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

|  |  |
| --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 2** | |
| **по дисциплине «Технологии и инструментарий анализа больших данных»** | |
|  | |
| Выполнил студент группы ИКБО-20-21 | Мухаметшин А. Р. |
| Проверил ассистент кафедры ПМ ИИТ | Тетерин Н.Н. |

**Задание 1**

Найти и выгрузить многомерные данные (с большим количеством признаков – столбцов) с использованием библиотеки `pandas`. В отчёте описать найденные данные. В качестве данных был взят файл, который относится к «Годовому обследованию предприятий» (Annual Enterprise Survey) за 2023 финансовый год, которое проводит Статистическое управление Новой Зеландии (Stats NZ). Это крупное статистическое обследование, предоставляющее информацию о финансовых результатах бизнеса по секторам экономики.

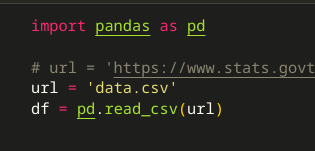


Рисунок 1 – выполнение задания 1

**Задание 2**

Вывести информацию о данных при помощи методов `.info()`, `.head()`. Проверить данные на наличие пустых значений. В случае их наличия удалить данные строки или интерполировать пропущенные значения. При необходимости дополнительно предобработать данные для дальнейшей работы с ними.

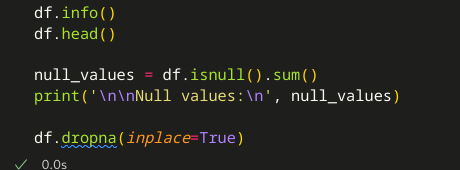


Рисунок 2 – Код к заданию 2

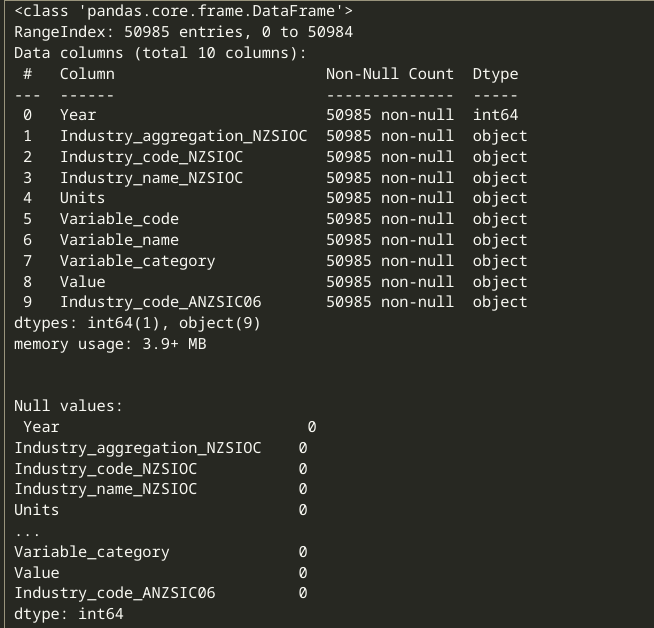


Рисунок 3 – Пример работы

**Задание 3**

Построить столбчатую диаграмму (`.bar`) с использованием модуля `graph\_objs` из библиотеки `Plotly` со следующими параметрами:

* По оси `Х` указать дату или название, по оси `У` указать количественный показатель.
* Сделать так, чтобы столбец принимал цвет в зависимости от значения показателя (`marker=dict(color=признак, coloraxis="coloraxis")`).
* Сделать так, чтобы границы каждого столбца были выделены чёрной линией с толщиной равной `2`.
* Отобразить заголовок диаграммы, разместив его по центру сверху, с `20` размером текста.
* Добавить подписи для осей `X` и `Y` с размером текста, равным `16`. Для оси абсцисс развернуть метки так, чтобы они читались под углом, равным `315`.
* Размер текста меток осей сделать равным `14`.
* Расположить график во всю ширину рабочей области и присвоить высоту, равную `700` пикселей.
* Добавить сетку на график, сделать её цвет 'ivory' и толщину равную `2`. (Можно сделать это при настройке осей с помощью `gridwidth=2`, `gridcolor='ivory'`)
* Убрать лишние отступы по краям.

Код к заданию 3 представлен на рисунке 4.

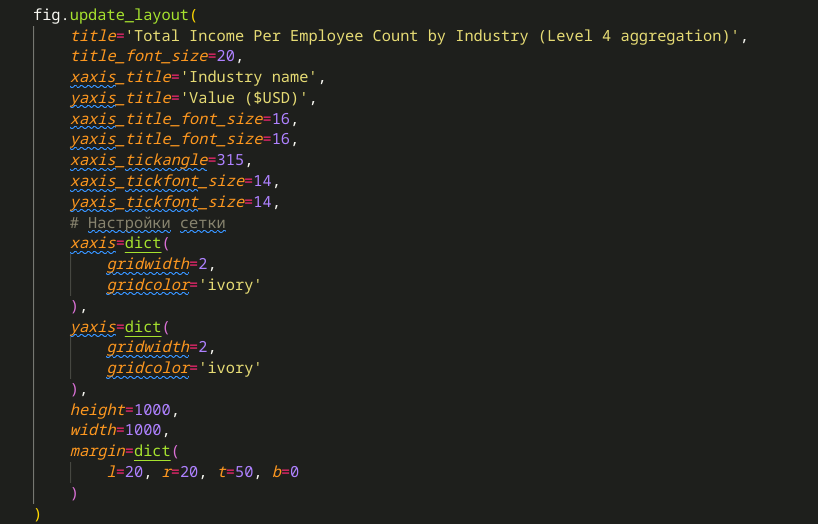
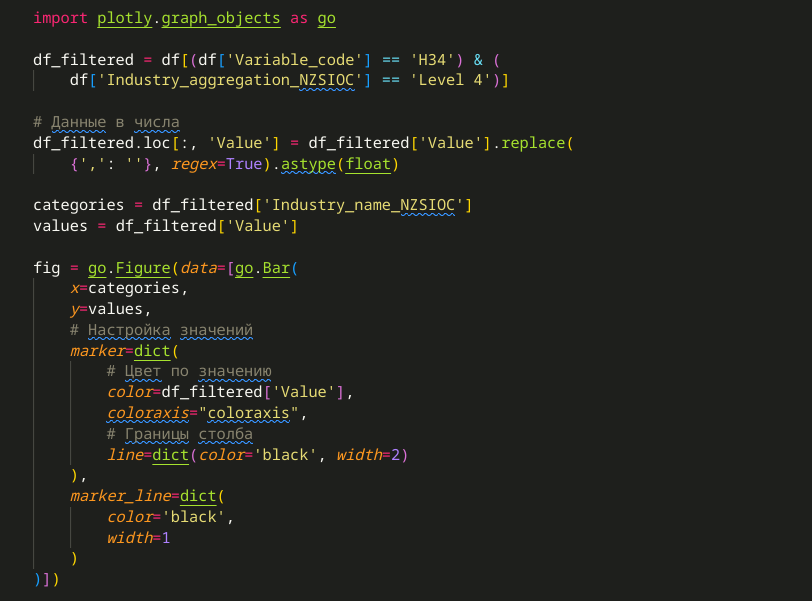


Рисунок 4 – Код к заданию 3

Пример работы представлен на рисунке 5.

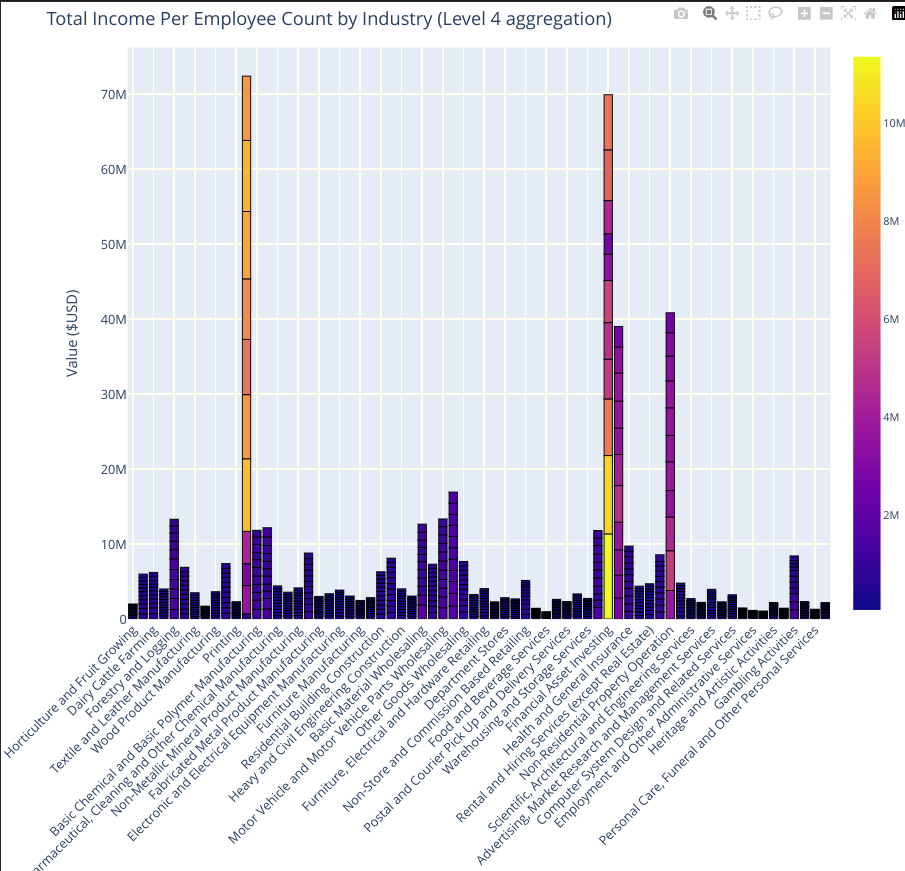


Рисунок 5 – Пример работы

**Задание 4**

Построить круговую диаграмму (`go.Pie`), использовав данные и стиль оформления из предыдущего графика. Сделать так, чтобы границы каждой доли были выделены чёрной линией с толщиной, равной `2` и категории круговой диаграммы были читаемы (к примеру, объединить часть объектов)

Код к заданию 4 представлен на рисунке 6.

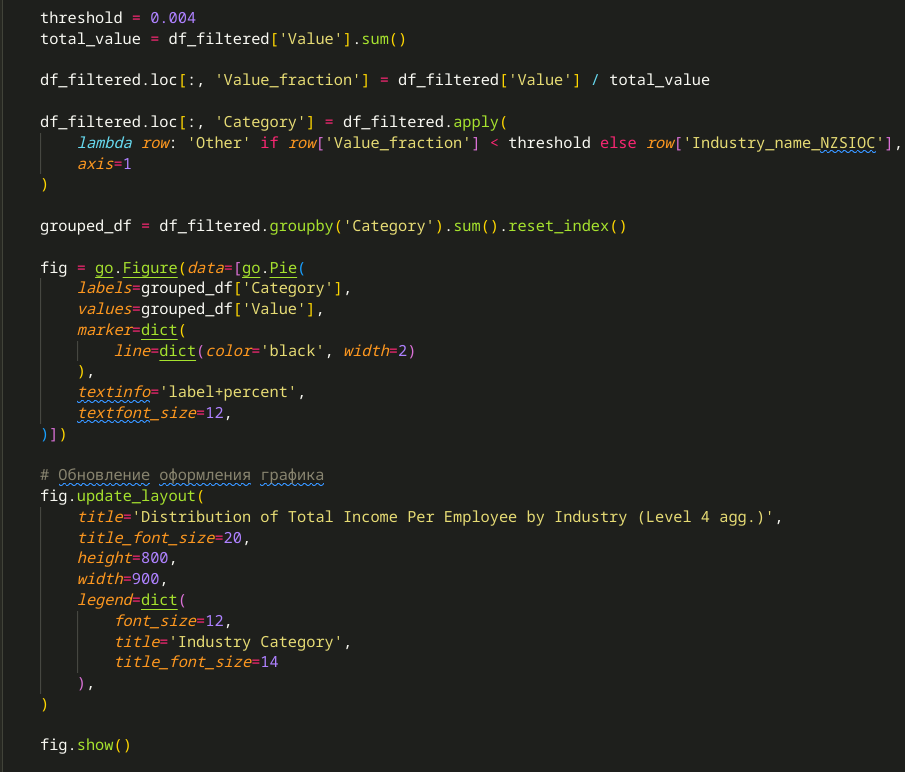


Рисунок 6 – Код к заданию 4

Пример работы представлен на рисунке 7.

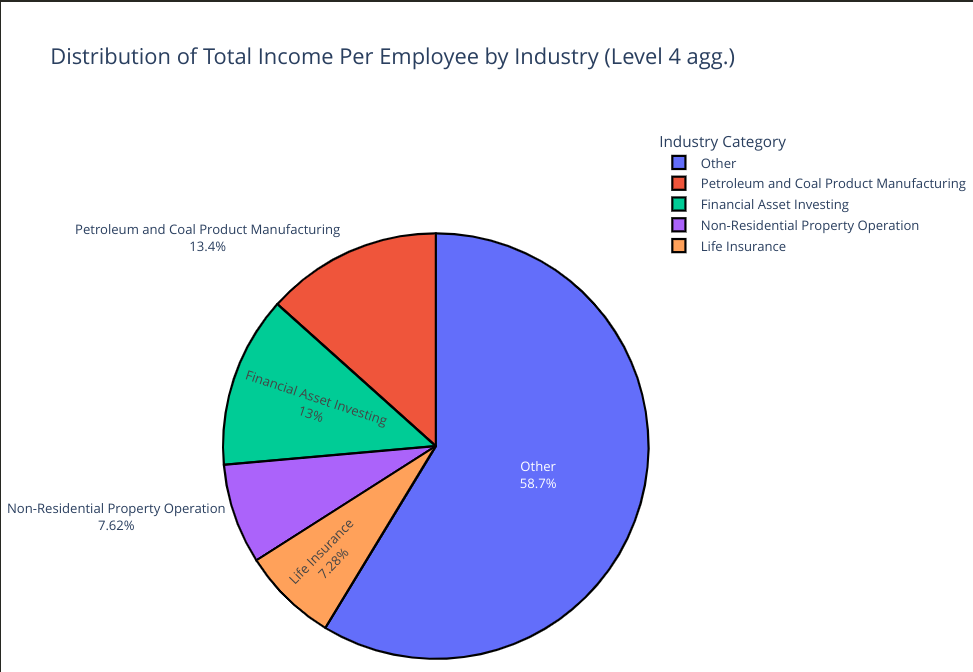


Рисунок 7 – Пример работы

**Задание 5**

Построить линейные графики, взять один из параметров и определить зависимость между `другими` несколькими (от `2` до `5`) показателями с использованием библиотеки `matplotlib`. Сделать вывод.

- Сделать график с линиями и маркерами, цвет линии `'crimson'`, цвет точек `'white'`, цвет границ точек `'black'`, толщина границ точек равна `2`.

- Добавить сетку на график, сделать её цвет `'mistyrose'` и толщину равную `2`. (Можно сделать это при настройке осей с помощью `linewidth=2`, `color='mistyrose'`).

Код к заданию 5 представлен на рисунке 8.

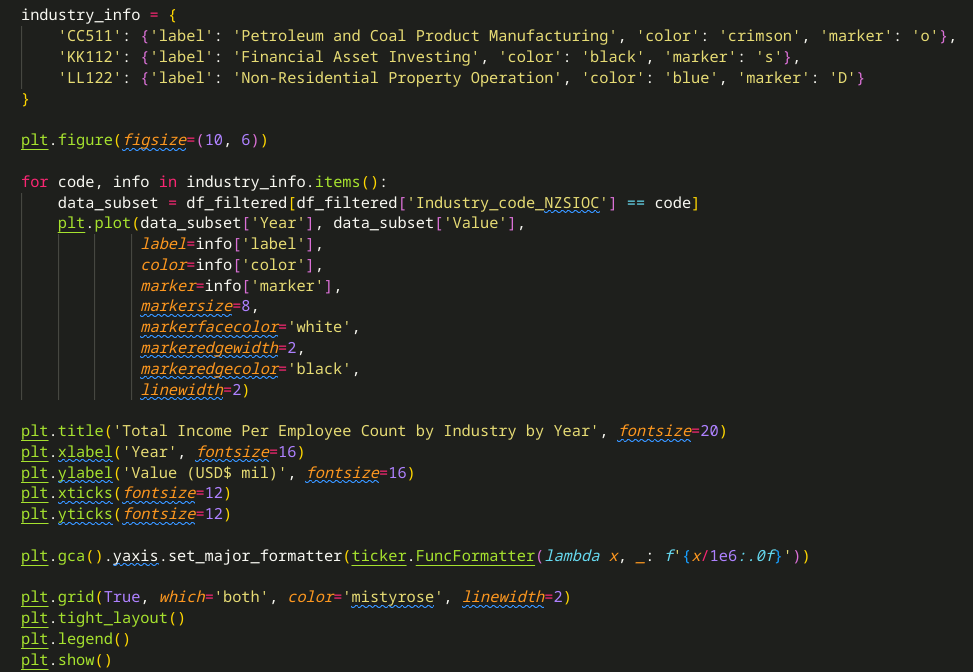


Рисунок 8 – Код к заданию 5

Пример работы представлен на рисунке 9.

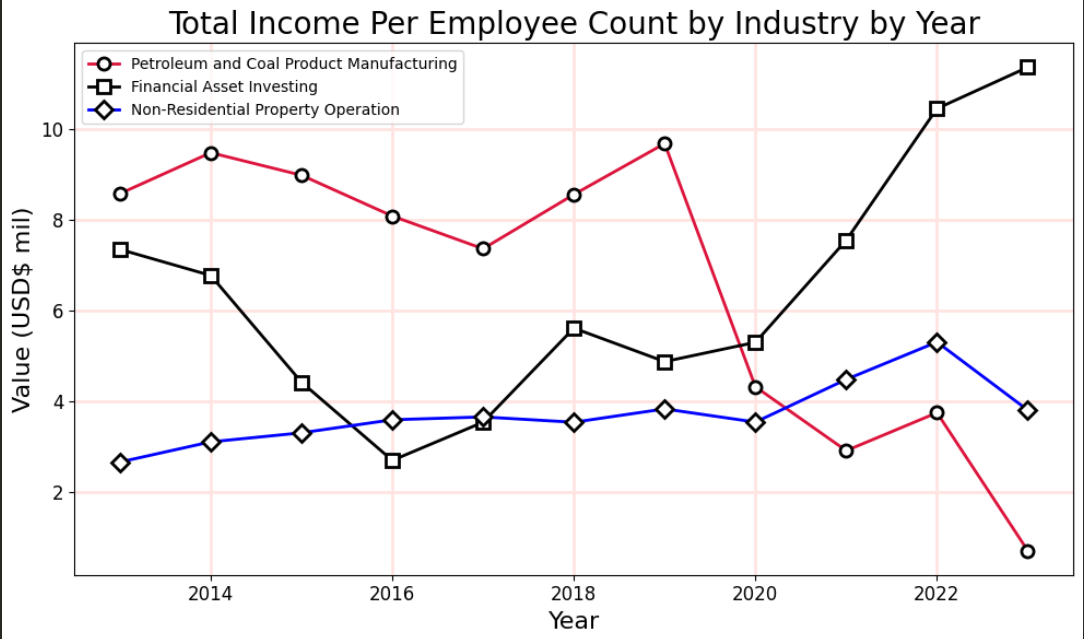


Рисунок 9 – Пример работы

**Задание 6**

Выполнить визуализацию многомерных данных, используя `t-SNE`. Необходимо использовать набор данных `MNIST` или `fashion MNIST` (можно использовать и другие готовые наборы данных, где можно наблюдать разделение объектов по кластерам). Рассмотреть результаты визуализации для разных значений перплексии.

Код к заданию 6 представлен на рисунке 10.

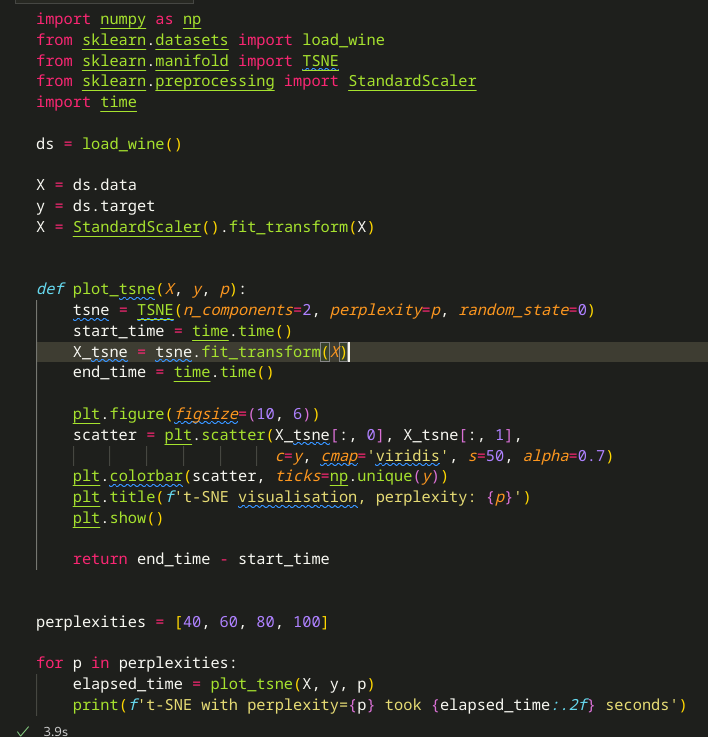


Рисунок 10 – Код к заданию 6

Пример работы представлен на рисунке 11.

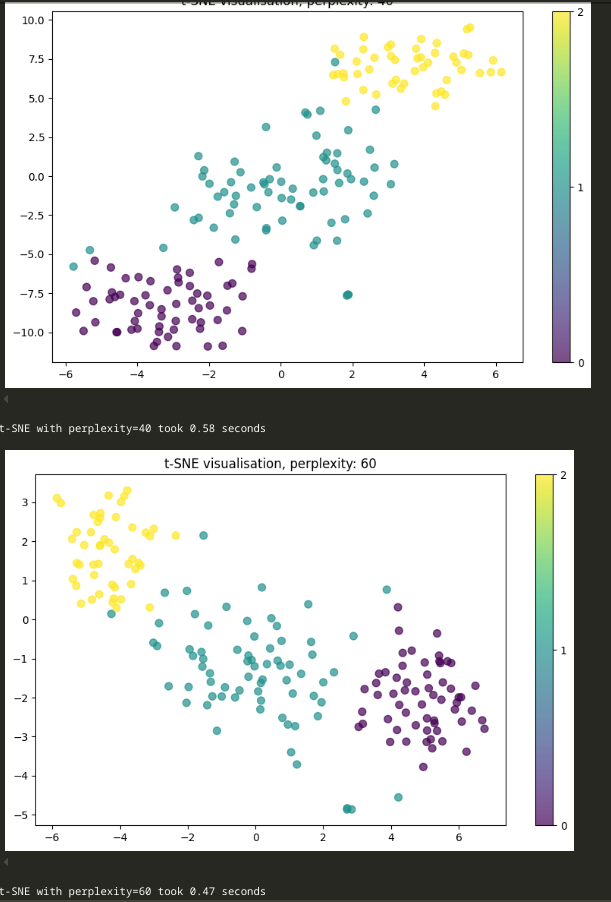


Рисунок 11 – Пример работы

**Задание 7**

Выполнить визуализацию многомерных данных, используя `UMAP` с различными параметрами `n\_neighbors` и `min\_dist`. Рассчитать время работы алгоритма с помощью библиотеки `time` и сравнить его с временем работы `t-SNE`.

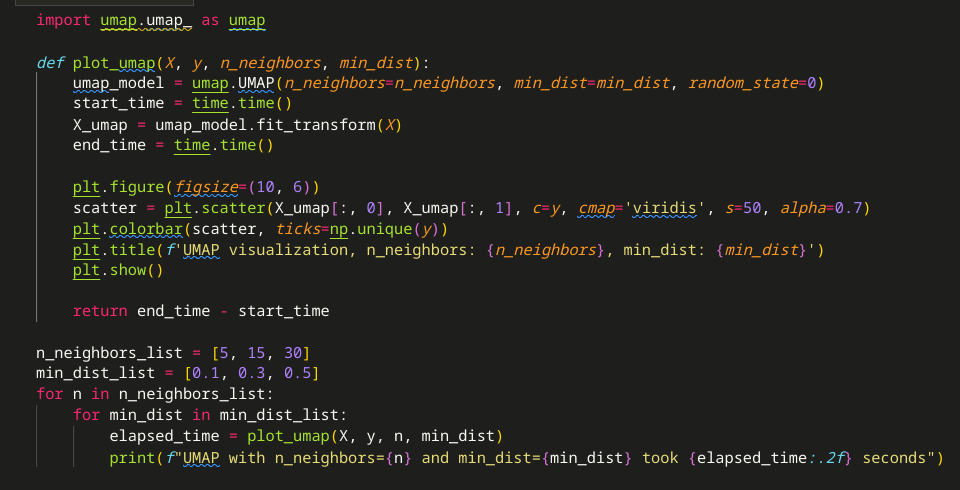


Рисунок 12 – выполнение задания

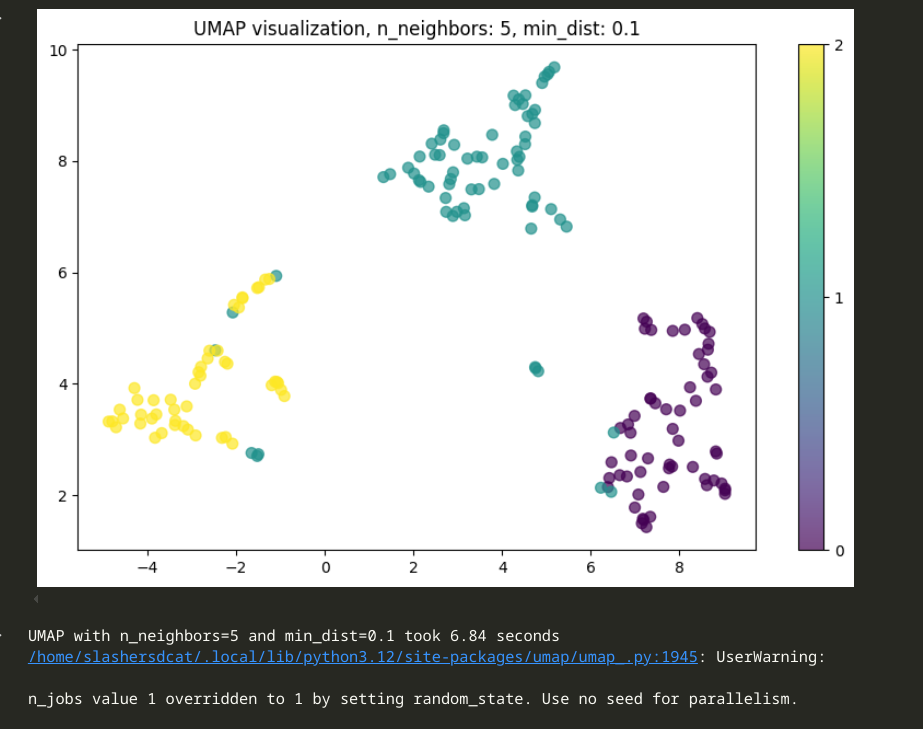


Рисунок 13 – результат работы