**LAPORAN TUGAS PEMBELAJARAN MESIN**

**MENJELASKAN, IMPLEMENTASI, ANALISIS, SERTA DESAIN TEKNIK PEMBELAJARAN MESIN SUPERVISED LEARNING**

laporan ini dibuat untuk memenuhi kebutuhan nilai mata kuliah Pembelajaran Mesin

Dosen mata kuliah : **Tjokorda Agung Budi Wirayuda, S.M., M.T.**

**CII3C3-IF-44-11**



disusun oleh

**NAUFAL ADRIAN HIDAYAT – 1301204056**

**IF-44-11**

**FAKULTAS INFORMATIKA**

**TELKOM UNIVERSITY**

**2022**

**BAB 1**

**PENDAHULUAN**

**Case-Based 2**

**Scenario**

Mengikuti keberhasilan tugas sebelumnya, Anda diberi kesempatan lebih lanjut untuk mengesankan atasan Anda mengenai kemampuan Anda untuk menganalisis data. Anda diminta untuk melakukan beberapa analisis dan menghasilkan seperangkat aturan yang berguna menggunakan dataset berikut: Kumpulan data berikut tersedia online, tautan ke kumpulan data adalah sebagai berikut:

*https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data*

Dataset masih terdapat missing value atau outlier. Harap lakukan perbaikan terhadap hal ini, selanjutnya anda harus menganalisa data tersebut. Jika perlu konversi variable kategori menjadi integer. Jika perlu lakukan normalisasi data melalui fitur rescaling. Jika perlu lakukan analisa elbow. Jika perlu lakukan analisa dengan plot data secara visual. Jika perlu lakukan transformasi data secara logaritmik. Dan masih banyak kemungkinan Analisa data yang dapat anda lakukan.

Hint: Anda bebas memilih satu dari tiga alat analisis data yaitu Weka, R, atau Python untuk membantu Anda menganalisis data dan menunjukkan pra-pemrosesan data yang diperlukan.

**Tugas Anda**

Tujuan dari tugas ini yaitu Anda diharapkan mampu menjelaskan, mengimplementasikan, menganalisis, dan mendesain teknik pembelajaran mesin unsupervised learning yaitu KMEANS/DBSCAN/Hierarchical.

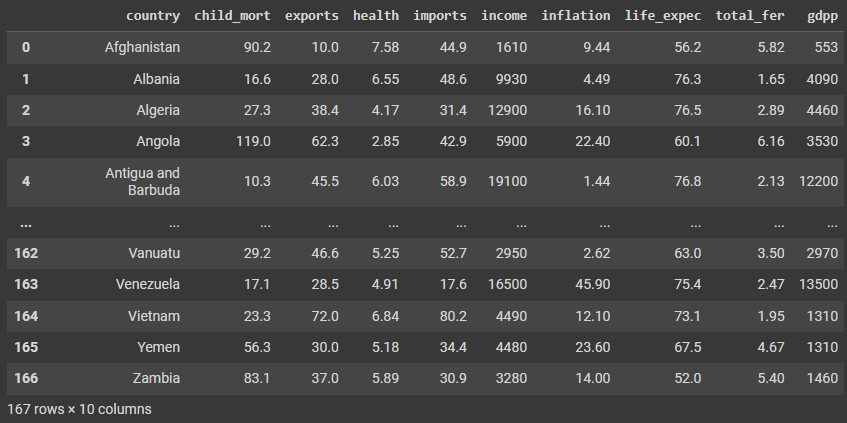
Pertama, selidiki masalah kualitas data yang telah diberikan di atas. Jelaskan keputusan Anda mengenai pendekatan pra-pemrosesan data. Jelajahi kumpulan data dengan meringkas data menggunakan statistik dan mengidentifikasi masalah kualitas data apa pun. Tidak ada batasan jumlah ringkasan yang akan dilaporkan tetapi Anda diharapkan hanya melaporkan yang paling relevan.

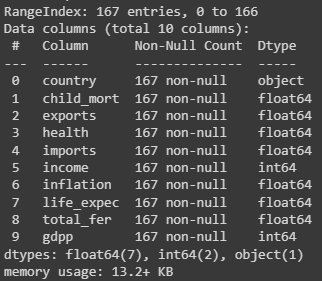
Kedua, pilih salah satu dari metode unsupervised learning yang telah dipelajari yaitu K-means/dbscan/hierarchical. Anda hanya perlu memilih satu metode untuk diterapkan. Gunakan algoritma tersebut untuk memberikan beberapa output/outcome dengan menggunakan variasi parameter, kemudian membuat laporannya dan menganalisis hasilnya.

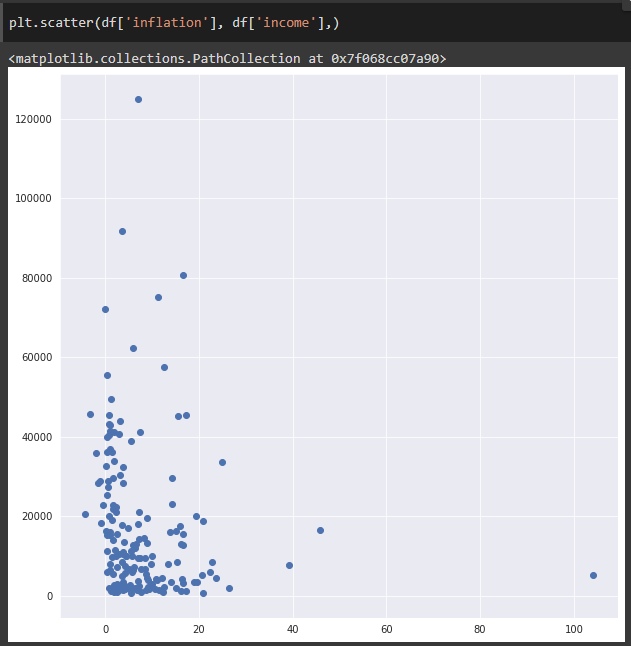
**BAB 2**

**PEMBAHASAN**

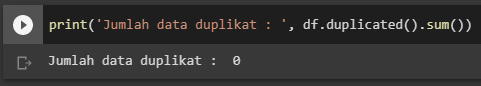
1. **Ikhtisar Data yang dipilih**   
    Case-based 2 yang saya dapat ialah data dengan NIM akhir genap, yakni “6”. Maka, saya memakai data country-data.csv dan data data-dictionary.csv sebagai acuan pada pengerjaan Case-Based berikut. data yang digunakan merupakan data yang sifatnya *unsupervised learning* punya nilai target yang ada pada kolom *class*, data tersebut merupakan data negara yang memiliki angka seperti *child\_mort* (Kematian anak-anak di bawah umur 5 tahun per 1000 live birth), export, health (total health spending per capita), imports, income, inflation, life\_expec (Rata-rata jumlah tahun seorang anak yang baru lahir akan hidup jika pola kematian saat ini tetap sama), total\_fer(angka kelahiran anak), dan gdpp.
2. **Bentuk Data Table**

Berikut adalah dataset dari *Country-Data.csv* :

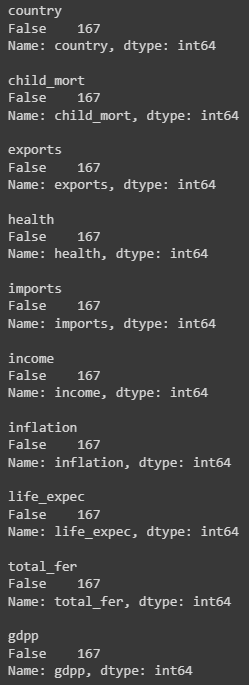
Kemudian, pada tabel *Country-Data.csv* memiliki baris sebanyak 167 record dan juga kolom sebanyak 10 atribut serta pada dataset ini kita mendapat informasi terdiri sebanyak 7 kolom dengan tipe data float, sebanyak 2 kolom data integer, dan 1 kolom data object.

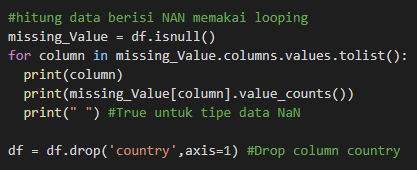
Lalu, saya mencoba menampilkan sebaran data dari column [‘Inflation’] dan [‘Income’]. Berikut adalah tampilan datanya sebelum dinormalisasikan datanya. Data tersebut masih tersebar dan tidak begitu dekat.

1. **Kualitas Data**
   1. **Mengecek Duplikasi Data**

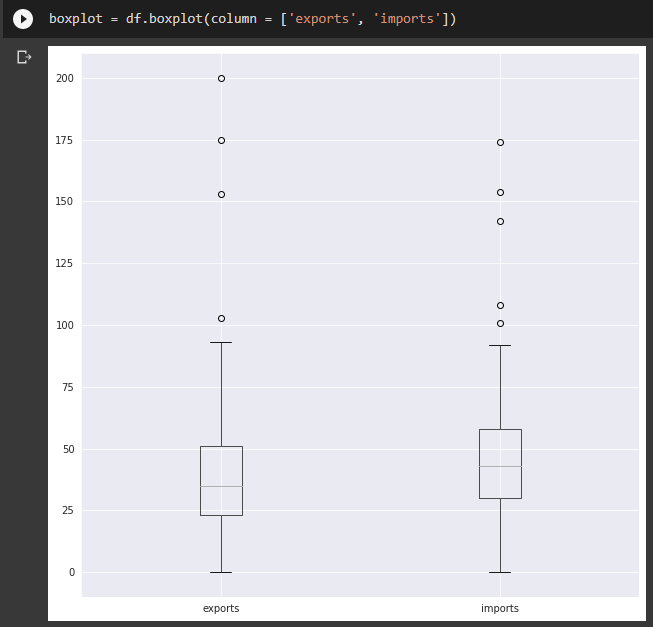
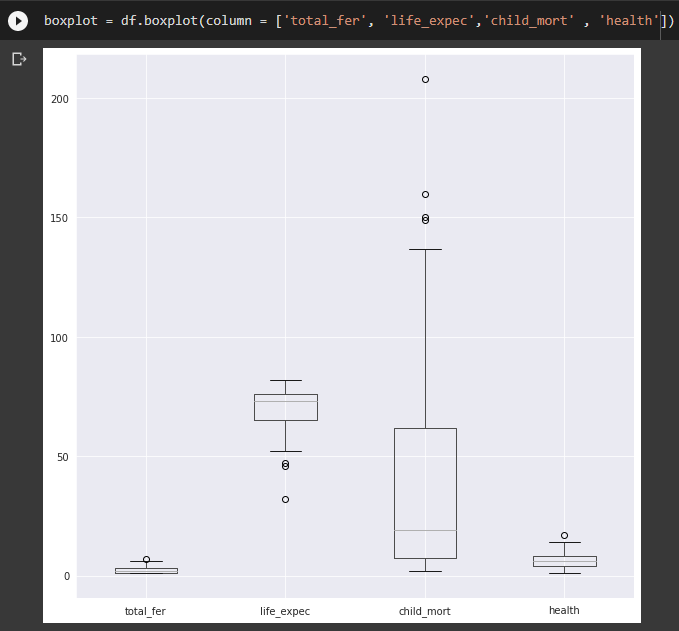
Pada dataset *Country-Data.csv*, tidak ada data yang terduplikasi. Ketika dilakukan pengecekan data yang duplikasi, keluar *outputnya* ialah “0” yang menandakan tidak adanya data yang terduplikat pada *Country-Data.csv* tersebut.

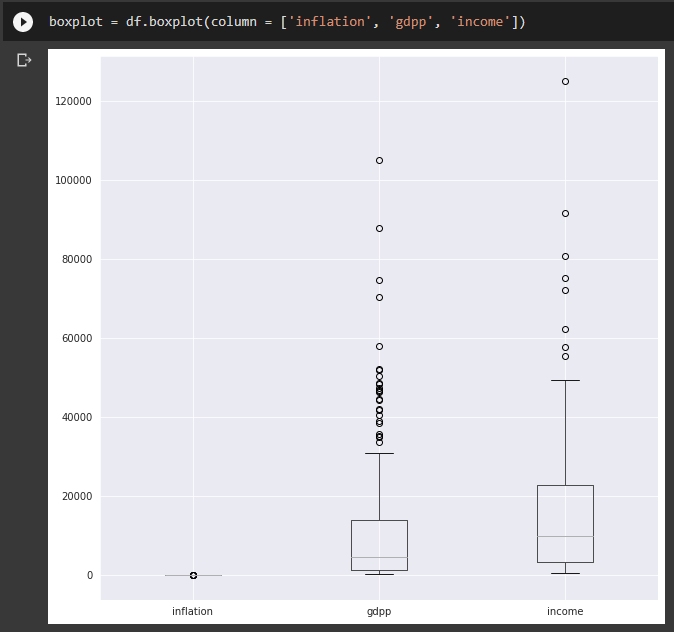
* 1. **Mengecek *Missing Value* Data**

Pada dataset *Country-Data.csv*, tidak terdapat data yang memiliki nilai kosong ataupun NaN Value. Hal tersebut bisa dibuktikan dengan mengecek datanya menggunakan code sepertinya berikut. Jika memang ada data kosong, maka outputnya akan mengeluarkan True, jika tidak ada, maka akan mengeluarkan False.



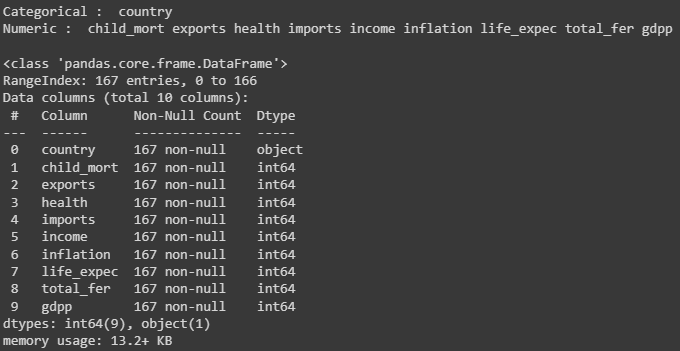
1. **Visualisasi Data**

Pada bagian ini, ditampilkannya bagaimana penyebaran data pada dataset *Country-Data.csv* dengan memakai boxplot. Jika dilihat dari boxplot, tampak banyak sekali pencilan data (Outliers) yang jarak antar datanya jauh satu sama lain pada salah satu kolom yakni [gdpp]. Karena itu, mengatasi hal tersebut saya memakai penghitungan quartil agar jarak pencilannya tidak banyak dan berdekatan.

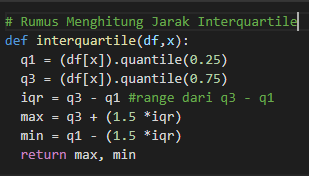


1. **Ringkasan Pra-Pemrosesan Data**
2. **Penghapusan Kolom Serta Ubah Tipe Data Kolom**

Pada sub bab ini, saya akan menghapus salah satu kolom yang berbentuk tipe data kategorikal atau objek serta mengubah seluruh tipe data yang float menjadi integer. Karena tipe data kategorikal tidak akan bisa digunakan untuk tahap penghitungan ataupun pemrosesan data dan juga mengubah float menjadi integer supaya lebih mudah dalam melakukan proses penghitungan apapun itu. Berikut adalah cara untuk menghapus serta mengubah tipe data kolom. 

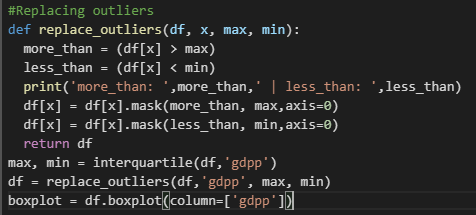


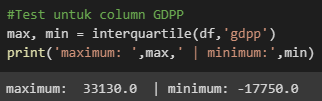
1. **Membuat Penghitungan Interquartile**

Pada Subbab ini, saya menghitung interquartile dari kolom [‘gdpp’] dan juga [‘income’] yang mana pada sub bab sebelumnya terlihat bahwa outliers pada kolom tersebut masih terlalu banyak dan jaraknya terlalu jauh. Pada penghitungan interquartile, saya menggunakan rumus sebagai berikut. 

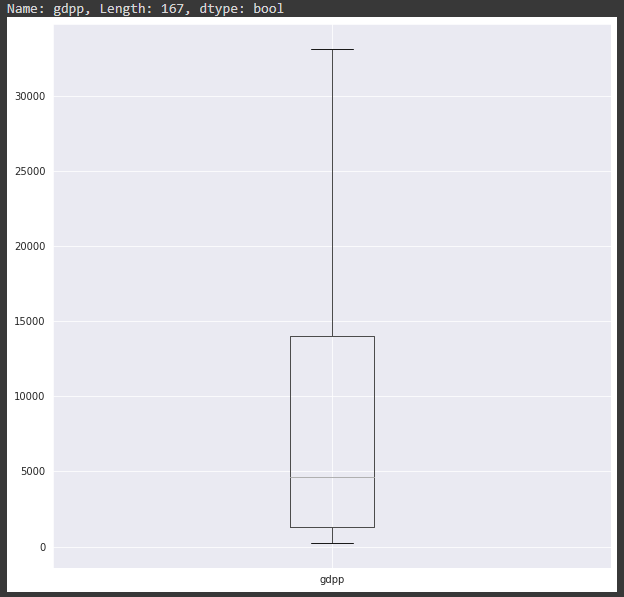
Rumus tersebut yang akan membagi data menjadi empat bagian, kemudian pencilan data (outliers) bisa ditentukan dengan nilai yang kurang dari 1.5 \* interquartile (yang mana selisih antara kuartil 1 dan kuartil 3) terhadap kuartil 1 dan nilai yang lebih dari 1.5 \* interquartile terhadap kuartil 3.

* 1. **Testing dan Replace kolom [‘gdpp’]**

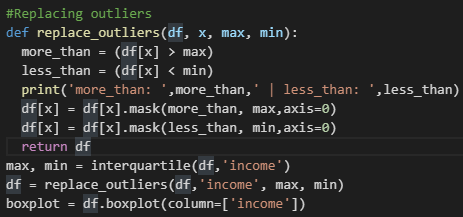
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“gdpp”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

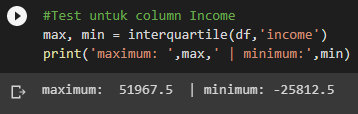


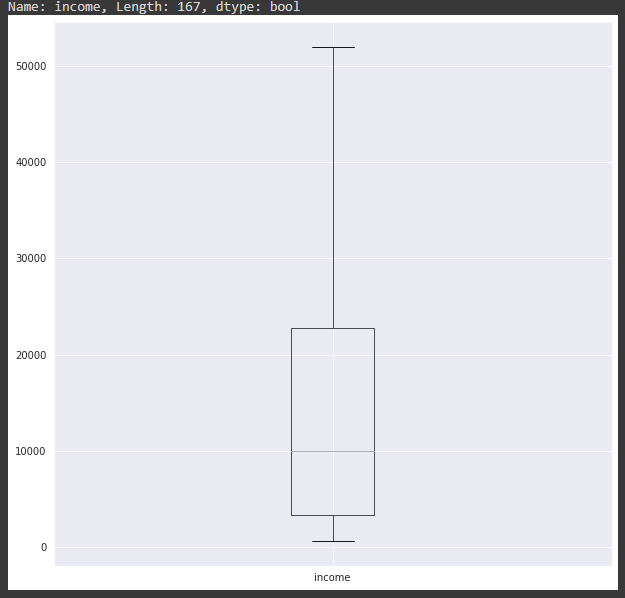
Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘gdpp’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘gdpp’]-nya.



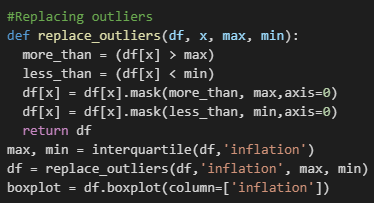
* 1. **Testing dan Replace kolom [‘income’]**

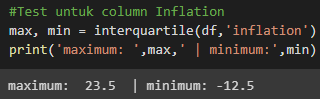
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“income”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

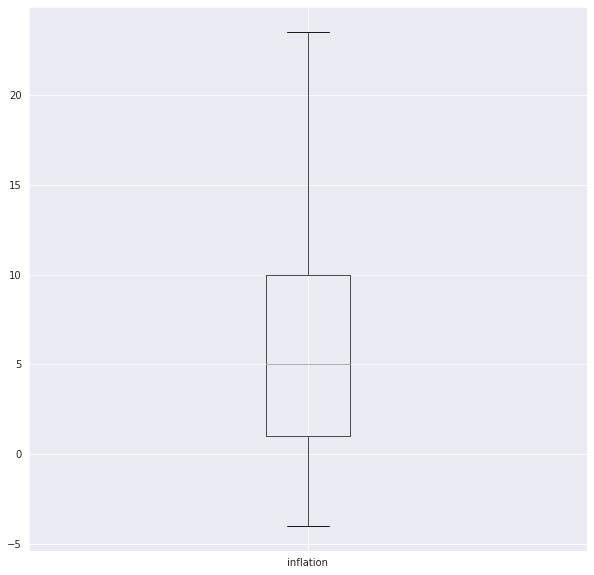


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘income’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘income’]-nya. 

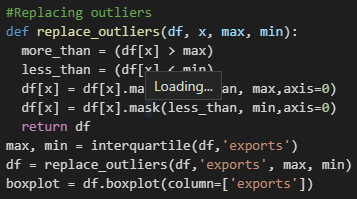
* 1. **Testing dan Replace kolom [‘inflation’]**

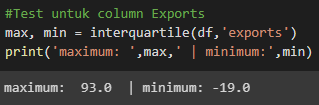
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“inflation”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

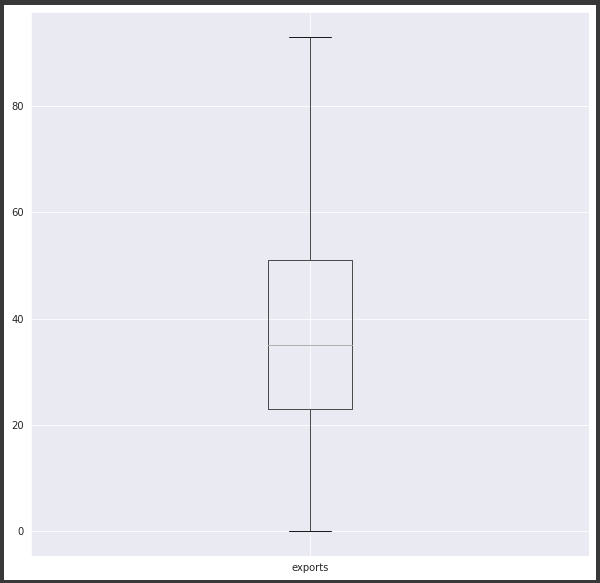


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘income’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘inflation’]-nya.

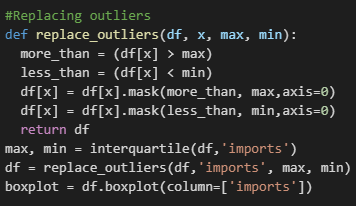
* 1. **Testing dan Replace kolom [‘exports’]**

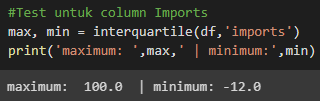
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“exports”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

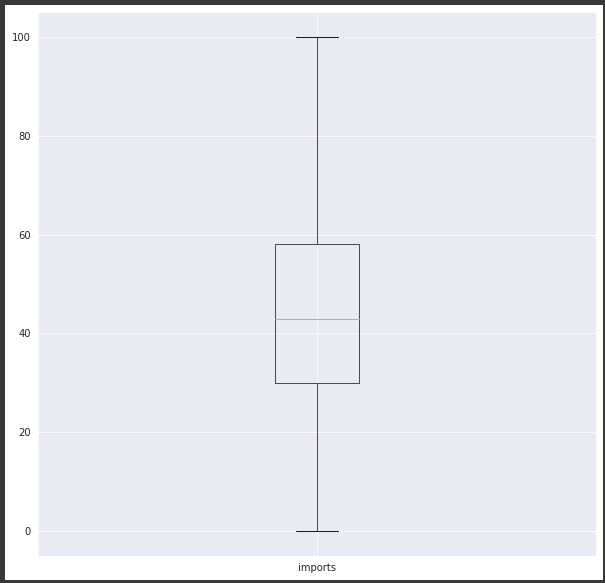


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘exports’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘exports’]-nya.

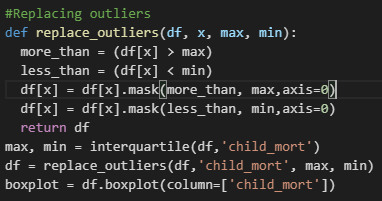
* 1. **Testing dan Replace kolom [‘imports’]**

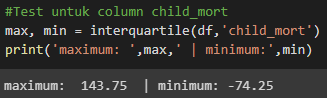
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“imports”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

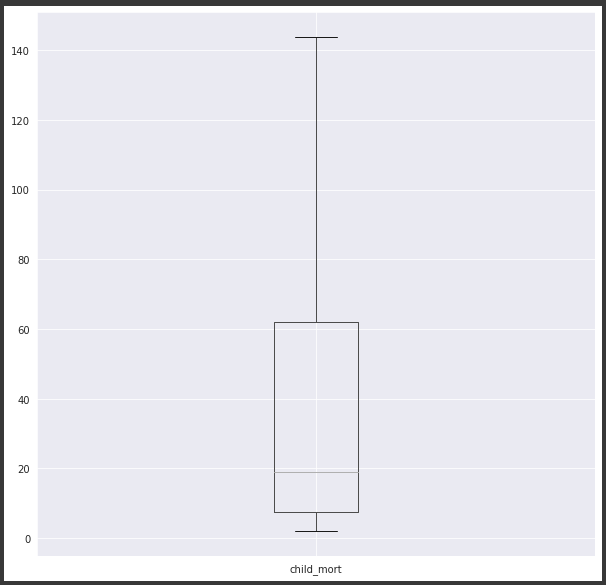


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘imports’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘imports’]-nya. 

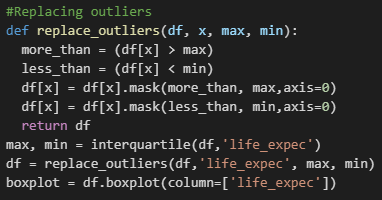
* 1. **Testing dan Replace kolom [‘child\_mort’]**

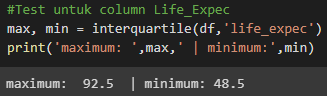
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“child\_mort”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

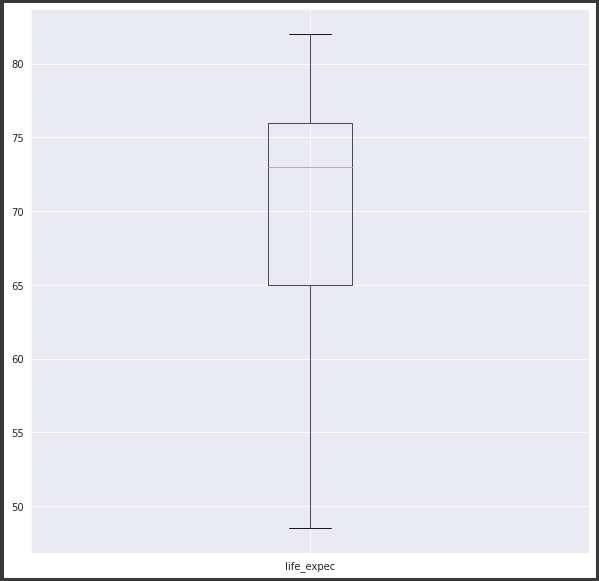


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘child\_mort’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘child\_mort’]-nya. 

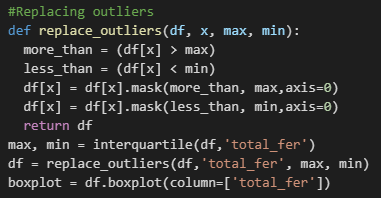
* 1. **Testing dan Replace kolom [‘life\_expec’]**

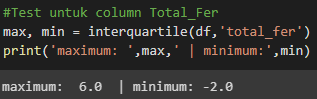
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“life\_expec”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

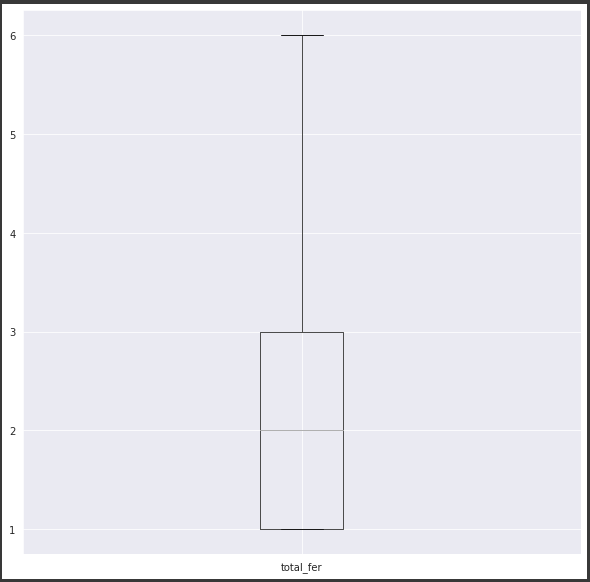


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘life\_expec’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘life\_expec’]-nya.

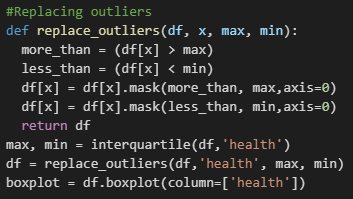
* 1. **Testing dan Replace kolom [‘total\_fer’]**

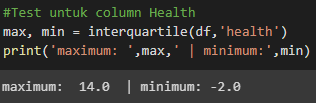
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“total\_fer”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

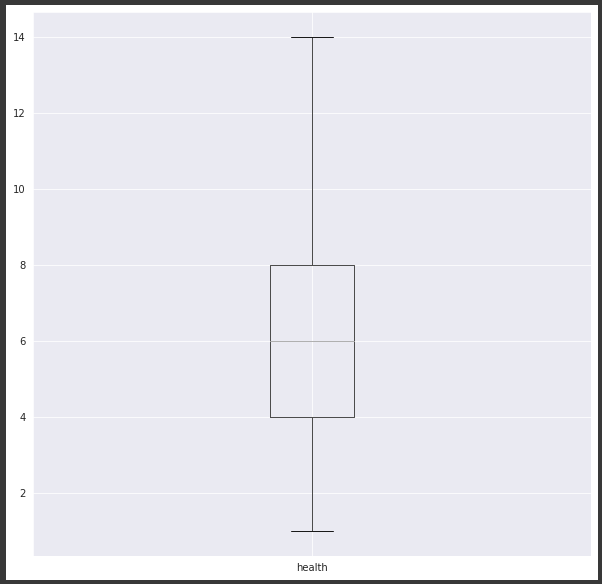


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘total\_fer’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘total\_fer’]-nya.

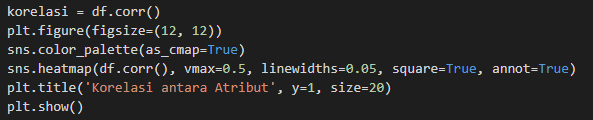
* 1. **Testing dan Replace kolom [‘health’]**

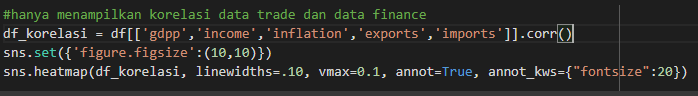
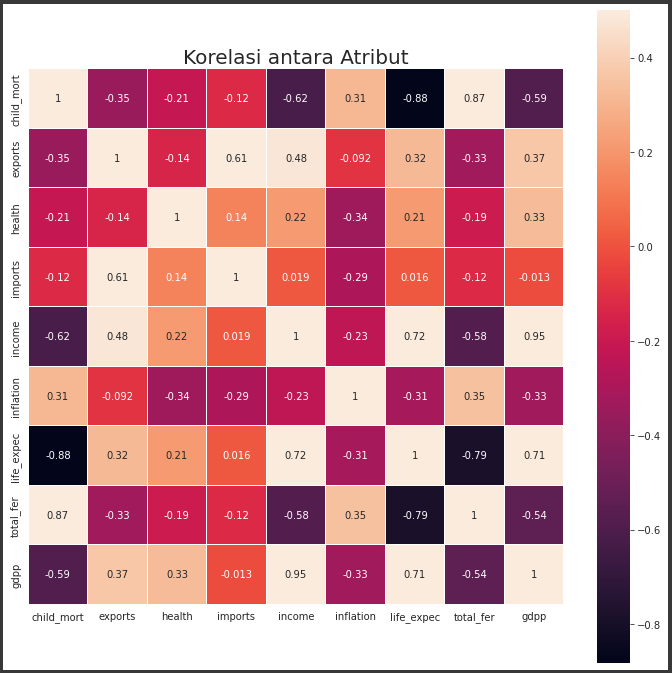
Setelah memakai rumus untuk membuat interquartile, dalam sub bab ini akan melakukan testing dan replace nilai ke dalam kolom [“health”] yang diharapkan supaya nilai dari data pencilan atau outliers tersebut bisa diatasi. Berikut adalah penggunaan untuk testingnya dan replace value kolomnya.

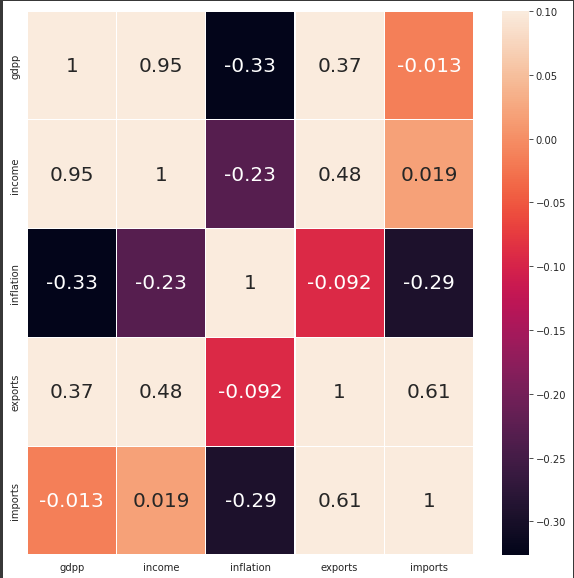


Penggunaan codingan di atas untuk menunjukkan maximal dan minimal dari penghitungan interkuartil dari kolom [‘health’] serta visualisasi dari boxplot kolum gdpp. Pada visualisasinya didapatkan tidak ada sama sekali pencilan ataupun outlier yang berada di garis atas maupun bawah. Berikut tampilan boxplot kolom [‘health’]-nya.

1. **Mengecek Korelasi Antar Data**

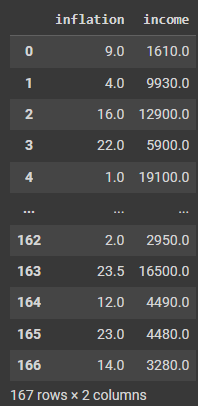
Pada sub bab ini, saya akan mengecek keterhubungan antar data menggunakan seaborn yang menampilkan seperti heatmap data dari persebaran atributnya. Atribut yang dipakai ialah seluruh dari kolom dataset yang ada terkecuali kolom [‘country’] yang mana kolom tersebut sudah dihapus pada bagian pemrosesan data di sub bab pertama. Berikut adalah data yang dipakai untuk mengecek korelasi.

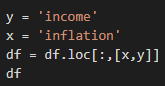
Adapun jika dilihat dari hasil heatmap tersebut. Map tersebut menampilkan seluruh keterhubungan data. Karena saya ingin semua data yang ditampilkan ialah data yang ada hubungannya dengan keuangan dan pendapatan, maka saya akan coba tampilkan heatmap dengan kolom gdpp, income, inflation, imports, dan exports serta mengabaikan empat kolom lainnya untuk melihat kembali keterhubungan antara kolom yang disebutkan.



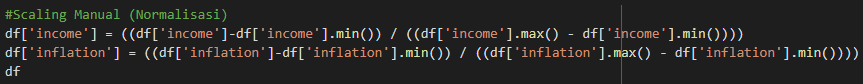
Adapun dari map ataupun korelasi tersebut, saya simpulkan kolom yang paling kecil yakni gdpp - inflasi dengan nilai -0.33 dan paling besar gdpp - income dengan nilai 0.95. Maka saya ambil kolom inflasi dan income untuk splitting data dan proses selanjutnya.

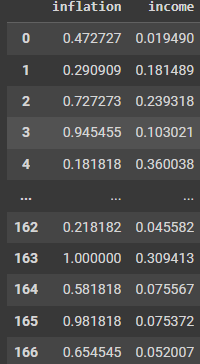
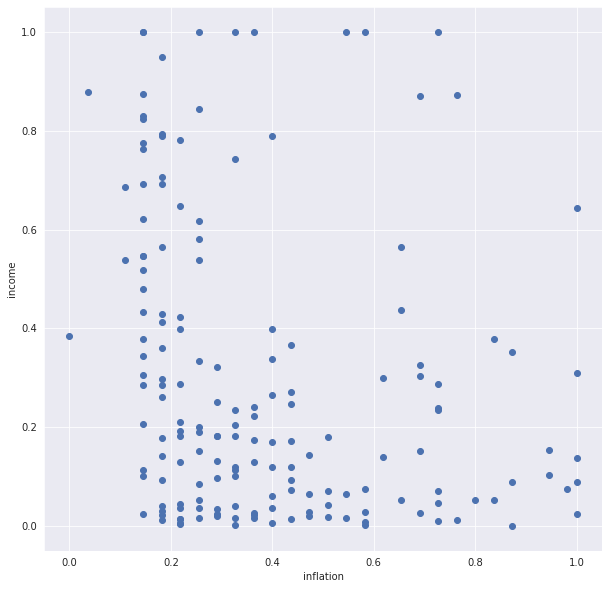
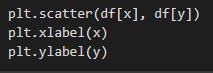
1. **Splitting Data**

Pada bagian ini, saya mencoba split data dari kolom yang sudah saya pilih pada bagian korelasi data, yakni kolom inflasi dan kolom income. Splitting pada bagian ini juga bertujuan untuk memastikan bahwa data yang dipakai sudah bagus dan akurat supaya bisa menghasilkan data yang lebih baik pada proses berikutnya. Berikut adalah splitting kolom yang dipakai.



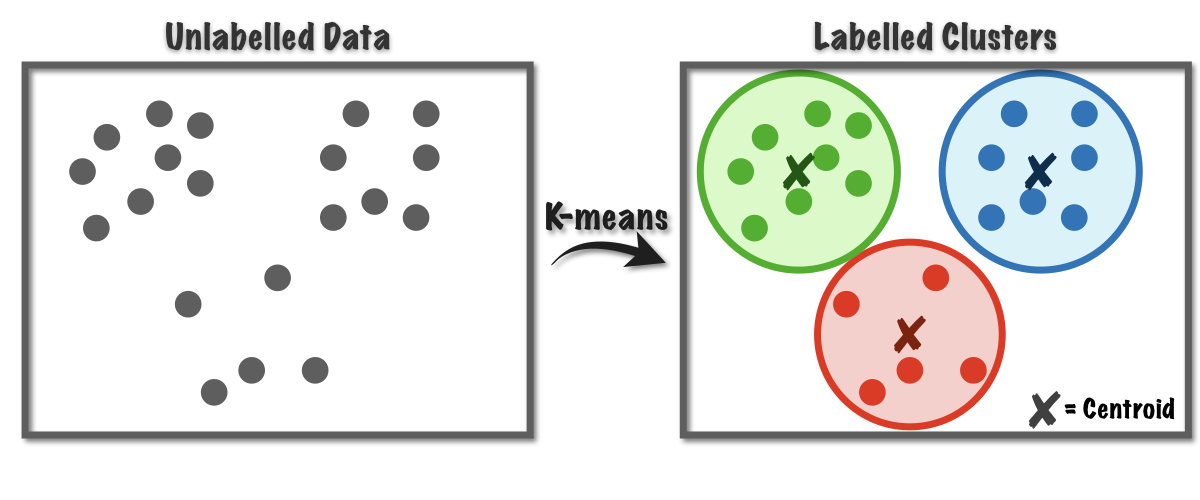
1. **Scalling Data**

Setelah selesai membagi data ataupun kolom, saya mencoba normalisasi data tersebut dengan memakai rumus minmax scaler yang berfungsi sebagai normalisasi, yakni membuat data dapat normal atau jarak pencilan dan data tersebut tidak terlalu jauh dan supaya data berada pada interval 0 dan 1. Berikut adalah rumus codingan yang saya gunakan untuk melakukan normalisasi serta visualisasi data dari normalisasi.

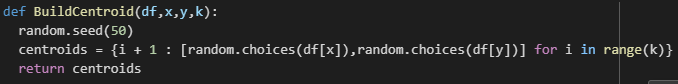


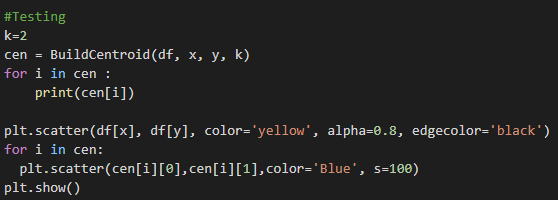
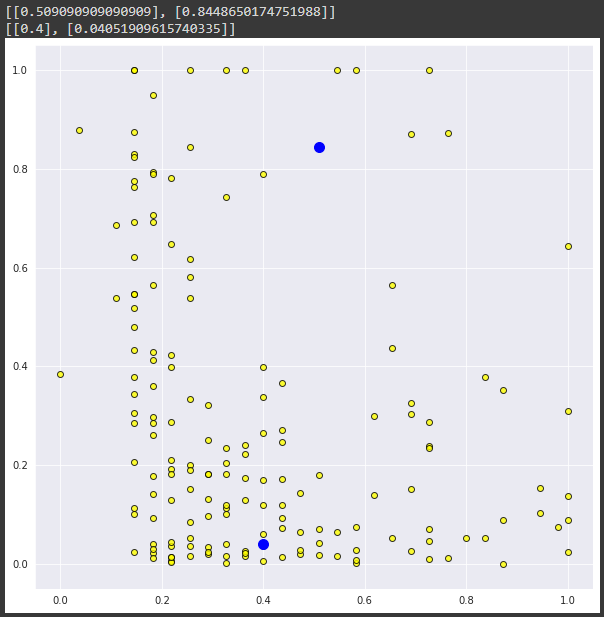
1. **Menerapkan Algoritma yang dipilih**
2. **Algoritma Clustering K-Means**

K-Means merupakan algoritma yang bersifat unsupervised learning. K-Means memiliki fungsi untuk mengelompokkan data kedalam data cluster. Algoritma ini dapat menerima data tanpa ada label kategori. K-Means Clustering Algoritma juga merupakan metode non-hierarchy. Metode Clustering Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain.

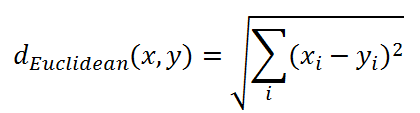
Cluster Sampling adalah teknik pengambilan sampel di mana unit-unit populasi dipilih secara acak dari kelompok yang sudah ada yang disebut ‘cluster’. K-Means Clustering memiliki objektivitas yaitu meminimalisasi objek function yang telah diatur pada proses clusterisasi dengan cara minimalisasi variasi antar satu cluster dengan maksimalisasi variasi pada data di cluster lainnya.

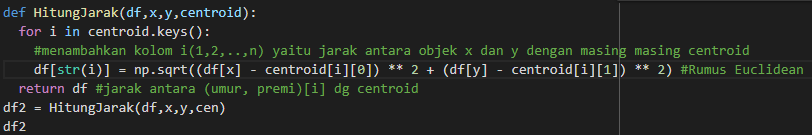
1. **Membuat Centroid**

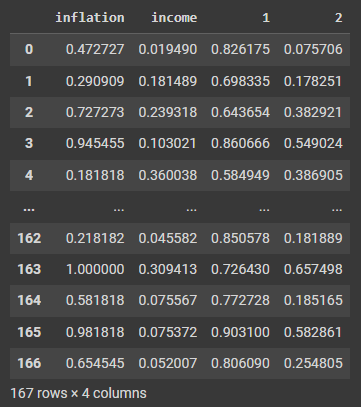
Pada sub bab ini, mulai pemrosesan data menggunakan K-Means Clustering dengan pertama membuat atau inisiasi dari centroidnya terlebih dahulu. Saya menggunakan data yang ada untuk random koordinat dari titik centroid yang akan dibuat dengan mencoba test nilai K adalah 2. Berikut cara membuat centroid serta visualisasinya. 



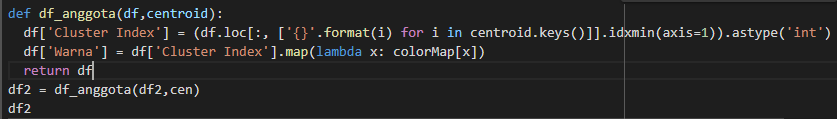
1. **Menghitung Jarak**

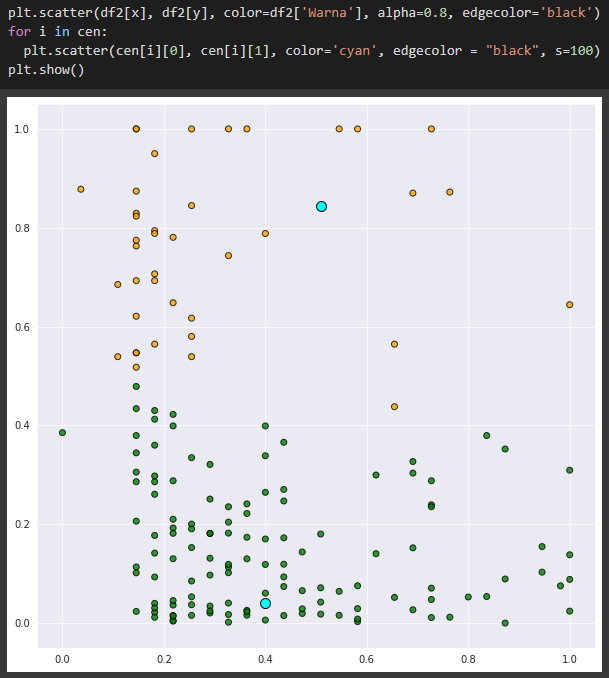
Setelah didapat centroidnya, langsung saja di bagian ini akan menghitung jarak dari centroid dengan setiap atribut-atribut yang ada. Rumus untuk menghitung jarak ini adalah memakai metode euclidean. Rumus tersebut yang nanti data selanjutnya bisa diproses pada bagian selanjutnya. Berikut adalah penerapannya pada codingan google collabs yang saya buat.

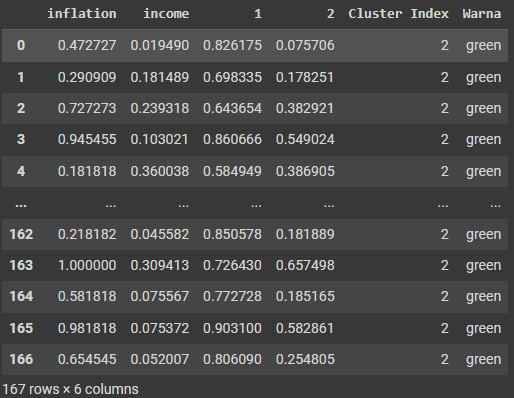


Setelah datanya telah dihitung. pada code di atas, saya langsung menentukan tiap baris mana yang akan dimasukkan ke dalam kolom baru yang akan menyimpan nilai jarak antara kolom dengan masing-masing centroid terdekat. Berikut adalah visualisasi dari penghitungan data serta penambahan kolom baru.

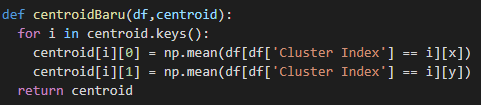
1. **Membuat Keanggotaan dari Penghitungan Jarak**

Setelah selesai menghitung jarak dan memasukkan hasilnya pada kolom baru yakni centroid 1 dan centroid 2, pada sub bab ini akan membuat Keanggotaan yang menyatakan antara kolom [‘income’] dan kolom [‘inflation’] akan masuk ke dalam index cluster yang sudah dibuat. Apakah cluster 1 atau cluster 2? Hal tersebut dilihat dari nilai terkecil ataupun terdekat pada baris, misalkan nilai terkecil adalah pada cluster 1, maka kolom [‘inflation’] dan [‘income’] pada baris 1 akan dimasukkan ke dalam index cluster 1. Berikut cara dan visualisasi dari keanggotaan yang telah dibuat.

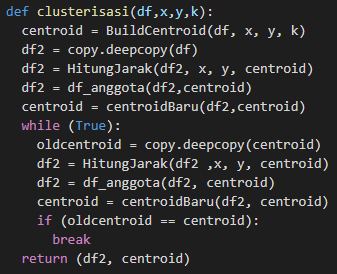


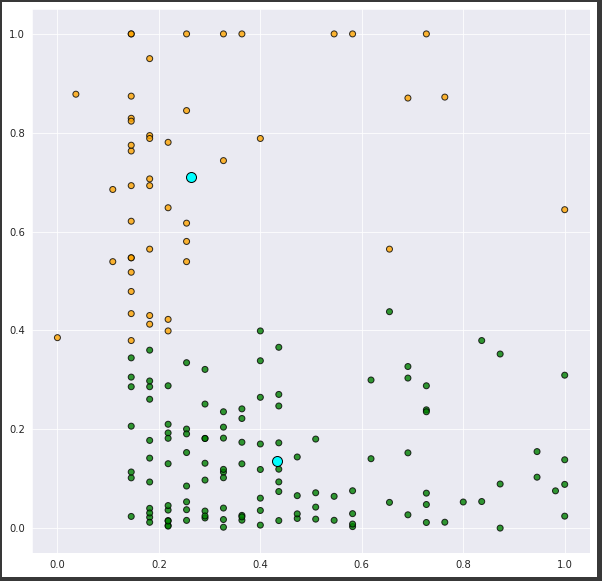


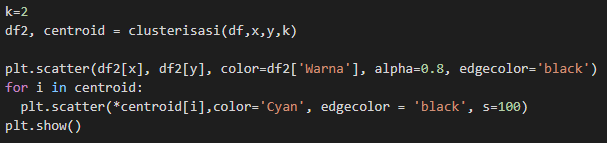
1. **Menentukan Centroid Baru**

Pada Subbab ini, hanya melakukan pembuatan centroid baru yang mana centroid ini merupakan centroid akhir agar centroid yang dilakukan di awal tidak berubah-ubah dan tetap dalam menentukan letak centroid pada atributnya. Berikut adalah cara untuk menentukan centroid baru.

1. **Melakukan Clusterisasi Fungsi K-Means**

Setelah membuat centroid baru, saatnya melakukan clusterisasi K-Means. Clusterisasi ini memakai beberapa fungsi yang digunakan pada fungsi-fungsi sebelumnya seperti menghitung jarak, membuat centroid, membuat keanggotaan, serta membuat centroid baru. Pada main fungsi ini, jika masih bertemu dengan centroid yang lama, maka akan berhenti jika sampai bertemu dengan centroid yang baru. Berikut adalah cara menggunakannya.

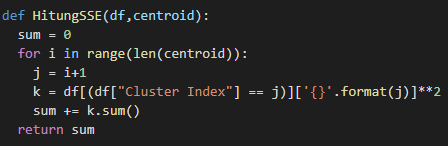
Setelah membuat bagian code berikut, maka saya akan mencoba test untuk memakai jumlah K = 2. Berikut adalah visualisasi K = 2. 



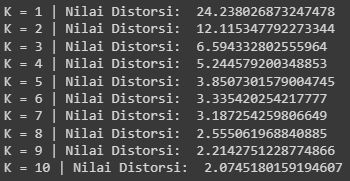
1. **Elbow Method**

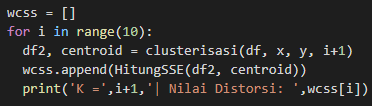
Pada sub bab ini, saya akan menggunakan elbow method untuk melihat nilai K terbaik yang akan digunakan untuk clusterisasi K-Means dengan meminimumkan jarak antara data point dan centroid serta memaksimalkan jarak antara centroid yang dihitung memakai WCSS (within-cluster sum of Squares). Tujuan memakai WCSS adalah agar bisa dilakukannya interpretasi data, jika WCSS hasilnya 0 berarti semua data point berada pada cluster yang berbeda. Ketika WCSS hasilnya 1 berarti semua data point berada dalam satu cluster.

* 1. **Menghitung SSE (Sum Square Error)**

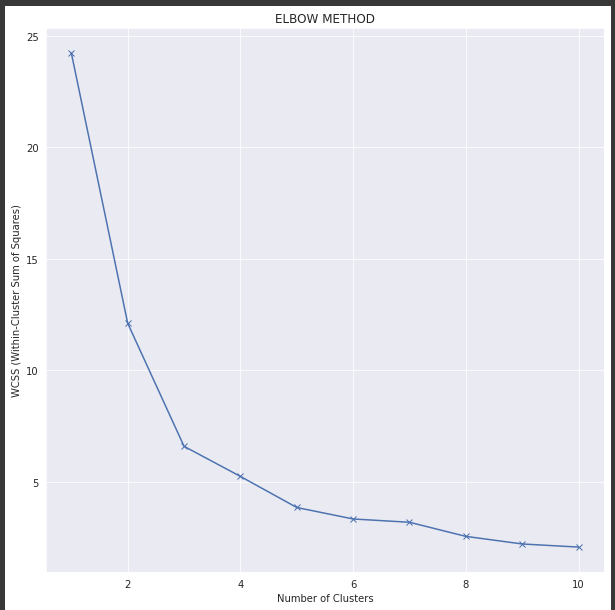
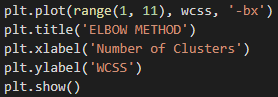
Sebelum membuat Elbow Methodnya, pertama kita menghitung dulu SSE-nya untuk menghitung jumlah jarak kuadrat dari titik-titik yang ada ke pusat yang telah ditetapkan atau bisa disebut dengan distorsi data. Berikut adalah menghitung SSE untuk menerapkan ke dalam elbow methodnya.

* 1. **Menghitung WCSS (within-cluster sum of Squares)**

Setelah membuat SSE, sekarang membuat rumus untuk menghitung WCSS, WCSS ini akan menghitung nilai distorsi dan disimpan pada array. Berikut adalah menghitung WCSS.



* 1. **Visualisasi Elbow Method**

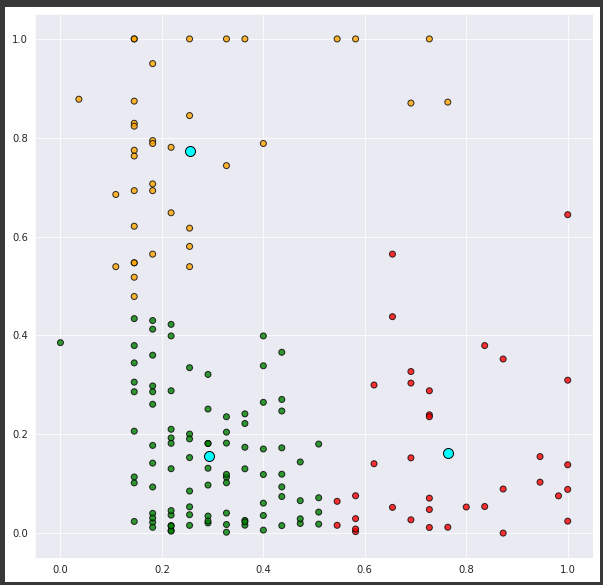
Setelah melakukan penghitungan WCSS dan SSE di atas, sekarang hanya menampilkan bentuk dari elbow methodnya dari range 1 hingga 10. Berikut adalah cara untuk menampilkan hasil dari Elbow Method yang sudah diperoleh.

Jika dilihat dari grafik elbow method di atas, terlihat bahwa lekukan paling patah yakni terjadi pada K = 3, K = 5, serta K = 7.

1. **Evaluasi Hasil**

Pada bagian ini, tahap yang dilakukan ialah evaluasi hasil dari penghitungan nilai K berdasarkan fungsi K-Means sebelumnya untuk mengetahui apakah clustering yang sudah dibuat pada tahap-tahap sebelumnya akurat dan bagus. Berikut beberapa uji coba dan hasil evaluasi dari data Coutry-data.csv yang saya buat

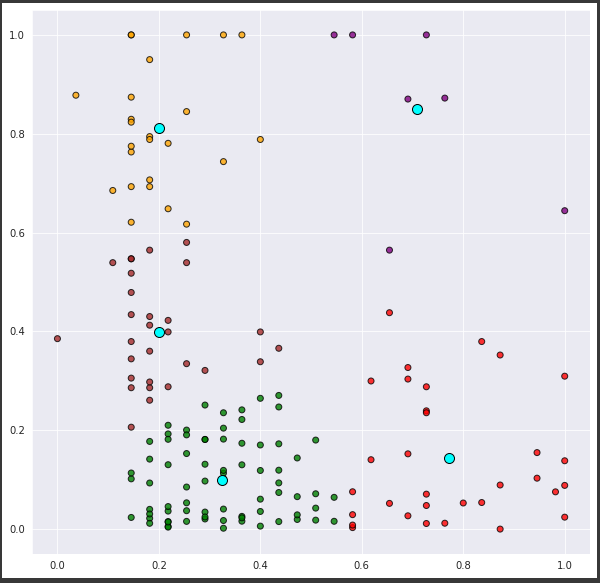
1. **Testing Hasil K = 3**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 3.

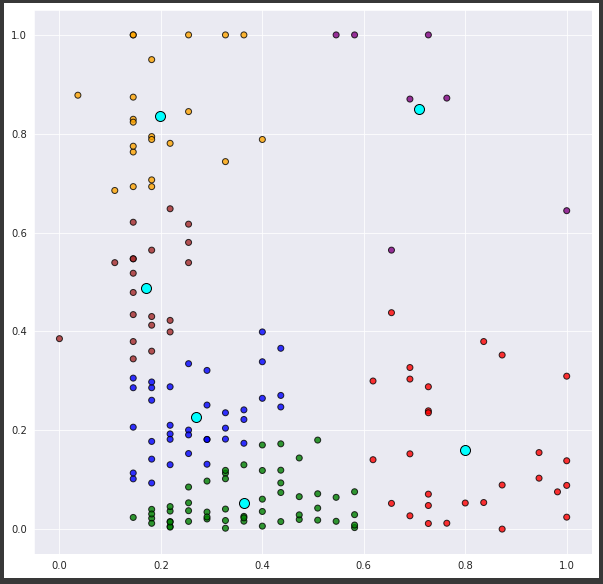
1. **Testing Hasil K = 4**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 4. 

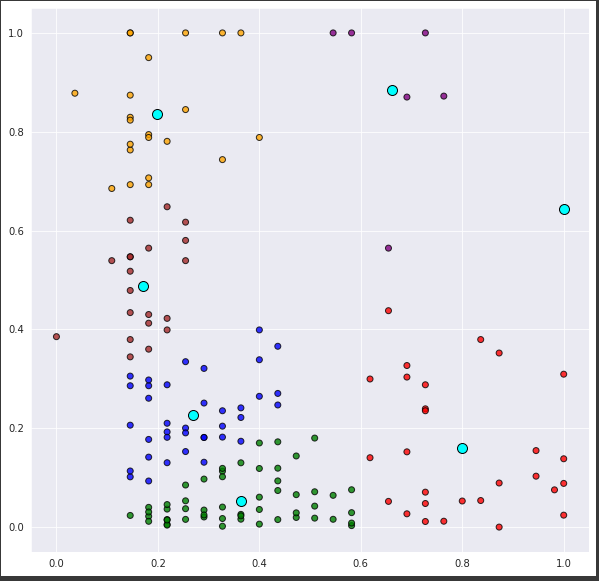
1. **Testing Hasil K = 5**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 5.

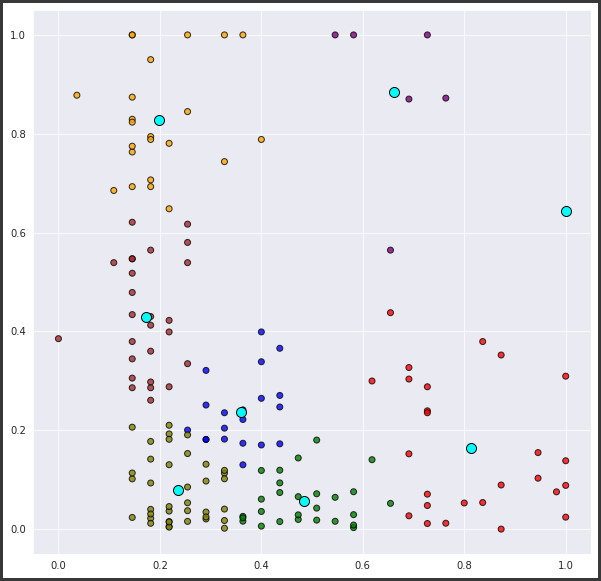
1. **Testing Hasil K = 6**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 6. 

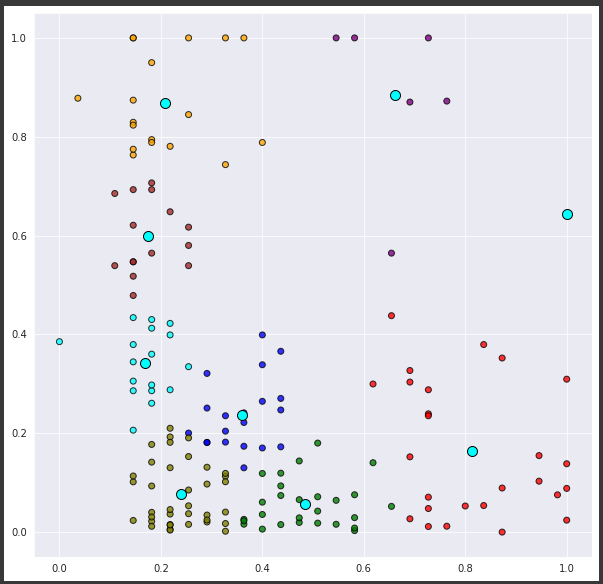
1. **Testing Hasil K = 7**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 7.

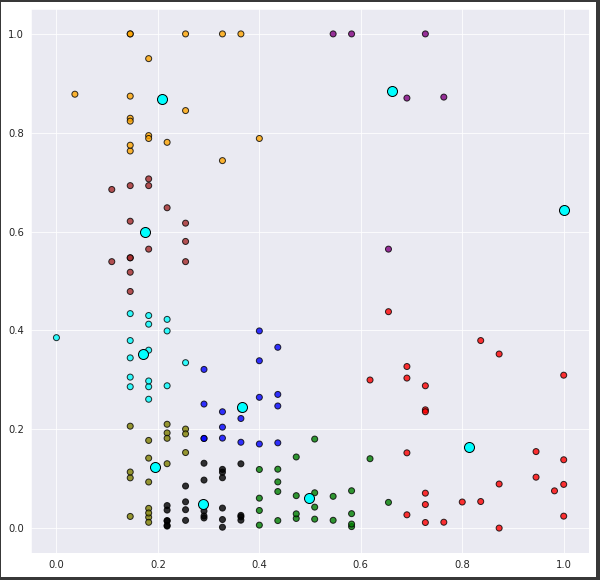
1. **Testing Hasil K = 8**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 8. 

1. **Testing Hasil K = 9**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 9.

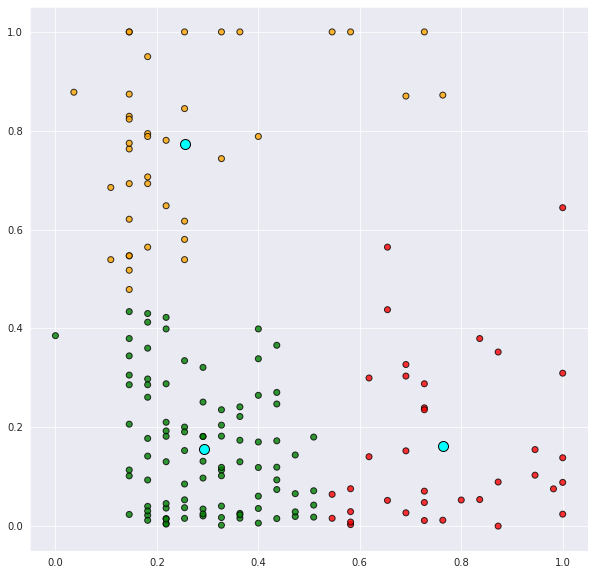
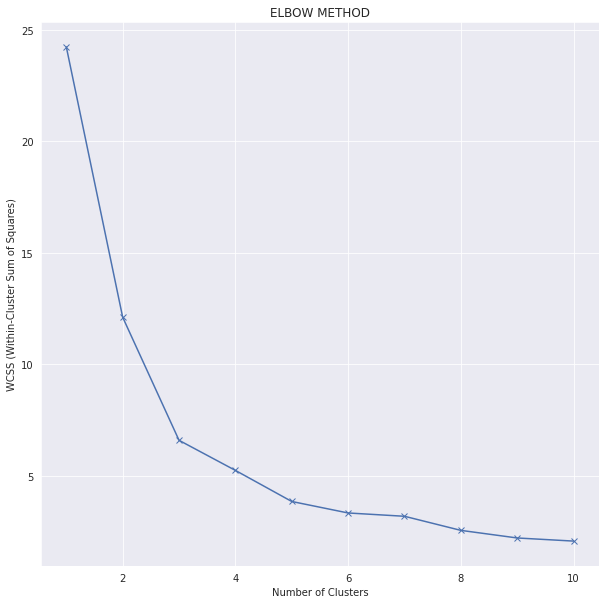
1. **Testing Hasil K = 10**

Berikut hasil visualisasi clustering K-means ketika saya buat nilai K adalah 10.

**BAB 3**

**KESIMPULAN**

1. **Kesimpulan**

Menurut hasil dari evaluasi yang saya amati, hasil clustering terbaik ketika nilai K adalah 3, dikarenakan jika dilihat kembali bahwa data yang saling berdekatan dan pembagian clusteringnya juga rata dan jelas dilihatnya. Hasil pengamatan saya diambil juga dari visualisasi elbow method yang mana garis yang sangat terlihat siku adalah pada nilai K = 3, yang bisa diartikan pada titik tersebut akan mengubah cluster menjadi lebih baik dan pembagiannya sama rata dari jumlah data yang akan dibuat clusterisasi.

1. **Lampiran**

**Link Codingan (Google Collabs) :**

[*https://colab.research.google.com/drive/19egJB\_jHSXSMdLBzbG0KmDWqEpvzrRke?usp=sharing*](https://colab.research.google.com/drive/19egJB_jHSXSMdLBzbG0KmDWqEpvzrRke?usp=sharing)

**Link PowerPoint :**

[*https://docs.google.com/presentation/d/1ivmXZxBWoG9Noq5ADjEPF8cPegH74k6T/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true*](https://docs.google.com/presentation/d/1ivmXZxBWoG9Noq5ADjEPF8cPegH74k6T/edit?usp=sharing&ouid=100359807533626157288&rtpof=true&sd=true)

**Link Video (Youtube) :**

[*https://youtu.be/OLC67CJuQZs*](https://youtu.be/OLC67CJuQZs)