Министерство образования Республики Беларусь  
Учреждение образования

«Брестский государственный технический университет»

ФАКУЛЬТЕТ ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ  
Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №1

Выполнил

В.Д.Головкина,

студент группы АС66

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 1

Выборка Titanic. Содержит информацию о пассажирах лайнера, включая их

возраст, пол, класс каюты и факт выживания.

Задачи:

1. Загрузите данные и выведите первые 5 строк, а также общую

информацию о столбцах (.info()).

2. Найдите и визуализируйте количество выживших и погибших

пассажиров с помощью столбчатой диаграммы.

3. Обработайте пропуски в столбце Age, заполнив их медианным

значением.

4. Преобразуйте категориальные признаки Sex и Embarked в числовые с

помощью One-Hot Encoding.

5. Постройте гистограмму распределения возрастов пассажиров.

6. Создайте новый признак FamilySize путем сложения значений из

столбцов SibSp и Parch.

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from pathlib import Path

iris = sns.load\_dataset("iris")

iris.rename(columns={

    "sepal\_length":  "sepal.length",

    "sepal\_width":   "sepal.width",

    "petal\_length":  "petal.length",

    "petal\_width":   "petal.width",

    "species":       "variety"

}, inplace=True)

print("Shape:", iris.shape)

print("Columns:", iris.columns)

print(iris.head())

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"])

report\_path = Path(\_\_file\_\_).parent / "iris\_report.txt"

with report\_path.open("w", encoding="utf-8") as f:

    f.write("Исходные данные (все строки):\n")

    f.write(iris.to\_string(index=False) + "\n\n")

    f.write("Проверка пропущенных значений:\n")

    f.write(missing.to\_string() + "\n\n")

    f.write("Количество образцов по каждому виду:\n")

    f.write(counts.to\_string() + "\n\n")

    f.write("Средние значения признаков по каждому виду:\n")

    f.write(means.to\_string() + "\n\n")

    f.write("Стандартизованные данные (первые 5 строк):\n")

    f.write(iris\_scaled\_df.head().to\_string(index=False) + "\n\n")

    f.write("One-Hot Encoding (первые 5 строк):\n")

    f.write(iris\_encoded.head().to\_string(index=False) + "\n\n")

print(f"Отчёт сохранён в {report\_path.name}")

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")

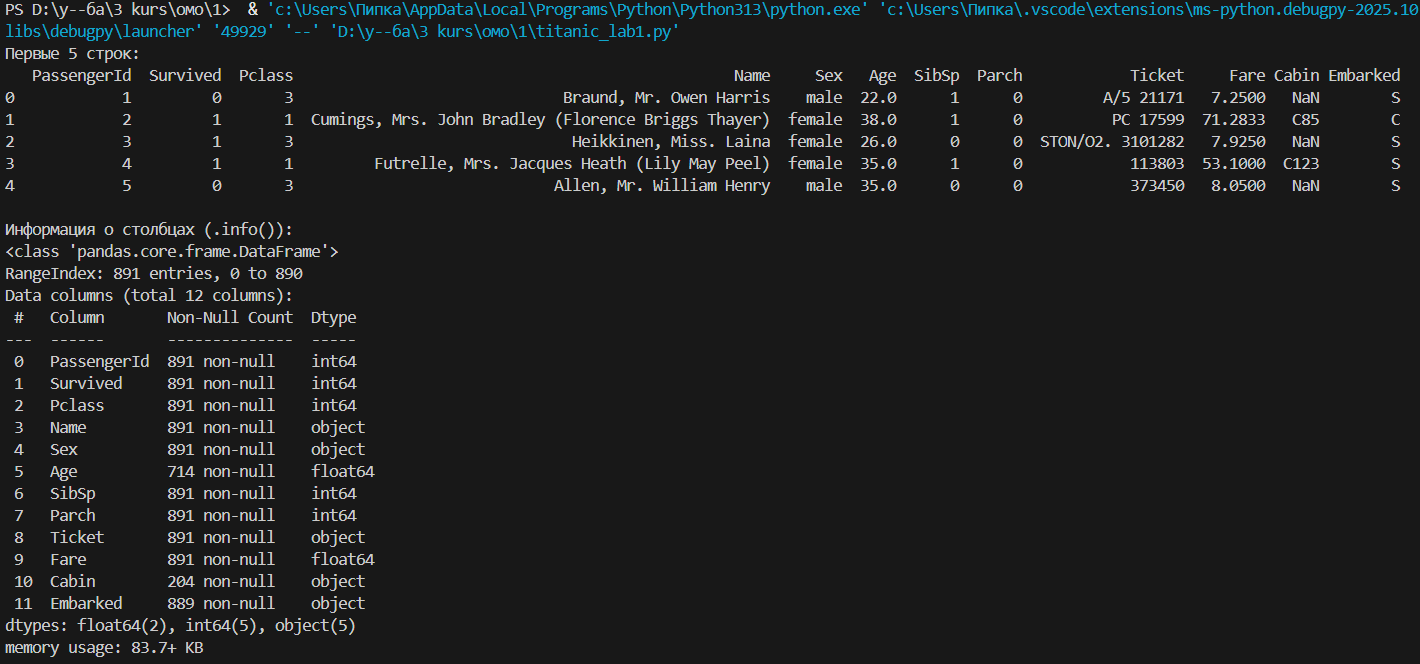
plt.savefig("boxplot.png")

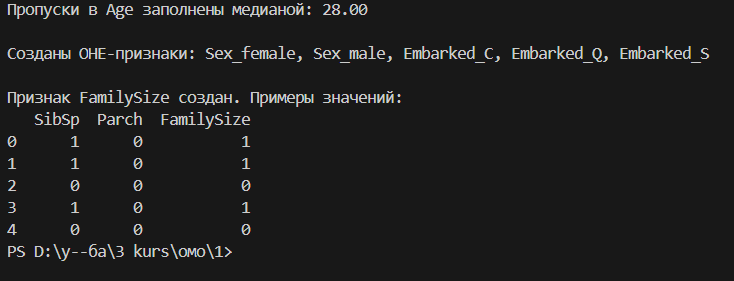
plt.close()

print("Графики сохранены в pairplot.png и boxplot.png")









Вывод: я получила практические навыки работы с данными с использованием

библиотек Pandas для манипуляции и Matplotlib для визуализации.

Научилась выполнять основные шаги предварительной обработки данных,

такие как очистка, нормализация и работа с различными типами признаков.