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

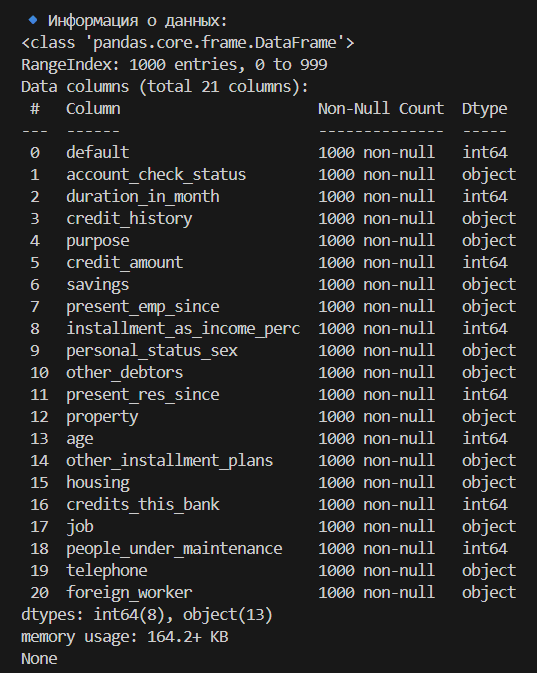
from sklearn.preprocessing import MinMaxScaler

# 1. Загрузка данных и вывод информации

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

print("🔹 Информация о данных:")

print(df.info())



# 2. Анализ и визуализация целей кредита

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

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

purpose\_df = purpose\_counts.reset\_index()

purpose\_df.columns = ['purpose', 'count']

sns.barplot(data=purpose\_df, x='purpose', y='count', hue='purpose', palette="Blues", legend=False)

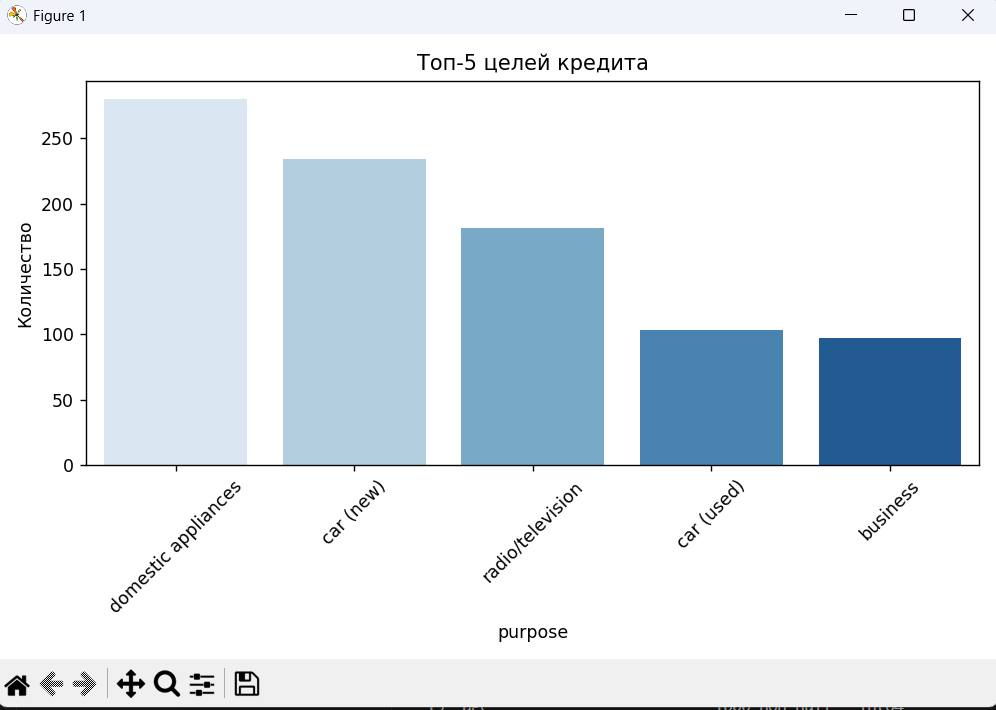
plt.title("Топ-5 целей кредита")

plt.ylabel("Количество")

plt.xticks(rotation=45)

plt.tight\_layout()

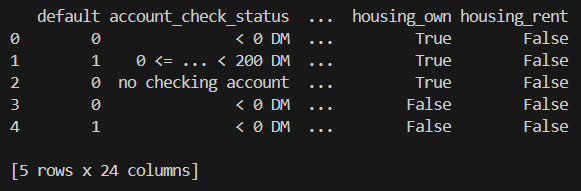
plt.show()



# 3. Преобразование категориальных признаков Sex и Housing

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

print(df\_encoded.head())



# 4. Ящик с усами для Credit amount по default

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

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

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

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

plt.ylabel("Сумма кредита")

plt.tight\_layout()

plt.show()



# 5. Сводная таблица по Credit history

pivot\_table = df.pivot\_table(

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

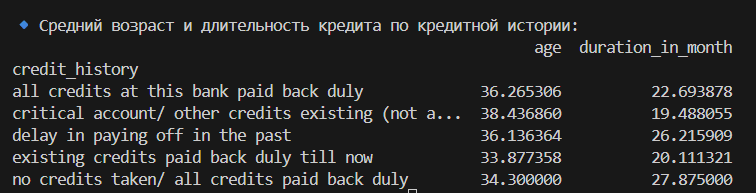
index='credit\_history',

aggfunc='mean'

)

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

print(pivot\_table)



# 6. Нормализация числовых признаков

scaler = MinMaxScaler()

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

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

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