МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «ВГТУ», ВГТУ)

Факультет информационных технологий и информационной безопасности

Кафедра автоматизированных и вычислительных систем

ОТЧЕТ

по лабораторной работе № 1

по дисциплине «Интеллектуальные системы»

Тема «Введение в Python для анализа данных»

Выполнил

студент гр. мУПИ-251

Подпись, дата Инициалы, фамилия

Руководитель С.А. Олейникова

Подпись, дата Инициалы, фамилия

Защищена \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата

ВОРОНЕЖ 2025

**Цель лабораторной работы**

Получение практических навыков работы с языком Python и библиотеками для анализа данных (pandas, numpy, matplotlib), а также знакомство с базовыми инструментами обработки и анализа данных.

**Краткие теоретические сведения**

Python является универсальным языком программирования, широко используемым в области анализа данных и машинного обучения. Для анализа данных активно применяются библиотеки pandas (работа с табличными данными), numpy (численные вычисления), matplotlib и seaborn (визуализация). Основные этапы анализа данных включают загрузку датасета, предварительный просмотр, обнаружение и обработку пропусков, исследование статистических характеристик и визуализацию.

**Ход выполнения работы**

1. В Jupyter Notebook был загружен датасет о пассажирах Титаника.

2. Выполнен просмотр первых строк таблицы, что позволило ознакомиться со структурой данных.

3. Определено общее количество строк и столбцов (887 строк, 8 признаков).

4. Проверено наличие пропусков — пропущенные значения отсутствуют.

5. Определены типы данных для каждого признака (Survived, Pclass — целые числа, Name — строка, Sex — строка, Age — вещественное число, Siblings/Spouses Aboard и Parents/Children Aboard — целые числа, Fare — вещественное число).

6. Выполнен базовый анализ данных, рассчитаны основные статистические показатели.

7. Построены графики для наглядной визуализации распределений признаков.

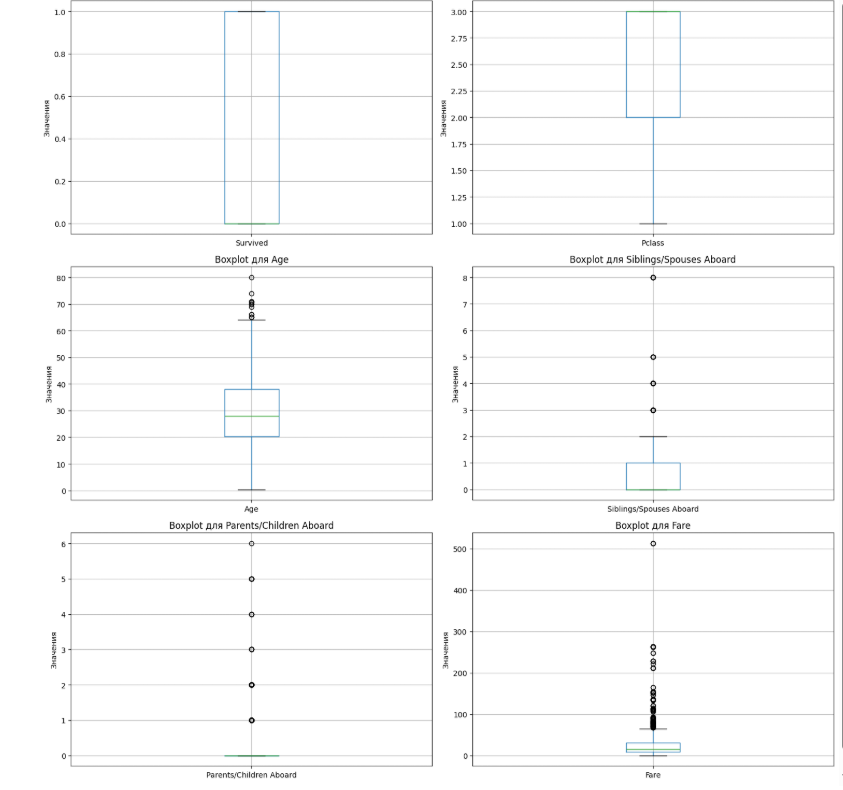


Рисунок 1 – Использование Boxplot для анализа выбросов

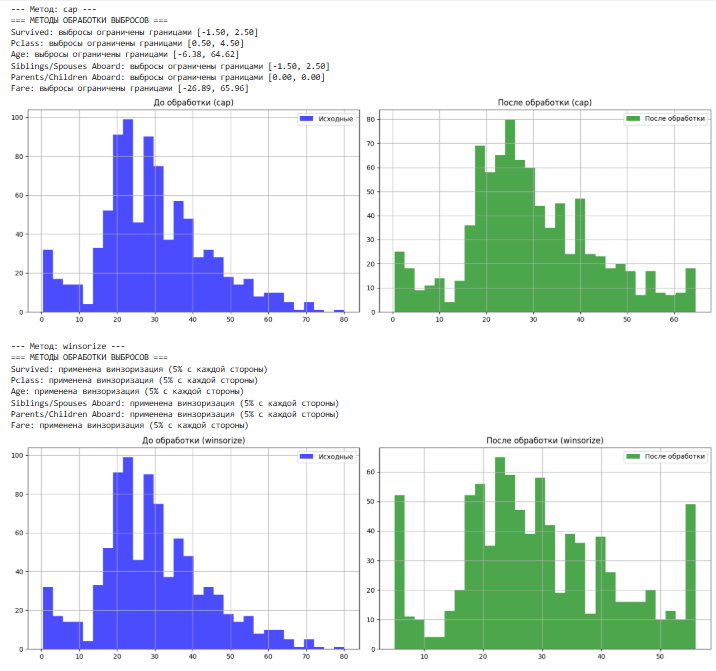


Рисунок 2 –Работа с выбросами при помощи разных методов

**Результаты работы**

В результате выполнения лабораторной работы был проведён анализ данных датасета Titanic. Удалось освоить основные инструменты библиотеки pandas для обработки и анализа табличных данных, а также познакомиться с методами визуализации.

**Вывод**

В ходе лабораторной работы получены практические навыки использования Python для анализа данных. Были изучены операции загрузки, первичного анализа и визуализации данных. Полученные знания являются основой для дальнейшего изучения методов интеллектуального анализа данных и машинного обучения.