|  |  |  |
| --- | --- | --- |
|  | |  |
| МИНОБРНАУКИ РОССИИ | |  |
| Федеральное государственное бюджетное образовательное учреждение  высшего образования  **"МИРЭА - Российский технологический университет"**  **РТУ МИРЭА** | |  |
|  | Институт информационных технологий (ИТ) | |
|  | Кафедра прикладной математики | |

|  |  |
| --- | --- |
| **Отчет по шестой практической работе.** | |
| **по дисциплине** | |
| **«Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО- 22-20 | Гологузов К.А. |
| Принял | Парамонов А.А. |

Москва 2023 г.

**Практическое задание**

1. Найти данные для кластеризации. Данные в группе не должны повторяться. Если признаки в данных имеют очень сильно разные масштабы, то необходимо данные предварительно нормализовать.

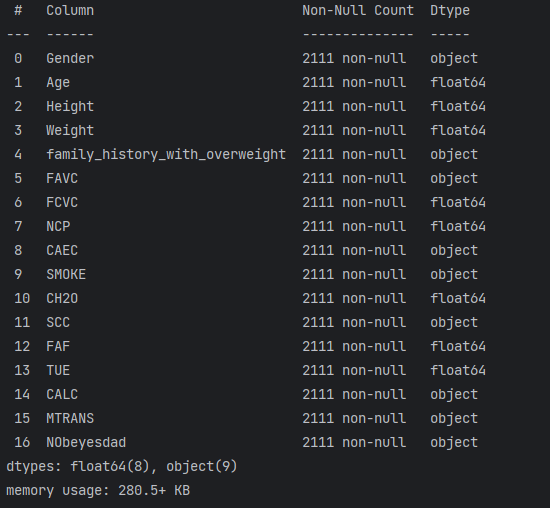


Рис.1 – Данные

Для выполнения практического задания был выбран датасет Obesity or CVD risk (Classify/Regressor/Cluster), содержащий данные о людях с ожирением.

Перед началом работы данные были предобработаны: значения в столбцах Gender и SMOKE были заменены на 0 и 1, столбцы с нечисловыми значениями были удалены.

2. Провести кластеризацию данных с помощью алгоритма k-means. Использовать «правило локтя» и коэффициент силуэта для поиска оптимального количества кластеров.

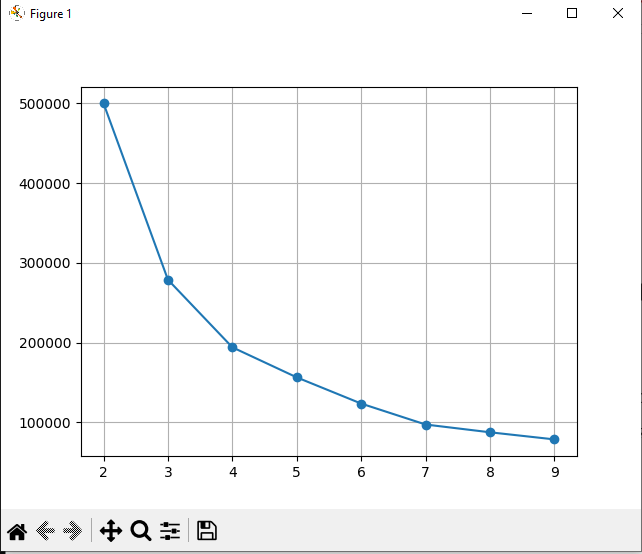


Рис.2 – Правило локтя

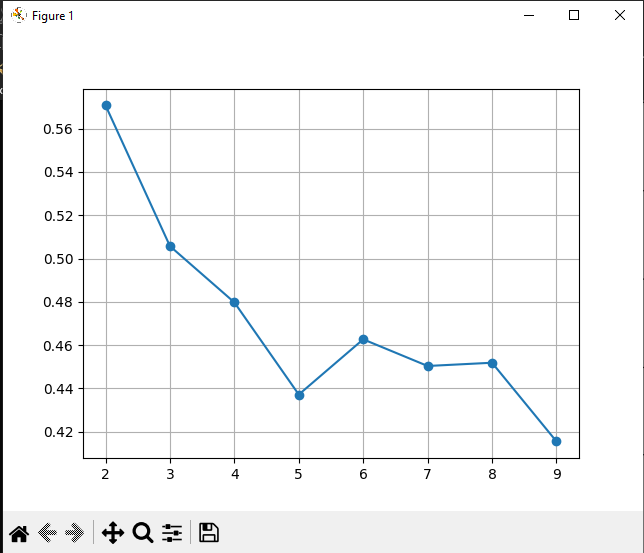


Рис.3 – Коэффициент силуэта

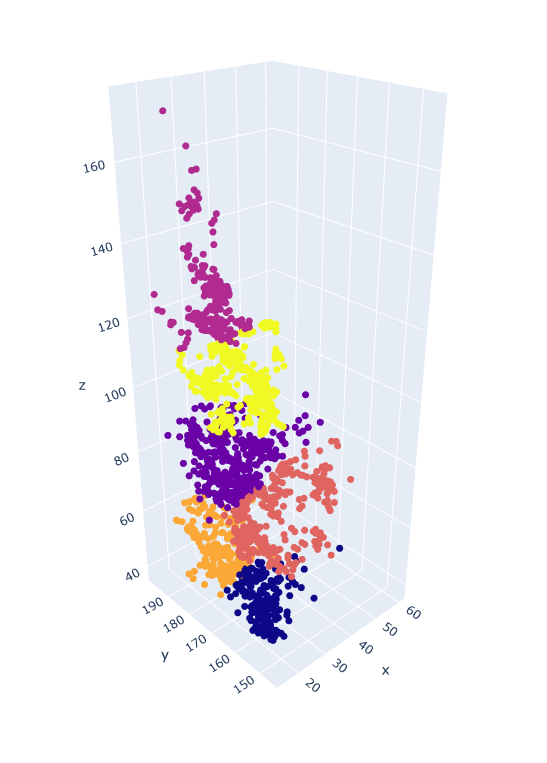


Рис.4 – Визуализация данных с помощью plotly

3. Провести кластеризацию данных с помощью алгоритма иерархической кластеризации.

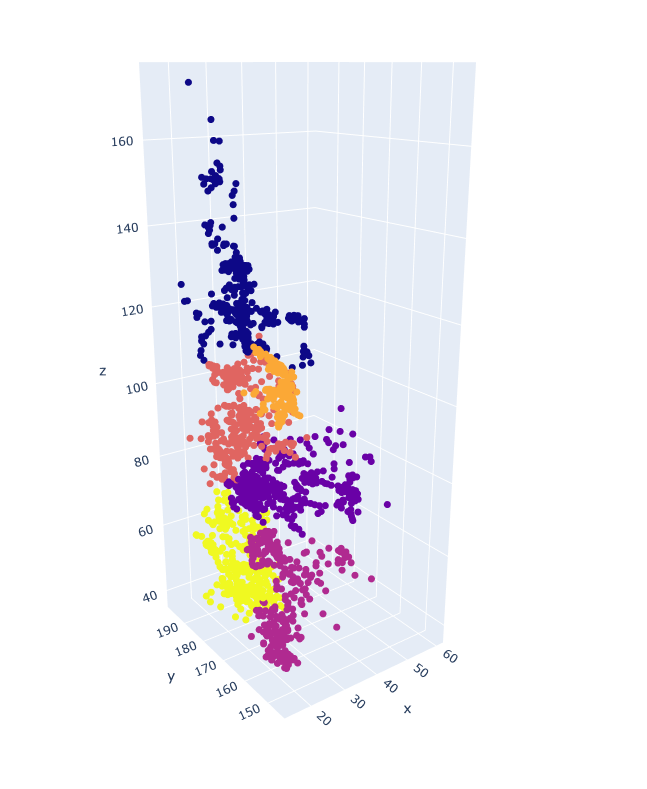


Рис.5 – Иерархическая кластеризация

4. Провести кластеризацию данных с помощью алгоритма DBSCAN.

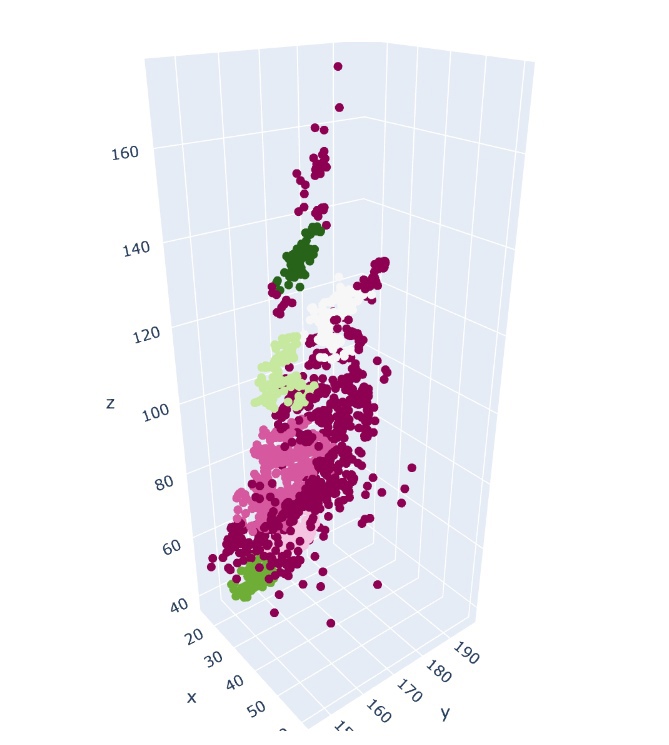


Рис.6 - Кластеризация DBSCAN