

ANALISIS PENGARUH FAKTOR DEMOGRAFIS TERHADAP TINGKAT PENGANGGURAN TERBUKA (TPT) DI INDONESIA TAHUN 2015–2022

MULAI SEKARANG



SELENGKAPNYA





DISUSUN OLEH :

● SEPTIANI AMALIA WULANDARI
1314623069

● DARISTA WARDHANI
24031554146



● LATAR BELAKANG

Pengangguran merupakan salah satu permasalahan ekonomi dan sosial yang masih dihadapi oleh banyak negara, termasuk Indonesia. Tingginya tingkat pengangguran tidak hanya mencerminkan ketidakseimbangan antara permintaan dan penawaran tenaga kerja, tetapi juga menjadi indikator ketidakoptimalan pembangunan ekonomi dan kualitas sumber daya manusia. Berdasarkan proyeksi penduduk oleh BPS, jumlah penduduk Indonesia akan melonjak hingga 300 jiwa. Kondisi ini dapat menjadi peluang untuk mendorong pertumbuhan ekonomi jika dimanfaatkan secara optimal, namun juga dapat meningkatkan angka pengangguran bila tidak dikelola dengan baik. Berbagai penelitian terdahulu menunjukkan bahwa perempuan memiliki peluang lebih tinggi untuk menganggur dibanding laki-laki, dan tingkat pengangguran juga bervariasi berdasarkan pendidikan dan usia. Oleh karena itu, penting untuk menganalisis pengaruh faktor demografis terhadap TPT agar kebijakan ketenagakerjaan dapat lebih tepat sasaran.

● **TUJUAN**

Tujuan penelitian ini adalah untuk memahami hubungan antara faktor-faktor demografis seperti pendidikan, umur, dan gender dengan tingkat pengangguran terbuka di Indonesia. Selain itu, juga bertujuan untuk memberikan gambaran mengenai kondisi tenaga kerja di masa bonus demografi serta pentingnya peningkatan kualitas sumber daya manusia agar dapat bersaing di pasar kerja.

● **MANFAAT**

Manfaat dari pembahasan ini adalah memberikan wawasan bagi masyarakat mengenai pentingnya pendidikan, keterampilan, dan kesiapan kerja dalam menghadapi tantangan ketenagakerjaan. Selain itu, hasil kajian ini dapat menjadi bahan pemahaman bagi pemerintah dan lembaga terkait dalam menyusun program peningkatan kualitas tenaga kerja.





PROSES WRANGLING



TEKNIK PENGAMBILAN

Teknik pengambilan (sering disebut sebagai *data extraction technique*) adalah cara atau metode yang digunakan untuk mengambil, memilih, menggabungkan, atau mendapatkan data tertentu dari suatu sumber data baik itu berasal dari tabel, kolom, database, file CSV, maupun kumpulan dataset yang lebih besar.

● DIRECTORY SCANNING DAN FILTERING

Bagian code 1 menggunakan teknik pengambilan data berbasis pencarian file dalam folder. program memeriksa folder /content (folder Google Colab), kemudian:

- Menemukan semua file CSV dengan pengecekan `endswith('.csv')`
- Menemukan semua file Excel dengan pengecekan `endswith('.xlsx')`
- Menampilkan file apa saja yang ditemukan

● DATAFRAME DAN SAMPLING

Bagian ini menggunakan teknik pengambilan data berupa pembacaan file. `pd.read_csv()` adalah teknik yang digunakan untuk mengambil data tabel dari file CSV dan memuatnya ke dalam DataFrame.

Bagian `print("=== Tabel Data 1 ===")` dan `print(df1.head())` melakukan teknik pengambilan data berupa sampling, yaitu mengambil 5 baris pertama sebagai contoh.

Tujuannya untuk:

- Mengecek apakah data berhasil dibaca
- Melihat struktur kolom
- Memastikan data tidak rusak



CSV ditemukan: ['TPT_Age.csv', 'Data_pengangguran.csv', 'Data_Pendidikan.csv', 'TPT_Gender.csv']
Excel ditemukan: []

```

=== Tabel Data 1 ===
  Periode  Bulan  Tidak/belum pernah sekolah  Tidak/belum tamat SD \
0    2006  Februari                234465                614960
1    2006   Agustus                170666                611254
2    2007  Februari                145750                520316
3    2007   Agustus                 94301                438519
4    2008  Februari                 79764                448431

      SD    SLTP  SLTA Umum/SMU  SLTA Kejuruan/SMK  Akademi/Diploma \
0  2675459  2860007        2842876        1204140        297185
1  2589699  2730045        2851518        1305190        278074
2  2753548  2643062        2630360        1114675        330316
3  2179792  2264198        2532204        1538349        397191
4  2216748  2166619        2204377        1165582        519867

  Universitas  Total
0    375601  11104693
1    395554  10932000
2    409890  10547917
3    566588  10011142
4    626202   9427590

=== Tabel Data 2 ===
   Date  Value
0 31-Dec-82  3.00%
1 31-Dec-83   NaN
2 31-Dec-84   NaN
3 31-Dec-85  2.20%
4 31-Dec-86  2.70%

=== Tabel Data 3 ===
  Kelompok_Umur  TPT  Year
0      15-19    17.71  2015
1      20-24    12.86  2015
2      25-29    10.65  2015
3      30-34     8.86  2015
4      35-39     7.86  2015

=== Tabel Data 4 ===
  Year  Laki - Laki  Perempuan
0  2015         6.07         6.37
1  2016         5.70         5.45
2  2017         5.53         5.44
3  2018         5.34         5.25
4  2019         5.24         5.22

```

Pada hasil interpretasi ni terlihat bahwa kode menampilkan output data yang berbentuk csv dan excel. selain itu, pada hasil tersebut juga membaca dan menampilkan tabel yang berisi data tentang pengangguran, gender, umur dan pendidikan.



INTEGRASI DATA

Integrasi data adalah proses menggabungkan beberapa sumber data yang berbeda menjadi satu dataset yang lengkap, konsisten, dan mudah dianalisis. Tujuannya adalah agar data yang tersebar di berbagai file atau tabel dapat dipadukan sehingga memberikan gambaran yang utuh.

- 01 METODE DATA CLEANING
- 02 INTEGRASI SKEMA
- 03 INTEGRASI TRANSFORMATION
- 04 OUTER JOIN

PROSES WRANLING

● INTEGRASI DATA

Pada data ini output yang dihasilkan berhasil dilakukan, di mana kolom dari berbagai sumber data (umur, pendidikan, gender, dan TPT) telah digabung menjadi satu tabel.

Isi tabel mencakup:

- Kolom umur (15-19, 20-24, 25-29, dst.)
- Kolom pendidikan (SD, SLTP, SLTA, SMK, Akademi/Diploma, Universitas, Total)
- Kolom gender (Laki-laki, Perempuan)
- Nilai TPT (Value)

BERANDA

LAYANAN



```
Integrasi selesai! Kolom 60 Keatas & Rata-Rata sudah diatur posisinya.
  Year  15-19  20-24  25-29  30-34  35-39  40-44  45-49  50-54  55-59  ...  \
33  2015  17.71  12.86  10.65   8.86   7.86   7.59   6.97   5.93   6.05  ...
34  2016  28.09  15.80   7.08   3.63   2.21   2.05   1.35   1.66   1.55  ...
35  2017  27.54  16.62   6.76   3.40   2.45   1.86   1.51   1.54   1.73  ...
36  2018  26.93  16.79   6.97   3.44   2.48   1.80   1.58   1.39   1.25  ...
37  2019  26.12  15.64   7.19   3.52   2.25   2.06   1.81   1.65   1.30  ...
38  2020  24.34  18.71   9.77   5.75   4.32   3.92   3.54   3.61   3.21  ...
39  2021  23.91  17.73   9.26   5.43   4.02   3.42   3.30   2.18   1.98  ...
40  2022  29.08  17.02   7.13   3.70   2.65   2.43   2.33   2.38   2.37  ...

      SD      SLTP  SLTA Umum/SMU  SLTA Kejuruan/SMK  \
33  1162676.5  1.512153e+06  2.021220e+06      1372028.0
34  1127342.5  1.304149e+06  1.748662e+06      1434438.0
35  1098397.5  1.277828e+06  1.731862e+06      1502212.0
36   941944.5  1.203794e+06  1.809214e+06      1598790.5
37   915709.5  1.186197e+06  1.849281e+06      1568453.0
38  1208640.5  1.436435e+06  2.205639e+06      1885060.5
39  1306493.0  1.559768e+06  2.388976e+06      2100237.5
40  1259740.0  1.487278e+06  2.402635e+06      1733215.0

      Akademi/Diploma  Universitas      Total  Laki - Laki  Perempuan  Value
33   252926.500000     609494.0  7507794.5      6.07      6.37      4.51
34   234549.000000     631269.5  7027973.5      5.70      5.45      4.30
35   246321.000000     612848.5  7005262.0      5.53      5.44      3.78
36   264100.000000     771997.0  7018421.0      5.34      5.25      4.39
37   246665.500000     801104.0  7001610.0      5.24      5.22      3.59
38   286422.000000     903057.5  8346620.0      7.46      6.46      4.26
39   235240.500000     924100.0  8924030.0      6.74      6.11      3.83
40   184779.666667     743913.0  8418005.0      5.93      5.75      3.46
```

[8 rows x 23 columns]



CLEANING DATA

Data cleaning (pembersihan data) adalah proses memperbaiki data agar lebih akurat, konsisten, dan siap digunakan untuk analisis. Secara lebih spesifik, data cleaning mencakup kegiatan seperti menghapus data yang tercatat dua kali, mengisi atau menangani nilai yang kosong, serta memperbaiki format yang salah.

1. Standardisasi Nama Kolom

Membersihkan nama kolom dengan menghapus spasi dan mengganti karakter tidak valid agar lebih konsisten dan mudah diproses.

2. Konversi Tipe Data

Mengubah kolom menjadi tipe numerik untuk memastikan data dapat dihitung dengan benar.

3. Penanganan Missing Values

Mengisi nilai kosong (NaN) dengan angka 0 untuk menghindari error pada perhitungan.

4. Penghapusan Duplikasi

Menghapus baris data yang sama persis untuk mencegah data terhitung dua kali.

5. Filtering Data

Memilih hanya data tahun 2015–2022 agar analisis fokus pada periode yang relevan.

6. Pembulatan Angka

Membulatkan nilai numerik menjadi dua desimal agar data lebih rapi.

7. Validasi Data

Melakukan pengecekan ulang seperti statistik ringkas dan jumlah missing value untuk memastikan data siap dianalisis.

pada hasil ini output menampilkan statistik ringkas (descriptive statistics) setelah proses pembersihan data.

Statistik yang ditampilkan mencakup: count (jumlah data)

- mean (rata-rata)
- std (standar deviasi)
- min – max (nilai minimum dan maksimum)
- 25%, 50%, 75% (kuartil)

```
Tidak_belum_pernah_sekolah    0
Tidak_belum_tamat_SD         0
SD                             0
SLTP                           0
SLTA_Umum_SMU                 0
SLTA_Kejuruan_SMK             0
Akademi_Diploma               0
Universitas                   0
Total                          0
Laki__Laki                     0
Perempuan                     0
Value                          0
dtype: int64

=== Statistik Ringkas Setelah Cleaning ===

```

	Year	15_19	20_24	25_29	30_34	35_39
count	8.00000	8.000000	8.000000	8.000000	8.000000	8.000000
mean	2018.50000	25.465000	16.396250	8.101250	4.716250	3.530000
std	2.44949	3.598877	1.738595	1.536177	1.918101	1.928626
min	2015.00000	17.710000	12.860000	6.760000	3.400000	2.210000
25%	2016.75000	24.232500	15.760000	7.052500	3.500000	2.400000
50%	2018.50000	26.525000	16.705000	7.160000	3.665000	2.565000
75%	2020.25000	27.677500	17.197500	9.387500	5.510000	4.095000
max	2022.00000	29.080000	18.710000	10.650000	8.860000	7.860000

	40_44	45_49	50_54	55_59	...	SD
count	8.000000	8.000000	8.000000	8.000000	...	8.000000e+00
mean	3.141250	2.798750	2.542500	2.430000	...	1.127618e+06
std	1.956012	1.876261	1.543704	1.596988	...	1.401428e+05
min	1.800000	1.350000	1.390000	1.250000	...	9.157095e+05
25%	2.002500	1.562500	1.622500	1.487500	...	1.059284e+06
50%	2.245000	2.070000	1.920000	1.855000	...	1.145010e+06
75%	3.545000	3.360000	2.687500	2.580000	...	1.221415e+06
max	7.590000	6.970000	5.930000	6.050000	...	1.306493e+06

	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma
count	8.000000e+00	8.000000e+00	8.000000e+00	8.000000
mean	1.370950e+06	2.019686e+06	1.649304e+06	243875.521250
std	1.457100e+05	2.796015e+05	2.448723e+05	29200.744081
min	1.186197e+06	1.731862e+06	1.372028e+06	184779.670000
25%	1.259320e+06	1.794076e+06	1.485268e+06	235067.625000
50%	1.370292e+06	1.935250e+06	1.583622e+06	246493.250000
75%	1.493497e+06	2.251473e+06	1.771176e+06	255719.875000
max	1.559768e+06	2.402635e+06	2.100238e+06	286422.000000

	Universitas	Total	Laki__Laki	Perempuan	Value
count	8.000000	8.000000e+00	8.000000	8.000000	8.000000
mean	749722.937500	7.656214e+06	6.001250	5.756250	4.015000
std	125082.106868	7.874083e+05	0.757165	0.497822	0.397312
min	609494.000000	7.001610e+06	5.240000	5.220000	3.460000
25%	626664.250000	7.015131e+06	5.482500	5.392500	3.732500
50%	757955.000000	7.267884e+06	5.815000	5.600000	4.045000
75%	826592.375000	8.364466e+06	6.237500	6.175000	4.322500
max	924100.000000	8.924030e+06	7.460000	6.460000	4.510000

[8 rows x 23 columns]

Data berhasil dibersihkan dan disimpan sebagai 'Data_Terintegrasi_2015_2022_CLEAN.csv'



EKSPLORASI DATA

Exploratory Data Analysis (EDA) merupakan proses awal dalam analisis data yang bertujuan untuk Mengidentifikasi pola, struktur, dan elemen penting dalam data sebagai dasar sebelum melanjutkan ke tahap analisis statistik atau prediksi berikutnya.



EKSPLORASI TABEL

Dilakukan untuk melihat struktur dan karakteristik dataset menggunakan fungsi dari library pandas

● RINGKASAN DAN STATISTIK DATA

	Year	15_19	20_24	25_29	30_34	35_39	\
count	8.00000	8.000000	8.000000	8.000000	8.000000	8.000000	
mean	2018.50000	25.465000	16.396250	8.101250	4.716250	3.530000	
std	2.44949	3.598877	1.738595	1.536177	1.918101	1.928626	
min	2015.00000	17.710000	12.860000	6.760000	3.400000	2.210000	
25%	2016.75000	24.232500	15.760000	7.052500	3.500000	2.400000	
50%	2018.50000	26.525000	16.705000	7.160000	3.665000	2.565000	
75%	2020.25000	27.677500	17.197500	9.387500	5.510000	4.095000	
max	2022.00000	29.080000	18.710000	10.650000	8.860000	7.860000	

	40_44	45_49	50_54	55_59	...	SD	\
count	8.000000	8.000000	8.000000	8.000000	...	8.000000e+00	
mean	3.141250	2.798750	2.542500	2.430000	...	1.127618e+06	
std	1.956012	1.876261	1.543704	1.596988	...	1.401428e+05	
min	1.800000	1.350000	1.390000	1.250000	...	9.157095e+05	
25%	2.002500	1.562500	1.622500	1.487500	...	1.059284e+06	
50%	2.245000	2.070000	1.920000	1.855000	...	1.145010e+06	
75%	3.545000	3.360000	2.687500	2.580000	...	1.221415e+06	
max	7.590000	6.970000	5.930000	6.050000	...	1.306493e+06	

	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	\
count	8.000000e+00	8.000000e+00	8.000000e+00	8.000000	
mean	1.370950e+06	2.019686e+06	1.649304e+06	243875.521250	
std	1.457100e+05	2.796015e+05	2.448723e+05	29200.744081	
min	1.186197e+06	1.731862e+06	1.372028e+06	184779.670000	
25%	1.259320e+06	1.794076e+06	1.485268e+06	235067.625000	
50%	1.370292e+06	1.935250e+06	1.583622e+06	246493.250000	
75%	1.493497e+06	2.251473e+06	1.771176e+06	255719.875000	
max	1.559768e+06	2.402635e+06	2.100238e+06	286422.000000	

	Universitas	Total	Laki__Laki	Perempuan	Value
count	8.000000	8.000000e+00	8.000000	8.000000	8.000000
mean	749722.937500	7.656214e+06	6.001250	5.756250	4.015000
std	125082.106868	7.874083e+05	0.757165	0.497822	0.397312
min	609494.000000	7.001610e+06	5.240000	5.220000	3.460000
25%	626664.250000	7.015131e+06	5.482500	5.392500	3.732500
50%	757955.000000	7.267884e+06	5.815000	5.600000	4.045000
75%	826592.375000	8.364466e+06	6.237500	6.175000	4.322500
max	924100.000000	8.924030e+06	7.460000	6.460000	4.510000

● SHORTING DATA (PENDIDIKAN TERAKHIR)

	Year	15_19	20_24	25_29	30_34	35_39	40_44	45_49	50_54	55_59	...	SD	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	Universitas	Total	Laki__Laki	Perempuan	Value
7	2022	29.08	17.02	7.13	3.70	2.65	2.43	2.33	2.38	2.37	...	1259740.0	1487278.33	2402634.67	1733215.0	184779.67	743913.0	8418005.0	5.93	5.75	3.46
6	2021	23.91	17.73	9.26	5.43	4.02	3.42	3.30	2.18	1.98	...	1306493.0	1559768.50	2388976.00	2100237.5	235240.50	924100.0	8924030.0	6.74	6.11	3.83
5	2020	24.34	18.71	9.77	5.75	4.32	3.92	3.54	3.61	3.21	...	1208640.5	1436435.00	2205639.00	1885060.5	286422.00	903057.5	8346620.0	7.46	6.46	4.26
0	2015	17.71	12.86	10.65	8.86	7.86	7.59	6.97	5.93	6.05	...	1162676.5	1512153.00	2021220.00	1372028.0	252926.50	609494.0	7507794.5	6.07	6.37	4.51
4	2019	26.12	15.64	7.19	3.52	2.25	2.06	1.81	1.65	1.30	...	915709.5	1186197.00	1849281.00	1568453.0	246665.50	801104.0	7001610.0	5.24	5.22	3.59
3	2018	26.93	16.79	6.97	3.44	2.48	1.80	1.58	1.39	1.25	...	941944.5	1203794.50	1809213.50	1598790.5	264100.00	771997.0	7018421.0	5.34	5.25	4.39
1	2016	28.09	15.80	7.08	3.63	2.21	2.05	1.35	1.66	1.55	...	1127342.5	1304149.00	1748662.50	1434438.0	234549.00	631269.5	7027973.5	5.70	5.45	4.30
2	2017	27.54	16.62	6.76	3.40	2.45	1.86	1.51	1.54	1.73	...	1098397.5	1277828.50	1731861.50	1502212.0	246321.00	612848.5	7005262.0	5.53	5.44	3.78

● AGGREGATING DATA (PENDIDIKAN TERAKHIR)

	Tidak_belum_pernah_sekolah	Tidak_belum_tamat_SD	SD	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	Universitas
min	18421.33000	387031.5000	915709.5	1.186197e+06	1.731862e+06	1.372028e+06	184779.67000	609494.0000
max	89928.50000	588023.0000	1306493.0	1.559768e+06	2.402635e+06	2.100238e+06	286422.00000	924100.0000
mean	49400.47875	447848.0625	1127618.0	1.370950e+06	2.019686e+06	1.649304e+06	243875.52125	749722.9375
median	38312.00000	433173.5000	1145009.5	1.370292e+06	1.935250e+06	1.583622e+06	246493.25000	757955.0000

● SHORTING DATA (KELOMPOK UMUR)

	Year	15_19	20_24	25_29	30_34	35_39	40_44	45_49	50_54	55_59	...	SD	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	Universitas	Total	Laki__Laki	Perempuan	Value
7	2022	29.08	17.02	7.13	3.70	2.65	2.43	2.33	2.38	2.37	...	1259740.0	1487278.33	2402634.67	1733215.0	184779.67	743913.0	8418005.0	5.93	5.75	3.46
1	2016	28.09	15.80	7.08	3.63	2.21	2.05	1.35	1.66	1.55	...	1127342.5	1304149.00	1748662.50	1434438.0	234549.00	631269.5	7027973.5	5.70	5.45	4.30
2	2017	27.54	16.62	6.76	3.40	2.45	1.86	1.51	1.54	1.73	...	1098397.5	1277828.50	1731861.50	1502212.0	246321.00	612848.5	7005262.0	5.53	5.44	3.78
3	2018	26.93	16.79	6.97	3.44	2.48	1.80	1.58	1.39	1.25	...	941944.5	1203794.50	1809213.50	1598790.5	264100.00	771997.0	7018421.0	5.34	5.25	4.39
4	2019	26.12	15.64	7.19	3.52	2.25	2.06	1.81	1.65	1.30	...	915709.5	1186197.00	1849281.00	1568453.0	246665.50	801104.0	7001610.0	5.24	5.22	3.59
5	2020	24.34	18.71	9.77	5.75	4.32	3.92	3.54	3.61	3.21	...	1208640.5	1436435.00	2205639.00	1885060.5	286422.00	903057.5	8346620.0	7.46	6.46	4.26
6	2021	23.91	17.73	9.26	5.43	4.02	3.42	3.30	2.18	1.98	...	1306493.0	1559768.50	2388976.00	2100237.5	235240.50	924100.0	8924030.0	6.74	6.11	3.83
0	2015	17.71	12.86	10.65	8.86	7.86	7.59	6.97	5.93	6.05	...	1162676.5	1512153.00	2021220.00	1372028.0	252926.50	609494.0	7507794.5	6.07	6.37	4.51

● AGGREGATING DATA (KELOMPOK UMUR)

	15_19	20_24	25_29	30_34	35_39	40_44	45_49	50_54	55_59	60_Keatas
min	17.710	12.86000	6.76000	3.40000	2.210	1.80000	1.35000	1.3900	1.250	0.6100
max	29.080	18.71000	10.65000	8.86000	7.860	7.59000	6.97000	5.9300	6.050	4.7400
mean	25.465	16.39625	8.10125	4.71625	3.530	3.14125	2.79875	2.5425	2.430	2.0425
median	26.525	16.70500	7.16000	3.66500	2.565	2.24500	2.07000	1.9200	1.855	1.6100

● SHORTING DATA (JENIS KELAMIN)



	Year	15_19	20_24	25_29	30_34	35_39	40_44	45_49	50_54	55_59	...	SD	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	Universitas	Total	Laki___Laki	Perempuan	Value
5	2020	24.34	18.71	9.77	5.75	4.32	3.92	3.54	3.61	3.21	...	1208640.5	1436435.00	2205639.00	1885060.5	286422.00	903057.5	8346620.0	7.46	6.46	4.26
0	2015	17.71	12.86	10.65	8.86	7.86	7.59	6.97	5.93	6.05	...	1162676.5	1512153.00	2021220.00	1372028.0	252926.50	609494.0	7507794.5	6.07	6.37	4.51
6	2021	23.91	17.73	9.26	5.43	4.02	3.42	3.30	2.18	1.98	...	1306493.0	1559768.50	2388976.00	2100237.5	235240.50	924100.0	8924030.0	6.74	6.11	3.83
7	2022	29.08	17.02	7.13	3.70	2.65	2.43	2.33	2.38	2.37	...	1259740.0	1487278.33	2402634.67	1733215.0	184779.67	743913.0	8418005.0	5.93	5.75	3.46
1	2016	28.09	15.80	7.08	3.63	2.21	2.05	1.35	1.66	1.55	...	1127342.5	1304149.00	1748662.50	1434438.0	234549.00	631269.5	7027973.5	5.70	5.45	4.30
2	2017	27.54	16.62	6.76	3.40	2.45	1.86	1.51	1.54	1.73	...	1098397.5	1277828.50	1731861.50	1502212.0	246321.00	612848.5	7005262.0	5.53	5.44	3.78
3	2018	26.93	16.79	6.97	3.44	2.48	1.80	1.58	1.39	1.25	...	941944.5	1203794.50	1809213.50	1598790.5	264100.00	771997.0	7018421.0	5.34	5.25	4.39
4	2019	26.12	15.64	7.19	3.52	2.25	2.06	1.81	1.65	1.30	...	915709.5	1186197.00	1849281.00	1568453.0	246665.50	801104.0	7001610.0	5.24	5.22	3.59

● AGGREGATING DATA (JENIS KELAMIN)

	Laki___Laki	Perempuan
min	5.24000	5.22000
max	7.46000	6.46000
mean	6.00125	5.75625
median	5.81500	5.60000



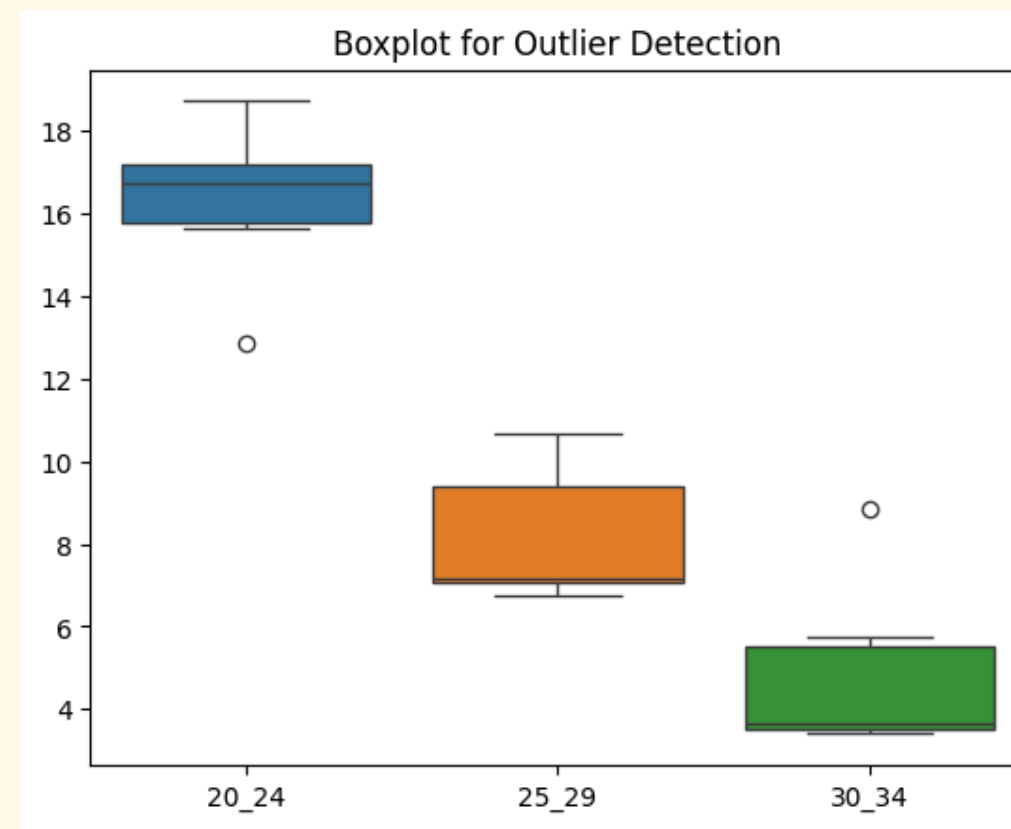
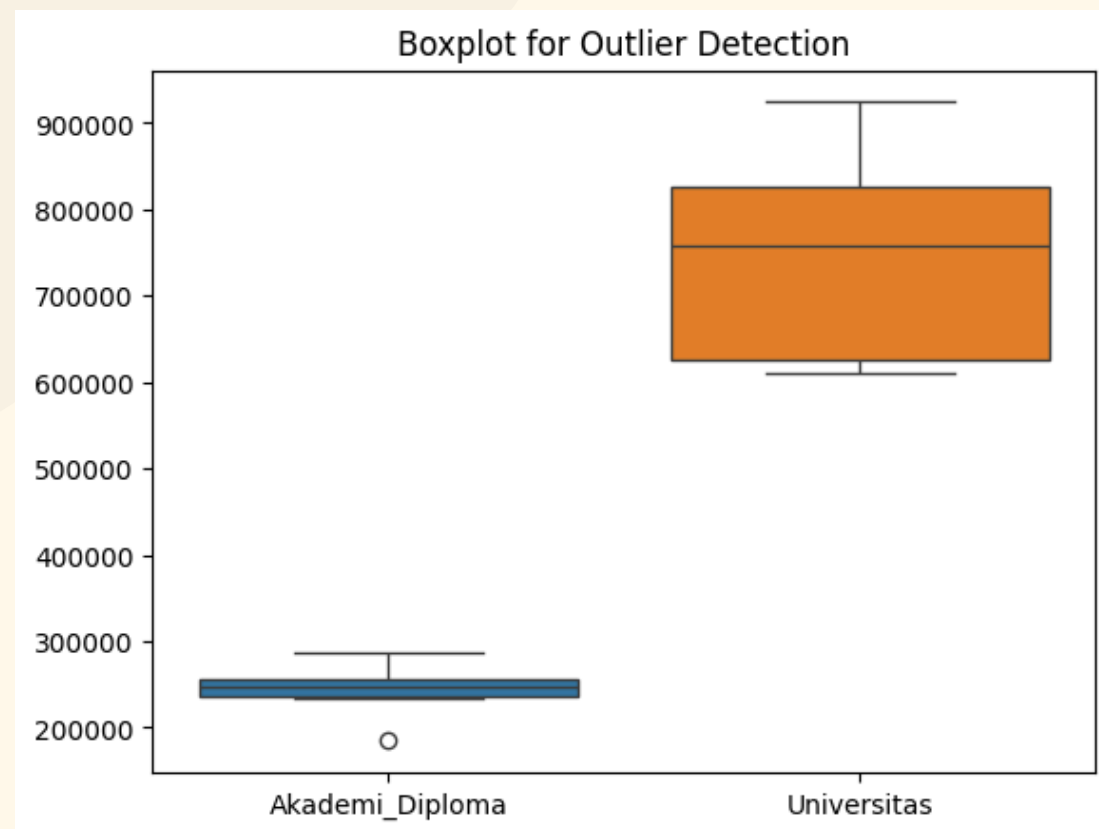
PENGECEKAN MISSING VALUES DAN DUPLICATES

Pengecekan missing values dilakukan untuk memastikan kelengkapan data, jika ditemukan ada nilai yang kosong pada setiap kolomnya, maka akan dilakukan proses imputasi maupun pembersihan data. Untuk pengecekan duplicates digunakan untuk menghitung jumlah baris yang nilainya sama lebih dari sekali

Berdasarkan output, tidak ditemukan missing values dan duplicates pada data, hal tersebut karena data sebelumnya sudah dilakukan proses cleaning

PENGECEKAN OUTLIER

Outlier atau nilai pencilan perlu dideteksi dalam proses eksplorasi data, sebab nilai yang menyimpang jauh dari sebaran data dapat memengaruhi hasil analisis dan berpotensi menghasilkan kesimpulan yang tidak akurat.



MENGIDENTIFIKASI KORELASI

Korelasi merupakan ukuran statistik yang digunakan untuk mengetahui hubungan antara variabel (positif, negatif, atau tidak ada hubungannya). Korelasi biasanya juga digunakan untuk memahami pola hubungan antar variabel.

	15_19	20_24	25_29	30_34	35_39	40_44	45_49	50_54	55_59	60_Keatas
15_19	1.000000	0.547551	-0.895975	-0.958531	-0.956932	-0.945225	-0.942859	-0.878509	-0.862485	-0.704088
20_24	0.547551	1.000000	-0.195756	-0.479279	-0.523162	-0.560738	-0.517466	-0.509441	-0.559597	-0.497382
25_29	-0.895975	-0.195756	1.000000	0.945861	0.912663	0.902732	0.911599	0.874466	0.830796	0.711699
30_34	-0.958531	-0.479279	0.945861	1.000000	0.993616	0.991453	0.988141	0.958349	0.945864	0.832228
35_39	-0.956932	-0.523162	0.912663	0.993616	1.000000	0.994334	0.993070	0.963346	0.961324	0.845805
40_44	-0.945225	-0.560738	0.902732	0.991453	0.994334	1.000000	0.994285	0.978130	0.973690	0.857825
45_49	-0.942859	-0.517466	0.911599	0.988141	0.993070	0.994285	1.000000	0.970342	0.963696	0.866661
50_54	-0.878509	-0.509441	0.874466	0.958349	0.963346	0.978130	0.970342	1.000000	0.991781	0.832556
55_59	-0.862485	-0.559597	0.830796	0.945864	0.961324	0.973690	0.963696	0.991781	1.000000	0.867297
60_Keatas	-0.704088	-0.497382	0.711699	0.832228	0.845805	0.857825	0.866661	0.832556	0.867297	1.000000





MENGIDENTIFIKASI KORELASI

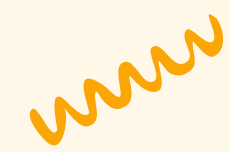
	Tidak_belum_pernah_sekolah	Tidak_belum_tamat_SD	SD	SLTP	SLTA_Umum_SMU	SLTA_Kejuruan_SMK	Akademi_Diploma	Universitas
Tidak_belum_pernah_sekolah	1.000000	0.179546	-0.193062	-0.163330	-0.665173	-0.808214	0.231360	-0.847899
Tidak_belum_tamat_SD	0.179546	1.000000	0.356956	0.295748	0.191862	-0.347971	-0.779991	-0.585094
SD	-0.193062	0.356956	1.000000	0.925743	0.785476	0.549486	-0.349201	0.219548
SLTP	-0.163330	0.295748	0.925743	1.000000	0.835394	0.490962	-0.266640	0.234620
SLTA_Umum_SMU	-0.665173	0.191862	0.785476	0.835394	1.000000	0.765863	-0.363981	0.606688
SLTA_Kejuruan_SMK	-0.808214	-0.347971	0.549486	0.490962	0.765863	1.000000	-0.018671	0.902418
Akademi_Diploma	0.231360	-0.779991	-0.349201	-0.266640	-0.363981	-0.018671	1.000000	0.213400
Universitas	-0.847899	-0.585094	0.219548	0.234620	0.606688	0.902418	0.213400	1.000000

	Laki____Laki	Perempuan
Laki____Laki	1.000000	0.877732
Perempuan	0.877732	1.000000



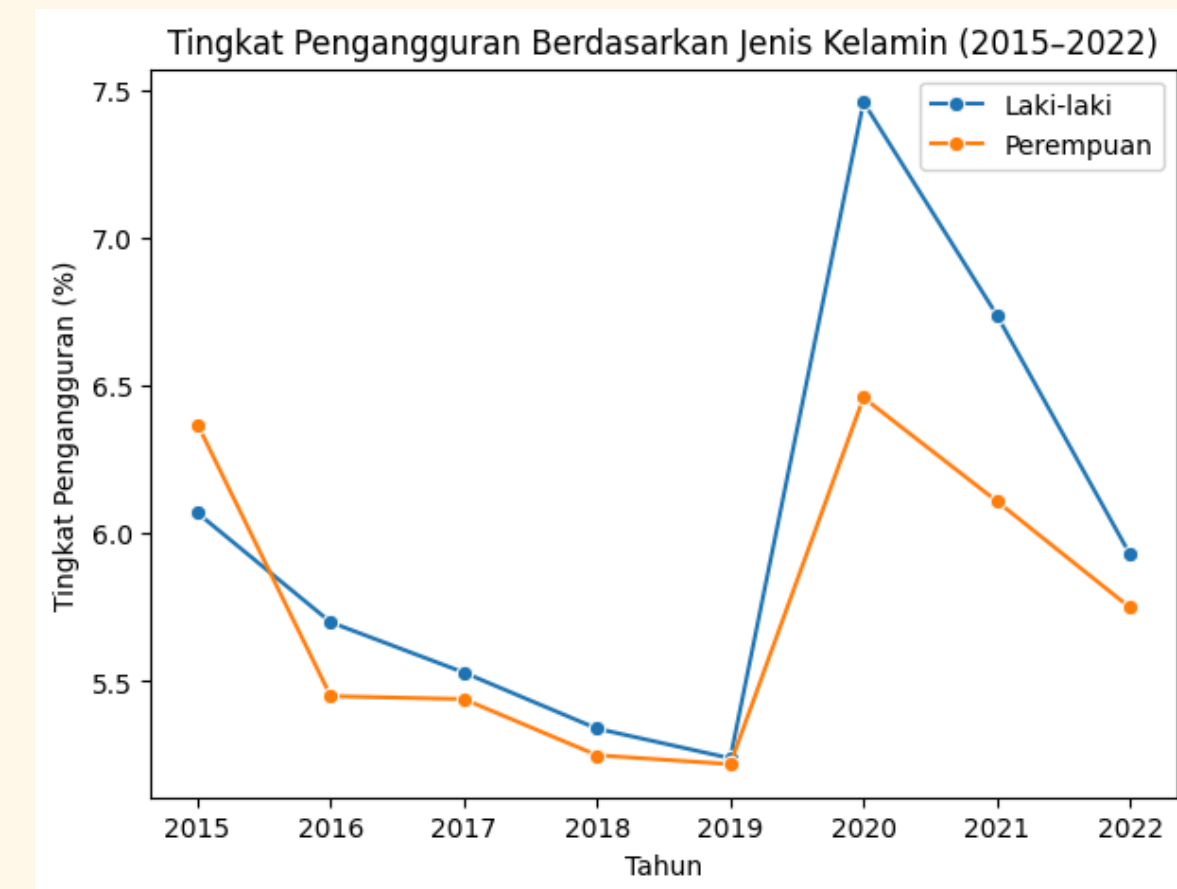
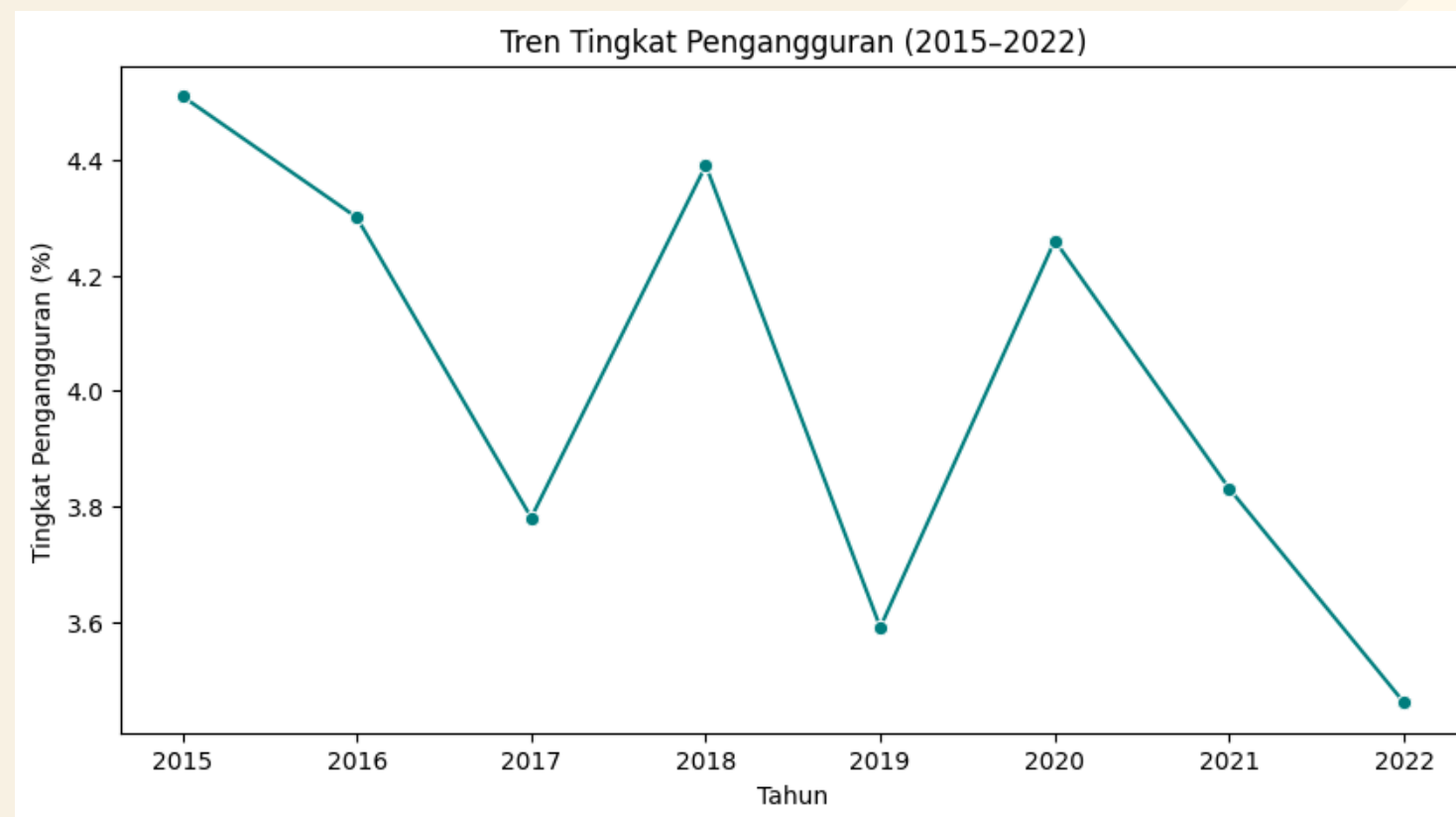
VISUALISASI

Visualisasi dalam EDA digunakan untuk memperkuat analisis dengan mengubah data menjadi gambar yang bermakna sehingga dapat menghasilkan pemahaman yang lebih mendalam. Visualisasi dalam EDA umumnya disajikan dalam grafik maupun diagram, visualisasi tersebut dapat menyajikan pola, tren, serta hubungan antar variabel dalam data.



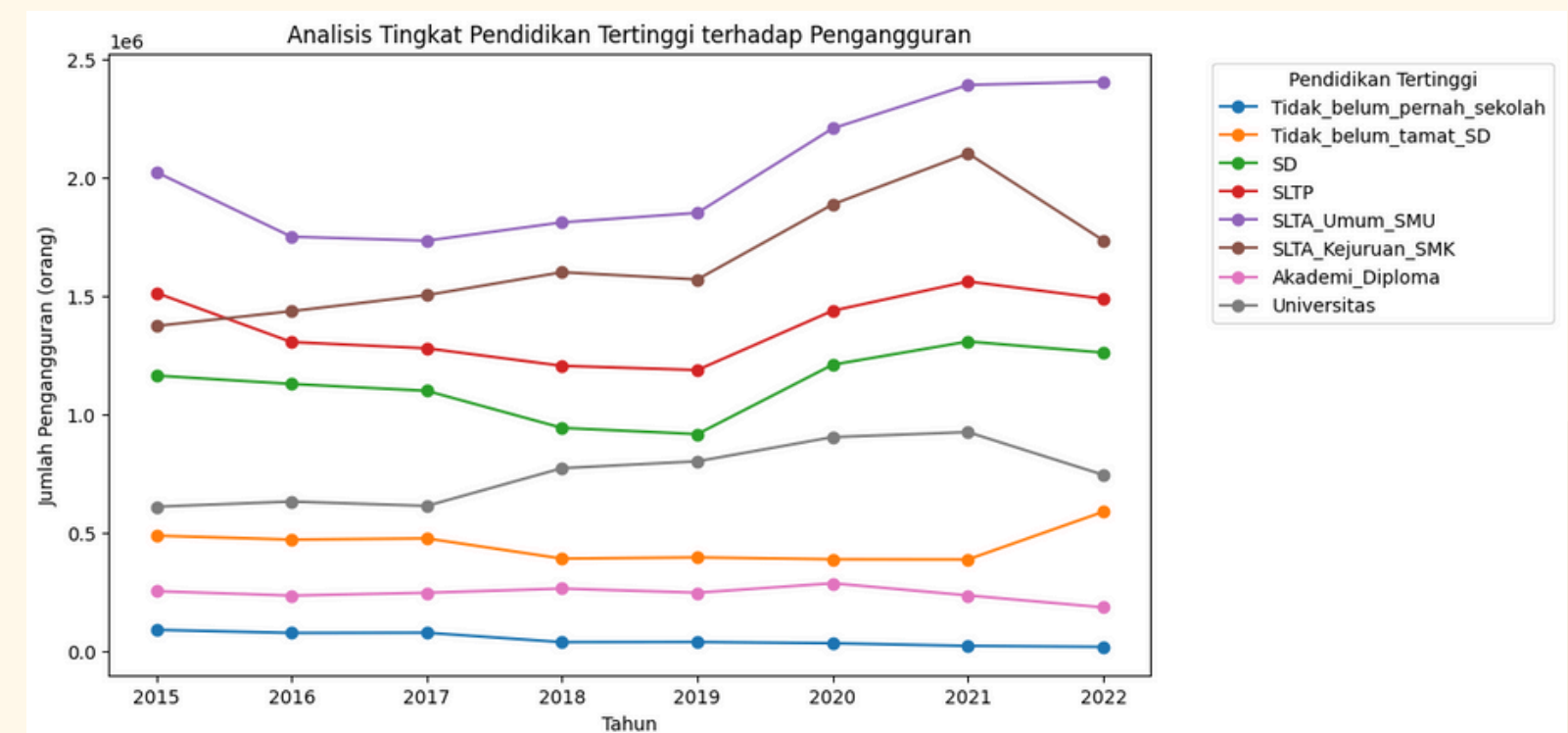
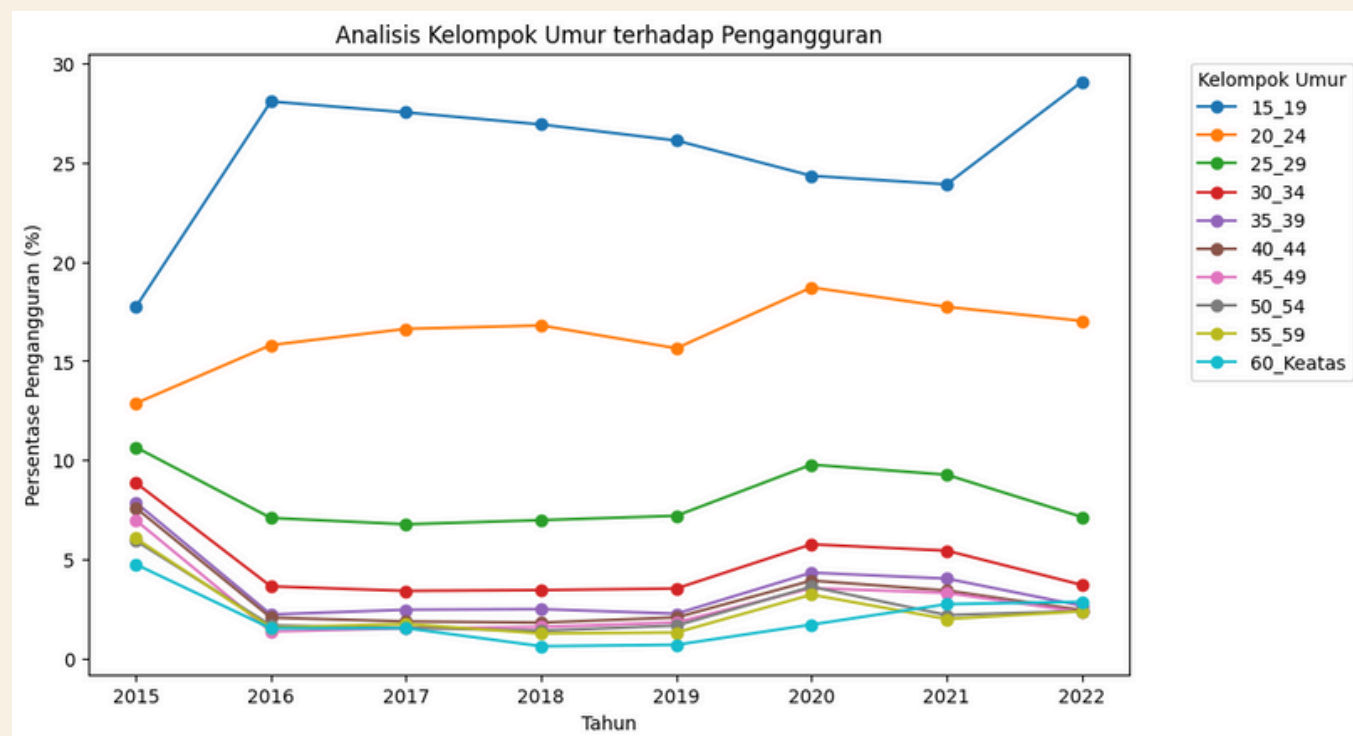
● LINE CHART

Digunakan untuk melihat tren waktu.



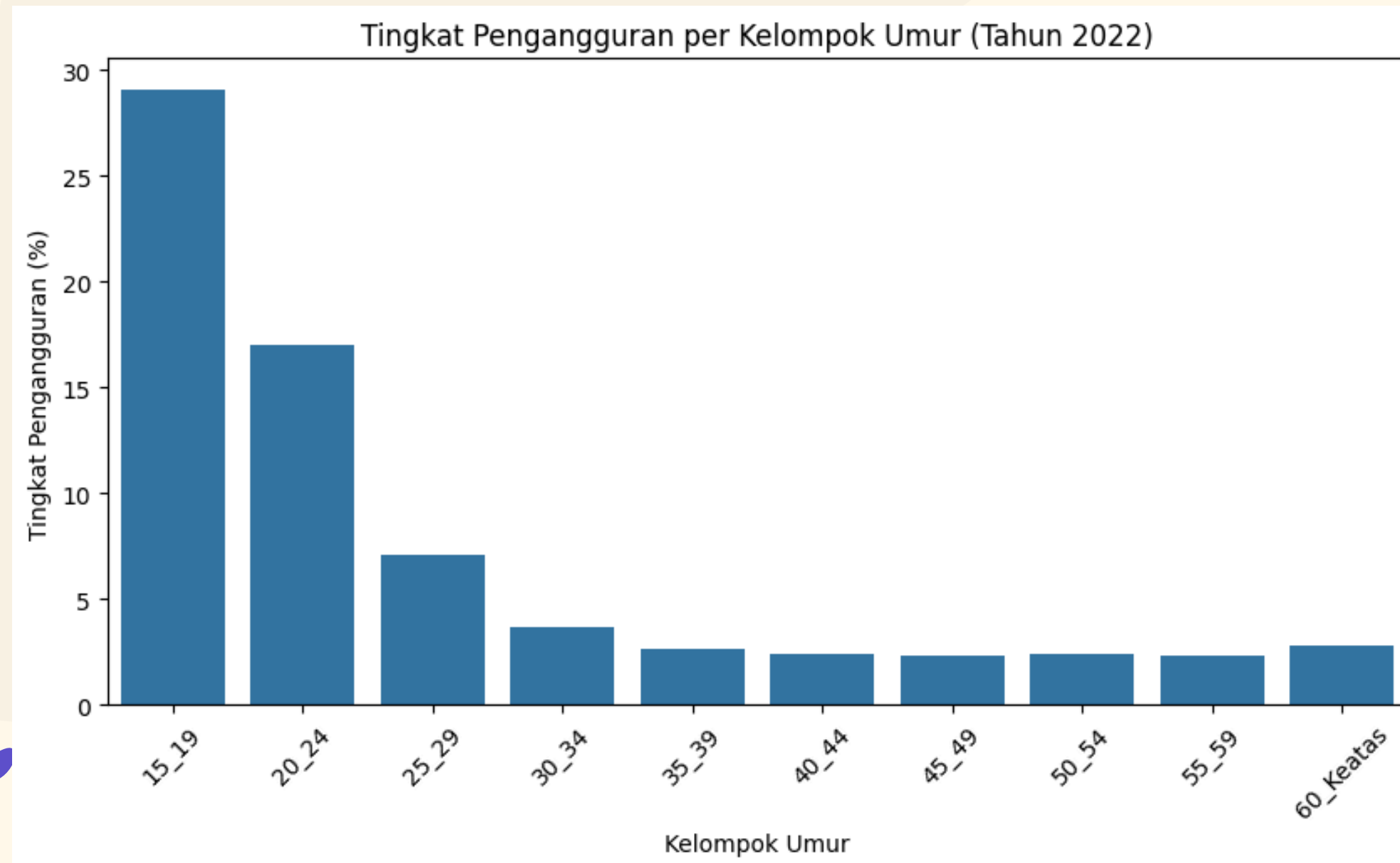
● LINE CHART

Digunakan untuk melihat tren waktu.

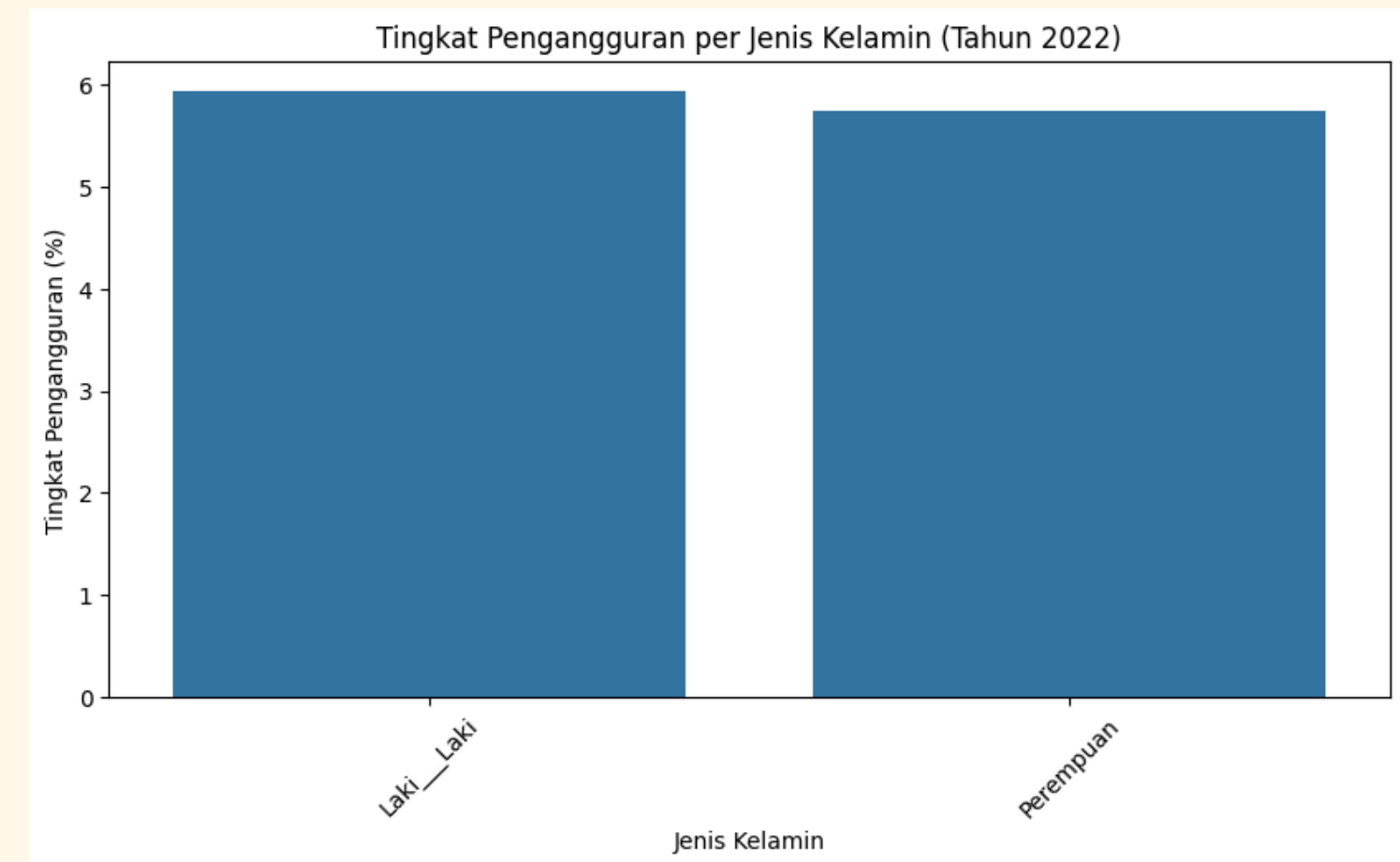
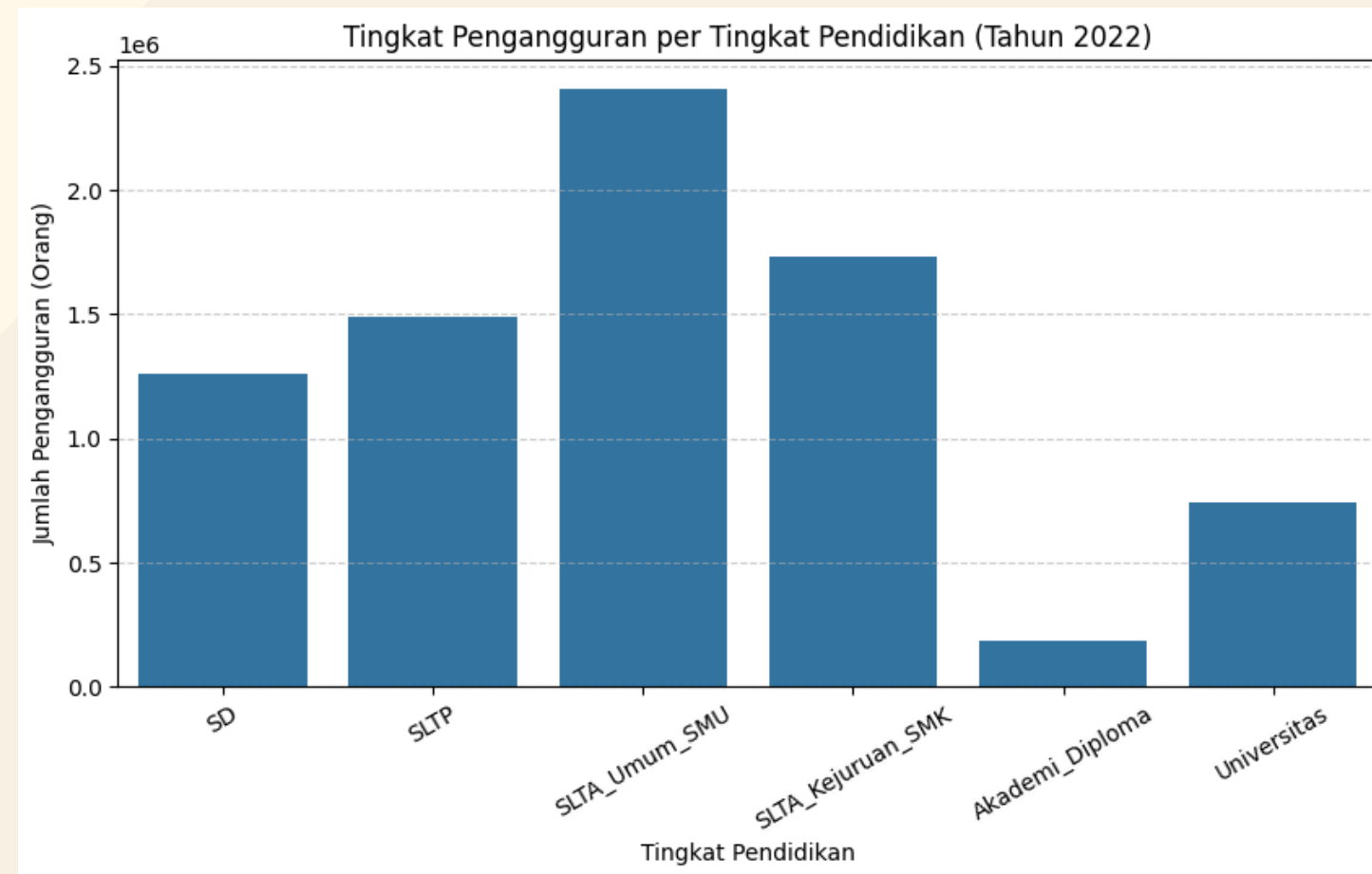


● BAR CHART

Digunakan untuk membandingkan nilai kategori tertentu pada periode tertentu.

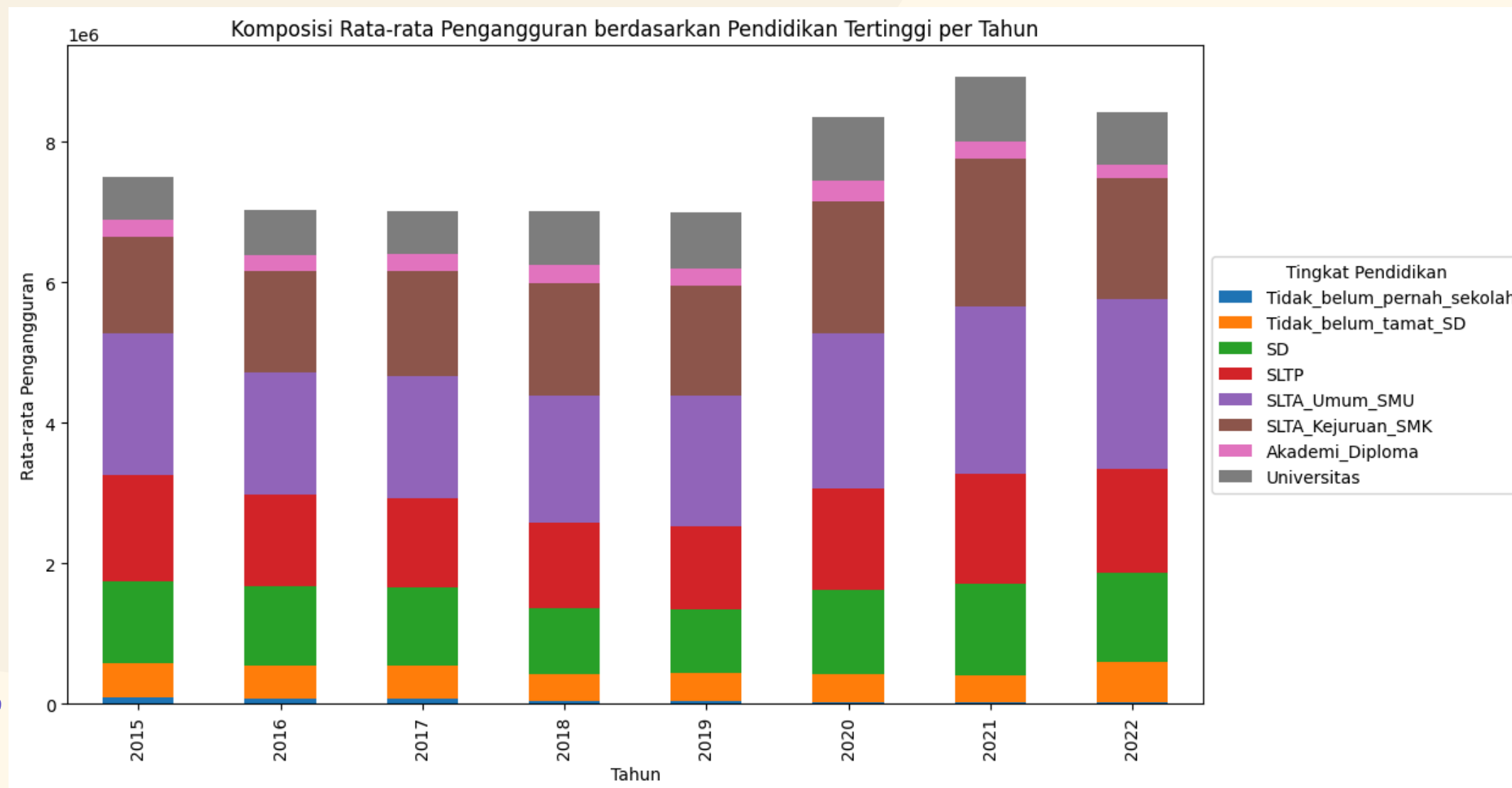


● BAR CHART

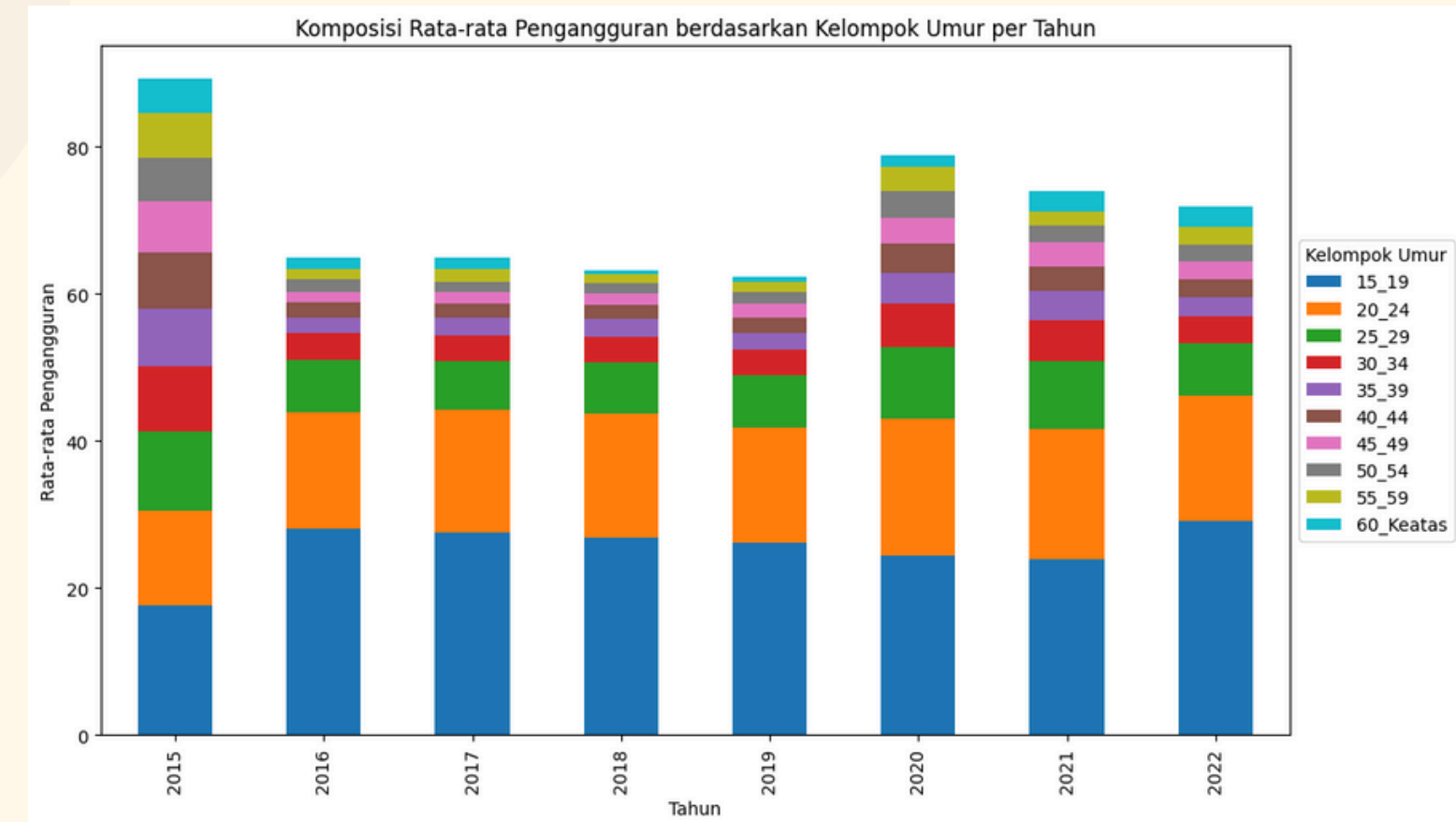
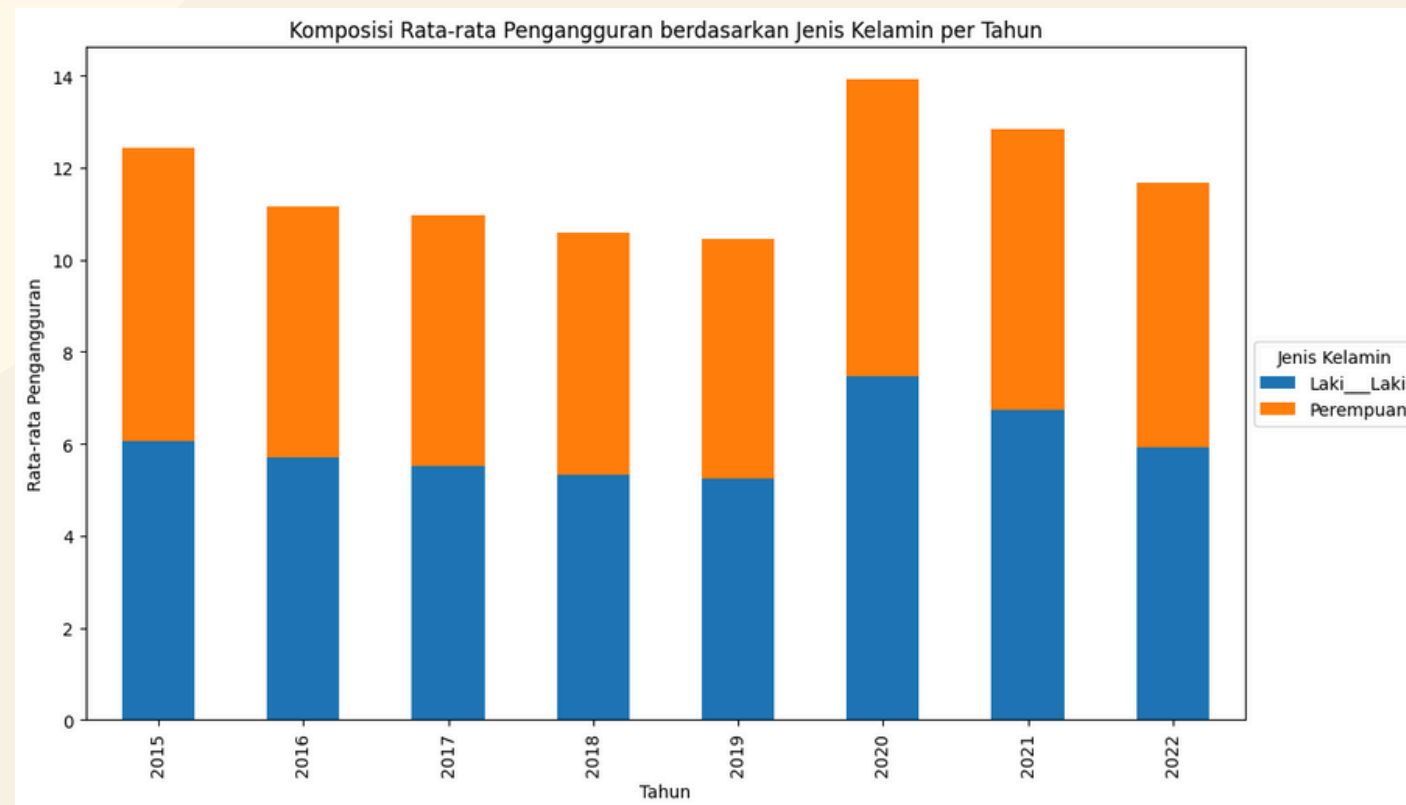


● STACKED BAR CHART

Digunakan untuk melihat perbandingan antar kategori dalam bentuk proporsi.

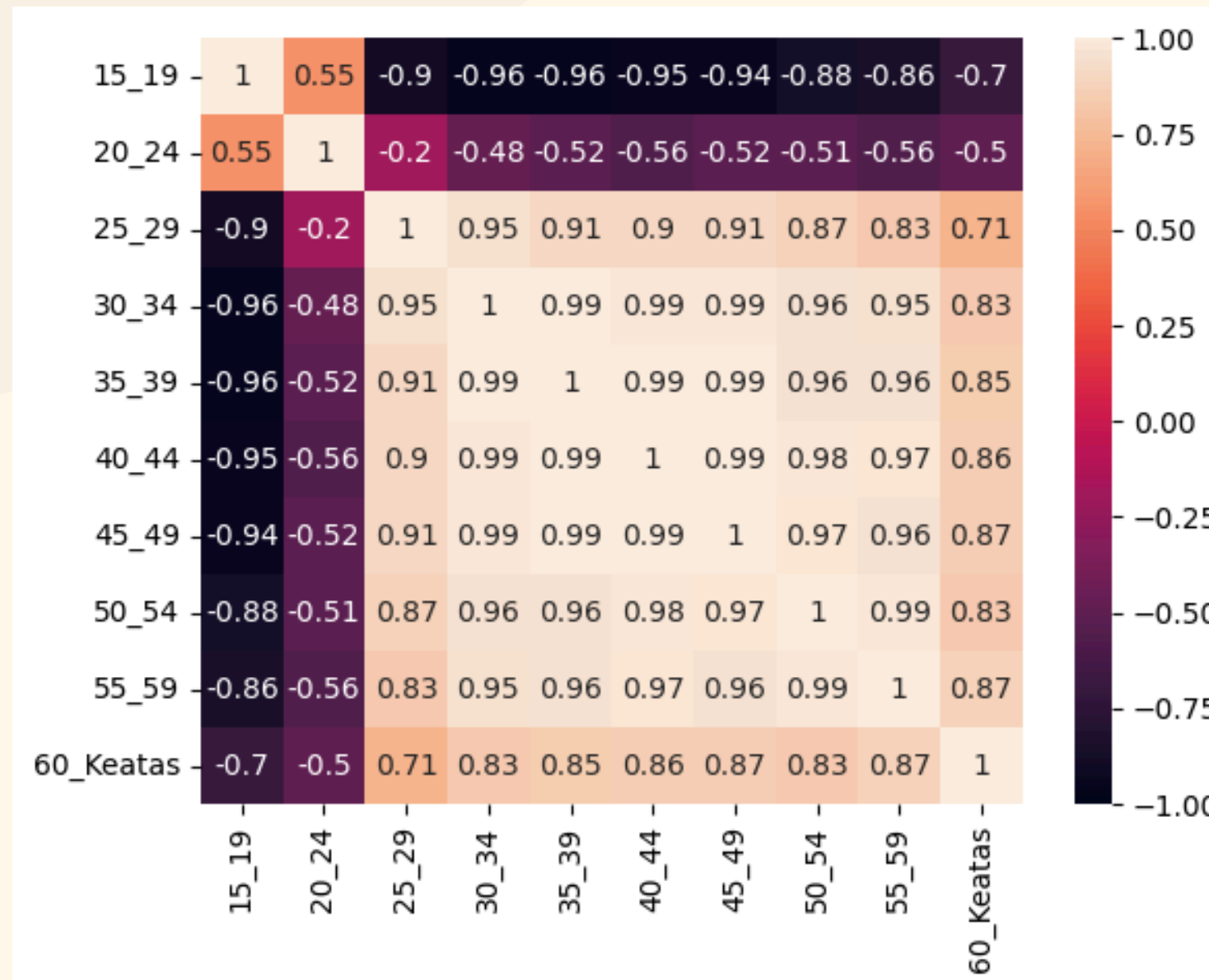


● STACKED BAR CHART

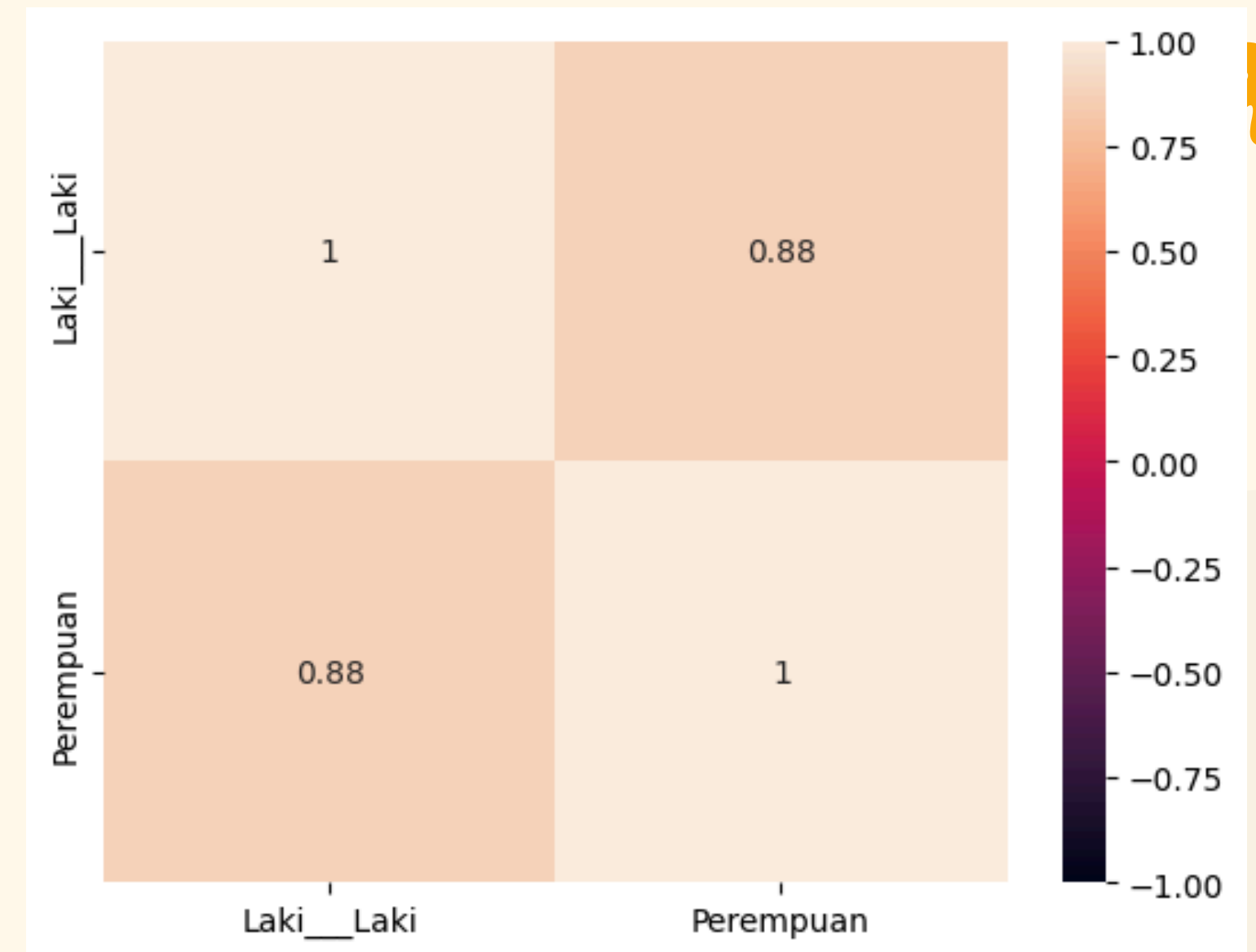
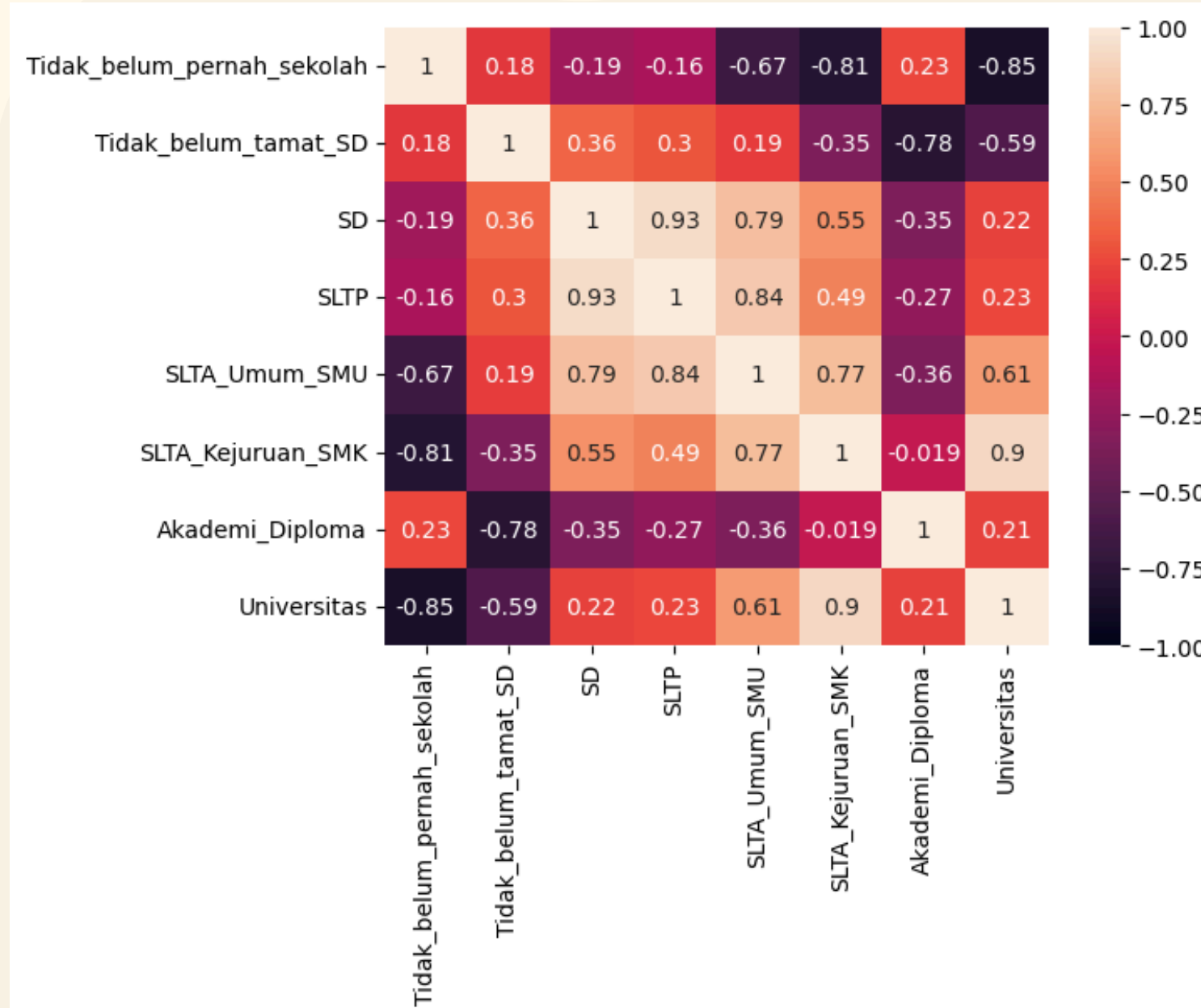


● HEATMAP

Digunakan untuk melihat korelasi antar variabel dalam bentuk visualisasi matriks berwarna.



● HEATMAP





DATA PUBLISHING

[Link GitHub Projek](#)
[Data Wrangling](#)



● DOKUMENTASI PIPELINE



Pengumpulan Data

■ Pengumpulan Data

- Import "Data_Pendidikan.csv"
- Import "Data_pengangguran.csv"
- Import "TPT_Age.csv"
- Import "TPT_Gender.csv"



Integrating Data

■ Integrating Data

- Menyeragamkan format numerik dan tahun
- Pivot dataset TPT Kelompok Umur
- Agregasi TPT Pendidikan per tahun
- Membersihkan simbol dan tipe data pada TPT Gender dan DataPengangguran
- Merge semua dataset ke dalam satu data integrasi



Cleaning Data

■ Cleaning Data

- Pembersihan nama kolom
- Konversi tipe data ke numerik
- Cek missing value
- Hapus duplikasi
- Filter tahun sesuai periode (2015-2022)
- Pembulatan
- Menyimpan dataset final



Eksplorasi Data

■ Eksplorasi Data

- Eksplorasi tabel (ringkasan dan struktur data)
- Cek missing values dan duplicate
- Finding Outlier
- Identifikasi Korelasi
- Visualisasi (line chart, bar chart, stacked bar chart, heatmap)

KENDALA

- 01 Perubahan ide dan data yang digunakan
- 02 Keterbatasan sumber data tingkat provinsi
- 03 Perubahan ide dan data yang digunakan
- 04 Kesulitan dalam menggabungkan data
- 05 Kesulitan dalam mencari dataset





RENCANA

- Melakukan Data Validation (Pemeriksaan Kembali Hasil Wrangling)
- Melakukan Feature Engineering secara mendalam
- Melakukan Transformasi Data



KESIMPULAN

Berdasarkan analisis TPT Indonesia tahun 2015–2022, tingkat pengangguran dipengaruhi oleh faktor umur, pendidikan, dan gender. Kelompok usia muda 15–24 tahun memiliki pengangguran tertinggi karena kurang pengalaman dan keterampilan, sementara lulusan SMA/SMK juga lebih rentan akibat ketidaksesuaian kompetensi dengan kebutuhan kerja. Laki-laki cenderung mengalami pengangguran lebih tinggi dibanding perempuan karena persaingan yang lebih ketat pada sektor-sektor pekerjaan yang didominasi tenaga kerja laki-laki. Secara keseluruhan, dinamika pengangguran sangat dipengaruhi karakteristik demografis sehingga kebijakan perlu disesuaikan dengan kelompok sasaran.

TERIMA KASIH PERHATIANNYA

ANALISIS PENGARUH FAKTOR DEMOGRAFIS TERHADAP TINGKAT
PENGANGGURAN TERBUKA (TPT) DI INDONESIA TAHUN 2015-2022

SAMPAI JUMPA LAGI 

SELENGKAPNYA

