

LAPORAN UAS MINI PROJEK
BIG DATA
“Clustering Segmentation Data Customer”



KELOMPOK :

Arif fathi falah / Membuat PPT

Erson Fairman Weryadi / Membuat Clustering Data Customer

Izzudin Ayyash / Membuat Makalah

Arrafi Thirza Vidiawan / Mencari Refrensi

Hafidzul Wahid / Membuat Makalah

STT TERPADU NURUL FIKRI
PROGRAM STUDI TEKNIK INFORMATIKA/SISTEM INFORMASI
DEPOK
2022

- **Deskripsi / Latar Belakang**

Dalam kasus segmentasi ini terdapat cara penyelesaian yaitu dengan teknik Clustering. Clustering merupakan proses untuk mengelompokkan suatu data dengan cara membandingkan kemiripan antar data, sehingga setiap data masuk ke dalam Cluster (grup). Clustering termasuk kedalam unsupervised learning (pembelajaran tidak terbimbing), karena ketika pengelompokan data didasarkan atas kemiripan dan ketidakmiripan antar datanya. Analisis cluster telah banyak digunakan termasuk riset pasar, pengenalan pola, analisis data, dan pengolahan citra. Dalam bisnis, clustering dapat membantu pemasar untuk menemukan kepentingan pelanggan mereka berdasarkan pada pola pembelian dan ciri kelompok customer, oleh karena itu dalam kasus segmentasi customer, maka diperlukan algoritma clustering yang cocok berdasarkan karakteristik dari datanya. Untuk mengetahui algoritma clustering apa yang cocok digunakan pada segmentasi pelanggan, maka perlu mengetahui karakteristik dari beberapa algoritma clustering. Pada teknik partisi clustering terdapat algoritma K-means,

algoritma K-means merupakan suatu algoritma clustering yang mempartisi dataset kedalam beberapa k cluster, algoritma K-means cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan.

- **Tujuan**

Data tersebut akan digunakan oleh tim marketing sebagai penunjang membuat keputusan untuk mengoptimalkan strategi pemasaran

- **Spesifikasi Aplikasi**

Tools yang dipakai kelompok kami untuk menyelesaikan proyek ini yaitu mini anaconda (orange),

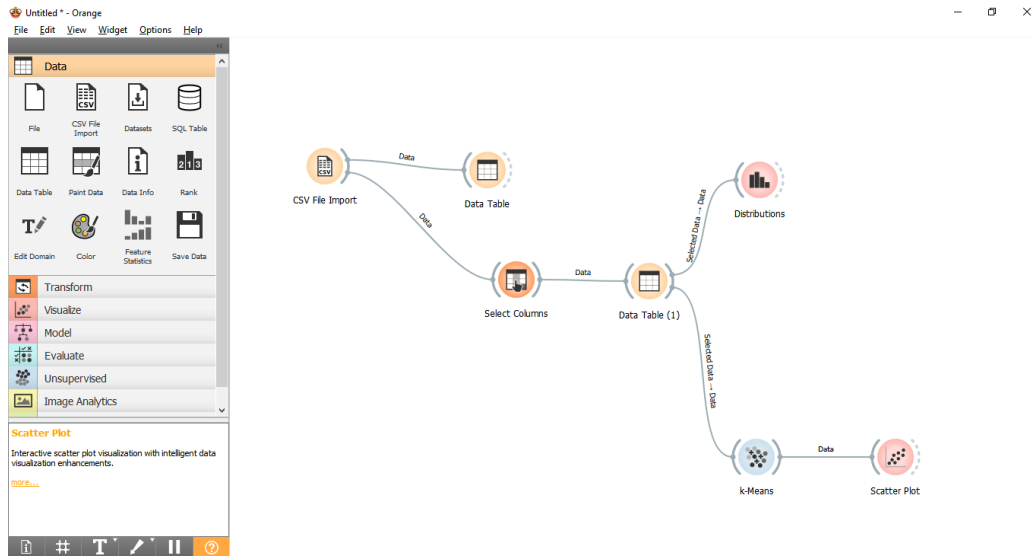
Di dalam orange ada :

- K-means
- Distribution

- **Timeline Proyek**

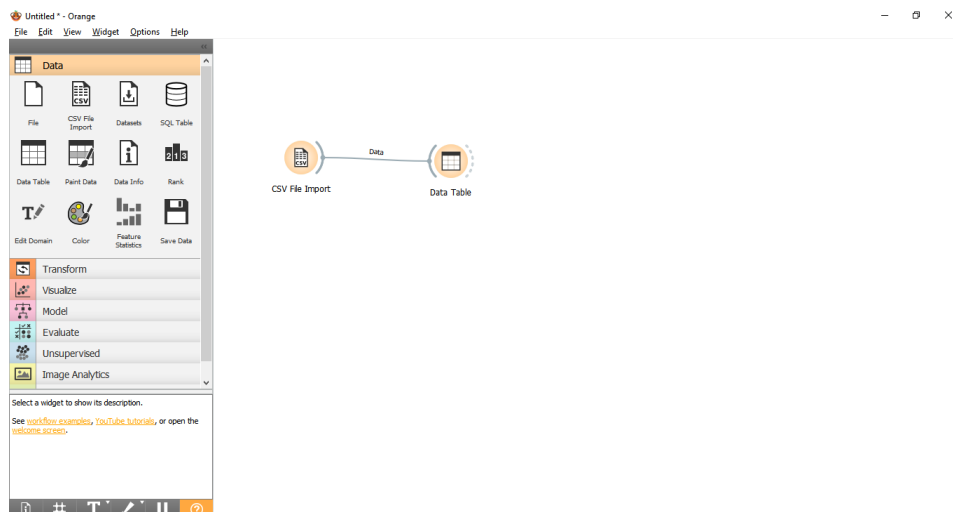
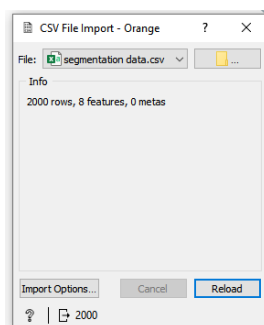
27 Desember 2022 : Menentukan Tema
28 Desember 2022 : Pembagian Tugas
29 Desember 2022 : Eksekusi dan selesai

- **Desain Proyek**
Desain yang kami buat yaitu clustering data

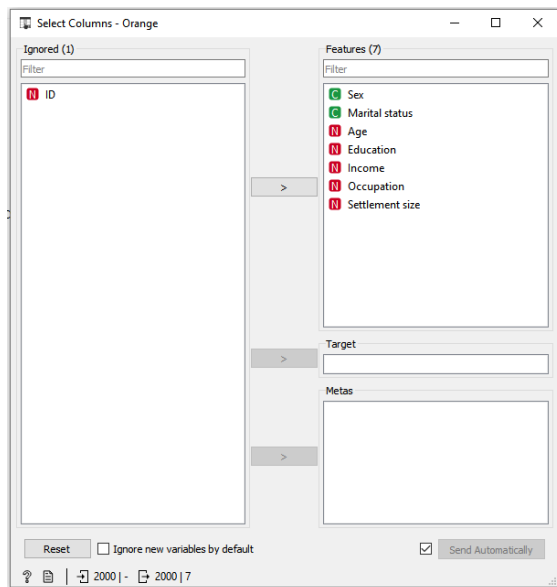
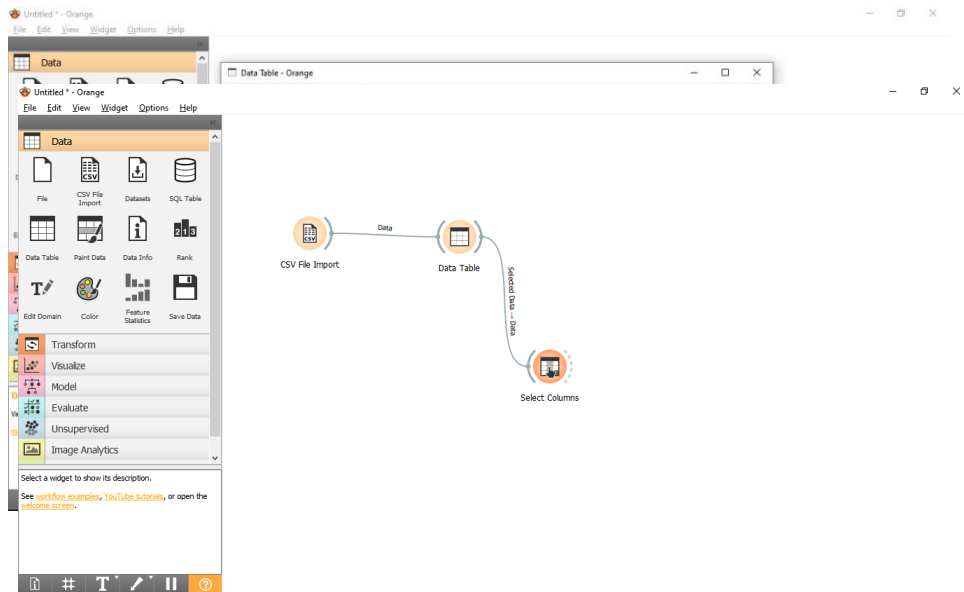


- **Hasil Pengujian**

- Memasukkan dataset kedalam tools orange

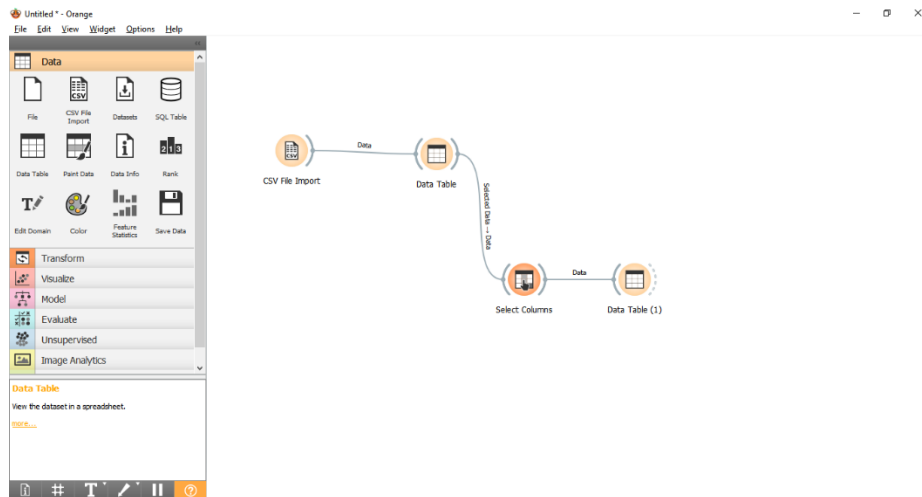


- Melihat data table menggunakan tools data table



- Selanjutnya kita menggunakan tools select columns untuk menghilangkan id dari data table

- Setelah itu kita dapat melihat columns “id” sudah tidak ada di data table1



Data Table (1) - Orange

Info
2000 instances (no missing data)
7 features
No target variable.
No meta attributes.

Variables
☒ Show variable labels (if present)
☐ Visualize numeric values
☒ Color by instance classes

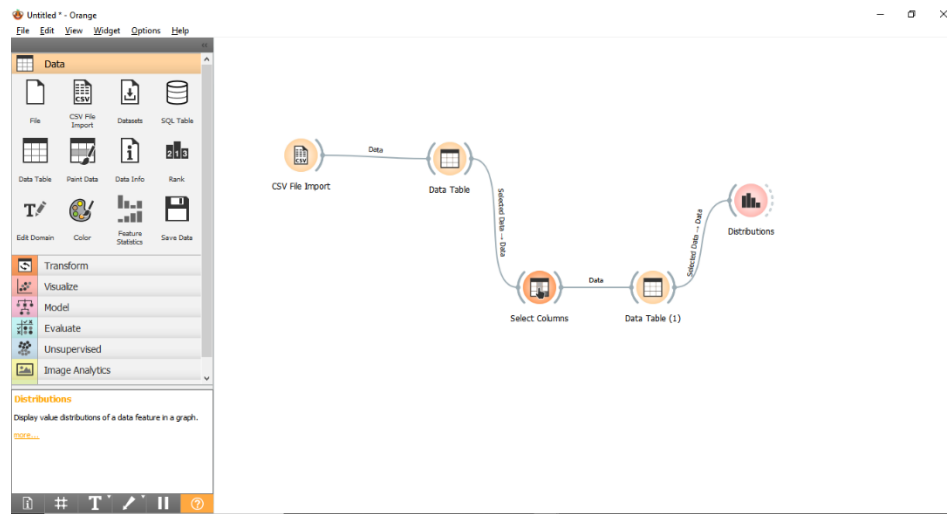
Selection
☒ Select full rows

Restore Original Order

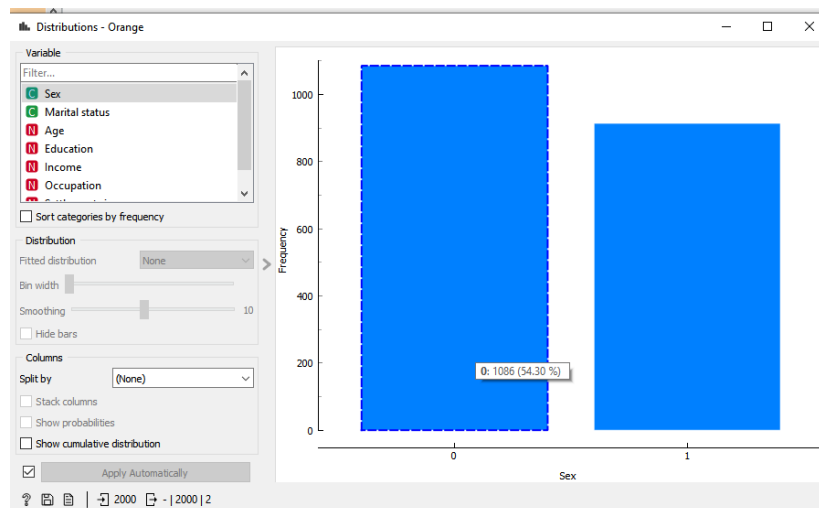
☒ Send Automatically

| | Sex | Marital status | Age | Education | Income |
|----|-----|----------------|-----|-----------|--------|
| 1 | 0 | 0 | 67 | 2 | 124670 |
| 2 | 1 | 1 | 22 | 1 | 150773 |
| 3 | 0 | 0 | 49 | 1 | 89210 |
| 4 | 0 | 0 | 45 | 1 | 171565 |
| 5 | 0 | 0 | 53 | 1 | 149031 |
| 6 | 0 | 0 | 35 | 1 | 144848 |
| 7 | 0 | 0 | 53 | 1 | 156495 |
| 8 | 0 | 0 | 35 | 1 | 193621 |
| 9 | 0 | 1 | 61 | 2 | 151591 |
| 10 | 0 | 1 | 28 | 1 | 174646 |
| 11 | 1 | 1 | 25 | 1 | 108469 |
| 12 | 1 | 1 | 24 | 1 | 127596 |
| 13 | 1 | 1 | 22 | 1 | 108687 |
| 14 | 0 | 0 | 60 | 2 | 89374 |
| 15 | 1 | 1 | 28 | 1 | 102899 |
| 16 | 1 | 1 | 32 | 1 | 88428 |
| 17 | 0 | 0 | 53 | 1 | 125550 |
| 18 | 0 | 0 | 25 | 0 | 157434 |
| 19 | 1 | 1 | 44 | 2 | 261952 |
| 20 | 0 | 0 | 31 | 0 | 144657 |

- Selanjutnya kita dapat distribusikan data table1

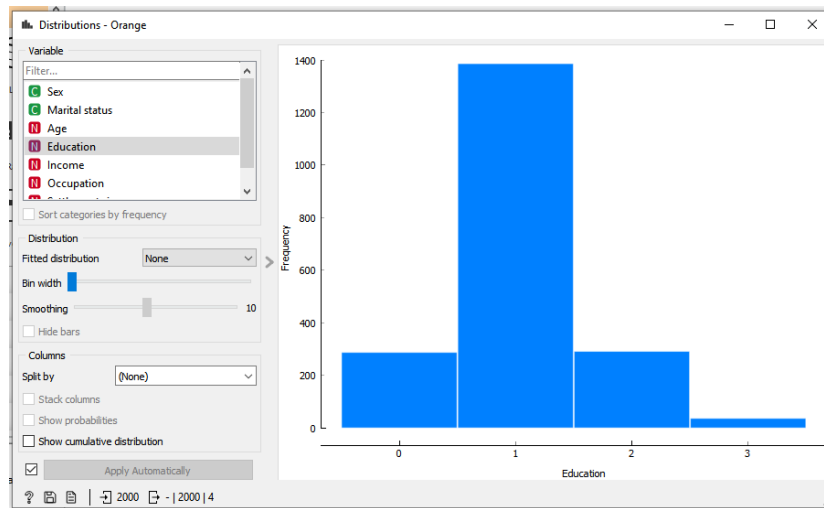


- Di tools distribusi ini kita dapat melihat perbandingan semua columns contohnya

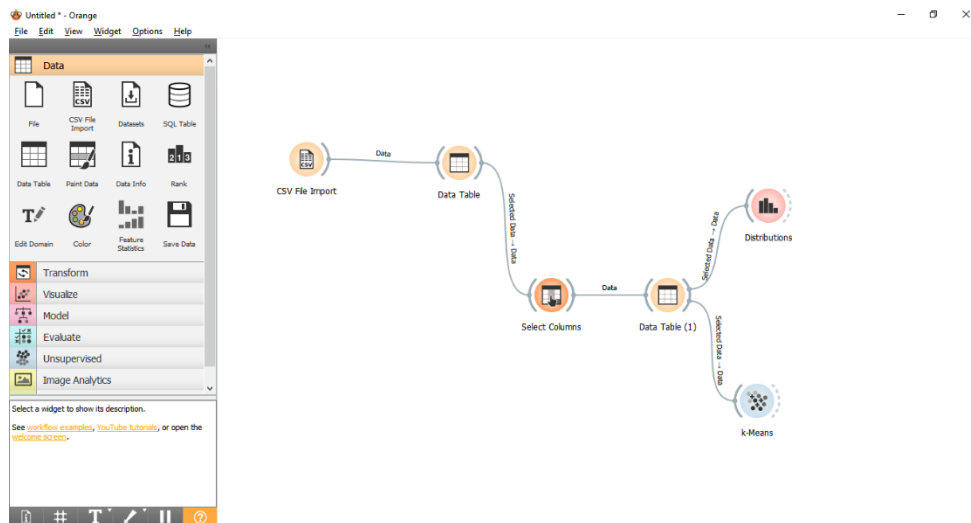


- Total 1086 atau 54.30% customer dengan jenis kelamin laki-laki
- Total 914 atau 45.70% customer dengan jenis kelamin perempuan

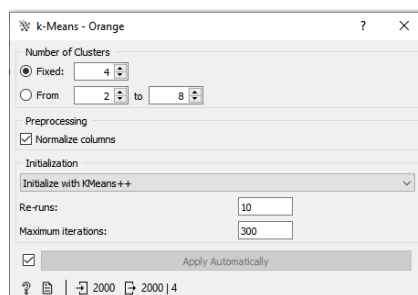
selanjutnya untuk education(pendidikan) di data tersebut didominasi oleh poin 1 yang artinya Customer dengan tingkat pendidikan high school yang mendominasi



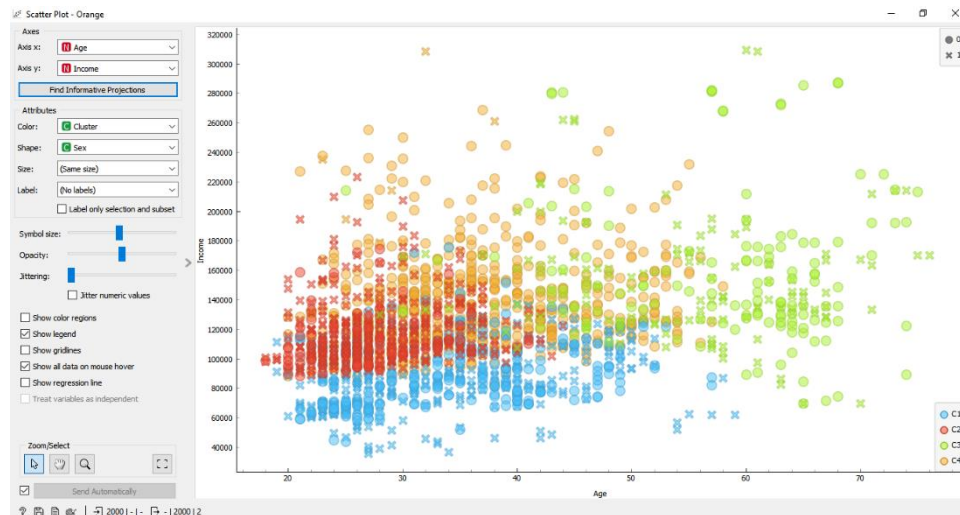
- Selanjutnya kita menambahkan tools k-mens untuk mengatur clustering dari data tersebut



- Di tools k-mens ini kita membuat 4 cluster dari data tersebut

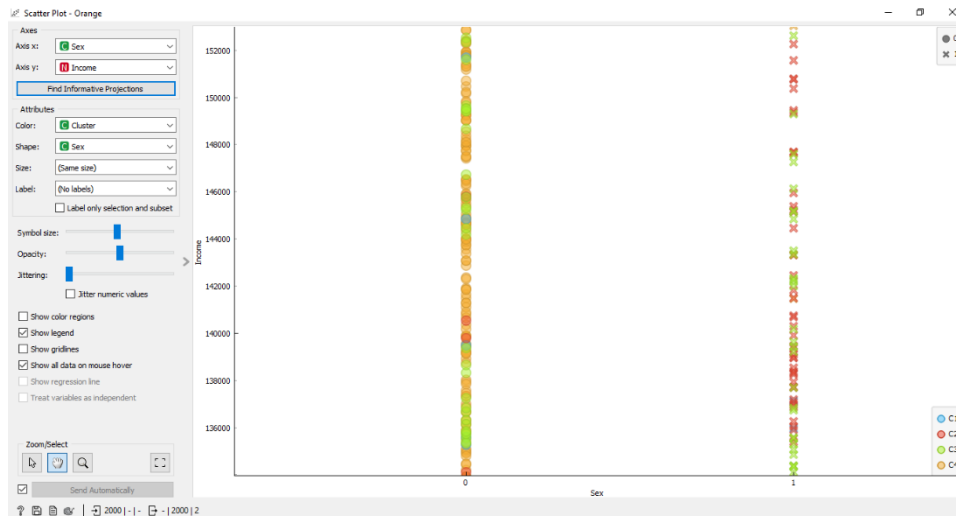


- Selanjutnya kita menambahkan tools scatter plot untuk melihat clustering yang ada di data tersebut

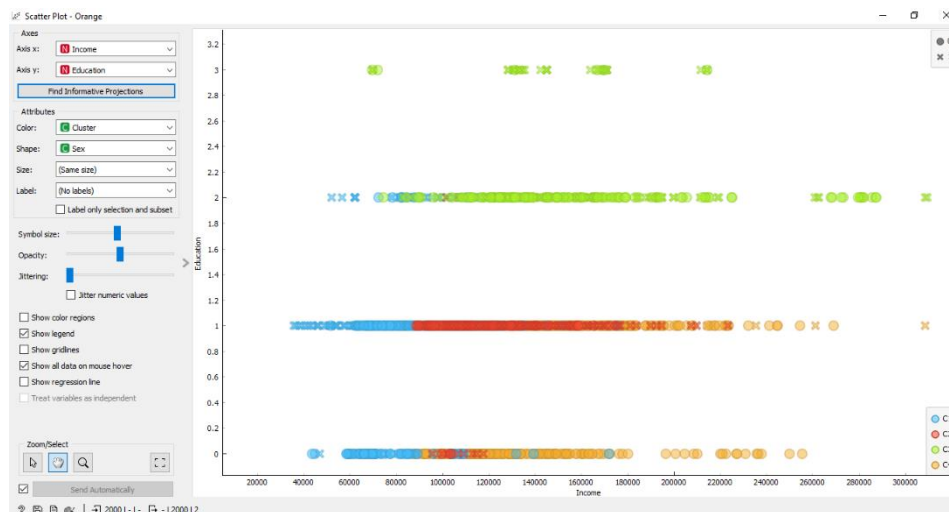


- Di data tersebut kita membuat menjadi 4 cluster diantara nya adalah
- Cluster **C1**
C2
C3
C4
- Cluster C1 yaitu, customer dengan rata-rata memiliki pendapatan yang besar dan pengeluarannya pun juga besar.
- Cluster C2 yaitu, customer dengan rata-rata memiliki keseimbangan antara pengeluaran serta penghasilan.
- Clustering berwarna hijau (C3) yaitu, customer dengan rata-rata memiliki pengeluaran yang besar akan tetapi memiliki pendapatan yang kecil.
- Clustering berwarna orange (C4) yaitu, customer dengan rata-rata memiliki pengeluaran yang kecil dan penghasilan nya yang besar

- Perbandingan pendapatan customer sesuai jenis kelamin



- Ini adalah status Pendidikan dan pendapatan yang di peroleh customer dari data tersebut



• Evaluasi Sistem Proyek

Saat pengerjaan kelompok kami mengalami Kendal dalam membaca informasi yang dihasilkan dari pembagian clustering tersebut

• Kendala Yang Dihadapi

Kurang nya refrensi yang akurat
Kesulitan mencari dataset yang relevan

• Kesimpulan

Dari dataset di atas dapat disimpulkan sebagai berikut :

Jenis Kelamin
Laki-laki : 1086
Perempuan : 914

Status Pernikahan
Single : 1007
non-single : 993

Umur

Didominasi Oleh Rentan umur 26-28th
Dan yang paling sedikit rentan umur 72-74th

Pendidikan

Didominasi oleh customer yang berpendidikan di tingkat Sekolah menengah atas
Dengan total customer : **1386**

Pendapatan

Pendapatan costumer didominasi dengan jarak antara \$100.00 - \$110.00

Pekerjaan

Didominasi oleh customer yang bekerja sebagai pegawai kantor
dengan total : **1113 Customer**

Ukuran Pemukiman

Didominasi oleh customer yang bertempat tinggal di kota kecil
Dengan total : **989 Customer**

- **Referensi**

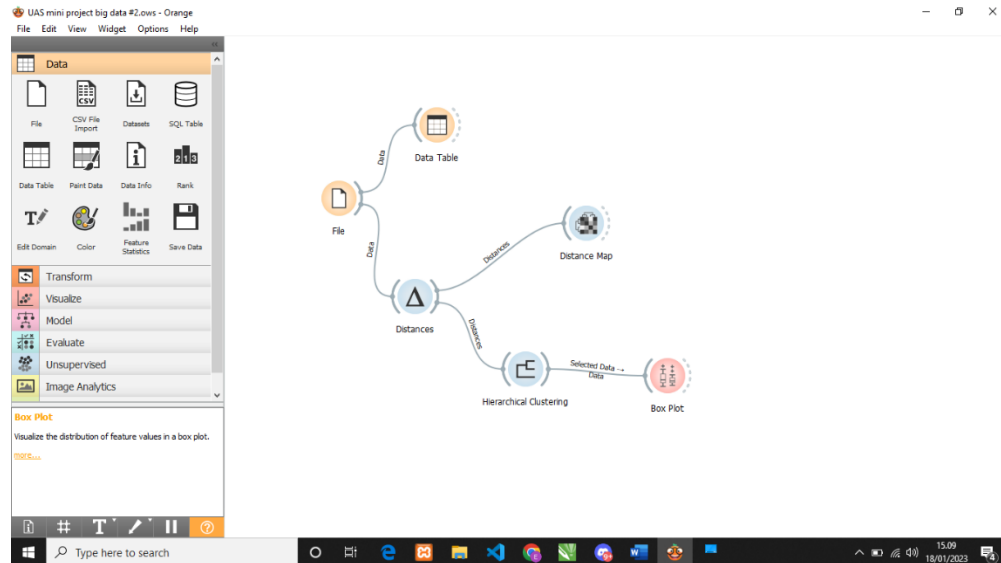
https://elib.unikom.ac.id/files/disk1/747/jbptunikompp-gdl-gumilarakb-37349-1-unikom_g-1.pdf

<https://www.kaggle.com/datasets/dev0914sharma/customer-clustering>

Hierarchical Segmentation Data customer

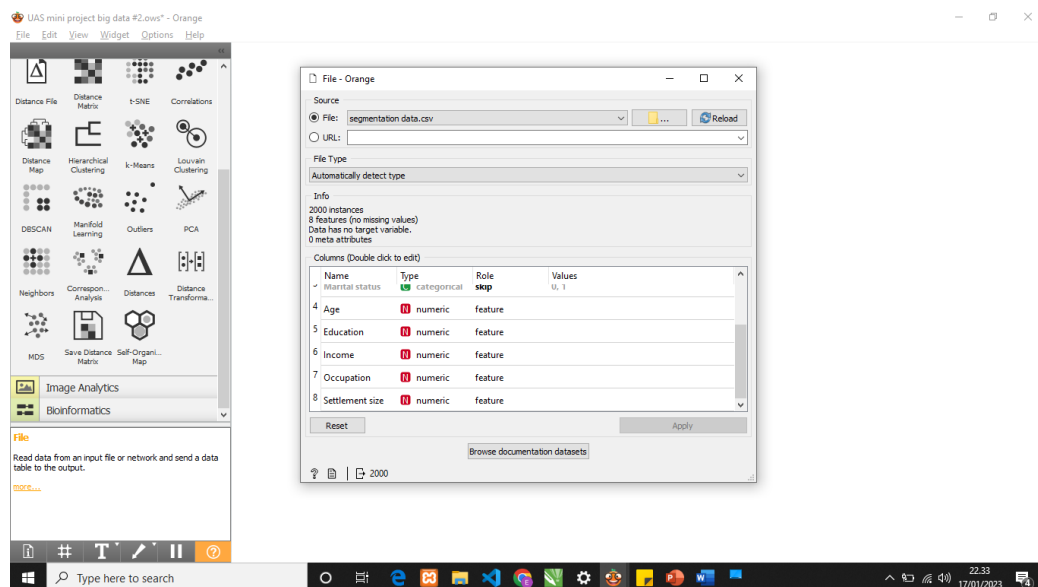
- Disini kita mencoba juga menggunakan metode Hierarchical dan Distance Map

Ini desain yang kami buat untuk metode hierarchical clustering dan box plot

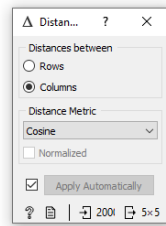


• Hasil Pengujian

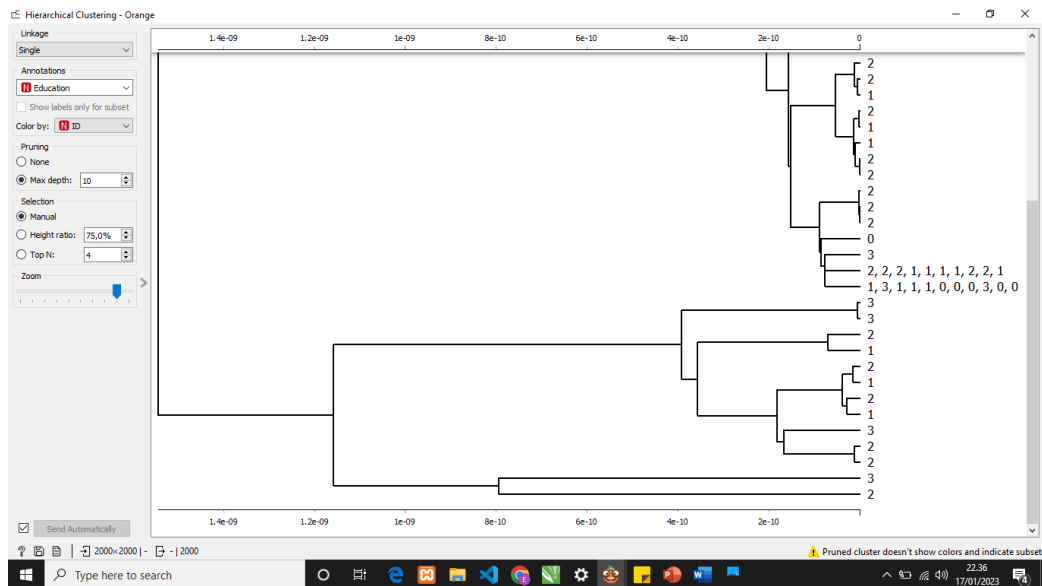
- Langkah awal masukkan dataset yang sudah tersedia



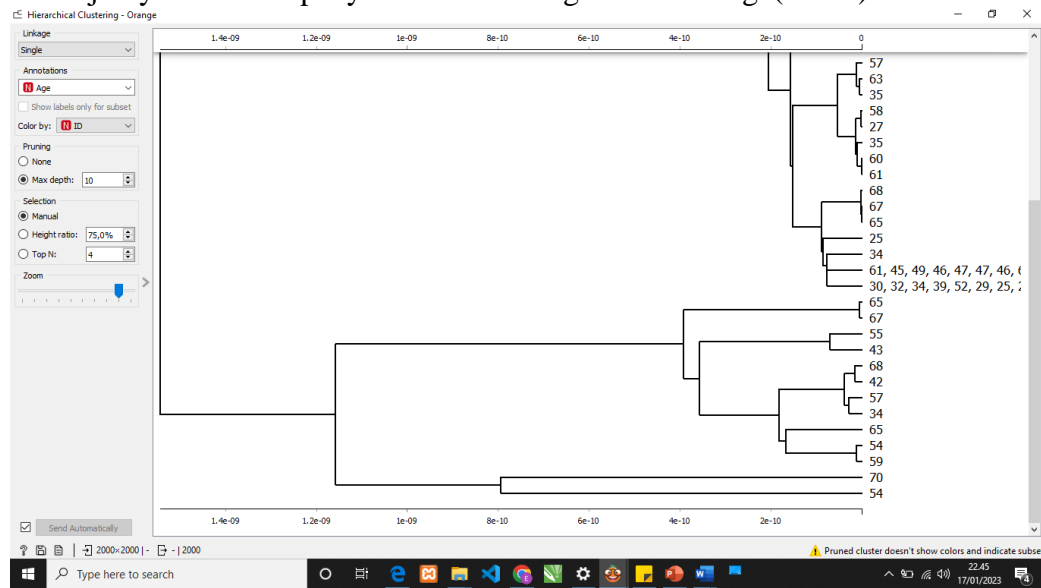
- Selanjutnya kita menggunakan tools Distance untuk mengatur hierarchical yang akan kita buat



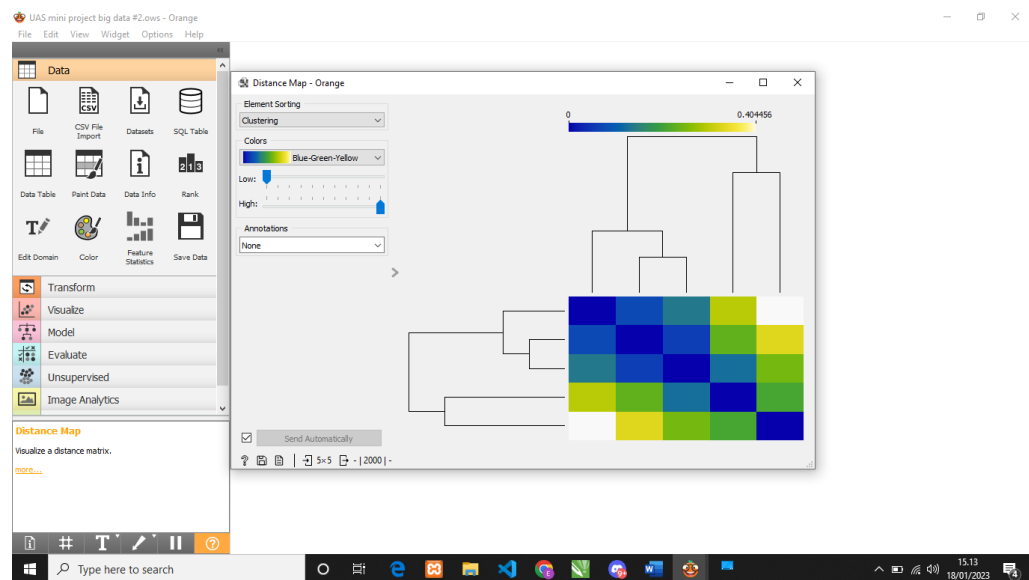
- Selanjutnya kita tambahkan tools hierarchical untuk melihat penyebaran data tersebut Dan ini contoh penyebaran data untuk variable Education



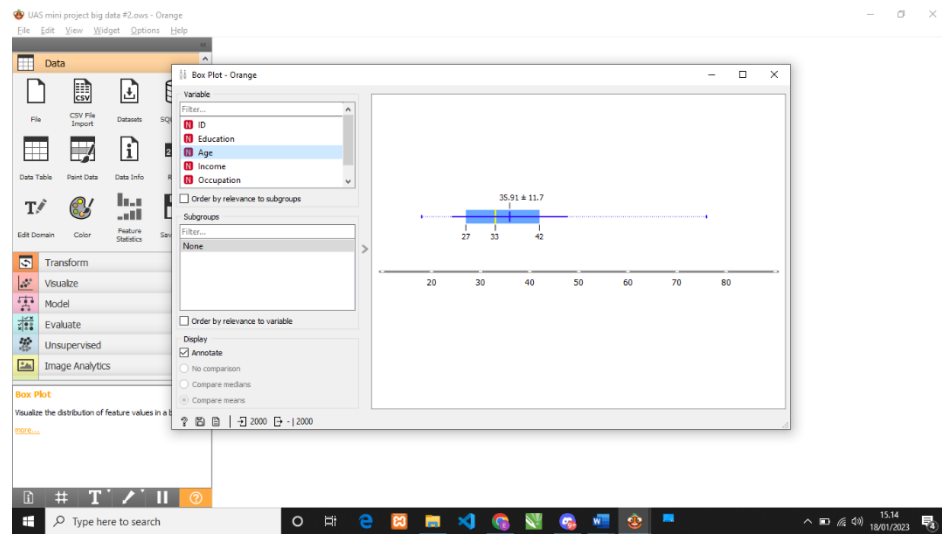
- selanjutnya ini untuk penyebaran data dengan variable age(Umur)



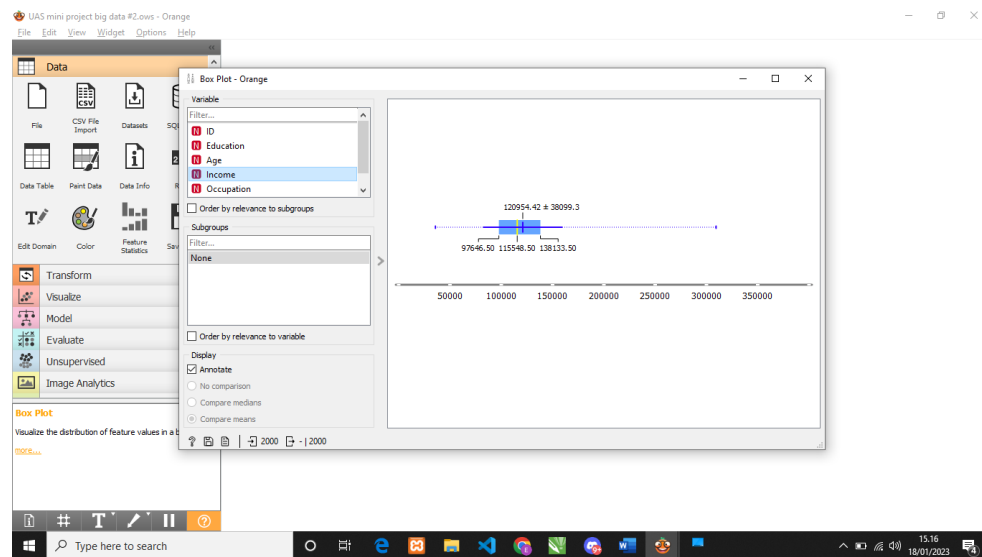
- Selanjutnya kita menggunakan tools distance maps untuk melihat penyebaran clustering dari data tersebut



- Selanjutnya kita tambah kan tools box plot untuk menyimpulkan data tersebut



- Dari diagram diatas terdapat informasi untuk dominan umur di angka 27-42Tahun



- **Kesimpulan**

Kesimpulannya adalah Clustering data menggunakan k-Mens lebih mudah di gunakan menggunakan tools distributions, kekurangan dari metode k-mens adalah sulitnya membaca informasi dari pembagian clustering

Link Google Drive Kelompok 15 :

https://drive.google.com/drive/folders/1UL4UQwdSWtMHEA5F6uztYxDf1sCe_tbd?usp=sharing