**MODUL ANALISIS DATA**

**AGGLOMERATIVE CLUSTERING**

****

**Dosen Pengampu:**

**Prizka Rismawati Arum, S.Si., M.Stat**

**Disusun oleh :**

Aqsal Maulana (B2A020043)

**PROGRAM STUDI S1 STATISTIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS MUMAMMADIYAH SEMARANG**

**202****3**

# KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Allah SWT atas berkat dan rahmat-Nya, sehingga laporan dengan judul **“Modul Agglomerative Clustering Menggunakan Google Collabs (Python)”** dapat terselesaikan dengan baik tanpa halangan suatu apapun.

Penulisan makalah ini bertujuan guna memenuhi UTS mata kuliah Analisis Data dengan dosen pengampu Ibu Prizka Rismawati Arum, S.Si., M.Stat. Makalah ini juga bertujuan untuk dapat meningkatkan pengetahuan, baik penyusun maupun pembaca.

Penyusun juga menyadari atas ketidaksempurnaan makalah ini. Oleh sebab itu, penyusun menerima dengan senang hati apabila terdapat kritik dan saran untuk membuat makalah ini menjadi lebih baik.

Semarang, 22 Mei 2023

Penyusun

# DAFTAR ISI

[KATA PENGANTAR ii](#_Toc135717981)

[DAFTAR ISI iii](#_Toc135717982)

[BAB I 1](#_Toc135717983)

[Bahasa pemrograman merupakan bahasa yang berisi sintaks yang membentuk sebuah perintah tertentu dan akan diterjemahkan menjadi logika yang dimengerti oleh komputer. Hingga saat ini sudah banyak sekali jumlah dari bahasa pemrograman. Python ini sering digunakan developer karena dapat digunakan dalam membuat/membangun aplikasi baik itu untuk desktop, web atau mobile .Python merupakan salah satu bahasa pemrograman yang sering digunakan. Python termasuk jenis bahasa pemrograman *high level language.* 1](#_Toc135717984)

[BAB II 2](#_Toc135717985)

[**A.** **Pengertian Python** 2](#_Toc135717986)

[**B.** **Function dalam Python** 3](#_Toc135717987)

# BAB I

**PENDAHULUAN**

**A**. Latar Belakang

Bahasa pemrograman merupakan bahasa yang berisi sintaks yang membentuk sebuah perintah tertentu dan akan diterjemahkan menjadi logika yang dimengerti oleh komputer. Hingga saat ini sudah banyak sekali jumlah dari bahasa pemrograman. Python ini sering digunakan developer karena dapat digunakan dalam membuat/membangun aplikasi baik itu untuk desktop, web atau mobile .Python merupakan salah satu bahasa pemrograman yang sering digunakan. Python termasuk jenis bahasa pemrograman *high level language.*

Python selain digunakan para developer, para Data science untuk melakukan analisis statistic terhadap suatu data. Banyak sekali package – package untuk mendukung proses analisis data. Dari uji normalitas, uji anova hingga metode yang kompleks seperti analisis spasial. Pada topik pembahasan kali ini saya akan membahas bagaimana cara membuat cluster / clustering menggunakan Bahasa python.

Clustering merupakan salah satu teknik data mining yang digunakan untuk mendapatkan kelompok-kelompok dari obyek-obyek yang mempunyai karakteristik yang umum di data yang cukup besar. Tujuan utama dari metode clustering adalah pengelompokan sejumlah data/obyek ke dalam cluster (group) sehingga dalam setiap cluster akan berisi data yang semirip mungkin [1].

Clustering terbagi menjadi banyak sekali metode salah satunya metodologi pengelompokan berbasis konektivitas yang disebut hierarchical clustering. Metode ini bekerja dengan mengatur titik data yang berdekatan satu sama lain menurut kesamaan atau kedekatan. Titik data yang berdekatan satu sama lain dianggap lebih mirip atau terhubung daripada yang berjauhan.

Hubungan hierarkis antar kelompok dapat dilihat dalam dendrogram, representasi seperti pohon yang dibuat oleh pengelompokan hierarkis. Grup terbesar dendrogram, yang terdiri dari semua titik data, ditemukan di bagian atas, sedangkan titik data individu ditemukan di bagian bawah. Dendrogram dapat dipotong pada ketinggian yang berbeda untuk menghasilkan jumlah cluster yang bervariasi.

**B**. Rumusan Masalah

1. Bagaimana cara analisis Agglomerative Clustering dalam python ?

**C**. Tujuan

1. Untuk mengetahui bagaimana cara analisis Agglomerative Clustering dalam python.

# BAB II

**PEMBAHASAN**

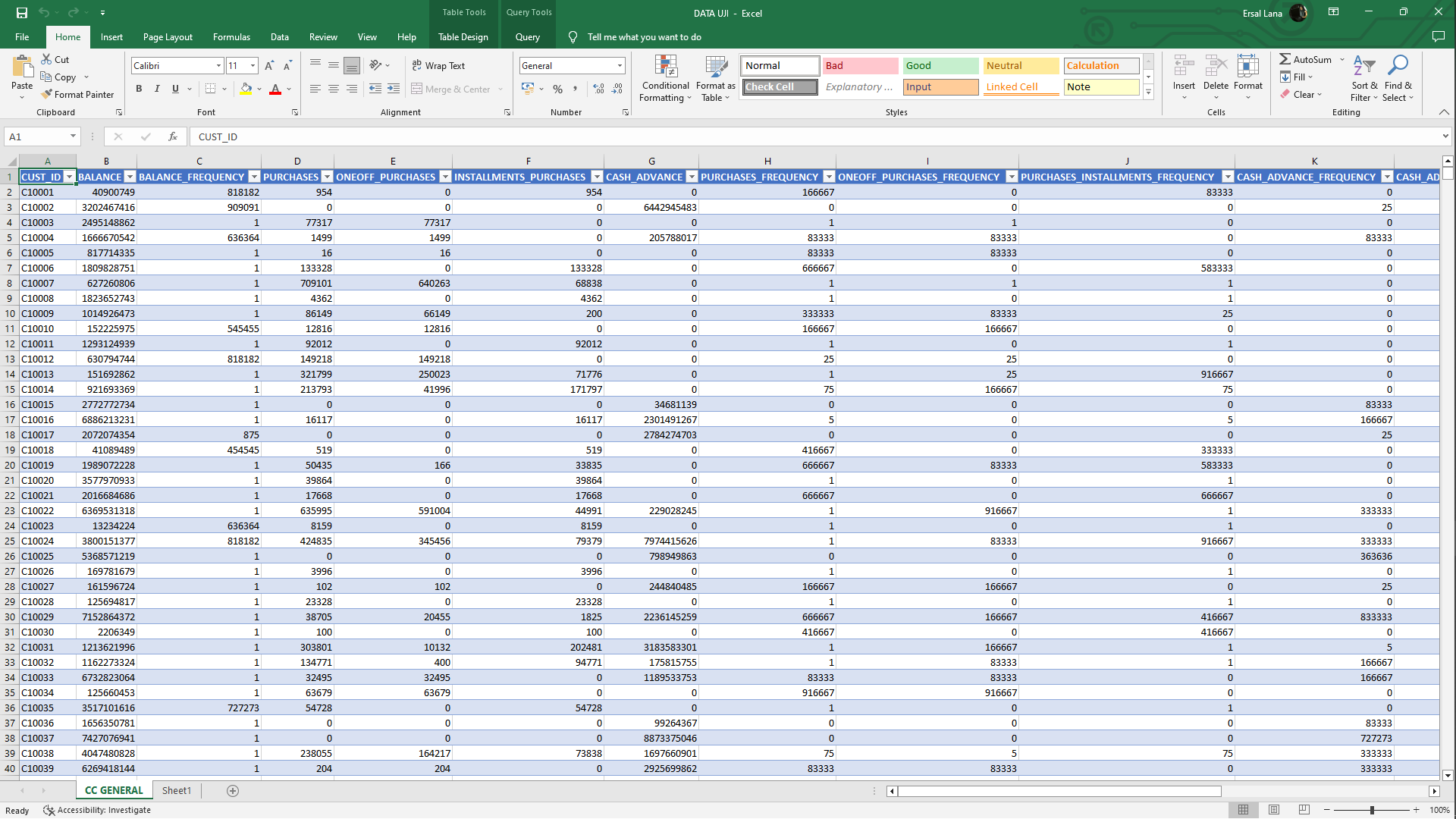
1. **Hierarchical Clustering**

Hierarchical Clustering adalah metode clustering yang umum digunakan. Metode ini bekerja dengan mengatur titik data yang berdekatan satu sama lain menurut kesamaan atau kedekatan. Titik data yang berdekatan satu sama lain dianggap lebih mirip atau terhubung daripada yang berjauhan.

Dendrogram dibentuk dengan menggabungkan atau membagi cluster secara iteratif menurut metrik jarak atau kesamaan antara titik data. Pemisahan atau penggabungan klaster berulang kali terjadi hingga semua titik data berada dalam satu klaster, atau hingga sejumlah klaster tercapai.

Hierarchical Agglomerative Clustering (HAC) atau yang dikenal dengan pendekatan bawah ke atas adalah sebuah metode yang memberikan informasi yang lebih berguna daripada kumpulan cluster tidak terstruktur dari metode pengelompokan datar. Jumlah cluster tidak perlu ditentukan terlebih dahulu untuk proses clustering ini. Algoritme bottom-up awalnya menganggap setiap potongan data sebagai cluster tunggal sebelum menggabungkan pasangan cluster sampai hanya tersisa satu cluster yang berisi semua data.

1. Cara melakukan Analisis Agglomerative Clustering
2. Memasukkan data ke Excel

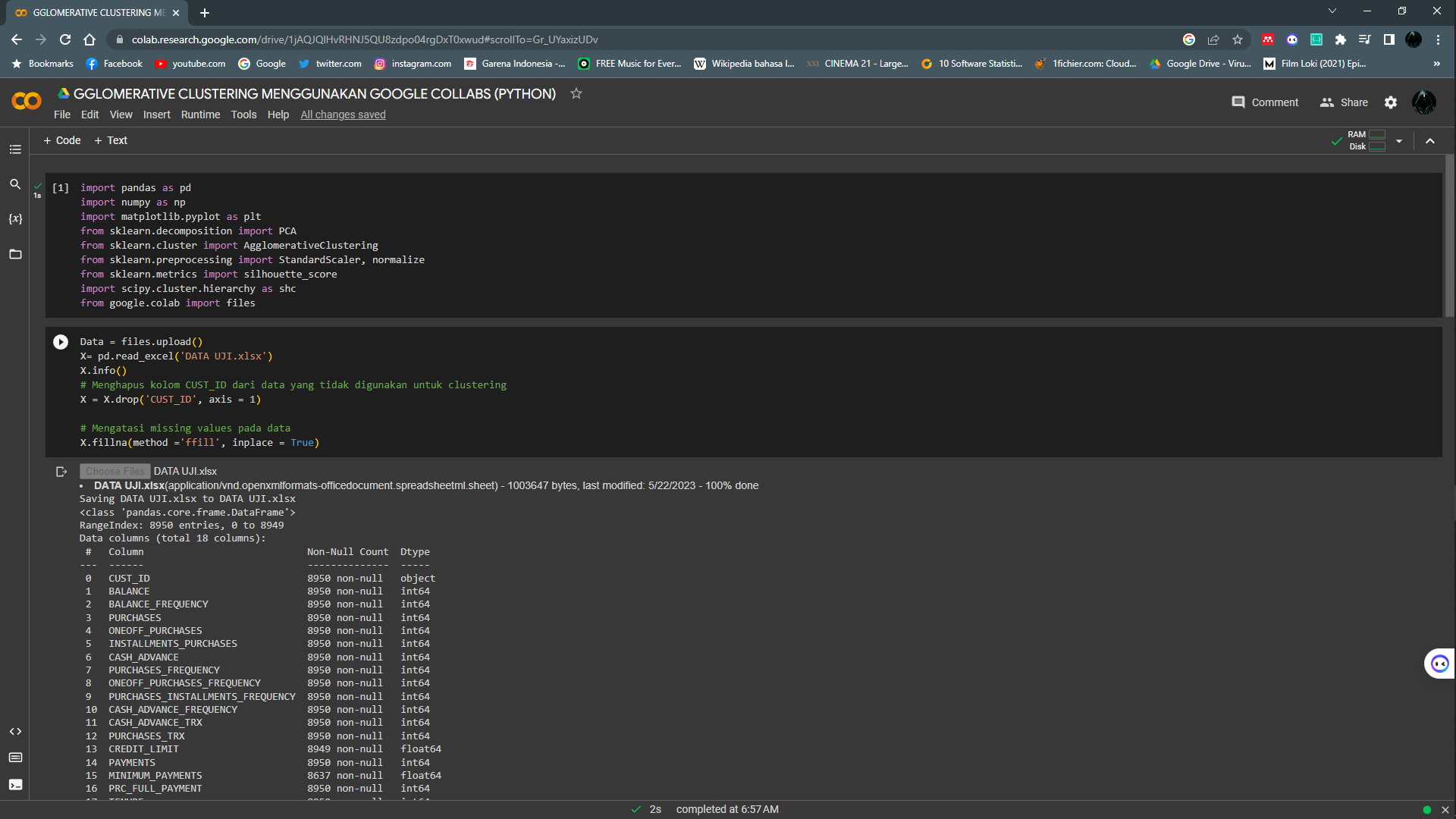


Bersumber : Kaggle.com

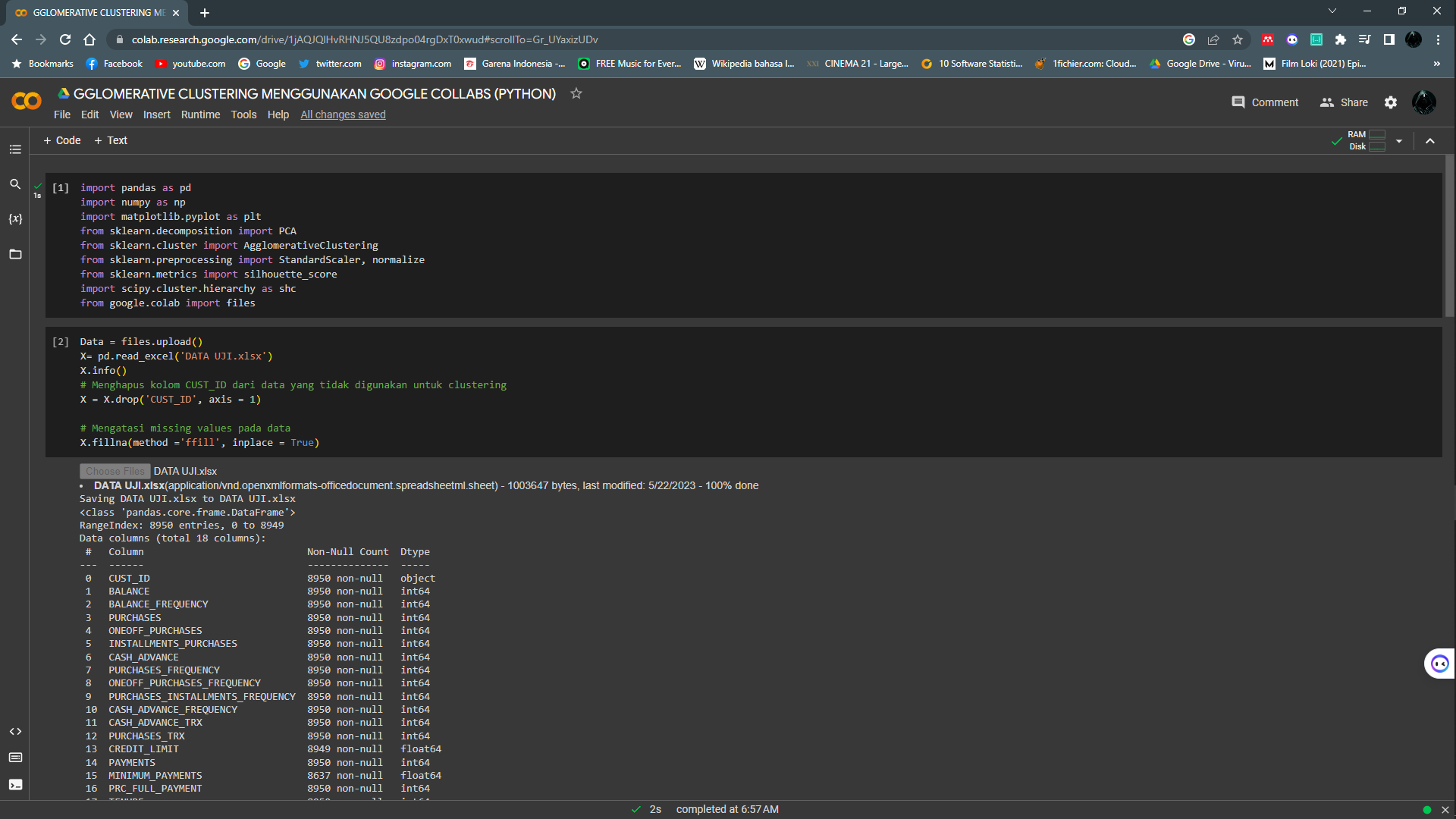
<https://docs.google.com/spreadsheets/d/1UzK_14e5RKjBmYBS9ZzDalbQ9zWMtPos/edit?usp=sharing&ouid=116133336834214899503&rtpof=true&sd=true>

untuk data bisa diunduh di link diatas.

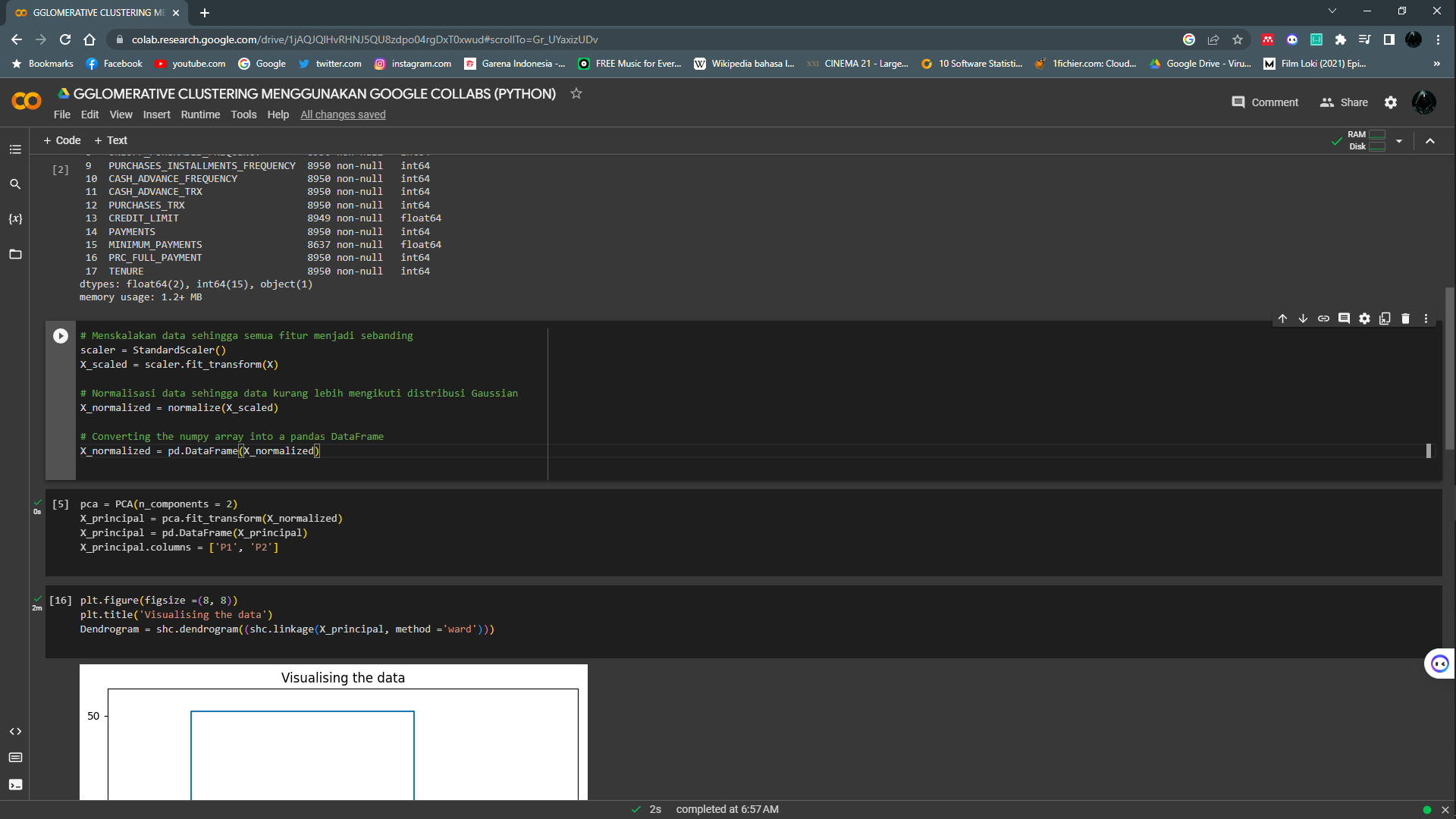
1. Menginport package



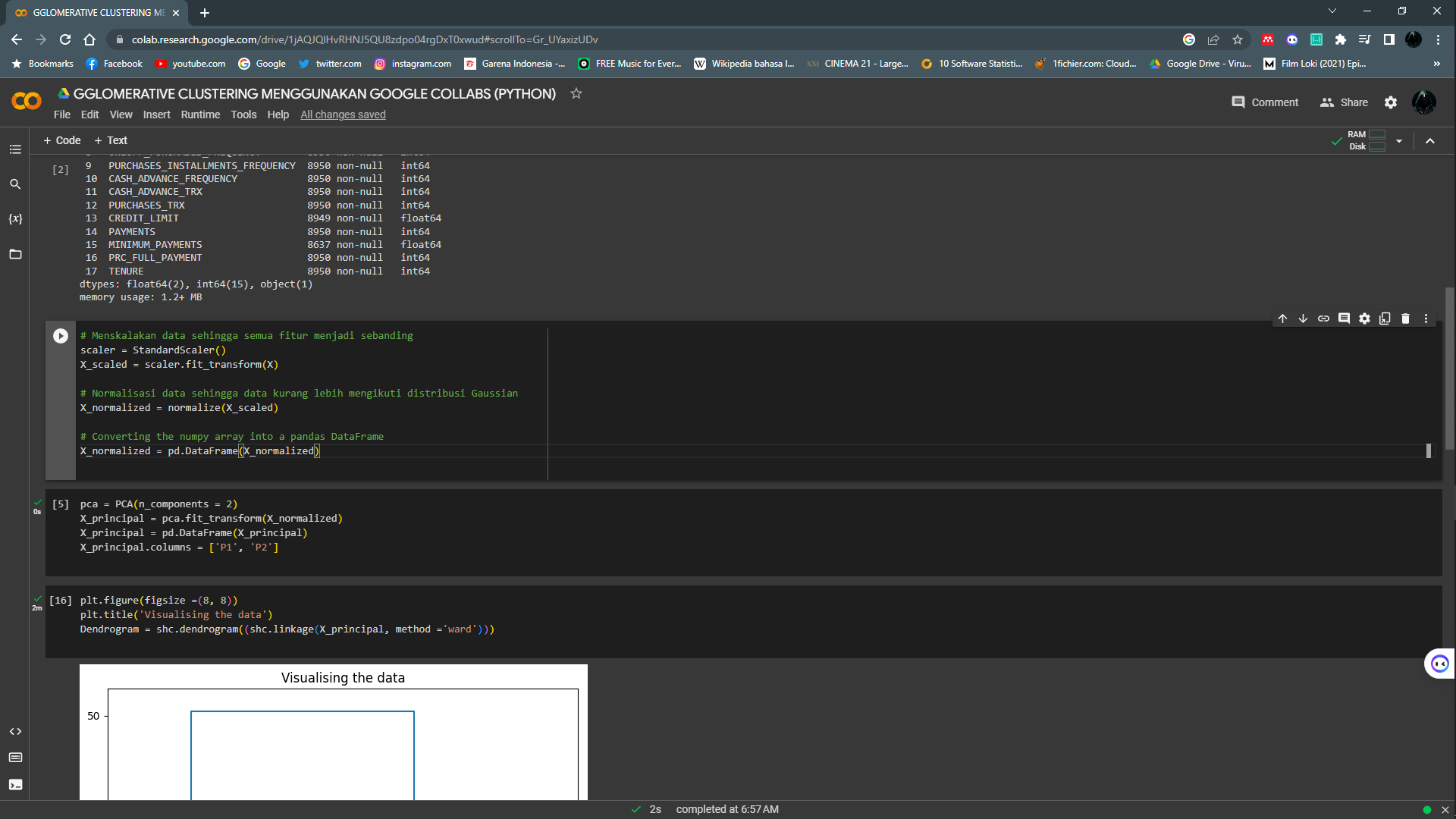
1. Upload data dan membersihkan data



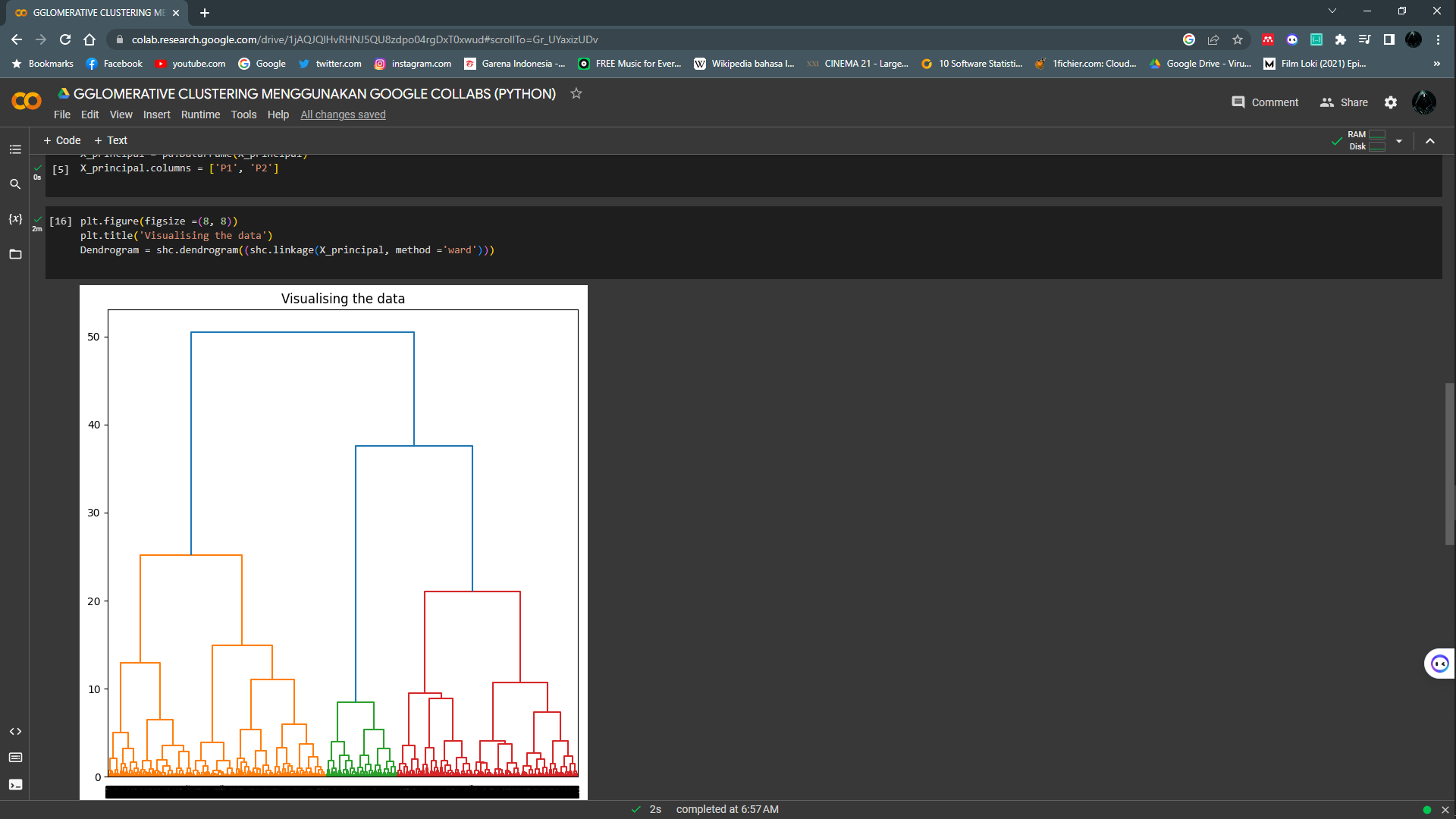
1. Mengubah data agar sebanding



1. Mengurangi dimensi Data



1. Membuat Dendogram



1. Membuat Visual Cluster

