

KELOMPOK 17

DATA SCIENCE

ANALISIS ANALISIS KASUS COVID-19 DI INDONESIA

THE TEAM



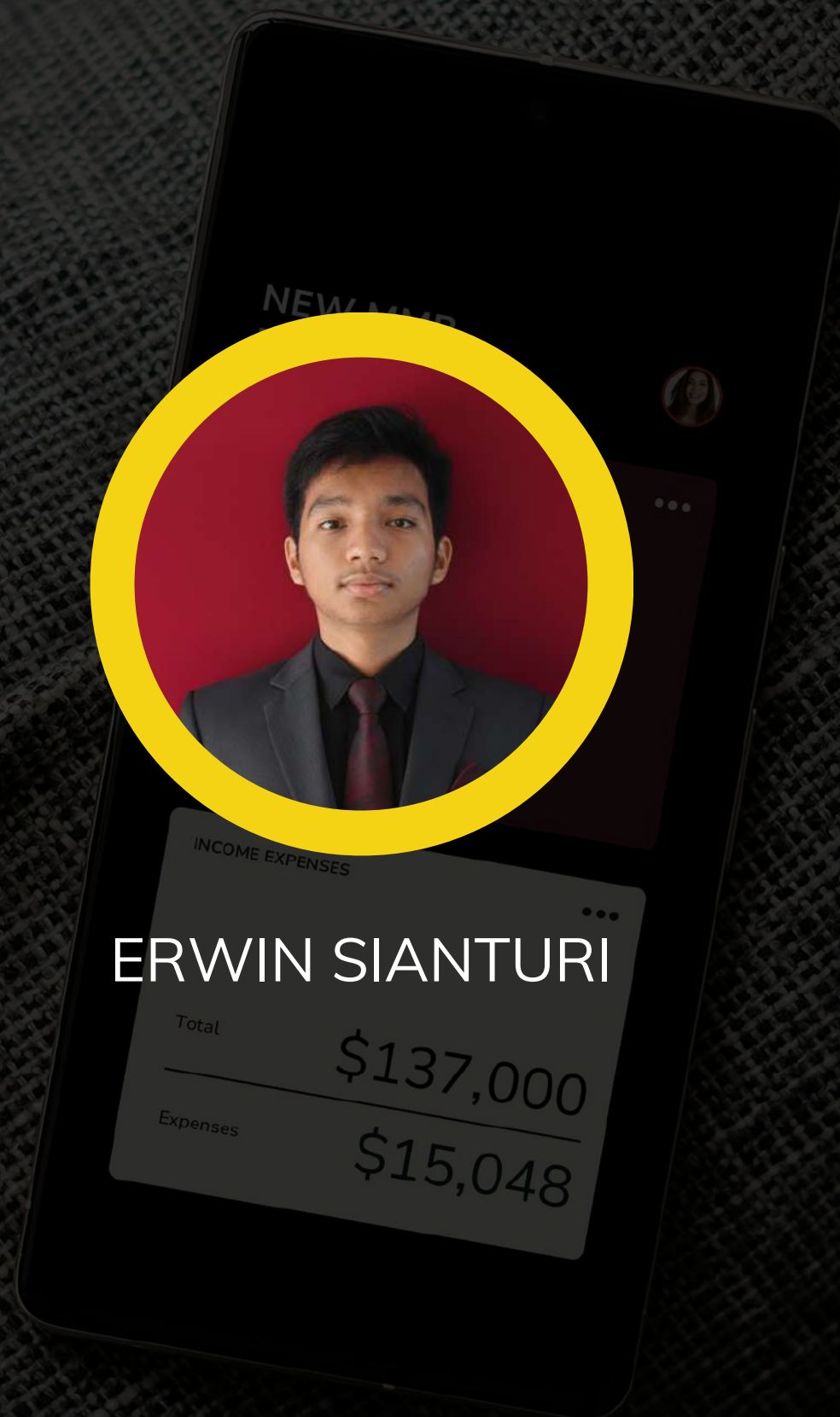
YONATHAN PURBA



ERWIN SIANTURI



JAPPY SIRAIT



LATAR BELAKANG

Analisis data deret waktu COVID-19 di Indonesia sangat penting untuk memahami dinamika penyebaran virus, mengidentifikasi pola kasus harian, kematian, kesembuhan, dan progres vaksinasi, yang semuanya berfungsi sebagai landasan pengambilan keputusan berbasis bukti bagi pemerintah. Proyek analisis eksploratif dan visualisasi ini bertujuan memberikan wawasan mendalam mengenai tren, memprediksi potensi lonjakan, menilai efektivitas intervensi kesehatan publik, dan mengukur kesiapan sistem kesehatan melalui visualisasi data yang mudah dipahami, sehingga mendukung pengambilan keputusan strategis yang cepat, adaptif, dan proaktif dalam manajemen krisis kesehatan nasional.

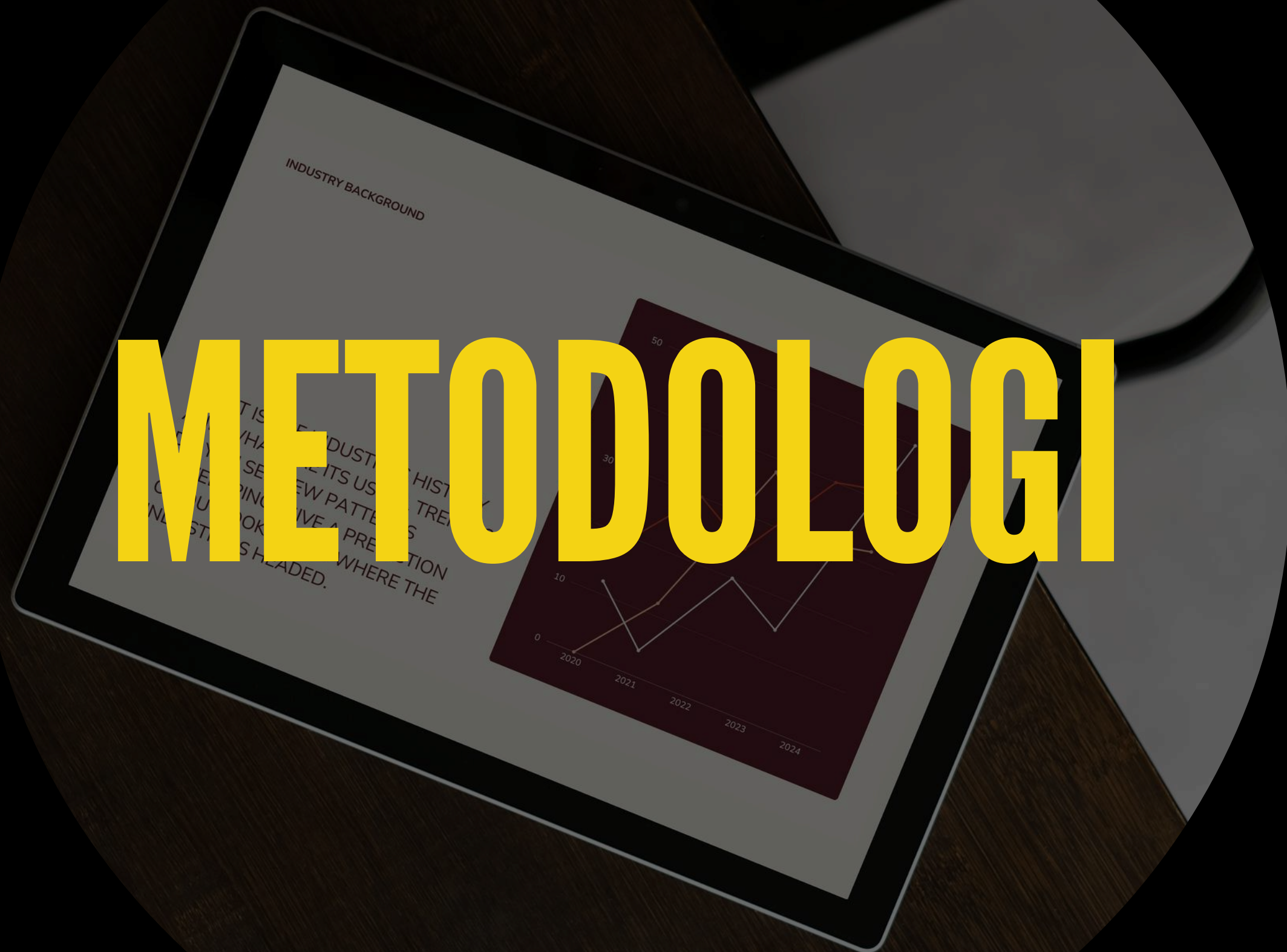


RUMUSAN MASALAH

- Bagaimana tren jumlah kasus positif COVID-19 di setiap provinsi di Indonesia dari waktu ke waktu?
- Apakah terdapat hubungan antara tingkat kematian (Case Fatality Rate) dengan tingkat kesembuhan (Case Recovered Rate) di setiap provinsi?
- Apakah terdapat perbedaan signifikan antara provinsi dengan jumlah kasus tinggi dan rendah terhadap tingkat kematian?
- Bagaimana pengaruh faktor pertumbuhan kasus baru (Growth Factor of New Cases) terhadap pertumbuhan kematian (Growth Factor of New Deaths)?
- Provinsi mana yang memiliki tingkat kesembuhan tertinggi dan tingkat kematian terendah selama periode pandemi COVID-19?



METODOLOGI



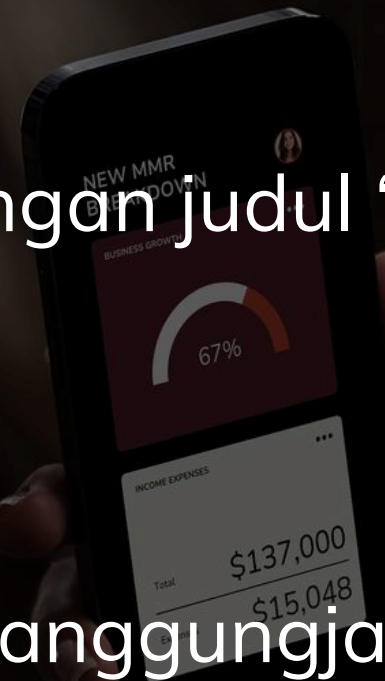
DATA COLLECTION

Dataset diperoleh dari Kaggle dengan judul “COVID-19 Indonesia Dataset”

Deskripsi singkat:

- Jumlah baris: 31.822 Baris
- Jumlah kolom: 38 Attribut

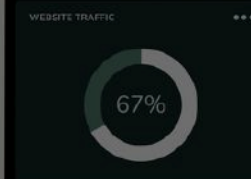
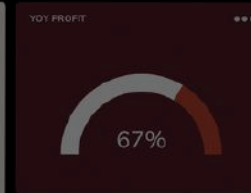
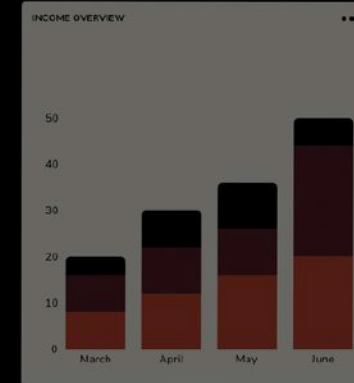
Dataset ini valid dan dapat dipertanggungjawabkan karena berasal dari sumber terbuka terpercaya (open data platform).



DATA UNDERSTANDING



FINANCIAL OUTLOOK



INCOME EXPENSES

Total	\$137,000
Expenses	\$15,048

INDUSTRY BACKGROUND

WHAT IS THE INDUSTRY'S HISTORY AND WHAT ARE ITS USUAL TRENDS? DO YOU SEE NEW PATTERNS DEVELOPING? GIVE A PREDICTION OR OUTLOOK ABOUT WHERE THE INDUSTRY IS HEADED.



STRUKTUR DAN BENTUK DATA

NO	Nama Kolom	Deskripsi
1	Date	Tanggal pencatatan data
2	Location	Nama Provinsi atau wilayah
3	New Cases	Jumlah kasus baru yang
4	New Deaths	Jumlah Kematian Baru yang
5	New Recovered	Jumlah Pasien yang dinyatakan
6	Total Cases	Jumlah Kasus hingga tanggal
7	Tota Deaths	Jumlah Kumulatif Kematian
8	Total Active Cases	Jumlah pasien yang masih
9	Total Vaccine 1	Jumlah orang yang telah
10	Total Vaccine 2	Jumlah orang yang telah
11	Population	Jumlah Penduduk di wilayah
13	Island	Pulau besar tempat provinsi

PEMERIKSAAN AWAL KUALITAS DATA

- TIPE DATA
- PEMERIKSAAN DUPLIKASI
- PEMERIKSAAN MISSING VALUES
- KONSISTENSI DATA



STATISTIK DESKRIPTIF AWAL

Statistik	New Cases	New Deaths	New Recovered	Total Cases	Total Deaths	Total Recovered
Mean	950.4	45.2	870.3	254,300	9,430	230,000
Median	620	28	590	230,000	8,200	210,000
Max	56,757	2,069	43,649	6,000,000	160,000	5,700,000
Min	0	0	0	0	0	0

INTERPRETASI

- Nilai rata-rata kasus baru harian (New Cases) menunjukkan pola fluktuatif dengan puncak signifikan selama gelombang besar pandemi (Juli 2021 & Februari 2022).
- New Deaths memiliki korelasi kuat terhadap New Cases, menandakan bahwa peningkatan kasus diikuti oleh peningkatan angka kematian dengan lag waktu tertentu.
- Jumlah kasus sembuh (New Recovered) umumnya sebanding dengan kasus baru, menandakan sistem kesehatan mampu menyesuaikan kapasitas perawatan.
- Tren Total Cases dan Total Deaths menunjukkan pertumbuhan eksponensial pada fase awal pandemi dan mulai melambat setelah program vaksinasi berjalan.

DATA CLEANING

&

ADVANCED PREPROCESSING



DATA CLEANING

```
# Filter hanya level provinsi
df = df[df["Location Level"] == "Province"]

# Hapus kolom tidak relevan
df.drop(columns=["City or Regency"], errors="ignore", inplace=True)

# Konversi kolom tanggal
df["Date"] = pd.to_datetime(df["Date"], errors="coerce")
df = df.dropna(subset=["Date"])

# Bersihkan kolom persen dan ubah jadi float
df["Case Fatality Rate"] = df["Case Fatality Rate"].astype(str).str.replace("%", "").astype(float)
df["Case Recovered Rate"] = df["Case Recovered Rate"].astype(str).str.replace("%", "").astype(float)

# Isi missing values sesuai konteks
df["Total Rural Villages"] = df["Total Rural Villages"].fillna(0)
df["Growth Factor of New Deaths"] = df["Growth Factor of New Deaths"].fillna(1.0)
df["Growth Factor of New Cases"] = df["Growth Factor of New Cases"].fillna(1.0)
```


ADVANCED PREPROCESSING

```
from sklearn.preprocessing import StandardScaler

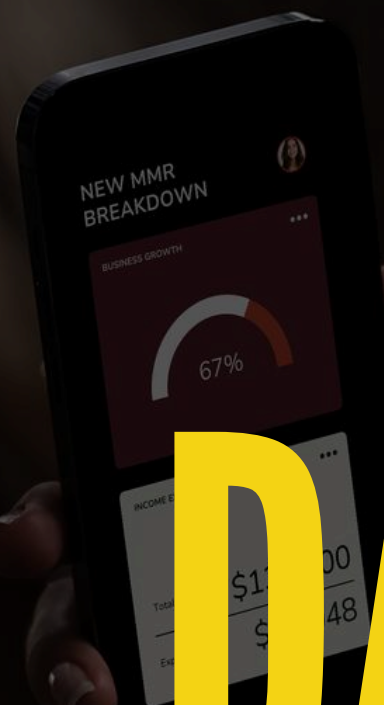
# --- Normalisasi ---
scaler = StandardScaler()
cols_to_scale = ["Total Cases", "Total Deaths", "Total Recovered", "Population Density"]
df_scaled = df.copy()
df_scaled[cols_to_scale] = scaler.fit_transform(df_scaled[cols_to_scale])

print("\n✅ Normalisasi berhasil dilakukan pada kolom numerik utama.")

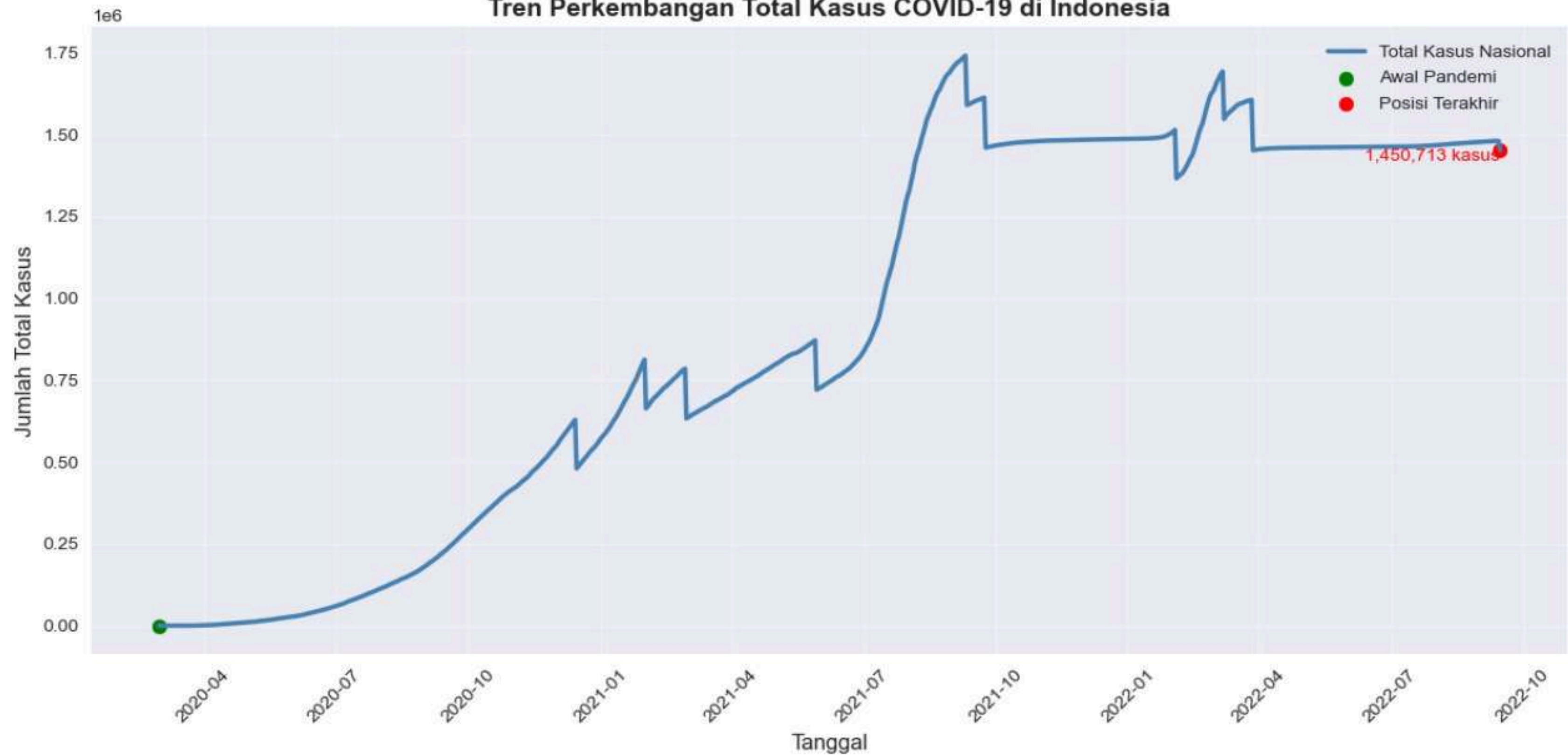
# --- Deteksi dan hapus outlier (IQR) ---
Q1 = df["Total Cases"].quantile(0.25)
Q3 = df["Total Cases"].quantile(0.75)
IQR = Q3 - Q1
lower, upper = Q1 - 1.5 * IQR, Q3 + 1.5 * IQR

df = df[(df["Total Cases"] >= lower) & (df["Total Cases"] <= upper)]
print("✅ Outlier dihapus berdasarkan IQR (Total Cases)")
```

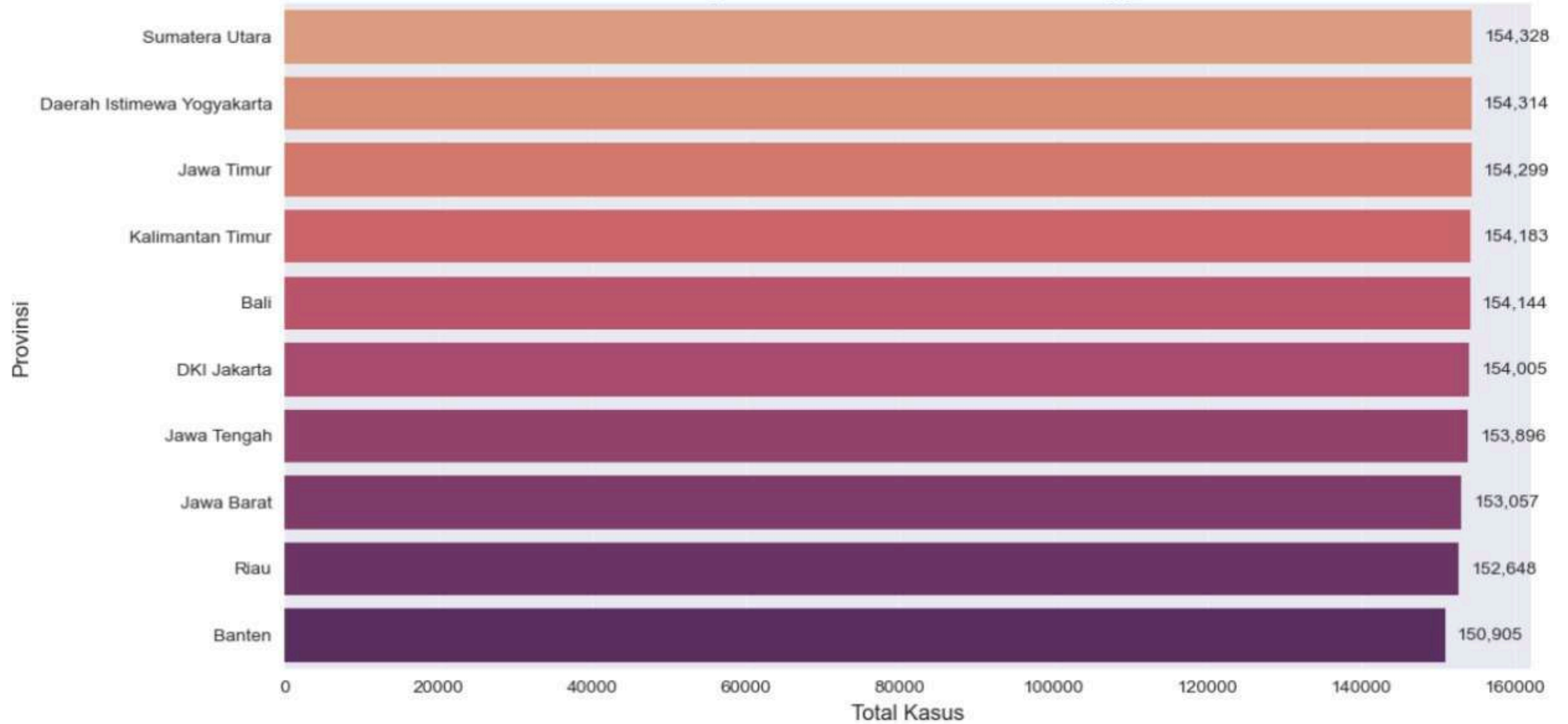

DATA VISUALIZATION

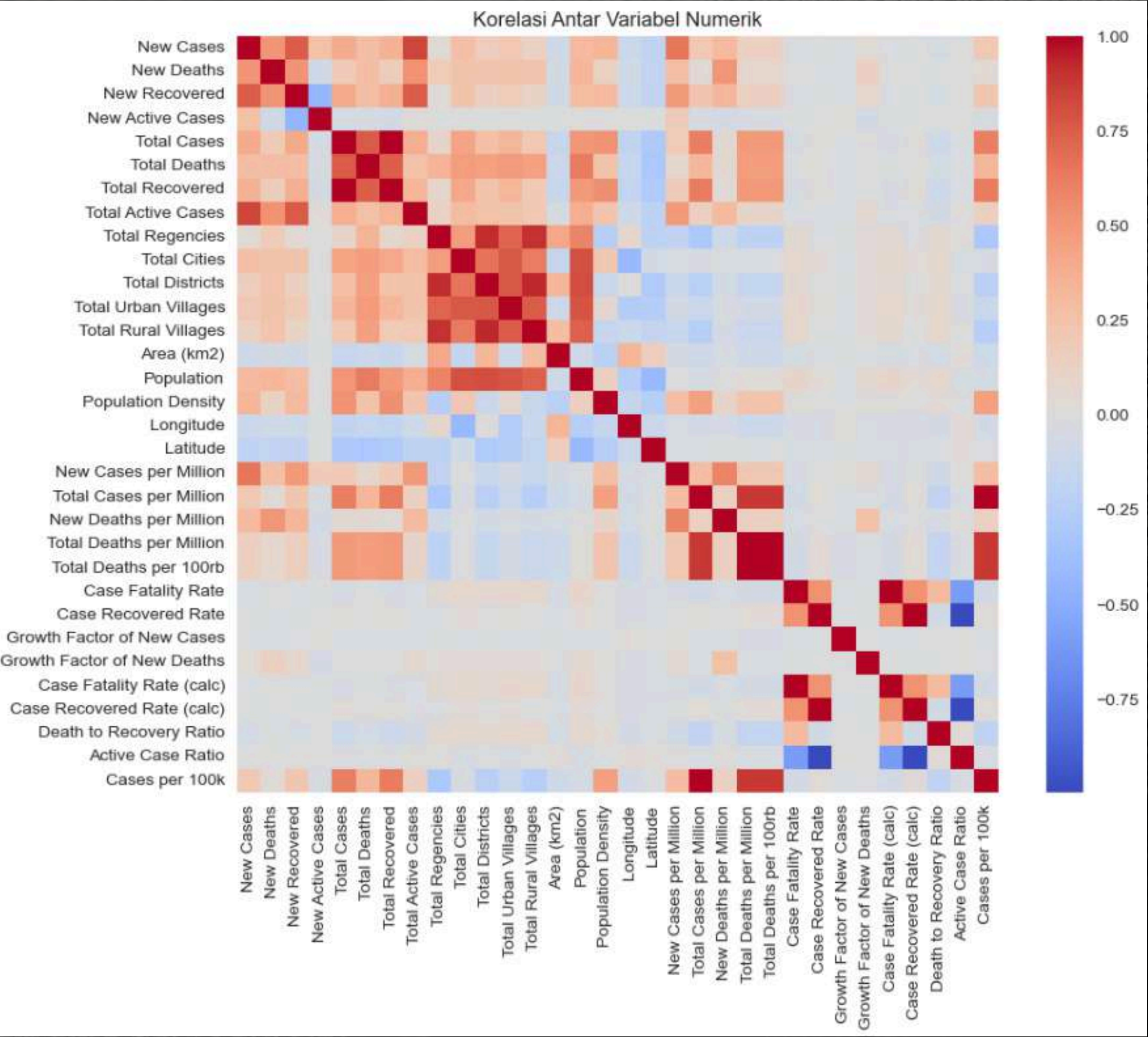


Tren Perkembangan Total Kasus COVID-19 di Indonesia

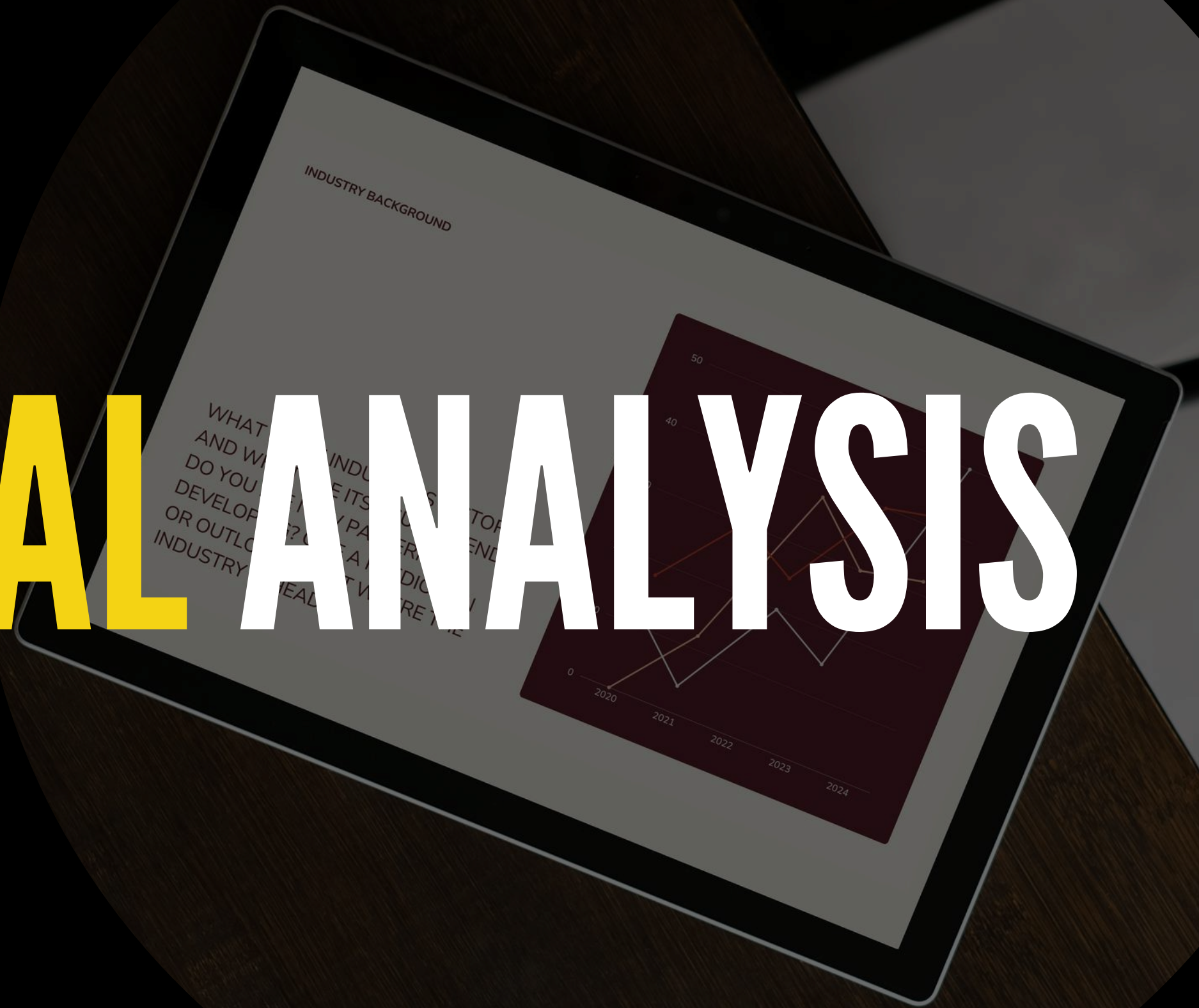
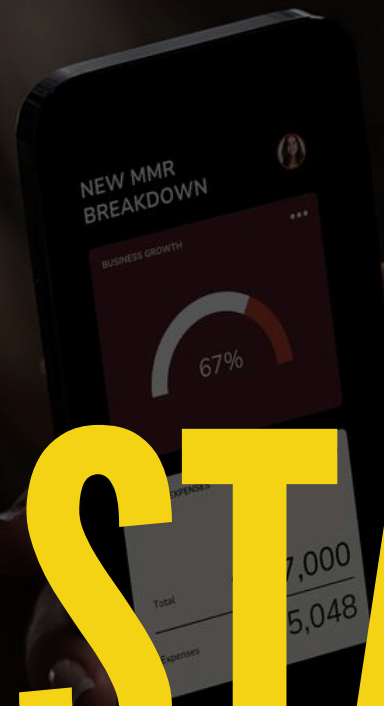


10 Provinsi dengan Total Kasus COVID-19 Tertinggi di Indonesia





STATISTICAL ANALYSIS



ANOVA (ONE-WAY ANOVA)

```
# --- ANOVA antar 3 provinsi ---
prov_list = ['DKI Jakarta', 'Jawa Barat', 'Jawa Tengah']
df_prov = df[df["Province"].isin(prov_list)]

data_jakarta = df_prov[df_prov["Province"] == "DKI Jakarta"]["Total Cases"]
data_jabar = df_prov[df_prov["Province"] == "Jawa Barat"]["Total Cases"]
data_jateng = df_prov[df_prov["Province"] == "Jawa Tengah"]["Total Cases"]

f_stat, p_val = f_oneway(data_jakarta, data_jabar, data_jateng)
print("\n=== Uji ANOVA (Total Kasus antar Provinsi) ===")
print(f"F-Statistic: {f_stat:.4f}")
print(f"P-value: {p_val:.4e}")
if p_val < 0.05:
    print("👉 Ada perbedaan signifikan jumlah kasus antar provinsi.")
else:
    print("👉 Tidak ada perbedaan signifikan jumlah kasus antar provinsi.")
```

HASIL:

F-Statistic: 8.6112

P-value: 1.9621e-04

👉 Ada perbedaan signifikan jumlah kasus antar provinsi.

MANN-WHITNEY U TEST

```
# --- Uji t-test & Mann-Whitney ---
prov1 = df[df["Province"] == "DKI Jakarta"]["New Cases"]
prov2 = df[df["Province"] == "Jawa Barat"]["New Cases"]

if prov1.empty or prov2.empty:
    print("\n⚠ Data tidak tersedia untuk provinsi tertentu.")
else:
    t_stat, p_val_t = ttest_ind(prov1, prov2, nan_policy='omit')
    print(f"\nT-test New Cases DKI Jakarta vs Jawa Barat: t={t_stat:.3f}, p={p_val_t:.3e}")

    u_stat, p_val_u = mannwhitneyu(prov1, prov2, alternative="two-sided")
    print(f"Mann-Whitney U Test: U={u_stat:.3f}, p={p_val_u:.3e}")
```

HASIL:

T-test New Cases DKI Jakarta vs Jawa Barat: $t=1.601$, $p=1.100e-01$
Mann-Whitney U Test: $U=59088.000$, $p=1.973e-06$

UJI KORELASI PEARSON DAN SPEARMAN

```
# --- Korelasi ---
cases = df["Total Cases"]
deaths = df["Total Deaths"]
recovered = df["Total Recovered"]

pearson_corr, _ = pearsonr(cases, deaths)
spearman_corr, _ = spearmanr(cases, recovered)

print(f"\nKorelasi Pearson (Kasus vs Kematian): {pearson_corr:.3f}")
print(f"Korelasi Spearman (Kasus vs Sembuh): {spearman_corr:.3f}")
```

HASIL:

```
Korelasi Pearson (Kasus vs Kematian): 0.850
Korelasi Spearman (Kasus vs Sembuh): 0.996
```


HASIL &

PEMBAHASAN



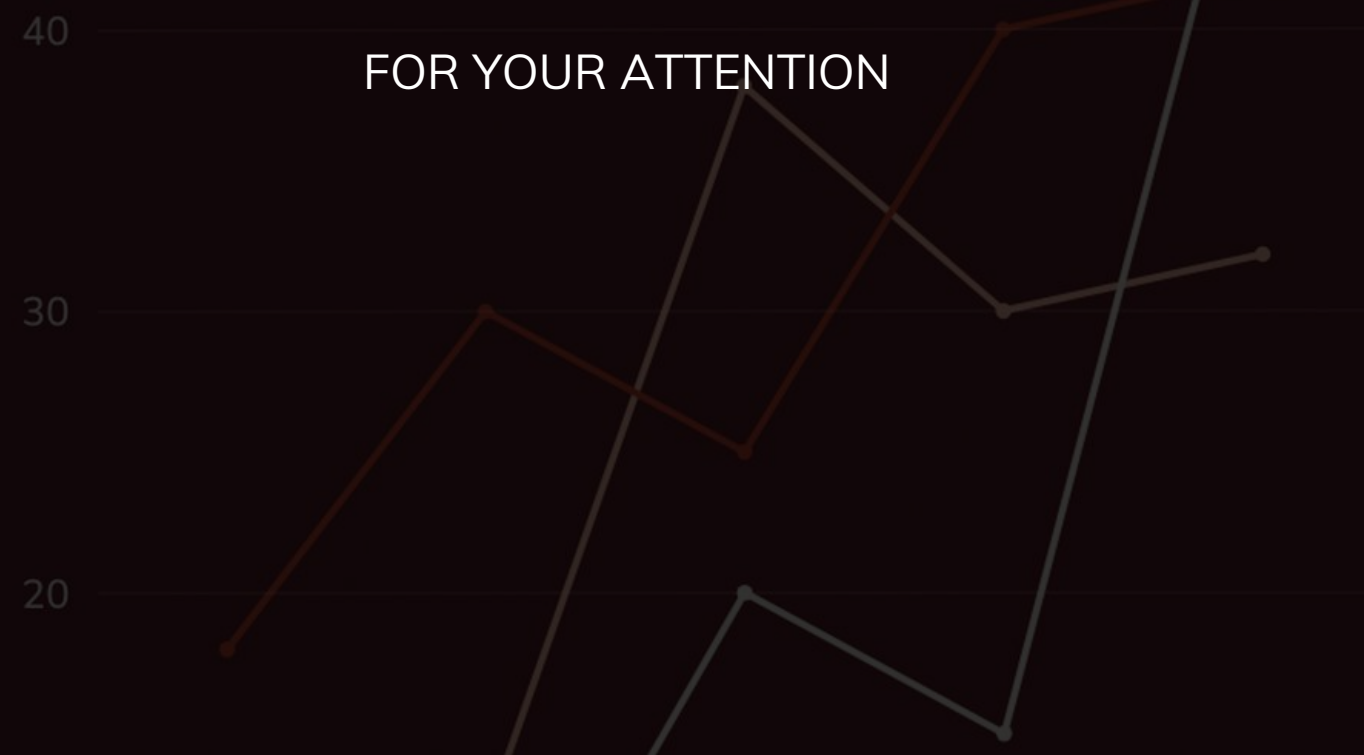
KESIMPULAN



WE WANT TO SAY

THANK YOU

FOR YOUR ATTENTION



GROUND

THE INDUSTRY'S HISTORY