

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
“БРЕСТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ”

ФАКУЛЬТЕТ ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ

Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №1

Специальность АС-66

Выполнил
А. С. Езепчук,
студент группы АС-66

Проверил
А. А. Крощенко,
ст. преп. кафедры ИИТ,
«___» _____ 2025 г.

Брест 2025

Цель работы: Получить практические навыки работы с данными с использованием библиотек **Pandas** для манипуляции и **Matplotlib** для визуализации. Научиться выполнять основные шаги предварительной обработки данных, такие как очистка, нормализация и работа с различными типами признаков.

Вариант 3.

Выборка Iris. Классический набор данных для классификации, содержащий измерения длины и ширины чашелистиков и лепестков для трех видов ирисов.

Задачи:

1. Загрузите данные и проверьте, есть ли в них пропущенные значения.

```
Проверка пропущенных значений:
```

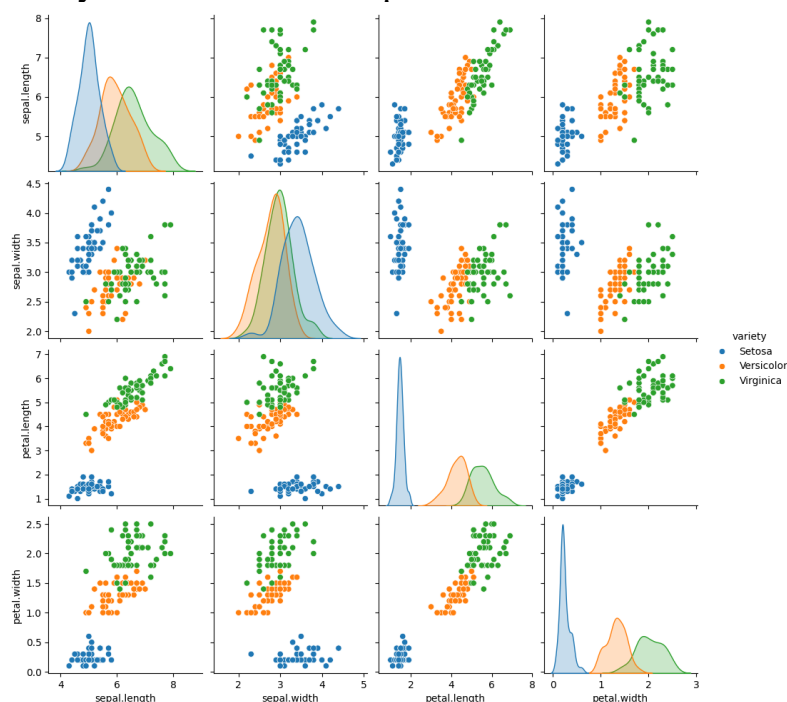
```
sepal.length    0
sepal.width     0
petal.length    0
petal.width     0
variety         0
dtype: int64
```

2. Выведите количество образцов каждого вида ириса.

```
Количество образцов по каждому виду:
```

```
variety
Setosa      50
Versicolor  50
Virginica   50
Name: count, dtype: int64
```

3. Постройте парные диаграммы рассеяния (pair plot) для всех признаков, чтобы визуально оценить их разделимость.

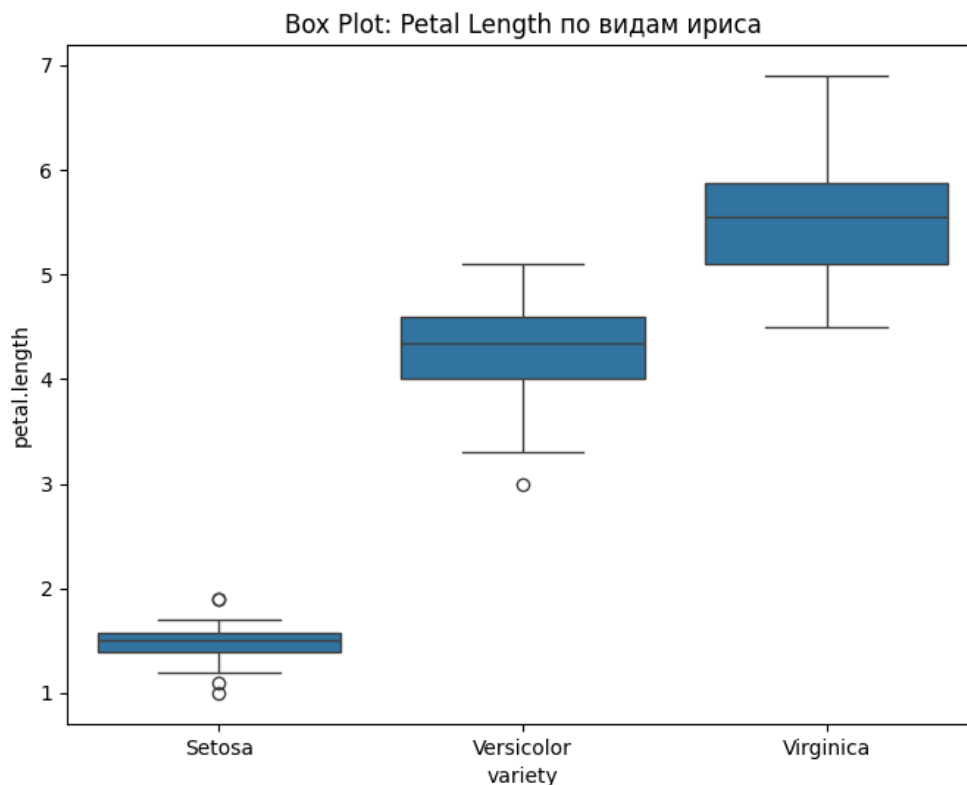


4. Для каждого вида ириса рассчитайте среднее значение по каждому из четырех признаков.

Средние значения признаков по каждому виду:

	sepal.length	sepal.width	petal.length	petal.width
variety				
Setosa	5.006	3.428	1.462	0.246
Versicolor	5.936	2.770	4.260	1.326
Virginica	6.588	2.974	5.552	2.026

5. Создайте "ящик с усами" (box plot) для признака Petal Length (cm), чтобы сравнить его распределение по разным видам ирисов.



6. Стандартизируйте данные (приведите к нулевому среднему и единичному стандартному отклонению)

Стандартизованные данные:

	sepal.length	sepal.width	petal.length	petal.width	variety
0	-0.897674	1.015602	-1.335752	-1.311052	Setosa
1	-1.139200	-0.131539	-1.335752	-1.311052	Setosa
2	-1.380727	0.327318	-1.392399	-1.311052	Setosa
3	-1.501490	0.097889	-1.279104	-1.311052	Setosa
4	-1.018437	1.245030	-1.335752	-1.311052	Setosa
5	-0.535384	1.933315	-1.165809	-1.048667	Setosa

Код программы:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```

iris = pd.read_csv("iris.csv")

missing = iris.isnull().sum()
counts = iris["variety"].value_counts()
means = iris.groupby("variety").mean(numeric_only=True)

features = ["sepal.length", "sepal.width", "petal.length",
"petal.width"]
iris_scaled_df = iris.copy()
iris_scaled_df[features] = (iris[features] - iris[features].mean()) /
iris[features].std()

iris_encoded = pd.get_dummies(iris_scaled_df, columns=["variety"])

with open("iris_report.txt", "w", encoding="utf-8") as f:
    f.write("Исходные данные (все строки):\n")
    f.write(iris.to_string() + "\n\n")
    f.write("Проверка пропущенных значений:\n")
    f.write(str(missing) + "\n\n")
    f.write("Количество образцов по каждому виду:\n")
    f.write(str(counts) + "\n\n")
    f.write("Средние значения признаков по каждому виду:\n")
    f.write(str(means) + "\n\n")
    f.write("Стандартизованные данные:\n")
    f.write(iris_scaled_df.to_string() + "\n\n")
    f.write("Данные после One-Hot Encoding (первые 5 строк):\n")
    f.write(iris_encoded.head().to_string() + "\n\n")

print("Отчёт сохранён в iris_report.txt")

sns.pairplot(iris, hue="variety", diag_kind="kde")
plt.suptitle("Pair Plot признаков Iris", y=1.02)
plt.savefig("pairplot.png")
plt.close()

plt.figure(figsize=(8, 6))
sns.boxplot(x="variety", y="petal.length", data=iris)
plt.title("Box Plot: Petal Length по видам ириса")
plt.savefig("boxplot.png")
plt.close()

```

Вывод: научился разрабатывать простые программы на Python с использованием библиотек **Pandas** для манипуляции и **Matplotlib** для визуализации, получил практический опыт работы с данными.