```
Рк №1, Шушпанов В.О., ИУ5-62Б
      Вариант 21 Для одной из колонок провести масштабирование. Произвести кодирование двумя способами ещё для одной колонки. Для одной из
      колонок сделать ящик с усами.
In [52]: import numpy as np
      import pandas as pd
      import seaborn as sns
      import matplotlib.pyplot as plt
      from sklearn.preprocessing import MinMaxScaler
      from sklearn.preprocessing import LabelEncoder, OneHotEncoder
In [46]: data = pd.read_csv('states_all_extended.csv', sep=",")
In [47]: data.shape
Out[47]: (1715, 266)
In [48]: data.columns
Out[48]: Index(['PRIMARY_KEY', 'STATE', 'YEAR', 'ENROLL', 'TOTAL_REVENUE',
           'FEDERAL_REVENUE', 'STATE_REVENUE', 'LOCAL_REVENUE',
          'TOTAL_EXPENDITURE', 'INSTRUCTION_EXPENDITURE',
          'G08_HI_A_READING', 'G08_HI_A_MATHEMATICS', 'G08_AS_A_READING',
          'G08_AS_A_MATHEMATICS', 'G08_AM_A_READING', 'G08_AM_A_MATHEMATICS',
          'G08_HP_A_READING', 'G08_HP_A_MATHEMATICS', 'G08_TR_A_READING',
          'G08_TR_A_MATHEMATICS'],
          dtype='object', length=266)
In [49]: sc1 = MinMaxScaler()
In [67]: data['ENROLL'].replace('', np.nan, inplace=True)
      data.dropna(subset=['ENROLL'], inplace=True)
      data_norm_by_enroll = sc1.fit_transform(data[['ENROLL']])
      data_norm_by_enroll
Out[67]: array([[0.10918617],
          [0.01234042],
          [0.10097641],
          [0.03718541],
          [0.12994567],
          [0.00808618]])
In [78]: sns.boxplot(data_norm_by_enroll)
Out[78]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1e8e7790>
            0.2
In [69]: cat_enc = pd.DataFrame({'c1':data["STATE"]})
In [70]: cat_enc['c1'].unique()
Out[70]: array(['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'], dtype=object)
In [71]: le = LabelEncoder()
      cat_enc_le = le.fit_transform(cat_enc['c1'])
In [72]: | np.unique(cat_enc_le)
Out[72]: array([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
          17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
          34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50])
In [73]: ohe = OneHotEncoder()
      cat_enc_ohe = ohe.fit_transform(cat_enc[['c1']])
In [74]: cat_enc.shape
Out[74]: (1224, 1)
In [75]: cat_enc_ohe.shape
Out[75]: (1224, 51)
In [76]: cat_enc_ohe.todense()[0:10]
0., 0., 0.],
           0., 0., 0.],
           0., 0., 0.],
           0., 0., 0.],
           0., 0., 0.],
           0., 0., 0.],
           [0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
           0., 0., 0.],
           [0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
           0., 0., 0.],
           [0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
           0., 0., 0.],
           [0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0.,
           0., 0., 0.]])
In [77]: sns.boxplot(cat_enc_le)
Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1e720dd0>
In [ ]:
```

In []: