A logo for a science company

Description automatically generated

**Ardian - 2106638173**

**Data Mining**

**Assignment 7 - Text Mining**

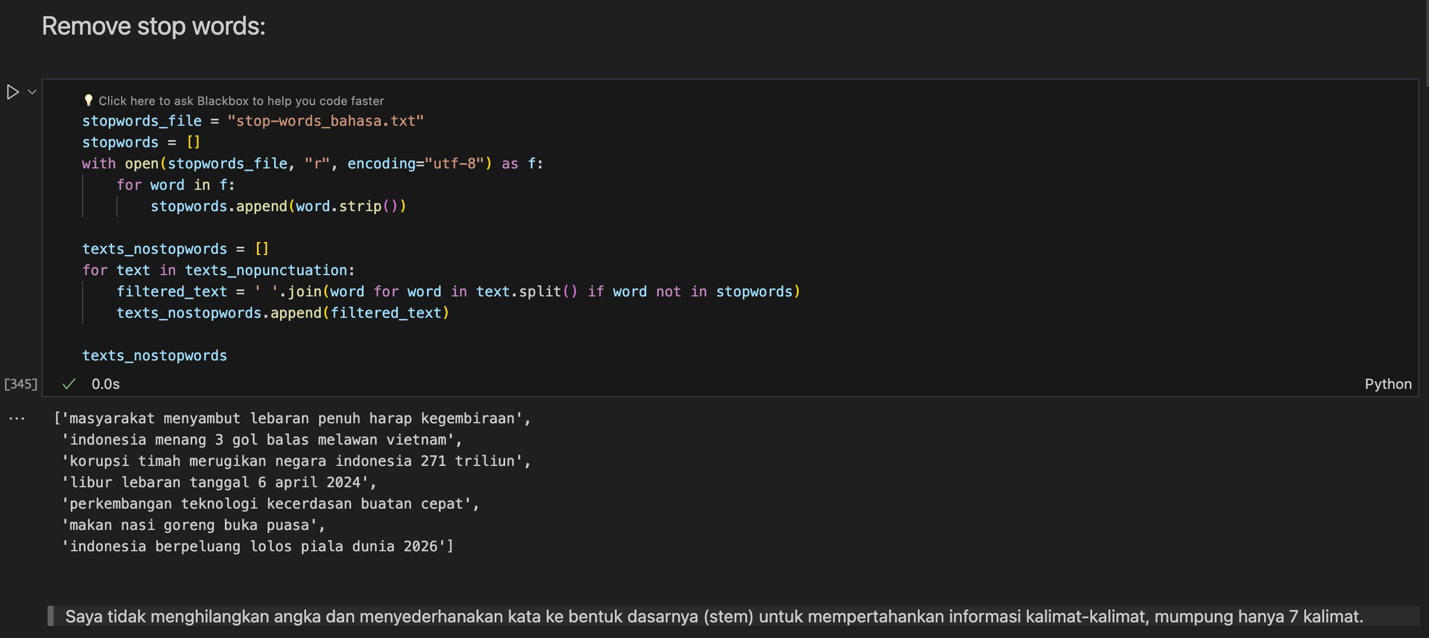
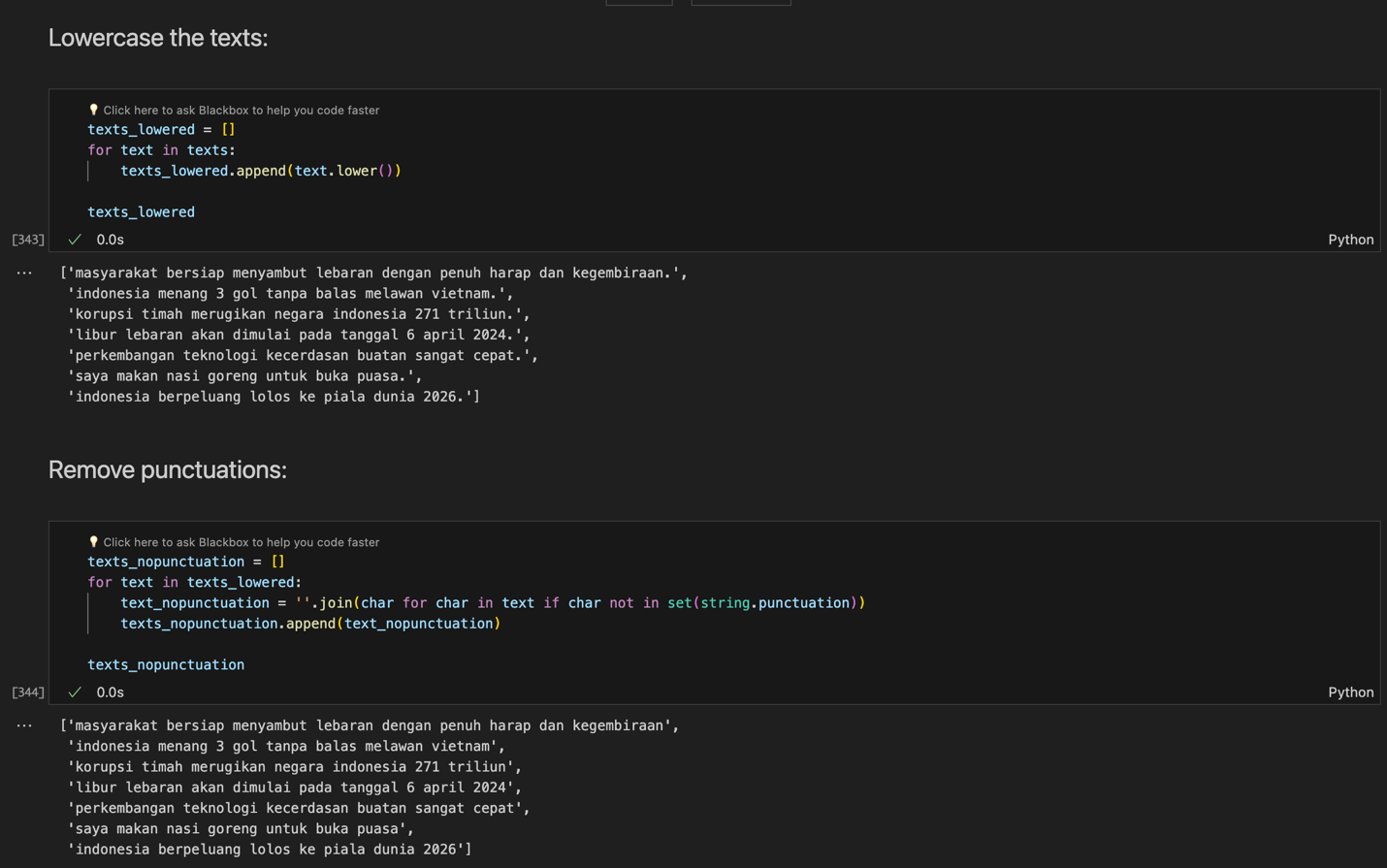
Pada tugas ini, Anda diberikan beberapa kalimat yang perlu Anda lakukan Text Clustering menggunakan Hierarchical Agglomerative Clustering. Berikut kalimatnya:

1. Masyarakat bersiap menyambut Lebaran dengan penuh harap dan kegembiraan.
2. Indonesia menang 3 gol tanpa balas melawan Vietnam.
3. Korupsi timah merugikan negara Indonesia 271 Triliun.
4. Libur lebaran akan dimulai pada tanggal 6 April 2024.
5. Perkembangan teknologi kecerdasan buatan sangat cepat.

\*karena tidak membaca soal dengan teliti, saya tidak sengaja membuat program untuk soal 1-3. Berikut notebooknya: <https://colab.research.google.com/drive/1K8_TCcoB2UnrWkHmtn8SS-BFMXuTyF0Q?usp=share_link>

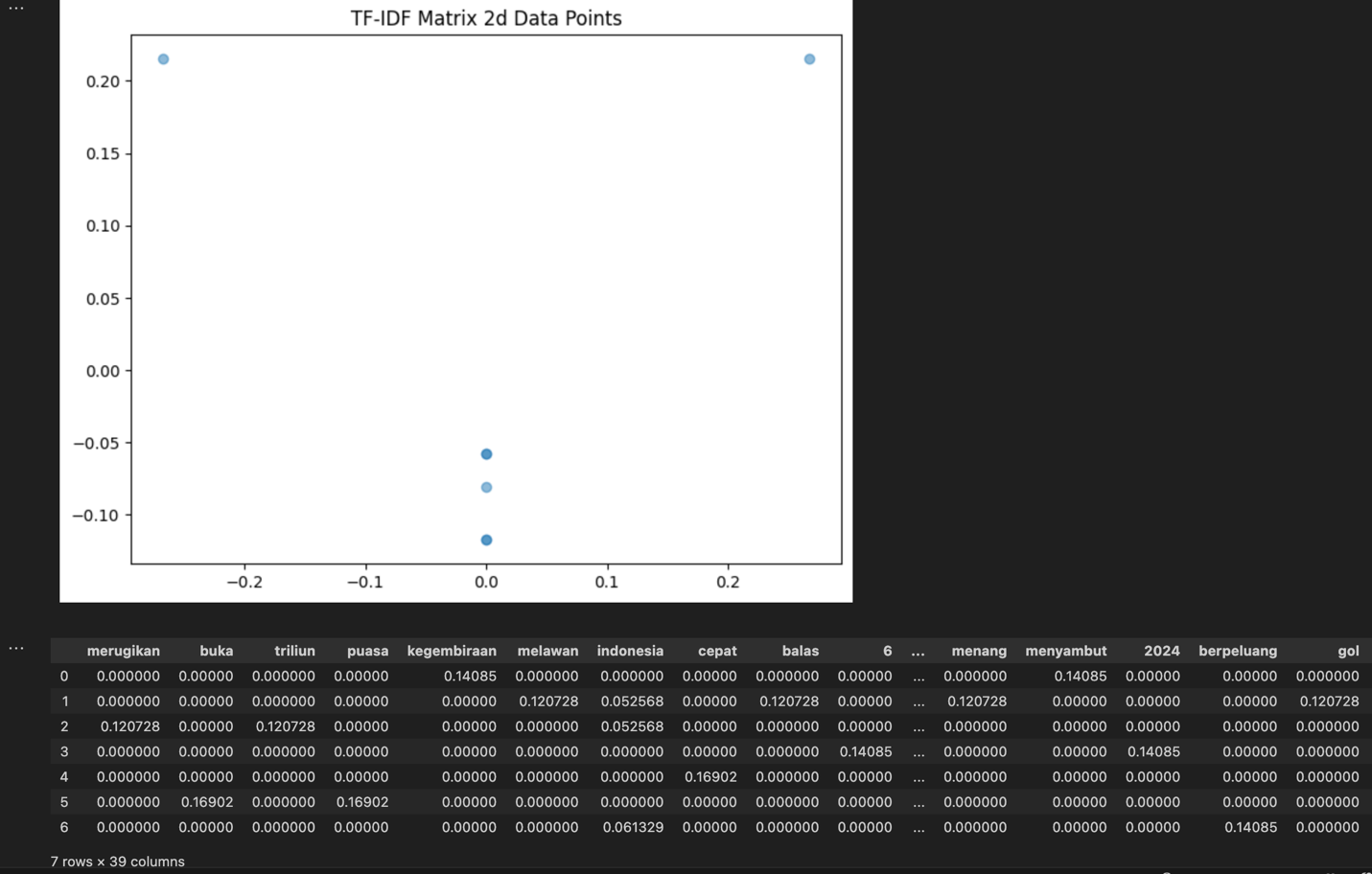
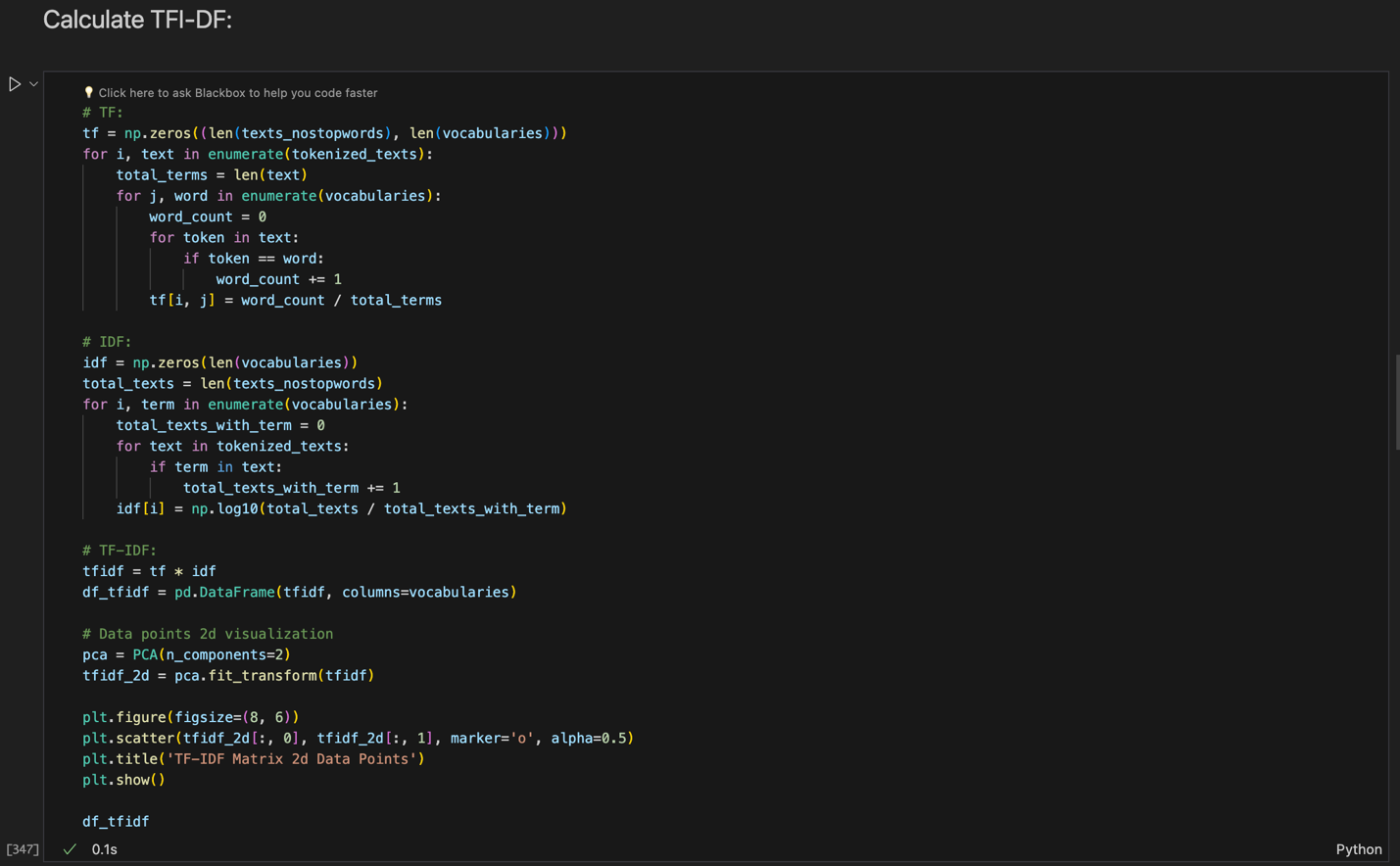
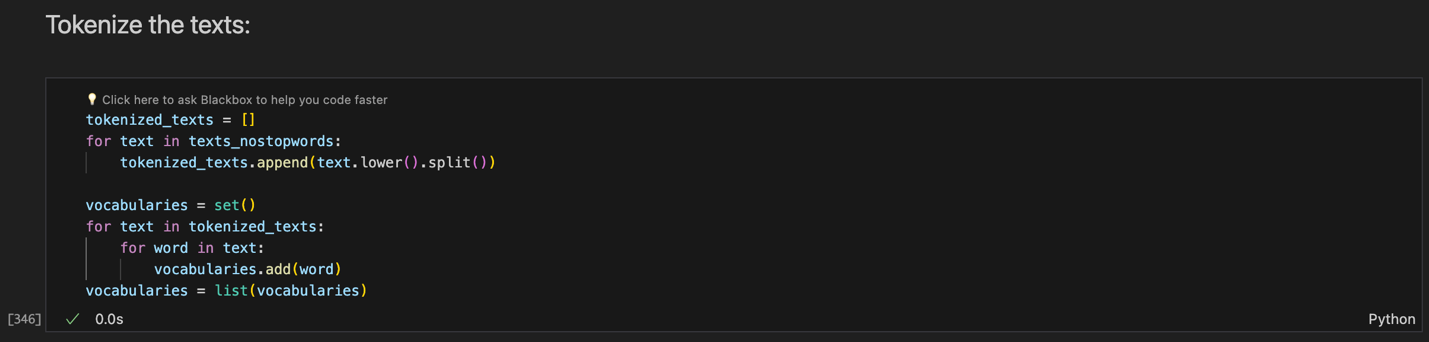
1. Lakukan tahap preprocessing untuk teks berita tersebut secara manual dan jelaskan setiap perubahan yang Anda lakukan dengan lengkap.

Jawaban:



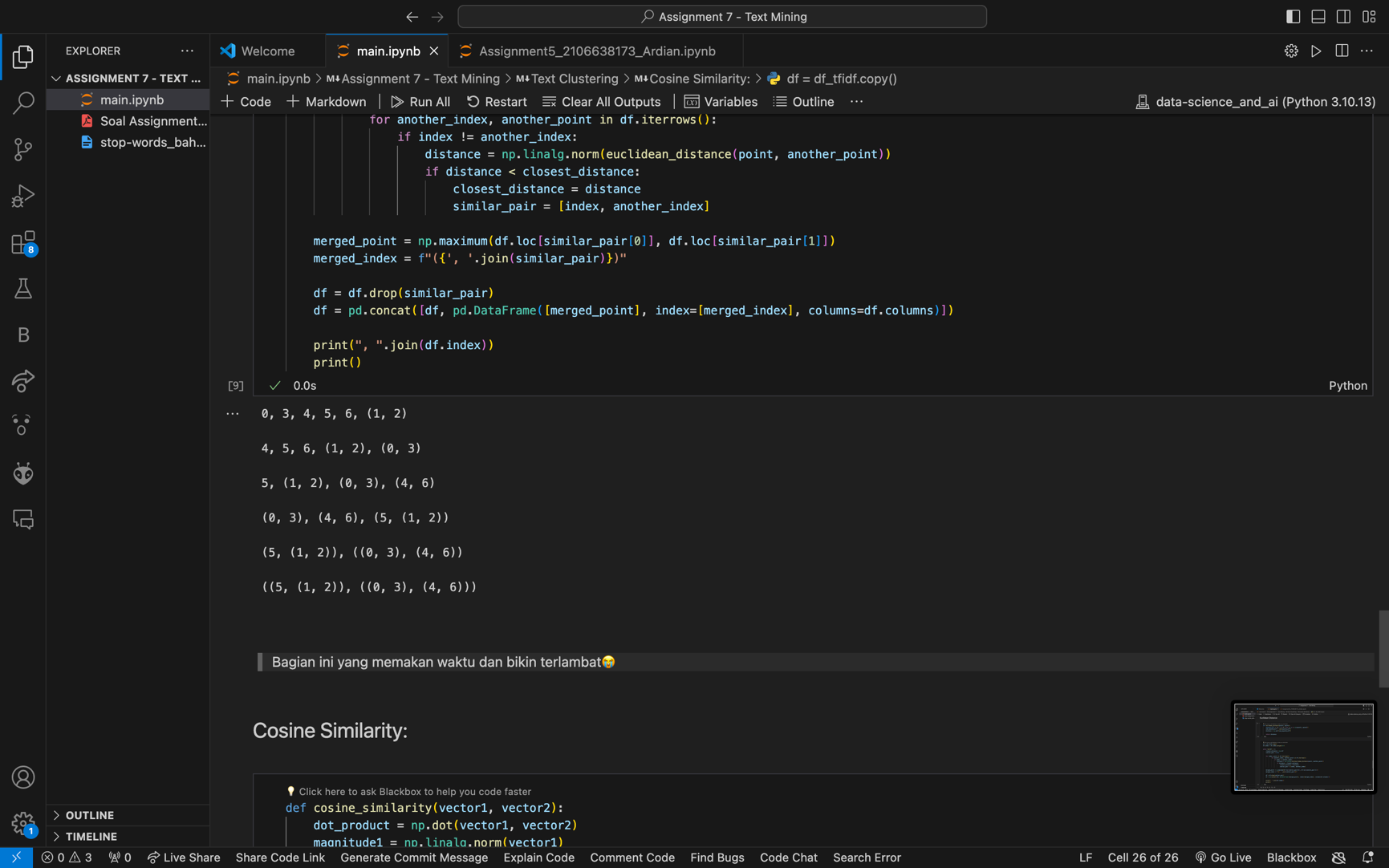
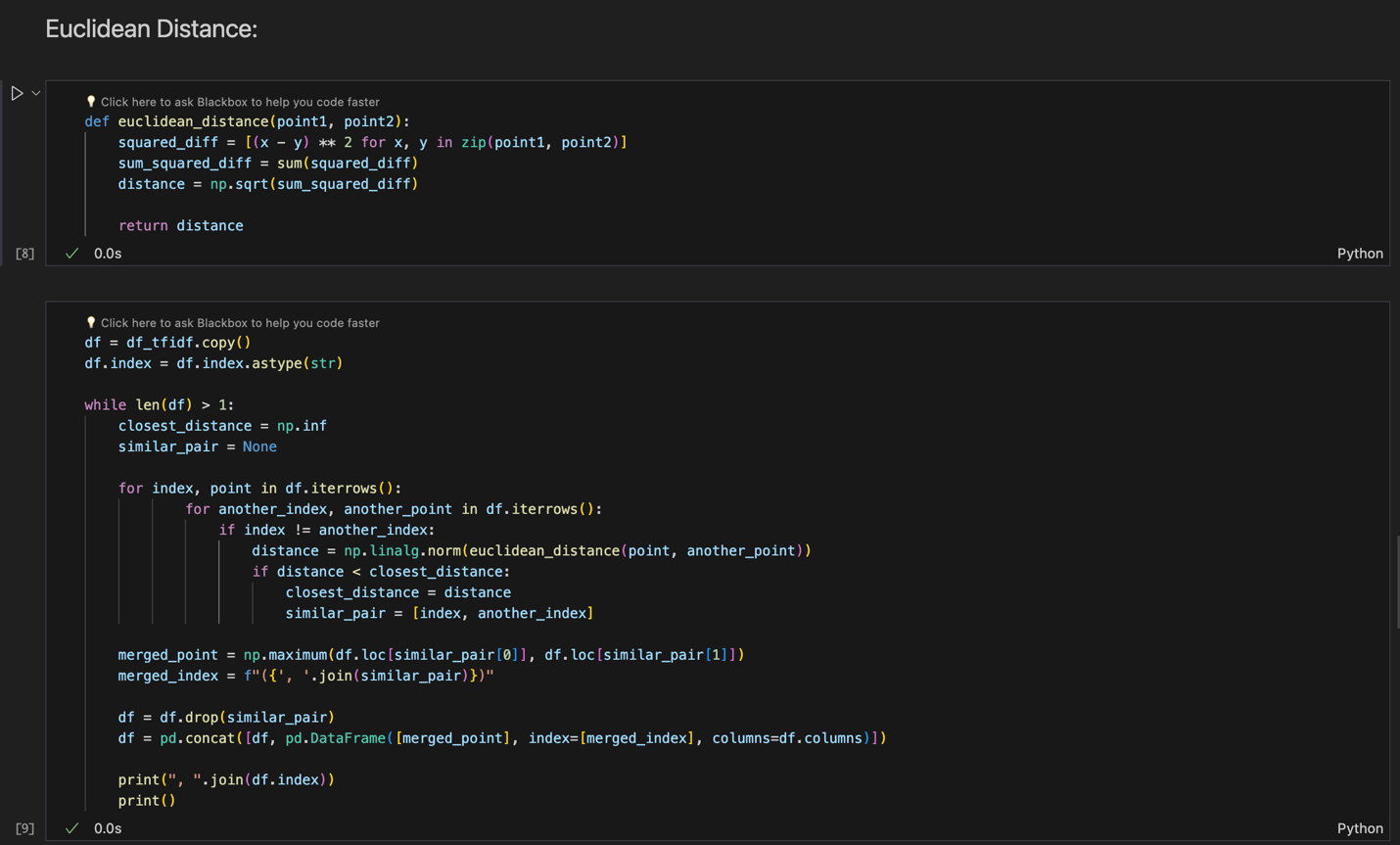
1. Lakukan vektorisasi unigram menggunakan algoritma TF-IDF secara manual dan jelaskan setiap langkah yang Anda lakukan dengan lengkap.

Jawaban:



1. Lakukan pengelompokkan kalimat menggunakan algoritma Hierarchical Agglomerative Clustering dengan menggunakan teknik pengelompokkan smallest distance dan rumus jarak euclidean distance dan cosine similarity (lakukan 2 kali pengelompokkan, 1 menggunakan euclidean distance dan 1 menggunakan cosine similarity) secara manual dan jelaskan setiap langkah yang Anda lakukan dengan lengkap.

Jawaban:

A screenshot of a computer program

Description automatically generatedA black rectangle with white text

Description automatically generated

1. Tuliskan hasil akhir yaitu dendrogram yang berhasil dibuat. Anda bisa menggambar menggunakan tulisan tangan kemudian difoto/dipindai untuk hal ini. Pastikan hasil gambar dapat terlihat dengan jelas. Anda cukup menuliskan nomor kalimat untuk mewakili kalimat tersebut dalam dendrogram.

Jawaban:

Dari hasil kalkulasi program manual di atas, berikut dendrogram yang dapat digambarkan:

A graph of a diagram of a city

Description automatically generated with medium confidence

Maaf atas keterlambatannya, saya struggling di nomor 3, saya sangat berharap jawaban saya diberi nilai. Terima kasih atas pengertiannya.😭