|  |
| --- |
|  |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** |
| Институт информационных технологий (ИТ) |
| Кафедра прикладной математики (ПМ) |

|  |  |
| --- | --- |
| **ОТЧЕТ**  **ПО ПРАКТИЧЕСКОЙ РАБОТЕ №4** | |
| **по дисциплине** |  |
| **«Технологии и инструментарий анализа больших данных»** | |
| Выполнил студент группы ИКБО-20-19 | Московка А.А. |
| Принял семинарист кафедры ИиППО | Десятников А.А. |

|  |  |  |
| --- | --- | --- |
| Практическая работы выполнены | « » 2022г. |  |
| «Зачтено» | « » 2022 г. |  |

Москва 2022

**Постановка задачи**

1. Загрузить данные из файла “insurance.csv”.
2. С помощью метода describe() посмотреть статистику по данным. Сделать выводы.
3. Построить гистограммы для числовых показателей. Сделать выводы.
4. Найти меры центральной тенденции и меры разброса для индекса массы тела (bmi) и расходов (charges). Отобразить результаты в виде текста и на гистограммах (3 вертикальные линии). Добавить легенду на графики. Сделать выводы.
5. Построить box-plot для числовых показателей. Названия графиков должны соответствовать названиям признаков. Сделать выводы.
6. Используя признак charges или imb, проверить, выполняется ли центральная предельная теорема. Использовать различные длины выборок n. Количество выборок = 300 Вывести результат в виде гистограмм. Найти стандартное отклонение и среднее для полученных распределений. Сделать выводы.
7. Построить 95% и 99% доверительный интервал для среднего значения расходов и среднего значения индекса массы тела.
8. Оформить отчет на основе проделанной работы. Написать выводы.

**Программный код с комментариями и результаты его выполнения**

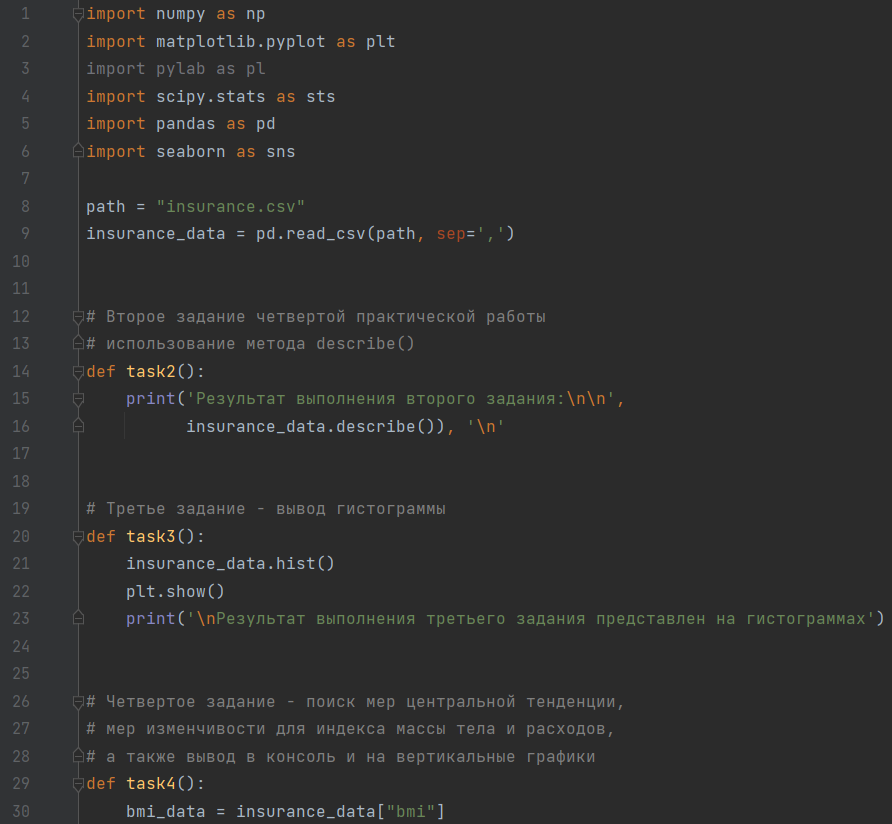


Рисунок 1 – Скриншот программного кода (Часть 1)



Рисунок 2 – Скриншот программного кода (Часть 2)

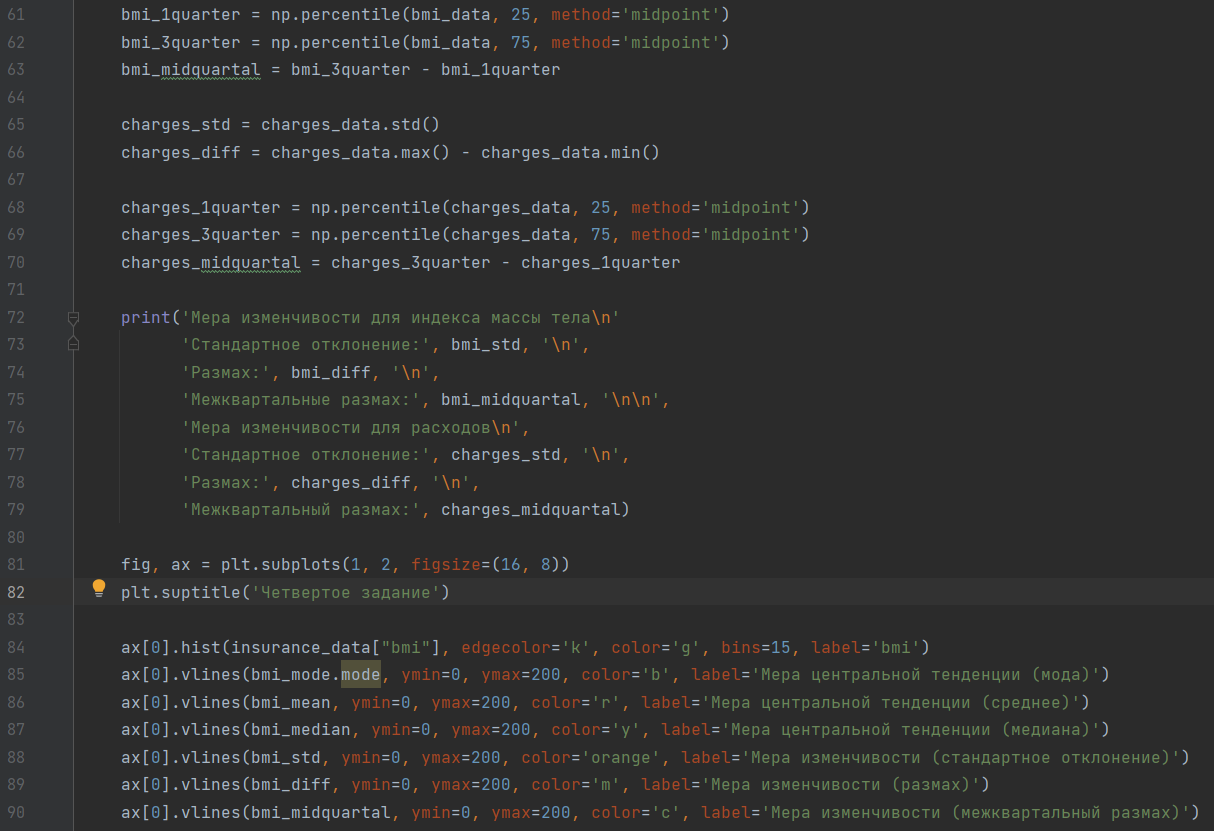


Рисунок 3 – Скриншот программного кода (Часть 3)

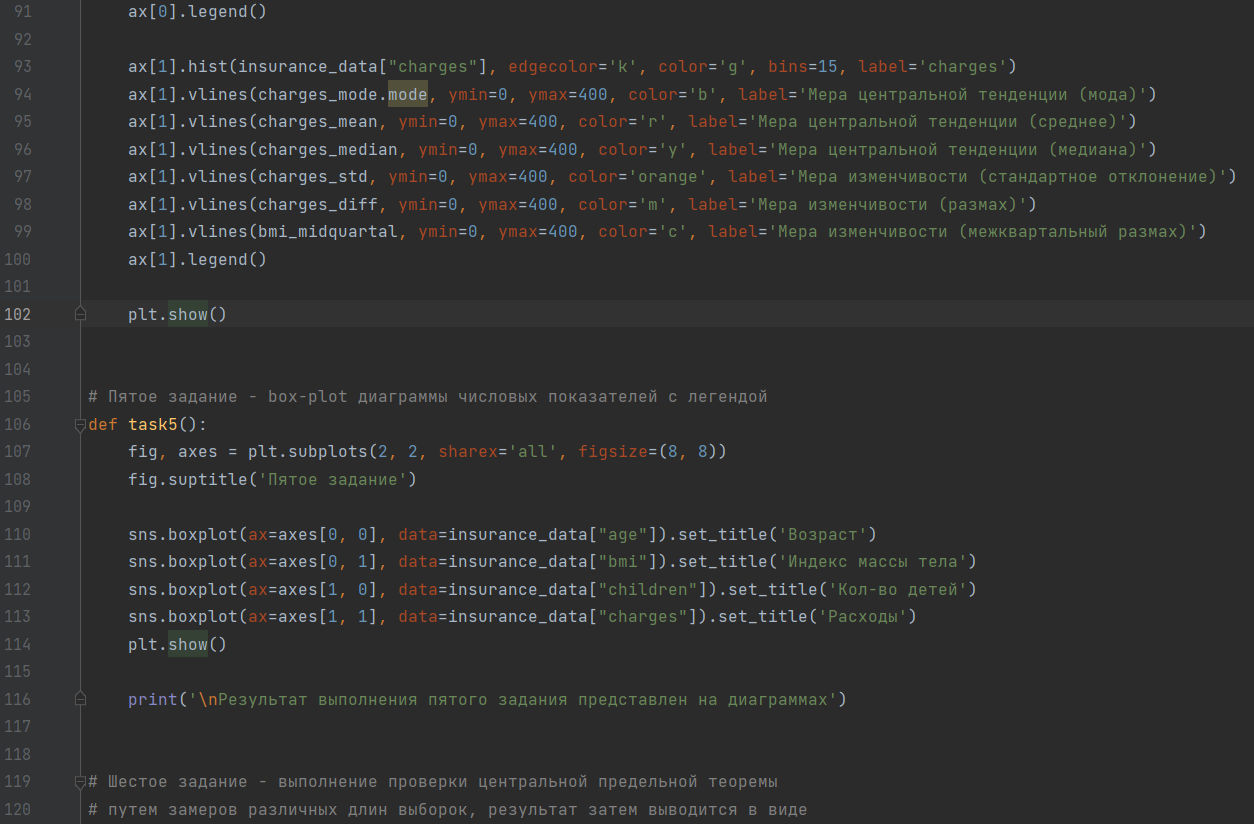


Рисунок 4 – Скриншот программного кода (Часть 4)

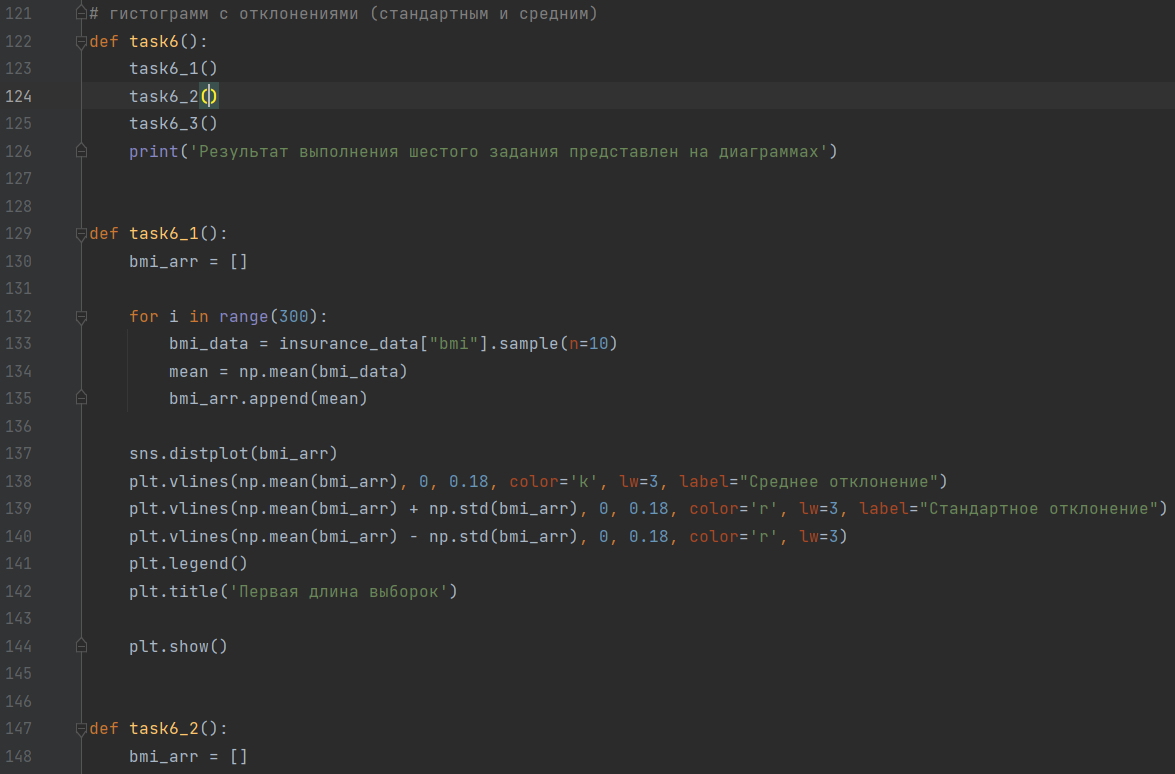


Рисунок 5 – Скриншот программного кода (Часть 5)



Рисунок 6 – Скриншот программного кода (Часть 6)

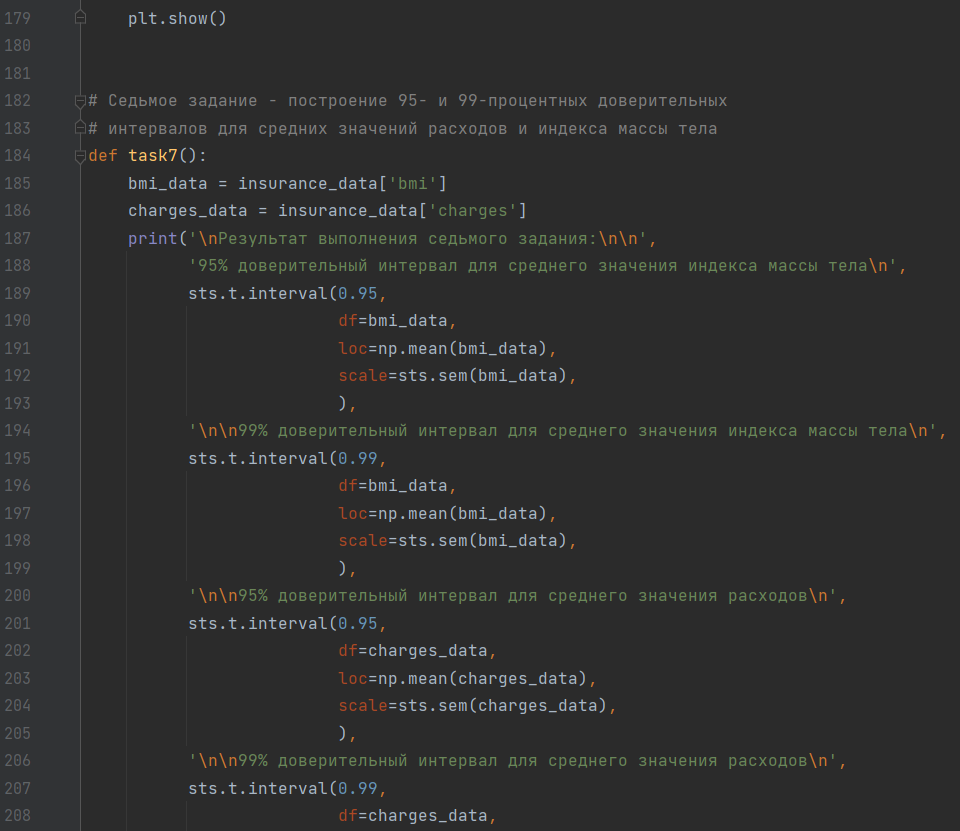


Рисунок 7 – Скриншот программного кода (Часть 7)

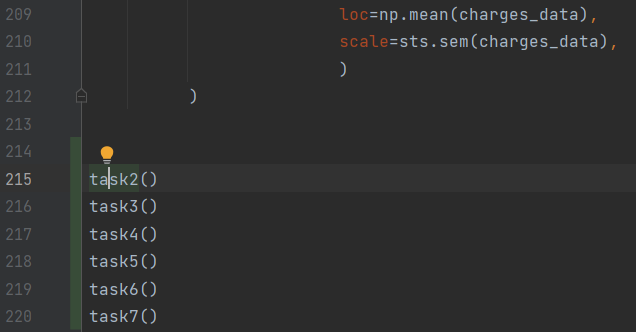


Рисунок 8 – Скриншот программного кода (Часть 8)

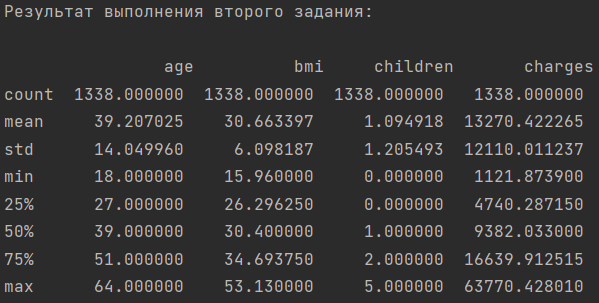


Рисунок 9 – Скриншот результата выполнения второго задания

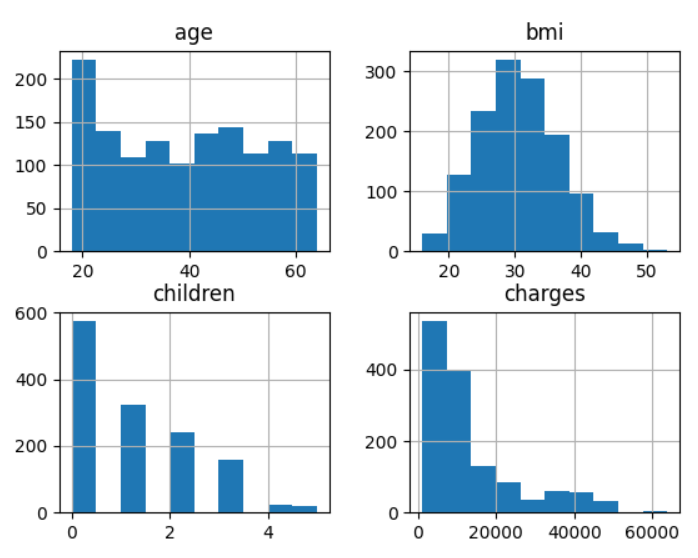


Рисунок 10 – Скриншот результата выполнения третьего задания

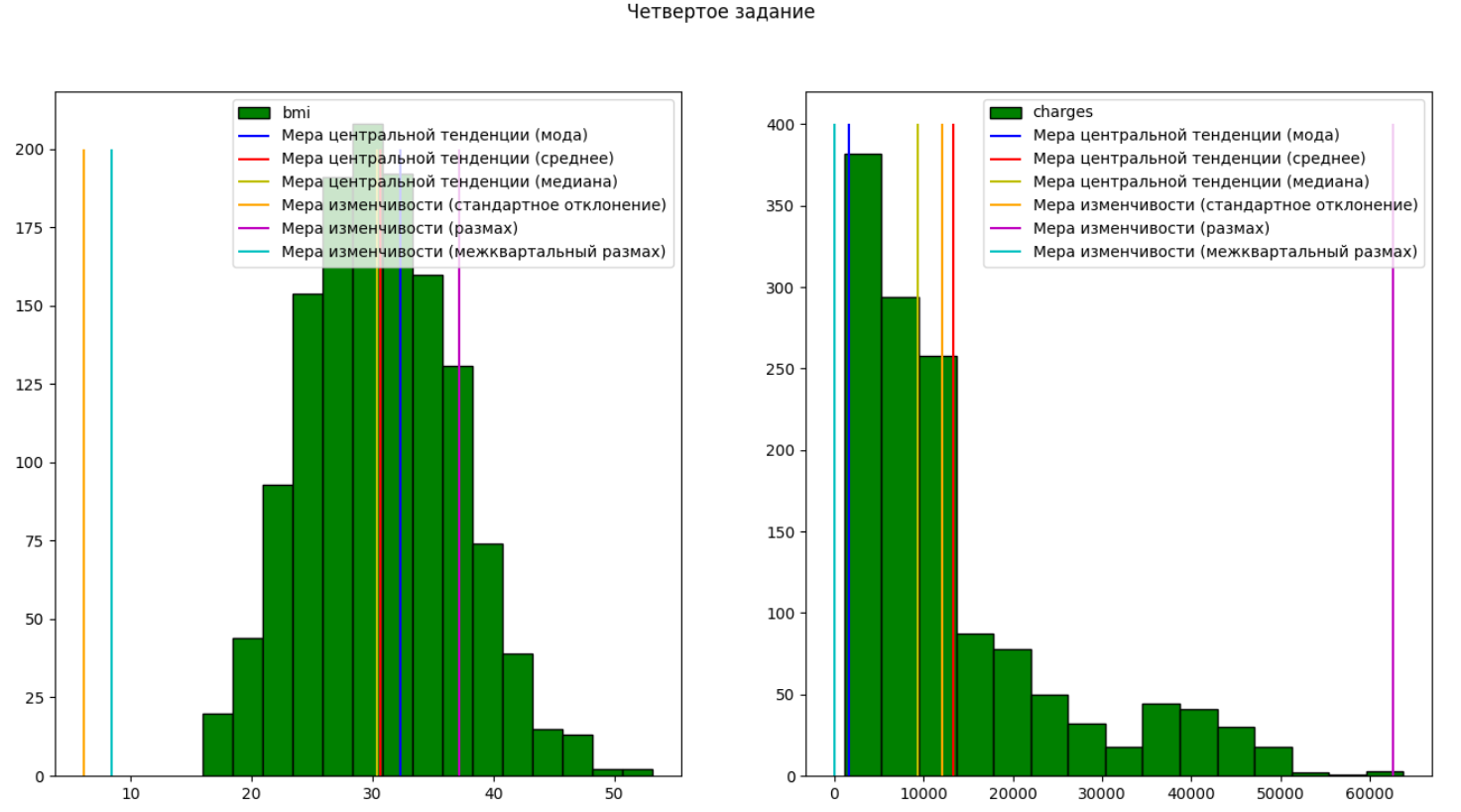


Рисунок 11 – Скриншот результата выполнения четвертого задания (диаграммы)

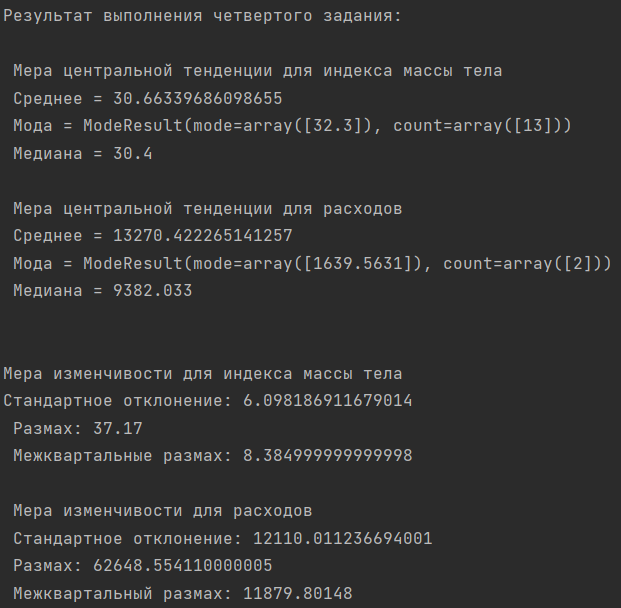


Рисунок 12 – Скриншот результата выполнения четвертого задания (консольный вывод)

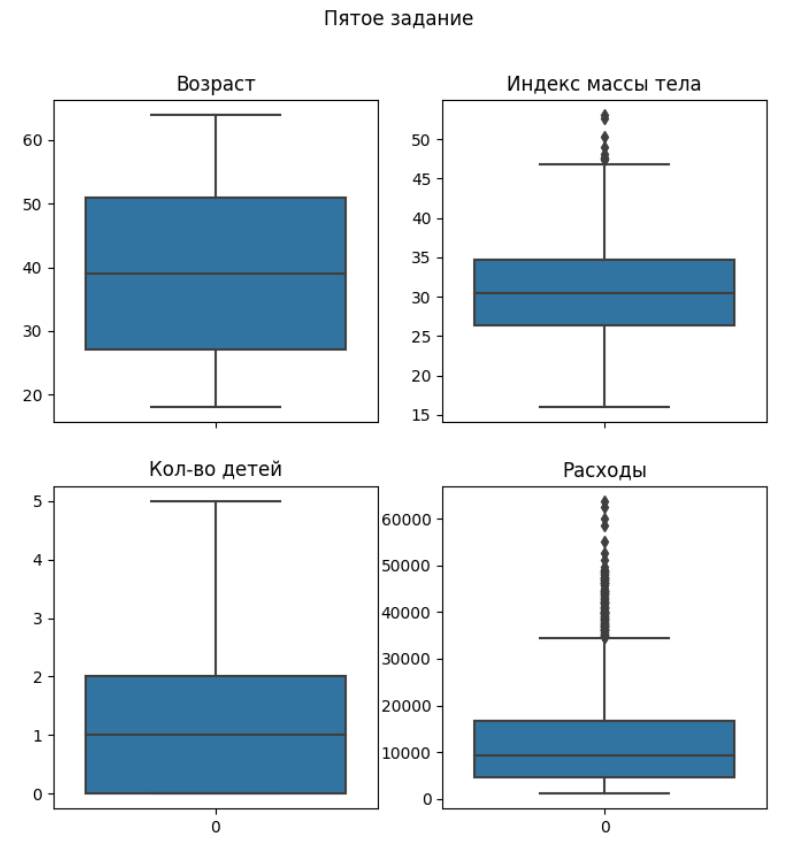


Рисунок 13 – Скриншот результата выполнения пятого задания

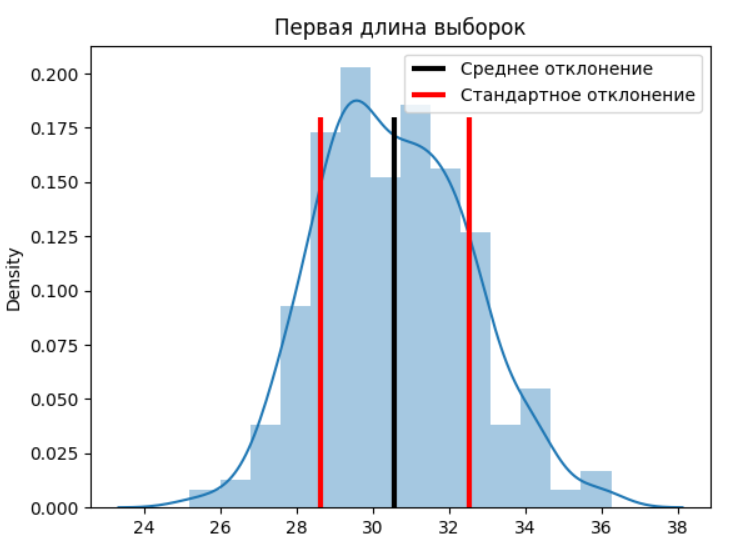


Рисунок 14 – Скриншот результата выполнения шестого задания (диаграмма первой длины выборки)

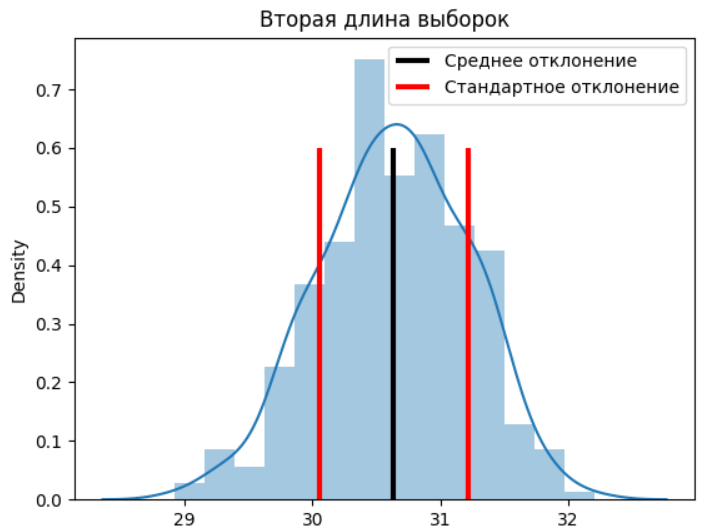


Рисунок 15 – Скриншот результата выполнения шестого задания (диаграмма второй длины выборки)

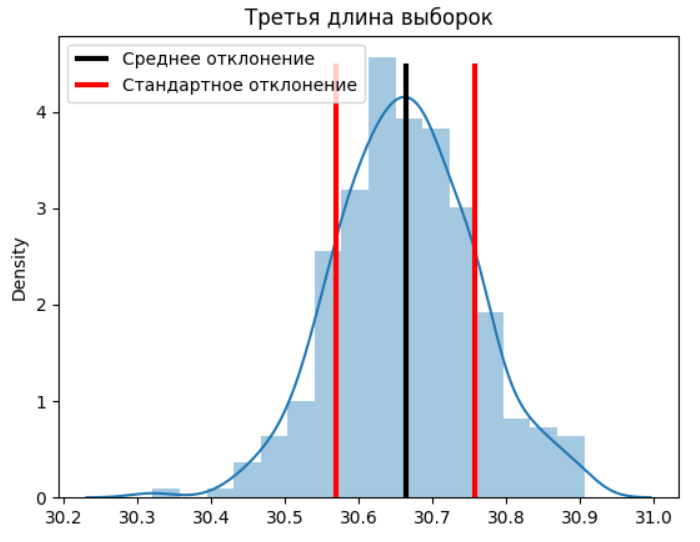


Рисунок 16 – Скриншот результата выполнения шестого задания (диаграмма третьей длины выборки)

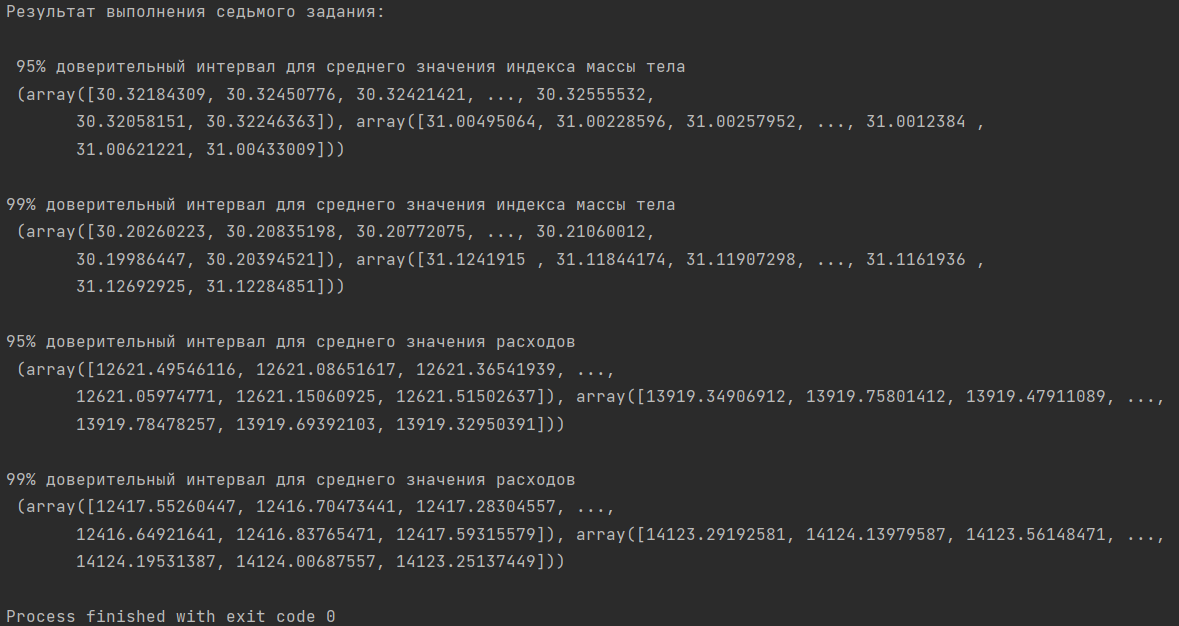


Рисунок 17 – Скриншот результата выполнения седьмого задания

**Вывод**

В результате выполнения данной практической работы была повторно и более основательно проведена работа с различными библиотеками визуализации данных (matplotlib, pyplot) и особенностями работы с ними в среде программирования Python, а также получен ценный практический и теоретический опыт использования вышеупомянутых инструментов обработки и анализа больших данных.