



LAPORAN ANALISA

Help International



JANUARY 29, 2022

SANBERCODE
Data Science

Laporan Analisa

HELP International telah berhasil mengumpulkan sekitar \$ 10 juta. Saat ini, CEO LSM perlu memutuskan bagaimana menggunakan uang ini secara strategis dan efektif. Jadi, CEO harus mengambil keputusan untuk memilih negara yang paling membutuhkan bantuan.

Untuk menyelesaikan masalah agak target negara tidak salah saya menganalisis berdasarkan data yang diberikan, penilaian negara terbelakang itu bisa dilihat dari berbagi aspek ekonomi, Kesehatan, dan lain sebagainya

Disini saya menggunakan data yang di berikan dan menganalisisnya

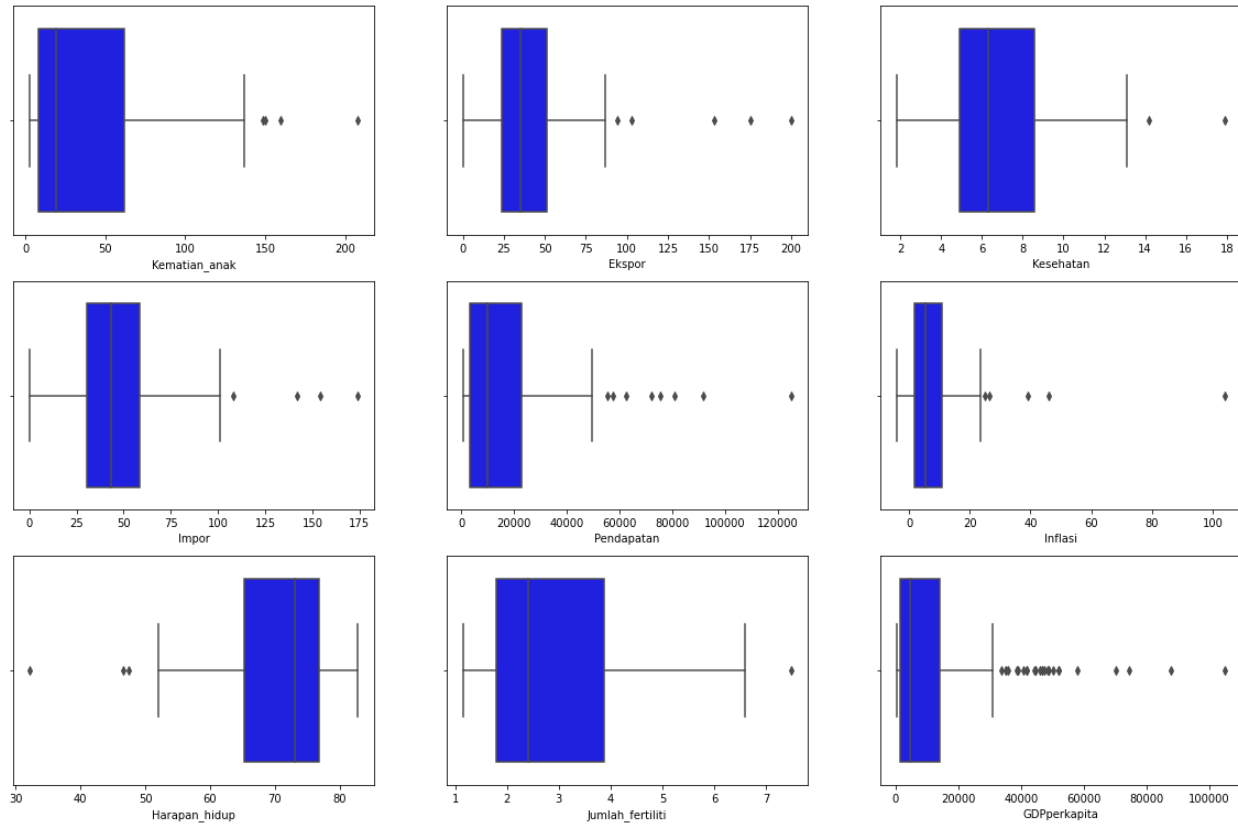
Pertama data tersebut harus di cek apak terdapat missing value atau tidak dan total negara yang ada di data tersebut ada sebanyak 167 negara

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200
...
162	Vanuatu	29.2	46.6	5.25	52.7	2950	2.62	63.0	3.50	2970
163	Venezuela	17.1	28.5	4.91	17.6	16500	45.90	75.4	2.47	13500
164	Vietnam	23.3	72.0	6.84	80.2	4490	12.10	73.1	1.95	1310
165	Yemen	56.3	30.0	5.18	34.4	4480	23.60	67.5	4.67	1310
166	Zambia	83.1	37.0	5.89	30.9	3280	14.00	52.0	5.40	1460

167 rows × 10 columns

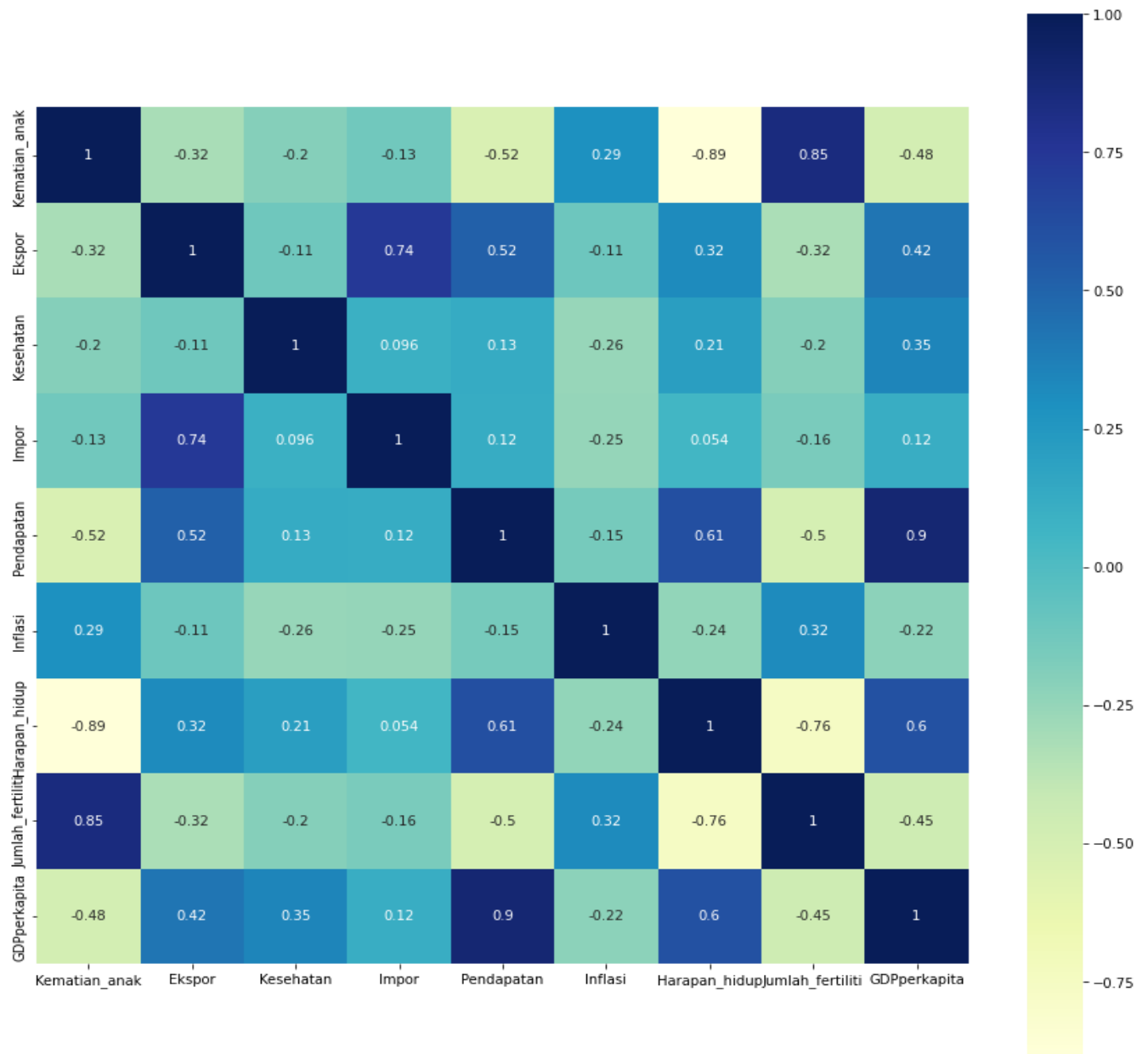
Dilihata dari data tersebut ada kolom kematian anak, ekspor, Kesehatan, impor, pendapatan, inflasi, harapan hidup, jumlah fertility, GDPperkapita.

Dari data di atas saya cek missing value, data duplicated dan tidak ada yang missing dan data duplikat, setelah itu kita cek data outlier nya dan di dapatkan boxplot seperti ini



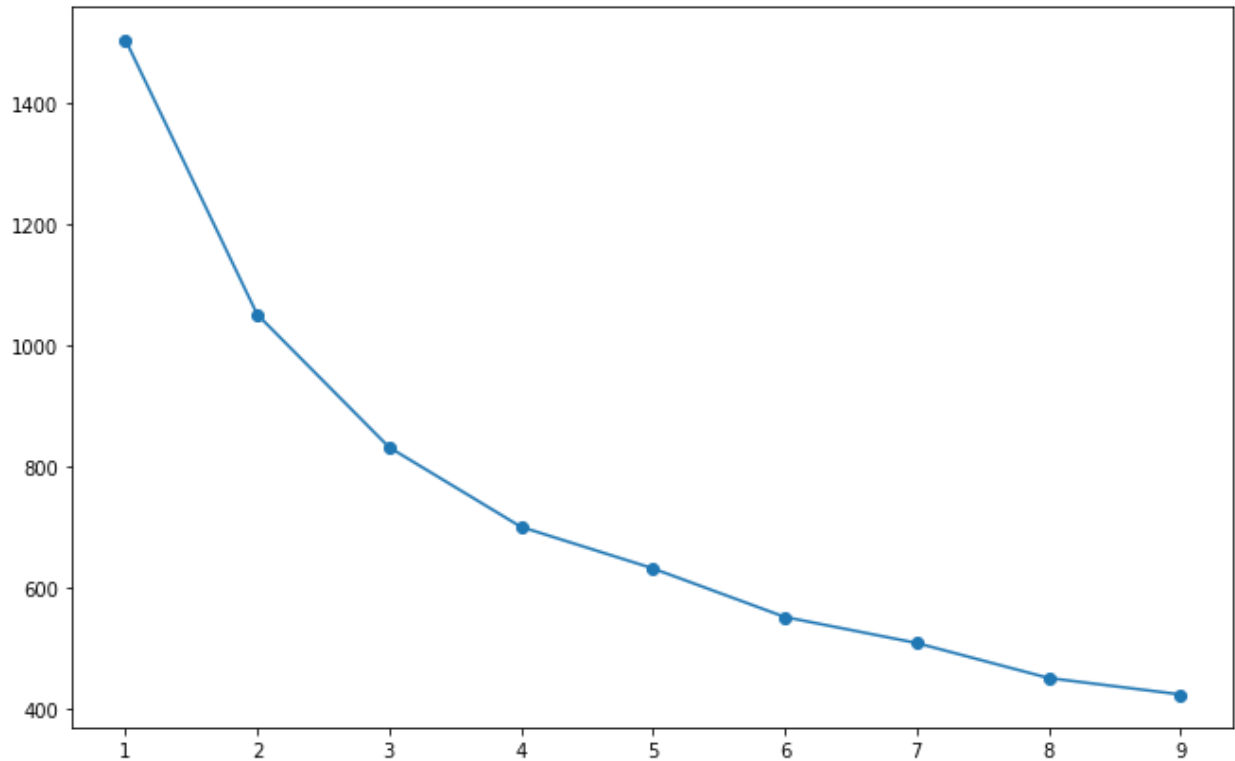
Kita bisa lihat ada banyak sekali data outlier , nilai outlier atau data pencilan tersebut bisa berdampak serius pada tidak kesesuaian data Analisa dan ketidak akuratan model machine learning yang di kerjakan. Untuk itu saya bisa hapus / drop data pencilan tersebut. Tapi saya mengambil keputusan untuk tidak di hapus, karena kemungkinan data terebut benar dan sangat mungkin jika ada outlier pada kolom-kolom tersebut.

Setelah cek data outlier, saya melihat korelasi atau hubungan antara sepasang data. nilai korelasi berkisar antara (-1) hingga (+1) bergantung pada nilai kovariansinya. Jika nilai korelasi mendekati -1 atau mendekati +1 maka data tersebut memiliki korelasi linier yang tinggi

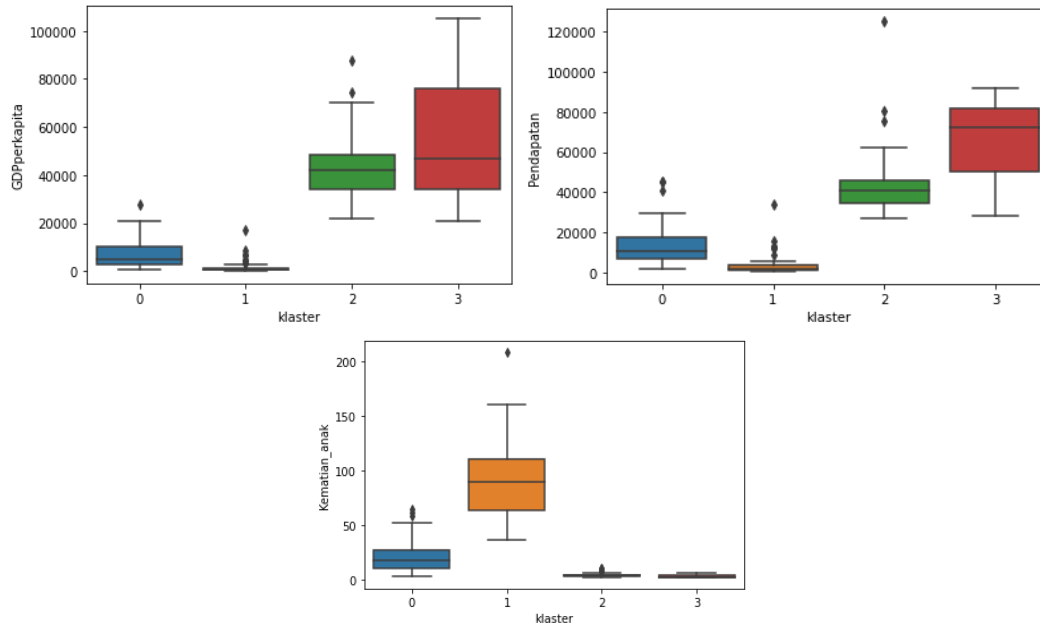


Dilihat dari visualisasi di atas korelasi dari setiap kolom, kolom jumlah_fertiliti dan kematian_anak cukup bagus karena mendekati angka 1, kolom impor dan ekspor juga memiliki korelasi bagus

Pada Analisis ini saya menggunakan algoritma K-Means clustering, algoritma ini digunakan untuk mencari grup yang belum di beri label secara eksplisit dalam data. Ini dapat digunakan untuk mengkonfirmasi asumsi tentang jenis grup yang ada atau untuk mengidentifikasi grup yang tidak dikenal dalam kumpulan data yang kompleks.

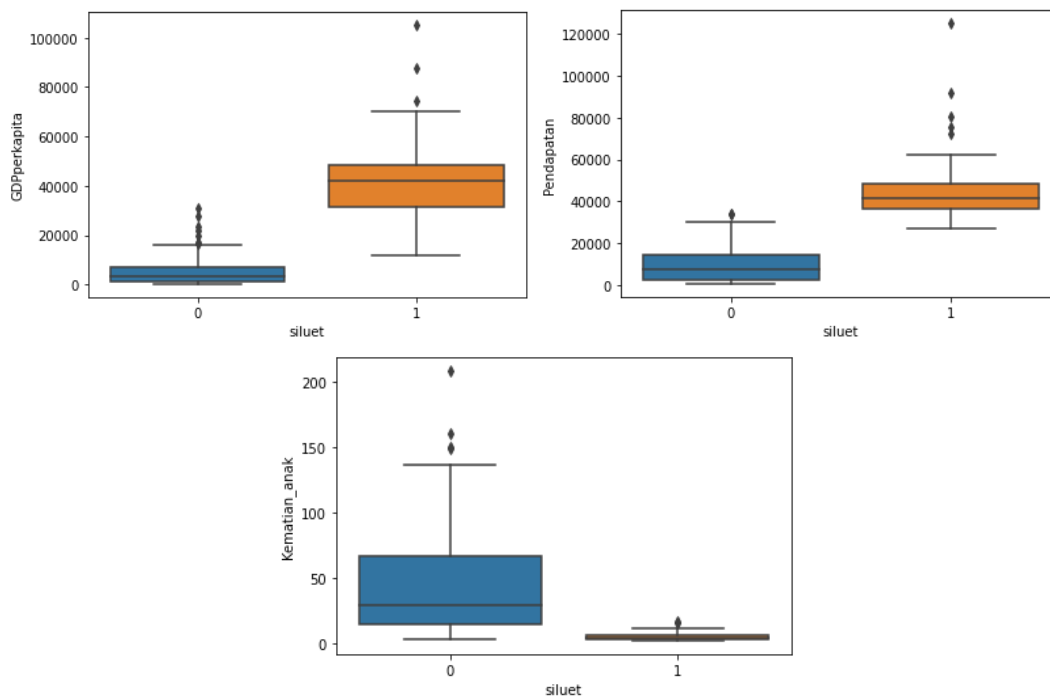


Di lihat dari visualisasi di atas clustering ini menggunakan method elbow, dimana method ini membandingkan jumlah cluster dengan inersianya. Saya mengambil 4 kluster, disini saya mendapatkan nilai predict, artinya memprediksi jumlah cluster Negara atau kelompok mana yang termasuk dalam cluster mana.



Kluster 1 itu mendefinisikan untuk kolom GDPperkapita dan Pendapatan yang terkecil, untuk kolom kematian_anak kluster 1 ini mendefinisikan kematian anak tertinggi.

Kenapa saya menggunakan kolom tersebut dalam analisis ini, karena pendapatan itu berpengaruh terhadap ekonomi suatu negara, kematian_anak juga berpengaruh terhadap kemajuan negara karena yang dapat mengubah suatu negara agar lebih baik itu adalah generasi muda nya jika kematian anak dalam suatu negara tinggi otomatis kemajuan negara tersebut terhambat, GDPperkapita juga berpengaruh terhadap pertumbuhan ekonomi dalam suatu negara karena jika GDP suatu negara hanya sedikit perkembangannya juga lambat.



Saya juga menggunakan siluet cluster yang mengambil 2 cluster, fungsinya juga sama dengan yang KMeans tapi ini menggunakan AgglomerativeClustering.

Semua cluster KMeans dan AgglomerativeClustering saya masukan kekolom dengan nama kluster dan siluet

Dan saya filter data tersebut dengan kluster 0 dan siluet 1 dan keluar kolom seperti berikut

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	kluster	siluet
0	Afghanistan	90.2	10.00	7.58	44.9	1610	9.440	56.2	5.82	553	1	0
3	Angola	119.0	62.30	2.85	42.9	5900	22.400	60.1	6.16	3530	1	0
17	Benin	111.0	23.80	4.10	37.2	1820	0.885	61.8	5.36	758	1	0
21	Botswana	52.5	43.60	8.30	51.3	13300	8.920	57.1	2.88	6350	1	0
25	Burkina Faso	116.0	19.20	6.74	29.6	1430	6.810	57.9	5.87	575	1	0
26	Burundi	93.6	8.92	11.60	39.2	764	12.300	57.7	6.26	231	1	0
28	Cameroon	108.0	22.20	5.13	27.0	2660	1.910	57.3	5.11	1310	1	0
31	Central African Republic	149.0	11.80	3.98	26.5	888	2.010	47.5	5.21	446	1	0
32	Chad	150.0	36.80	4.53	43.5	1930	6.390	56.5	6.59	897	1	0
36	Comoros	88.2	16.50	4.51	51.7	1410	3.870	65.9	4.75	769	1	0
37	Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.800	57.5	6.54	334	1	0
38	Congo, Rep.	63.9	85.10	2.46	54.7	5190	20.700	60.4	4.95	2740	1	0
40	Cote d'Ivoire	111.0	50.60	5.30	43.3	2690	5.390	56.3	5.27	1220	1	0
49	Equatorial Guinea	111.0	85.80	4.48	58.9	33700	24.900	60.9	5.21	17100	1	0
50	Eritrea	55.2	4.79	2.66	23.3	1420	11.600	61.7	4.61	482	1	0
55	Gabon	63.7	57.70	3.50	18.9	15400	16.600	62.9	4.08	8750	1	0
56	Gambia	80.3	23.80	5.69	42.7	1660	4.300	65.5	5.71	562	1	0

Kolom tersebut hanya Sebagian dari 47 negara.

Dan setelah itu saya sorting lagi berdasarkan kolom GDPperkapita, Pendapatan, Kematian_anak dengan data 10 teratas, keluar table seperti ini

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita	klaster	siluet
26	Burundi	93.6	8.92	11.60	39.2	764	12.30	57.7	6.26	231	1	0
88	Liberia	89.3	19.10	11.80	92.6	700	5.47	60.8	5.02	327	1	0
37	Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.80	57.5	6.54	334	1	0
112	Niger	123.0	22.20	5.16	49.1	814	2.55	58.8	7.49	348	1	0
132	Sierra Leone	160.0	16.80	13.10	34.5	1220	17.20	55.0	5.20	399	1	0
93	Madagascar	62.2	25.00	3.77	43.0	1390	8.79	60.8	4.60	413	1	0
106	Mozambique	101.0	31.50	5.21	46.2	918	7.64	54.5	5.56	419	1	0
31	Central African Republic	149.0	11.80	3.98	26.5	888	2.01	47.5	5.21	446	1	0
94	Malawi	90.5	22.80	6.59	34.9	1030	12.10	53.1	5.31	459	1	0
50	Eritrea	55.2	4.79	2.66	23.3	1420	11.60	61.7	4.61	482	1	0

Kita dapat kesimpulan dari table tersebut, saya merekomendasikan 10 negara tersebut untuk di bantu dari segi ekomoni seperti pendapatan yang dapat di atur dari segi lapangan pekerjaan, memberikan pelatihan skill terhadap negara tersebut, dan Kesehatan seperti mengurangi angka kematian pada anak dengan menyuplai obat-obatan atau dokter ke negara tersebut. Dan untuk pembagian uang kita bisa membaginya itu untuk satu negara 1 juta dolar dan membaginya untuk kepentingan negara dalam segi peningkatan ekonomi dan Kesehatan.