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

Выполнила

Е. С. Неруш,

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 10**

Выборка German Credit Data. Содержит информацию о заемщиках, включая их кредитную историю, цель кредита, возраст, и оценку кредитоспособности (хороший/плохой).

Задачи:

1. Загрузите данные и выведите информацию о них.

2. Проанализируйте распределение цели кредита (Purpose). Визуализируйте 5 самых популярных целей.

3. Преобразуйте категориальные признаки Sex и Housing в числовой формат.

4. Постройте "ящик с усами" для Credit amount, чтобы сравнить суммы кредитов у "хороших" и "плохих" заемщиков.

5. Создайте сводную таблицу, показывающую средний возраст (Age) и среднюю длительность кредита (Duration) для каждой категории кредитной истории (Credit history).

6. Нормализуйте числовые столбцы Age, Credit amount, Duration.

Код программы:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import MinMaxScaler

# 1. Загрузка данных

df = pd.read\_csv("E:/Projects/ml\_as66/reports/Nerush/lab1/src/german\_credit.csv")

print("🔹 Информация о данных:")

print(df.info())

# 2. Исследовательский анализ

print("\n🔹 Количество пропущенных значений:")

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

print("\n🔹 Статистика:")

print(df.describe())

print("\n🔹 Медианы:")

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

print("\n🔹 Стандартные отклонения:")

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

# 3. Обработка пропущенных значений

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

for col in df.select\_dtypes(include='object').columns:

df[col] = df[col].fillna(df[col].mode()[0])

# 4. One-Hot Encoding для категориальных признаков

df\_encoded = pd.get\_dummies(df, columns=['personal\_status\_sex', 'housing'], drop\_first=True)

# 5. Нормализация числовых признаков

scaler = MinMaxScaler()

num\_cols = ['age', 'credit\_amount', 'duration\_in\_month']

df\_encoded[num\_cols] = scaler.fit\_transform(df\_encoded[num\_cols])

# 6. Визуализация

# 6.1 Гистограмма целей кредита

plt.figure(figsize=(8, 5))

purpose\_counts = df['purpose'].value\_counts().head(5)

sns.barplot(x=purpose\_counts.index, y=purpose\_counts.values, palette="Blues")

plt.title("Топ-5 целей кредита")

plt.ylabel("Количество")

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# 6.2 Ящик с усами по credit\_amount

plt.figure(figsize=(8, 5))

sns.boxplot(x='default', y='credit\_amount', data=df, palette="Set2")

plt.title("Сравнение суммы кредита по кредитоспособности")

plt.xlabel("Кредитоспособность (0 = плохой, 1 = хороший)")

plt.ylabel("Сумма кредита")

plt.tight\_layout()

plt.show()

# 6.3 Диаграмма рассеяния age vs duration\_in\_month

plt.figure(figsize=(8, 5))

sns.scatterplot(x='age', y='duration\_in\_month', hue='default', data=df, palette="coolwarm")

plt.title("Возраст vs Длительность кредита")

plt.tight\_layout()

plt.show()

# 7. Сводная таблица по credit\_history

pivot\_table = df.pivot\_table(

values=['age', 'duration\_in\_month'],

index='credit\_history',

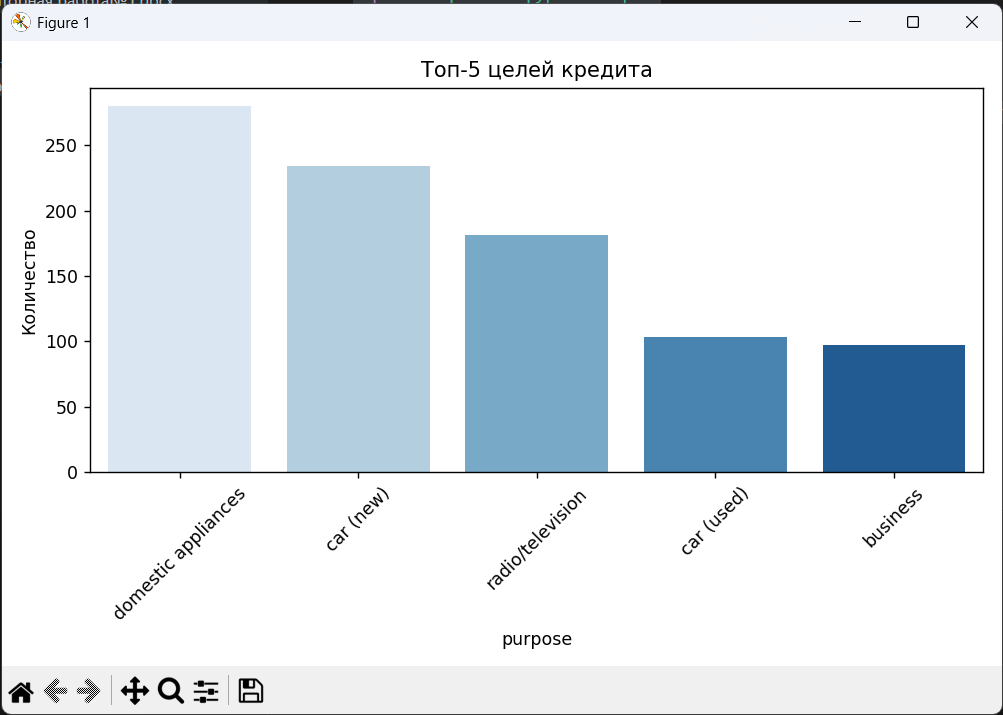
aggfunc='mean'

)

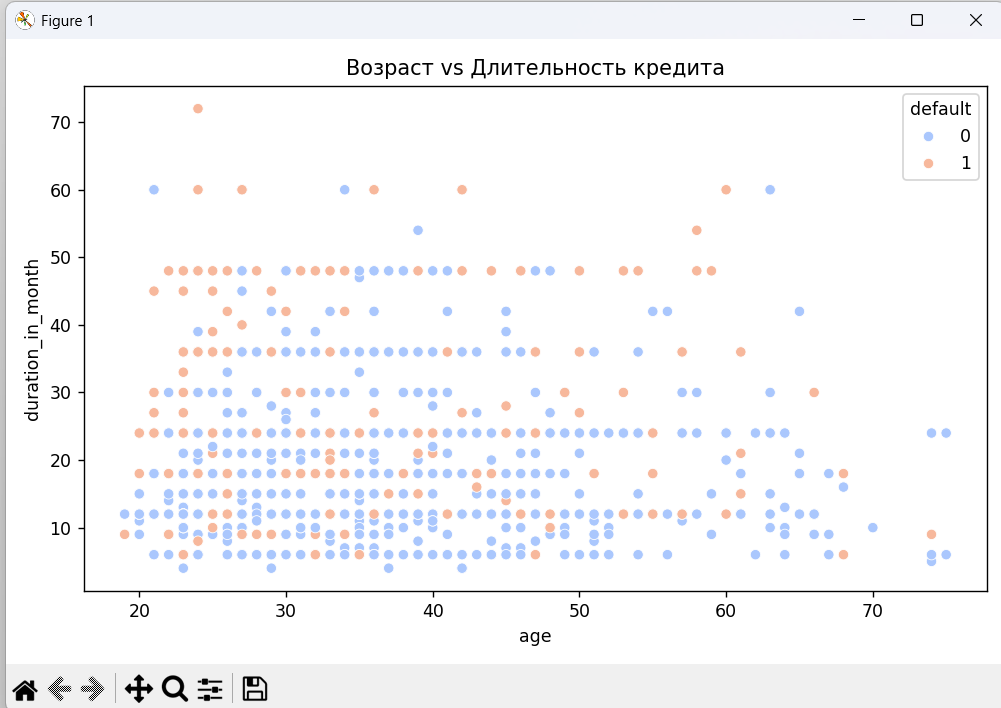
print("\n🔹 Средний возраст и длительность кредита по кредитной истории:")

print(pivot\_table)

# Рисунки с результатами работы программы

****

****

****

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