

Country Development Classification Using Clustering Algorithm for HELP International



Benny Septiawan Salim
Data Science Final Project

The Case

- Preparing HELP International to prioritize finance gift based on country's social economy crisis and health factor
- Feature Column Contains :
 - Country
 - Number of Children's Death
 - Export (good and services spending per capita)
 - (health spending per capita)
 - Import (importing good and services per capita)
 - Income
 - Inflation (GDP Growth Number)
 - Life Expectancy for newborn (based on children's death number)
 - Number of Fertilization (Total of Newborn)
 - GDP per capita



Tahap 1 : Exploratory Data Analysis (EDA)

Proses EDA Dimulai dengan membagi kesatuan dataset menjadi 2 jenis kategori data yang sudah difilter sesuai kondisi kasus, yaitu

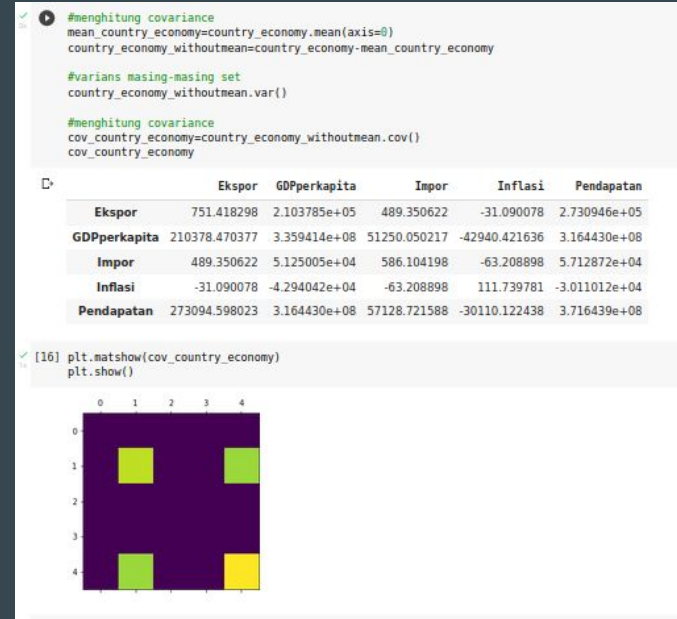
- data berdasarkan tingkat kesehatan melalui angka kematian dan harapan hidup (country_life)

- serta data berdasarkan tingkat ekonomi melalui data ekspor impor dan hubungannya (country_economy)

Berikutnya juga dilakukan penghitungan korelasi dengan menggunakan covarians kedua data secara terpisah

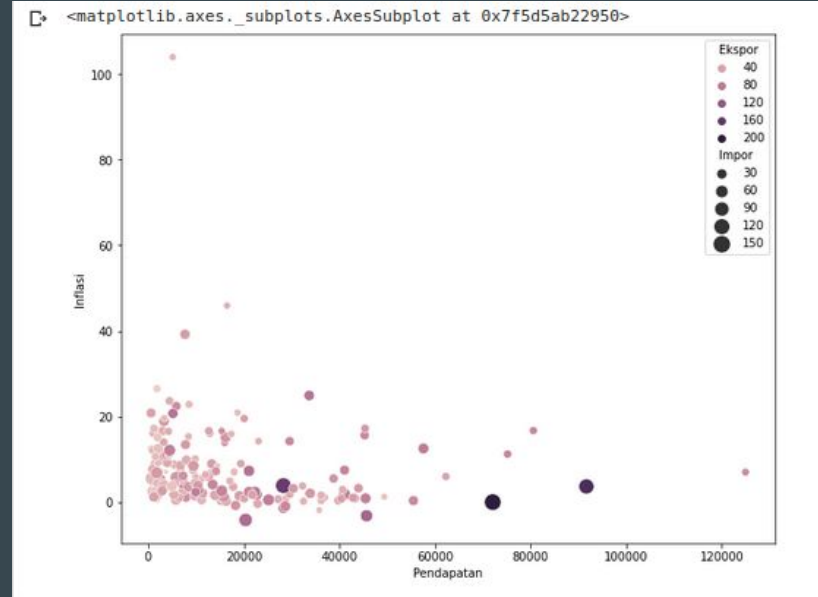
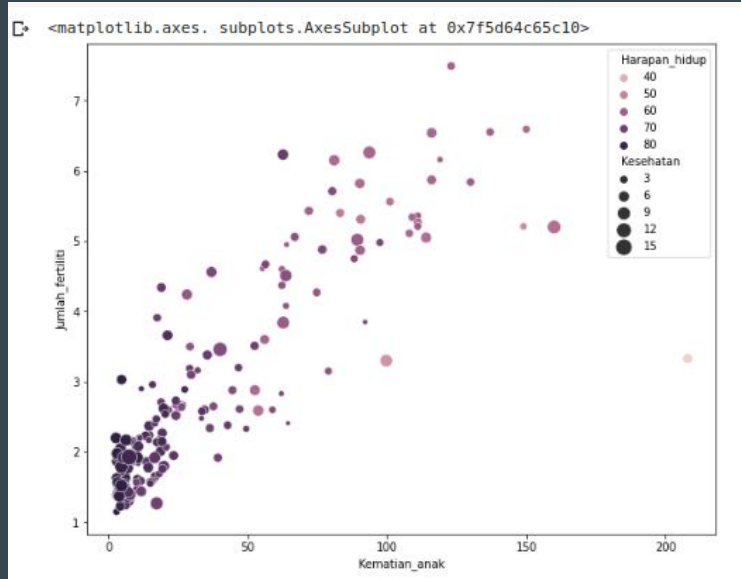
Korelasi ini dihitung dengan melihat koefisien antara -1 dan 1, sehingga menghasilkan data gambar berikut ini

Tahap 1 : Exploratory Data Analysis (EDA)



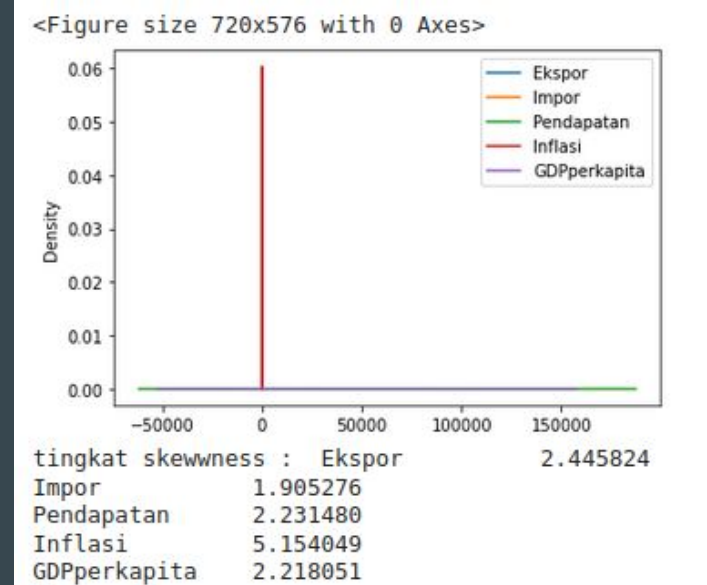
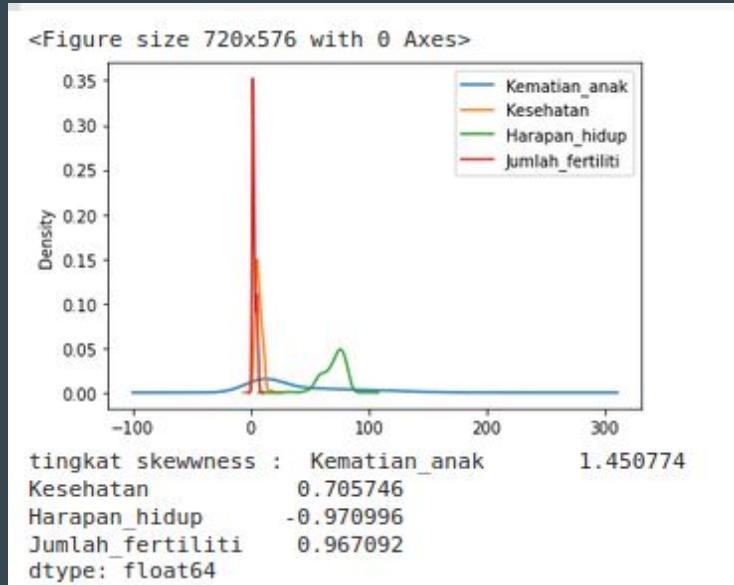
Gambar ini menunjukkan ukuran besarnya persebaran data, dimana terlihat bahwa data kedua memiliki angka covarians yang lebih besar daripada data country_life, serta hasil visualisasi data kedua lebih terlihat terang daripada data pertama

Tahap 1 : Exploratory Data Analysis (EDA)



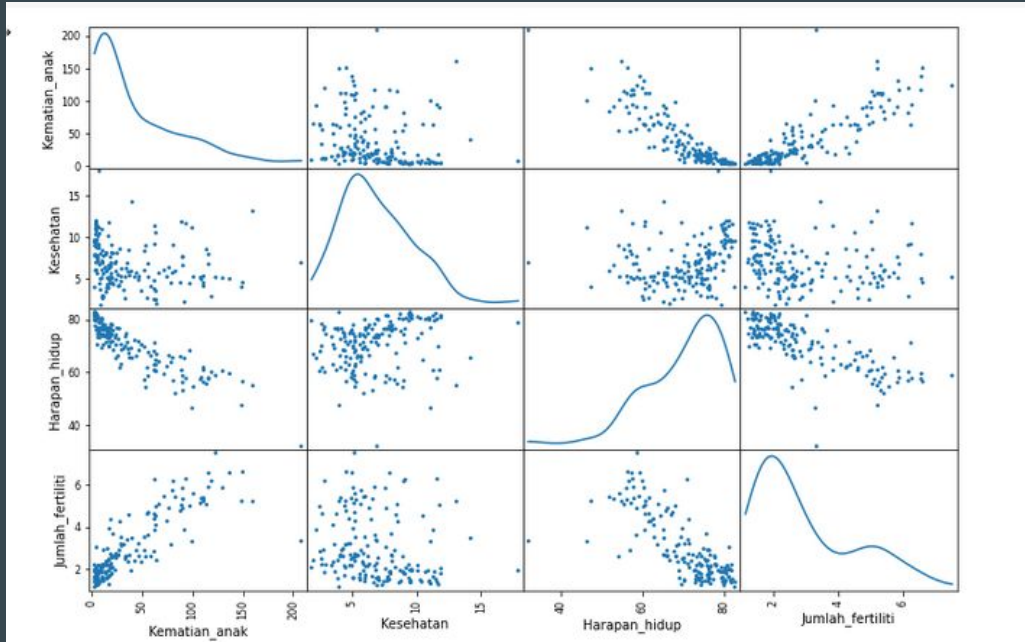
Covarians tersebut membantu data untuk menghasilkan angka Correlation tertinggi, dataset pertama memiliki hubungan correlation x y dengan angka yang baik, yaitu 0.8, sedangkan dataset kedua memiliki angka korelasi x y berkisar -0.1, angka paling maksimal yang bisa didapatkan.

Tahap 1 : Exploratory Data Analysis (EDA)



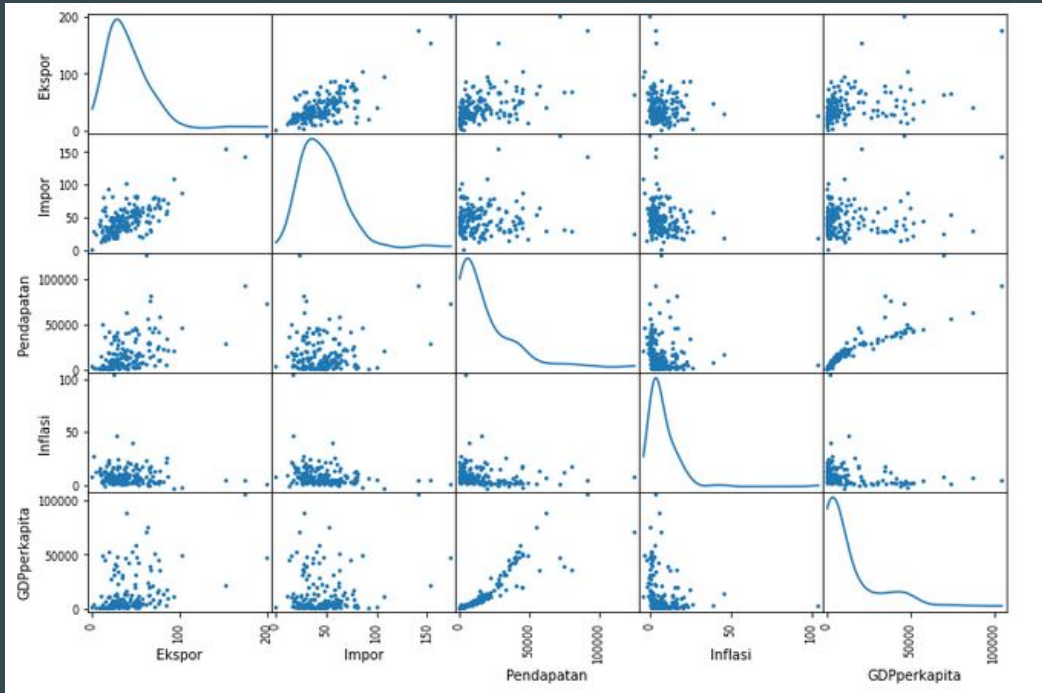
Berikut menunjukkan grafik skewness untuk mengecek apakah data tersebut sudah dalam bentuk normal distribution atau belum. Hasil skewness menunjukkan skewness yang terdistribusi normal, meskipun distribusi tersebut membentuk sudut lancip.

Tahap 2 : Multivariate Analysis



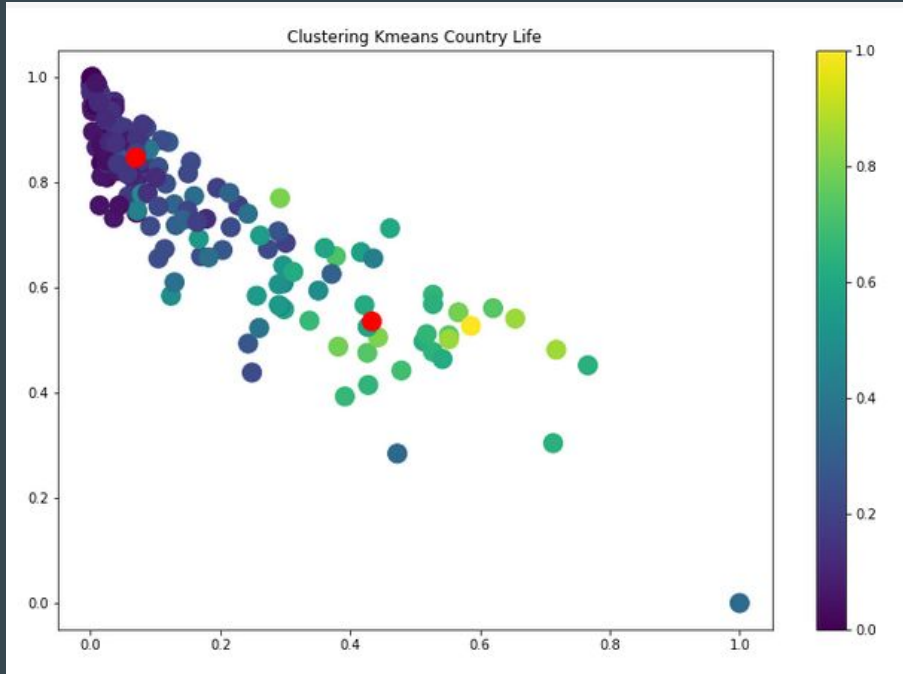
Grafik Multivariate ini menunjukkan skewness dari data *country_life*, bisa terlihat bahwa data point yang divisualisasikan dalam bentuk scatter masih tersebar meski minim outlier, namun positifnya adalah tidak adanya overlapping.

Tahap 2 : Multivariate Analysis



Grafik Multivariate ini menunjukkan skewness dari data `country_economy`, bisa terlihat bahwa data point yang divisualisasikan dalam bentuk scatter membentuk pola yang jelas meski adanya outlier, overlapping menurut saya terlihat sedikit, namun visualisasi skewness menunjukkan hasil yang konsisten dari kelima column data.

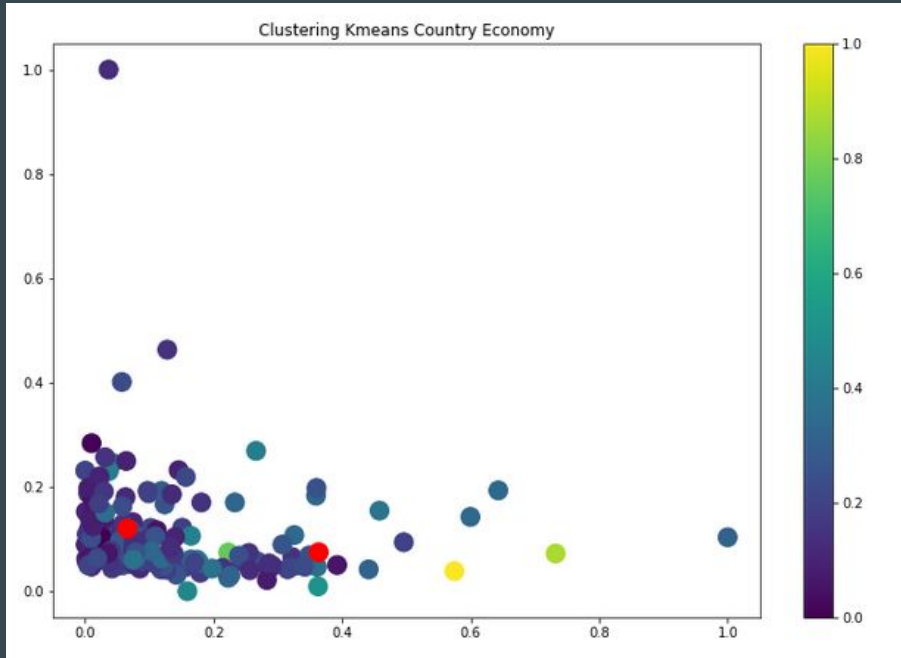
Tahap 3 : Clustering



Berikut merupakan hasil clustering dari country life dengan x y mengikuti dengan scatterplot pada analisis EDA sebelumnya.

Clustering membagi data menjadi 2 bagian terpisah, sehingga menunjukkan bahwa mayoritas jumlah kematian anak lebih dominan pada negara dengan fertilisasi rendah.

Tahap 3 : Clustering



Berikut merupakan hasil clustering dari country_economy dengan x y mengikuti dengan scatterplot pada analisis EDA sebelumnya.

Dapat disimpulkan bahwa pendapatan terendah bisa terjadi karena adanya angka inflasi yang tinggi dari tiap negara

Thank You For Your Support
& Attention to This Final
Project :)

“Benny Septiawan Salim”

