



**Министерство науки и высшего образования Российской Федерации**  
**Федеральное государственное бюджетное образовательное учреждение**  
**высшего образования**  
**«Московский государственный технический университет**  
**имени Н.Э. Баумана**  
**(национальный исследовательский университет)»**  
**(МГТУ им. Н.Э. Баумана)**

---

ФАКУЛЬТЕТ \_\_\_\_\_ Информатика и системы управления \_\_\_\_\_  
КАФЕДРА \_\_\_\_\_ Системы обработки информации и управления \_\_\_\_\_

Рубежный контроль №1  
«Технологии разведочного анализа и обработки данных»  
по курсу «Технологии машинного обучения»

Вариант №26

Выполнил:  
Студент группы ИУ5Ц-81Б  
Булыгина С.А.

\_\_\_\_\_

Проверил:  
Преподаватель кафедры ИУ5  
Гапанюк Ю.Е.

\_\_\_\_\_

---

**Москва 2021**

## Данные варианта:

Номер варианта	Номер задачи	Номер набора данных
26	4	2

## Задача:

Для заданного набора данных постройте основные графики, входящие в этап разведочного анализа данных. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Какие графики Вы построили и почему? Какие выводы о наборе данных Вы можете сделать на основании построенных графиков?

Датасет: [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_iris.html#sklearn.datasets.load\\_iris](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris)

## Дополнительное задание:

Для пары произвольных колонок данных построить график «Диаграмма рассеяния».

## Выполнение рубежного контроля:

```
In [3]: 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
from sklearn import datasets
from sklearn.datasets import load_iris
from sklearn import linear_model
from sklearn.cluster import KMeans
from sklearn import metrics
from pandas import DataFrame
%pylab inline

Populating the interactive namespace from numpy and matplotlib
```

```
In [4]: boston = load_iris()
data = pd.DataFrame(boston.data, columns=boston.feature_names)
data['TARGET'] = boston.target
```

```
In [6]: data.head()
```

```
Out[6]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	TARGET
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.8	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```
In [7]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype  
---  --
 0   sepal length (cm)    150 non-null   float64
 1   sepal width (cm)     150 non-null   float64
 2   petal length (cm)    150 non-null   float64
 3   petal width (cm)     150 non-null   float64
 4   TARGET               150 non-null   int32   
dtypes: float64(4), int32(1)
memory usage: 5.3 KB
```

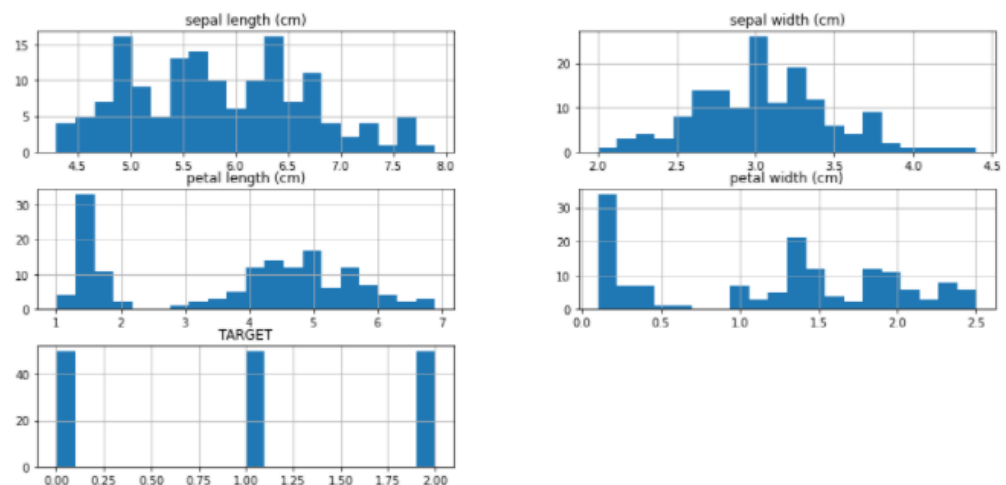
```
In [8]: data.describe()
```

```
Out[8]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	TARGET
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.198333	1.000000
std	0.828066	0.435866	1.765298	0.762238	0.819232
min	4.300000	2.000000	1.000000	0.100000	0.000000
25%	5.100000	2.800000	1.600000	0.300000	0.000000
50%	5.800000	3.000000	4.350000	1.300000	1.000000
75%	6.400000	3.300000	5.100000	1.800000	2.000000
max	7.900000	4.400000	6.900000	2.500000	2.000000

```
In [12]: data.hist(bins=20, figsize = (15,7))
```

```
Out[12]: array([[<AxesSubplot:title={'center':'sepal length (cm)'}>,
  <AxesSubplot:title={'center':'sepal width (cm)'}>],
  [<AxesSubplot:title={'center':'petal length (cm)'}>,
  <AxesSubplot:title={'center':'petal width (cm)'}>],
  [<AxesSubplot:title={'center':'TARGET'}>], <AxesSubplot:>]],
  dtype=object)
```

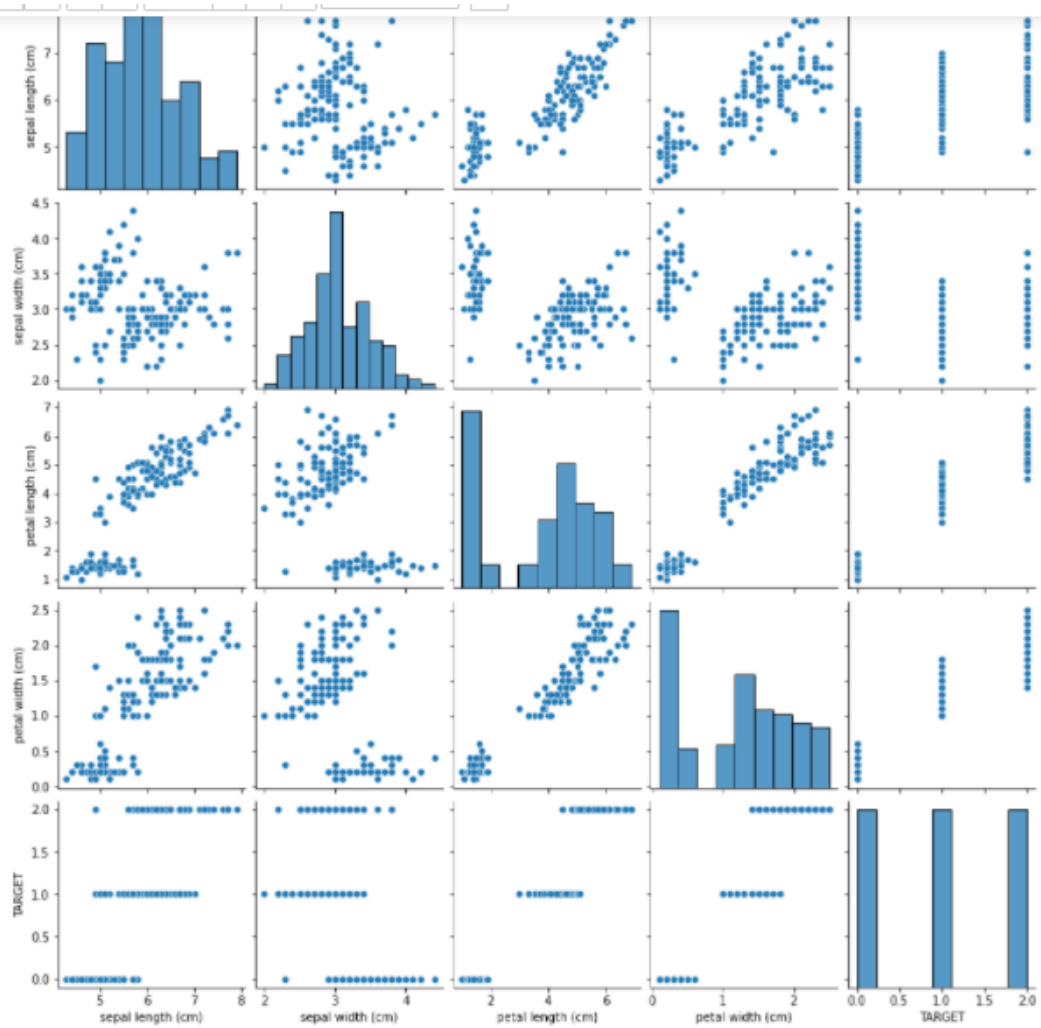


```
In [14]: # Диаграммы рассеяния для всех признаков
plt.figure(figsize=(10,5))
sns.pairplot(data)
```

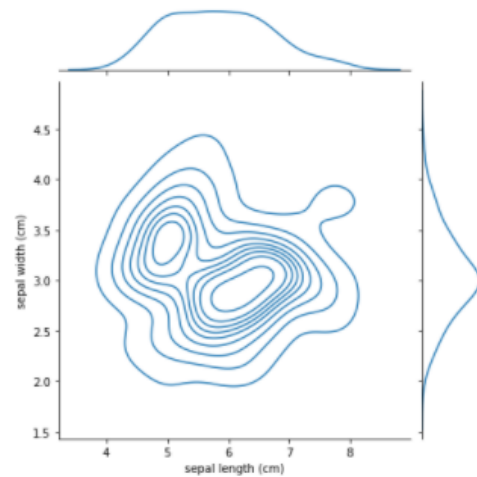
```
Out[14]: <seaborn.axisgrid.PairGrid at 0x1399df28>
```

```
<Figure size 720x360 with 0 Axes>
```

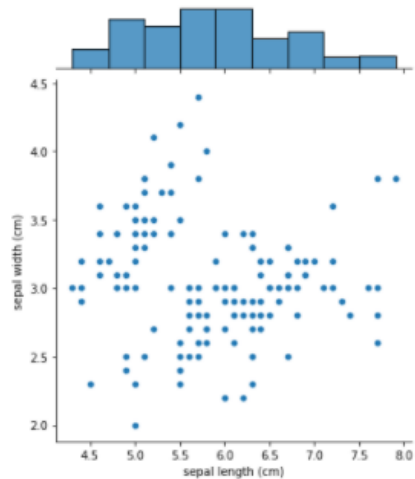




```
In [16]: sns.jointplot(x = "sepal length (cm)", y = "sepal width (cm)", data = data, kind="kde")
Out[16]: <seaborn.axisgrid.JointGrid at 0x15976b98>
```



```
In [19]: sns.jointplot(x = "sepal length (cm)", y = "sepal width (cm)", data=data, kind="scatter")
Out[19]: <seaborn.axisgrid.JointGrid at 0x16047640>
```



**Выполним дополнительное задание: для пары произвольных колонок данных построим график «Диаграмма рассеяния», используя колонки `sepal length (cm)` и `sepal width (cm)`**

```
In [20]: fig, ax = plt.subplots(figsize=(8,8))
sns.scatterplot(ax=ax, x = "sepal length (cm)", y = "sepal width (cm)", data = data)
Out[20]: <AxesSubplot:xlabel='sepal length (cm)', ylabel='sepal width (cm)'\>
```

