

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
А. А. Петrusевич,
студентка 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 1.

Выборка Titanic. Содержит информацию о пассажирах лайнера, включая их возраст, пол, класс каюты и факт выживания.

Задачи:

1. Загрузите данные и выведите первые 5 строк, а также общую информацию о столбцах (.info()).
2. Найдите и визуализируйте количество выживших и погибших пассажиров с помощью столбчатой диаграммы.
3. Обработайте пропуски в столбце Age, заполнив их медианным значением.
4. Преобразуйте категориальные признаки Sex и Embarked в числовые с помощью One-Hot Encoding.
5. Постройте гистограмму распределения возрастов пассажиров.
6. Создайте новый признак FamilySize путем сложения значений из столбцов SibSp и Parch.

Код программы:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv("G:/ЛАБЫ 3 КУРС/ОМО/titanic.csv")

print(" Первые 5 строк выборки:")
print(df.head())

print("\n └── Информация о столбцах:")
df.info()

survival_counts = df["Survived"].value_counts()

plt.figure(figsize=(6, 4))
sns.barplot(x=survival_counts.index, y=survival_counts.values,
palette="viridis")
plt.xticks([0, 1], ["Погибли", "Выжили"])
plt.title("Количество выживших и погибших пассажиров")
plt.xlabel("Статус")
plt.ylabel("Количество")
plt.savefig("G:/ЛАБЫ 3 КУРС/ОМО/survival_barplot.png")
plt.close()

age_missing_before = df["Age"].isnull().sum()
df["Age"].fillna(df["Age"].median(), inplace=True)
age_missing_after = df["Age"].isnull().sum()

df_encoded = pd.get_dummies(df, columns=["Sex", "Embarked"])
```

```

plt.figure(figsize=(8, 5))
sns.histplot(df["Age"], bins=30, kde=True, color="steelblue")
plt.title("Распределение возрастов пассажиров")
plt.xlabel("Возраст")
plt.ylabel("Количество")
plt.savefig("G:/ЛАБЫ 3 КУРС/ОМО/age_histogram.png")
plt.close()

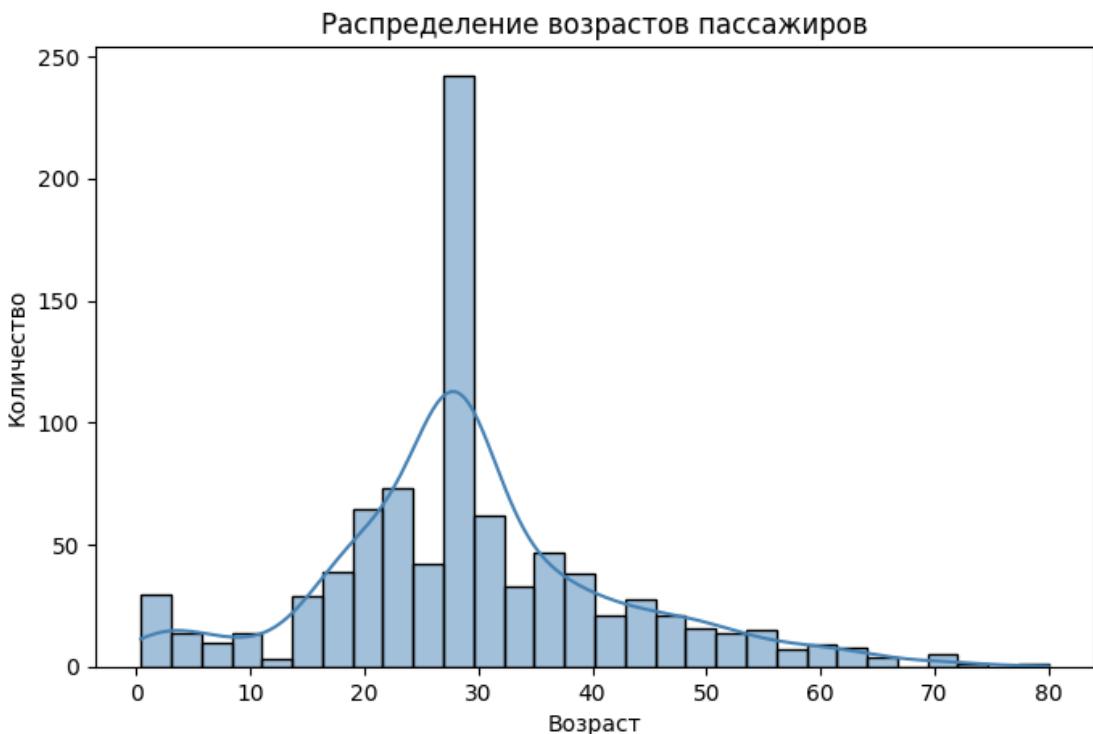
df["FamilySize"] = df["SibSp"] + df["Parch"]

with open("G:/ЛАБЫ 3 КУРС/ОМО/titanic_report.txt", "w",
encoding="utf-8") as f:
    f.write(" Первые 5 строк:\n")
    f.write(df.head().to_string() + "\n\n")
    f.write(" Информация о столбцах:\n")
    f.write(str(df.info()) + "\n\n")
    f.write("Количество выживших и погибших:\n")
    f.write(str(survival_counts) + "\n\n")
    f.write(f" Пропущенные значения в Age до: {age_missing_before},\n"
после: {age_missing_after}\n\n")
    f.write(" Данные после One-Hot Encoding (первые 5 строк):\n")
    f.write(df_encoded.head().to_string() + "\n\n")
    f.write("Новый признак FamilySize (первые 5 строк):\n")
    f.write(df[['SibSp', "Parch", "FamilySize"]].head().to_string()
+ "\n\n")

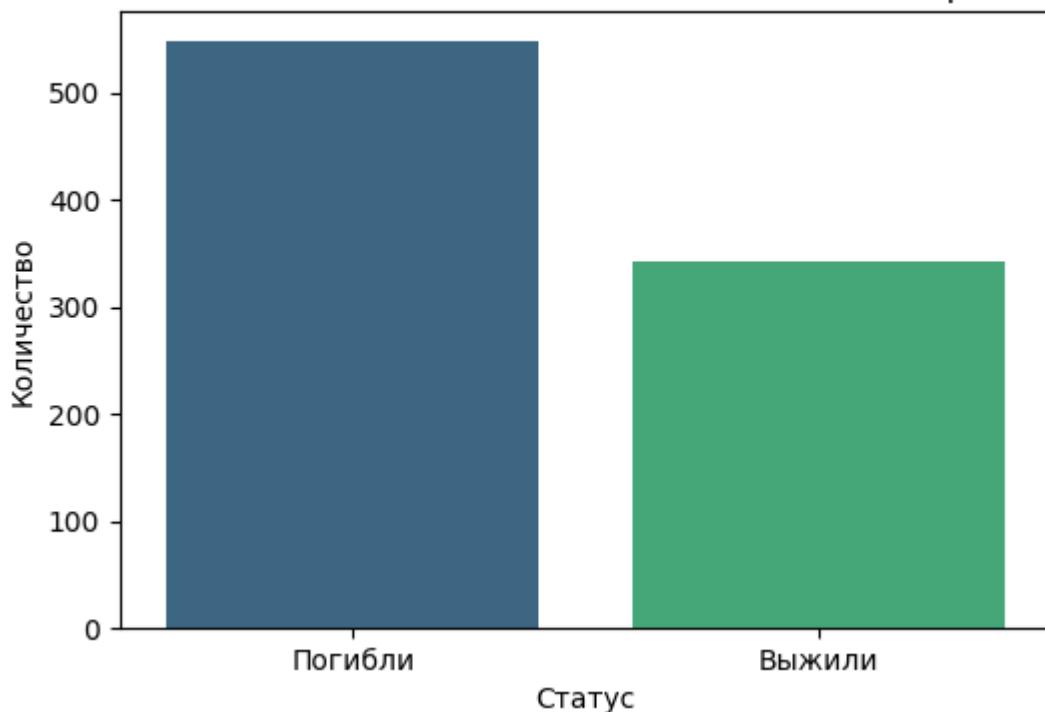
print(" Отчёт сохранён в titanic_report.txt")
print(" Графики сохранены в survival_barplot.png и
age_histogram.png")

```

Рисунки с результатами работы программы



Количество выживших и погибших пассажиров



```

🕒 Первые 5 строк:
  PassengerId Survived Pclass          Name     Sex   Age SibSp Parch     Ticket  Fare Cabin Embarked FamilySize
0            1       0      3    Braund, Mr. Owen Harris   male  22.0      1    0        A/5 21171  7.2500   NaN      S       1
1            2       1      1  Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38.0      1    0         PC 17599 71.2833   C85      C       1
2            3       1      3           Heikkinen, Miss. Laina  female  26.0      0    0        STON/O2. 3101282  7.9250   NaN      S       0
3            4       1       1        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35.0      1    0        113803 53.1000  C123      S       1
4            5       0      3    Allen, Mr. William Henry   male  35.0      0    0        373450  8.0500   NaN      S       0

💡 Информация о столбцах:
None

📊 Количество выживших и погибших:
Survived
0      549
1      342
Name: count, dtype: int64

🕒 Пропущенные значения в Age до: 177, после: 0

📅 данные после One-Hot Encoding (первые 5 строк):
  PassengerId Survived Pclass          Name     Sex   Age SibSp Parch     Ticket  Fare Cabin Embarked Sex_female Sex_male Embarked_C Embarked_Q Embarked_S
0            1       0      3    Braund, Mr. Owen Harris   male  22.0      1    0        A/5 21171  7.2500   NaN      False      True    False    False    True
1            2       1      1  Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38.0      1    0         PC 17599 71.2833   C85      True      False   True    False    False
2            3       1      3           Heikkinen, Miss. Laina  female  26.0      0    0        STON/O2. 3101282  7.9250   NaN      True      False   False    False    True
3            4       1       1        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35.0      1    0        113803 53.1000  C123      True      False   False    False    True
4            5       0      3    Allen, Mr. William Henry   male  35.0      0    0        373450  8.0500   NaN      False      True    False    False    True

💡 Новый признак FamilySize (первые 5 строк):
  SibSp Parch FamilySize
0      1      0       1
1      1      0       1
2      0      0       0
3      1      0       1
4      0      0       0

```

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

