

# **Clustering the Countries by Using K-Means For HELP International**

**D**

**i**

**s**

**u**

**s**

**u**

**n**

**oleh:**

**Renta Sitorus**

## I. Pendahuluan

### A. Latar Belakang

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. Oleh karena itu kita akan mengkategorikan negara menggunakan beberapa faktor sosial ekonomi dan kesehatan yang menentukan perkembangan negara secara keseluruhan.

### B. Tujuan

Untuk mengkategorikan negara menggunakan faktor sosial ekonomi dan kesehatan yang menentukan pembangunan negara secara keseluruhan.

## II. Pembahasan

### C. Membaca dan Memahami Data

- Mengimport data

	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

Data memiliki 167 baris dan 10 kolom

- Mencari informasi data

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 167 entries, 0 to 166
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Negara                167 non-null   object
1   Kematian_anak         167 non-null   float64
2   Ekspor                167 non-null   float64
3   Kesehatan             167 non-null   float64
4   Impor                 167 non-null   float64
5   Pendapatan            167 non-null   int64
6   Inflasi               167 non-null   float64
7   Harapan_hidup         167 non-null   float64
8   Jumlah_fertiliti      167 non-null   float64
9   GDPperkapita          167 non-null   int64
dtypes: float64(7), int64(2), object(1)
memory usage: 13.2+ KB
```

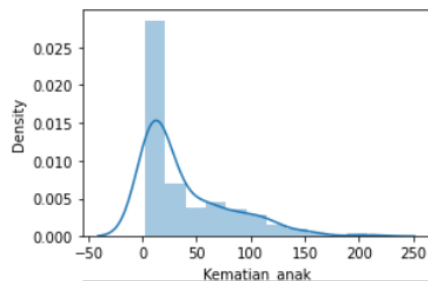
Dari gambar diatas dapat kita lihat informasi bahwa kolom-kolom pada data semuanya berisi.

	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000
mean	38.270060	41.108976	6.815689	46.890215	17144.688623	7.781832	70.555689	2.947964	12964.155689
std	40.328931	27.412010	2.746837	24.209589	19278.067698	10.570704	8.893172	1.513848	18328.704809
min	2.600000	0.109000	1.810000	0.065900	609.000000	-4.210000	32.100000	1.150000	231.000000
25%	8.250000	23.800000	4.920000	30.200000	3355.000000	1.810000	65.300000	1.795000	1330.000000
50%	19.300000	35.000000	6.320000	43.300000	9960.000000	5.390000	73.100000	2.410000	4660.000000
75%	62.100000	51.350000	8.600000	58.750000	22800.000000	10.750000	76.800000	3.880000	14050.000000
max	208.000000	200.000000	17.900000	174.000000	125000.000000	104.000000	82.800000	7.490000	105000.000000

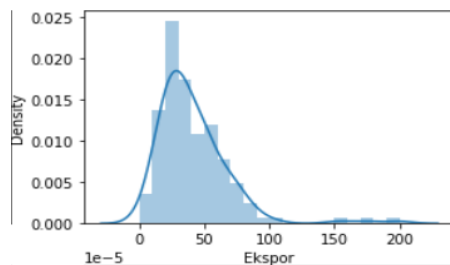
Pada tabel diatas terdapat informasi tentang count dari tiap kolom seperti jumlah kematian anak yaitu 167.000000, dan terdapat informasi mean, standar deviasi, min, dan lain sebagainya.

#### D. Exploratory Data Analyst

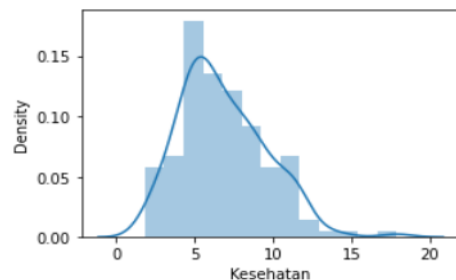
- **Univariate analysis**



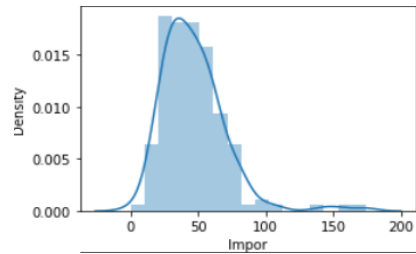
Dari gambar disamping terdapat informasi mengenai Jumlah kematian anak terbanyak antara 0 – 50.



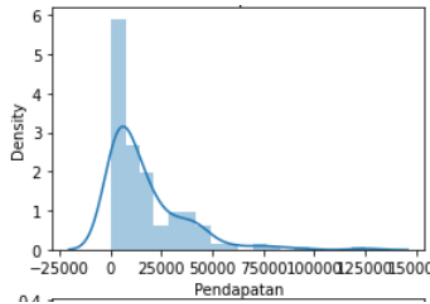
Dari gambar disamping terdapat informasi berupa mengenai Jumlah ekspor terbanyak pada rentang 0 - 50



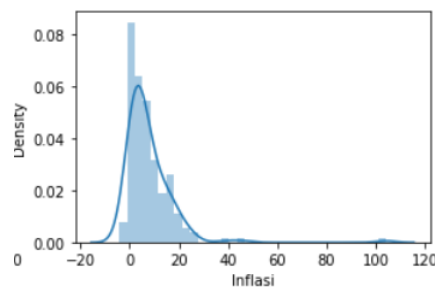
Dari gambar disamping terdapat informasi berupa mengenai Jumlah orang sehat pada rentang 5-10



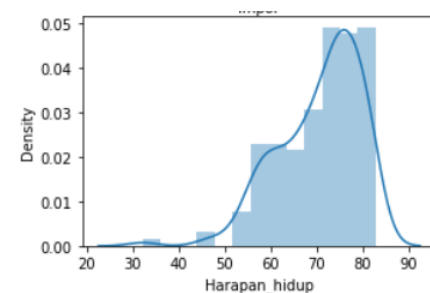
Dari gambar disamping terdapat informasi berupa mengenai Jumlah Impor terbanyak pada rentang 0 – 50



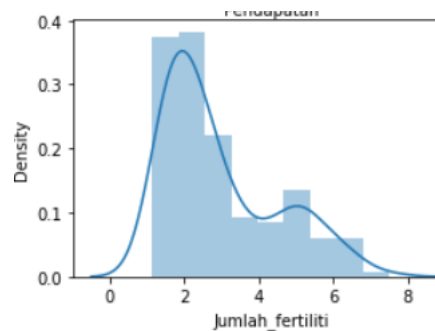
Dari gambar disamping terdapat informasi berupa mengenai Jumlah Pendapatan terbanyak yaitu pada rentang 0 - 25000



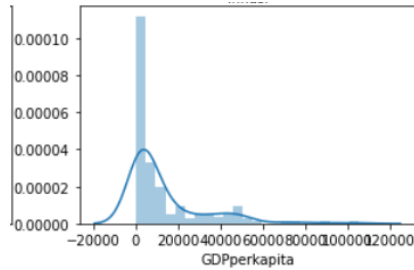
Dari gambar disamping terdapat informasi berupa mengenai Jumlah Inflasi tertinggi yaitu pada rentang 0 - 20



Dari gambar disamping terdapat informasi berupa mengenai Jumlah Harapan Hidup tertinggi pada rentang 70-80



Dari gambar disamping terdapat informasi berupa mengenai Jumlah ferliti tertinggi pada rentang 2-4



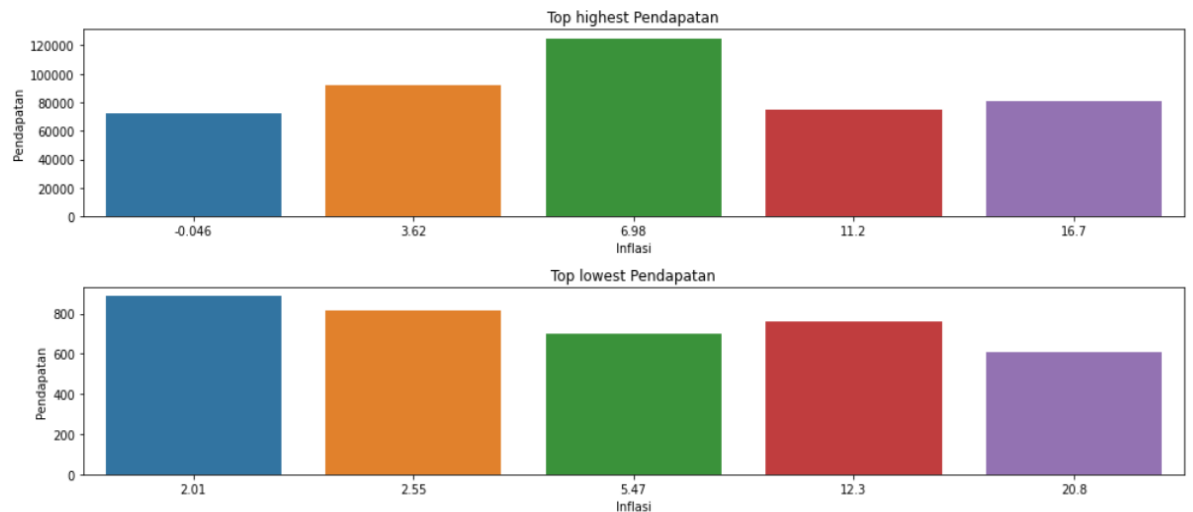
Dari gambar disamping terdapat informasi berupa mengenai Jumlah GDPperkapita tertinggi pada rentang 0-20000

## • Bivariate analysis

	Negara	Kematian_anak	Ekspor	Kesehatan	Impor	Pendapatan	Inflasi	Harapan_hidup	Jumlah_fertiliti	GDPperkapita
123	Qatar	9.0	62.30	1.81	23.8	125000	6.980	79.5	2.07	70300
91	Luxembourg	2.8	175.00	7.77	142.0	91700	3.620	81.3	1.63	105000
23	Brunei	10.5	67.40	2.84	28.0	80600	16.700	77.1	1.84	35300
82	Kuwait	10.8	66.70	2.63	30.4	75200	11.200	78.2	2.21	38500
133	Singapore	2.8	200.00	3.96	174.0	72100	-0.046	82.7	1.15	46600
...	...	...	...	...	...	...	...	...	...	...
31	Central African Republic	149.0	11.80	3.98	26.5	888	2.010	47.5	5.21	446
112	Niger	123.0	22.20	5.16	49.1	814	2.550	58.8	7.49	348
26	Burundi	93.6	8.92	11.60	39.2	764	12.300	57.7	6.26	231
88	Liberia	89.3	19.10	11.80	92.6	700	5.470	60.8	5.02	327
37	Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.800	57.5	6.54	334

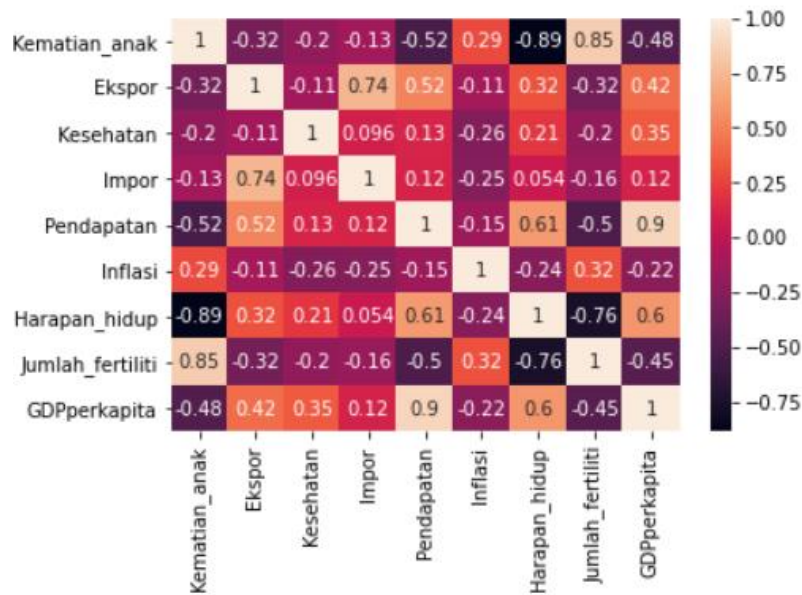
167 rows × 10 columns

Dari tabel diatas disusun berdasarkan Pendapatan tertinggi hingga terendah. Pada posisi tertinggi yaitu negara Qatar dengan Pendapatan sebesar 125000 dan pada posisi terendah yaitu negara Congo, Dem, Rep sebesar 609.



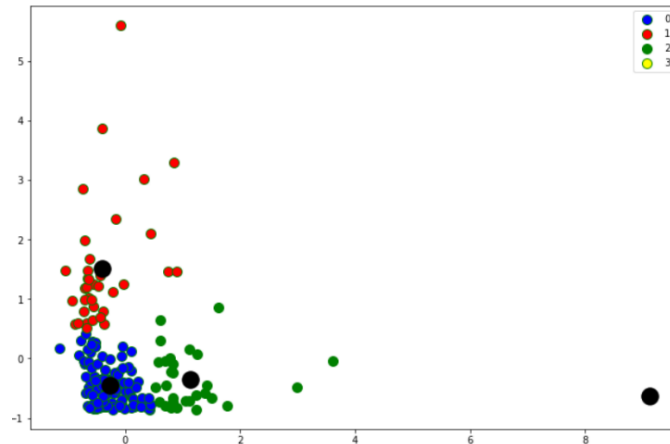
Dari gambar diatas terdapat perbandingan antar jumlah Pendapatan tertinggi dan terendah. Dimana jumlah Pendapatan Tertinggi yaitu pada 120000 dan Inflasi sebanyak 6.98. Jumlah Pendapatan Terendah sebanyak 600 dengan Inflasi sebanyak 5.47

- **Multivariate Analysis**



Dari gambar diatas terdapat kolom yang memiliki korelasi yang tinggi seperti Kematian anak, GDPperkapita dan Pendapatan.

- **Scaling data**



Gambar disamping merupakan hasil Clustering antara Inflasi dan Pendapatan tiap negara. Dapat kita lihat bahwa cluster dari 2 variabel terdapat pada cluster 0, 1, 2. Maka HELP Internasional akan berfokus ke pada negara tersebut.

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

Tabel diatas merupakan hasil pengurutan terendah dari GDPperkapita, dimana HELP internasional harus berfokus kepada negara-negara pada tabel diatas seperti Burundi, Liberia, Congo, Dem, Rep dan lain sebagainya.

### III. Penutup

#### Kesimpulan

HELP International harus mengambil keputusan yang tepat untuk membantu negara-negara yang pendapatan nya rendah, GDPperkapita nya juga rendah dan faktor lainnya. Maka dari itu dari hasil analisis data yang telah dilakukan maka HELP International harus berfokus pada negara – negara pada tabel dibawah ini.

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