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

Вавдийчик Н.Д.

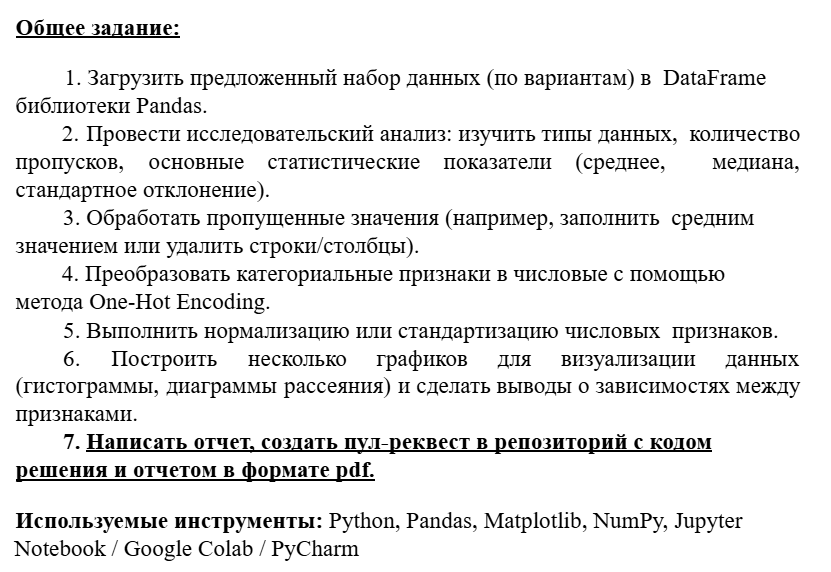
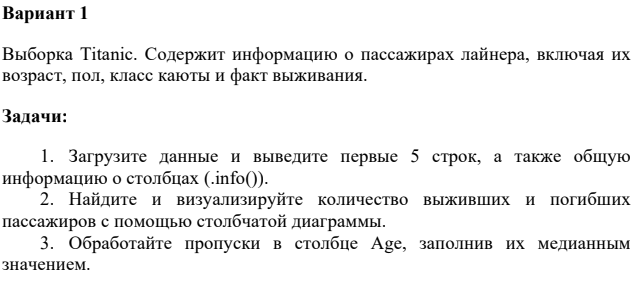
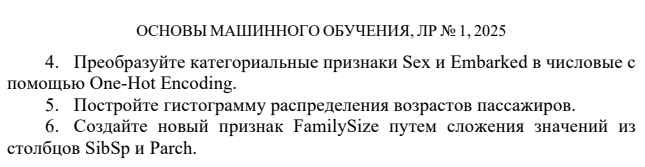
Проверил:

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

import numpy as np

data = pd.read\_csv("Titanic-Dataset.csv")

print("Первые 5 строк:")

print(data.head())

print("\nИнформация о данных:")

print(data.info())

print("\nКоличество выживших и погибших:")

print(data['Survived'].value\_counts())

data['Survived'].value\_counts().plot(kind='bar')

plt.title("Выжившие (1) и погибшие (0)")

plt.xlabel("Статус")

plt.ylabel("Количество")

plt.show()

print("\nПропуски в Age до обработки:", data['Age'].isnull().sum())

median\_age = data['Age'].median()

data['Age'] = data['Age'].fillna(median\_age)

print("Пропуски в Age после обработки:", data['Age'].isnull().sum())

data = pd.get\_dummies(data, columns=['Sex', 'Embarked'], drop\_first=True)

print("\nДанные после One-Hot Encoding:")

print(data.head())

plt.hist(data['Age'], bins=20, color='skyblue', edgecolor='black')

plt.title("Распределение возраста пассажиров")

plt.xlabel("Возраст")

plt.ylabel("Количество")

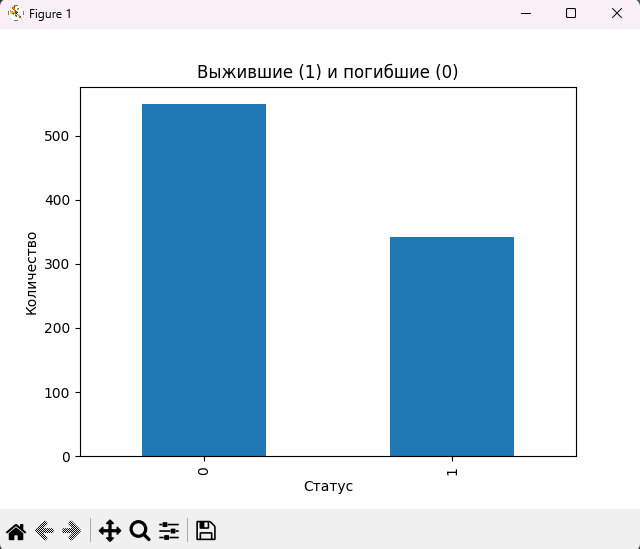
plt.show()

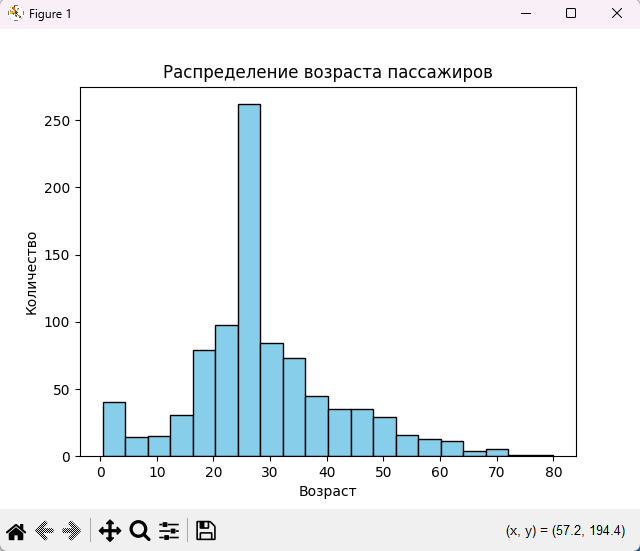
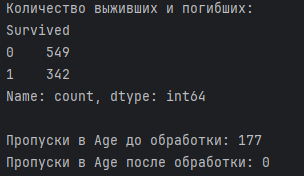
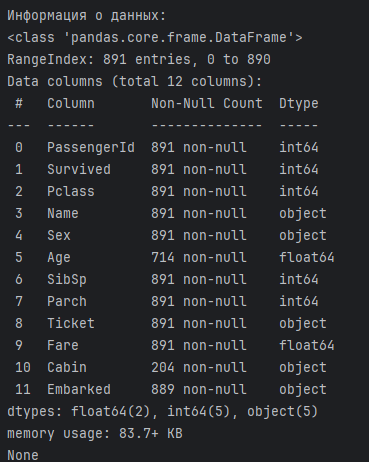
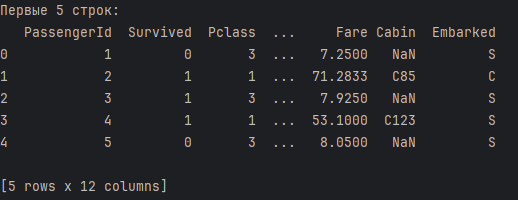
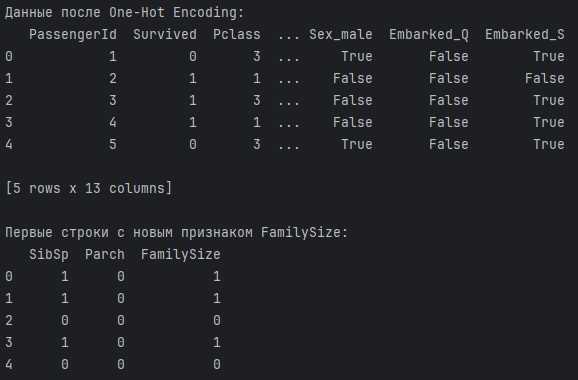
data['FamilySize'] = data['SibSp'] + data['Parch']

print("\nПервые строки с новым признаком FamilySize:")

print(data[['SibSp', 'Parch', 'FamilySize']].head())

Графики:



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