

РК №1 Зудин А.М. ИУ5-25М Вариант №7

In [9]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
from sklearn.impute import MissingIndicator
%matplotlib inline
```

In [13]:

```
data = pd.read_csv('/content/uk_universities.csv')
print(data)
```

	University_name	Region	Founded_year	\
0	University of Cambridge	East of England	1209	
1	University of Oxford	South East England	1096	
2	University of St Andrews	Scotland	1413	
3	Imperial College London	London	1907	
4	Loughborough University	East Midlands	1966	
..	
126	University of East London	London	1992	
127	University of Bedfordshire	East of England	2006	
128	University of Suffolk	East of England	2007	
129	Wrexham Glyndwr University	Wales	2008	
130	Ravensbourne University London	London	1962	

	Motto	UK_rank	World_rank	\
0	From here, light and sacred draughts	1	4	
1	The Lord is my light	2	2	
2	Ever to excel	3	86	
3	Knowledge is the adornment and safeguard of th...	4	8	
4	With Truth, Knowledge and Labour	5	404	
..	
126	Science and fulfillment of vows	127	971	
127	NaN	128	1281	
128	Honesty and diligence	129	4030	
129	Confidence through Education	130	2397	
130	Designed for industry	131	2759	

	CWUR_score	Minimum_IELTS_score	UG_average_fees_(in_pounds)	\
0	94.1	6.5	21750	
1	93.3	6.5	21770	
2	75.8	6.5	17040	
3	86.6	6.5	23500	
4	72.8	5.5	16400	
..	
126	NaN	4.5	10700	
127	NaN	4.5	9750	
128	NaN	4.5	9495	
129	NaN	4.5	10250	
130	NaN	5.5	9250	

	PG_average_fees_(in_pounds)	...	Student_satisfaction	Student_enrollment	\
0	23187	...	85.50%	20,000-24,999	
1	19888	...	86.50%	25,000-29,999	
2	15440	...	87.90%	10,000-14,999	
3	29900	...	77.90%	15,000-19,999	
4	16400	...	85.80%	15,000-19,999	
..	
126	11700	...	76.10%	10,000-14,999	
127	9900	...	77.40%	10,000-14,999	
128	9495	...	76.00%	7,000-7,999	
129	10500	...	74.30%	5,000-5,999	
130	9250	...	76.10%	2,000-2,999	

```

130      Academic_staff Control_type Academic_Calender Campus_setting \
0      over-5,000      Public      Trimesters      Urban
1      over-5,000      Public      Trimesters      Urban
2      1,000-1,499      Public      Semesters      Suburban
3      4,000-4,499      Public      Continuous      Urban
4      1,500-1,999      Public      Semesters      Suburban
..      ...      ...      ...      ...
126      900-999      Public      Semesters      Suburban
127      500-599      Public      Semesters      Urban
128      200-299      Public      NaN      NaN
129      200-299      Public      NaN      Urban
130      100-199      Public      Semesters      Urban

```

```

      Estimated_cost_of_living_per_year_(in_pounds) Latitude Longitude \
0      12000      52.2054      0.1132
1      11500      51.7548      -1.2544
2      12000      56.3417      -2.7943
3      10700      51.4988      -0.1749
4      9398      52.7650      -1.2321
..      ...      ...      ...
126      10229      51.5076      0.0651
127      9415      51.8779      -0.4093
128      9415      52.0523      1.1629
129      7771      53.0526      -3.0062
130      10229      51.5017      0.0055

```

```

      Website
0      www.cam.ac.uk
1      www.ox.ac.uk
2      www.st-andrews.ac.uk
3      www.ic.ac.uk
4      www.lboro.ac.uk/
..      ...
126      www.uel.ac.uk
127      www.beds.ac.uk
128      www.ucs.ac.uk/
129      www.glyndwr.ac.uk/
130      www.ravensbourne.ac.uk

```

[131 rows x 21 columns]

Задача №7

In [14]:

```
data.isnull().sum()
```

Out[14]:

```

University_name      0
Region               0
Founded_year         0
Motto               17
UK_rank              0
World_rank           0
CWUR_score          47
Minimum_IELTS_score  0
UG_average_fees_(in_pounds)  0
PG_average_fees_(in_pounds)  0
International_students  0
Student_satisfaction  0
Student_enrollment   0
Academic_staff       0
Control_type         0
Academic_Calender    26
Campus_setting       18
Estimated_cost_of_living_per_year_(in_pounds)  0
Latitude             0
Longitude            0

```

```
Website  
dtype: int64
```

0

```
In [15]:
```

```
data["Minimum_IELTS_score"] = SimpleImputer(strategy = "median").fit_transform(data[["Minimum_IELTS_score"]])
```

```
display(data[["Minimum_IELTS_score"]].isnull().sum())
```

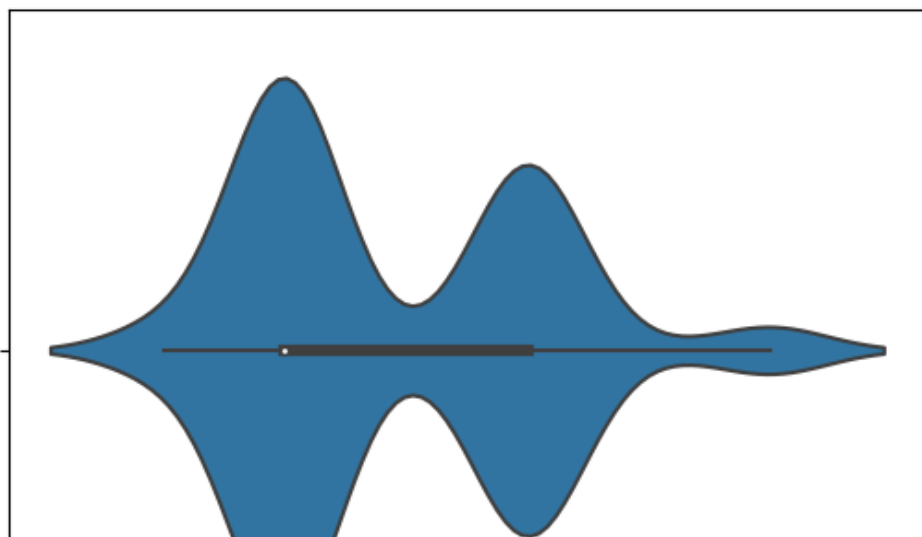
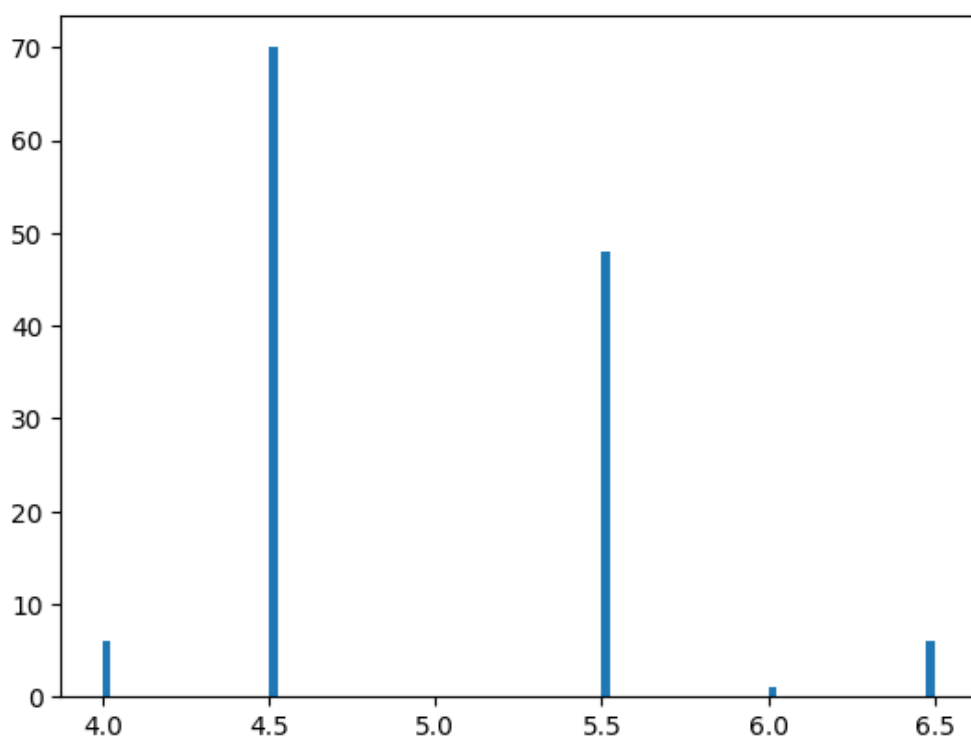
```
Minimum_IELTS_score    0  
dtype: int64
```

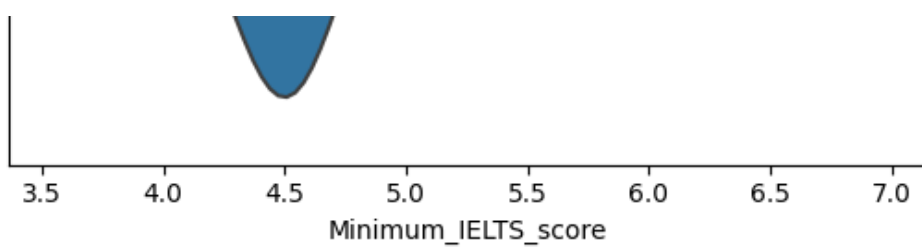
Выполнена импьютация медианным значением столбца **Minimum_IELTS_score**. Число нулевых значений стало **0**.

Задача №27

```
In [16]:
```

```
import matplotlib.pyplot as plt  
import seaborn as sns  
plt.hist(data[["Minimum_IELTS_score"]], 100)  
plt.show()  
sns.violinplot(x=data["Minimum_IELTS_score"]);
```





In [19]:

```
# вычисление 5% и 95% квантилей для столбца 'A'
q05 = data['Minimum_IELTS_score'].quantile(0.05)
q95 = data['Minimum_IELTS_score'].quantile(0.95)
print('\n5% квантиль: ', q05)
print('95% квантиль: ', q95)
```

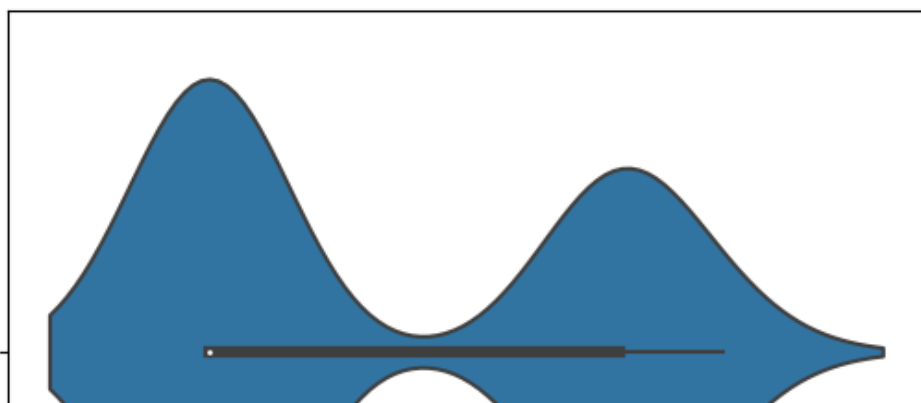
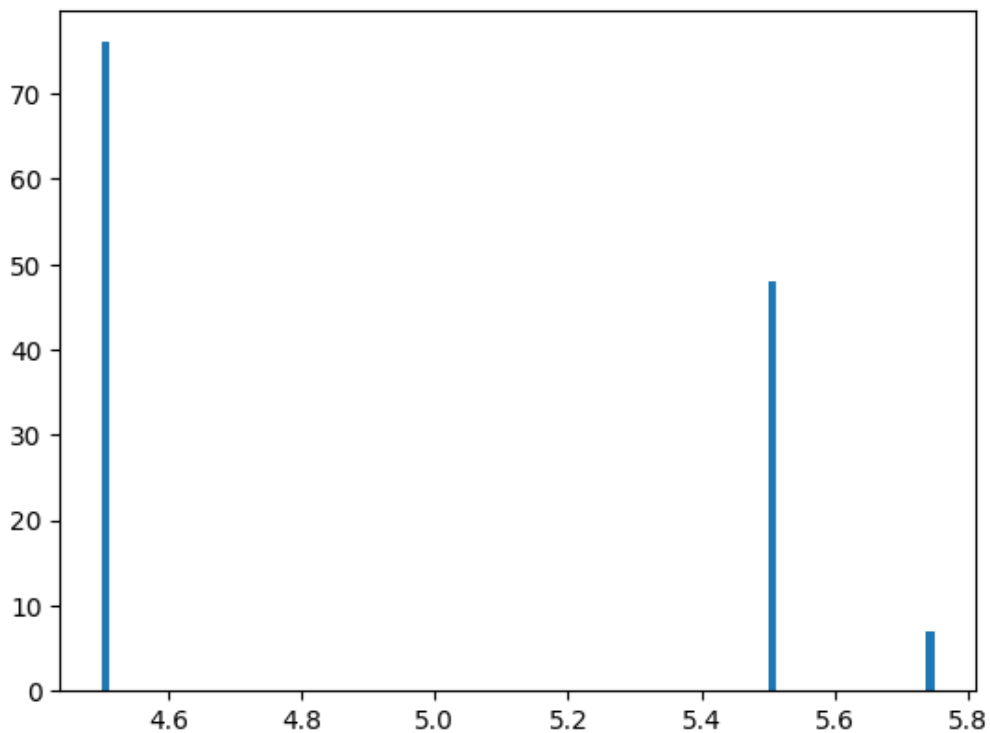
```
5% квантиль: 4.5
95% квантиль: 5.75
```

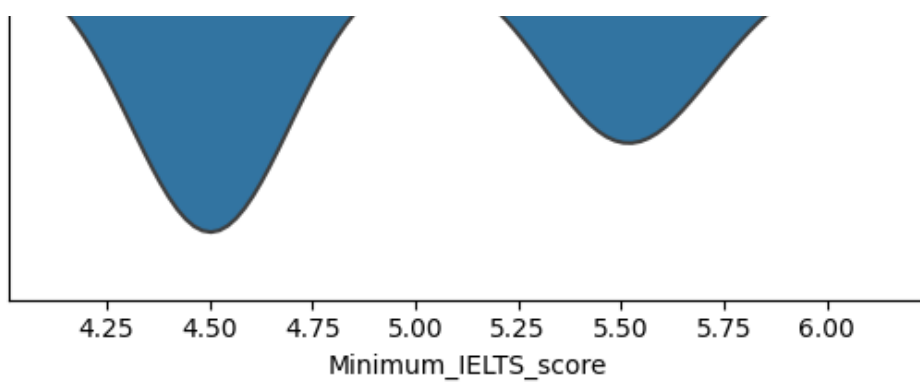
In [21]:

```
data.loc[data['Minimum_IELTS_score'] < q05, 'Minimum_IELTS_score'] = q05
data.loc[data['Minimum_IELTS_score'] > q95, 'Minimum_IELTS_score'] = q95
```

In [22]:

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.hist(data[["Minimum_IELTS_score"]], 100)
plt.show()
sns.violinplot(x=data["Minimum_IELTS_score"]);
```





Доп Задание

In [26]:

```
sns.pairplot(data, hue='Founded_year')
```

Out[26]:

<seaborn.axisgrid.PairGrid at 0x7f8f8bc10dc0>

