данных float64. Количество пустых значений
440, 25.66%.
Колонка FEDERAL_REVENUE. Тип данных float64. Количество пустых
значений 440, 25.66%.
Колонка STATE_REVENUE. Тип данных float64. Количество пустых значений
440, 25.66%.
Колонка LOCAL_REVENUE. Тип данных float64. Количество пустых значений
440, 25.66%.
Колонка TOTAL_EXPENDITURE. Тип данных float64. Количество пустых
значений 440, 25.66%.
Колонка INSTRUCTION_EXPENDITURE. Тип данных float64. Количество пустых
значений 440, 25.66%.
Колонка SUPPORT_SERVICES_EXPENDITURE. Тип данных float64. Количество
пустых значений 440, 25.66%.
Колонка OTHER_EXPENDITURE. Тип данных float64. Количество пустых
значений 491, 28.63%.
Колонка CAPITAL_OUTLAY_EXPENDITURE. Тип данных float64. Количество
пустых значений 440, 25.66%.
Колонка A_A_A. Тип данных float64. Количество пустых значений 83,
4.84%.
```

Колонка G01_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G02_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G03_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G04_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G05_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G07_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G08_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G09_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G10_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G11_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка G12_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка KG_A_A. Тип данных float64. Количество пустых значений 83, 4.84%.
Колонка PK_A_A. Тип данных float64. Количество пустых значений 173, 10.09%.
Колонка G01-G08_A_A. Тип данных float64. Количество пустых значений 695, 40.52%.
Колонка G09-G12_A_A. Тип данных float64. Количество пустых значений 644, 37.55%.
Колонка G01_AM_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G01_AM_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_HI_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G01_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_HP_F. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G01_HP_M. Тип данных float64. Количество пустых значений 1352, 78.83%.

Колонка G01_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G01_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G01_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G01_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_AM_F. Тип данных float64. Количество пустых значений 1309, 76.33%.
Колонка G02_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G02_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_HP_F. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G02_HP_M. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G02_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G02_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G02_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G02_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_AM_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G03_AM_M. Тип данных float64. Количество пустых значений 1309, 76.33%.
Колонка G03_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.

Колонка G03_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_HP_F. Тип данных float64. Количество пустых значений 1352, 78.83%.
Колонка G03_HP_M. Тип данных float64. Количество пустых значений 1349, 78.66%.
Колонка G03_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G03_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G03_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G03_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_AM_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G04_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G04_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_HP_F. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G04_HP_M. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G04_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G04_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G04_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G04_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_AM_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G05_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G05_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.

Колонка G05_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_HP_F. Тип данных float64. Количество пустых значений 1352, 78.83%.
Колонка G05_HP_M. Тип данных float64. Количество пустых значений 1349, 78.66%.
Колонка G05_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G05_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G05_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G05_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_AM_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G06_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_HP_F. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G06_HP_M. Тип данных float64. Количество пустых значений 1349, 78.66%.
Колонка G06_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G06_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G06_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G06_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_AM_F. Тип данных float64. Количество пустых значений 1307, 76.21%.

Колонка G07_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G07_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_HP_F. Тип данных float64. Количество пустых значений 1350, 78.72%.
Колонка G07_HP_M. Тип данных float64. Количество пустых значений 1352, 78.83%.
Колонка G07_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G07_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G07_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G07_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_AM_F. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G08_AM_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_HP_F. Тип данных float64. Количество пустых значений 1350, 78.72%.
Колонка G08_HP_M. Тип данных float64. Количество пустых значений 1349, 78.66%.
Колонка G08_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G08_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.

Колонка G08_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G08_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_AM_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G09_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.

Колонка G09_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_HP_F. Тип данных float64. Количество пустых значений 1352, 78.83%.
Колонка G09_HP_M. Тип данных float64. Количество пустых значений 1351, 78.78%.
Колонка G09_TR_F. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G09_TR_M. Тип данных float64. Количество пустых значений 1344, 78.37%.
Колонка G09_WH_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G09_WH_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_AM_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_AM_M. Тип данных float64. Количество пустых значений 1308, 76.27%.
Колонка G10_AS_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_AS_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_BL_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_BL_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_HI_F. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_HI_M. Тип данных float64. Количество пустых значений 1307, 76.21%.
Колонка G10_HP_F. Тип данных float64. Количество пустых значений 1352,

78.83%.
Колонка G10_HP_M. Тип данных float64. Количество пустых значений 1351,
78.78%.
Колонка G10_TR_F. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G10_TR_M. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G10_WH_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G10_WH_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_AM_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_AM_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_AS_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_AS_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_BL_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_BL_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_HI_F. Тип данных float64. Количество пустых значений 1308,
76.27%.
Колонка G11_HI_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_HP_F. Тип данных float64. Количество пустых значений 1352,
78.83%.
Колонка G11_HP_M. Тип данных float64. Количество пустых значений 1354,
78.95%.
Колонка G11_TR_F. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G11_TR_M. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G11_WH_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G11_WH_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_AM_F. Тип данных float64. Количество пустых значений 1309,
76.33%.
Колонка G12_AM_M. Тип данных float64. Количество пустых значений 1310,
76.38%.
Колонка G12_AS_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_AS_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_BL_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_BL_M. Тип данных float64. Количество пустых значений 1307,

76.21%.
Колонка G12_HI_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_HI_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_HP_F. Тип данных float64. Количество пустых значений 1352,
78.83%.
Колонка G12_HP_M. Тип данных float64. Количество пустых значений 1352,
78.83%.
Колонка G12_TR_F. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G12_TR_M. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка G12_WH_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка G12_WH_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_AM_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_AM_M. Тип данных float64. Количество пустых значений 1308,
76.27%.
Колонка KG_AS_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_AS_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_BL_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_BL_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_HI_F. Тип данных float64. Количество пустых значений 1308,
76.27%.
Колонка KG_HI_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_HP_F. Тип данных float64. Количество пустых значений 1349,
78.66%.
Колонка KG_HP_M. Тип данных float64. Количество пустых значений 1350,
78.72%.
Колонка KG_TR_F. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка KG_TR_M. Тип данных float64. Количество пустых значений 1344,
78.37%.
Колонка KG_WH_F. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка KG_WH_M. Тип данных float64. Количество пустых значений 1307,
76.21%.
Колонка PK_AM_F. Тип данных float64. Количество пустых значений 1332,
77.67%.
Колонка PK_AM_M. Тип данных float64. Количество пустых значений 1321,
77.03%.
Колонка PK_AS_F. Тип данных float64. Количество пустых значений 1321,

77.03%.
Колонка PK_AS_M. Тип данных float64. Количество пустых значений 1323, 77.14%.
Колонка PK_BL_F. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка PK_BL_M. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка PK_HI_F. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка PK_HI_M. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка PK_HP_F. Тип данных float64. Количество пустых значений 1387, 80.87%.
Колонка PK_HP_M. Тип данных float64. Количество пустых значений 1384, 80.7%.
Колонка PK_TR_F. Тип данных float64. Количество пустых значений 1357, 79.13%.
Колонка PK_TR_M. Тип данных float64. Количество пустых значений 1357, 79.13%.
Колонка PK_WH_F. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка PK_WH_M. Тип данных float64. Количество пустых значений 1321, 77.03%.
Колонка G04_A_A_READING. Тип данных float64. Количество пустых значений 1065, 62.1%.
Колонка G04_A_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1150, 67.06%.
Колонка G04_A_M_READING. Тип данных float64. Количество пустых значений 1065, 62.1%.
Колонка G04_A_M_MATHEMATICS. Тип данных float64. Количество пустых значений 1150, 67.06%.
Колонка G04_A_F_READING. Тип данных float64. Количество пустых значений 1065, 62.1%.
Колонка G04_A_F_MATHEMATICS. Тип данных float64. Количество пустых значений 1150, 67.06%.
Колонка G04_WH_A_READING. Тип данных float64. Количество пустых значений 1450, 84.55%.
Колонка G04_WH_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1450, 84.55%.
Колонка G04_BL_A_READING. Тип данных float64. Количество пустых значений 1489, 86.82%.
Колонка G04_BL_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1486, 86.65%.
Колонка G04_HI_A_READING. Тип данных float64. Количество пустых значений 1465, 85.42%.
Колонка G04_HI_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1465, 85.42%.
Колонка G04_AS_A_READING. Тип данных float64. Количество пустых значений 1551, 90.44%.
Колонка G04_AS_A_MATHEMATICS. Тип данных float64. Количество пустых

значений 1547, 90.2%.
Колонка G04_AM_A_READING. Тип данных float64. Количество пустых значений 1651, 96.27%.
Колонка G04_AM_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1652, 96.33%.
Колонка G04_HP_A_READING. Тип данных float64. Количество пустых значений 1699, 99.07%.
Колонка G04_HP_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1700, 99.13%.
Колонка G04_TR_A_READING. Тип данных float64. Количество пустых значений 1532, 89.33%.
Колонка G04_TR_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1532, 89.33%.
Колонка G08_A_A_READING. Тип данных float64. Количество пустых значений 1153, 67.23%.
Колонка G08_A_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1113, 64.9%.
Колонка G08_A_M_READING. Тип данных float64. Количество пустых значений 1153, 67.23%.
Колонка G08_A_M_MATHEMATICS. Тип данных float64. Количество пустых значений 1113, 64.9%.
Колонка G08_A_F_READING. Тип данных float64. Количество пустых значений 1153, 67.23%.
Колонка G08_A_F_MATHEMATICS. Тип данных float64. Количество пустых значений 1113, 64.9%.
Колонка G08_WH_A_READING. Тип данных float64. Количество пустых значений 1450, 84.55%.
Колонка G08_WH_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1450, 84.55%.
Колонка G08_BL_A_READING. Тип данных float64. Количество пустых значений 1493, 87.06%.
Колонка G08_BL_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1494, 87.11%.
Колонка G08_HI_A_READING. Тип данных float64. Количество пустых значений 1469, 85.66%.
Колонка G08_HI_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1467, 85.54%.
Колонка G08_AS_A_READING. Тип данных float64. Количество пустых значений 1562, 91.08%.
Колонка G08_AS_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1558, 90.85%.
Колонка G08_AM_A_READING. Тип данных float64. Количество пустых значений 1654, 96.44%.
Колонка G08_AM_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1655, 96.5%.
Колонка G08_HP_A_READING. Тип данных float64. Количество пустых значений 1701, 99.18%.
Колонка G08_HP_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1702, 99.24%.
Колонка G08_TR_A_READING. Тип данных float64. Количество пустых

значений 1574, 91.78%.
Колонка G08_TR_A_MATHEMATICS. Тип данных float64. Количество пустых значений 1570, 91.55%.

Возьмем в качестве количественного признака признак G06_A_A - общее количество учащихся шестого класса Заменим пропуски на медианное значение:

```
med = df['G06_A_A'].median()
print(med)
df['G06_A_A'] = df['G06_A_A'].fillna(med)
```

```
48672.0
```

```
for col in df.columns:
    pct_missing = np.mean(df[col].isnull())
    print('{} - {}%'.format(col, round(pct_missing*100)))
```

```
for col in df.columns:
    missing = df[col].isnull()
    num_missing = np.sum(missing)
```

```
PRIMARY_KEY - 0%
STATE - 0%
YEAR - 0%
ENROLL - 29%
TOTAL_REVENUE - 26%
FEDERAL_REVENUE - 26%
STATE_REVENUE - 26%
LOCAL_REVENUE - 26%
TOTAL_EXPENDITURE - 26%
INSTRUCTION_EXPENDITURE - 26%
SUPPORT_SERVICES_EXPENDITURE - 26%
OTHER_EXPENDITURE - 29%
CAPITAL_OUTLAY_EXPENDITURE - 26%
A_A_A - 5%
G01_A_A - 5%
G02_A_A - 5%
G03_A_A - 5%
G04_A_A - 5%
G05_A_A - 5%
G06_A_A - 0%
G07_A_A - 5%
G08_A_A - 5%
G09_A_A - 5%
G10_A_A - 5%
G11_A_A - 5%
G12_A_A - 5%
KG_A_A - 5%
PK_A_A - 10%
G01-G08_A_A - 41%
```

G09-G12_A_A - 38%
G01_AM_F - 76%
G01_AM_M - 76%
G01_AS_F - 76%
G01_AS_M - 76%
G01_BL_F - 76%
G01_BL_M - 76%
G01_HI_F - 76%
G01_HI_M - 76%
G01_HP_F - 79%
G01_HP_M - 79%
G01_TR_F - 78%
G01_TR_M - 78%
G01_WH_F - 76%
G01_WH_M - 76%
G02_AM_F - 76%
G02_AM_M - 76%
G02_AS_F - 76%
G02_AS_M - 76%
G02_BL_F - 76%
G02_BL_M - 76%
G02_HI_F - 76%
G02_HI_M - 76%
G02_HP_F - 79%
G02_HP_M - 79%
G02_TR_F - 78%
G02_TR_M - 78%
G02_WH_F - 76%
G02_WH_M - 76%
G03_AM_F - 76%
G03_AM_M - 76%
G03_AS_F - 76%
G03_AS_M - 76%
G03_BL_F - 76%
G03_BL_M - 76%
G03_HI_F - 76%
G03_HI_M - 76%
G03_HP_F - 79%
G03_HP_M - 79%
G03_TR_F - 78%
G03_TR_M - 78%
G03_WH_F - 76%
G03_WH_M - 76%
G04_AM_F - 76%
G04_AM_M - 76%
G04_AS_F - 76%
G04_AS_M - 76%
G04_BL_F - 76%
G04_BL_M - 76%
G04_HI_F - 76%

G04_HI_M - 76%
G04_HP_F - 79%
G04_HP_M - 79%
G04_TR_F - 78%
G04_TR_M - 78%
G04_WH_F - 76%
G04_WH_M - 76%
G05_AM_F - 76%
G05_AM_M - 76%
G05_AS_F - 76%
G05_AS_M - 76%
G05_BL_F - 76%
G05_BL_M - 76%
G05_HI_F - 76%
G05_HI_M - 76%
G05_HP_F - 79%
G05_HP_M - 79%
G05_TR_F - 78%
G05_TR_M - 78%
G05_WH_F - 76%
G05_WH_M - 76%
G06_AM_F - 76%
G06_AM_M - 76%
G06_AS_F - 76%
G06_AS_M - 76%
G06_BL_F - 76%
G06_BL_M - 76%
G06_HI_F - 76%
G06_HI_M - 76%
G06_HP_F - 79%
G06_HP_M - 79%
G06_TR_F - 78%
G06_TR_M - 78%
G06_WH_F - 76%
G06_WH_M - 76%
G07_AM_F - 76%
G07_AM_M - 76%
G07_AS_F - 76%
G07_AS_M - 76%
G07_BL_F - 76%
G07_BL_M - 76%
G07_HI_F - 76%
G07_HI_M - 76%
G07_HP_F - 79%
G07_HP_M - 79%
G07_TR_F - 78%
G07_TR_M - 78%
G07_WH_F - 76%
G07_WH_M - 76%
G08_AM_F - 76%

G08_AM_M - 76%
G08_AS_F - 76%
G08_AS_M - 76%
G08_BL_F - 76%
G08_BL_M - 76%
G08_HI_F - 76%
G08_HI_M - 76%
G08_HP_F - 79%
G08_HP_M - 79%
G08_TR_F - 78%
G08_TR_M - 78%
G08_WH_F - 76%
G08_WH_M - 76%
G09_AM_F - 76%
G09_AM_M - 76%
G09_AS_F - 76%
G09_AS_M - 76%
G09_BL_F - 76%
G09_BL_M - 76%
G09_HI_F - 76%
G09_HI_M - 76%
G09_HP_F - 79%
G09_HP_M - 79%
G09_TR_F - 78%
G09_TR_M - 78%
G09_WH_F - 76%
G09_WH_M - 76%
G10_AM_F - 76%
G10_AM_M - 76%
G10_AS_F - 76%
G10_AS_M - 76%
G10_BL_F - 76%
G10_BL_M - 76%
G10_HI_F - 76%
G10_HI_M - 76%
G10_HP_F - 79%
G10_HP_M - 79%
G10_TR_F - 78%
G10_TR_M - 78%
G10_WH_F - 76%
G10_WH_M - 76%
G11_AM_F - 76%
G11_AM_M - 76%
G11_AS_F - 76%
G11_AS_M - 76%
G11_BL_F - 76%
G11_BL_M - 76%
G11_HI_F - 76%
G11_HI_M - 76%
G11_HP_F - 79%

G11_HP_M - 79%
G11_TR_F - 78%
G11_TR_M - 78%
G11_WH_F - 76%
G11_WH_M - 76%
G12_AM_F - 76%
G12_AM_M - 76%
G12_AS_F - 76%
G12_AS_M - 76%
G12_BL_F - 76%
G12_BL_M - 76%
G12_HI_F - 76%
G12_HI_M - 76%
G12_HP_F - 79%
G12_HP_M - 79%
G12_TR_F - 78%
G12_TR_M - 78%
G12_WH_F - 76%
G12_WH_M - 76%
KG_AM_F - 76%
KG_AM_M - 76%
KG_AS_F - 76%
KG_AS_M - 76%
KG_BL_F - 76%
KG_BL_M - 76%
KG_HI_F - 76%
KG_HI_M - 76%
KG_HP_F - 79%
KG_HP_M - 79%
KG_TR_F - 78%
KG_TR_M - 78%
KG_WH_F - 76%
KG_WH_M - 76%
PK_AM_F - 78%
PK_AM_M - 77%
PK_AS_F - 77%
PK_AS_M - 77%
PK_BL_F - 77%
PK_BL_M - 77%
PK_HI_F - 77%
PK_HI_M - 77%
PK_HP_F - 81%
PK_HP_M - 81%
PK_TR_F - 79%
PK_TR_M - 79%
PK_WH_F - 77%
PK_WH_M - 77%
G04_A_A_READING - 62%
G04_A_A_MATHEMATICS - 67%
G04_A_M_READING - 62%

```
G04_A_M_MATHEMATICS - 67%
G04_A_F_READING - 62%
G04_A_F_MATHEMATICS - 67%
G04_WH_A_READING - 85%
G04_WH_A_MATHEMATICS - 85%
G04_BL_A_READING - 87%
G04_BL_A_MATHEMATICS - 87%
G04_HI_A_READING - 85%
G04_HI_A_MATHEMATICS - 85%
G04_AS_A_READING - 90%
G04_AS_A_MATHEMATICS - 90%
G04_AM_A_READING - 96%
G04_AM_A_MATHEMATICS - 96%
G04_HP_A_READING - 99%
G04_HP_A_MATHEMATICS - 99%
G04_TR_A_READING - 89%
G04_TR_A_MATHEMATICS - 89%
G08_A_A_READING - 67%
G08_A_A_MATHEMATICS - 65%
G08_A_M_READING - 67%
G08_A_M_MATHEMATICS - 65%
G08_A_F_READING - 67%
G08_A_F_MATHEMATICS - 65%
G08_WH_A_READING - 85%
G08_WH_A_MATHEMATICS - 85%
G08_BL_A_READING - 87%
G08_BL_A_MATHEMATICS - 87%
G08_HI_A_READING - 86%
G08_HI_A_MATHEMATICS - 86%
G08_AS_A_READING - 91%
G08_AS_A_MATHEMATICS - 91%
G08_AM_A_READING - 96%
G08_AM_A_MATHEMATICS - 97%
G08_HP_A_READING - 99%
G08_HP_A_MATHEMATICS - 99%
G08_TR_A_READING - 92%
G08_TR_A_MATHEMATICS - 92%

print(df['G06_A_A'])

0        59929.0
1         9542.0
2        53832.0
3        35017.0
4       399776.0
          ...
1710     48672.0
1711     48672.0
1712     48672.0
1713     48672.0
```

```
1714     48672.0
Name: G06_A_A, Length: 1715, dtype: float64
```

В качестве категориального признака можно было бы взять, например, state, но так как в этом столбце нет пропущенных значений, обработаем пропусти еще для одного количественного признака: GRADES 9_12 - Число учащихся с девятого по двенадцатый классы. Заполним пропуски средним значением.

```
mean = df['G09-G12_A_A'].mean()
print(mean)
df['G09-G12_A_A'] = df['G09-G12_A_A'].fillna(mean)
```

```
282069.08496732026
```

```
for col in df.columns:
    pct_missing = np.mean(df[col].isnull())
    print('{} - {}%'.format(col, round(pct_missing*100)))
```

```
for col in df.columns:
    missing = df[col].isnull()
    num_missing = np.sum(missing)
```

```
PRIMARY_KEY - 0%
STATE - 0%
YEAR - 0%
ENROLL - 29%
TOTAL_REVENUE - 26%
FEDERAL_REVENUE - 26%
STATE_REVENUE - 26%
LOCAL_REVENUE - 26%
TOTAL_EXPENDITURE - 26%
INSTRUCTION_EXPENDITURE - 26%
SUPPORT_SERVICES_EXPENDITURE - 26%
OTHER_EXPENDITURE - 29%
CAPITAL_OUTLAY_EXPENDITURE - 26%
A_A_A - 5%
G01_A_A - 5%
G02_A_A - 5%
G03_A_A - 5%
G04_A_A - 5%
G05_A_A - 5%
G06_A_A - 0%
G07_A_A - 5%
G08_A_A - 5%
G09_A_A - 5%
G10_A_A - 5%
G11_A_A - 5%
G12_A_A - 5%
KG_A_A - 5%
PK_A_A - 10%
```

G01-G08_A_A - 41%
G09-G12_A_A - 0%
G01_AM_F - 76%
G01_AM_M - 76%
G01_AS_F - 76%
G01_AS_M - 76%
G01_BL_F - 76%
G01_BL_M - 76%
G01_HI_F - 76%
G01_HI_M - 76%
G01_HP_F - 79%
G01_HP_M - 79%
G01_TR_F - 78%
G01_TR_M - 78%
G01_WH_F - 76%
G01_WH_M - 76%
G02_AM_F - 76%
G02_AM_M - 76%
G02_AS_F - 76%
G02_AS_M - 76%
G02_BL_F - 76%
G02_BL_M - 76%
G02_HI_F - 76%
G02_HI_M - 76%
G02_HP_F - 79%
G02_HP_M - 79%
G02_TR_F - 78%
G02_TR_M - 78%
G02_WH_F - 76%
G02_WH_M - 76%
G03_AM_F - 76%
G03_AM_M - 76%
G03_AS_F - 76%
G03_AS_M - 76%
G03_BL_F - 76%
G03_BL_M - 76%
G03_HI_F - 76%
G03_HI_M - 76%
G03_HP_F - 79%
G03_HP_M - 79%
G03_TR_F - 78%
G03_TR_M - 78%
G03_WH_F - 76%
G03_WH_M - 76%
G04_AM_F - 76%
G04_AM_M - 76%
G04_AS_F - 76%
G04_AS_M - 76%
G04_BL_F - 76%
G04_BL_M - 76%

G04_HI_F - 76%
G04_HI_M - 76%
G04_HP_F - 79%
G04_HP_M - 79%
G04_TR_F - 78%
G04_TR_M - 78%
G04_WH_F - 76%
G04_WH_M - 76%
G05_AM_F - 76%
G05_AM_M - 76%
G05_AS_F - 76%
G05_AS_M - 76%
G05_BL_F - 76%
G05_BL_M - 76%
G05_HI_F - 76%
G05_HI_M - 76%
G05_HP_F - 79%
G05_HP_M - 79%
G05_TR_F - 78%
G05_TR_M - 78%
G05_WH_F - 76%
G05_WH_M - 76%
G06_AM_F - 76%
G06_AM_M - 76%
G06_AS_F - 76%
G06_AS_M - 76%
G06_BL_F - 76%
G06_BL_M - 76%
G06_HI_F - 76%
G06_HI_M - 76%
G06_HP_F - 79%
G06_HP_M - 79%
G06_TR_F - 78%
G06_TR_M - 78%
G06_WH_F - 76%
G06_WH_M - 76%
G07_AM_F - 76%
G07_AM_M - 76%
G07_AS_F - 76%
G07_AS_M - 76%
G07_BL_F - 76%
G07_BL_M - 76%
G07_HI_F - 76%
G07_HI_M - 76%
G07_HP_F - 79%
G07_HP_M - 79%
G07_TR_F - 78%
G07_TR_M - 78%
G07_WH_F - 76%
G07_WH_M - 76%

G08_AM_F - 76%
G08_AM_M - 76%
G08_AS_F - 76%
G08_AS_M - 76%
G08_BL_F - 76%
G08_BL_M - 76%
G08_HI_F - 76%
G08_HI_M - 76%
G08_HP_F - 79%
G08_HP_M - 79%
G08_TR_F - 78%
G08_TR_M - 78%
G08_WH_F - 76%
G08_WH_M - 76%
G09_AM_F - 76%
G09_AM_M - 76%
G09_AS_F - 76%
G09_AS_M - 76%
G09_BL_F - 76%
G09_BL_M - 76%
G09_HI_F - 76%
G09_HI_M - 76%
G09_HP_F - 79%
G09_HP_M - 79%
G09_TR_F - 78%
G09_TR_M - 78%
G09_WH_F - 76%
G09_WH_M - 76%
G10_AM_F - 76%
G10_AM_M - 76%
G10_AS_F - 76%
G10_AS_M - 76%
G10_BL_F - 76%
G10_BL_M - 76%
G10_HI_F - 76%
G10_HI_M - 76%
G10_HP_F - 79%
G10_HP_M - 79%
G10_TR_F - 78%
G10_TR_M - 78%
G10_WH_F - 76%
G10_WH_M - 76%
G11_AM_F - 76%
G11_AM_M - 76%
G11_AS_F - 76%
G11_AS_M - 76%
G11_BL_F - 76%
G11_BL_M - 76%
G11_HI_F - 76%
G11_HI_M - 76%

G11_HP_F - 79%
G11_HP_M - 79%
G11_TR_F - 78%
G11_TR_M - 78%
G11_WH_F - 76%
G11_WH_M - 76%
G12_AM_F - 76%
G12_AM_M - 76%
G12_AS_F - 76%
G12_AS_M - 76%
G12_BL_F - 76%
G12_BL_M - 76%
G12_HI_F - 76%
G12_HI_M - 76%
G12_HP_F - 79%
G12_HP_M - 79%
G12_TR_F - 78%
G12_TR_M - 78%
G12_WH_F - 76%
G12_WH_M - 76%
KG_AM_F - 76%
KG_AM_M - 76%
KG_AS_F - 76%
KG_AS_M - 76%
KG_BL_F - 76%
KG_BL_M - 76%
KG_HI_F - 76%
KG_HI_M - 76%
KG_HP_F - 79%
KG_HP_M - 79%
KG_TR_F - 78%
KG_TR_M - 78%
KG_WH_F - 76%
KG_WH_M - 76%
PK_AM_F - 78%
PK_AM_M - 77%
PK_AS_F - 77%
PK_AS_M - 77%
PK_BL_F - 77%
PK_BL_M - 77%
PK_HI_F - 77%
PK_HI_M - 77%
PK_HP_F - 81%
PK_HP_M - 81%
PK_TR_F - 79%
PK_TR_M - 79%
PK_WH_F - 77%
PK_WH_M - 77%
G04_A_A_READING - 62%
G04_A_A_MATHEMATICS - 67%

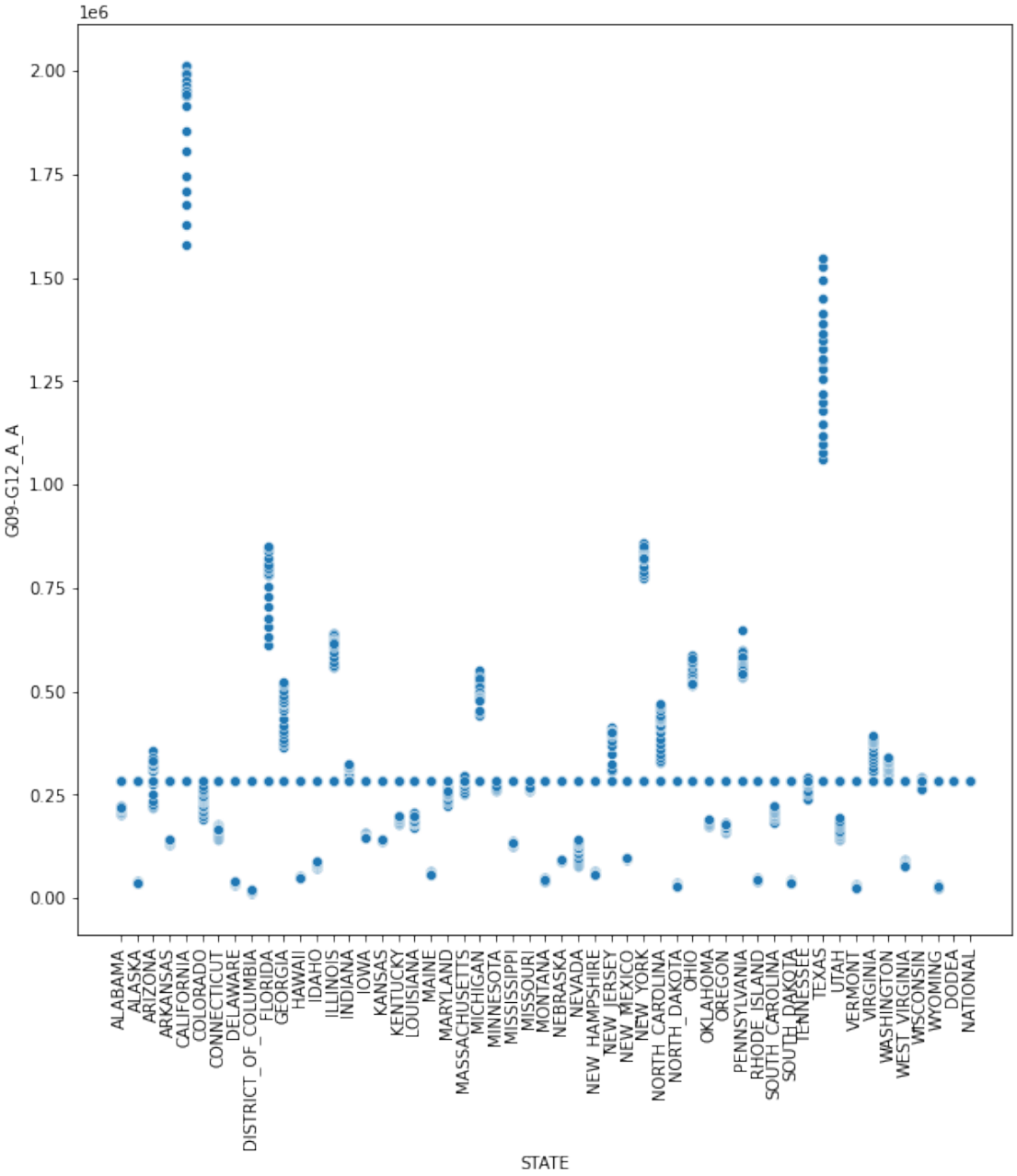
```
G04_A_M_READING - 62%
G04_A_M_MATHEMATICS - 67%
G04_A_F_READING - 62%
G04_A_F_MATHEMATICS - 67%
G04_WH_A_READING - 85%
G04_WH_A_MATHEMATICS - 85%
G04_BL_A_READING - 87%
G04_BL_A_MATHEMATICS - 87%
G04_HI_A_READING - 85%
G04_HI_A_MATHEMATICS - 85%
G04_AS_A_READING - 90%
G04_AS_A_MATHEMATICS - 90%
G04_AM_A_READING - 96%
G04_AM_A_MATHEMATICS - 96%
G04_HP_A_READING - 99%
G04_HP_A_MATHEMATICS - 99%
G04_TR_A_READING - 89%
G04_TR_A_MATHEMATICS - 89%
G08_A_A_READING - 67%
G08_A_A_MATHEMATICS - 65%
G08_A_M_READING - 67%
G08_A_M_MATHEMATICS - 65%
G08_A_F_READING - 67%
G08_A_F_MATHEMATICS - 65%
G08_WH_A_READING - 85%
G08_WH_A_MATHEMATICS - 85%
G08_BL_A_READING - 87%
G08_BL_A_MATHEMATICS - 87%
G08_HI_A_READING - 86%
G08_HI_A_MATHEMATICS - 86%
G08_AS_A_READING - 91%
G08_AS_A_MATHEMATICS - 91%
G08_AM_A_READING - 96%
G08_AM_A_MATHEMATICS - 97%
G08_HP_A_READING - 99%
G08_HP_A_MATHEMATICS - 99%
G08_TR_A_READING - 92%
G08_TR_A_MATHEMATICS - 92%
```

При обработке пропусков в данных для категориального признака можно было бы использовать стратегии "most_frequent" или "constant"

Диаграмма рассеивания

```
fig, ax = plt.subplots(figsize=(10,10))
plt.xticks(rotation=90)
sns.scatterplot(ax=ax, x='STATE', y='G09-G12_A_A', data=df)
```

```
<AxesSubplot:xlabel='STATE', ylabel='G09-G12_A_A'>
```


1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
G09-G12_A_A
ALABAMA
ALASKA
ARIZONA
ARKANSAS
CALIFORNIA
COLORADO
CONNECTICUT
DELAWARE
DISTRICT_OF_COLUMBIA
FLORIDA
GEORGIA
HAWAII
IDAHO
ILLINOIS
INDIANA
IOWA
KANSAS
KENTUCKY
LOUISIANA
MAINE
MARYLAND
MASSACHUSETTS
MICHIGAN
MINNESOTA
MISSISSIPPI
MISSOURI
MONTANA
NEBRASKA
NEVADA
NEW_HAMPSHIRE
NEW_JERSEY
NEW_MEXICO
NEW_YORK
NORTH_CAROLINA
NORTH_DAKOTA
OHIO
OKLAHOMA
OREGON
PENNSYLVANIA
RHODE_ISLAND
SOUTH_CAROLINA
SOUTH_DAKOTA
TENNESSEE
TEXAS
UTAH
VERMONT
VIRGINIA
WASHINGTON
WEST_VIRGINIA
WISCONSIN
WYOMING
DODEA
NATIONAL
STATE