

EXPLANATORY DATA ANALYSIS: WORLD HAPPINESS REPORT 2024

Dimas Faza Julian Putra

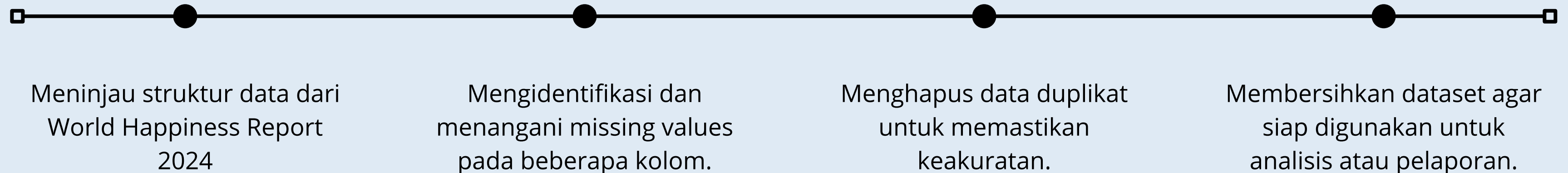
Dataset ini mencakup skor kebahagiaan dari 143 negara berdasarkan indikator ekonomi, sosial, dan kesehatan yang digunakan dalam laporan World Happiness Report 2024.



EXPLANATORY DATA ANALYSIS

World Happiness Report 2024

EDA dilakukan untuk memahami struktur dan isi dataset World Happiness Report 2024, mengidentifikasi potensi masalah dalam data seperti nilai yang hilang atau duplikat, serta menemukan insight awal dari indikator yang memengaruhi kebahagiaan di berbagai negara.



GOALS:

1

Observasi dataset World Happiness Report 2024

2

Cek data duplicate dan missing value

3

Handling duplicate dan missing value



INTRODUCTION TO DATASET

Background

World Happiness Report merupakan laporan tahunan yang mengukur tingkat kebahagiaan penduduk di berbagai negara. Laporan tahun 2024 ini menilai 143 negara berdasarkan indikator sosial, ekonomi, dan kesehatan seperti GDP per kapita, dukungan sosial, dan harapan hidup sehat.

Problem Statement

Meskipun tersedia data tahunan yang cukup lengkap, pemahaman tentang distribusi kebahagiaan antar negara dan kontribusi tiap indikator belum sepenuhnya jelas. Oleh karena itu, perlu dilakukan eksplorasi lebih lanjut untuk mengetahui faktor apa saja yang paling berpengaruh terhadap skor kebahagiaan global.



Urgency

Analisis ini penting untuk memberikan gambaran tentang kualitas hidup global, serta membantu pembuat kebijakan dan organisasi internasional dalam mengidentifikasi area yang perlu perhatian lebih, baik dari sisi ekonomi, sosial, maupun kesehatan masyarakat.

DATA OBSERVASI: DATA OVERVIEW

Dataset ini berisi informasi kebahagiaan dari 143 negara menurut World Happiness Report 2024. Negara dengan skor tertinggi seperti Finlandia, Denmark, dan Islandia ada di awal, sementara negara dengan skor terendah seperti Lebanon dan Afghanistan di akhir.

Indikator kebahagiaan mencakup GDP per kapita, dukungan sosial, harapan hidup sehat, kebebasan, kemurahan hati, dan persepsi terhadap korupsi. Negara dengan kebahagiaan tinggi umumnya dari Eropa Barat, sementara yang rendah berasal dari Sub-Sahara Afrika dan Asia Selatan.

Data akan dibersihkan dengan menangani nilai kosong dan duplikat.

```
[4] # 3. Lihat 5 baris pertama  
df.head()
```

	Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Log GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity
0	Finland	Western Europe	7.741	7.815	7.667	1.844	1.572	0.695	0.859	
1	Denmark	Western Europe	7.583	7.665	7.500	1.908	1.520	0.699	0.823	
2	Iceland	Western Europe	7.525	7.618	7.433	1.881	1.617	0.718	0.819	
3	Sweden	Western Europe	7.344	7.422	7.267	1.878	1.501	0.724	0.838	
4	Israel	Middle East and North Africa	7.341	7.405	7.277	1.803	1.513	0.740	0.641	

```
[6] # 4. Lihat 5 baris terakhir  
df.tail()
```

	Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Log GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
138	Congo (Kinshasa)	Sub-Saharan Africa	3.295	3.462	3.128	0.534	0.665	0.262	0.473	0.189	
139	Sierra Leone	Sub-Saharan Africa	3.245	3.366	3.124	0.654	0.566	0.253	0.469	0.181	
140	Lesotho	Sub-Saharan Africa	3.186	3.469	2.904	0.771	0.851	0.000	0.523	0.082	
141	Lebanon	Middle East and North Africa	2.707	2.797	2.616	1.377	0.577	0.556	0.173	0.068	
142	Afghanistan	South Asia	1.721	1.775	1.667	0.628	0.000	0.242	0.000	0.091	

UKURAN DATA

Seperti pada kode disamping, jumlah dataset World Happiness Report 2024 memiliki 143 baris dan terdapat 13 kolom.

Kode **df.shape** disamping digunakan untuk mengecek isi datasetnya, **df.shape[0]** untuk mengambil jumlah baris, sedangkan **df.shape[1]** untuk mengambil jumlah kolom pada dataset.

```
# 6. Ukuran data (jumlah baris & kolom)
print("\n=== Ukuran Data ===")
print(f"Jumlah baris: {df.shape[0]}")
print(f"Jumlah kolom: {df.shape[1]}")
```

```
=== Ukuran Data ===
Jumlah baris: 143
Jumlah kolom: 12
```

INFO DATASET

- Jumlah Data:
 - Dataset terdiri dari 143 entri (baris) dengan indeks dari 0 hingga 142.
- Terdapat 12 kolom dengan rincian sebagai berikut:
 - Country name: 143 non-null, tipe object (string/teks).
 - Regional indicator: 143 non-null, tipe object (string/teks).
 - Ladder score: 143 non-null, tipe float64.
 - upperwhisker: 143 non-null, tipe float64.
 - lowerwhisker: 143 non-null, tipe float64.
 - Log GDP per capita: 140 non-null, tipe float64 (terdapat missing values, karena hanya 140 dari 143 yang terisi).
 - Social support: 140 non-null, tipe float64 (terdapat missing values).
 - Healthy life expectancy: 140 non-null, tipe float64 (terdapat missing values).
 - Freedom to make life choices: 140 non-null, tipe float64 (terdapat missing values).
 - Generosity: 140 non-null, tipe float64 (terdapat missing values).
 - Perceptions of corruption: 140 non-null, tipe float64 (terdapat missing values).
 - Dystopia + residual: 140 non-null, tipe float64 (terdapat missing values).

```
=== Info Dataset ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 12 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Country name                        143 non-null    object
 1   Regional indicator                  143 non-null    object
 2   Ladder score                        143 non-null    float64
 3   upperwhisker                       143 non-null    float64
 4   lowerwhisker                       143 non-null    float64
 5   Log GDP per capita                  140 non-null    float64
 6   Social support                      140 non-null    float64
 7   Healthy life expectancy             140 non-null    float64
 8   Freedom to make life choices        140 non-null    float64
 9   Generosity                         140 non-null    float64
10   Perceptions of corruption           140 non-null    float64
11   Dystopia + residual                 140 non-null    float64
dtypes: float64(10), object(2)
memory usage: 13.5+ KB
```

INFO DATASET

- Missing Value
 - Kolom Log GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, Perceptions of corruption, dan Dystopia + residual memiliki missing values, dengan 3 entri kosong di setiap kolom.
- Kesimpulan Dataset berikut:
 - Dataset ini berisi informasi mengenai kebahagiaan negara berdasarkan berbagai indikator seperti GDP, dukungan sosial, harapan hidup sehat, kebebasan dalam memilih hidup, kemurahan hati, dan persepsi terhadap korupsi. Kolom dengan missing values perlu diperhatikan dan ditangani sebelum analisis lebih lanjut. Tipe data sudah sesuai, dengan sebagian besar kolom berisi data numerik dan beberapa kolom berisi data teks.

```
=== Info Dataset ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 12 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Country name                          143 non-null    object
 1   Regional indicator                    143 non-null    object
 2   Ladder score                           143 non-null    float64
 3   upperwhisker                          143 non-null    float64
 4   lowerwhisker                          143 non-null    float64
 5   Log GDP per capita                     140 non-null    float64
 6   Social support                         140 non-null    float64
 7   Healthy life expectancy                140 non-null    float64
 8   Freedom to make life choices           140 non-null    float64
 9   Generosity                            140 non-null    float64
10   Perceptions of corruption              140 non-null    float64
11   Dystopia + residual                    140 non-null    float64
dtypes: float64(10), object(2)
memory usage: 13.5+ KB
```


PROSES CEK DATA

Sebelum memuat dataset ke dalam google colabnya, langkah pertama yang harus dilakukan yaitu import library dahulu.

Kemudian gunakan kode `df = pd.read_csv()` agar data bisa diolah dan dianalisis.

```
# 1. Import library
import pandas as pd
import numpy as np
```

```
# 2. Load dataset
df = pd.read_csv('World-happiness-report-2024.csv')
```

MELIHAT TAMPILAN DATA

Kode **df.head()** atau seperti pada namanya yang merupakan kepala, yang berarti menunjukkan 5 baris pertama dari kumpulan data untuk memberikan gambaran umum data awal.

kemudian kode **df.tail()** atau kebalikan kode dari yang tadi dan sesuai namanya ekor, **df.tail()** menunjukkan 5 baris terakhir dari kumpulan data untuk memahami data di akhir.

3. Lihat 5 baris pertama
df.head()

	Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Log GDP per capita	Social support
0	Finland	Western Europe	7.741	7.815	7.667	1.844	1.572
1	Denmark	Western Europe	7.583	7.665	7.500	1.908	1.520
2	Iceland	Western Europe	7.525	7.618	7.433	1.881	1.617
3	Sweden	Western Europe	7.344	7.422	7.267	1.878	1.501
4	Israel	Middle East and North	7.341	7.405	7.277	1.803	1.513

4. Lihat 5 baris terakhir
df.tail()

	Country name	Regional indicator	Ladder score	upperwhisker	lowerwhisker	Log GDP per capita	Social support	Healthy life expectancy
138	Congo (Kinshasa)	Sub-Saharan Africa	3.295	3.462	3.128	0.534	0.665	0.262
139	Sierra Leone	Sub-Saharan Africa	3.245	3.366	3.124	0.654	0.566	0.253
140	Lesotho	Sub-Saharan Africa	3.186	3.469	2.904	0.771	0.851	0.000
141	Lebanon	Middle East and North Africa	2.707	2.797	2.616	1.377	0.577	0.556
142	Afghanistan	South Asia	1.721	1.775	1.667	0.628	0.000	0.242

MELIHAT TAMPILAN DATA

Pada kode berikut, untuk mengecek isi atau info dataset bisa menggunakan **df.info()**, bisa untuk mengetahui jumlah kolom, jumlah data non-null, dan tipe data pada setiap kolom.

Untuk dataset disamping, terdapat 143 baris dan 12 kolom, dengan beberapa kolom memiliki missing values.

```
# 5. Cek Info dataset
print("\n=== Info Dataset ===")
df.info()
```

```
=== Info Dataset ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143 entries, 0 to 142
Data columns (total 12 columns):
 #   Column                                Non-Null Count  Dtype  
---  -
 0   Country name                         143 non-null   object 
 1   Regional indicator                   143 non-null   object 
 2   Ladder score                         143 non-null   float64
 3   upperwhisker                        143 non-null   float64
 4   lowerwhisker                        143 non-null   float64
 5   Log GDP per capita                   140 non-null   float64
 6   Social support                       140 non-null   float64
 7   Healthy life expectancy              140 non-null   float64
 8   Freedom to make life choices         140 non-null   float64
 9   Generosity                           140 non-null   float64
10   Perceptions of corruption            140 non-null   float64
11   Dystopia + residual                  140 non-null   float64
dtypes: float64(10), object(2)
memory usage: 13.5+ KB
```


PENANGANAN DATA DUPLICATE

World Happiness Report 2024

```
# 10. Handling data duplicate  
print(f"Jumlah duplikat: {df.duplicated().sum()}")
```

Jumlah duplikat: 0

HANDLING

Untuk mengecek data yang duplikat, perlu menjalankan **df.isnull().sum()**

TEMUAN

Dikarenakan tidak ada data yang terduplikat, jadi tidak perlu menjalankan kode:
df = df.drop_duplicates()

PENANGANAN MISSING VALUE

```
# 7. Cek missing values
print("\n=== Missing Values ===")
print(df.isnull().sum())
```

```
=== Missing Values ===
Country name                0
Regional indicator          0
Ladder score                0
upperwhisker                0
lowerwhisker                0
Log GDP per capita          3
Social support              3
Healthy life expectancy     3
Freedom to make life choices 3
Generosity                  3
Perceptions of corruption   3
Dystopia + residual          3
dtype: int64
```

```
# 11. Handling data missing
df.dropna(inplace=True)
```

HANDLING

Untuk menghilangkan data yang null atau kosong, perlu menjalankan **df.isnull().sum()** untuk menghitung jumlah nilai kosong per kolom. Lalu untuk membersihkan data yang null, jalankan **df.dropna(inplace=True)**.

TEMUAN

Dalam dataset ini, 7 kolom dari masing masing kolom LoGDP per capita, Social support, Healthy life expectancy, dan kolom lainnya pada di gambar, memiliki 3 nilai kosong

INSIGHT DATA

Berdasarkan analisis statistik deskriptif terhadap World Happiness Report 2024:

- Rata-rata skor kebahagiaan (ladder score) adalah 5.52, dengan nilai tertinggi 7.74 dan terendah 1.72. Artinya, terdapat kesenjangan signifikan antar negara dalam hal kesejahteraan.
- Kolom Log GDP per capita, Social support, dan Healthy life expectancy memiliki 3 nilai kosong (missing), yang sudah ditangani dalam proses pembersihan.
- Beberapa kolom seperti GDP, Support, dan Freedom mengandung nilai 0, yang kemungkinan merupakan representasi dari data yang tidak tersedia.
- Kolom Dystopia + residual memiliki kontribusi cukup besar (rata-rata 1.57) dalam menyusun skor kebahagiaan di hampir semua negara.
- Secara umum, negara-negara dengan skor kebahagiaan tinggi juga memiliki skor tinggi pada GDP, dukungan sosial, dan harapan hidup sehat, meskipun belum dianalisis secara visual.

THANK YOU

Any Question?

Semoga proyek ini bisa memberi insight baru dan jadi langkah kecil menuju pemahaman data yang lebih baik.



[Github Dimas Faza Julian Putra](#)



dimasfazajulianputra@gmail.com



+62 856-2446-1018



d.faza_jp