

Московский государственный технический университет имени Н.Э.Баумана

Кафедра «Системы обработки информации и управления»

ОТЧЕТ

Лабораторная работа №1  
по дисциплине  
«Методы машинного обучения»  
на тему  
«Разведочный анализ данных. Исследование и визуализация  
данных»

Выполнил:  
Студент ИУ5-24М  
Гаврилюк А.Г.

Москва, 2020

In [14]:

```
import numpy as np
```

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.datasets import load_boston
```

In [21]:

```
boston_ds = load_boston()
data = pd.DataFrame(boston_ds.data)
data.head()
```

Out[21]:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33

In [22]:

```
data.describe()
```

Out[22]:

	0	1	2	3	4	5	6	7	8	9
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407	408.237154
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259	168.537116
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000	279.000000
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000	330.000000
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000	666.000000
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000

In [23]:

```
data.shape
```

Out[23]:

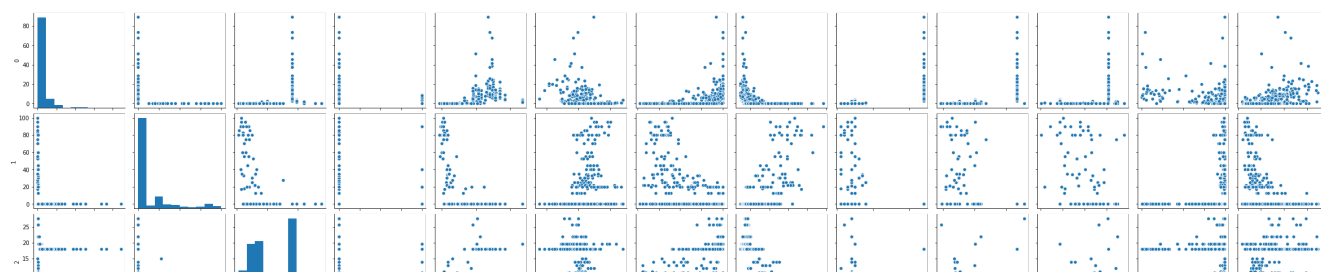
(506, 13)

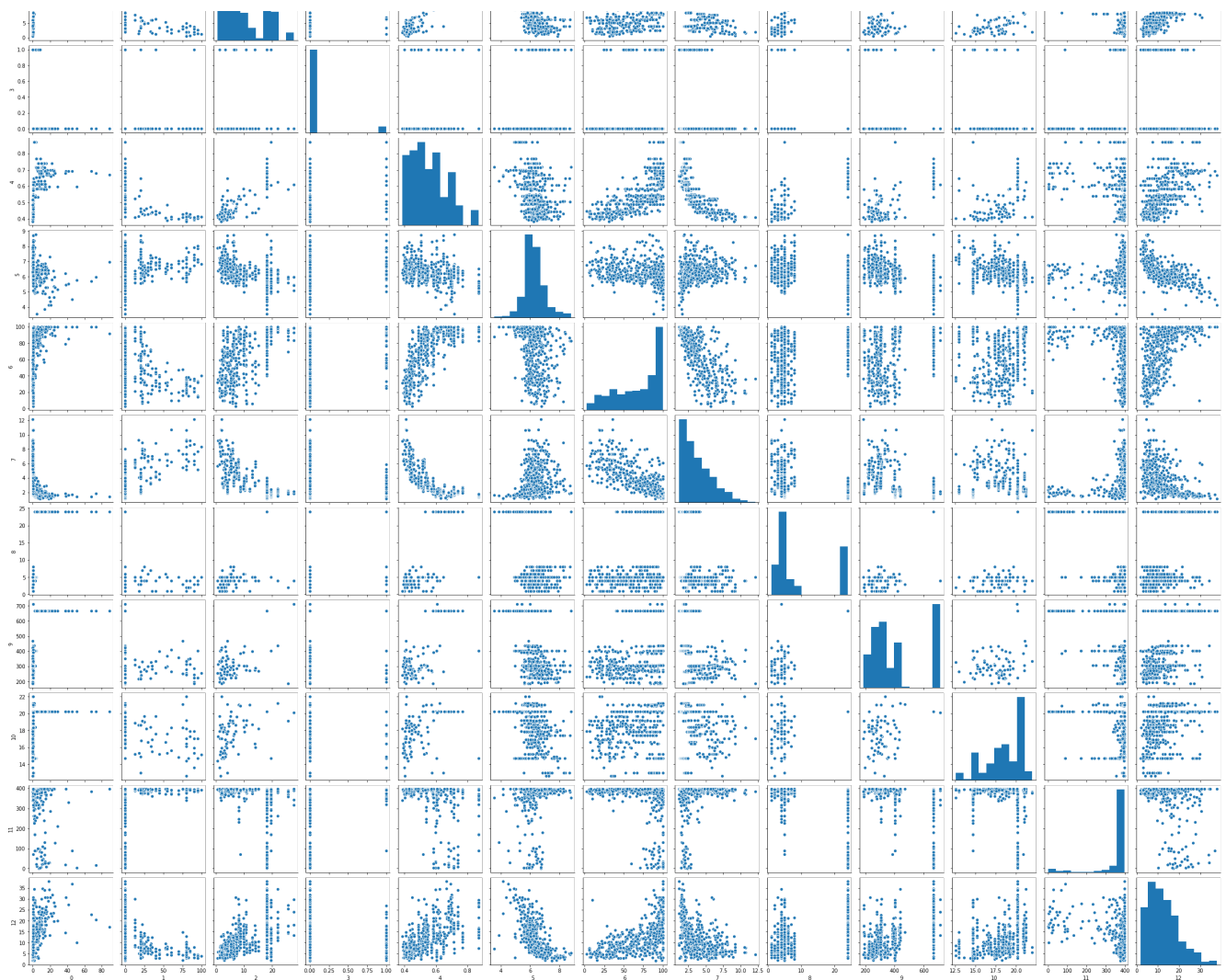
In [24]:

```
sns.pairplot(data)
```

Out[24]:

<seaborn.axisgrid.PairGrid at 0x12ad274d0>



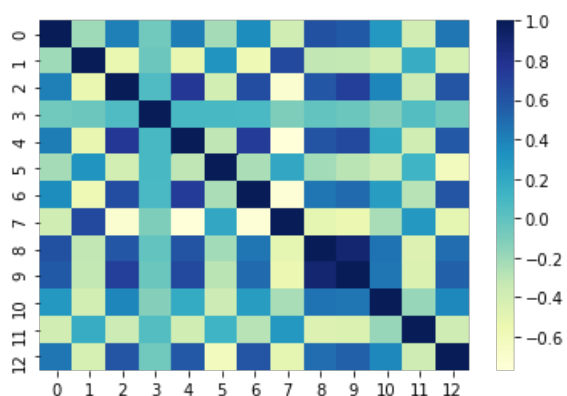


In [33]:

```
corr = data.corr()
# sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns)
sns.heatmap(data.corr(), cmap='YlGnBu')
```

Out[33]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x1322de6d0>

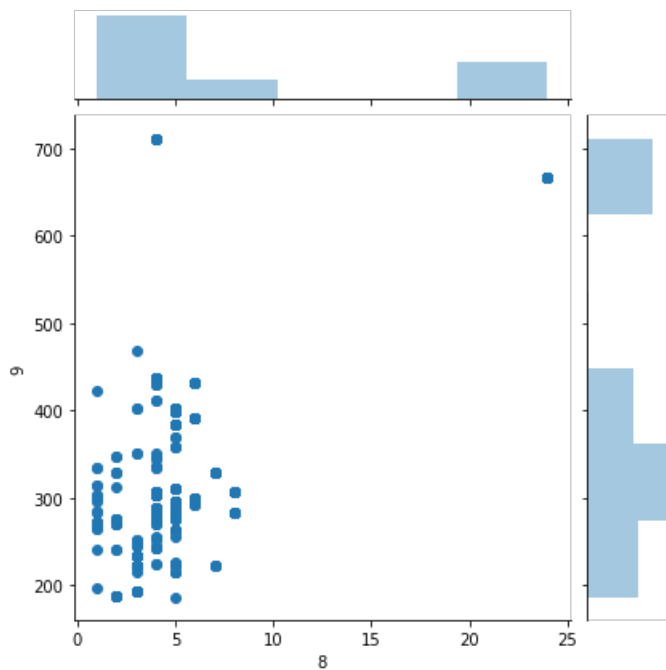


In [26]:

```
sns.jointplot(x=8, y=9, data=data)
```

Out[26]:

<seaborn.axisgrid.JointGrid at 0x1305f9a50>

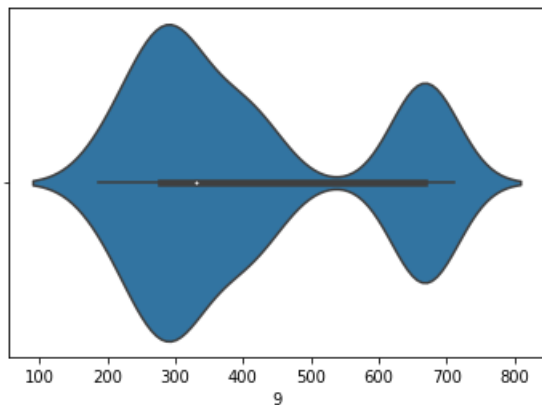


In [28]:

```
sns.violinplot(x=data[9])
```

Out[28]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x131fa7b10>

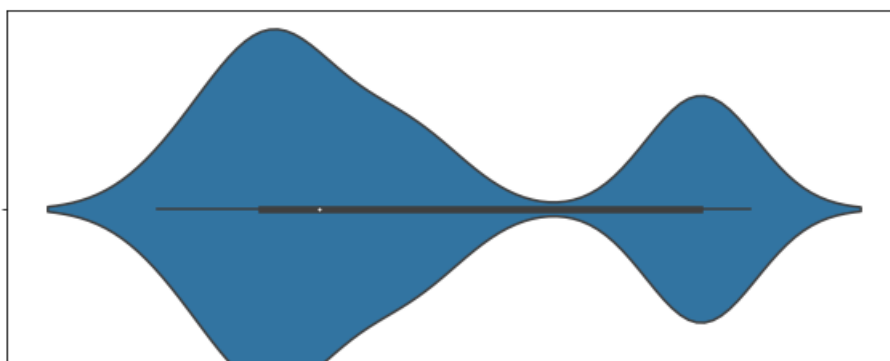


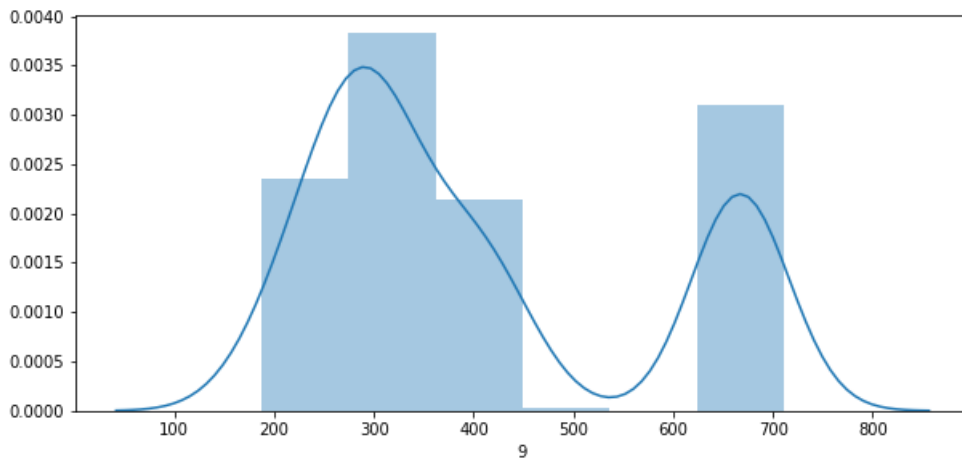
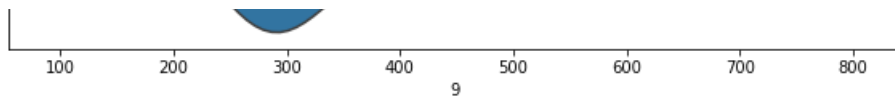
In [34]:

```
fig, ax = plt.subplots(2, 1, figsize=(10,10))
sns.violinplot(ax=ax[0], x=data[9])
sns.distplot(data[9], ax=ax[1])
```

Out[34]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x1324f3390>





In [ ]: