

## EKONOMETRIKA TERAPAN

# REGRESI VARIABEL DUMMY

**Studi Kasus : Analisa Pengaruh Indikator Kependudukan dan  
Kemiskinan terhadap PDRB Provinsi Jawa Timur**

Dosen Pengampu : Ronny Susetyoko S.Si., M.Si



## Latar Belakang

Perekonomian suatu wilayah sangat dipengaruhi oleh berbagai faktor, termasuk kondisi penduduk dan tingkat kemiskinan di wilayah tersebut. Provinsi Jawa Timur, sebagai salah satu provinsi terbesar di Indonesia, memiliki karakteristik beragam dalam hal indikator kependudukan dan kemiskinan. **Penelitian ini bertujuan untuk menggali lebih dalam dampak indikator kependudukan dan kemiskinan terhadap Produk Domestik Regional Bruto (PDRB) di Provinsi Jawa Timur.**

Analisis ini menggunakan regresi variabel dummy untuk memahami bagaimana faktor-faktor seperti distribusi penduduk, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, persentase penduduk miskin, garis kemiskinan, dan pengeluaran riil per kapita dapat memengaruhi PDRB di wilayah tersebut. Selain itu, variabel dummy yang mengkategorikan wilayah sebagai Kabupaten atau Kota digunakan sebagai variabel independen.

**Melalui analisis ini, kita akan dapat mengidentifikasi faktor-faktor apa yang paling berpengaruh terhadap PDRB di Provinsi Jawa Timur dan apakah terdapat perbedaan signifikan antara Kabupaten dan Kota dalam kontribusi mereka terhadap pertumbuhan ekonomi regional.**



# Tujuan

Penelitian ini bertujuan untuk mencapai beberapa tujuan utama:

- 1. Menganalisis hubungan antara variabel-variabel seperti Distribusi Penduduk, Indeks Kedalaman Kemiskinan P1, Indeks Keparahan Kemiskinan P2, Persentase Penduduk Miskin, Garis Kemiskinan, Pengeluaran ril per Kapita, dan Kabupaten/Kota (variabel dummy) dengan PDRB di Provinsi Jawa Timur.**
- 2. Mengidentifikasi apakah variabel dummy "Kabupaten/Kota" mempengaruhi PDRB dan apakah terdapat perbedaan signifikan antara wilayah Kabupaten dan Kota dalam hubungan ini.**



# Dataset

Dalam penelitian ini dataset yang digunakan bersumber dari Tabel Dinamis pada website Badan Pusat Statistik Jawa Timur tahun 2022. Berikut penjelasan variabel yang digunakan :

- **Variabel X (Independen) :**
  - 1. Distribusi Penduduk :** Variabel ini menggambarkan cara distribusi penduduk di wilayah Jawa Timur. Distribusi penduduk dapat berpengaruh signifikan pada tingkat kemiskinan dan perkembangan ekonomi.
  - 2. Indeks Kedalaman Kemiskinan P1 :** Indeks ini digunakan untuk mengukur kedalaman kemiskinan di wilayah tersebut. Semakin dalam indeks ini, semakin parah kemiskinan yang dialami oleh penduduk.
  - 3. Indeks Keparahan Kemiskinan P2 :** Variabel ini mengukur sejauh mana kemiskinan di wilayah tersebut memengaruhi penduduk. Semakin tinggi nilai indeks ini, semakin parah dampak kemiskinan tersebut.
  - 4. Persentase Penduduk Miskin (PPM) :** Persentase penduduk miskin menggambarkan proporsi penduduk yang hidup di bawah garis kemiskinan. Ini adalah indikator penting dalam analisis kemiskinan.



# Dataset

**5. Garis Kemiskinan :** Garis kemiskinan adalah ambang batas pendapatan yang digunakan untuk mengklasifikasikan seseorang sebagai penduduk miskin atau tidak miskin.

**6. Pengeluaran Rata-rata per Kapita:** Variabel ini mengukur tingkat pengeluaran rata-rata per individu atau per kapita di wilayah tersebut. Ini dapat mencerminkan daya beli dan tingkat kesejahteraan masyarakat.

**7. Kabupaten/Kota (Variabel Dummy) :** Ini adalah variabel dummy yang mengidentifikasi apakah wilayah tersebut merupakan kabupaten atau kota. Variabel ini digunakan untuk mengkategorikan wilayah ke dalam dua kelompok.

- **Variabel Y (Dependen) :**

**1. PDRB (Produk Domestik Regional Bruto):** sebuah indikator yang mengukur total nilai produksi barang dan jasa dalam sebuah wilayah atau daerah.



# Hasil dan Pembahasan

## EDA - DESKRIPSI DATA

Sample Data

KabKota	Distribusi_Penduduk	Indeks_Kedalaman_Kemiskinan_P1	Indeks_Keparahan_Kemiskinan_P2	PPM	Garis_Kemiskinan	Pengeluaran_per_Kapita	PDRB
0	0	1.44	1.35	0.23	13.80	327758	9184 0.66
1	0	2.34	0.99	0.16	9.32	362461	10199 0.84
2	0	1.80	1.38	0.28	10.96	381448	10042 0.76
3	0	2.69	0.66	0.09	6.71	391888	11162 1.62
4	0	3.01	1.33	0.30	8.71	355511	11001 1.52

Statistika Deskriptif

KabKota	Distribusi_Penduduk	Indeks_Kedalaman_Kemiskinan_P1	Indeks_Keparahan_Kemiskinan_P2	PPM	Garis_Kemiskinan	Pengeluaran_per_Kapita	PDRB
count	38.000000	38.000000	38.000000	38.000000	38.000000	38.000000	38.000000
mean	0.236842	2.631053	1.502632	0.351842	10.325789	446972.631579	11837.973684 2.631842
std	0.430851	1.641064	0.735430	0.227689	4.270218	80226.343920	2237.915240 4.074330
min	0.000000	0.330000	0.400000	0.060000	3.790000	327758.000000	8944.000000 0.280000
25%	0.000000	1.710000	0.992500	0.220000	7.265000	392002.250000	10278.000000 0.785000
50%	0.000000	2.665000	1.345000	0.295000	9.630000	424584.500000	11564.000000 1.385000
75%	0.000000	3.230000	1.670000	0.407500	12.450000	509745.500000	12978.750000 3.200000
max	1.000000	7.020000	3.720000	1.160000	21.610000	652540.000000	18345.000000 24.010000

Informasi Kolom

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38 entries, 0 to 37
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   KabKota          38 non-null    int64  
 1   Distribusi_Penduduk 38 non-null  float64 
 2   Indeks_Kedalaman_Kemiskinan_P1 38 non-null  float64 
 3   Indeks_Keparahan_Kemiskinan_P2 38 non-null  float64 
 4   PPM              38 non-null    float64 
 5   Garis_Kemiskinan 38 non-null    int64  
 6   Pengeluaran_per_Kapita       38 non-null  int64  
 7   PDRB             38 non-null    float64 

dtypes: float64(5), int64(3)
memory usage: 2.5 KB

```

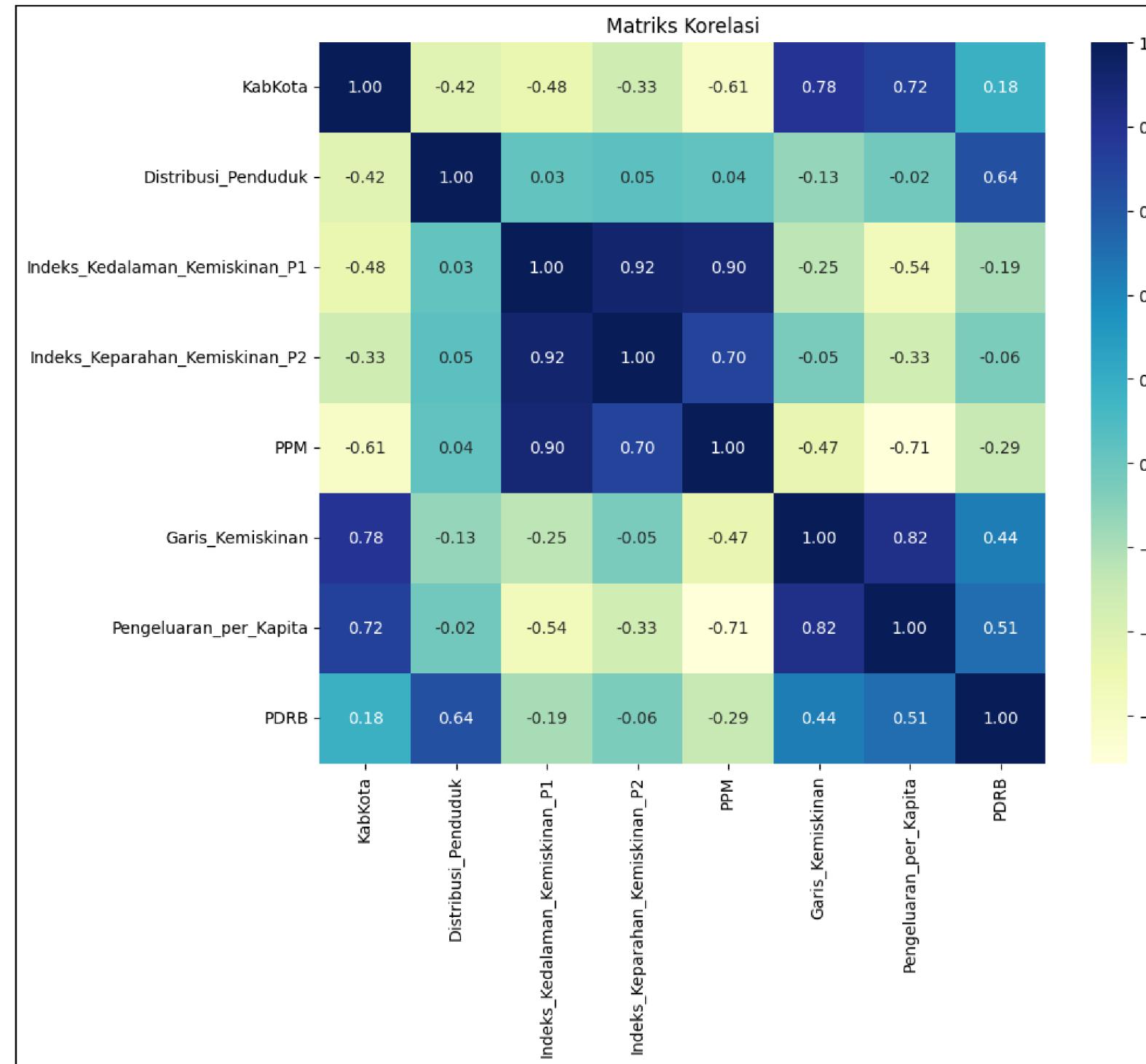
Mengecek Missing Values

KabKota	0
Distribusi_Penduduk	0
Indeks_Kedalaman_Kemiskinan_P1	0
Indeks_Keparahan_Kemiskinan_P2	0
PPM	0
Garis_Kemiskinan	0
Pengeluaran_per_Kapita	0
PDRB	0
dtype:	int64

Output diatas merupakan hasil analisis terhadap data yang digunakan, kita bisa melihat overview dari sampel data, mengamati statistik deskriptif, informasi kolom, serta ada atau tidaknya missing values pada data tersebut.



## EDA - MATRIKS KORELASI

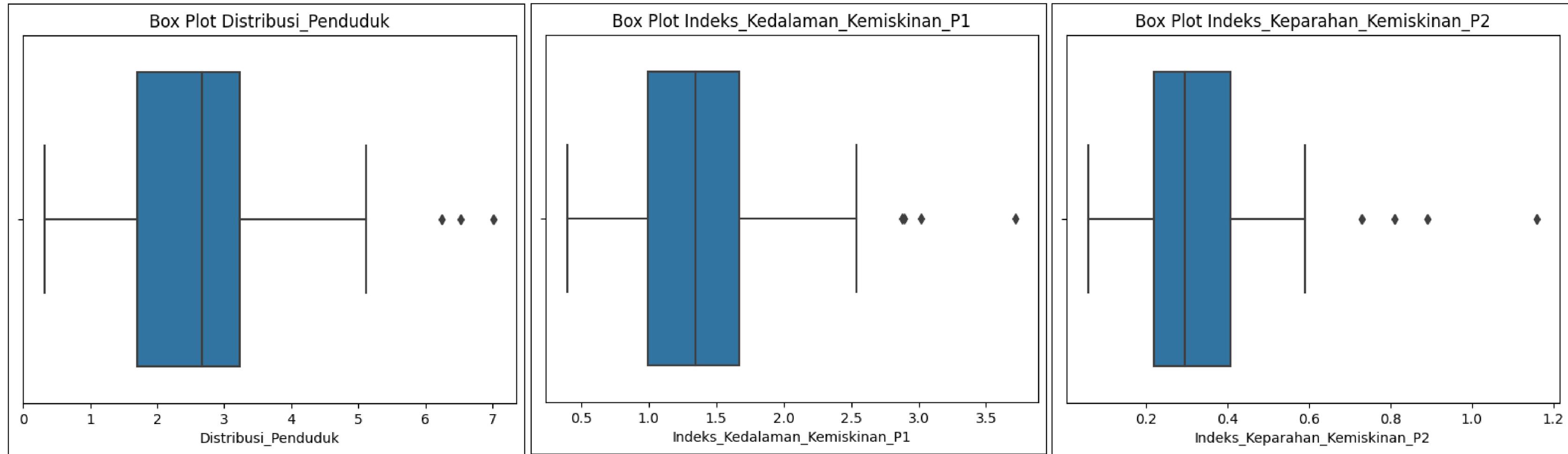


Matriks korelasi disamping dapat memberikan wawasan awal tentang hubungan antara variabel-variabel independen dan PDRB Provinsi Jawa Timur. Variabel 'Distribusi Penduduk', 'Garis Kemiskinan', dan 'Pengeluaran per Kapita' memiliki pengaruh positif yang kuat terhadap PDRB, sementara variabel lainnya memiliki pengaruh negatif atau relatif lemah.



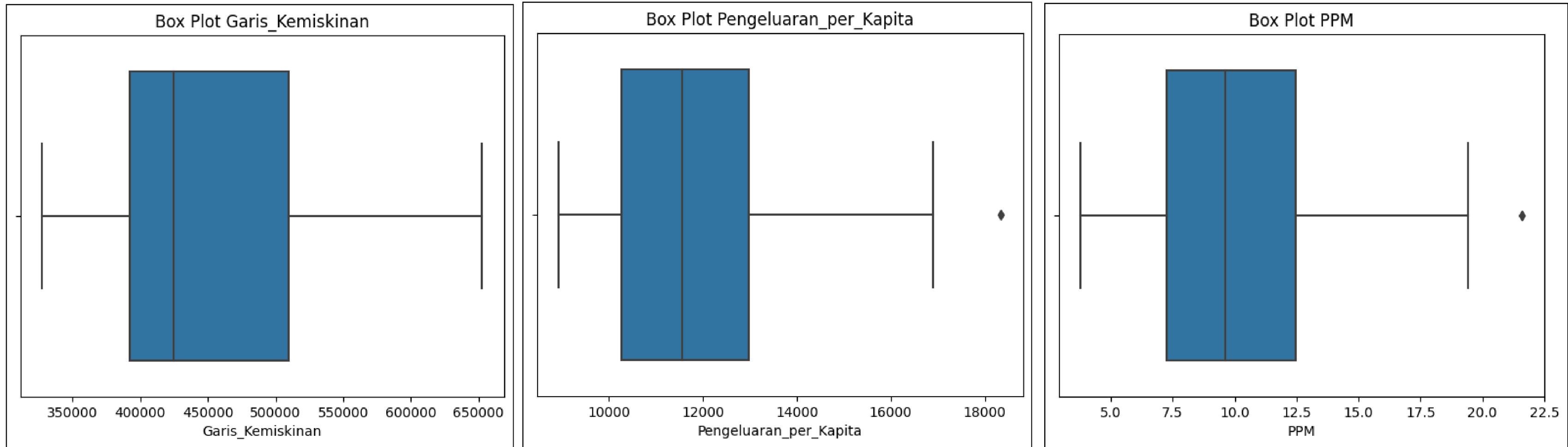
# Hasil dan Pembahasan

## EDA - ANALISA OUTLIER



## Hasil dan Pembahasan

### EDA - ANALISA OUTLIER



Berdasarkan box plot untuk analisa outlier dapat kita lihat bahwa masing-masing variabel independen yang digunakan antara lain Distribusi Penduduk, Indeks Kedalaman Kemiskinan P1, Indeks Keparahan Kemiskinan P2, dan Pengeluaran per Kapita memiliki nilai outlier. Hanya Garis Kemiskinan saja yang tidak memiliki nilai outlier.



## Hasil dan Pembahasan

# EDA - PENANGANAN OUTLIER

```
# Fungsi untuk menangani outlier dengan median
def handle_outliers_with_median(df, column_name, threshold=1.5):
    Q1 = df[column_name].quantile(0.25)
    Q3 = df[column_name].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - threshold * IQR
    upper_bound = Q3 + threshold * IQR
    df[column_name] = np.where((df[column_name] < lower_bound) | (df[column_name] > upper_bound), df[column_name].median(), df[column_name])

# Panggil fungsi handle_outliers_with_median untuk kolom tertentu
handle_outliers_with_median(df, 'Distribusi_Penduduk')
handle_outliers_with_median(df, 'Indeks_Kedalaman_Kemiskinan_P1')
handle_outliers_with_median(df, 'Indeks_Keparahan_Kemiskinan_P2')
handle_outliers_with_median(df, 'PPM')
handle_outliers_with_median(df, 'Pengeluaran_per_Kapita')
```

Kode program diatas digunakan untuk penanganan nilai outlier dimana nilai yang melebihi batas minimum/maksimum akan di-assign sebagai nilai median dari data di setiap variabel.



# Hasil dan Pembahasan

# **NORMALISASI DATA**

Distribusi_Penduduk	Indeks_Kedalaman_Kemiskinan_P1	Indeks_Keparahan_Kemiskinan_P2	PPM	Garis_Kemiskinan	Pengeluaran_per_Kapita	KabKota	PDRB
0	0.165919	0.286145	0.154545	0.561728	0.000000	0.025529	0 0.66
1	0.300448	0.177711	0.090909	0.310325	0.106850	0.133496	0 0.84
2	0.219731	0.295181	0.200000	0.402357	0.165311	0.116796	0 0.76
3	0.352765	0.078313	0.027273	0.163861	0.197456	0.235932	0 1.62
4	0.400598	0.280120	0.218182	0.276094	0.085451	0.218807	0 1.52

Normalisasi data dilakukan untuk menyamakan range nilai data setiap variabel karena sebelumnya satuan variabel bervariatif. Normalisasi ini menggunakan MinMaxScaler dimana range nilai setiap variabel akan di-assign antara nilai 0 dan 1.



## Hasil dan Pembahasan

### PARTISI VARIABEL X DAN Y

```
# Separate predictor and target variables  
  
x = result_df[['KabKota', 'Distribusi_Penduduk', 'Indeks_Kedalaman_Kemiskinan_P1', 'Indeks_Keparahan_Kemiskinan_P2', 'PPM', 'Garis_Kemiskinan','Pengeluaran_per_Kapita']]  
y = result_df[['PDRB']]
```

- **Variabel (x)** : Berisi variabel prediktor, yaitu 'KabKota', 'Distribusi\_Penduduk', 'Indeks\_Kedalaman\_Kemiskinan\_P1', 'Indeks\_Keparahan\_Kemiskinan\_P2', 'PPM', 'Garis\_Kemiskinan' dan 'Pengeluaran\_per\_Kapita'. Variabel ini akan digunakan untuk memprediksi nilai PDRB (produk domestik regional bruto).
- **Variabel (y)** : Berisi variabel target, yaitu 'PDRB'. Variabel target adalah apa yang akan diprediksi oleh model statistik atau mesin learning.



# Hasil dan Pembahasan

## REGRESI OLS

OLS Regression Results											
Dep. Variable:	PDRB	R-squared:	0.728								
Model:	OLS	Adj. R-squared:	0.665								
Method:	Least Squares	F-statistic:	11.48								
Date:	Sun, 15 Oct 2023	Prob (F-statistic):	5.41e-07								
Time:	19:48:31	Log-Likelihood:	-82.044								
No. Observations:	38	AIC:	180.1								
Df Residuals:	30	BIC:	193.2								
Df Model:	7										
Covariance Type:	nonrobust										
		coef	std err	t	P> t	[0.025	0.975]				
const		-5.7083	1.711	-3.336	0.002	-9.203	-2.213				
KabKota		2.1733	2.068	1.051	0.302	-2.050	6.397				
Distribusi_Penduduk		12.6401	2.010	6.288	0.000	8.535	16.746				
Indeks_Kedalaman_Kemiskinan_P1		-0.2109	11.733	-0.018	0.986	-24.173	23.751				
Indeks_Keparahan_Kemiskinan_P2		0.3792	7.796	0.049	0.962	-15.542	16.301				
PPM		2.2094	6.312	0.350	0.729	-10.682	15.101				
Garis_Kemiskinan		2.2920	3.943	0.581	0.565	-5.761	10.345				
Pengeluaran_per_Kapita		5.8330	3.836	1.521	0.139	-2.000	13.666				
Omnibus:	6.255	Durbin-Watson:	2.116								
Prob(Omnibus):	0.044	Jarque-Bera (JB):	5.103								
Skew:	0.635	Prob(JB):	0.0779								
Kurtosis:	4.269	Cond. No.	51.2								
<b>Notes:</b>											
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.											



## ANALISA REGRESI OLS

- Persamaan regresi linier hasil output adalah sebagai berikut :
  - **PDRB = -5.7083 + 2.1733 . KabKota + 12.6401 . Distribusi Penduduk - 0.2109 . Indeks\_Kedalaman\_Kemiskinan\_P1 + 0.3792 . Indeks\_Keparahan\_Kemiskinan\_P2 + 2.2094 . PPM + 2.2920 . Garis\_Kemiskinan + 5.8330 . Pengeluaran\_per\_Kapita**
- Interpretasi :
  1. **Dalam kasus ini, R-squared adalah 0.728, yang berarti sekitar 72.8% variasi dalam PDRB dapat dijelaskan oleh variabel-variabel dalam model.**
  2. **KabKota (Variabel Dummy) :**
    - Koefisien **KabKota** (2.1733) adalah koefisien regresi untuk variabel dummy **KabKota**. Ini menunjukkan pengaruh rata-rata perubahan **PDRB** ketika wilayah berubah dari Kabupaten (0) menjadi Kota (1).
    - Nilai **P> | t |** (0.302) untuk **KabKota** menunjukkan bahwa variabel dummy ini tidak signifikan pada tingkat signifikansi 0.05. Oleh karena itu, dalam model ini, perubahan dari Kabupaten ke Kota tidak memiliki pengaruh yang signifikan pada **PDRB**.



## Hasil dan Pembahasan

### ANALISA REGRESI OLS

- Interpretasi :

#### 3. Variabel Lain:

- Distribusi\_Penduduk (Variabel kontinu) memiliki koefisien regresi positif yang signifikan (12.6401), menunjukkan bahwa perubahan dalam distribusi penduduk memiliki dampak positif yang signifikan pada PDRB.
- Variabel lainnya, seperti Indeks\_Kedalaman\_Kemiskinan\_P1, Indeks\_Keparahan\_Kemiskinan\_P2, PPM, Garis\_Kemiskinan, dan Pengeluaran\_per\_Kapita, tidak menunjukkan pengaruh signifikan pada PDRB pada tingkat signifikansi 0.05.

#### 4. Dalam interpretasi hasil regresi, kita dapat menyusun pernyataan interpretatif untuk variabel Distribusi\_Penduduk, yang merupakan variabel yang signifikan dalam hubungannya dengan PDRB, seperti yang dijelaskan sebelumnya.

- Dengan asumsi bahwa semua variabel lainnya dalam model tetap konstan, setiap kenaikan satu poin persen dalam Distribusi\_Penduduk (misalnya dari 30% menjadi 31%) terkait dengan peningkatan rata-rata PDRB sebesar 12.6401 persen.



## Kesimpulan

### KESIMPULAN REGRESI OLS

Dalam analisis regresi ini, model menunjukkan bahwa **variabel "KabKota"** (sebagai variabel dummy untuk Kabupaten atau Kota) tidak memiliki pengaruh signifikan terhadap **Produk Domestik Regional Bruto (PDRB)** Provinsi Jawa Timur. Namun, variabel **"Distribusi Penduduk"** memiliki pengaruh positif yang signifikan pada PDRB. Setiap kenaikan satu poin persen dalam distribusi penduduk berhubungan dengan peningkatan rata-rata PDRB sebesar 12.64 persen. Variabel lainnya dalam model tidak memiliki pengaruh yang signifikan terhadap PDRB. **Keseluruhan, model ini dapat digunakan untuk memprediksi PDRB berdasarkan distribusi penduduk, meskipun faktor lain juga dapat memengaruhi PDRB yang tidak dimasukkan dalam model.**



## Penutup

**“Sejatinya orang yang mau belajar, tidak akan merasa pintar. Tapi semakin merasa dirinya bodoh.”**

**- Gus Iqdam**



# THANK YOU



Moch. Toriqul Muchlissin | 3321600001 | 3 D4 Sains Data Terapan