



FINAL PROJECT

SANBERCODE
PYTHON – DATA SCIENCE

DATA SCIENCE

Data Negara yang Memerlukan Bantuan dari Segi Sosial Ekonomi dan Kesehatan

Data yang diperoleh berasal dari suatu LSM internasional bernama HELP, dimana LSM tersebut berkomitmen untuk memerangi kemiskinan dan menyediakan fasilitas dan bantuan bagi masyarakat di negara-negara terbelakang. Saat ini HELP telah berhasil mengumpulkan dana sekitar \$ 10 juta, dimana uang tersebut akan diberikan kepada negara terbelakang untuk meningkatkan taraf negara tersebut baik dari segi ekonomi atau kesehatan.

Data tersebut berisi beberapa indikator yaitu, Kematian anak, ekspor, kesehatan, impor, Pendapatan, inflasi, harapan hidup, jumlah fertilitas, GDP perkapita. Terdiri dari 167 baris dan 10 kolom.

Langkah dalam Olah Data

1

Data Cleaning

Mengidentifikasi apakah pada data tersebut terdapat Missing value dan men-sorting data apa saja yang akan digunakan untuk dianalisis.

2

Analisis Univariate

Menganalisis 1 variabel data

3

Analisis Bivariate

Menganalisis data dengan membandingkan 2 variabel data

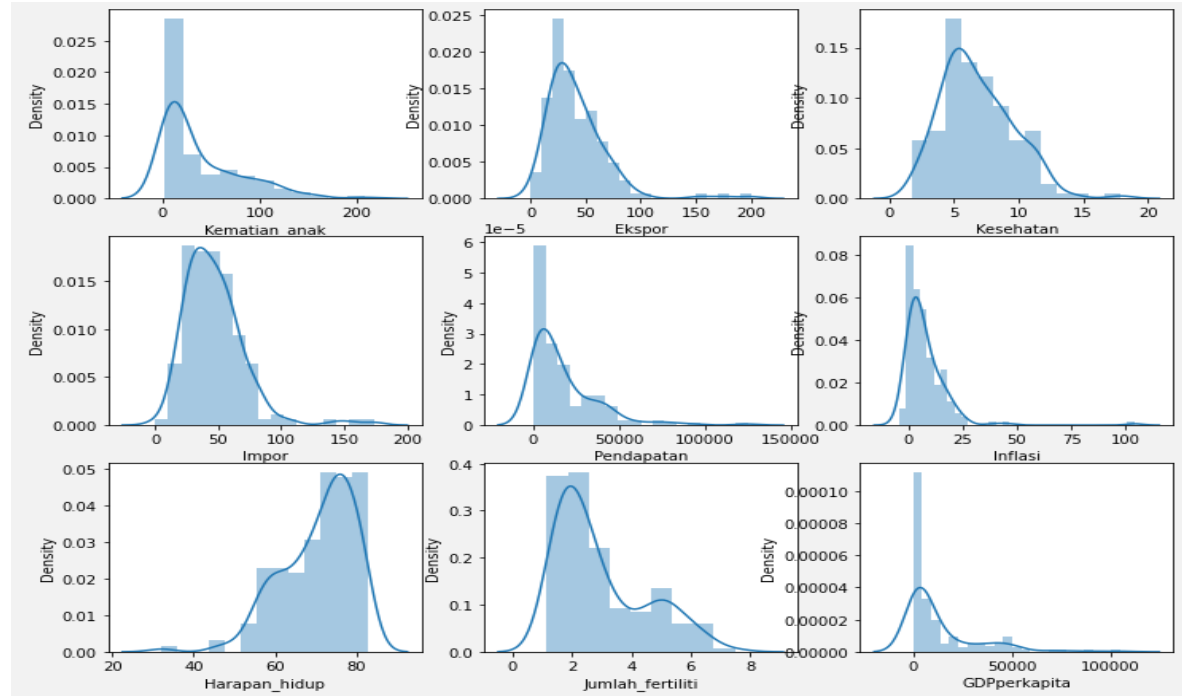
4

Analisis Multivariate

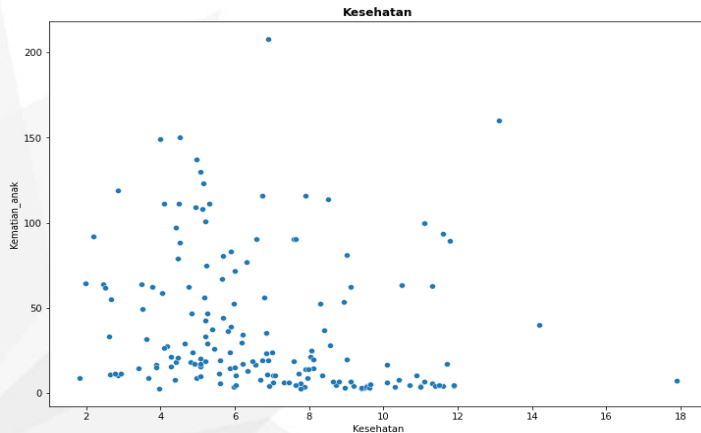
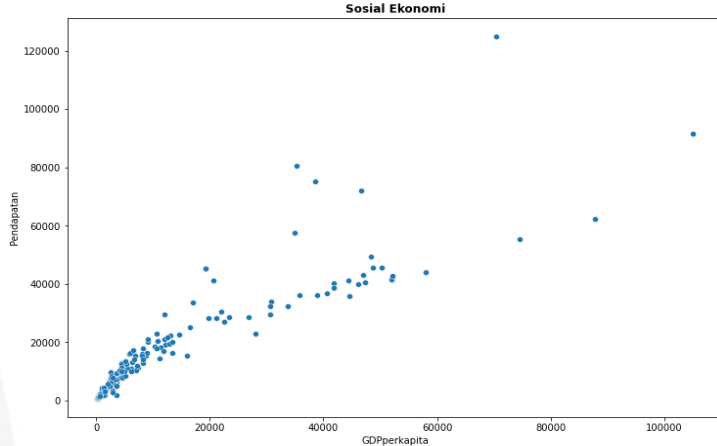
Menganalisis data dengan membandingkan lebih dari 2 variabel data.

Analisis Univariante

- Pada tabel disamping dapat dilihat bahwa sebaran data yang tinggi terdapat pada nilai yang rendah, seperti pada data kematian anak, ekspor, kesehatan, impor, pendapatan, inflasi, jumlah fertiliti, GDP perkapita. Hanya data harapan hidup yang memiliki sebaran data yang tidnggi pada nilali yang tinggi juga.



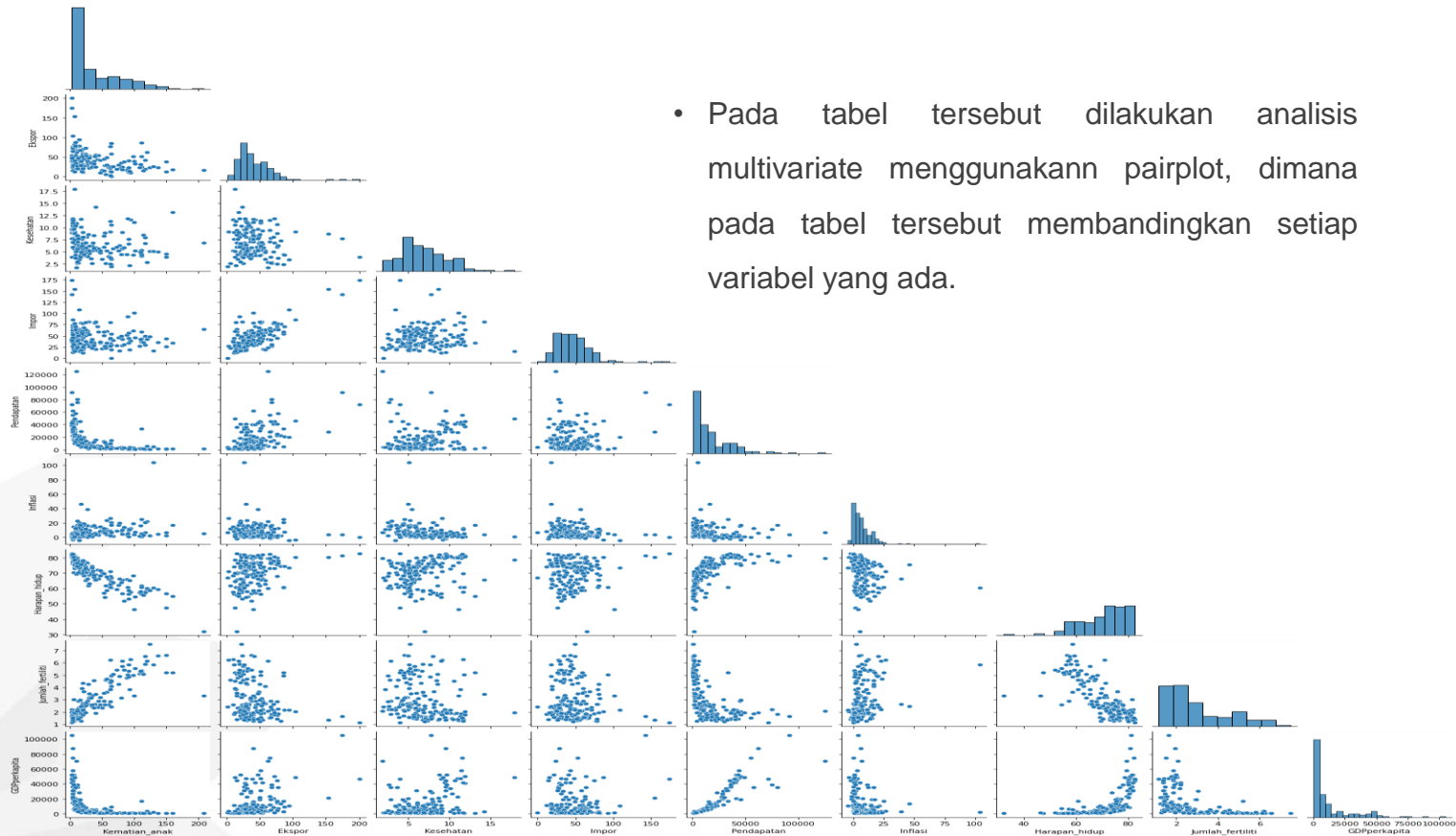
Analisis Bivariate



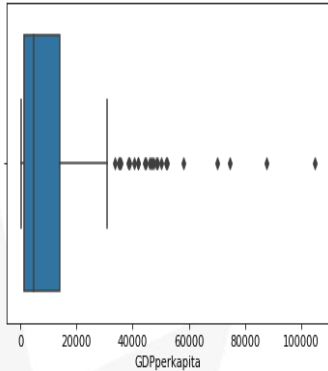
- Pada tabel disamping terdapat 2 tabel hasil Analisis Bivariate faktor sosial ekonomi dan kesehatan. Faktor sosial ekonomi membandingkan antara data GDP perkapita dengan data pendapatan. Pada tabel tersebut terlihat korelasi positif yang kuat antara GDP dan pendapatan.

- Pada tabel disamping menunjukkan hubungan antara data Kesehatan dan kematian anak. Pada data tersebut tidak menunjukkan adanya korelasi. Bahkan cenderung lemah korelasi antara Kesehatan dan Kematian Anak.

Analisis Multivariate

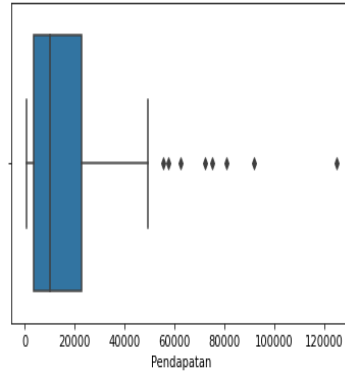


Identifikasi Data Penciran



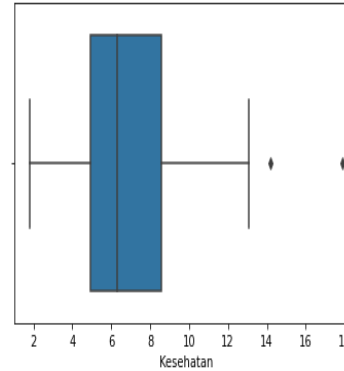
01

GDP perkapita



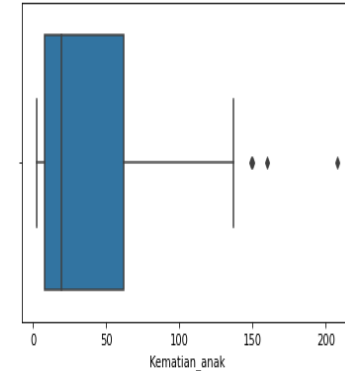
02

Pendapatan



03

Kesehatan

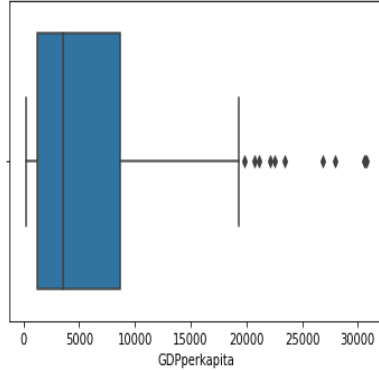


04

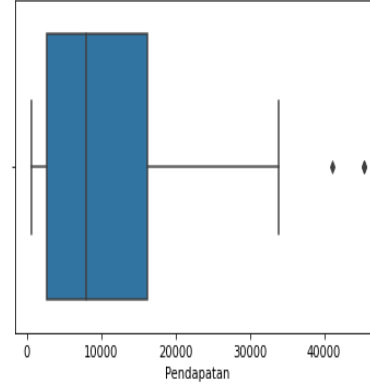
Kematian Anak

Outliers Treetment

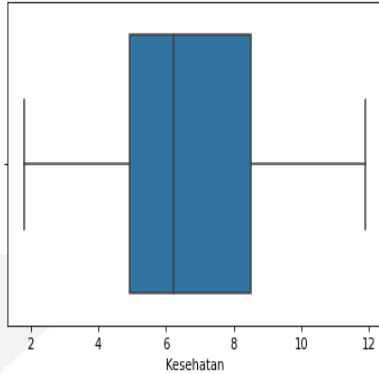
01
GDP perkapita



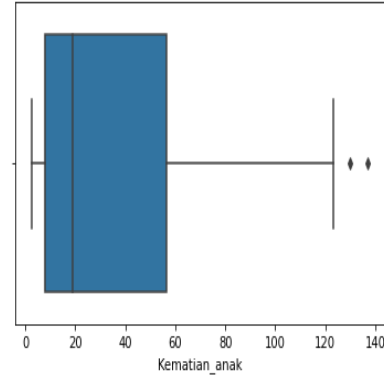
02
Pendapatan



03
Kesehatan



04
Kematian Anak



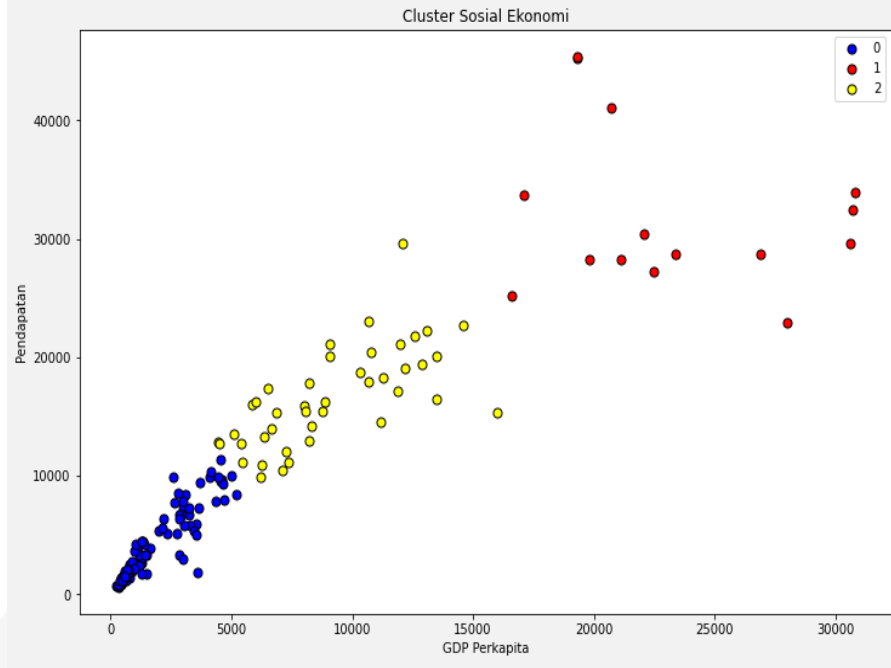
	GDPperkapita	Pendapatan	label1_kmeans
0	-0.816938	-0.945783	0
1	-0.317709	-0.101991	0
2	-0.265485	0.199219	2
3	-0.396750	-0.510703	0
4	0.826975	0.828006	2
...
137	-0.475791	-0.809884	0
138	1.010463	0.564321	2
139	-0.710091	-0.653701	0
140	-0.710091	-0.654715	0
141	-0.688919	-0.776416	0

142 rows × 3 columns

Scaling Data

	Kesehatan	Kematian_anak	label2_kmeans
0	0.352348	1.561948	2
1	-0.056882	-0.530491	1
2	-1.002483	-0.226291	1
3	-1.526934	2.380729	2
4	-0.263484	-0.709599	1
...
156	-0.573387	-0.172274	1
157	-0.708473	-0.516276	1
158	0.058338	-0.340011	1
159	-0.601199	0.598175	1
160	-0.319108	1.360096	2

161 rows × 3 columns

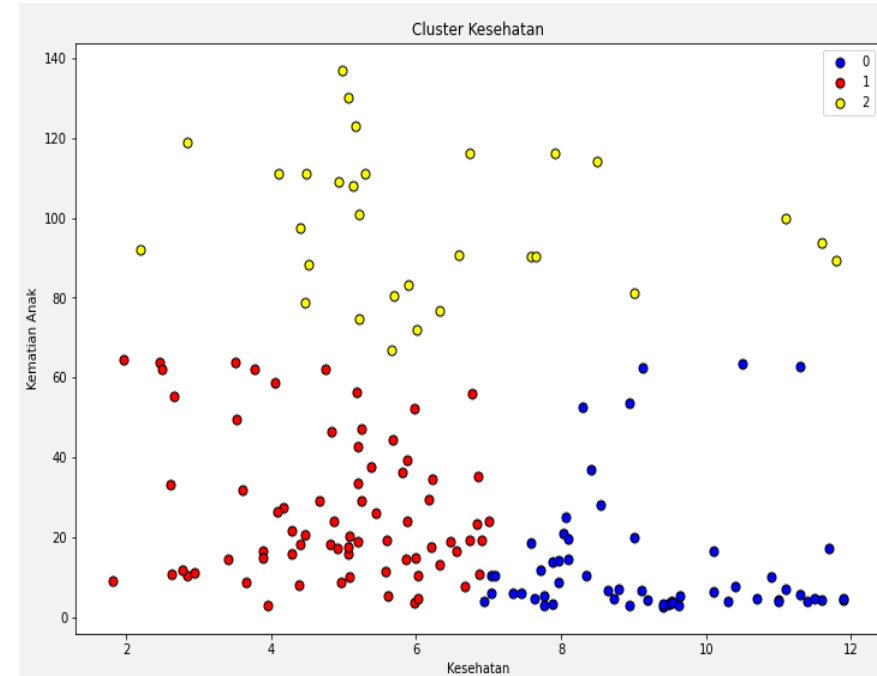


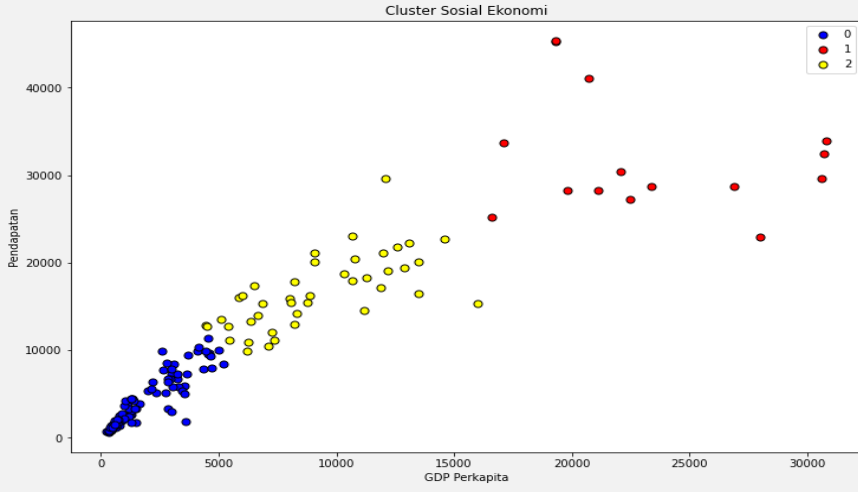
Clustering Sosial Ekonomi

Pada tabel data dibagi menjadi 3 cluster cluster dengan anotasi 0 merupakan cluster dengan negara yang memiliki GDP dan Pendapatan yang rendah

Clustering Kesehatan

Pada tabel data dibagi menjadi 3 cluster dimana anotasi 0 dan 1 merupakan negara dengan tingkat kematian yang rendah.





Rekomendasi Negara yang Mendapatkan Bantuan Dana

Berdasarkan hasil analisis dari 2 faktor yaitu Faktor sosial ekonomi dan Faktor Kesehatan, muncul nama-nama negara yang dapat menerima bantuan dana antara lain: **Rwanda dan Madagascar.** Dimana Rwanda dan Madagascar masuk dalam 10 teratas daftar negara dengan Faktor sosial ekonomi dan Faktor kesehatan terendah.

