

Министерство образования Республики Беларусь
Учреждение образования
«Брестский государственный технический университет»
ФАКУЛЬТЕТ ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ
Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №1

Выполнил
М.С.Николова,
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 6

Задание 1. Загрузите данные и выведите информацию о них. Проверьте на наличие пропусков.

```
import pandas as pd
pd.set_option('display.max_rows',None)
pd.set_option('display.max_columns',None)
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')
print(df)
print("Есть ли пропуски: ", df.isnull().values.any())
```

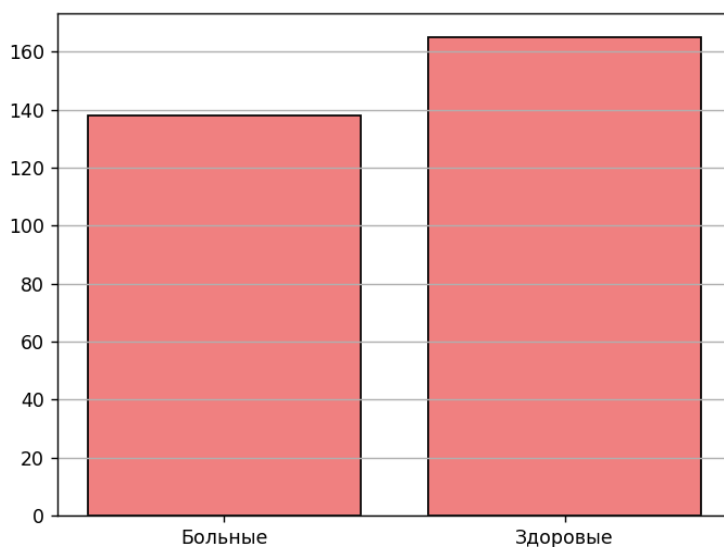
```
Есть ли пропуски: False
```

Задание 2. Постройте столбчатую диаграмму, сравнивающую количество здоровых и больных пациентов.

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')
```

```
counts = df['target'].value_counts()
```

```
plt.bar(counts.index, counts.values, color='lightcoral', edgecolor='black')
plt.xticks([0,1],['Больные','Здоровые'])
plt.grid(axis='y')
plt.show(block=True)
```



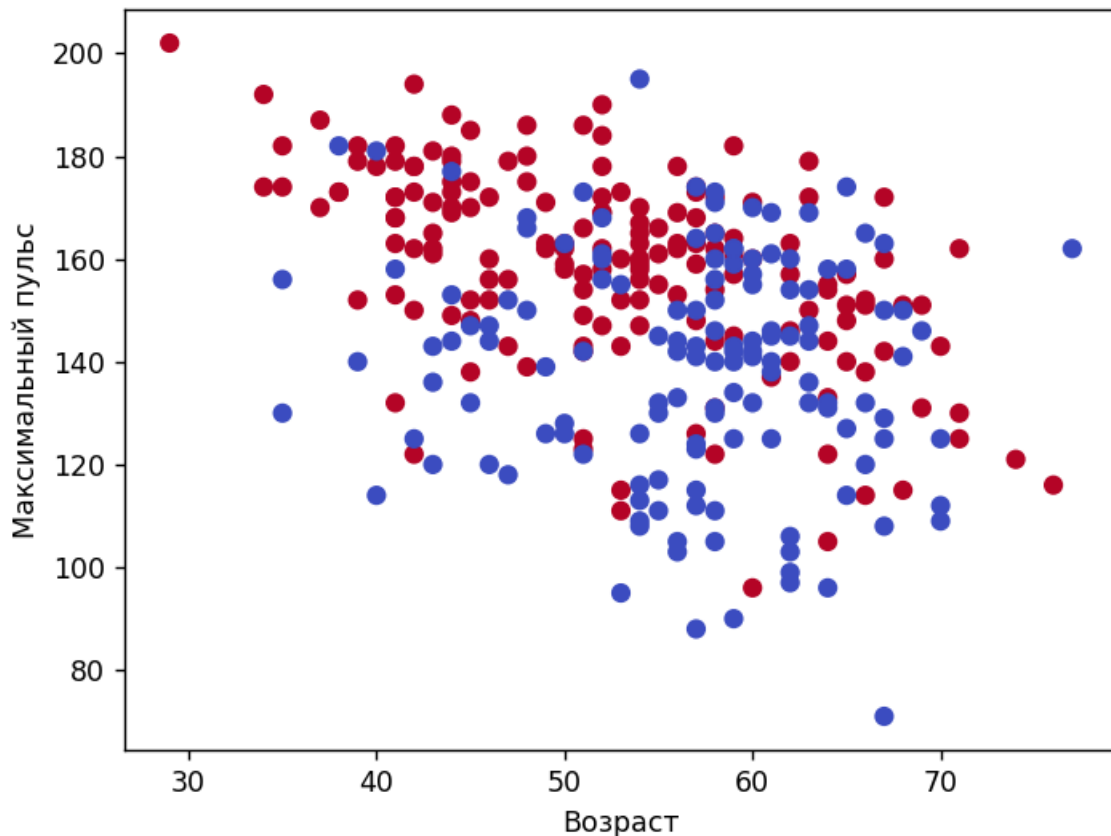
Задание 3. Создайте диаграмму рассеяния, показывающую зависимость максимального пульса (thalach) от возраста (age). Раскрасьте точки в зависимости от наличия болезни.

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')
```

```
plt.scatter(df['age'], df['thalach'], c=df['target'], cmap='coolwarm')
```

```
plt.xlabel('Возраст')
```

```
plt.ylabel('Максимальный пульс')
plt.show(block=True)
```



Задание 4. Преобразуйте признак sex (0 = женщина, 1 = мужчина) в более читаемый формат с категориями 'female' и 'male', а затем примените к нему One-Hot Encoding.

```
import pandas as pd
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')
df['sex'] = df['sex'].map({0: 'female', 1: 'male'})
df_ = pd.get_dummies(df, columns=['sex'])
df_ = df_.astype(int)
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
print(df_)
```

sex_female	sex_male
0	1
0	1
1	0
0	1
1	0
0	1
1	0
0	1
0	1
0	1
0	1
1	0
0	1
0	1
1	0
1	0
1	0

Задание 5. Рассчитайте средний уровень холестерина (chol) для больных и здоровых пациентов.

```
import pandas as pd
df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')
chol_diseased = df[df['target'] == 1]['chol'].mean()
chol_healthy = df[df['target'] == 0]['chol'].mean()
print(f"Средний уровень холестерина у больных пациентов: {chol_diseased:.2f}")
print(f"Средний уровень холестерина у здоровых пациентов: {chol_healthy:.2f}")
Средний уровень холестерина у больных пациентов: 242.23
Средний уровень холестерина у здоровых пациентов: 251.09
```

Задание 6. Выполните нормализацию признаков age, trestbps, chol и thalach.

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

df = pd.read_csv('C:\\Users\\z3594\\OneDrive\\Документы\\лабы\\heart.csv')

features_to_normalize = ['age', 'trestbps', 'chol', 'thalach']

scaler = MinMaxScaler()

df[features_to_normalize] = scaler.fit_transform(df[features_to_normalize])

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)

print(df)
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	0.708333	1	3	0.481132	0.244292	1	0	0.603053	0	2.3	0	0	1	1
1	0.166667	1	2	0.339623	0.283105	0	1	0.885496	0	3.5	0	0	2	1
2	0.250000	0	1	0.339623	0.178082	0	0	0.770992	0	1.4	2	0	2	1
3	0.562500	1	1	0.245283	0.251142	0	1	0.816794	0	0.8	2	0	2	1
4	0.583333	0	0	0.245283	0.520548	0	1	0.702290	1	0.6	2	0	2	1
5	0.583333	1	0	0.433962	0.150685	0	1	0.587786	0	0.4	1	0	1	1
6	0.562500	0	1	0.433962	0.383562	0	0	0.625954	0	1.3	1	0	2	1
7	0.312500	1	1	0.245283	0.312785	0	1	0.778626	0	0.0	2	0	3	1
8	0.479167	1	2	0.735849	0.166667	1	1	0.694656	0	0.5	2	0	3	1

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.