

**Московский государственный технический
университет им. Н.Э. Баумана**

Факультет Радиотехнический
Кафедра РТ5

Курс «Сети и телекоммуникации»

Отчет по лабораторной работе №1

«Разведочный анализ данных.

Исследование и визуализация данных»

Выполнил:

студент группы РТ5-61Б:

Бабасанова Н. С.

Руководитель:

преподаватель каф. ИУ5

Нардид А. Н.

Москва, 2025г.

Задание:

- Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь](#).
- Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из [Scikit-learn](#).
- Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть [здесь](#).

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

- Создать ноутбук, который содержит следующие разделы:
 1. Текстовое описание выбранного Вами набора данных.
 2. Основные характеристики датасета.
 3. Визуальное исследование датасета.
 4. Информация о корреляции признаков.
- Сформировать отчет и разместить его в своем репозитории на github.

Текст программы и формы с примерами выполнения программы:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
iris_sk = load_iris()

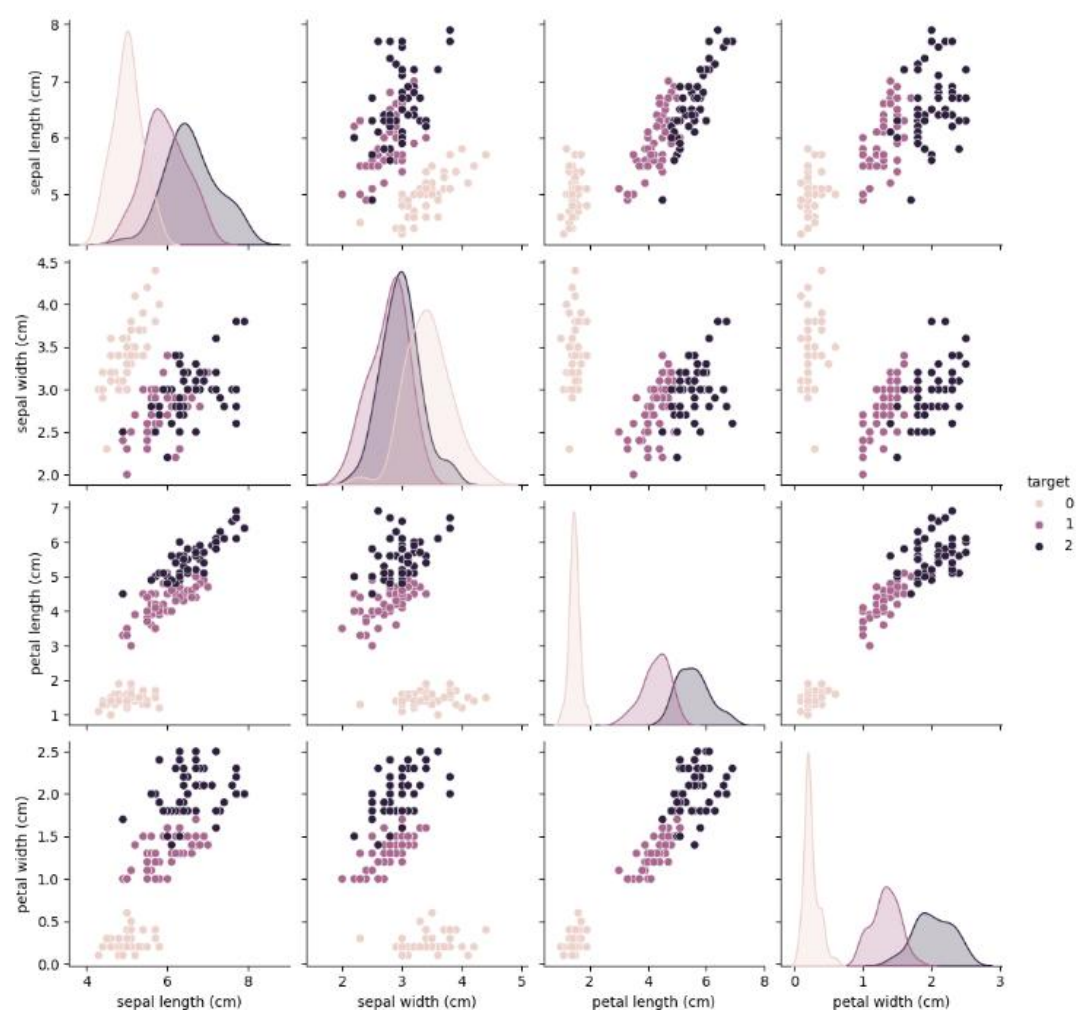
df = pd.DataFrame(data = iris_sk.data, columns = iris_sk.feature_names)
df["target"] = iris_sk.target
df
```

	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	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

150 rows × 5 columns

Визуализация:

```
sns.pairplot(df, hue="target")
```



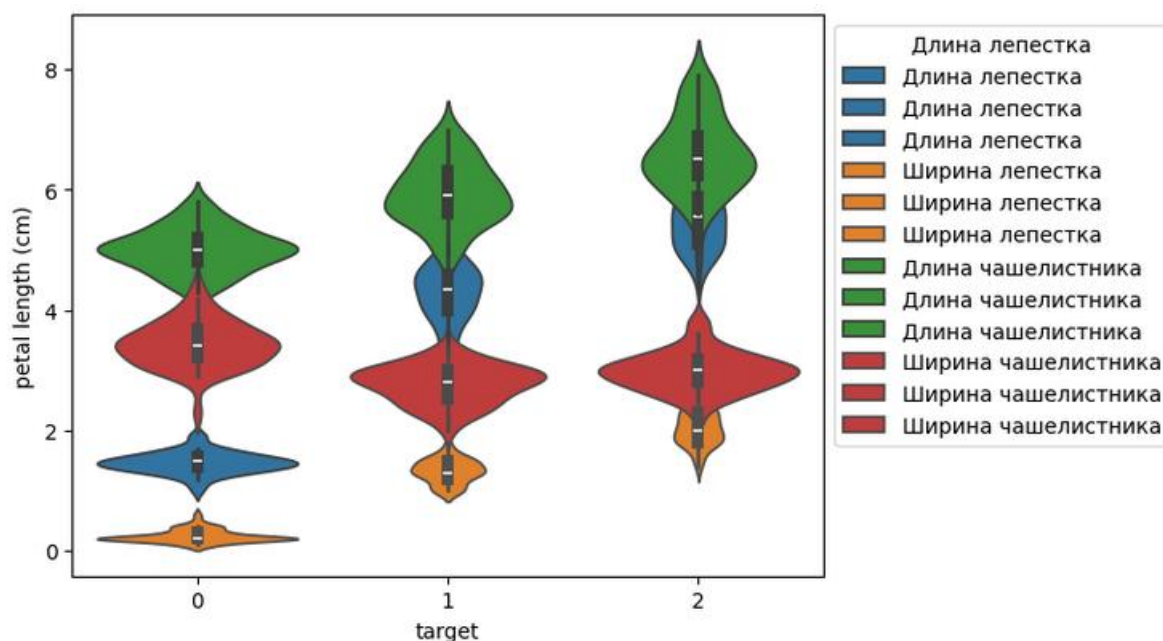
```
# Распределение параметра Petal Length, width и sepal l, w
сгруппированные по Target.

sns.violinplot(x='target', y= 'petal length (cm)', data=df,
label="Длина лепестка")

sns.violinplot(x='target', y= 'petal width (cm)', data=df,
label="Ширина лепестка")

sns.violinplot(x='target', y= 'sepal length (cm)', data=df,
label="Длина чашелистника")

sns.violinplot(x='target', y= 'sepal width (cm)', data=df,
label="Ширина чашелистника")
```



Корреляция признаков:

```
df.corr()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941	0.782561
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126	-0.426658
petal length (cm)	0.871754	-0.428440	1.000000	0.962865	0.949035
petal width (cm)	0.817941	-0.366126	0.962865	1.000000	0.956547
target	0.782561	-0.426658	0.949035	0.956547	1.000000

```
df.corr()["target"]
```

	target
sepal length (cm)	0.782561
sepal width (cm)	-0.426658
petal length (cm)	0.949035
petal width (cm)	0.956547
target	1.000000