

PERTEMUAN 3

Seputar Data and Website Analytics





PENGOLAHAN DATA



Pengolahan data adalah proses mengumpulkan, menyusun, menganalisis, menyimpan, dan menginterpretasikan data untuk menghasilkan informasi yang berguna. Proses ini dapat dilakukan secara manual atau menggunakan teknologi komputer.

TAHAPAN PENGOLAHAN DATA

Pengumpulan Data 1

Mengumpulkan data dari berbagai sumber seperti survei, sensor, transaksi, atau database.



Pembersihan Data 2

Memeriksa dan membersihkan data dari kesalahan, duplikasi, atau nilai yang tidak valid.



Pengorganisasian Data 3

Menyusun data dalam format yang sesuai, seperti tabel atau basis data.



Analisis Data 4

Mengolah dan menganalisis data untuk menemukan pola, tren, atau informasi yang berguna.



Penyimpanan Data 7

Menyimpan data dalam database atau sistem penyimpanan untuk digunakan kembali di masa depan.



Penyajian Data 6

Menyajikan data dalam bentuk laporan, grafik, atau visualisasi untuk memudahkan pemahaman.



Interpretasi Data 5

Mengubah hasil analisis menjadi wawasan yang dapat digunakan untuk pengambilan keputusan.





JENIS PENGOLAHAN DATA

Pengolahan Manual

#1

Menggunakan tenaga manusia tanpa bantuan alat elektronik.



Pengolahan Mekanis

#2

Menggunakan alat sederhana seperti kalkulator atau mesin hitung.



Pengolahan Elektronik

#3

Menggunakan komputer dan perangkat lunak seperti Excel, SQL, atau Python untuk mengolah data secara efisien.



CONTOH PENGOLAHAN DATA



BISNIS

Menganalisis data penjualan untuk menentukan strategi pemasaran.



PENDIDIKAN

Mengolah nilai mahasiswa untuk menghasilkan laporan akademik.



KESEHATAN

Menganalisis data pasien untuk meningkatkan layanan medis.

PENGOLAHAN DATA

Pengolahan data sangat penting dalam berbagai bidang karena membantu pengambilan keputusan yang lebih baik berdasarkan fakta dan analisis.






MANFAAT PENGOLAHAN DATA #1

Automasi Proses Bisnis

#1

Mengurangi pekerjaan manual dengan sistem otomatis seperti chatbot, sistem ERP (Enterprise Resource Planning), dan CRM (Customer Relationship Management)

 Mempercepat pemrosesan transaksi dan administrasi, seperti otomatisasi penggajian atau pencatatan stok barang.



Prediksi dan Tren Masa Depan

#2

Dengan teknik machine learning dan big data analytics, pengolahan data bisa membantu memprediksi tren pasar, kebutuhan pelanggan, atau kondisi ekonomi.



Perusahaan e-commerce menganalisis data pembelian untuk merekomendasikan produk kepada pelanggan.





MANFAAT PENGOLAHAN DATA #2



Pengambilan Keputusan yang Lebih Akurat

#3

Data yang diolah dengan baik mengurangi subjektivitas dalam pengambilan keputusan.

-  Rumah sakit menggunakan data pasien untuk menentukan metode perawatan yang paling efektif.



Personalisasi Layanan

#4

Dengan menganalisis preferensi pengguna, perusahaan bisa memberikan layanan yang lebih personal.



- Netflix dan Spotify merekomendasikan konten berdasarkan riwayat penggunaan.





MANFAAT PENGOLAHAN DATA #3

Efisiensi Operasional

#5

Mengoptimalkan penggunaan sumber daya, baik tenaga kerja, waktu, maupun biaya.



Perusahaan logistik menggunakan data lalu lintas untuk menentukan rute tercepat dan menghemat bahan bakar.



Deteksi Kecurangan dan Keamanan Data

#6

Analisis data dapat digunakan untuk mendeteksi aktivitas mencurigakan dalam transaksi keuangan atau sistem keamanan



Bank menggunakan data untuk mendeteksi transaksi mencurigakan dan mencegah penipuan kartu kredit.





MANFAAT PENGOLAHAN DATA #4



Inovasi dan Pengembangan Produk

#7

Perusahaan dapat mengembangkan produk baru berdasarkan kebutuhan pasar yang teridentifikasi melalui analisis data.



Produsen smartphone menganalisis tren penggunaan fitur tertentu untuk merancang model baru..



Pengelolaan Risiko

#8

Dalam bidang keuangan dan asuransi, pengolahan data digunakan untuk menganalisis risiko dan menentukan kebijakan yang tepat.



Bank mengevaluasi riwayat kredit nasabah sebelum menyetujui pinjaman.





MANFAAT PENGOLAHAN DATA #5

Pengukuran Kinerja dan Evaluasi

#9

Data dapat digunakan untuk mengukur efektivitas strategi yang telah diterapkan.



Perusahaan menilai keberhasilan kampanye pemasaran berdasarkan data penjualan dan interaksi pelanggan.



Mendukung Penelitian dan Pengembangan

#10

Dalam dunia akademik dan ilmiah, pengolahan data membantu dalam riset yang lebih akurat dan mendalam.



Peneliti menganalisis data cuaca untuk memprediksi perubahan iklim..



PHYTON

bahasa pemrograman tingkat tinggi yang populer karena sintaksnya yang sederhana dan mudah dipahami. Python banyak digunakan untuk berbagai keperluan, seperti pengembangan web, analisis data, kecerdasan buatan (AI), otomatisasi, dan lainnya.

CONTOH DATA KEPENDUDUKAN

NO	WILAYAH	2021	2022	2023
1	Teluk Betung Barat	41761	42642	38527
2	Teluk Betung Timur	54746	55901	49926
3	Teluk Betung Selatan	43564	44483	39359
4	Teluk Betung Utara	54419	55567	50587



PANDAS

pandas untuk mengolah data ini dan menentukan pemeringkatan berdasarkan jumlah penduduk.



Membuat DataFrame dari data yang diberikan.



Mengurutkan berdasarkan jumlah penduduk secara descending (dari yang terbesar ke terkecil).

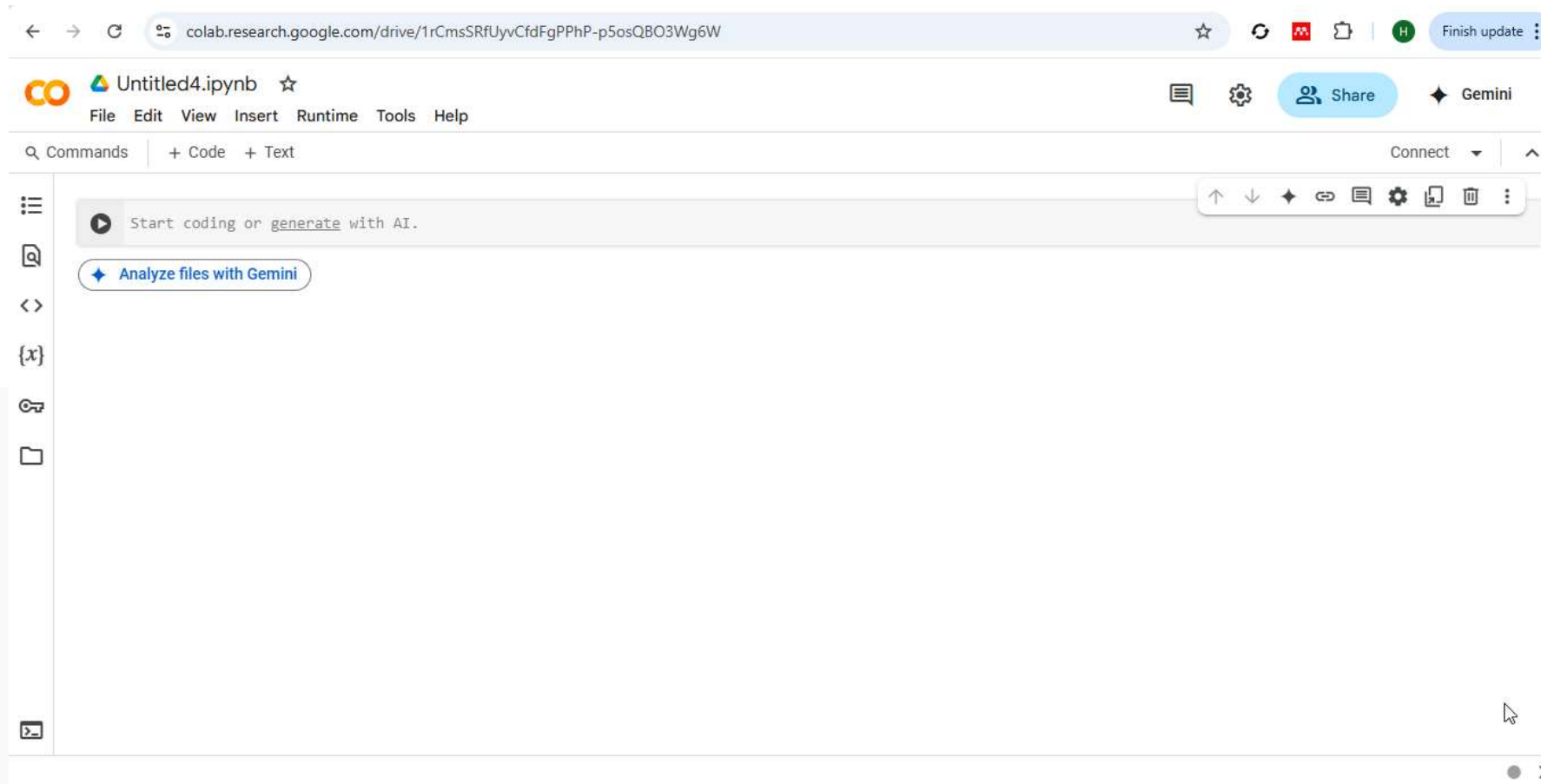


Menambahkan kolom peringkat berdasarkan urutan jumlah penduduk.



KODE PROGRAM DAN HASIL

<https://colab.research.google.com/>



KODE PROGRAM DAN HASIL

```
!pip install pandas numpy scikit-learn
```

```
[2] import pandas as pd  
import numpy as np  
import sklearn  
  
print("Semua library berhasil diimpor!")
```

```
→ Semua library berhasil diimpor!
```

KODE PROGRAM DAN HASIL

MASUKAN DATA YANG SEBELUMNYA KE DALAM PROGRAM

```
[3] import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

# Data penduduk dari 2021-2023
data = {
    "Wilayah": ["Teluk Betung Barat", "Teluk Betung Timur", "Teluk Betung Selatan", "Teluk Betung Utara"],
    2021: [41761, 54746, 43564, 54419],
    2022: [42642, 55901, 44483, 55567],
    2023: [38527, 49926, 39359, 50587],
}

# Konversi ke DataFrame
df = pd.DataFrame(data)
print(df)
```

```
Wilayah  2021  2022  2023
0  Teluk Betung Barat  41761  42642  38527
1  Teluk Betung Timur  54746  55901  49926
2  Teluk Betung Selatan  43564  44483  39359
3  Teluk Betung Utara  54419  55567  50587
```

KODE PROGRAM DAN HASIL

PENENTUAN PERINGKAT BERDASARKAN JUMLAH TERBESAR TAHUN TERAKHIR #1

```
[8] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Data penduduk dari 2021-2023
data = {
    "Wilayah": ["Teluk Betung Barat", "Teluk Betung Timur", "Teluk Betung Selatan", "Teluk Betung Utara"],
    2021: [41761, 54746, 43564, 54419],
    2022: [42642, 55901, 44483, 55567],
    2023: [38527, 49926, 39359, 50587],
}

# Konversi ke DataFrame
df = pd.DataFrame(data)

# Menentukan peringkat berdasarkan jumlah penduduk tahun 2023
df["Peringkat"] = df[2023].rank(ascending=False, method="min").astype(int)

# Mengurutkan berdasarkan peringkat
df = df.sort_values(by="Peringkat").reset_index(drop=True)
```

KODE PROGRAM DAN HASIL

PENENTUAN PERINGKAT BERDASARKAN JUMLAH TERBESAR TAHUN TERAKHIR #2

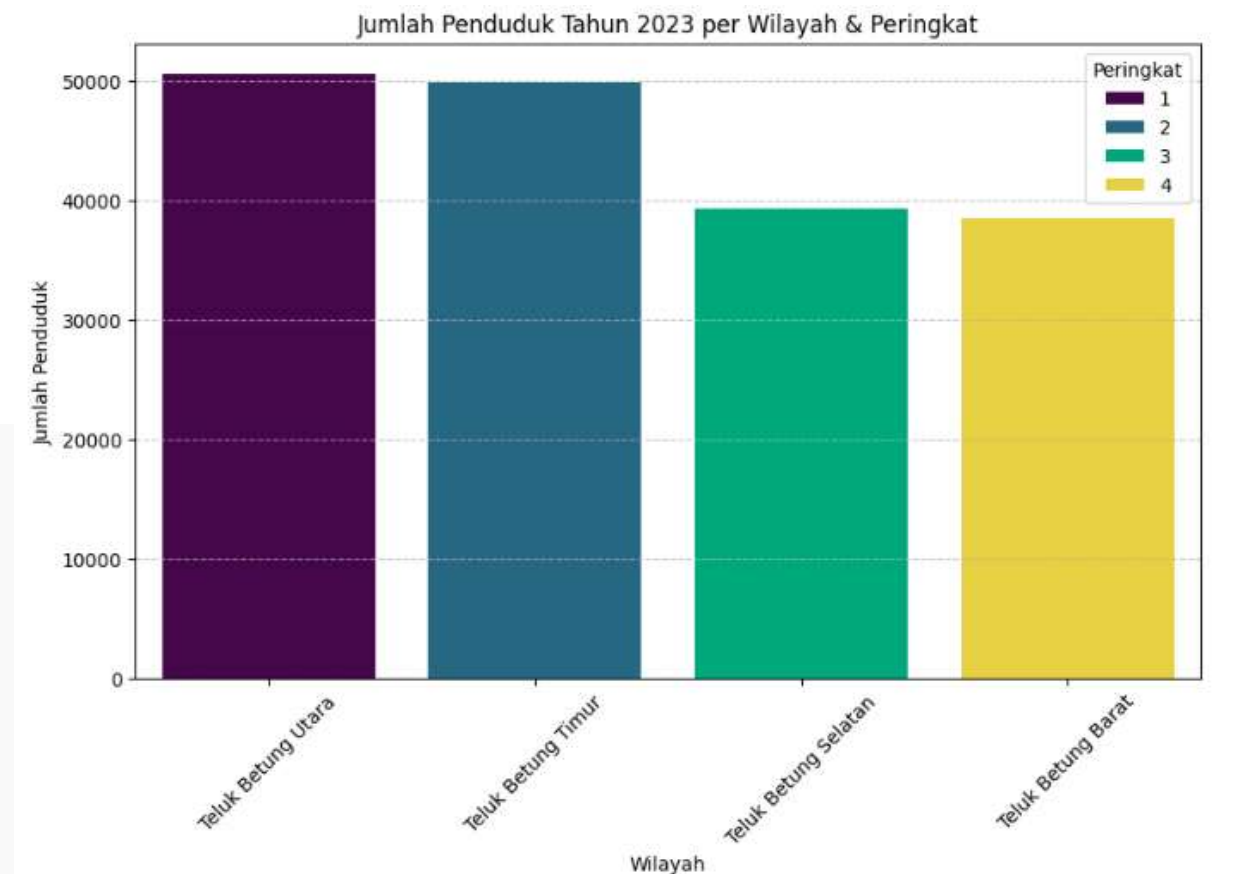
```
# Menampilkan tabel akhir dengan total penduduk per wilayah dan peringkatnya
print("\n🏆 **Tabel Wilayah dengan Jumlah Penduduk & Peringkat Tahun 2023:**")
print(df)

# Grafik Bar untuk Perbandingan Jumlah Penduduk Tahun 2023
plt.figure(figsize=(10, 6))
sns.barplot(x=df["Wilayah"], y=df[2023], hue=df["Peringkat"], palette="viridis")
plt.xlabel("Wilayah")
plt.ylabel("Jumlah Penduduk")
plt.title("Jumlah Penduduk Tahun 2023 per Wilayah & Peringkat")
plt.xticks(rotation=45)
plt.legend(title="Peringkat")
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.show()
```

```
🏆 **Tabel Wilayah dengan Jumlah Penduduk & Peringkat Tahun 2023:**
   Wilayah  2021  2022  2023  Peringkat
0  Teluk Betung Utara  54419  55567  50587         1
1  Teluk Betung Timur  54746  55901  49926         2
2  Teluk Betung Selatan  43564  44483  39359         3
3  Teluk Betung Barat  41761  42642  38527         4
```

KODE PROGRAM DAN HASIL

GRAFIK PENENTUAN PERINGKAT BERDASARKAN JUMLAH TERBESAR TAHUN TERAKHIR #3



KODE PROGRAM DAN HASIL

PROGRAM PENDUDUK PERWILAYAH DAN PERTAHUN #1

```
[10] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Data penduduk dari 2021-2023
data = {
    "Wilayah": ["Teluk Betung Barat", "Teluk Betung Timur", "Teluk Betung Selatan", "Teluk Betung Utara"],
    2021: [41761, 54746, 43564, 54419],
    2022: [42642, 55901, 44483, 55567],
    2023: [38527, 49926, 39359, 50587],
}

# Konversi ke DataFrame
df = pd.DataFrame(data)

# Ubah format DataFrame agar cocok untuk visualisasi
df_melted = df.melt(id_vars=["Wilayah"], var_name="Tahun", value_name="Jumlah Penduduk")

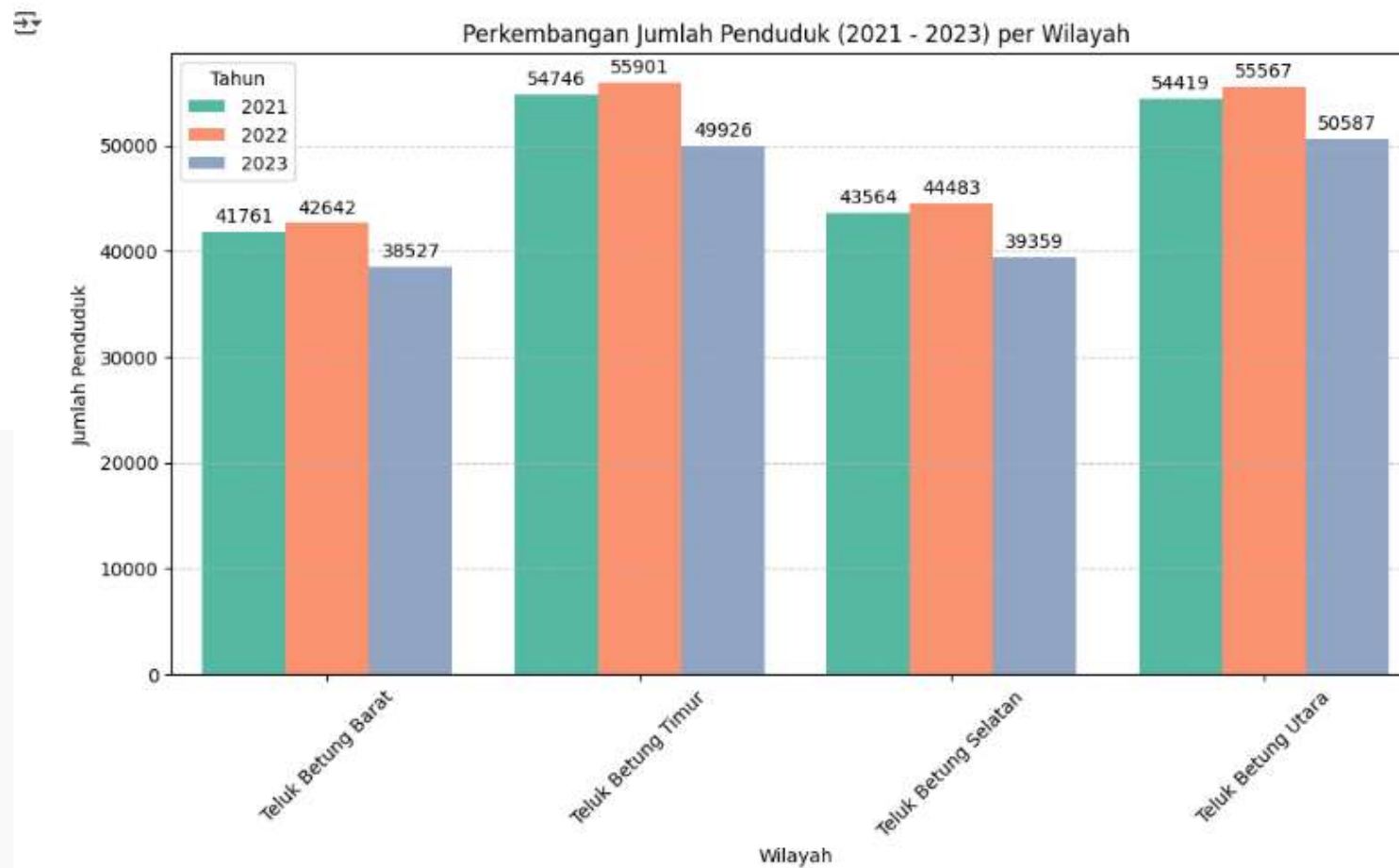
# Grafik Bar untuk Perkembangan Jumlah Penduduk per Tahun & Wilayah
plt.figure(figsize=(12, 6))
sns.barplot(x="Wilayah", y="Jumlah Penduduk", hue="Tahun", data=df_melted, palette="Set2")

# Menambahkan label pada batang
for container in plt.gca().containers:
    plt.bar_label(container, fmt='%d', fontsize=10, padding=3)

# Konfigurasi tampilan grafik
plt.xlabel("Wilayah")
plt.ylabel("Jumlah Penduduk")
plt.title("Perkembangan Jumlah Penduduk (2021 - 2023) per Wilayah")
plt.xticks(rotation=45)
plt.legend(title="Tahun")
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.show()
```

KODE PROGRAM DAN HASIL

GRAFIK PENDUDUK PERWILAYAH DAN PERTAHUN #2



KODE PROGRAM DAN HASIL

PRESENTASE PENDUDUK PERWILAYAH DAN PERTAHUN #1

```
[11] import pandas as pd

# Data penduduk dari 2021-2023
data = {
    "Wilayah": ["Teluk Betung Barat", "Teluk Betung Timur", "Teluk Betung Selatan", "Teluk Betung Utara"],
    "2021": [41761, 54746, 43564, 54419],
    "2022": [42642, 55901, 44483, 55567],
    "2023": [38527, 49926, 39359, 50587],
}

# Konversi ke DataFrame
df = pd.DataFrame(data)

# Hitung total penduduk tahun 2023
total_penduduk_2023 = df[2023].sum()

# Menentukan peringkat berdasarkan jumlah penduduk tahun 2023
df["Peringkat"] = df[2023].rank(ascending=False, method="min").astype(int)

# Hitung persentase jumlah penduduk tahun 2023
df["Persentase (%)"] = (df[2023] / total_penduduk_2023) * 100

# Urutkan berdasarkan peringkat
df = df.sort_values(by="Peringkat").reset_index(drop=True)

# Tampilkan tabel
print("\n👉 **Tabel Peringkat & Persentase Penduduk Tahun 2023:**")
print(df)
```

KODE PROGRAM DAN HASIL

PRESENTASE PENDUDUK PERWILAYAH DAN PERTAHUN #2



****Tabel Peringkat & Persentase Penduduk Tahun 2023:****

	Wilayah	2021	2022	2023	Peringkat	Persentase (%)
0	Teluk Betung Utara	54419	55567	50587	1	28.356101
1	Teluk Betung Timur	54746	55901	49926	2	27.985583
2	Teluk Betung Selatan	43564	44483	39359	3	22.062343
3	Teluk Betung Barat	41761	42642	38527	4	21.595973

KODE PROGRAM DAN HASIL

MENGHITUNG RATA-RATA, TERBESAR DAN TERKECIL PENDUDUK PERWILAYAH #1

```
# Hitung Mean, Min, dan Max untuk setiap tahun
summary_df = pd.DataFrame({
    "Statistik": ["Mean", "Min", "Max"],
    2021: [df[2021].mean(), df[2021].min(), df[2021].max()],
    2022: [df[2022].mean(), df[2022].min(), df[2022].max()],
    2023: [df[2023].mean(), df[2023].min(), df[2023].max()]
})

# Tampilkan tabel statistik
print("\n📊 **Tabel Mean, Min & Max Penduduk per Tahun:**")
print(summary_df)
```

```
📊 **Tabel Mean, Min & Max Penduduk per Tahun:**
  Statistik  2021    2022    2023
0      Mean  48622.5  49648.25  44599.75
1       Min  41761.0  42642.00  38527.00
2       Max  54746.0  55901.00  50587.00
```

KODE PROGRAM DAN HASIL

PREDIKSI JUMLAH PENDUDUK PERWILAYAH #1 (MASUKAN DATA AWAL PENDUDUK)

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

# Data penduduk dari 2021-2023
data = {
    "Wilayah": ["Teluk Betung Barat", "Teluk Betung Timur", "Teluk Betung Selatan", "Teluk Betung Utara"],
    "2021": [41761, 54746, 43564, 54419],
    "2022": [42642, 55901, 44483, 55567],
    "2023": [38527, 49926, 39359, 50587],
}

# Konversi ke DataFrame
df = pd.DataFrame(data)
print(df)
```

	Wilayah	2021	2022	2023
0	Teluk Betung Barat	41761	42642	38527
1	Teluk Betung Timur	54746	55901	49926
2	Teluk Betung Selatan	43564	44483	39359
3	Teluk Betung Utara	54419	55567	50587

KODE PROGRAM DAN HASIL

TAHUN PREDIKSI JUMLAH PENDUDUK PERWILAYAH #2

```
# Tahun yang tersedia
tahun = np.array([2021, 2022, 2023]).reshape(-1, 1)
tahun_prediksi = np.array([2024, 2025, 2026]).reshape(-1, 1)

# Menyimpan hasil prediksi
prediksi_penduduk = {}

# Loop untuk setiap wilayah
for i, wilayah in enumerate(df["Wilayah"]):
    # Data jumlah penduduk dari tahun 2021-2023
    penduduk = np.array(df.iloc[i, 1:]).reshape(-1, 1)

    # Model regresi linier
    model = LinearRegression()
    model.fit(tahun, penduduk)

    # Prediksi untuk 2024-2026
    prediksi = model.predict(tahun_prediksi)

    # Simpan hasil
    prediksi_penduduk[wilayah] = prediksi.flatten().astype(int)

# Konversi ke DataFrame hasil prediksi
df_prediksi = pd.DataFrame(prediksi_penduduk, index=[2024, 2025, 2026])
print("\nPrediksi Jumlah Penduduk (2024-2026):")
print(df_prediksi)
```

KODE PROGRAM DAN HASIL

HASIL PREDIKSI JUMLAH PENDUDUK PERWILAYAH #2

Prediksi Jumlah Penduduk (2024-2026):

	Teluk Betung Barat	Teluk Betung Timur	Teluk Betung Selatan	\
2024	37742	48704	38263	
2025	36125	46294	36161	
2026	34508	43884	34058	
	Teluk Betung Utara			
2024	49692			
2025	47776			
2026	45860			

KODE PROGRAM DAN HASIL

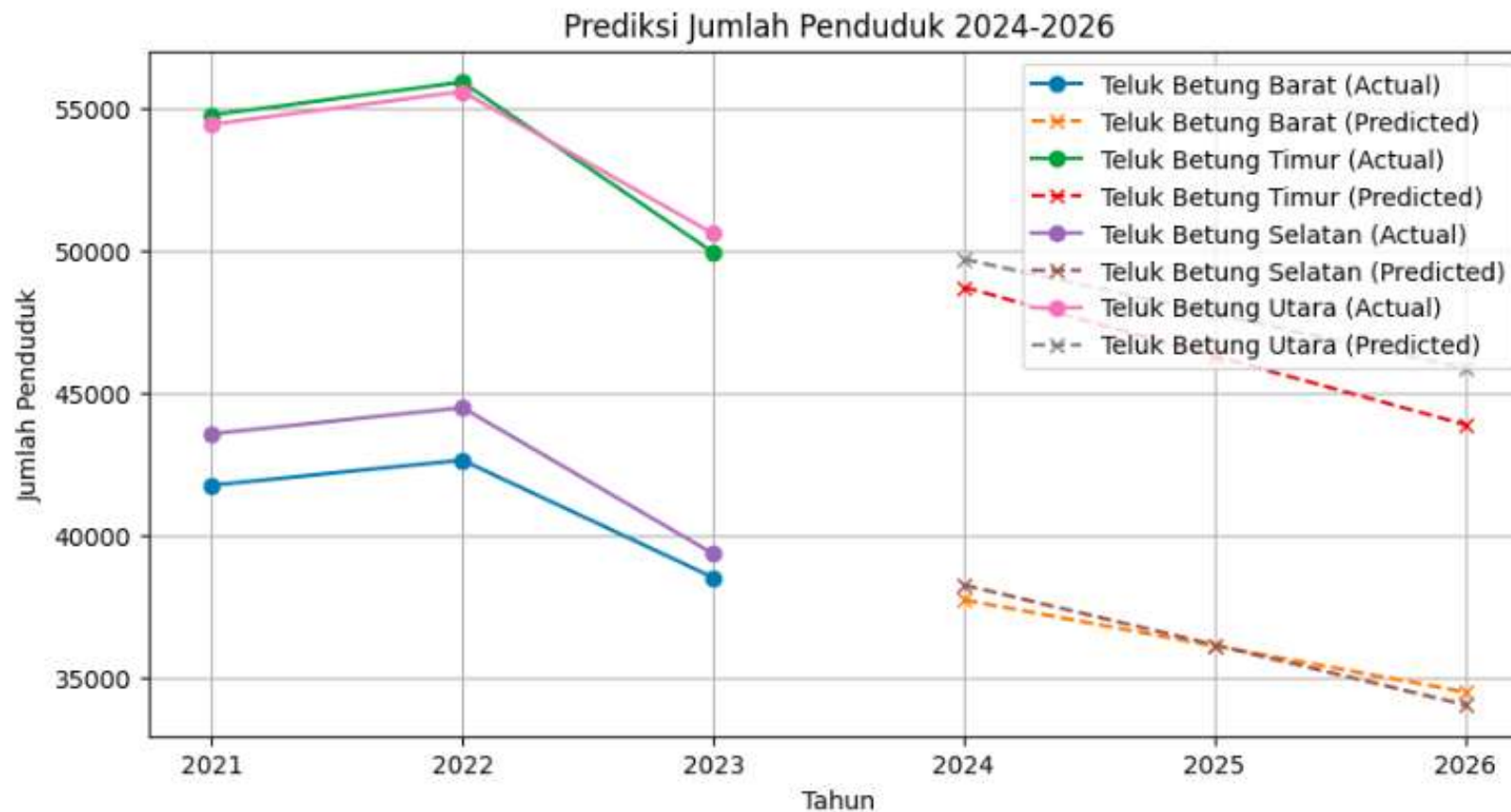
GRAFIK HASIL PREDIKSI JUMLAH PENDUDUK PERWILAYAH #2

```
# Plot hasil prediksi
plt.figure(figsize=(10, 5))
for wilayah in df["Wilayah"]:
    plt.plot([2021, 2022, 2023], df[df["Wilayah"] == wilayah].iloc[:, 1:].values.flatten(), marker='o', label=f"{wilayah} (Actual)")
    plt.plot([2024, 2025, 2026], df_prediksi[wilayah].values, marker='x', linestyle="dashed", label=f"{wilayah} (Predicted)")

plt.xlabel("Tahun")
plt.ylabel("Jumlah Penduduk")
plt.title("Prediksi Jumlah Penduduk 2024-2026")
plt.legend()
plt.grid(True)
plt.show()
```

KODE PROGRAM DAN HASIL

GRAFIK HASIL PREDIKSI JUMLAH PENDUDUK PERWILAYAH #2





**SEKIAN
TERIMAKASIH**

Tugas LIBURAN

Tugas Liburan / KP Tanggal 4 April 2025 :

Berdasarkan data minggu lalu, buatlah pengolahan data dengan phyton untuk :

1. Grafik data yang sudah dibuat
2. Peringkat dan persentasenya
3. Nilai rata-rata, nilai terbesar, nilai terkecil.
4. Predisi untuk 3 tahun kedepan beserta grafiknya

Tugas dikumpulkan dengan file .doc dan terakhir hari Sabtu, 4 April 2025 23.00.00 WIB di LMS
Gunakan chatGPT atau AI lainnya. Kode program di copy dan bukan di screenshot dan berikan nomor urut seperti permintaan diatas.

Terimakasih