

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

Отчет по лабораторной работе №1

Выполнил  
В.Д.Головкина,  
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 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.

```
import io
from typing import NoReturn
from pathlib import Path

import matplotlib.pyplot as plt
import pandas as pd

def main() -> None:
    file_path = Path(r"D:\y--ба\3 kurs\omo\1\Titanic-Dataset.csv")
    if not file_path.exists():
        print(f"❌ Файл не найден: {file_path}")
        return

    df = pd.read_csv(file_path)

    print("Первые 5 строк:")
    pd.set_option('display.max_columns', None)
    pd.set_option('display.width', None)
    pd.set_option('display.max_colwidth', None)
    print(df.head(5))

    print("\nИнформация о столбцах (.info()):")
    buf = io.StringIO()
    df.info(buf=buf)
    print(buf.getvalue())

# 2) Ст. диаграмма выж/пог
plt.figure(figsize=(6, 4))
survived_counts = df["Survived"].value_counts().reindex([0, 1], fill_value=0)
survived_counts.index = ["Мертвы (0)", "Живы (1)"]
survived_counts.plot(kind="bar", color=["#d9534f", "#5cb85c"])
plt.title("Количество погибших и выживших")
```

```
plt.ylabel("Число пассажиров")
```

```
plt.tight_layout()
```

```
plt.show()
```

```
df["Age"] = pd.to_numeric(df["Age"], errors="coerce")
```

```
age_median = df["Age"].median()
```

```
df["Age"] = df["Age"].fillna(age_median)
```

```
print(f"\nПропуски в Age заполнены медианой: {age_median:.2f}")
```

```
df_ohe = pd.get_dummies(df, columns=["Sex", "Embarked"], drop_first=False)
```

```
created_ohe = [c for c in df_ohe.columns if c.startswith("Sex_") or c.startswith("Embarked_")]
```

```
print("\nСозданы ОНЕ-признаки:", ", ".join(created_ohe))
```

```
plt.figure(figsize=(6, 4))
```

```
plt.hist(df_ohe["Age"], bins=30, color="#5bc0de", edgecolor="white")
```

```
plt.title("Распределение возраста пассажиров (Age)")
```

```
plt.xlabel("Возраст")
```

```
plt.ylabel("Частота")
```

```
plt.tight_layout()
```

```
plt.show()
```

```
df_ohe["FamilySize"] = df_ohe["SibSp"].fillna(0) + df_ohe["Parch"].fillna(0)
```

```
print("\nПризнак FamilySize создан. Примеры значений:")
```

```
print(df_ohe[["SibSp", "Parch", "FamilySize"]].head())
```

```
if __name__ == "__main__":
```

```
    main()
```

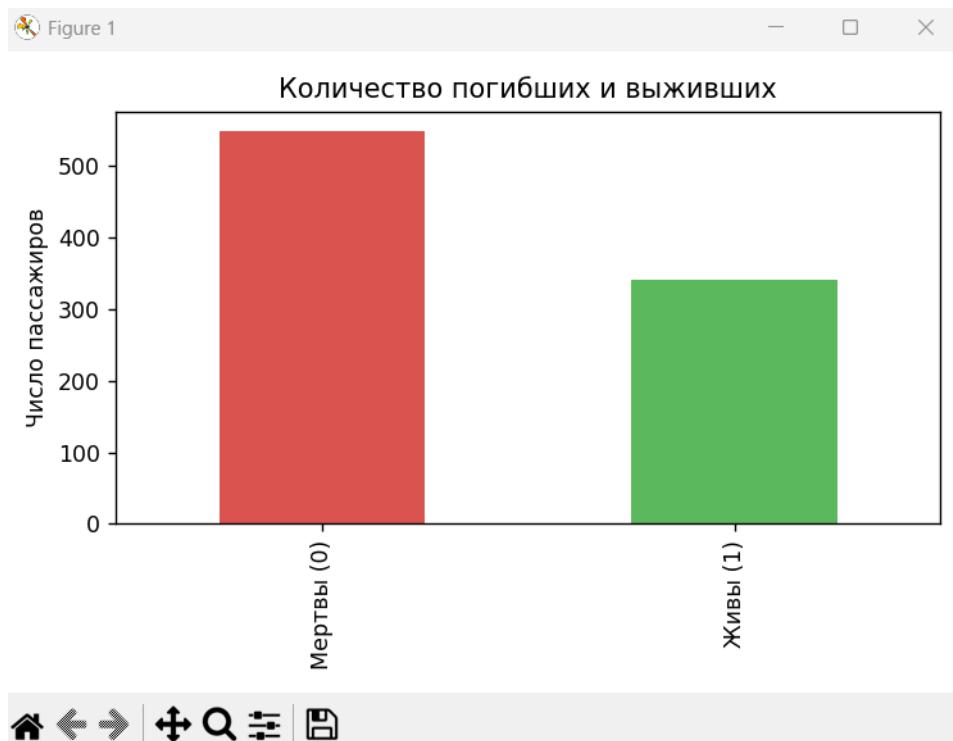
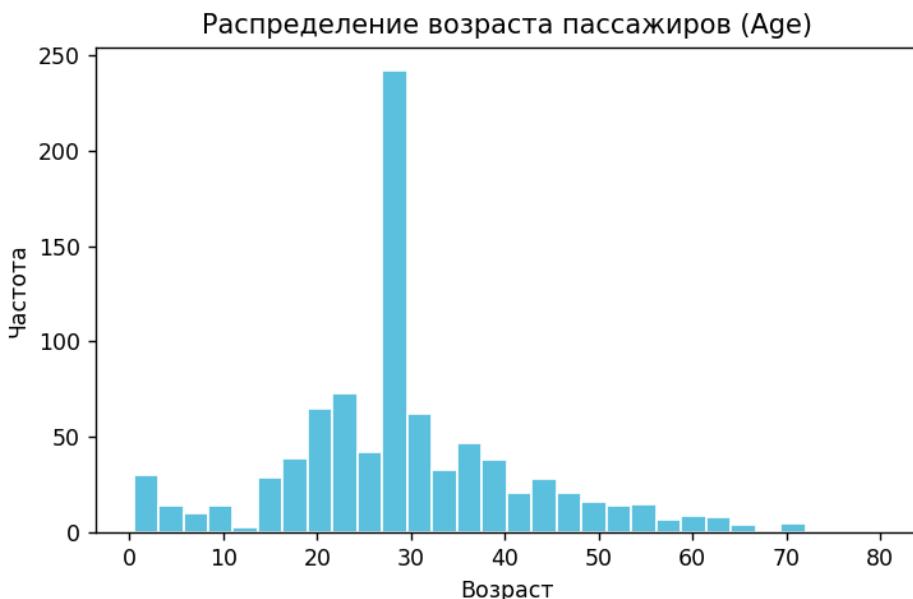


Figure 1

- □ ×



```
PS D:\y--бa\3 kurs\omo\1> & 'c:\Users\Пинка\AppData\Local\Programs\Python\Python313\python.exe' 'c:\Users\Пинка\.vscode\extensions\ms-python.debugpy-2025.10
libs\debugpy\launcher' '49929' '--' 'D:\y--бa\3 kurs\omo\1\titanic_lab1.py'
Первые 5 строк:
   PassengerId  Survived  Pclass          Name     Sex   Age  SibSp  Parch     Ticket  Fare Cabin Embarked
0            1         0      3  Braund, Mr. Owen Harris   male  22.0      1     0    A/5 21171  7.2500   NaN      S
1            2         1      1  Cumings, Mrs. John Bradley (Florence Briggs Thayer)  female  38.0      1     0        PC 17599  71.2833  C85      C
2            3         1      3           Heikkinen, Miss. Laina  female  26.0      0     0  STON/O2. 3101282  7.9250   NaN      S
3            4         1      1  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1     0   113803  53.1000 C123      S
4            5         0      3           Allen, Mr. William Henry   male  35.0      0     0        373450  8.0500   NaN      S

Информация о столбцах (.info()):
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId  891 non-null   int64  
 1   Survived     891 non-null   int64  
 2   Pclass       891 non-null   int64  
 3   Name         891 non-null   object 
 4   Sex          891 non-null   object 
 5   Age          714 non-null   float64 
 6   SibSp        891 non-null   int64  
 7   Parch        891 non-null   int64  
 8   Ticket       891 non-null   object 
 9   Fare         891 non-null   float64 
 10  Cabin        204 non-null   object 
 11  Embarked     889 non-null   object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

Пропуски в Age заполнены медианой: 28.00

Созданы ОНЕ-признаки: Sex\_female, Sex\_male, Embarked\_C, Embarked\_Q, Embarked\_S

Признак FamilySize создан. Примеры значений:

	SibSp	Parch	FamilySize
0	1	0	1
1	1	0	1
2	0	0	0
3	1	0	1
4	0	0	0

PS D:\y--бa\3 kurs\omo\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.