

LAPORAN HASIL AKHIR FINAL PROJECT
Pengaruh Kelompok Usia Ibu Pada Angka Total Kelahiran di
Provinsi/Kabupaten/Kota



Disusun Oleh :

Ayat Tulloh Rahulloh K. 5026211156

Analitika Data & Diagnostik (B)

DEPARTEMEN SISTEM INFORMASI
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER

2023

BAB 1

PENDAHULUAN

A. Latar Belakang

Kelahiran merupakan kemampuan memproduksi yang sebenarnya dari penduduk. Angka kelahiran total (total fertility rate) adalah rata-rata anak yang dilahirkan seorang wanita selama masa usia suburnya yakni usia 15-49 tahun. Angka kelahiran total ini merupakan indikator penting dan strategis untuk mengetahui sejauh mana keberhasilan suatu negara atau seluruh negara dalam mengendalikan jumlah penduduknya melalui program keluarga berencana (KB). Dikutip dari hasil survei Badan Pusat Statistik tahun 2020, angka kelahiran total (total fertility rate) sebesar 2,10. Ini berarti rata-rata perempuan di Indonesia akan melahirkan dua anak pada masa reproduksinya, dan dengan terjadinya covid-19 juga berdampak pada angka kelahiran total, walaupun dikatakan telah terjadi penurunan, beberapa wilayah masih memiliki tingkat kelahiran yang tinggi, diantaranya Nusa Tenggara Timur, Papua, Papua Barat, dan Maluku. namun belum pernah disebutkan secara spesifik pada usia berapa yang memiliki pengaruh yang paling besar dan usia berapa yang memiliki kestabilan distribusi. Hal ini patut untuk diperhatikan, karena pada usia yang paling subur bisa menjadi faktor pengubah yang paling besar.

B. Penelitian

Penelitian ini menggunakan metode analitika diagnostik komprehensif untuk meningkatkan keakuratan hasil dalam tingkat pengaruh kelompok tertentu usia ibu di Indonesia. Penelitian ini bertujuan untuk mengetahui pengaruh kelompok usia ibu pada angka kelahiran tahun 2020. Diharapkan dengan hasil penelitian ini menjadi bahan pertimbangan dalam memutuskan pendapat untuk menyelesaikan masalah kenaikan kelahiran pada daerah tertentu

BAB 2

PERSIAPAN PENELITIAN

A. Persiapan Dataset

Peneliti menggunakan dataset dari hasil survey Badan Pusat Statistik (BPS) yang terdiri dari 2 dataset yang relevan dengan Angka Kelahiran Indonesia. Dataset yang pertama adalah "Angka Kelahiran Total / Total Fertility Rate (TFR) Hasil Long Form (LF) SP2020 Menurut Provinsi/Kabupaten/Kota, 2020", yang memberikan informasi Total angka kelahiran tiap daerah selama satu dekade. Dan yang kedua adalah "Angka Kelahiran Hasil Long Form SP2020 Menurut Kelompok Umur Ibu (Age Spesific Fertility Rate/ASFR) dan Provinsi/Kabupaten/Kota, 2020", berisikan informasi Banyaknya jumlah kelahiran hidup dari perempuan dengan kelompok umur tertentu per 1000 perempuan dalam kelompok umur yang sama. Perempuan yang dicakup adalah perempuan dalam masa subur (15-49 tahun) yang dikelompokkan berdasarkan kelompok umur 5 tahunan. Kedua dataset di merge dengan Provinsi/Kabupaten/Kota sebagai indexnya.

B. Persiapan Alat

Untuk menunjang tingkat keakuratan yang lebih baik karena harus melakukan analisis diagnostik, maka dalam kasus ini, penggunaan yang terbaik adalah menggunakan Python dengan IDE DataSpell/Jupyter serta disini Excel juga akan digunakan untuk rename kolom dataset yang bermasalah. Langkah pertama adalah mempersiapkan library yang dibutuhkan, seperti pandas, numpy, matplotlib, seaborn dan statsmodel.

BAB 3

METODE PENELITIAN

A. Input

a. Pengambilan data

Disini peneliti melakukan import dataset dengan library pandas, yang dimana dataset pertama akan diidentifikasi sebagai 'form', dan dataset kedua sebagai 'total'. Kemudian dijalankan fungsi merge untuk menggabungkan kedua dataset.

```
form=pd.read_csv('datasets/Angka Kelahiran Hasil  
Long Form SP2020 Menurut Kelompok Umur Ibu (Age  
Spesific Fertility Rate_ASFR) da.csv')  
total=pd.read_csv('datasets/Angka Kelahiran Total_  
Total Fertility Rate (TFR) Hasil Long Form (LF)  
SP2020 Menurut Provinsi_Kabupa.csv')  
  
df=pd.merge(form, total,  
on='Provinsi/Kabupaten/Kota')
```

Fungsi ini menjadikan 2 dataset menjadi satu data frame dengan "Provinsi/Kabupaten/Kota" sebagai kunci penggabungan yang nantinya digunakan sebagai index.

b. Pembersihan data

Pembersihan data dilakukan jika data itu memiliki anomali, nilai null, NaN. Dataset harus terhindari dari masalah-masalah tersebut untuk bisa dilakukan analisa diagnostik. Langkah pertama adalah dengan menggunakan fungsi .isna().sum(), yang berguna untuk mengetahui apakah dataset kita memiliki nilai NaN

```
df_indexed.isna().sum()
```

Dalam kasus ini, peneliti tidak menemukan adanya NaN. Sehingga dilanjutkan dengan proses drop kolom, tujuannya adalah untuk menghapus kolom yang tidak ada hubungannya dengan penelitian

```
drop_column=['idx', 'id', 'Unnamed: 3']  
df=df.drop(columns=drop_column)  
  
df_indexed=df.set_index('Provinsi/Kabupaten/Kota')  
df_indexed
```

Disini peneliti melakukan drop 3 kolom dan menggunakan kolom 'Provinsi/Kabupaten/Kota' sebagai index.

B. Proses

a. Statistik

Untuk mengetahui ringkasan statistik deskriptif dari DataFrame dilakukan fungsi .describe() untuk menghasilkan ringkasan statistik deskriptif dari DataFrame. Statistik tersebut melibatkan nilai rata-rata, standar deviasi,

nilai minimum, kuartil (25%, 50%, 75%), dan nilai maksimum dari setiap kolom numerik dalam DataFrame.

```
df_indexed.describe().round(3)
```

.round(3) digunakan untuk membulatkan 3 angka dibelakang koma, ini membantu membuat output lebih mudah dibaca dan dapat meminimalkan jumlah angka yang ditampilkan.

Selanjutnya dilakukan identifikasi mean ke dalam variabel untuk nantinya dijadikan visualisasi mencari usia yang rata-rata jumlah kelahiran hidup yang paling besar.

```
df_mean=df_indexed.mean()  
df_mean
```

b. Line Chart

Dalam usaha mengetahui tren, digunakanlah matplotlib untuk memvisualisasikan line chart, dimana untuk ukuran kanvasnya sebesar 17x9, serta dataset yang digunakan adalah 'df_mean' yang artinya akan menampilkan tren rata-rata Angka kelahiran per kelompok usia. Dengan detail, sumbu x = kelompok usia, y= nilai rata-rata.

```
plt.figure(figsize=(17,9))  
plt.plot(df_mean)  
plt.title("Nilai rata-rata Angka kelahiran")
```

c. Correlation

Untuk mengetahui hubungan antar variabel, dibuktikan dengan correlation. Langkah pertama adalah menghitung nilai korelasi antar kolom untuk menghasilkan sebuah matriks yang nantinya akan divisualkan ke bentuk heatmap. Selanjutnya menyentuh pada tahap visualisasi dengan seaborn heatmap, dengan data matriks.

```
corr_mat=df_indexed.corr()  
plt.figure(figsize=(8,6))  
sns.heatmap(corr_mat, cmap="crest", annot=True)  
plt.title('Correlation Plot')  
plt.show()  
  
corr_mat
```

d. Pairplot

Dalam upaya menggambarkan hubungan sekaligus distribusi antar variabel, langkah awal dilakukan dengan pembuatan pairplot menggunakan modul Seaborn pairplot. Proses dimulai dengan langsung dengan dataset (seluruh kolom) untuk divisualkan, tujuannya untuk mengetahui distribusi mana yang bagus.

```
sns.pairplot(df_indexed)  
plt.show()
```

Pairplot ini akan secara otomatis menggambarkan histogram yang dimana kita bisa melihat hasil distribusi.

e. Boxplot

Digunakannya boxplot adalah untuk mencari outlier dalam data, proses dimulai dengan seaborn boxplot

```
sns.boxplot(df['25-29'])
plt.title("25-29")
plt.show()

sns.boxplot(df['35-39'])
plt.title("35-39")
plt.show()
```

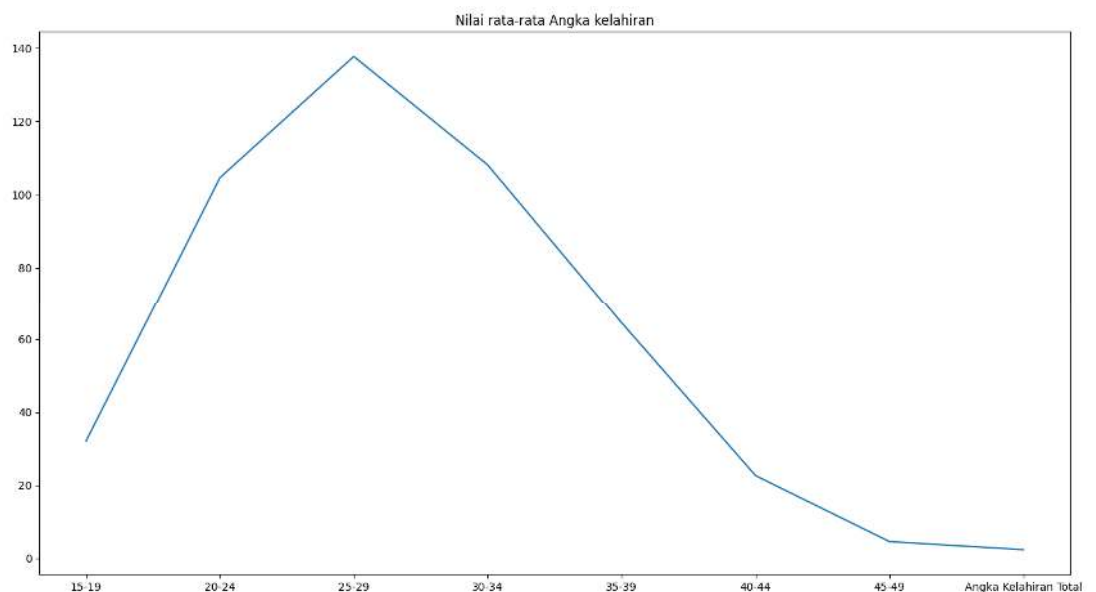
Disini peneliti menggunakan 2 variabel pilihan, untuk alasannya akan dijelaskan di bab selanjutnya.

C. Output

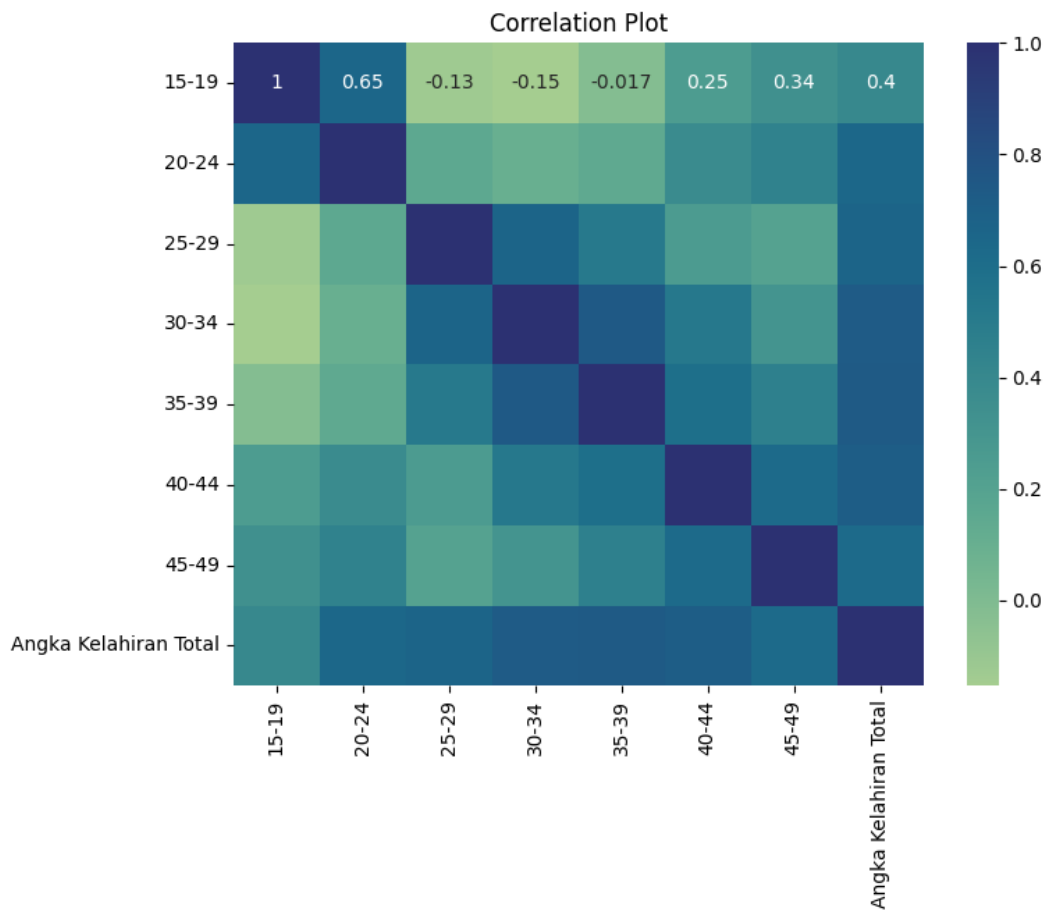
a. Tabel Statistik

	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Angka Kelahiran Total
Count	551.000	551.000	551.000	551.000	551.000	551.000	551.000	551.000
Mean	32.164	104.637	137.764	108.141	64.482	22.635	4.579	2.372
Std	17.215	22.706	22.202	20.858	16.132	11.189	5.439	0.375
Min	4.000	44.500	93.100	53.200	25.800	3.900	0.100	1.540
25%	18.500	92.200	123.250	95.050	54.150	15.850	1.800	2.150
50%	30.100	103.900	134.900	104.000	61.200	19.800	2.900	2.300
75%	43.750	116.500	147.500	117.700	74.050	25.800	5.000	2.490
Max	105.600	237.600	311.500	210.700	132.600	114.700	52.200	4.220

b. Line Chart

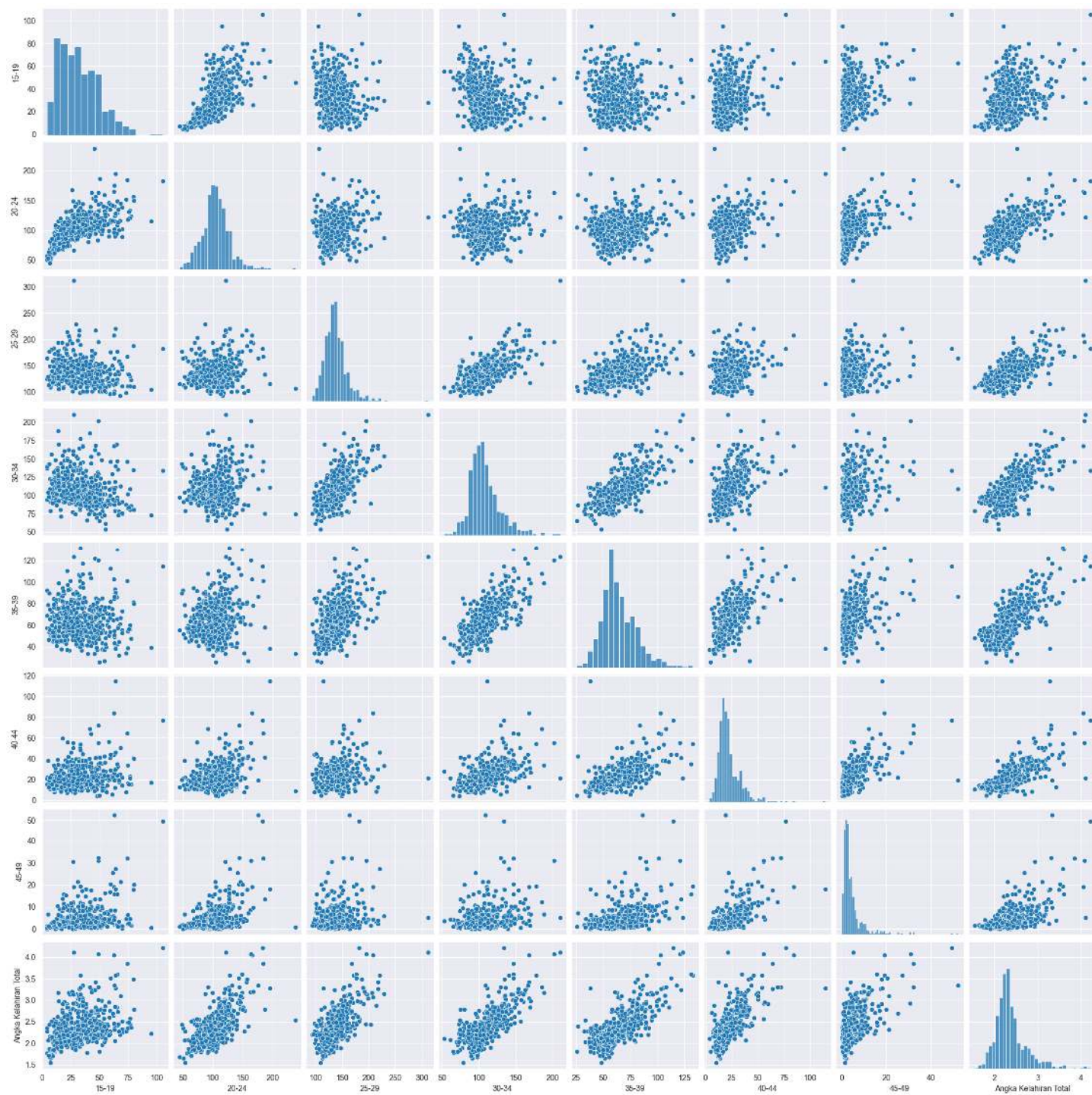


c. Correlation



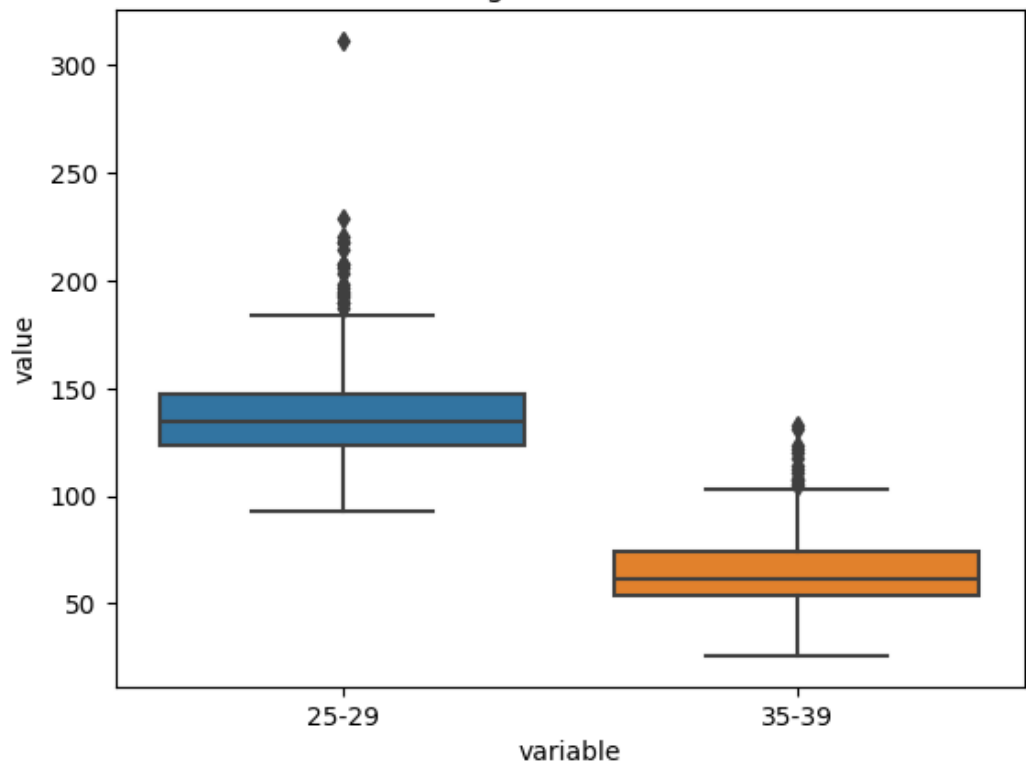
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Angka Kelahiran Total
15-19	1.000000	0.652937	-0.129266	-0.153082	-0.017089	0.249939	0.341239	0.404484
20-24	0.652937	1.000000	0.154378	0.102471	0.147155	0.377570	0.447702	0.646466
25-29	-0.129266	0.154378	1.000000	0.670381	0.517305	0.256708	0.201242	0.664622
30-34	-0.153082	0.102471	0.670381	1.000000	0.744160	0.520234	0.308669	0.733357
35-39	-0.017089	0.147155	0.517305	0.744160	1.000000	0.590752	0.459593	0.737764
40-44	0.249939	0.377570	0.256708	0.520234	0.590752	1.000000	0.621877	0.713653
45-49	0.341239	0.447702	0.201242	0.308669	0.459593	0.621877	1.000000	0.623297
Angka Kelahiran Total	0.404484	0.646466	0.664622	0.733357	0.737764	0.713653	0.623297	1.000000

d. Pairplot & Histogram



e. Boxplot

Angka Kelahiran



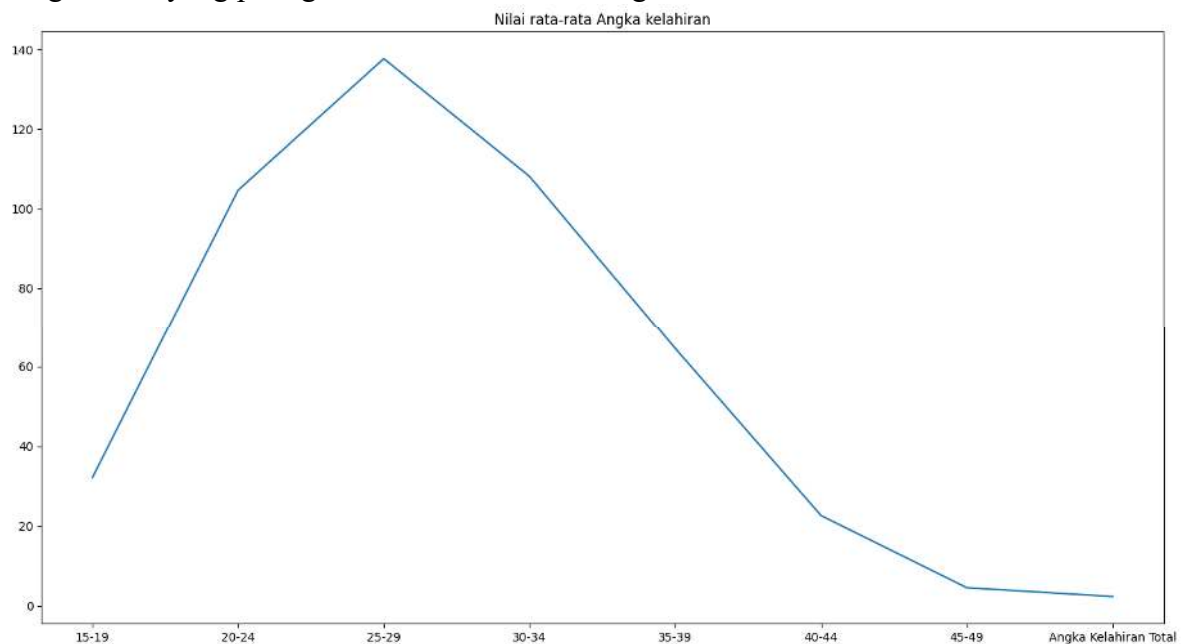
BAB 4

HASIL & ANALISIS

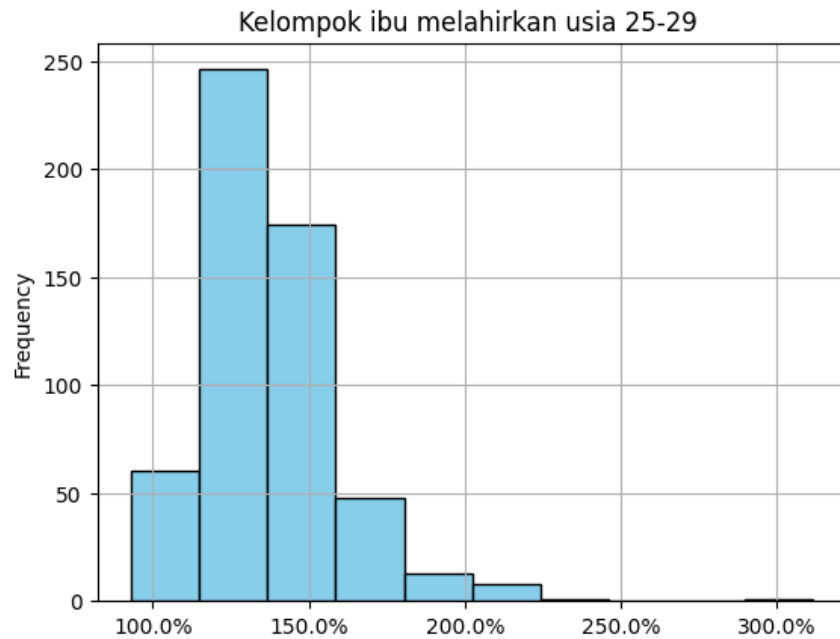
A. Kelompok Usia Yang Paling Produktif

	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Angka Kelahiran Total
Count	551.000	551.000	551.000	551.000	551.000	551.000	551.000	551.000
Mean	32.164	104.637	137.764	108.141	64.482	22.635	4.579	2.372
Std	17.215	22.706	22.202	20.858	16.132	11.189	5.439	0.375
Min	4.000	44.500	93.100	53.200	25.800	3.900	0.100	1.540
25%	18.500	92.200	123.250	95.050	54.150	15.850	1.800	2.150
50%	30.100	103.900	134.900	104.000	61.200	19.800	2.900	2.300
75%	43.750	116.500	147.500	117.700	74.050	25.800	5.000	2.490
Max	105.600	237.600	311.500	210.700	132.600	114.700	52.200	4.220

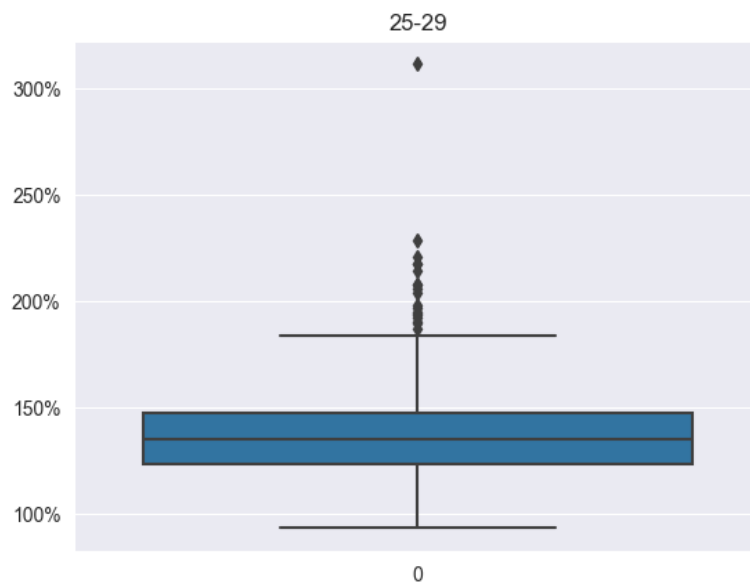
Dari hasil describe diatas, didapatkan rata-rata terbesar diraih oleh kelompok usia 25-29 tahun dengan nilai 137.764 dengan nilai standard deviasi sebesar 22.202. Jika dilihat nilai Min dan Max memiliki selisih 218.400. Nilai ini terbilang sangat besar dan perlu diawasi untuk kedepannya. Untuk persebarannya terlihat pada kuartil-kuartil dengan nilai yang paling relevan adalah 50% dengan nilai 134.900.



Bukti bahwa pada umur 25-29 tahun selama 1 dekade adalah yang paling tertinggi. Dilihat dari grafik diatas, bahwasannya **rata-rata** umur tersebut hampir menyentuh 140% dalam 1 dekade. Penurunan terlihat pada usia 30-49. Walaupun pada tahun 2020 indonesia dilanda pandemi Covid-19, dan terbilang mengalami penurunan yang signifikan dibanding dengan periode sebelumnya.



Berikut adalah tampilan persebaran **keseluruhannya** pada usia 25-29, Terlihat persebaran paling tinggi berada diantara 100%-150% yang hampir menyentuh angka 250. Hal ini menunjukkan adanya konsentrasi tingkat kelahiran yang tinggi pada nilai-nilai tersebut. Kemungkinan untuk terjadinya penurunan tingkat kelahiran hanya apabila nilai yang konsentrasinya rendah diperkuat.

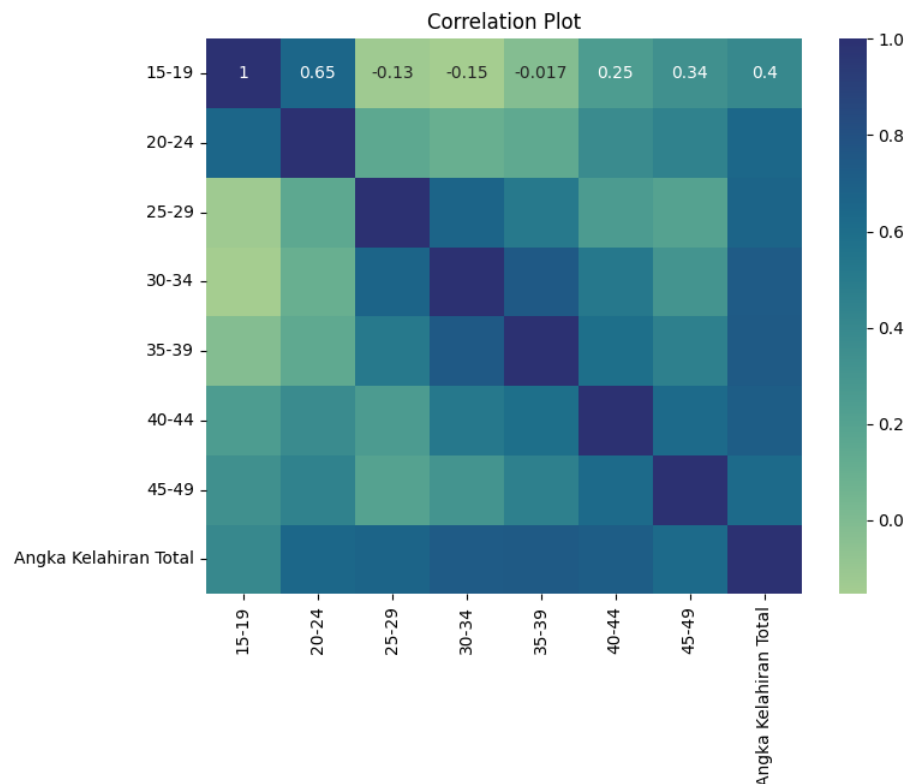


Boxplot diatas menunjukkan terdapat outlier data, menandakan bahwa terdapat nilai yang ekstrem pada pendistribusian data. Kemungkinan nilai maksimumnya <200% dan nilai minimumnya adalah 93.10% yang mengindikasikan rentang variasi tingkat kelahiran.

Terdapatnya outlier dalam data menyebabkan mean dan standard deviasi meningkatkan nilai secara signifikan.

B. Kelompok Usia Yang Memiliki Pengaruh Dalam Peningkatan Angka Kelahiran Total

Untuk mengetahui kelompok usia yang memiliki peranan mempengaruhi adalah dengan meninjau nilai korelasi. Berikut adalah Heatmap korelasi:

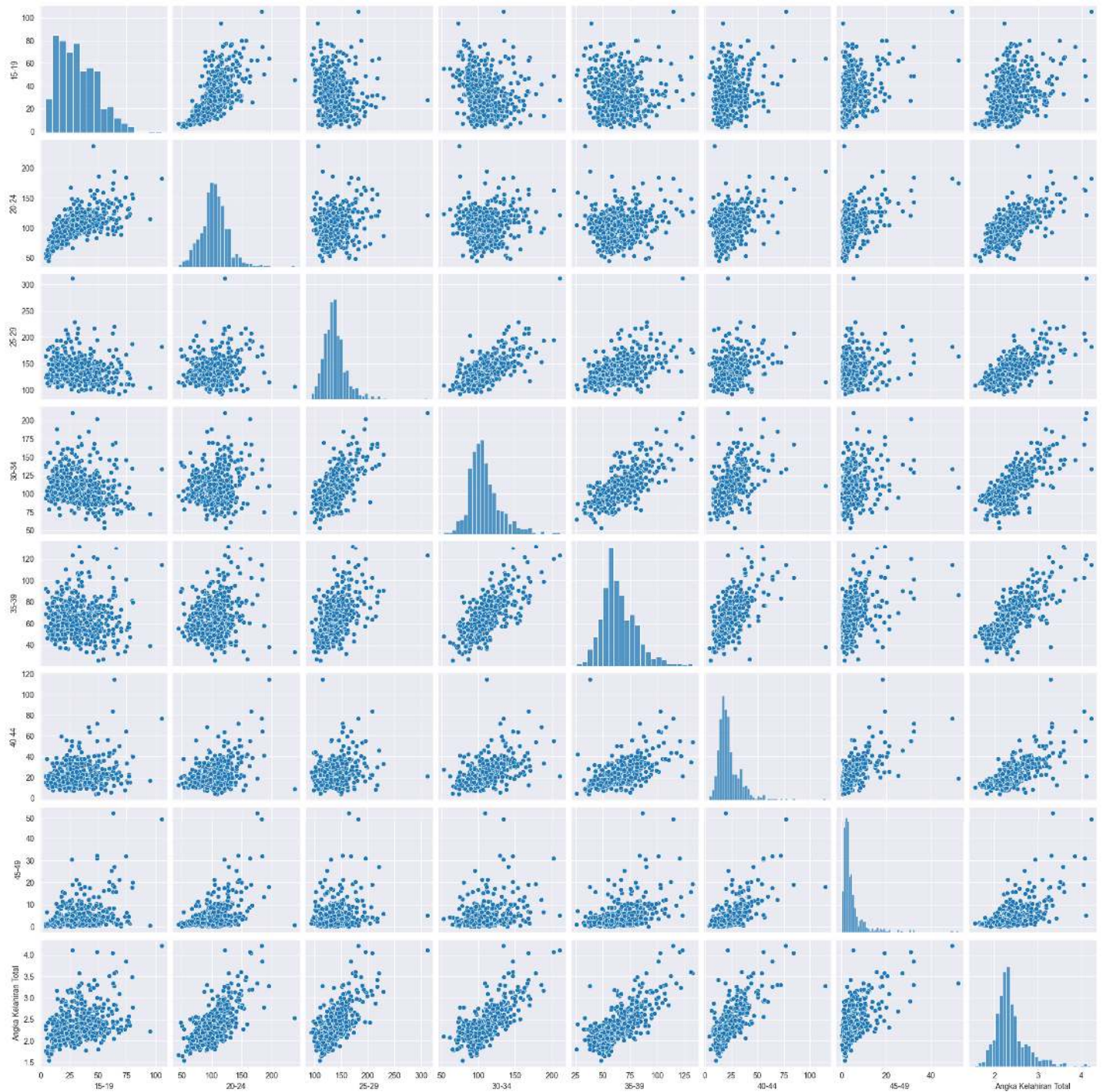


Analisis diagnostik dari correlation plot terkait kelompok usia yang memiliki pengaruh terhadap Angka Kelahiran Total menunjukkan pola positif pada kelompok usia 35-39 tahun. Hal ini juga terbukti dalam tabel dibawah ini:

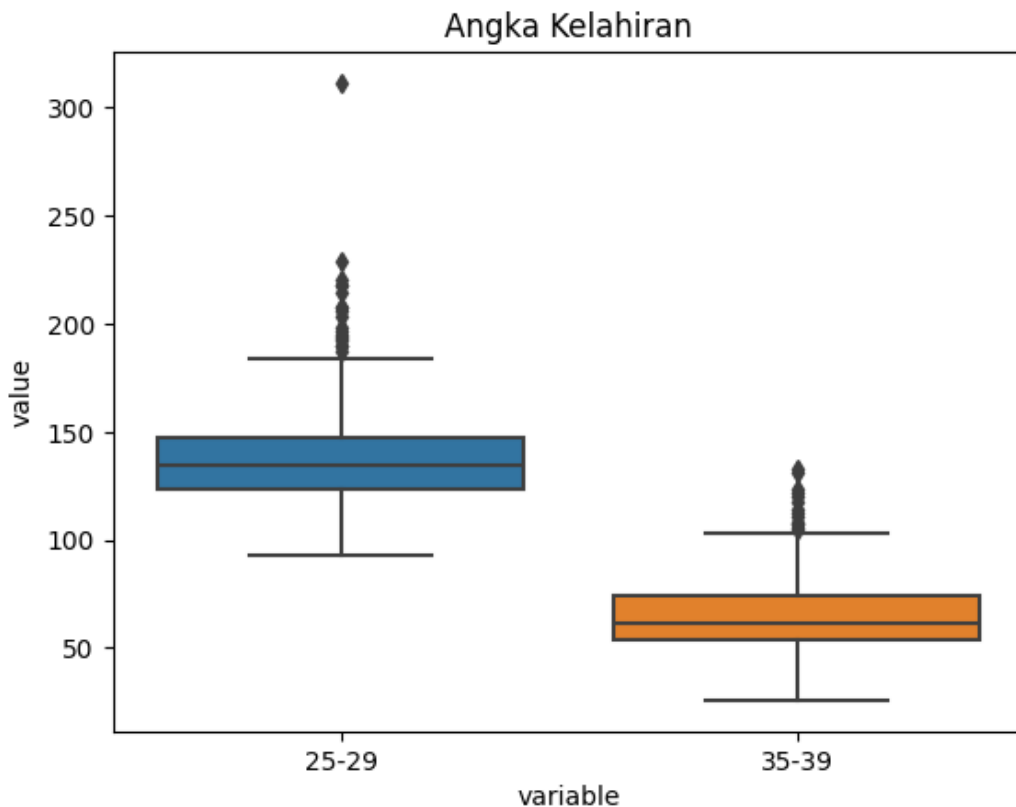
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	Angka Kelahiran Total
15-19	1.000000	0.652937	-0.129266	-0.153082	-0.017089	0.249939	0.341239	0.404484
20-24	0.652937	1.000000	0.154378	0.102471	0.147155	0.377570	0.447702	0.646466
25-29	-0.129266	0.154378	1.000000	0.670381	0.517305	0.256708	0.201242	0.664622
30-34	-0.153082	0.102471	0.670381	1.000000	0.744160	0.520234	0.308669	0.733357
35-39	-0.017089	0.147155	0.517305	0.744160	1.000000	0.590752	0.459593	0.737764
40-44	0.249939	0.377570	0.256708	0.520234	0.590752	1.000000	0.621877	0.713653
45-49	0.341239	0.447702	0.201242	0.308669	0.459593	0.621877	1.000000	0.623297
Angka Kelahir an Total	0.404484	0.646466	0.664622	0.733357	0.737764	0.713653	0.623297	1.000000

Terlihat dalam tabel bahwasannya nilai yang dimiliki kelompok usia 35-39 adalah yang tertinggi dari pada kelompok usia yang lain dengan nilai correlation 0.737764. Angka ini hampir menyentuh 1, yang dimana itu menunjukkan hubungan

positif tetapi tidak sempurna di antara dua variabel. Artinya, ketika satu variabel naik, variabel lainnya juga naik secara proporsional, dan sebaliknya.



Terlihat juga dalam histogram, bahwasannya kelompok 35-39 memiliki grafik distribusi yang hampir sempurna.



Jika dibandingkan dengan kelompok usia 25-29, kelompok usia 35-39 lebih minim nilai ekstrem. Bukti bahwasannya kelompok usia 25-29 memiliki kemungkinan nilai maksimumnya <200% dan nilai minimumnya adalah 93.10% dan memiliki nilai ekstrem pada >300%, serta, pada kelompok usia 35-39 dengan nilai maksimumnya sekitar 130% dan outlier ekstremnya hampir menyentuh 150%. Hal ini menandakan bahwasannya kelompok usia 35-39 sangat berperan besar dalam penurunan atau kenaikan terhadap Angka Kelahiran Total pada periode tersebut.

BAB 5

KESIMPULAN

Telah dilakukan analisa diagnostik pada dataset Angka Kelahiran Total dan dataset Angka Kelahiran Berdasarkan Kelompok Usia Ibu di Indonesia dengan Python sebagai bahasa pemrograman. Dengan proses yang diawali dengan import dataset, pembersihan data, statistik data, line chart, correlation matrix, pairplot, boxplot, sehingga terpenuhi semua tujuan yang diinginkan, Kelompok usia yang paling produktif pada periode tersebut diselesaikan dengan mengambil rata-rata, yakni sebesar 139% , dan pada tampilan barplot, persebaran **keseluruhannya** pada usia 25-29, Terlihat persebaran paling tinggi berada diantara 100%-150% yang hampir menyentuh angka 250. Hal ini menunjukkan adanya konsentrasi tingkat kelahiran yang tinggi pada nilai-nilai tersebut. Hal ini juga terjadi karena terdapat outlier dengan nilai yang ekstrem yang menyentuh >300%. Dapat ditarik kesimpulan bahwa kelompok usia 25-29 adalah kelompok usia yang paling banyak menyumbang Angka Kelahiran Total pada periode tersebut.

Kelompok usia 35-39 adalah kelompok usia yang memiliki pengaruh besar dalam penurunan atau kenaikan terhadap Angka Kelahiran Total pada periode tersebut. Hal ini dibuktikan dengan nilai korelasi yang paling tinggi dibanding kelompok usia lain, dengan nilai 0.737764, dimana itu menunjukkan hubungan positif tetapi tidak sempurna di antara dua variabel. Artinya, ketika satu variabel naik, variabel lainnya juga naik secara proporsional, dan sebaliknya. Serta memiliki distribusi yang hampir normal, outlier pun tidak terlalu ekstrem, sekitar hampir menyentuh 150% dengan nilai maksimumnya 132.60%.

LAMPIRAN

File ipynb:

https://colab.research.google.com/drive/1bZSi9UZ7fVRcPGDRLe1ERVLV6ZHH_nxC?usp=sharing