|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** | | | | |
| Институт информационных технологий (ИИТ) | |
| Кафедра Прикладной Математики (ПМ) | |

|  |  |  |  |
| --- | --- | --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №6** | | | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | | | |
|  | | | |
| Выполнил студент группы ИКБО-14-20 | | Вежновец Ф. Ю. | |
|  | |  | |
| Принял: асистент | | Горячев А. А. | |
| Практические работы выполнены | «\_\_\_»\_\_\_\_\_\_\_2023 г. | | (подпись студента) | |
| «Зачтено» | «\_\_\_»\_\_\_\_\_\_\_2023 г. | | (подпись руководителя) | |
|  |  | |  | |

Москва 2023

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.

Реализация

Результат работы программы и код предоставлена на рисунке 1 .

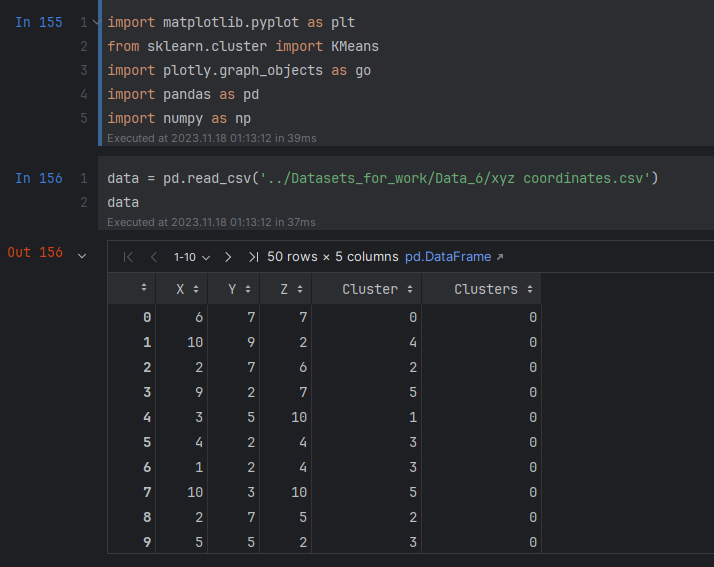


Рисунок 1 – Импорт данных

Задание 2

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

Реализация

Результат работы программы и код предоставлена на рисунке 2 - 6.

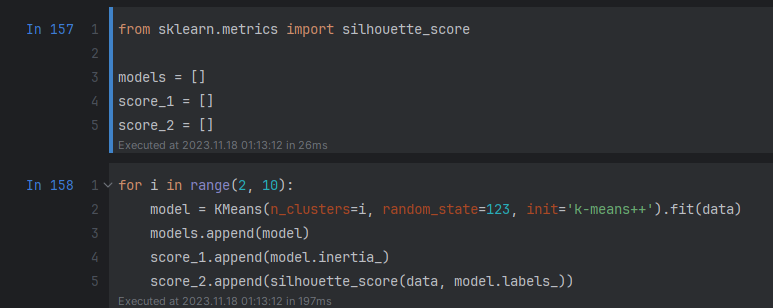


Рисунок 2 – Обучения для метода локтя и коэффициента силуэта

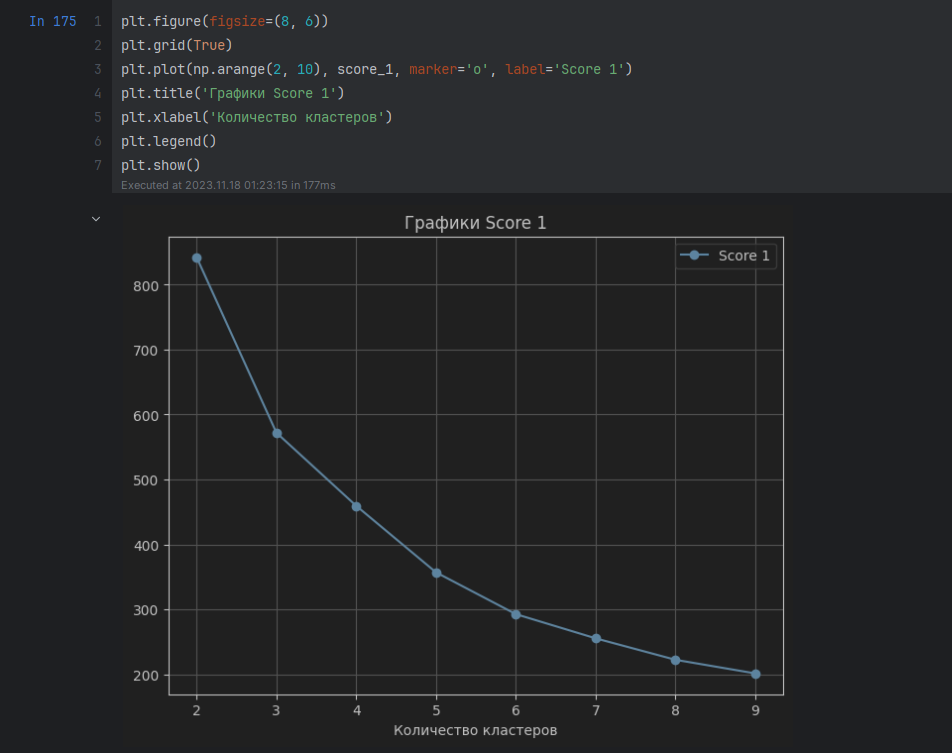


Рисунок 3 –Метод локтя

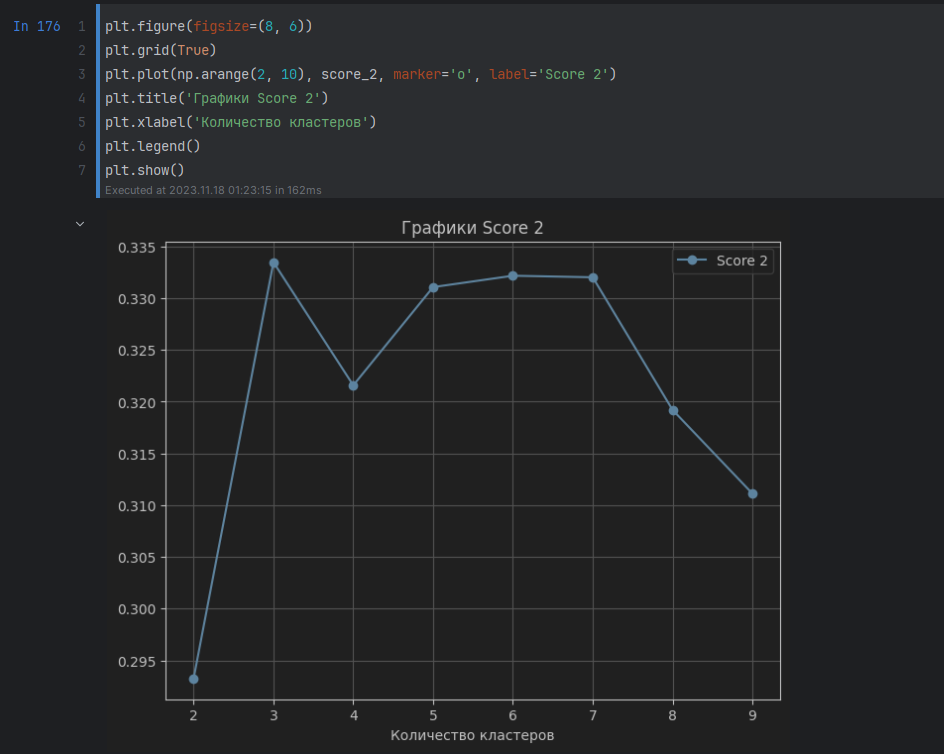


Рисунок 4 – Коэффициент силуэта

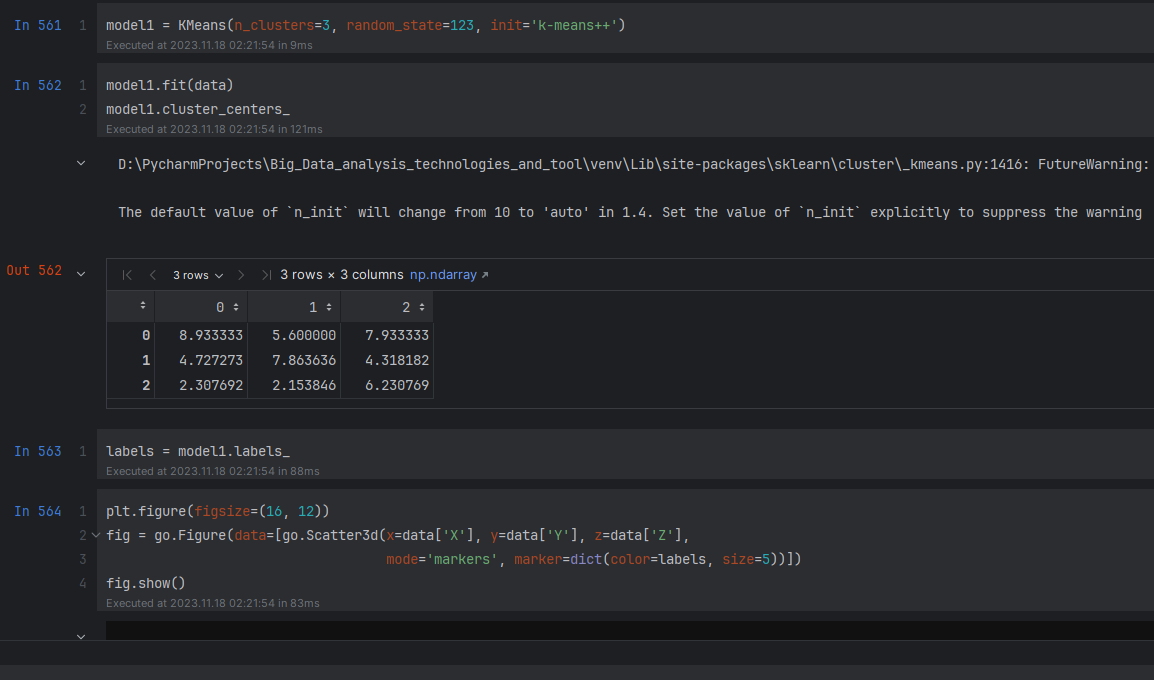


Рисунок 5 – K-means

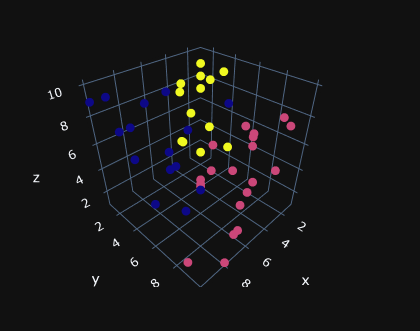


Рисунок 6 – K-means график

Задание 3

Провести кластеризацию данных с помощью алгоритма иерархической кластеризации.

Реализация

Результат работы программы и код предоставлена на рисунке 7 - 8.

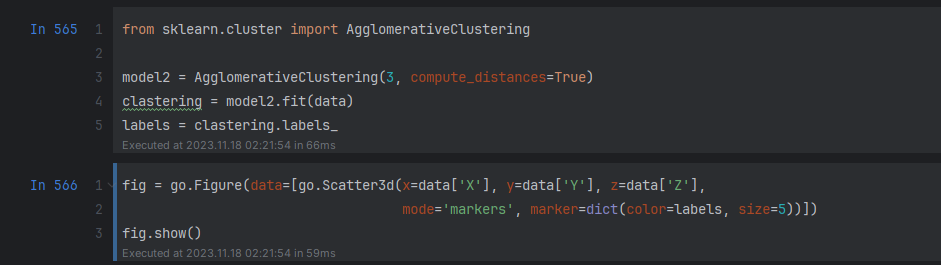


Рисунок 7 – Кластеризация алгоритма иерархической кластеризации

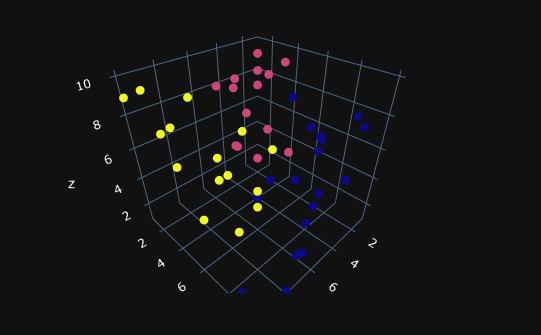


Рисунок 8 – График алгоритма иерархической кластеризации

Задание 4

Провести кластеризацию данных с помощью алгоритма DBSCAN.

Реализация

Результат работы программы и код предоставлена на рисунках 9 - 10.

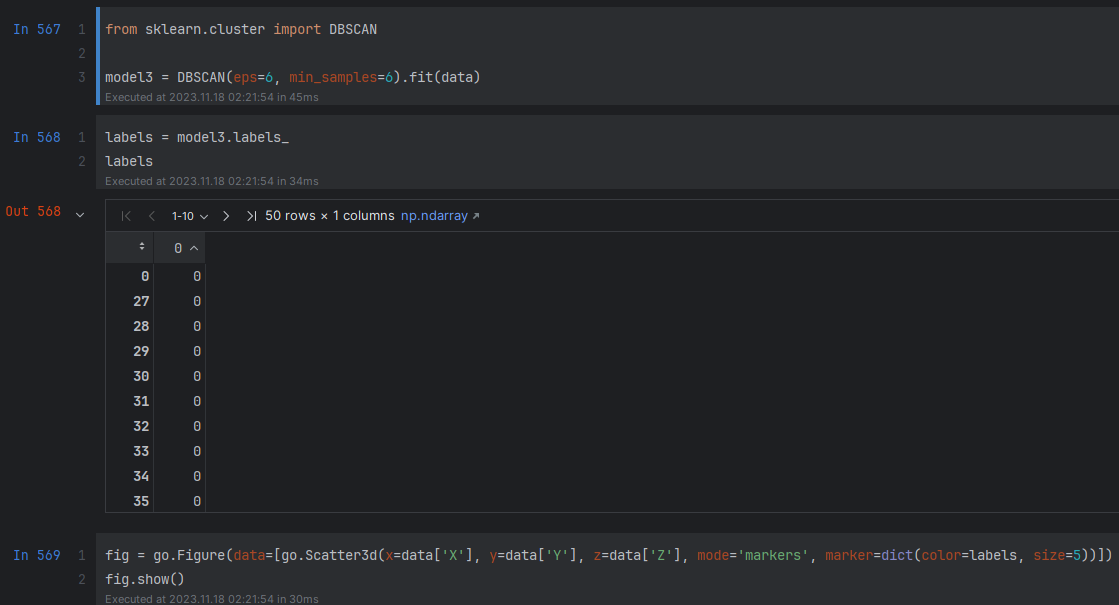


Рисунок 9 – DBSCAN

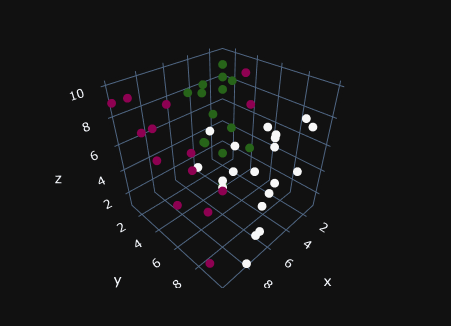


Рисунок 11 – График DBSCAN

Задание 5

Визуализировать кластеризованные данные с помощью t-SNE или UMAP, если необходимо. Если данные трехмерные, то можно использовать трехмерный точечный график.

Реализация

Результат работы программы и код предоставлена на рисунках 12 - 13.

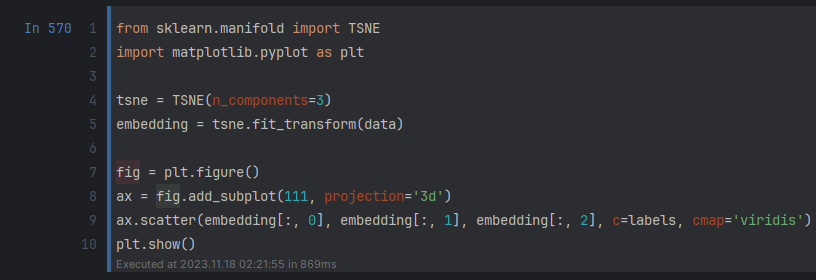


Рисунок 12 – UMAP

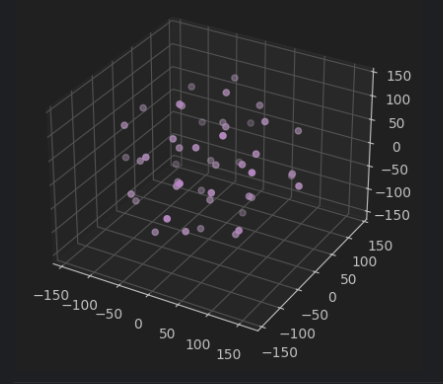


Рисунок 13 – График UMAP