**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: |  | Руководитель: |
| студент группы РТ5-61Б: |  | преподаватель каф. ИУ5 |
| Бабасанова Н. С. |  | Нардид А. Н. |

Москва, 2025г.

**Задание:**

* Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь.](https://github.com/ugapanyuk/courses_current/wiki/DSLIST)
* Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из [Scikit-learn.](https://scikit-learn.org/stable/datasets/toy_dataset.html)
* Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть [здесь.](https://github.com/ugapanyuk/courses_current/blob/main/notebooks/ds/sklearn_datasets.ipynb)

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

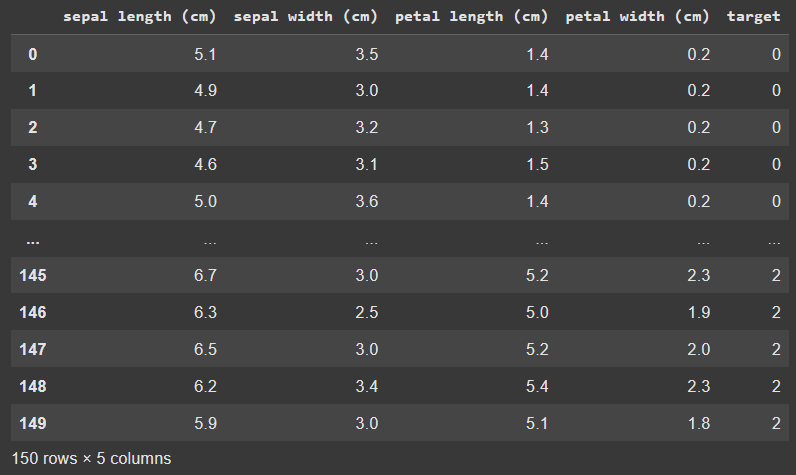
from sklearn.datasets import load\_iris

iris\_sk = load\_iris()

df = pd.DataFrame(data = iris\_sk.data, columns = iris\_sk.feature\_names)

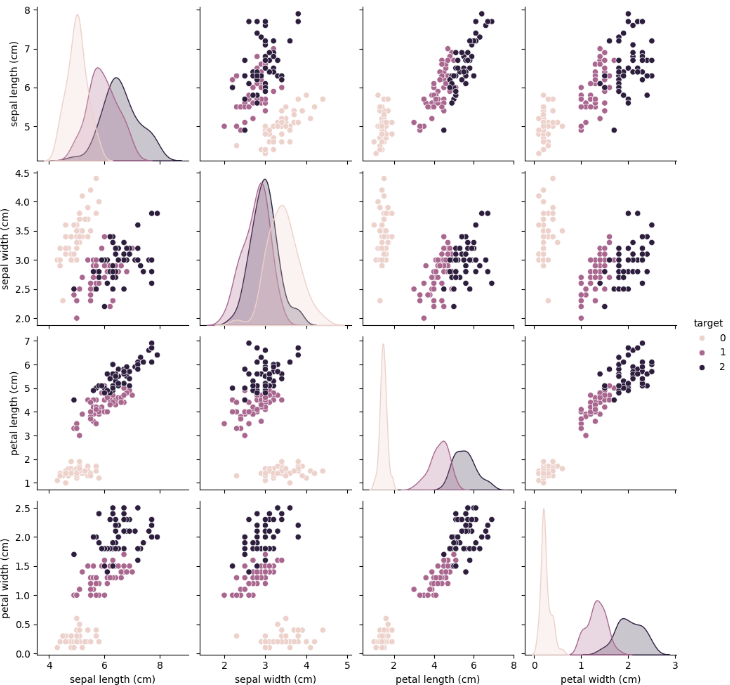
df["target"] = iris\_sk.target

df



Визуализация:

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



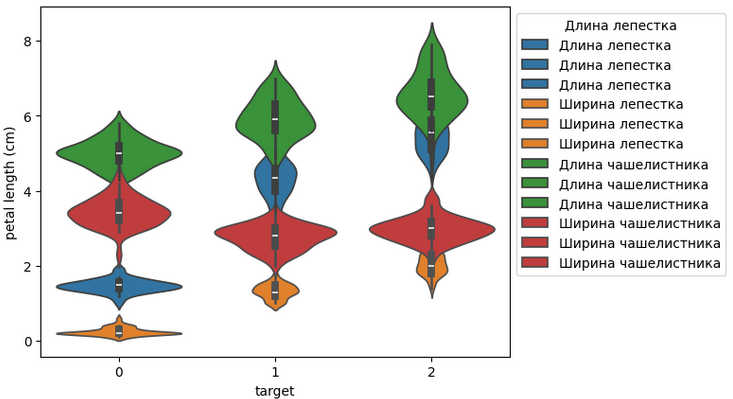
# Распределение параметра Petal Length, width и sepal l, w сгруппированные по Target.

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")

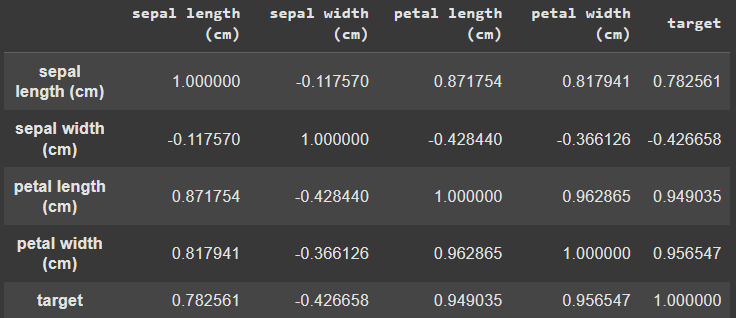
sns.violinplot(x='target', y= 'sepal length (cm)', data=df, label="Длина чашелистника")

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



Корреляция признаков:

df.corr()



df.corr()["target"]

