Министерство образования Республики Беларусь

Учреждение образования

«Брестский Государственный технический университет»

Кафедра ИИТ

Лабораторная работа №1

По дисциплине: «Основы машинного обучения»

Тема: «Знакомство с анализом данных: предварительная обработка и визуализация»

Выполнил:

Студент 3 курса Группы АС-65

Зинчук М.С.

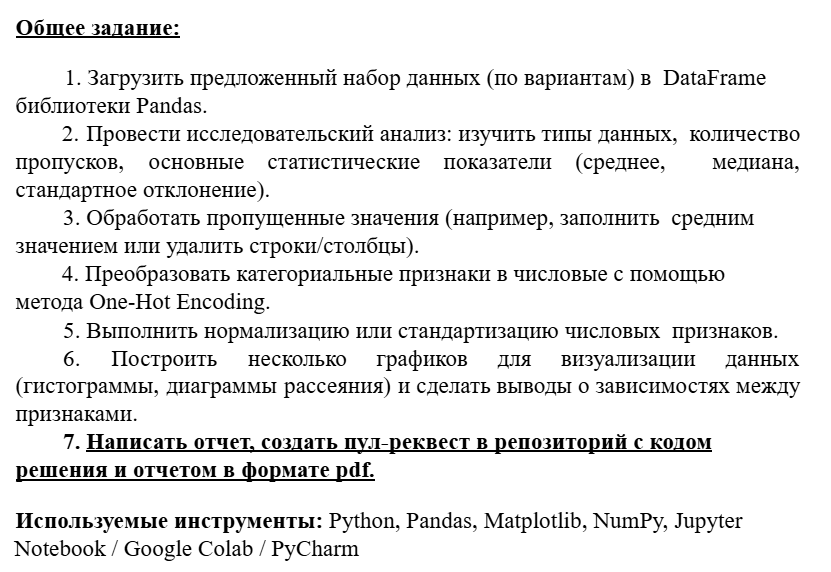
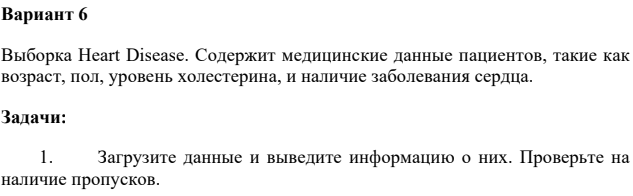
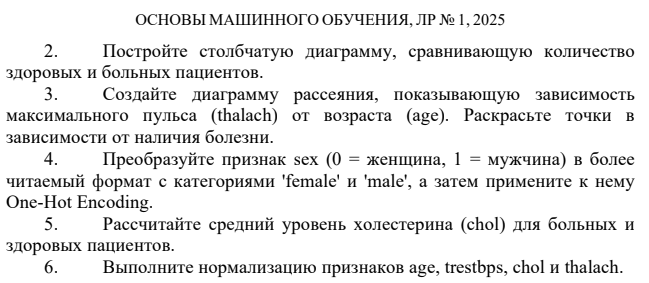
Проверил:

Крощенко А. А.

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.

Ход работы

Код программы:

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import MinMaxScaler

df = pd.read\_csv('heart.csv')

print("1. ИНФОРМАЦИЯ О ДАННЫХ:")

print(df.info())

print("\nПропуски в данных:")

print(df.isnull().sum())

print("\nОСНОВНЫЕ СТАТИСТИКИ:")

print(df.describe())

print("\nМедианы:")

print(df.median(numeric\_only=True))

print("\nСтандартные отклонения:")

print(df.std(numeric\_only=True))

if df.isnull().sum().sum() == 0:

print("\nПропусков нет, обработка не требуется.")

else:

df = df.fillna(df.mean(numeric\_only=True))

print("\nПропуски заполнены средними значениями.")

plt.figure(figsize=(6, 4))

df['target'].value\_counts().plot(kind='bar')

plt.title('Количество здоровых и больных пациентов')

plt.xlabel('Target (0-здоров, 1-болен)')

plt.ylabel('Количество')

plt.xticks(rotation=0)

plt.show()

print("Вывод: больных пациентов больше, чем здоровых.")

plt.figure(figsize=(8, 6))

colors = ['red' if x == 1 else 'blue' for x in df['target']]

plt.scatter(df['age'], df['thalach'], c=colors, alpha=0.6)

plt.title('Зависимость пульса от возраста')

plt.xlabel('Возраст')

plt.ylabel('Максимальный пульс')

plt.legend(['Больные', 'Здоровые'])

plt.show()

print("Вывод: у более молодых пациентов чаще встречается более высокий пульс.")

df['sex'] = df['sex'].map({0: 'female', 1: 'male'})

sex\_encoded = pd.get\_dummies(df['sex'], prefix='sex')

df = pd.concat([df, sex\_encoded], axis=1)

print("\n4. Результат One-Hot Encoding:")

print(df[['sex', 'sex\_female', 'sex\_male']].head())

mean\_chol\_sick = df[df['target'] == 1]['chol'].mean()

mean\_chol\_healthy = df[df['target'] == 0]['chol'].mean()

print(f"\n5. Средний уровень холестерина:")

print(f"Больные: {mean\_chol\_sick:.2f}")

print(f"Здоровые: {mean\_chol\_healthy:.2f}")

scaler = MinMaxScaler()

df[['age', 'trestbps', 'chol', 'thalach']] = scaler.fit\_transform(

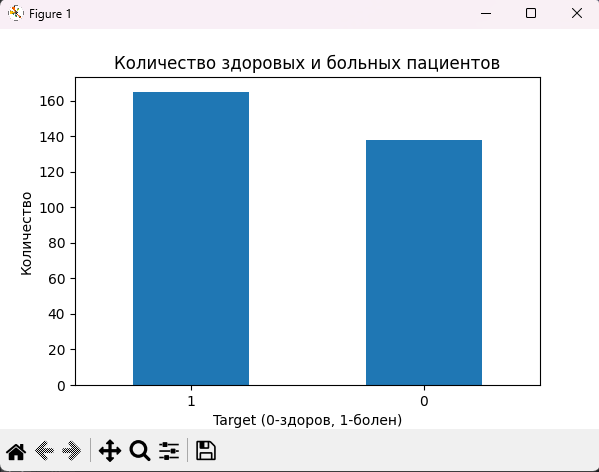
df[['age', 'trestbps', 'chol', 'thalach']]

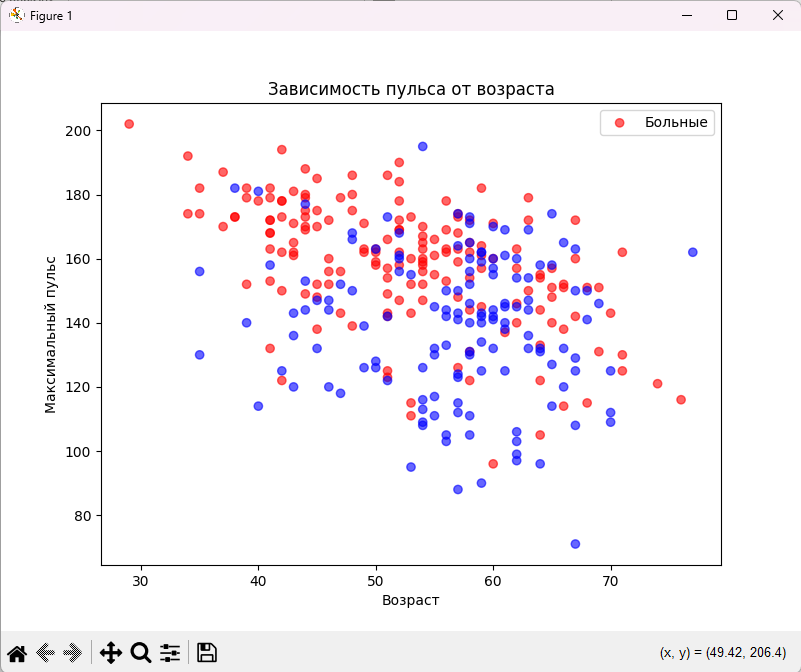
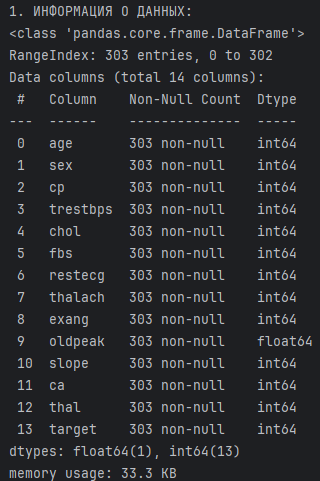
)

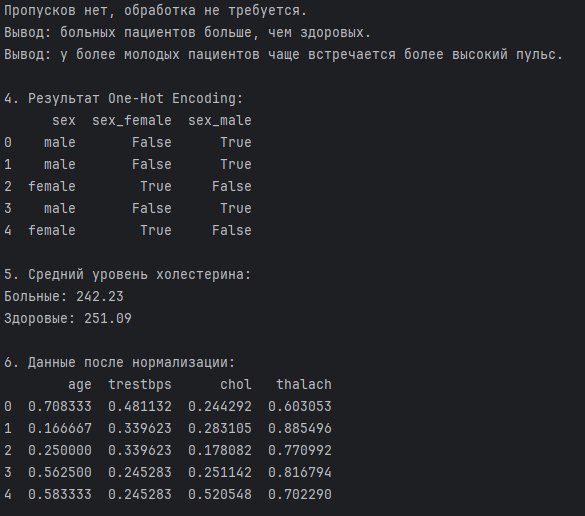
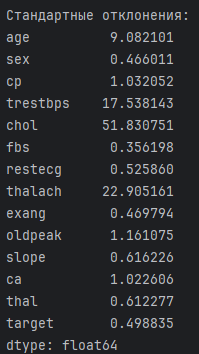
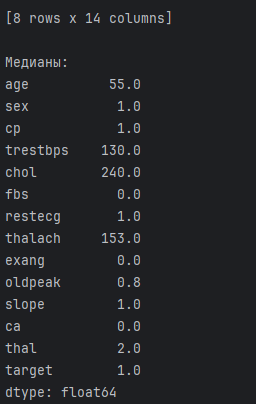
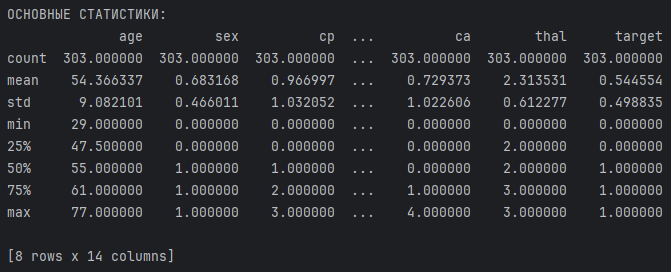
print("\n6. Данные после нормализации:")

print(df[['age', 'trestbps', 'chol', 'thalach']].head())

Графики:



  
  
Вывод: в результате выполнения данной лабораторной работы получили практические навыки работы с данными с использованием библиотек Pandas для манипуляции и Matplotlib для визуализации.