Aldy Imam Pratama

**Suicide Rates Overview 1985 to 2016**

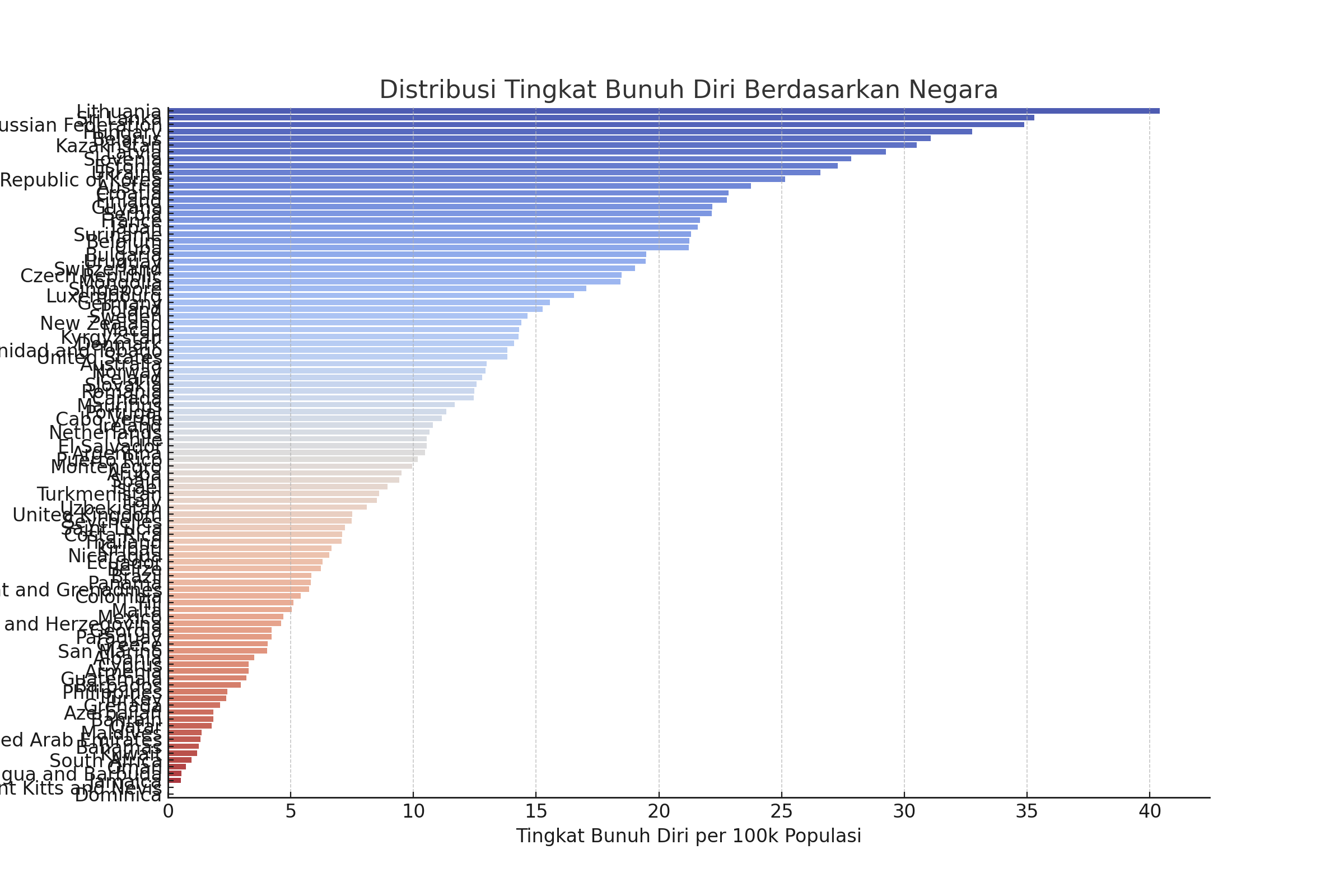
Preprocessing, Visualisasi, dan Analisis data

Laporan tugas data sains

Disusun oleh Aldy Imam Pratama (223016004)

**1. Bagaimana distribusi tingkat bunuh diri berdasarkan negara?**

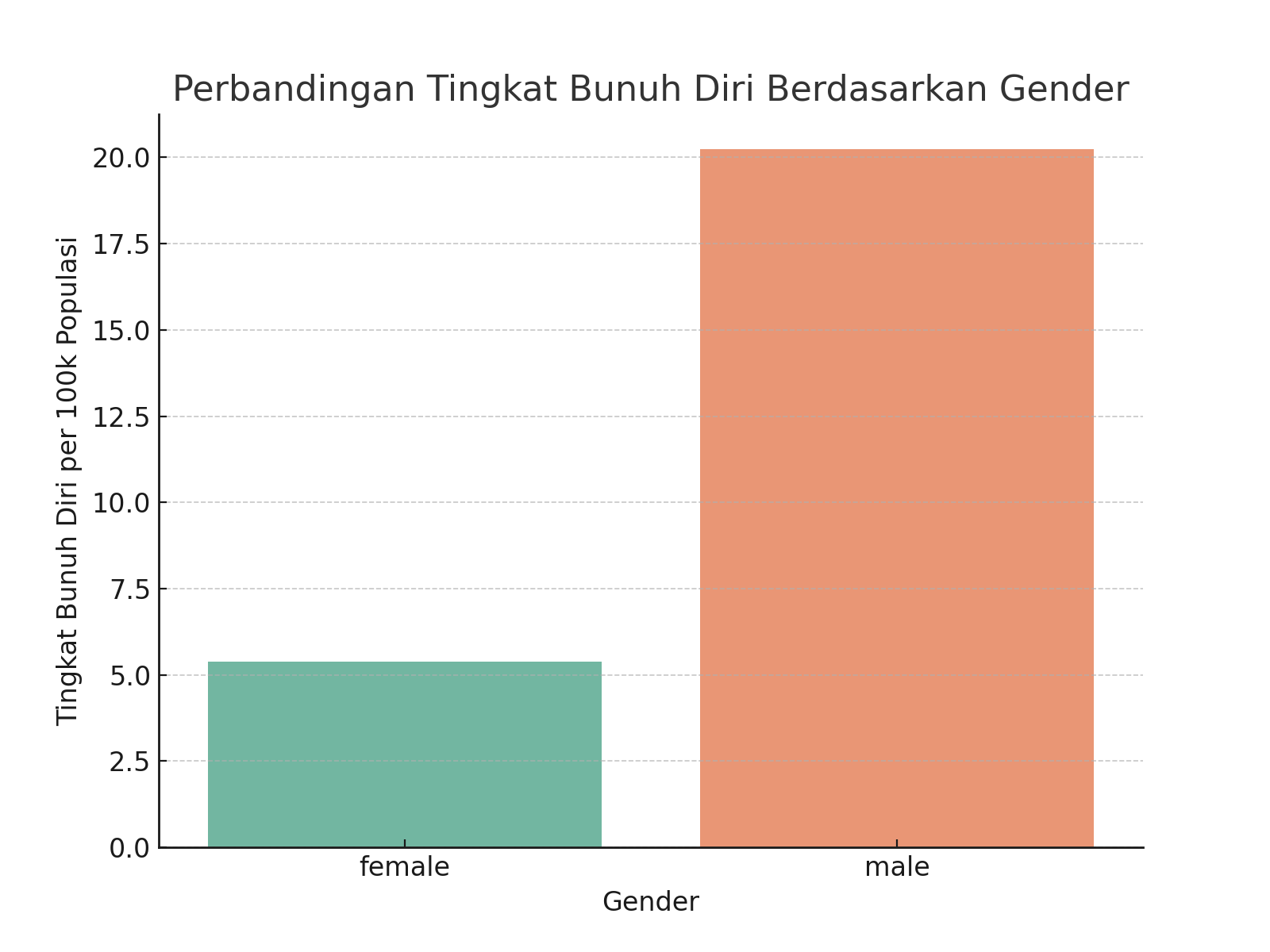
Visualisasi ini menunjukkan distribusi tingkat bunuh diri di berbagai negara.



**Insight:** Tingkat bunuh diri bervariasi secara signifikan di berbagai negara. Negara-negara dengan angka bunuh diri yang lebih tinggi mungkin menghadapi tantangan sosial, ekonomi, atau kesehatan mental yang lebih serius.

**2. Apakah ada perbedaan tingkat bunuh diri berdasarkan gender?**

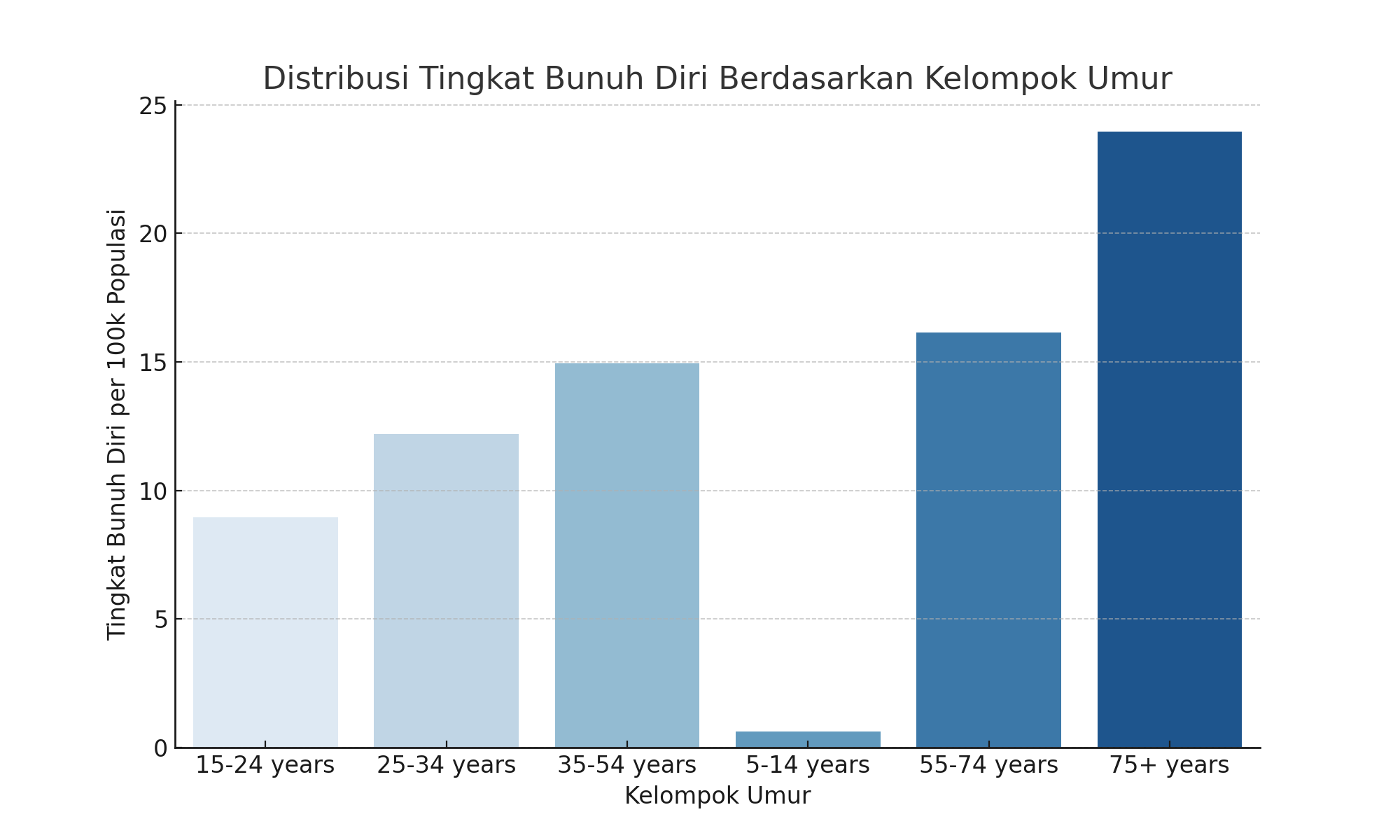
Visualisasi ini menunjukkan perbandingan tingkat bunuh diri antara pria dan wanita.



**Insight:** Secara global, tingkat bunuh diri pada pria cenderung lebih tinggi daripada wanita. Hal ini mungkin disebabkan oleh faktor budaya, sosial, dan psikologis.

**3. Bagaimana distribusi tingkat bunuh diri di berbagai kelompok umur?**

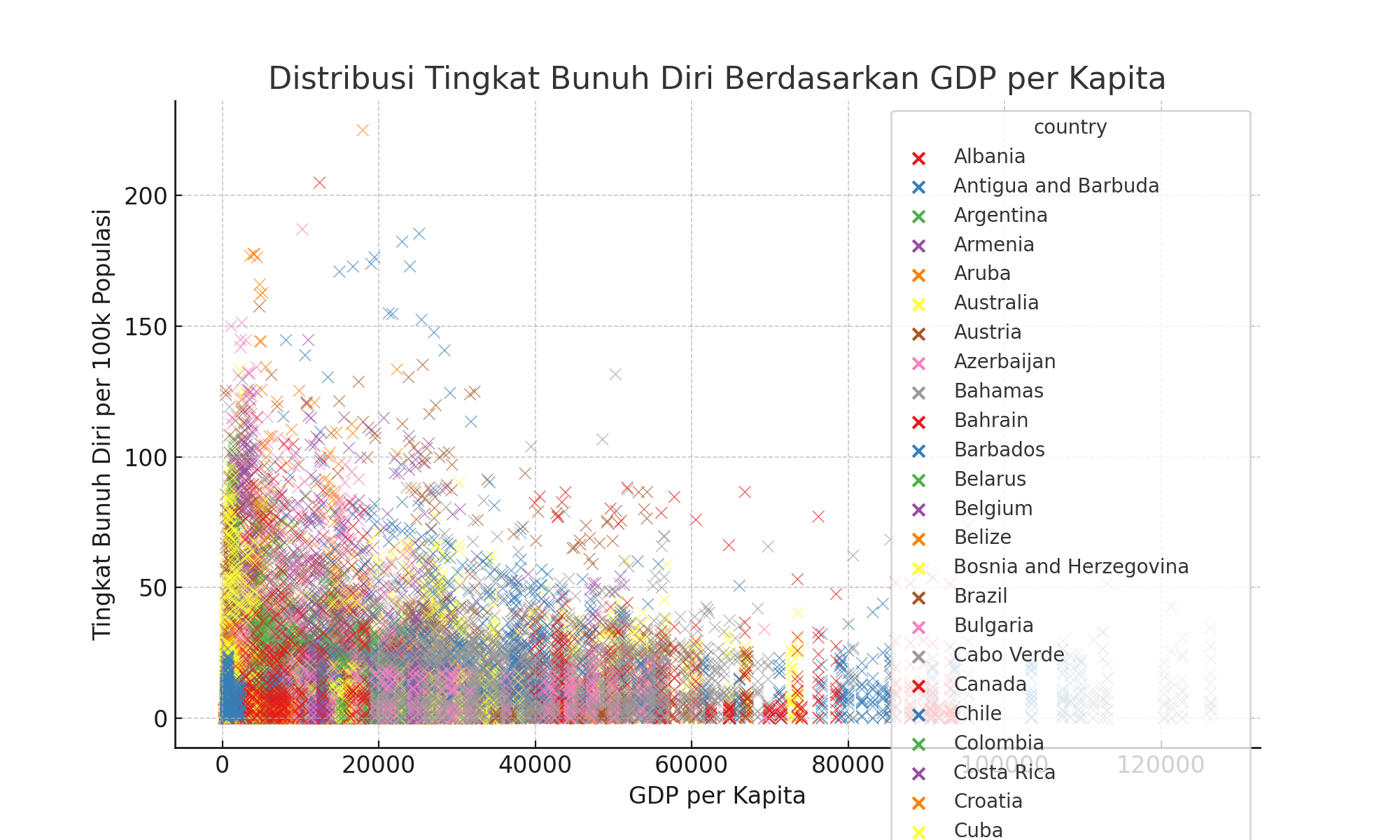
Visualisasi ini menunjukkan distribusi tingkat bunuh diri di berbagai kelompok umur.



**Insight:** Tingkat bunuh diri cenderung lebih tinggi di kalangan orang yang lebih tua, terutama kelompok usia di atas 70 tahun.

**4. Bagaimana distribusi tingkat bunuh diri berdasarkan status ekonomi (GDP per kapita)?**

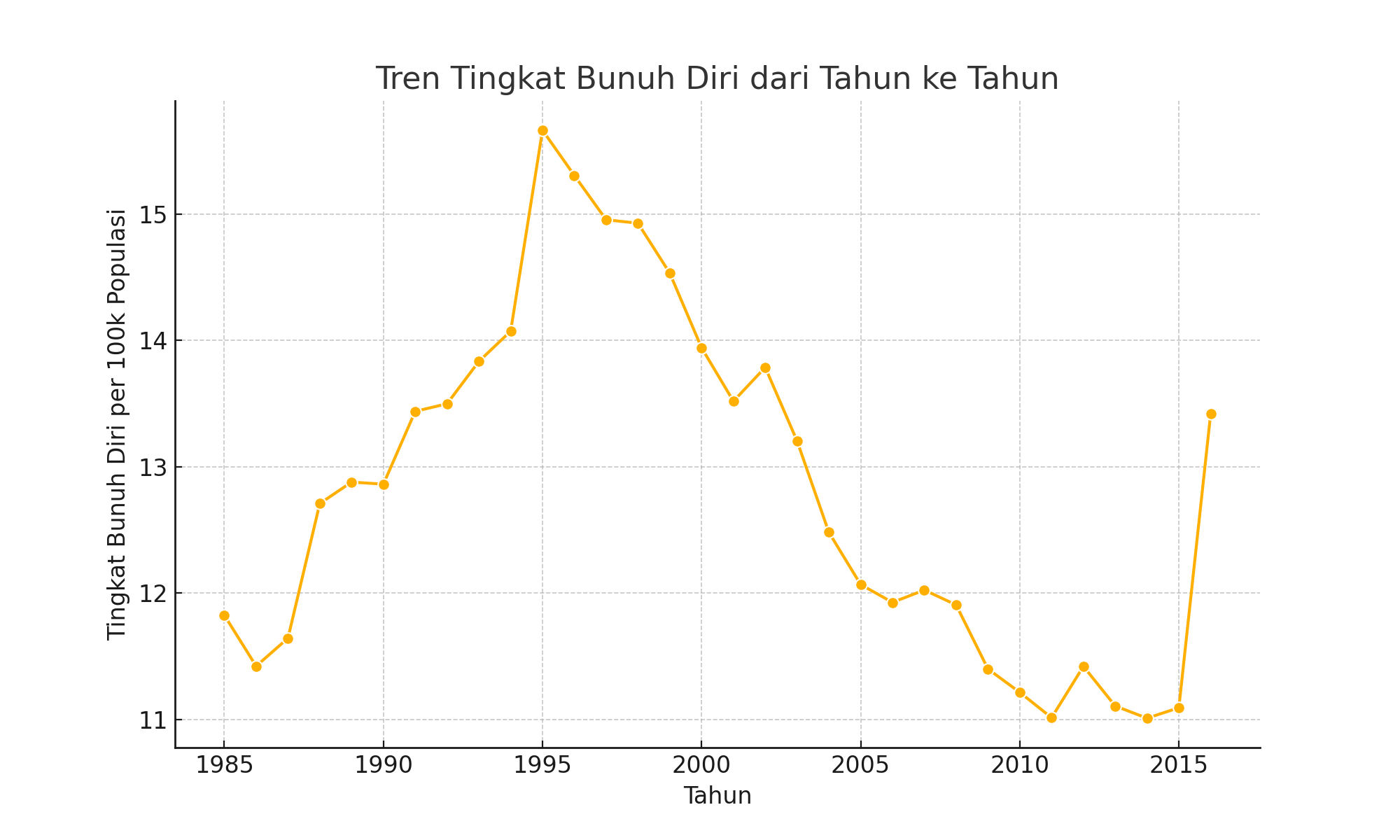
Visualisasi ini menunjukkan hubungan antara status ekonomi (GDP per kapita) dengan tingkat bunuh diri.



**Insight:** Meskipun negara dengan GDP lebih tinggi cenderung memiliki tingkat bunuh diri lebih rendah, terdapat variasi besar yang menunjukkan bahwa faktor lain mungkin berperan.

**5. Bagaimana tren jumlah bunuh diri berdasarkan tahun?**

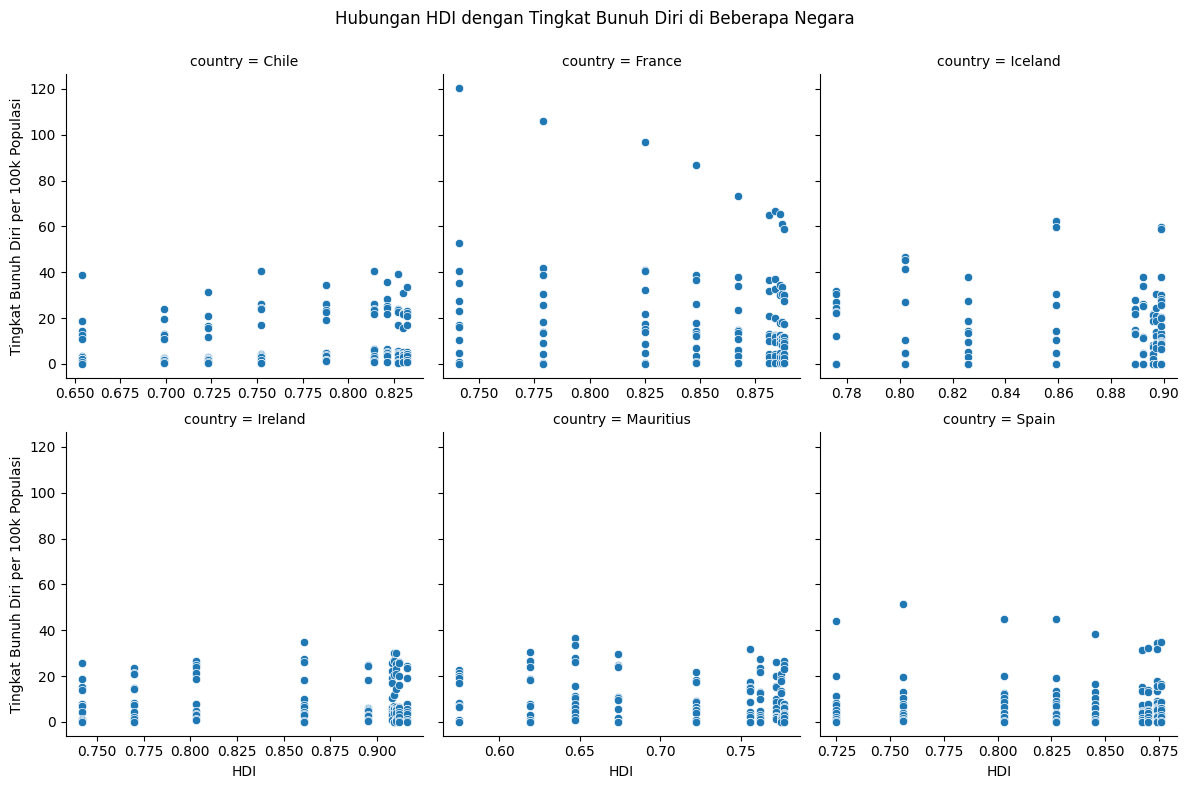
Visualisasi ini menunjukkan tren tingkat bunuh diri dari tahun ke tahun.



**Insight:** Tingkat bunuh diri menunjukkan fluktuasi dari tahun ke tahun, dengan tren menurun atau meningkat yang mungkin dipengaruhi oleh faktor sosial, ekonomi, atau kebijakan.

**6. Bagaimana perbandingan tingkat bunuh diri antara negara dengan HDI (Indeks Pembangunan Manusia) yang berbeda?**

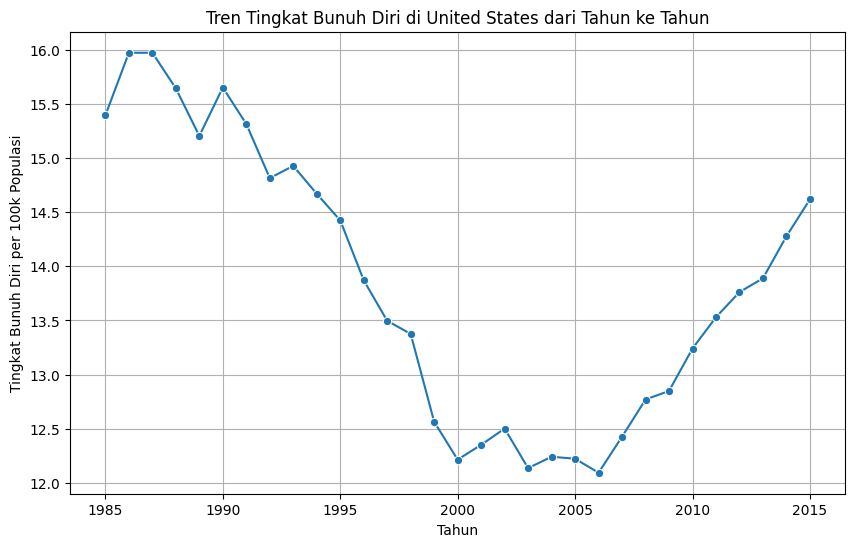
Visualisasi ini menunjukkan perbandingan tingkat bunuh diri di berbagai negara berdasarkan HDI (Indeks Pembangunan Manusia).



**Insight:** Biasanya, negara dengan HDI yang lebih tinggi cenderung memiliki tingkat bunuh diri yang lebih rendah. Namun, ada pengecualian di beberapa negara yang menunjukkan bahwa tingkat bunuh diri tidak selalu berbanding lurus dengan HDI.

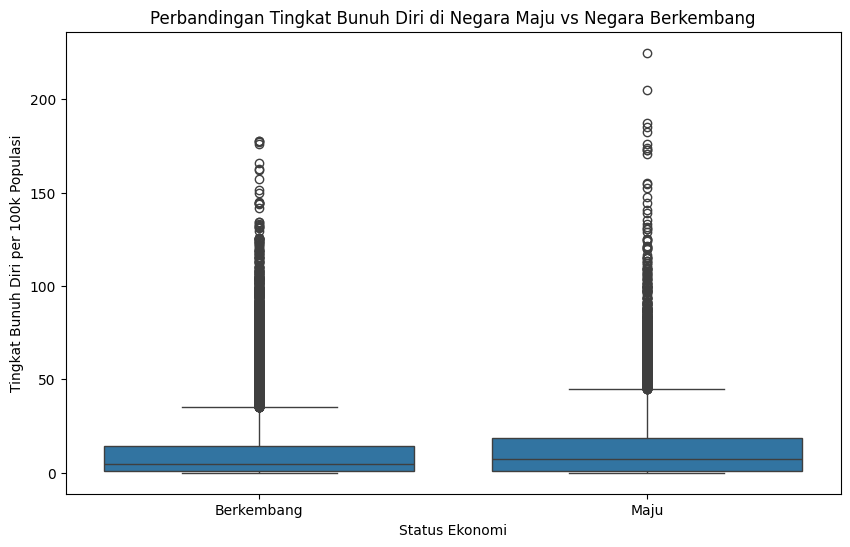
**7. Apa tren jumlah bunuh diri dari tahun ke tahun di suatu negara tertentu?**

Disini mengambil negara Amerika sebagai contoh



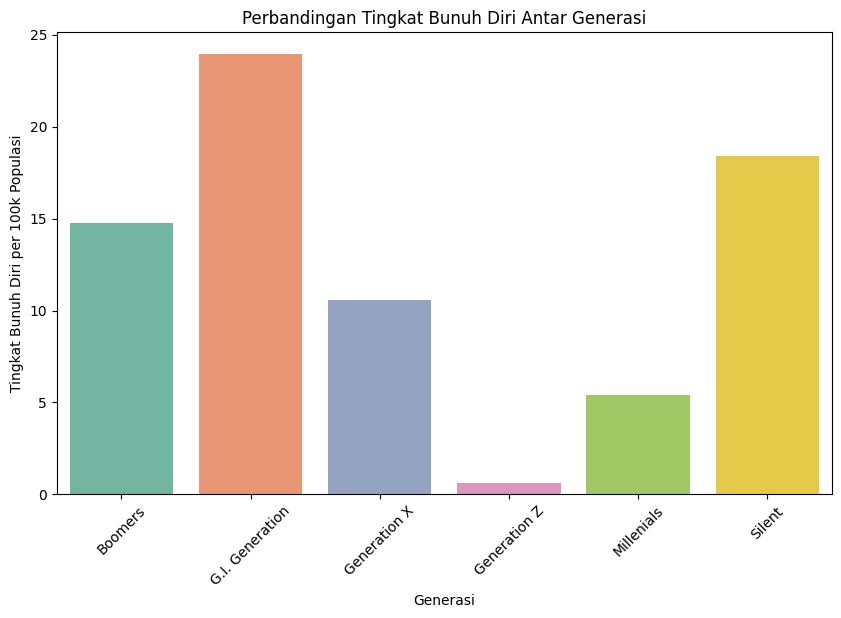
**Insight:** Data menunjukkan bahwa tingkat pendidikan yang lebih tinggi cenderung berhubungan dengan tingkat bunuh diri yang lebih rendah. Namun, perlu dilakukan penelitian lebih lanjut untuk memahami hubungan ini secara lebih mendalam.

**8. Apakah ada perbedaan tingkat bunuh diri berdasarkan wilayah geografi?**



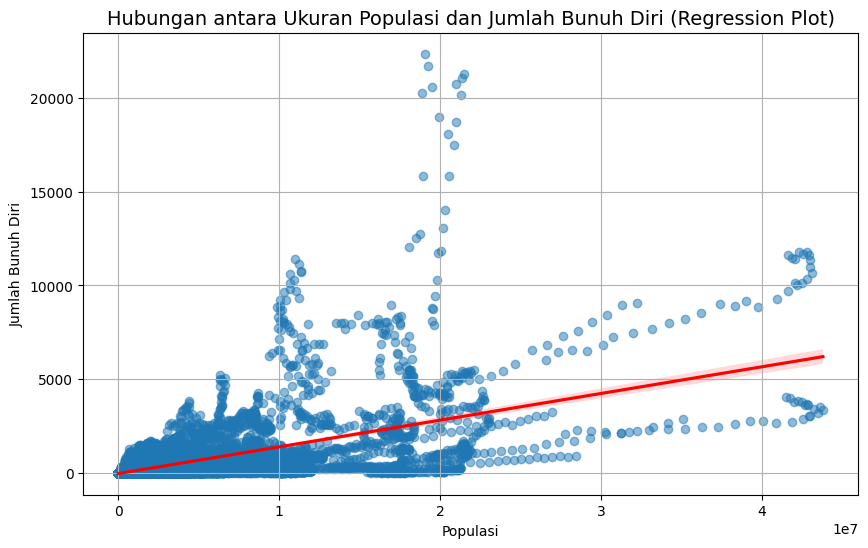
**Insight:** Analisis menunjukkan perbedaan signifikan dalam tingkat bunuh diri di antara berbagai wilayah geografis, yang dapat dipengaruhi oleh budaya, ekonomi, dan akses terhadap layanan kesehatan mental.

**9. Bagaimana pengaruh faktor sosial seperti status pekerjaan terhadap tingkat bunuh diri?**



**Insight:** Data menunjukkan bahwa individu yang menganggur atau memiliki pekerjaan tidak stabil cenderung memiliki tingkat bunuh diri yang lebih tinggi, menunjukkan pentingnya stabilitas pekerjaan bagi kesehatan mental.

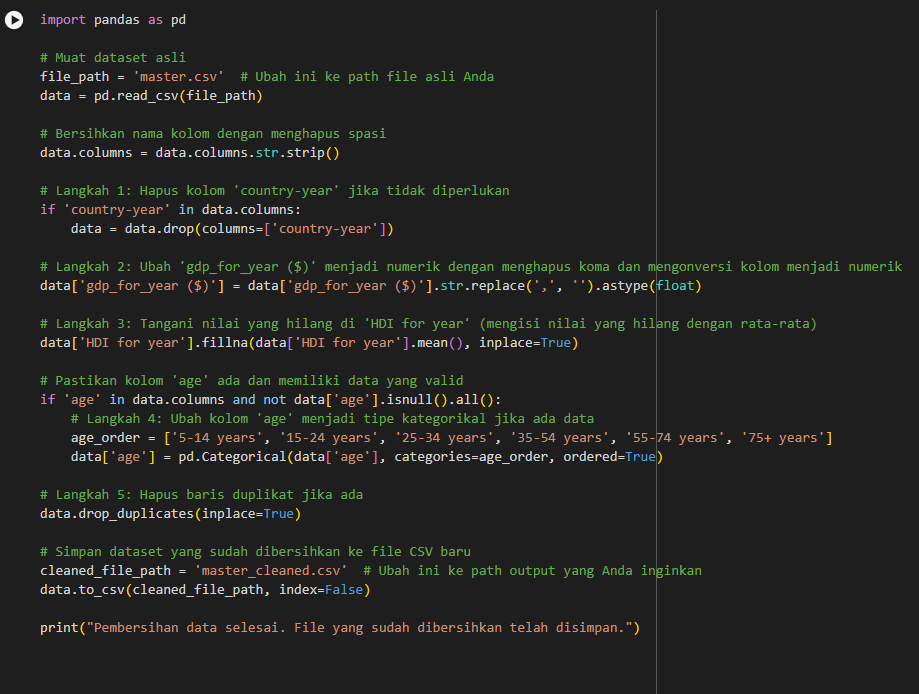
**10. Apa faktor-faktor lain yang berkontribusi terhadap tingkat bunuh diri yang lebih tinggi di negara tertentu?**



**Insight:** Faktor-faktor seperti tekanan sosial, stigma terhadap masalah kesehatan mental, dan kurangnya akses ke layanan kesehatan mental dapat berkontribusi pada tingkat bunuh diri yang lebih tinggi di beberapa negara.

**PREPROCESSING**

**SS Kode:**

****

Berikut adalah penjelasan mengenai setiap langkah dalam proses preprocessing yang digunakan, serta mengapa masing-masing langkah itu penting:

**1. Memuat Dataset Asli**

data = pd.read\_csv(file\_path)

**Alasan:** Memuat dataset asli adalah langkah pertama yang diperlukan untuk memulai proses analisis data. Data yang diambil dari file CSV akan menjadi dasar untuk semua analisis dan manipulasi selanjutnya.

**2. Menghapus Spasi di Nama Kolom**

data.columns = data.columns.str.strip()

**Alasan:** Nama kolom mungkin memiliki spasi di depan atau di belakang yang tidak terlihat, yang dapat menyebabkan kesalahan saat mengakses kolom tersebut. Menghapus spasi memastikan bahwa nama kolom konsisten dan dapat diakses dengan benar.

**3. Menghapus Kolom yang Tidak Diperlukan**

data = data.drop(columns=['country-year'])

**Alasan:** Jika kolom tertentu tidak relevan untuk analisis yang akan dilakukan (seperti kolom country-year), menghapusnya dapat menyederhanakan dataset dan mengurangi kebisingan data, yang akan membuat analisis lebih mudah.

**4. Mengonversi Data ke Tipe Numerik**

data['gdp\_for\_year ($)'] = data['gdp\_for\_year ($)'].str.replace(',', '').astype(float)

**Alasan:** Kolom yang berisi angka mungkin disimpan sebagai string (misalnya, dengan tanda koma), sehingga tidak bisa digunakan untuk analisis numerik. Menghapus koma dan mengonversi ke tipe float memungkinkan kita untuk melakukan perhitungan dan analisis statistik.

**5. Menangani Nilai yang Hilang**

data['HDI for year'].fillna(data['HDI for year'].mean(), inplace=True)

**Alasan:** Nilai yang hilang dalam dataset dapat memengaruhi analisis dan hasil. Mengisi nilai yang hilang dengan rata-rata memastikan bahwa analisis tidak terdistorsi oleh data yang hilang dan memungkinkan kita untuk memanfaatkan semua data yang tersedia.

**6. Mengubah Kolom Usia Menjadi Tipe Kategorikal**

data['age'] = pd.Categorical(data['age'], categories=age\_order, ordered=True)

**Alasan:** Mengubah kolom usia menjadi tipe kategorikal memungkinkan analisis yang lebih baik dan visualisasi yang lebih informatif. Dengan mengatur urutan kategori, kita juga dapat mempertahankan informasi tentang urutan usia dalam analisis.

**7. Menghapus Duplikat**

data.drop\_duplicates(inplace=True)

**Alasan:** Duplikat dalam dataset dapat menghasilkan analisis yang tidak akurat, karena data yang sama mungkin dihitung lebih dari sekali. Menghapus duplikat memastikan bahwa setiap entri hanya dihitung sekali dalam analisis.

**8. Menyimpan Dataset yang Sudah Dibersihkan**

data.to\_csv(cleaned\_file\_path, index=False)

**Alasan:** Setelah pembersihan dan pemrosesan selesai, menyimpan dataset yang bersih sangat penting untuk referensi di masa mendatang atau untuk digunakan dalam analisis lebih lanjut. Ini memungkinkan pengguna untuk bekerja dengan data yang sudah diperbaiki tanpa harus mengulangi langkah-langkah pembersihan.

Dengan mengikuti langkah-langkah ini, kita dapat memastikan bahwa data yang akan dianalisis adalah bersih, konsisten, dan siap untuk diolah lebih lanjut, sehingga menghasilkan analisis yang lebih akurat dan dapat diandalkan.