

LAPORAN PROYEK 2
KU1072 PENGENALAN TEKNOLOGI INFORMASI B
SEMESTER I 2018-2019

ANALISIS DESKRIPTIF SERTA VISUALISASI DATA PERTANIAN
DAN KEMISKINAN DENGAN PYTHON PANDAS DAN MATPLOTLIB

Disusun oleh:

Andhika Rahadian	16518123
Faris Muhammad Kautsar	16518327
Raras Pradnya Pramudita	16518021
Stefanus Gusega Gunawan	16518225
Fakultas :	STEI – K05 – Kelompok 4



PROGRAM TAHAP PERSIAPAN BERSAMA
INSTITUT TEKNOLOGI BANDUNG
NOVEMBER - 2018

DATA 1:
**Data Nilai Tukar Petani dan Indeks
Kepetanian di setiap provinsi di
Indonesia pada tahun 2008-2014**

BAB 1

DESKRIPSI DATA

Data ini menyajikan nilai tukar petani per provinsi di Indonesia dari tahun 2008 sampai tahun 2014. Nilai tukar petani (NTP) ini sendiri adalah perbandingan antara indeks harga yang diterima petani (it) dengan indeks harga yang dibayar petani (ib) yang dinyatakan dalam persentase. Nilai Tukar Petani (NTP) merupakan salah satu indikator untuk melihat tingkat kesejahteraan petani. Indeks ini juga dapat digunakan untuk menunjukkan daya tukar (*term of trade*) dari produk pertanian dengan barang dan jasa yang dikonsumsi maupun untuk biaya produksi. Sehingga, ada indikasi bahwa semakin tinggi NTP, relatif semakin sejahtera tingkat kehidupan petani.

Tabel ini diambil dalam format csv dari <https://data.go.id/dataset/nilai-tukar-petani-per-provinsi>, awalnya berisi 9 kolom dan 2475 baris. Namun setelah data ini dibersihkan dari data yang kosong, data ini memiliki 2400 baris. Data ini berukuran 167 KB.

```
import pandas as pd
dfkotor = pd.read_csv("processednilaitukarpetanitanamanpangan20082014.csv")
df = dfkotor.dropna()
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

print("Jumlah baris : ", df.shape[0])
print("Jumlah kolom : ", df.shape[1])

Jumlah baris : 2400
Jumlah kolom : 9
```

BAB 2 SAMPEL DATA

Berikut ini akan ditampilkan beberapa sampel data

A. Data di provinsi Jawa Barat pada tahun 2013

```
print ("Data di provinsi Jawa Barat pada tahun 2013")
print (df.loc[(df["nama_provinsi"] == "Prov. Jawa Barat") & (df["tahun"] == 2013)])
```

	tahun	bulan	kode_provinsi	nama_provinsi	it	ib	ntpp	\
1991	2013	1	32	Prov. Jawa Barat	165.37	149.80	110.40	
2024	2013	2	32	Prov. Jawa Barat	164.47	151.05	108.88	
2057	2013	3	32	Prov. Jawa Barat	162.38	151.94	106.87	
2090	2013	4	32	Prov. Jawa Barat	161.02	151.97	105.95	
2123	2013	5	32	Prov. Jawa Barat	160.75	151.85	105.86	
2156	2013	6	32	Prov. Jawa Barat	163.61	153.49	106.60	
2189	2013	7	32	Prov. Jawa Barat	168.96	158.16	106.83	
2222	2013	8	32	Prov. Jawa Barat	168.72	159.59	105.72	
2255	2013	9	32	Prov. Jawa Barat	170.35	159.69	106.68	
2288	2013	10	32	Prov. Jawa Barat	173.68	160.41	108.27	
2321	2013	11	32	Prov. Jawa Barat	174.43	160.31	108.81	
2354	2013	12	32	Prov. Jawa Barat	115.67	109.61	105.52	

	latitude	longitude
1991	-6.914722	107.6097
2024	-6.914722	107.6097
2057	-6.914722	107.6097
2090	-6.914722	107.6097
2123	-6.914722	107.6097
2156	-6.914722	107.6097
2189	-6.914722	107.6097
2222	-6.914722	107.6097
2255	-6.914722	107.6097
2288	-6.914722	107.6097
2321	-6.914722	107.6097
2354	-6.914722	107.6097

Dari data diatas, dapat jelas terlihat bahwa pada tahun 2013 di Provinsi Jawa Barat, nilai indeks yang dibayar petani memiliki kecenderungan untuk terus naik. Sedangkan nilai it dan ntp tidak stabil.

B. Data 10 teratas berdasarkan indeks yang diterima petani

```
df1 = df.sort_values(["it"], ascending = [0])
print ("Data 10 teratas berdasarkan indeks harga yang diterima petani")
print (df1[:10])
```

Data 10 teratas berdasarkan indeks harga yang diterima petani

	tahun	bulan	kode_provinsi	nama_provinsi	it	ib	ntpp	\
2251	2013	9	18	Prov. Lampung	197.25	145.92	135.17	
2284	2013	10	18	Prov. Lampung	196.33	146.63	133.90	
2317	2013	11	18	Prov. Lampung	195.91	146.30	133.90	
2218	2013	8	18	Prov. Lampung	195.65	146.01	134.00	
2185	2013	7	18	Prov. Lampung	193.56	143.77	134.64	
2152	2013	6	18	Prov. Lampung	190.09	140.01	135.77	
2119	2013	5	18	Prov. Lampung	187.40	139.43	134.40	
2086	2013	4	18	Prov. Lampung	185.91	139.41	133.35	
2053	2013	3	18	Prov. Lampung	185.40	139.44	132.96	
1954	2012	12	18	Prov. Lampung	185.28	135.56	136.67	

	latitude	longitude
2251	-5.429722	105.2625
2284	-5.429722	105.2625
2317	-5.429722	105.2625
2218	-5.429722	105.2625
2185	-5.429722	105.2625
2152	-5.429722	105.2625
2119	-5.429722	105.2625
2086	-5.429722	105.2625
2053	-5.429722	105.2625
1954	-5.429722	105.2625

Dari data diatas, dapat dilihat bahwa 10 nilai tertinggi indeks yang diterima petani adalah di Provinsi Lampung. Hal ini menunjukkan bahwa hasil tani petani-petani di Lampung memiliki nilai jual yang tinggi.

- C. Data 10 teratas berdasarkan indeks yang dibayar petani

```
df2 = df.sort_values(["ib"], ascending = [0])
```

```
print ("Data 10 teratas berdasarkan indeks harga yang dibayar petani")
```

```
print (df2[:10])
```

```
Data 10 teratas berdasarkan indeks harga yang dibayar petani
```

	tahun	bulan	kode_provinsi	nama_provinsi	it	ib	ntpp	\
2324	2013	11	35	Prov. Jawa Timur	170.42	162.56	104.83	
2291	2013	10	35	Prov. Jawa Timur	171.21	162.06	105.64	
2283	2013	10	17	Prov. Bengkulu	167.32	161.68	103.49	
2316	2013	11	17	Prov. Bengkulu	165.97	161.60	102.70	
2258	2013	9	35	Prov. Jawa Timur	168.37	161.58	104.20	
2225	2013	8	35	Prov. Jawa Timur	166.11	161.02	103.16	
2217	2013	8	17	Prov. Bengkulu	165.36	160.91	102.77	
2250	2013	9	17	Prov. Bengkulu	164.90	160.66	102.64	
2288	2013	10	32	Prov. Jawa Barat	173.68	160.41	108.27	
2321	2013	11	32	Prov. Jawa Barat	174.43	160.31	108.81	

	latitude	longitude
2324	-7.266667	112.7167
2291	-7.266667	112.7167
2283	-3.795556	102.2592
2316	-3.795556	102.2592
2258	-7.266667	112.7167
2225	-7.266667	112.7167
2217	-3.795556	102.2592
2250	-3.795556	102.2592
2288	-6.914722	107.6097
2321	-6.914722	107.6097

Dari data diatas, dapat dilihat bahwa mayoritas dari 10 nilai tertinggi indeks harga yang dibayar petani adalah dari provinsi-provinsi yang berada di Pulau Jawa. Hal ini menunjukkan bahwa biaya hidup di Pulau Jawa lebih tinggi disbanding pulau lain di Indonesia.

- D. Data 10 teratas berdasarkan nilai tukar petani

```
df3 = df.sort_values(["ntpp"], ascending = [0])
```

```
print ("Data 10 teratas berdasarkan Nilai Tukar Petani")
```

```
print (df3[:10])
```

	tahun	bulan	kode_provinsi	nama_provinsi	it	ib	ntpp	\
1921	2012	11	18	Prov. Lampung	184.96	135.15	136.85	
1855	2012	9	18	Prov. Lampung	183.85	134.40	136.79	
1954	2012	12	18	Prov. Lampung	185.28	135.56	136.67	
1888	2012	10	18	Prov. Lampung	184.07	134.94	136.41	
1822	2012	8	18	Prov. Lampung	181.70	133.81	135.79	
2152	2013	6	18	Prov. Lampung	190.09	140.01	135.77	
2251	2013	9	18	Prov. Lampung	197.25	145.92	135.17	
1987	2013	1	18	Prov. Lampung	184.76	137.04	134.83	
1789	2012	7	18	Prov. Lampung	179.18	132.93	134.79	
1591	2012	1	18	Prov. Lampung	176.07	130.71	134.70	

	latitude	longitude
1921	-5.429722	105.2625
1855	-5.429722	105.2625
1954	-5.429722	105.2625
1888	-5.429722	105.2625
1822	-5.429722	105.2625
2152	-5.429722	105.2625
2251	-5.429722	105.2625
1987	-5.429722	105.2625
1789	-5.429722	105.2625
1591	-5.429722	105.2625

Dari data diatas, dapat dilihat bahwa nilai tukar petani yang paling tinggi berasal dari Provinsi Lampung. Hal ini menunjukkan bahwa petani-petani di Lampung sejahtera, mapan, dan dapat mencukupi kebutuhan hidupnya dari hasil kerjanya sendiri.

BAB 3

MAKNA ATRIBUT DAN NILAI EKSTREMUM

Atribut yang terdapat dalam tabel data ini ada sembilan, yakni :

1. tahun : tahun diambilnya data tersebut. Data ini termasuk data kategorikal tipe nominal, dari tahun 2008 s.d. tahun 2014
2. bulan : bulan diambilnya data tersebut. Data ini termasuk data kategorikal tipe nominal, dari bulan 1 s.d. 12
3. kode_provinsi : kode provinsi ditulis berdasarkan data Kementrian Dalam Negeri RI. Data ini termasuk data kategorikal tipe nominal
4. nama_provinsi : nama provinsi tempat data tersbut diambil. Data ini termasuk data kategorikal tipe nominal
5. it : indeks harga yang diterima petani. Data ini termasuk data kuantitatif dengan range data dari 75.59 s.d. 197.25
6. ib : indeks harga yang dibayar petani. Data ini termasuk data kuantitatif dengan range data dari 102.2 s.d. 162.56
7. ntp : nilai tukar petani. Data ini termasuk data kuantitatif dengan range nilai data dari 63.39 s.d. 136.85
8. latitude : kedudukan provinsi tersebut di garis lintang. Data ini termasuk data kuantitatif dengan range nilai data dari -10.183 s.d. 5.55
9. longitude : kedudukan provinsi tersebut di garis bujur. Data ini termasuk data kuantitatif dengan range nilai data dari 95.316 s.d. 140.7169

```
print ("Nilai maksimum indeks harga yang diterima petani =", df["it"].max())
print ("Nilai minimum indeks harga yang diterima petani =", df["it"].min())
print ("Nilai maksimum indeks harga yang dibayar petani =", df["ib"].max())
print ("Nilai minimum indeks harga yang dibayar petani =", df["ib"].min())
print ("Nilai maksimum nilai tukar petani =", df["ntp"].max())
print ("Nilai minimum nilai tukar petani =", df["ntp"].min())
print ("Nilai maksimum latitude =", df["latitude"].max())
print ("Nilai minimum latitude =", df["latitude"].min())
print ("Nilai maksimum longitude =", df["longitude"].max())
print ("Nilai minimum longitude =", df["longitude"].min())
```

```
Nilai maksimum indeks harga yang diterima petani = 197.25
Nilai minimum indeks harga yang diterima petani = 75.59
Nilai maksimum indeks harga yang dibayar petani = 162.56
Nilai minimum indeks harga yang dibayar petani = 102.2
Nilai maksimum nilai tukar petani = 136.85
Nilai minimum nilai tukar petani = 63.39
Nilai maksimum latitude = 5.55
Nilai minimum latitude = -10.18333
Nilai maksimum longitude = 140.7169
Nilai minimum longitude = 95.31667
```

BAB 4

ANALISIS STATISTIK DATA

```
print ("Nilai rata-rata indeks harga yang diterima petani =", df["it"].mean())
print ("Standar deviasi indeks harga yang diterima petani =", df["it"].std())
print ("Nilai rata-rata indeks harga yang dibayar petani =", df["ib"].mean())
print ("Standar deviasi indeks harga yang dibayar petani =", df["ib"].std())
print ("Nilai rata-rata nilai tukar petani =", df["ntpp"].mean())
print ("Standar deviasi nilai tukar petani =", df["ntpp"].std())
print ("Nilai rata-rata latitude =", df["latitude"].mean())
print ("Standar deviasi latitude =", df["latitude"].std())
print ("Nilai rata-rata longitude =", df["longitude"].mean())
print ("Standar deviasi longitude =", df["longitude"].std())
```

Nilai rata-rata indeks harga yang diterima petani = 124.72559166666666
 Standar deviasi indeks harga yang diterima petani = 19.281728650672456
 Nilai rata-rata indeks harga yang dibayar petani = 127.86487916666668
 Standar deviasi indeks harga yang dibayar petani = 12.2434536917227
 Nilai rata-rata nilai tukar petani = 97.4670625
 Standar deviasi nilai tukar petani = 10.965210250805145
 Nilai rata-rata latitude = -2.863932190625
 Standar deviasi latitude = 3.679452362668222
 Nilai rata-rata longitude = 113.69997593749999
 Standar deviasi longitude = 10.61217078362237

Dari data yang diambil diatas, dapat disimpulkan bahwa di Indonesia, tingkat kesejahteraan petani relatif rendah. Hal ini ditunjukkan dari rata-rata nilai tukar petani yang kurang dari 100%, rata-rata indeks harga yang harus dibayar juga lebih tinggi daripada rata-rata indeks yang diterima. Artinya, hasil yang diterima petani lebih rendah dari pada harga yang harus dibayar untuk mencukupi kebutuhan hidupnya. Selain itu, standar deviasi indeks harga yang diterima petani cukup besar, hampir menyentuh 20%. Hal ini menunjukkan bahwa nilai tukar hasil tani di tiap-tiap daerah berbeda, tergantung daerahnya.

```
print ("Persentil 10% indeks harga yang diterima petani=", df["it"].quantile(0.1))
print ("Persentil 25% indeks harga yang diterima petani=", df["it"].quantile(0.25))
print ("Persentil 50% indeks harga yang diterima petani=", df["it"].quantile(0.5))
print ("Persentil 75% indeks harga yang diterima petani=", df["it"].quantile(0.75))
print ("Persentil 90% indeks harga yang diterima petani=", df["it"].quantile(0.9))
```

Persentil 10% indeks harga yang diterima petani= 103.69
 Persentil 25% indeks harga yang diterima petani= 111.4
 Persentil 50% indeks harga yang diterima petani= 121.09
 Persentil 75% indeks harga yang diterima petani= 136.84
 Persentil 90% indeks harga yang diterima petani= 151.07

Indeks harga yang diterima petani diperoleh dari perbandingan antara harga yang diterima petani pada tahun berlaku dengan harga tersebut pada tahun dasar. Pada persentil 10%, nilai berada diatas 100. Hal ini menunjukkan bahwa harga yang diterima petani lebih besar dari tahun sebelumnya.

```
print ("Persentil 10% indeks harga yang dibayar petani =", df["ib"].quantile(0.1))
print ("Persentil 25% indeks harga yang dibayar petani =", df["ib"].quantile(0.25))
print ("Persentil 50% indeks harga yang dibayar petani =", df["ib"].quantile(0.5))
print ("Persentil 75% indeks harga yang dibayar petani =", df["ib"].quantile(0.75))
print ("Persentil 90% indeks harga yang dibayar petani =", df["ib"].quantile(0.9))
```

Persentil 10% indeks harga yang dibayar petani = 110.9
 Persentil 25% indeks harga yang dibayar petani = 118.765
 Persentil 50% indeks harga yang dibayar petani = 127.765
 Persentil 75% indeks harga yang dibayar petani = 136.66750000000002
 Persentil 90% indeks harga yang dibayar petani = 144.20299999999997

Sedangkan untuk indeks harga yang dibayar petani, merupakan perbandingan antara harga yang dibayarkan petani pada tahun berlaku dengan harga yang dibayarkan petani pada tahun dasar. Persentil 10% data ib nilai nya lebih besar dari 100%, serta nilainya lebih besar dari persentil 10% it. Hal ini menunjukkan bahwa ib cenderung selalu lebih besar daripada it, atau artinya, harga yang diterima petani lebih kecil daripada harga yang harus dibayar petani.

```
print ("Persentil 10% nilai tukar petani =", df["ntpp"].quantile(0.1))
print ("Persentil 25% nilai tukar petani =", df["ntpp"].quantile(0.25))
print ("Persentil 50% nilai tukar petani =", df["ntpp"].quantile(0.5))
print ("Persentil 75% nilai tukar petani =", df["ntpp"].quantile(0.75))
print ("Persentil 90% nilai tukar petani =", df["ntpp"].quantile(0.9))
```

```
Persentil 10% nilai tukar petani = 85.89900000000002
Persentil 25% nilai tukar petani = 90.9
Persentil 50% nilai tukar petani = 96.56
Persentil 75% nilai tukar petani = 103.58500000000001
Persentil 90% nilai tukar petani = 110.83099999999999
```

Nilai tukar petani (ntpp) diperoleh dari perbandingan nilai it dan ib. Seperti yang sudah disimpulkan dari data sebelumnya, karena nilai ib cenderung lebih besar dari nilai it, maka ntp akan lebih kecil daripada 100. Hal ini menunjukkan bahwa harga yang harus dibayar petani lebih besar daripada harga yang diterima. Maka, dapat disimpulkan bahwa tingkat kesejahteraan petani di beberapa daerah di Indonesia masih rendah.

```
print ("Persentil 10% latitude =", df["latitude"].quantile(0.1))
print ("Persentil 25% latitude =", df["latitude"].quantile(0.25))
print ("Persentil 50% latitude =", df["latitude"].quantile(0.5))
print ("Persentil 75% latitude =", df["latitude"].quantile(0.75))
print ("Persentil 90% latitude =", df["latitude"].quantile(0.9))
Persentil 10% latitude = -7.801389
Persentil 25% latitude = -5.602291500000001
Persentil 50% latitude = -2.6008335000000002
Persentil 75% latitude = -0.380833325
Persentil 90% latitude = 1.083333
print ("Persentil 10% longitude =", df["longitude"].quantile(0.1))
print ("Persentil 25% longitude =", df["longitude"].quantile(0.25))
print ("Persentil 50% longitude =", df["longitude"].quantile(0.5))
print ("Persentil 75% longitude =", df["longitude"].quantile(0.75))
print ("Persentil 90% longitude =", df["longitude"].quantile(0.9))
Persentil 10% longitude = 101.4686
Persentil 25% longitude = 105.13605000000001
Persentil 50% longitude = 113.31835000000001
Persentil 75% longitude = 120.52364999999999
Persentil 90% longitude = 127.36669999999998
```

Untuk nilai latitude dan longitude, nilai-nilai ini hanya untuk menunjukkan kedudukan provinsi yang diambil datanya.

```
print(df["tahun"].value_counts())
2013    384
2011    384
2009    384
2012    384
2010    384
2008    384
2014     96
Name: tahun, dtype: int64
```

Jika seluruh ke-34 provinsi diambil datanya bulan 1 s.d. 12, maka seharusnya ada 408 data. Namun dari data yang dilihat diatas, ada kekurangan 24 data di tahun 2008 s.d. 2013, dan hanya ada 96 data yang diambil pada tahun 2014.

BAB 5

VISUALISASI DAN KORELASI DATA

```
jawabarat2008=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2008)]
print('Untuk Jawa Barat tahun 2008, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2008['it'].corr(jawabarat2008['bulan'])))
print('ib: ' +str(jawabarat2008['ib'].corr(jawabarat2008['bulan'])))
print('ntpp: ' + str(jawabarat2008['ntpp'].corr(jawabarat2008['bulan'])))
print('latitude: ' + str(jawabarat2008['latitude'].corr(jawabarat2008['bulan'])))
print('longitude: ' + str(jawabarat2008['longitude'].corr(jawabarat2008['bulan'])))

jawabarat2009=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2009)]
print('Untuk Jawa Barat tahun 2009, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2009['it'].corr(jawabarat2009['bulan'])))
print('ib: ' +str(jawabarat2009['ib'].corr(jawabarat2009['bulan'])))
print('ntpp: ' + str(jawabarat2009['ntpp'].corr(jawabarat2009['bulan'])))
print('latitude: ' + str(jawabarat2009['latitude'].corr(jawabarat2009['bulan'])))
print('longitude: ' + str(jawabarat2009['longitude'].corr(jawabarat2009['bulan'])))

jawabarat2010=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2010)]
print('Untuk Jawa Barat tahun 2010, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2010['it'].corr(jawabarat2010['bulan'])))
print('ib: ' +str(jawabarat2010['ib'].corr(jawabarat2010['bulan'])))
print('ntpp: ' + str(jawabarat2010['ntpp'].corr(jawabarat2010['bulan'])))
print('latitude: ' + str(jawabarat2010['latitude'].corr(jawabarat2010['bulan'])))
print('longitude: ' + str(jawabarat2010['longitude'].corr(jawabarat2010['bulan'])))

jawabarat2011=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2011)]
print('Untuk Jawa Barat tahun 2011, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2011['it'].corr(jawabarat2011['bulan'])))
print('ib: ' +str(jawabarat2011['ib'].corr(jawabarat2011['bulan'])))
print('ntpp: ' + str(jawabarat2011['ntpp'].corr(jawabarat2011['bulan'])))
print('latitude: ' + str(jawabarat2011['latitude'].corr(jawabarat2011['bulan'])))
print('longitude: ' + str(jawabarat2011['longitude'].corr(jawabarat2011['bulan'])))

jawabarat2012=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2012)]
print('Untuk Jawa Barat tahun 2012, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2012['it'].corr(jawabarat2012['bulan'])))
print('ib: ' +str(jawabarat2012['ib'].corr(jawabarat2012['bulan'])))
print('ntpp: ' + str(jawabarat2012['ntpp'].corr(jawabarat2012['bulan'])))
print('latitude: ' + str(jawabarat2012['latitude'].corr(jawabarat2012['bulan'])))
print('longitude: ' + str(jawabarat2012['longitude'].corr(jawabarat2012['bulan'])))

jawabarat2013=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2013)]
print('Untuk Jawa Barat tahun 2013, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2013['it'].corr(jawabarat2013['bulan'])))
print('ib: ' +str(jawabarat2013['ib'].corr(jawabarat2013['bulan'])))
print('ntpp: ' + str(jawabarat2013['ntpp'].corr(jawabarat2013['bulan'])))
print('latitude: ' + str(jawabarat2013['latitude'].corr(jawabarat2013['bulan'])))
print('longitude: ' + str(jawabarat2013['longitude'].corr(jawabarat2013['bulan'])))

jawabarat2014=df.loc[(df['nama_provinci']=='Prov. Jawa Barat') & (df['tahun']==2014)]
print('Untuk Jawa Barat tahun 2014, berikut koefisien korelasinya tiap bulan: ')
print('it: ' + str(jawabarat2014['it'].corr(jawabarat2014['bulan'])))
print('ib: ' +str(jawabarat2014['ib'].corr(jawabarat2014['bulan'])))
print('ntpp: ' + str(jawabarat2014['ntpp'].corr(jawabarat2014['bulan'])))
print('latitude: ' + str(jawabarat2014['latitude'].corr(jawabarat2014['bulan'])))
print('longitude: ' + str(jawabarat2014['longitude'].corr(jawabarat2014['bulan'])))
```

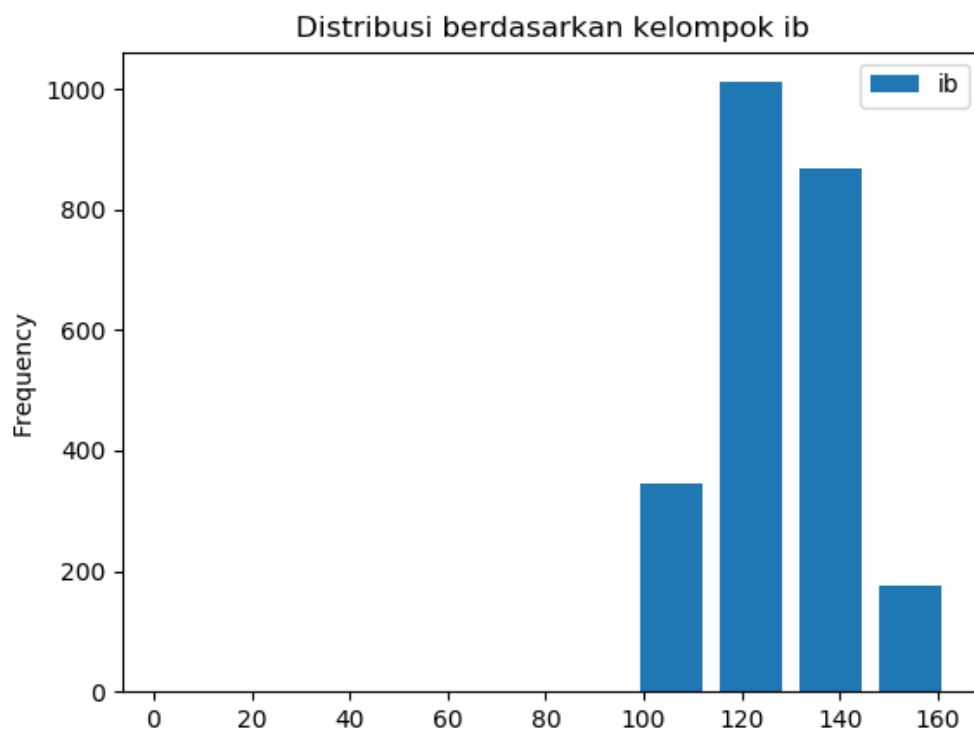
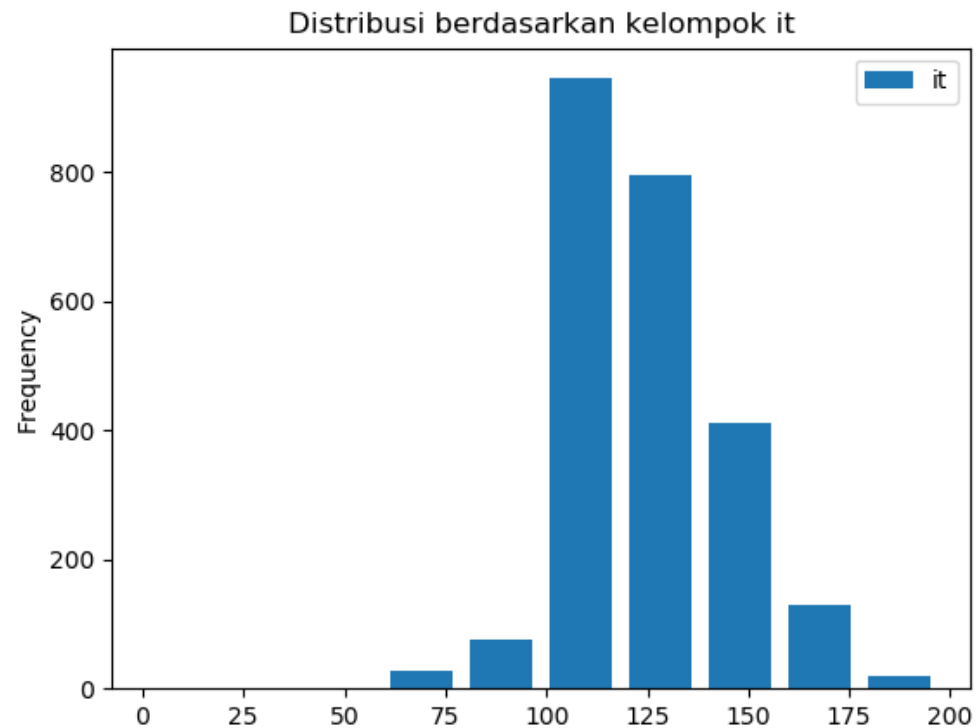
Untuk Jawa Barat tahun 2008, berikut koefisien korelasinya tiap bulan:
it: 0.6621078497897845
ib: 0.9734080628628686
ntpp: -0.1385314074359944
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2009, berikut koefisien korelasinya tiap bulan:
it: 0.8111637459029589
ib: 0.97648461124332
ntpp: -0.3492582862262574
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2010, berikut koefisien korelasinya tiap bulan:
it: 0.9117079614294684
ib: 0.9852874627201581
ntpp: 0.7268268792491996
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2011, berikut koefisien korelasinya tiap bulan:
it: 0.9643569645599562
ib: 0.8222577811567153
ntpp: 0.9805431214984544
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2012, berikut koefisien korelasinya tiap bulan:
it: 0.8886238160733461
ib: 0.9796239234280664
ntpp: 0.7279086980773599
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2013, berikut koefisien korelasinya tiap bulan:
it: -0.24668774366807456
ib: -0.21275561895944325
ntpp: -0.3363105799395459
latitude: nan
longitude: 0.0
Untuk Jawa Barat tahun 2014, berikut koefisien korelasinya tiap bulan:
it: 0.9824579845266985
ib: 0.9653959455373068
ntpp: 0.9966158955401254
latitude: nan
longitude: 0.0

Pada tahun 2008, nilai yang diterima petani (it) hampir berbanding lurus dengan semakin bertambahnya bulan, sedangkan nilai yang harus dibayar petani (ib) cenderung sangat berbanding lurus dengan bertambahnya bulan. Lalu, pada tahun 2009, nilai ib masih lebih berbanding lurus dengan bertambahnya bulan daripada it. Ini membuktikan bahwa pada tahun 2008 dan 2009, petani-petani di Jawa Barat masih kurang tingkat kesejahteraannya, jika dilihat dari kesebandingan ib, it, dan bulan, karena pertumbuhan ib lebih pesat daripada it. Di tahun 2010 juga sama seperti 2008 dan 2009. Namun, pada tahun 2011, koefisien korelasi it lebih besar dari koefisien korelasi dari ib, sehingga pada tahun 2011, petani-petani agaknya lebih sejahtera. Tahun 2012, keadaan kembali menyusahkan petani. Pada tahun 2013, koefisien korelasi it dan ib cenderung mendekati nol, sehingga antara keduanya dengan bulan cenderung hampir tidak ada korelasinya. Sehingga, keadaan saat itu tidak dapat diprediksi. Untuk tahun 2014, koefisien korelasi it lebih besar dari ib sehingga kesejahteraan petani agak sedikit meningkat. Jika dilihat dari nilai tukar petani (ntpp), pada tahun 2008, 2009, dan 2013, nilai tukar petani cenderung hampir tidak ada korelasinya dengan bulan. Sedangkan, pada tahun lainnya, nilai tukar petani cenderung berbanding lurus dengan bulan, apalagi pada tahun 2011 dan 2014, dikarenakan koefisien korelasi it yang lebih besar dari ib.

```

|
| df[['it']].plot(kind='hist',rwidth=0.8,title='Distribusi berdasarkan kelompok it')
| plt.show()
| df[['ib']].plot(kind='hist',rwidth=0.8,title='Distribusi berdasarkan kelompok ib')
| plt.show()

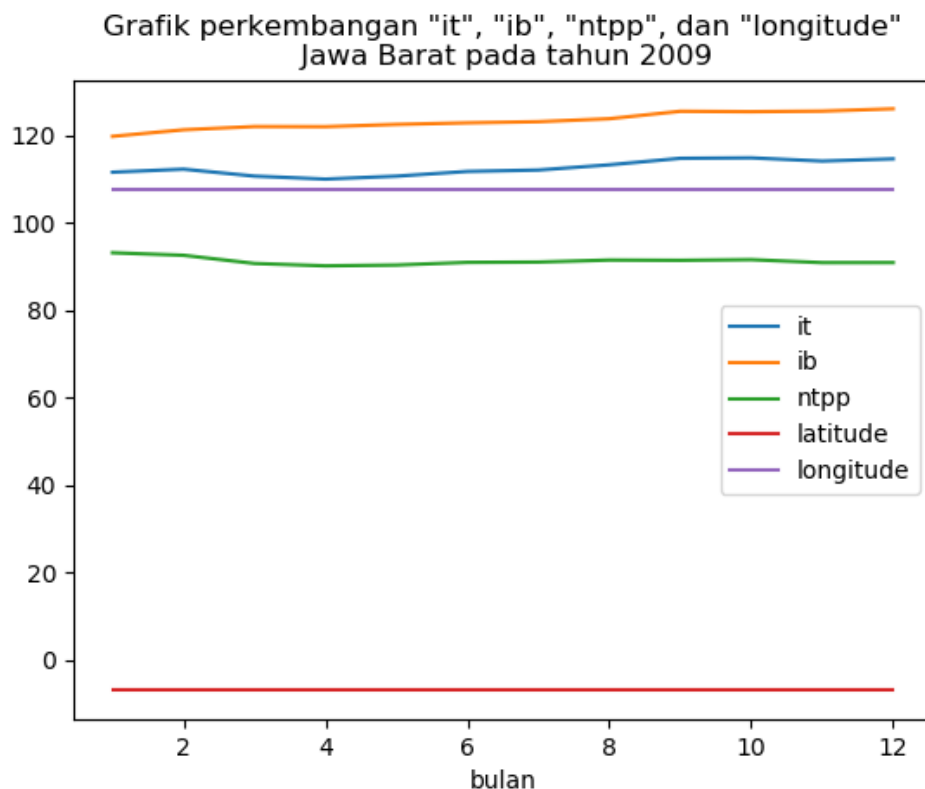
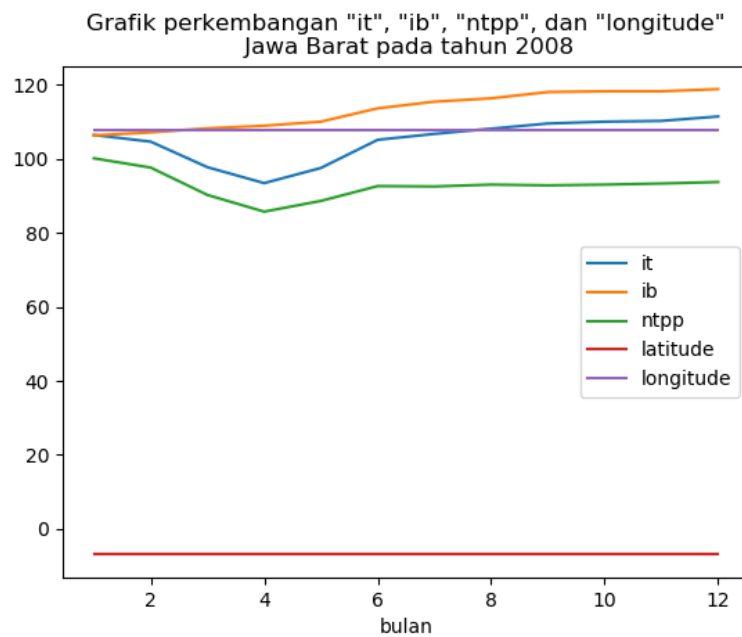
```



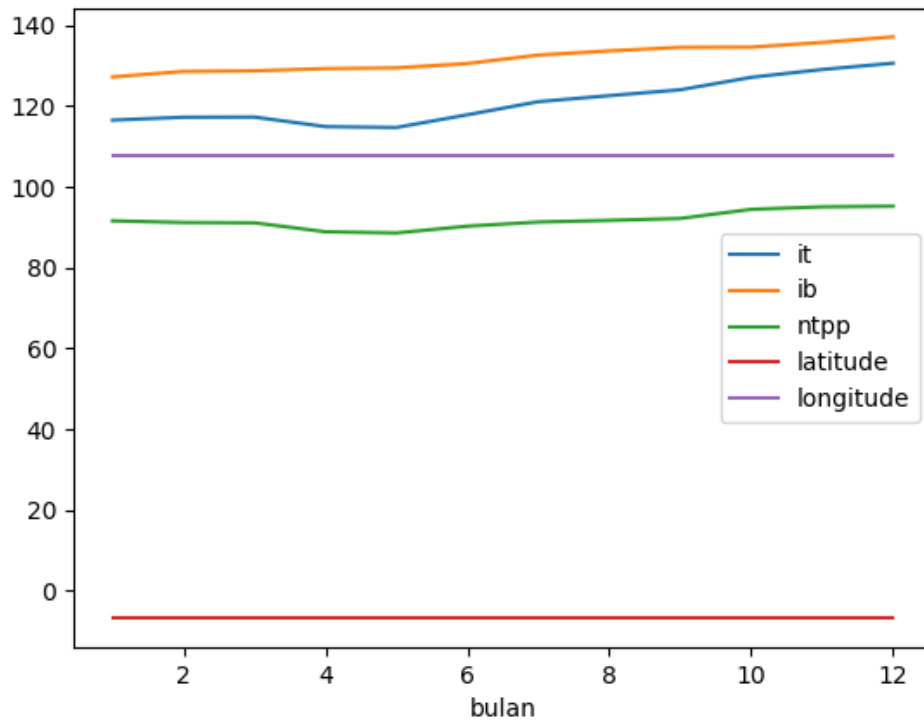
Kedua gambar di atas adalah histogram yang menggambarkan distribusi nilai it dan ib dari keseluruhan data. Jika dilihat, nilai it yang berfrekuensi paling banyak ada pada kisaran 100 – 115 yang terbilang

masih di bawah rata-rata dari data seluruhnya. Dan, jika dilihat, nilai ib yang berfrekuensi paling banyak ada pada kisaran 115 – 128. Yang merupakan kisaran rata-rata nilai yang harus dibayar petani. Sehingga, bisa disimpulkan ib relatif lebih tinggi daripada it: Nilai yang harus dibayar petani lebih tinggi daripada nilai yang diterima petani.

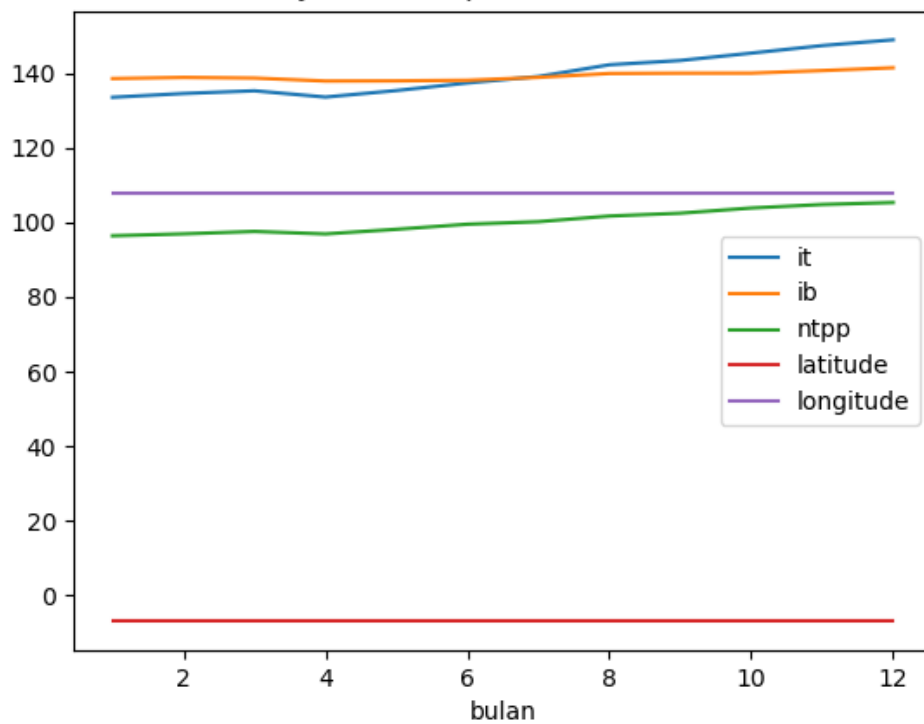
```
jawabarat2008.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2009.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2010.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2011.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2012.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2013.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
jawabarat2014.plot(kind='line',x='bulan',y=['it','ib','ntpp','latitude','longitude'],title='Grafik perkembangan "it", "ib", "ntpp", dan "longitude"\n Jawa Barat pada t
plt.show()
```



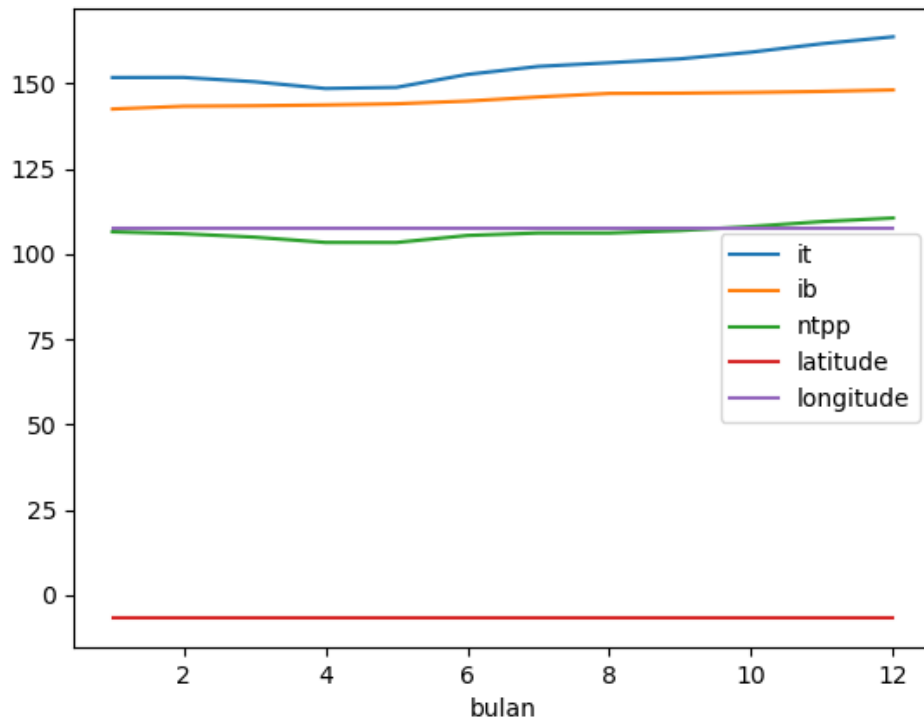
Grafik perkembangan "it", "ib", "ntpp", dan "longitude"
Jawa Barat pada tahun 2010



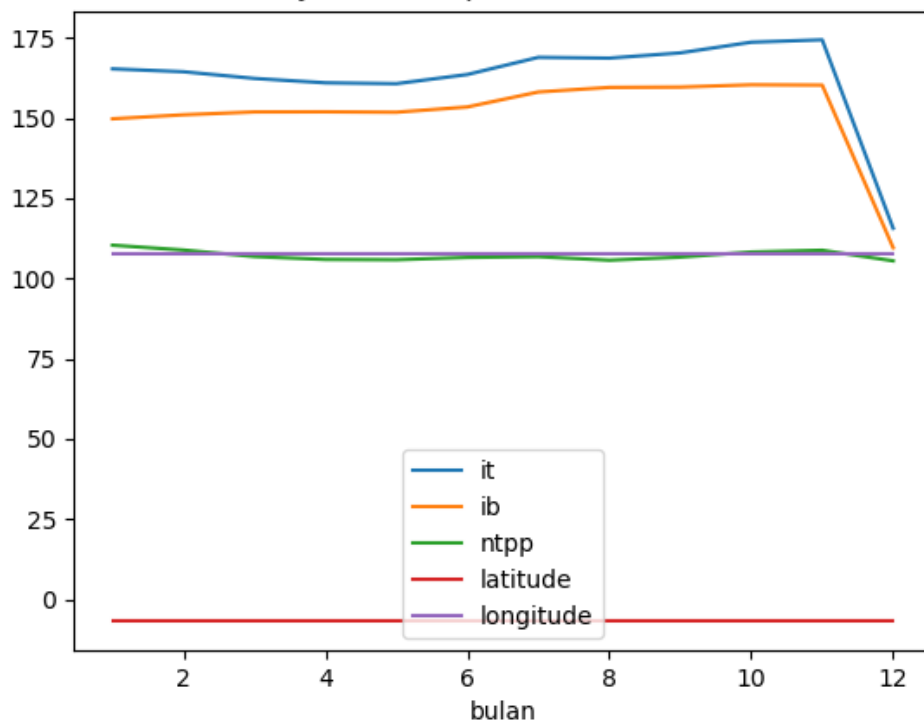
Grafik perkembangan "it", "ib", "ntpp", dan "longitude"
Jawa Barat pada tahun 2011

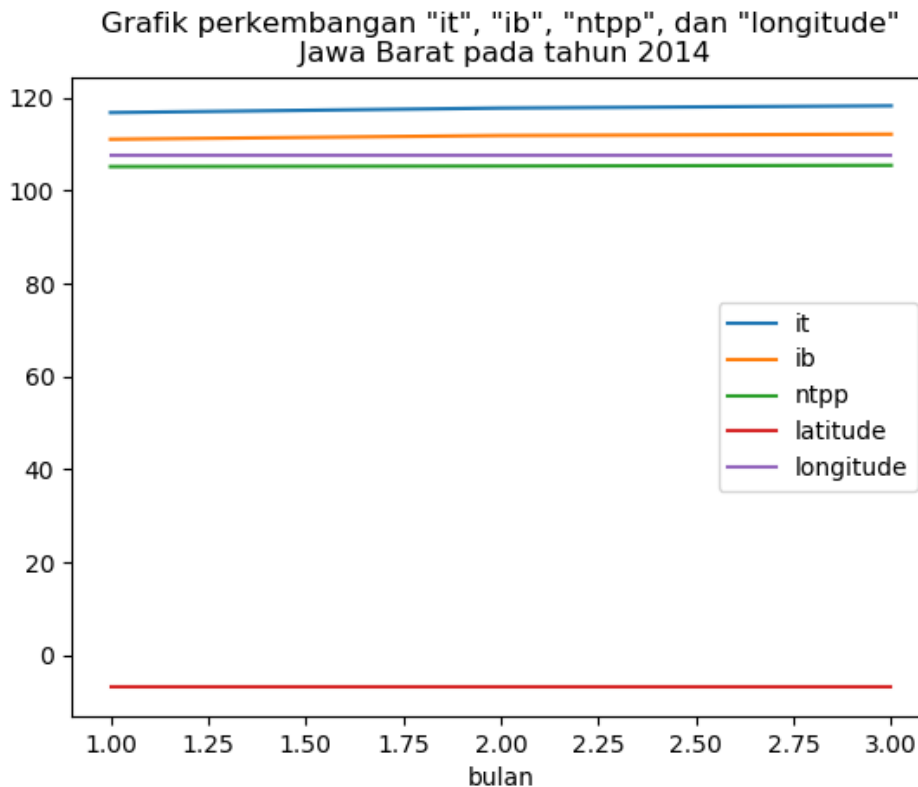


Grafik perkembangan "it", "ib", "ntpp", dan "longitude" Jawa Barat pada tahun 2012



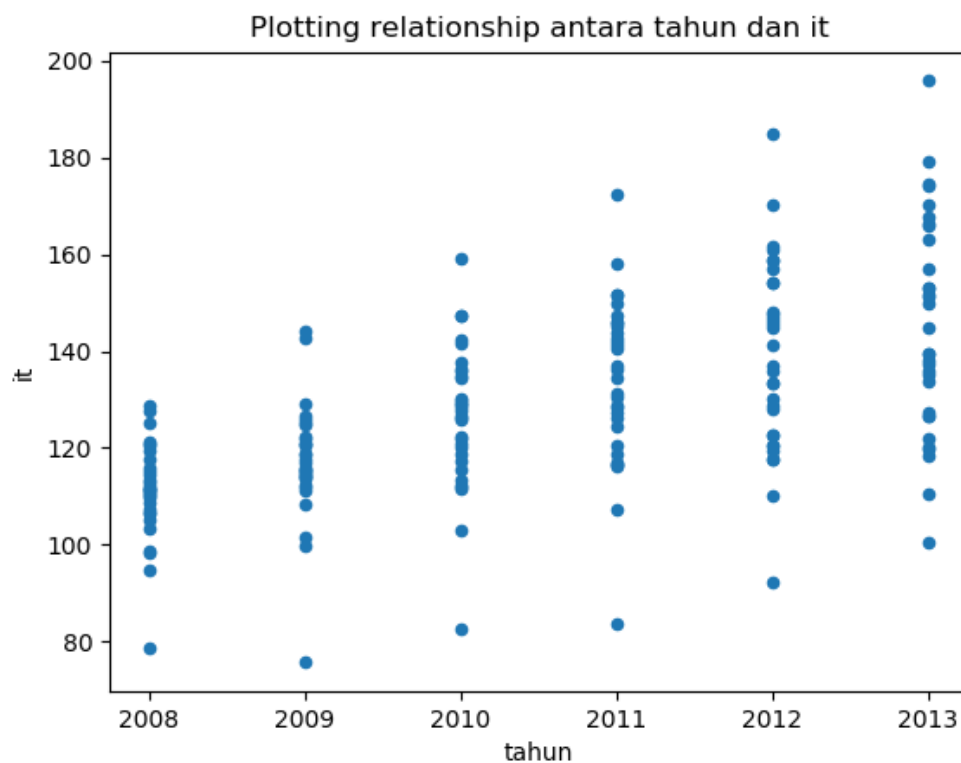
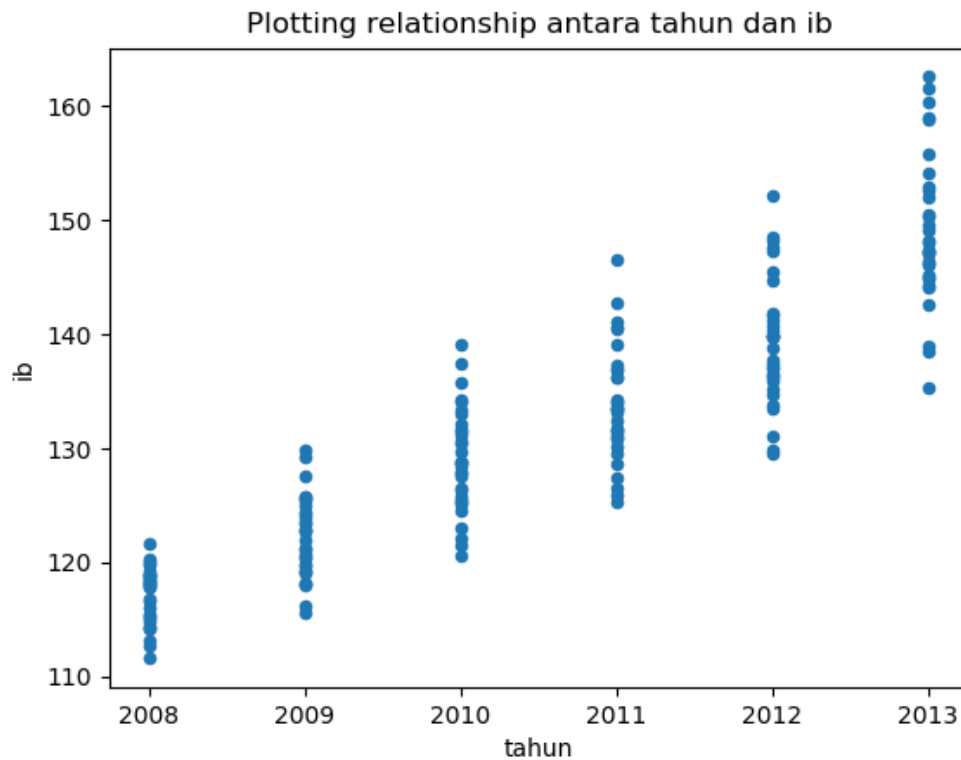
Grafik perkembangan "it", "ib", "ntpp", dan "longitude" Jawa Barat pada tahun 2013





Dapat dilihat juga dari grafik, semuanya: *ib*, *it*, *ntpp*. Pada tahun 2008, *ntpp* sempat menurun lalu cenderung stabil setelahnya. Sama dengan pada tahun 2009, *ntpp* sempat menurun namun tidak drastis, dan cenderung stabil setelahnya. Pada tahun-tahun berikutnya pun, ada naik turun, namun tidak signifikan, lalu cenderung stabil. Jika dilihat pada *it* dan *ib* nya, grafik *ib* cenderung di atas *it* pada tahun 2008, 2009, dan 2010. Pada tahun 2011, sempat di atas *it* lalu juga sempat di bawahnya. Pada tahun 2012, 2013, dan 2014 *it* cenderung di atas *ib*. Dapat disimpulkan bahwa, walau menurut interpretasi sebelumnya, para petani masih kurang sejahtera, tapi paling tidak perlahan-lahan, petani cenderung lebih sejahtera.

```
datanov=df.loc[df['bulan']==11]
datanov.plot(kind='scatter',x='tahun',y='it',title='Plotting relationship antara tahun dan it')
plt.show()
datanov.plot(kind='scatter',x='tahun',y='ib',title='Plotting relationship antara tahun dan ib')
plt.show()
```



Kedua grafik plotting di atas adalah plotting antara tahun dan juga it dan ib di seluruh provinsi di Indonesia, khususnya pada bulan November. Dapat dilihat dari trennya, it dan juga ib sama-sama naik. Bisa disimpulkan bahwa, nantinya *ntpp* juga terlihat cenderung stabil. Namun, plotting it sebenarnya sangatlah tidak rapi dibandingkan plotting ib. Maka, bisa disimpulkan ib stabil naik terus, namun it di setiap daerah cenderung stabil, naik sedikit.

DATA 2 :
Data Jumlah Penduduk Miskin dan
Indeks Kemiskinan pada seluruh
kabupaten di Indonesia Tahun 2007-
2010

BAB 1

DESKRIPSI DATA

Dataset ini menyajikan jumlah dan persentase penduduk miskin per kabupaten/kota beserta indeks kemiskinan dan perhitungan garis kemiskinan dari tahun 2007 hingga 2010. Garis kemiskinan adalah jumlah rupiah minimum untuk memenuhi kebutuhan pokok minimum makanan yang setara dengan 2100 kilokalori per kapita per hari, serta kebutuhan pokok non-makanan seperti sandang, papan, dan pendidikan. Indeks kemiskinan di dalam data ini terdiri dari dua komponen, yaitu Indeks Kedalaman(P1) dan Keparahan Kemiskinan(P2)

Yang pertama adalah Indeks Kedalaman Kemiskinan (*Poverty Gap Index*-P1), yang merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan. Indeks ini dapat mengindikasikan biaya minimum untuk menghapus kemiskinan, yaitu sebesar jumlah kedalaman kemiskinan (*poverty gap*), yaitu selisih antara garis kemiskinan dengan pendapatan

Yang kedua adalah Indeks Keparahan Kemiskinan (*Poverty Severity Index*-P2), yang memberikan gambaran penyebaran pengeluaran di antara penduduk miskin. Semakin tinggi nilai indeks, semakin tinggi ketimpangan pengeluaran di antara penduduk miskin.

Dataset ini diambil dalam format csv dari <https://data.go.id/dataset/penduduk-miskin-dan-indeks-kemiskinan> dan berisi 10 kolom dan 1880 baris. Ukuran data ini adalah 149 KB.

Kode:

```
import pandas as pd
df = pd.read_csv("processed-penduduk-miskin-indeks-kemiskinan-2007-2010.csv")

print("Jumlah baris : ", df.shape[0])
print("Jumlah kolom : ", df.shape[1])
```

BAB 2 SAMPEL DATA

E. 10 data teratas

```
print(df[:10])
```

	kode_provinsi	nama_provinsi	kode_kabkota	nama_kabkota	tahun	jumlah_penduduk_miskin	persen_penduduk_miskin	indeks_p1	indeks_p2	garis_kemiskinan
0	11	Prov. Nanggroe Aceh Darussalam	1101	Kab. Simeulue	2007	25100.0	32.26	5.24	1.22	216518.0
1	11	Prov. Nanggroe Aceh Darussalam	1101	Kab. Simeulue	2008	20600.0	26.45	4.51	1.19	253123.0
2	11	Prov. Nanggroe Aceh Darussalam	1101	Kab. Simeulue	2009	19100.0	24.72	3.95	1.01	255471.0
3	11	Prov. Nanggroe Aceh Darussalam	1101	Kab. Simeulue	2010	19000.0	23.61	4.29	1.19	278023.0
4	11	Prov. Nanggroe Aceh Darussalam	1102	Kab. Aceh Singkil	2007	26500.0	28.54	5.50	1.50	208087.0
5	11	Prov. Nanggroe Aceh Darussalam	1102	Kab. Aceh Singkil	2008	22200.0	23.27	4.82	1.38	213997.0
6	11	Prov. Nanggroe Aceh Darussalam	1102	Kab. Aceh Singkil	2009	20300.0	21.06	3.40	0.84	257778.0
7	11	Prov. Nanggroe Aceh Darussalam	1102	Kab. Aceh Singkil	2010	19900.0	19.38	2.91	0.65	280534.0
8	11	Prov. Nanggroe Aceh Darussalam	1103	Kab. Aceh Selatan	2007	49800.0	24.72	4.57	1.21	196167.0
9	11	Prov. Nanggroe Aceh Darussalam	1103	Kab. Aceh Selatan	2008	38800.0	19.40	3.62	1.01	203761.0

Dari sampel ini, kita mengambil 10 data teratas yang disusun berdasarkan kode provinsi.

F. 10 kabupaten/kota dengan indeks kedalaman kemiskinan tertinggi.

```
persenatas = df.loc[df["tahun"] == 2010].sort_values(["indeks_p1"], ascending = [0])
print ("10 kabupaten/kota dengan indeks kedalaman kemiskinan tertinggi:")
print(persenatas.iloc[:9,3])
```

10 kabupaten/kota dengan indeks kedalaman kemiskinan tertinggi:

1865	Kab. Lanny Jaya
1859	Kab. Supiori
1816	Kab. Puncak Jaya
1763	Kab. Teluk Bintuni
1759	Kab. Teluk Wondama
1874	Kab. Intan Jaya
1861	Kab. Mamberamo Raya
205	Kota Gunungsitoli
1832	Kab. Asmat

Dari sampel ini, kita dapat melihat 10 kabupaten/kota dengan indeks kedalaman kemiskinan tertinggi. Artinya, 10 kabupaten/kota ini memiliki kesenjangan antara pengeluaran rata-rata penduduk miskin dengan garis kemiskinan yang lebih tinggi dibandingkan kabupaten/kota lainnya di Indonesia.

G. 10 data dengan persen penduduk miskin terendah pada tahun 2010.

```
persenbawah = df.loc[df["tahun"] == 2010].sort_values(["persen_penduduk_miskin"], ascending = [1])
print ("10 data dengan persen penduduk miskin terendah pada tahun 2010:")
print (persenbawah[:10])
```

	kode_provinsi	nama_provinsi	kode_kabkota	nama_kabkota	tahun	jumlah_penduduk_miskin	persen_penduduk_miskin	indeks_p1	indeks_p2	garis_kemiskinan
1030	36	Prov. Banten	3674	Kota Tangerang Selatan	2010	21900.0	1.67	0.35	0.11	275643.0
1066	51	Prov. Bali	5171	Kota Denpasar	2010	17500.0	2.21	0.20	0.03	309672.0
265	13	Prov. Sumatera Barat	1373	Kota Sawah Lunto	2010	1400.0	2.47	0.19	0.02	207920.0
678	32	Prov. Jawa Barat	3276	Kota Depok	2010	49600.0	2.84	0.51	0.14	310279.0
1042	51	Prov. Bali	5103	Kab. Badung	2010	17700.0	3.23	0.39	0.06	312602.0
1304	63	Prov. Kalimantan Selatan	6303	Kab. Banjar	2010	17000.0	3.34	0.48	0.11	230759.0
574	31	Prov. DKI Jakarta	3172	Kab. Kodya Jakarta Timur	2010	91700.0	3.41	0.41	0.07	325980.0
367	15	Prov. Jambi	1572	Kota Sungai Penuh	2010	3000.0	3.64	0.46	0.11	225456.0
570	31	Prov. DKI Jakarta	3171	Kab. Kodya Jakarta Selatan	2010	78600.0	3.81	0.67	0.18	397415.0
582	31	Prov. DKI Jakarta	3174	Kab. Kodya Jakarta Barat	2010	87200.0	3.81	0.62	0.18	320072.0

Dari sampel ini, kita dapat melihat 10 data yang memiliki persentase penduduk miskin terendah di Indonesia pada tahun 2010. Kota Tangerang Selatan di Provinsi Banten adalah daerah yang persentase penduduknya paling sedikit berada di bawah garis kemiskinan.

H. Data Kota Bandung

```
print("Data Kota Bandung: ")
print (df.loc[df["nama_kabkota"] == "Kota Bandung"])
```

Data Kota Bandung:

	kode_provinsi	nama_provinsi	kode_kabkota	nama_kabkota	tahun	jumlah_penduduk_miskin	persen_penduduk_miskin	indeks_p1	indeks_p2	garis_kemiskinan
663	32	Prov. Jawa Barat	3273	Kota Bandung	2007	87200.0	3.68	0.66	0.18	181142.0
664	32	Prov. Jawa Barat	3273	Kota Bandung	2008	106800.0	4.42	0.78	0.22	209171.0
665	32	Prov. Jawa Barat	3273	Kota Bandung	2009	110300.0	4.50	0.55	0.11	262093.0
666	32	Prov. Jawa Barat	3273	Kota Bandung	2010	118600.0	4.95	0.98	0.29	279784.0

Sampel ini menunjukkan data kemiskinan Kota Bandung dari tahun ke tahun. Kita dapat melihat bahwa persentase penduduk miskin secara konsisten mengalami peningkatan di Bandung dari tahun ke tahun.

BAB 3

MAKNA ATRIBUT DAN NILAI EKSTREMUM

10. kode_provinsi: Kode tiap provinsi lokasi pengumpulan data sesuai dengan ketentuan Kementerian Dalam Negeri. Data kategorikal tipe nominal.
11. nama_provinsi: Nama provinsi lokasi pengumpulan data. Data kategorikal tipe nominal.
12. kode_kabkota: Kode tiap kabupaten/kota lokasi pengumpulan data. Data kategorikal tipe nominal.
13. nama_kabkota: Nama kabupaten/kota lokasi pengumpulan data. Data kategorikal tipe nominal.
14. tahun: Tahun diambilnya data. Data kategorikal tipe nominal, dengan range 2007-2010.
15. jumlah_penduduk_miskin: Jumlah penduduk yang berada di bawah garis kemiskinan pada daerah dan tahun tertentu yang dispesifikasikan. Data kuantitatif, dengan range 1000-519500.
16. persen_penduduk_miskin: Persentase penduduk yang berada di bawah garis kemiskinan pada daerah dan tahun tertentu yang dispesifikasikan. Data kuantitatif, dengan range 1.67-58.7.
17. indeks_p1: Indeks kedalaman kemiskinan, yaitu pengukuran rata-rata pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Data kuantitatif, dengan range 0.15-22.75.
18. indeks_p2: Indeks keparahan kemiskinan, pengukuran penyebaran pengeluaran di antara penduduk miskin. Data kuantitatif, dengan range 0.02-11.61.
19. garis_kemiskinan: Jumlah rupiah minimum untuk memenuhi kebutuhan pokok makanan dan non-makanan. Data kuantitatif, dengan range 90258-504235.

```
print("Nilai maksimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].max())
print("Nilai minimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].min())
print("Nilai maksimum persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].max())
print("Nilai minimum persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].min())
print("Nilai maksimum indeks kedalaman kemiskinan: ", df["indeks_p1"].max())
print("Nilai minimum indeks kedalaman kemiskinan: ", df["indeks_p1"].min())
print("Nilai maksimum indeks keparahan kemiskinan: ", df["indeks_p2"].max())
print("Nilai minimum indeks keparahan kemiskinan: ", df["indeks_p2"].min())
print("Nilai maksimum garis kemiskinan: ", df["garis_kemiskinan"].max())
print("Nilai minimum garis kemiskinan: ", df["garis_kemiskinan"].min())
```

```
Nilai maksimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 519500.0
Nilai minimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 1000.0
Nilai maksimum persen penduduk miskin tiap kabupaten/kota tiap tahun: 58.7
Nilai minimum persen penduduk miskin tiap kabupaten/kota tiap tahun: 1.67
Nilai maksimum indeks kedalaman kemiskinan: 22.72
Nilai minimum indeks kedalaman kemiskinan: 0.15
Nilai maksimum indeks keparahan kemiskinan: 11.61
Nilai minimum indeks keparahan kemiskinan: 0.02
Nilai maksimum garis kemiskinan: 504235.0
Nilai minimum garis kemiskinan: 90258.0
```

BAB 4

ANALISIS STATISTIK DATA

1. Jumlah penduduk miskin

```
print("Rata-rata jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].mean())
print("Standar deviasi jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].std())
print("Persentil 10% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].quantile(0.1))
print("Persentil 25% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].quantile(0.25))
print("Persentil 50% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].quantile(0.5))
print("Persentil 75% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].quantile(0.75))
print("Persentil 90% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].quantile(0.9))
print("Nilai maksimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].max())
print("Nilai minimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: ", df["jumlah_penduduk_miskin"].min())
```

Rata-rata jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 71643.01705756929
Standar deviasi jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 82922.66796969637
Persentil 10% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 10600.0
Persentil 25% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 18875.0
Persentil 50% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 36050.0
Persentil 75% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 90950.0
Persentil 90% jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 193300.0
Nilai maksimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 519500.0
Nilai minimum jumlah penduduk miskin tiap kabupaten/kota tiap tahun: 1000.0

Statistik di atas memberikan informasi mengenai jumlah absolut penduduk miskin di suatu daerah. Kesimpulan yang dapat diambil belum banyak, mengingat bahwa banyak faktor yang memengaruhi jumlah penduduk miskin. Contohnya, daerah yang memiliki penduduk banyak kemungkinan akan memiliki penduduk miskin yang lebih banyak dibandingkan daerah dengan penduduk yang lebih sedikit, meskipun belum tentu daerah yang pertama lebih terbelakang dibandingkan daerah yang kedua.

2. Persen penduduk miskin

```
print("Rata-rata persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].mean())
print("Standar deviasi persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].std())
print("Persentil 10% persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].quantile(0.1))
print("Persentil 25% persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].quantile(0.25))
print("Persentil 50% persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].quantile(0.5))
print("Persentil 75% persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].quantile(0.75))
print("Persentil 90% persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].quantile(0.9))
print("Nilai maksimum persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].max())
print("Nilai minimum persen penduduk miskin tiap kabupaten/kota tiap tahun: ", df["persen_penduduk_miskin"].min())
```

Rata-rata persen penduduk miskin tiap kabupaten/kota tiap tahun: 16.759472281449895
Standar deviasi persen penduduk miskin tiap kabupaten/kota tiap tahun: 9.946054638299112
Persentil 10% persen penduduk miskin tiap kabupaten/kota tiap tahun: 6.365
Persentil 25% persen penduduk miskin tiap kabupaten/kota tiap tahun: 9.3375
Persentil 50% persen penduduk miskin tiap kabupaten/kota tiap tahun: 14.58
Persentil 75% persen penduduk miskin tiap kabupaten/kota tiap tahun: 21.745
Persentil 90% persen penduduk miskin tiap kabupaten/kota tiap tahun: 30.229999999999997
Nilai maksimum persen penduduk miskin tiap kabupaten/kota tiap tahun: 58.7
Nilai minimum persen penduduk miskin tiap kabupaten/kota tiap tahun: 1.67

Dari statistik di atas, perbandingan standar deviasi dengan rata-rata kurang dari 1, yang menandakan bahwa sebagian besar data cukup mendekati rata-rata dan ketimpangan, meskipun ada, tidak bersifat drastis. Tetapi data ini belum memberikan informasi mengenai seberapa timpangnya pengeluaran penduduk yang hidup di bawah garis kemiskinan.

3. Indeks Kedalaman Kemiskinan

```
print("Rata-rata nilai indeks kedalaman kemiskinan: ", df["indeks_pl"].mean())
print("Standar deviasi nilai indeks kedalaman kemiskinan: ", df["indeks_pl"].std())
print("Persentil 10% indeks kedalaman kemiskinan: ", df["indeks_pl"].quantile(0.1))
print("Persentil 25% indeks kedalaman kemiskinan: ", df["indeks_pl"].quantile(0.25))
print("Persentil 50% indeks kedalaman kemiskinan: ", df["indeks_pl"].quantile(0.5))
print("Persentil 75% indeks kedalaman kemiskinan: ", df["indeks_pl"].quantile(0.75))
print("Persentil 90% indeks kedalaman kemiskinan: ", df["indeks_pl"].quantile(0.9))
print("Nilai maksimum indeks kedalaman kemiskinan: ", df["indeks_pl"].max())
print("Nilai minimum indeks kedalaman kemiskinan: ", df["indeks_pl"].min())
```



```

Rata-rata nilai indeks kedalaman kemiskinan: 3.212771855010661
Standar deviasi nilai indeks kedalaman kemiskinan: 2.761375636536152
Persentil 10% indeks kedalaman kemiskinan: 0.91
Persentil 25% indeks kedalaman kemiskinan: 1.4175
Persentil 50% indeks kedalaman kemiskinan: 2.4
Persentil 75% indeks kedalaman kemiskinan: 4.012499999999999
Persentil 90% indeks kedalaman kemiskinan: 6.455
Nilai maksimum indeks kedalaman kemiskinan: 22.72
Nilai minimum indeks kedalaman kemiskinan: 0.15

```

Indeks Ketimpangan Kemiskinan merepresentasikan total *gap* atau celah antara garis kemiskinan dengan pengeluaran individu yang hidup di bawah garis tersebut. Mengacu pada data Indeks Ketimpangan Kemiskinan Global (https://en.wikipedia.org/wiki/Poverty_gap_index#Poverty_gap_index_by_country), Indeks Ketimpangan Kemiskinan Indonesia secara rata-rata lebih rendah dibandingkan banyak negara lain seperti Amerika Serikat, Korea Selatan, atau Jepang. Di sisi lainnya, negara-negara tersebut memiliki garis kemiskinan yang jauh lebih tinggi dibandingkan garis kemiskinan Indonesia. Perbandingan antara standar deviasi dengan rata-rata berada di bawah 1, yang menandakan sebagian besar data cenderung mendekati rata-rata.

4. Indeks Keparahan Kemiskinan

```

print("Rata-rata nilai indeks keparahan kemiskinan: ", df["indeks_p2"].mean())
print("Standar deviasi nilai indeks keparahan kemiskinan: ", df["indeks_p2"].std())
print("Persentil 10% indeks keparahan kemiskinan: ", df["indeks_p2"].quantile(0.1))
print("Persentil 25% indeks keparahan kemiskinan: ", df["indeks_p2"].quantile(0.25))
print("Persentil 50% indeks keparahan kemiskinan: ", df["indeks_p2"].quantile(0.5))
print("Persentil 75% indeks keparahan kemiskinan: ", df["indeks_p2"].quantile(0.75))
print("Persentil 90% indeks keparahan kemiskinan: ", df["indeks_p2"].quantile(0.9))
print("Nilai maksimum indeks keparahan kemiskinan: ", df["indeks_p2"].max())
print("Nilai minimum indeks keparahan kemiskinan: ", df["indeks_p2"].min())

```

```

Rata-rata nilai indeks keparahan kemiskinan: 0.9480277185501066
Standar deviasi nilai indeks keparahan kemiskinan: 1.1329149429817194
Persentil 10% indeks keparahan kemiskinan: 0.2
Persentil 25% indeks keparahan kemiskinan: 0.34
Persentil 50% indeks keparahan kemiskinan: 0.61
Persentil 75% indeks keparahan kemiskinan: 1.11
Persentil 90% indeks keparahan kemiskinan: 1.98
Nilai maksimum indeks keparahan kemiskinan: 11.61
Nilai minimum indeks keparahan kemiskinan: 0.02

```

Pada statistik Indeks Keparahan Kemiskinan ini, terlihat ketimpangan yang amat besar antar daerah mengenai tingkat keparahan kemiskinan. Perbedaan antara nilai minimum dan maksimum dengan satu sama lain dan dengan rata-rata berukuran cukup besar, dan perbandingan antara standar deviasi dengan rata-rata menghasilkan nilai lebih dari 1.

5. Garis kemiskinan

```

print("Rata-rata garis kemiskinan: ", df["garis_kemiskinan"].mean())
print("Standar deviasi garis kemiskinan: ", df["garis_kemiskinan"].std())
print("Persentil 10% garis kemiskinan: ", df["garis_kemiskinan"].quantile(0.1))
print("Persentil 25% garis kemiskinan: ", df["garis_kemiskinan"].quantile(0.25))
print("Persentil 50% garis kemiskinan: ", df["garis_kemiskinan"].quantile(0.5))
print("Persentil 75% garis kemiskinan: ", df["garis_kemiskinan"].quantile(0.75))
print("Persentil 90% garis kemiskinan: ", df["garis_kemiskinan"].quantile(0.9))
print("Nilai maksimum garis kemiskinan: ", df["garis_kemiskinan"].max())
print("Nilai minimum garis kemiskinan: ", df["garis_kemiskinan"].min())

```

```
Rata-rata garis kemiskinan: 211480.1236673774
Standar deviasi garis kemiskinan: 53562.02696229887
Persentil 10% garis kemiskinan: 152262.5
Persentil 25% garis kemiskinan: 172819.5
Persentil 50% garis kemiskinan: 203834.5
Persentil 75% garis kemiskinan: 240600.25
Persentil 90% garis kemiskinan: 283221.5
Nilai maksimum garis kemiskinan: 504235.0
Nilai minimum garis kemiskinan: 90258.0
```

Ditinjau dari perbandingan standar deviasi dengan rata-rata, dapat disimpulkan bahwa garis kemiskinan tidak begitu berbeda dari satu daerah ke daerah lainnya, yang menandakan harga barang dari suatu daerah ke daerah lainnya, secara rata-rata, tidak bervariasi terlalu jauh.

6. Distribusi frekuensi data kategorikal

```
print(dfkotor["kode_provinsi"].value_counts())
print(dfkotor["nama_provinsi"].value_counts())
print(dfkotor["kode_kabkota"].value_counts())
print(dfkotor["nama_kabkota"].value_counts())
print(dfkotor["tahun"].value_counts())
```

Kode provinsi dan nama provinsi memiliki distribusi yang sama karena kedua data komplementer. Demikian juga dengan antara kode kabupaten/kota dengan nama kabupaten/kota.

Distribusi frekuensi data tiap tahun adalah sebagai berikut:

```
2010  497
2009  472
2008  456
2007  455
```

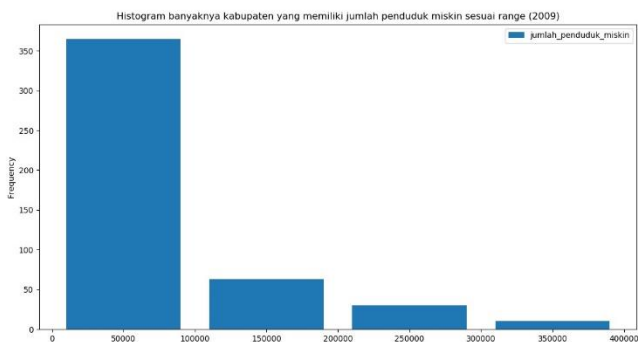
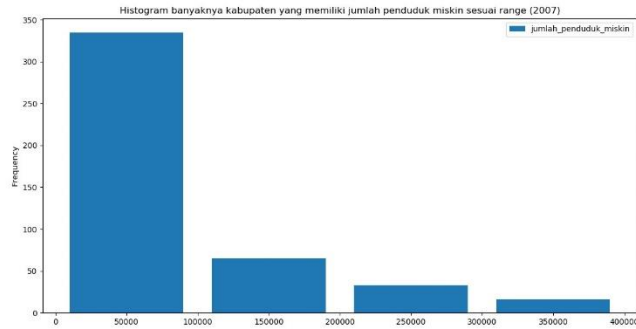
Dari statistik ini dapat disimpulkan bahwa data yang dikumpulkan berjumlah semakin banyak dari tahun ke tahun.

BAB 5

VISUALISASI DAN KORELASI DATA

1. Perbandingan Kategori / Comparing Categories

Kami membuat 4 histogram chart untuk Perbandingan Kategori. Adapun tujuan dibentuknya histogram ini adalah untuk mengetahui jumlah kabupaten yang memiliki jumlah penduduk miskin dalam suatu range tertentu. Dalam grafik dibawah ini akan dibagi setiap 50000 dari 0 hingga 400000 jiwa.



Dari 4 grafik disamping, kita dapat menyimpulkan dan mendapat insights bahwa pada tahun 2007-2010 terdapat 350-400 kabupaten/kota yang memiliki penduduk miskin berkisar 50000-100000 orang. Jumlah kabupaten dari masing masing range tidak menunjukkan kenaikan yang signifikan, melainkan naik turun sehingga grafik histogram yang dihasilkan diatas tidak menunjukkan perbedaan yang mencolok.

Source Code yang dipakai:

```
df2007[["jumlah_penduduk_miskin"]].plot(kind= "hist", bins =[0,100000,200000,300000,400000],rwidth=0.8,
title="Histogram banyaknya kabupaten yang memiliki jumlah penduduk miskin sesuai range (2007)")
plt.show()

df2008[["jumlah_penduduk_miskin"]].plot(kind= "hist", bins =[0,100000,200000,300000,400000],rwidth=0.8,
title="Histogram banyaknya kabupaten yang memiliki jumlah penduduk miskin sesuai range (2008)")
plt.show()

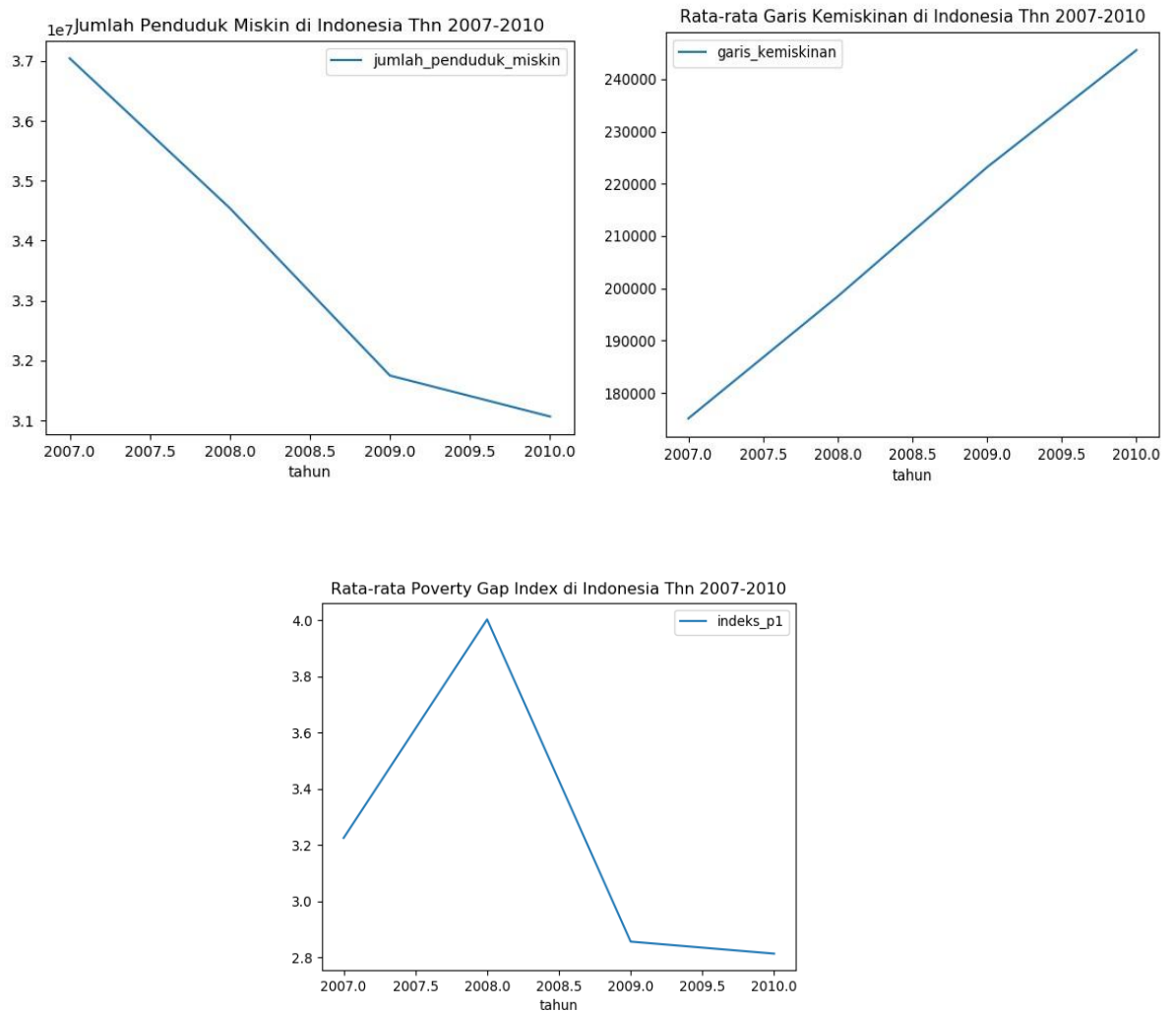
df2009[["jumlah_penduduk_miskin"]].plot(kind= "hist", bins =[0,100000,200000,300000,400000],rwidth=0.8,
title="Histogram banyaknya kabupaten yang memiliki jumlah penduduk miskin sesuai range (2009)")
plt.show()

df2010[["jumlah_penduduk_miskin"]].plot(kind= "hist", bins =[0,100000,200000,300000,400000],rwidth=0.8,
title="Histogram banyaknya kabupaten yang memiliki jumlah penduduk miskin sesuai range (2010)")
plt.show()
```

2. Penampilan Perubahan Tiap Waktu / Showing Over Time

Kami menggunakan 3 line chart untuk penampilan perubahan tiap waktu. Yang pertama menunjukkan jumlah penduduk miskin di Indonesia tahun 2007-2010, yang kedua menunjukkan

rata-rata garis kemiskinan Indonesia tahun 2007-2010, yang terakhir menunjukkan rata-rata poverty gap index/indeks kedalaman kemiskinan/indeks_p1 Indonesia tahun 2007-2010



Insight yang kami dapatkan dalam ketiga grafik ini adalah jumlah penduduk miskin Indonesia dari tahun 2007-2010 semakin menurun, namun rata-rata garis kemiskinan dari tahun 2007-2010 semakin tinggi. Menurut kami ini bisa saja terjadi apabila tingkat daya beli dari penduduk miskin naik secara drastis atau dengan kata lain, indeks p1 harus turun drastis. Namun disinilah keaneahannya. Apabila kita menganalisis grafik indeks p1, pada tahun 2008-2010 malah terjadi penurunan. Menurut kami ini disebabkan oleh adanya galat perhitungan dari pengukuran indeks tersebut.

Source Code:

```

df2007 = df.loc[df["tahun"] == 2007]
df2008 = df.loc[df["tahun"] == 2008]
df2009 = df.loc[df["tahun"] == 2009]
df2010 = df.loc[df["tahun"] == 2010]
#jumlah penduduk miskin di indonesia tahun 2007-2010 (setiap tahun)
a = df2007["jumlah_penduduk_miskin"].sum() #37062700
b = df2008["jumlah_penduduk_miskin"].sum() #34543900
c = df2009["jumlah_penduduk_miskin"].sum() #31772500
d = df2010["jumlah_penduduk_miskin"].sum() #31065300
#membuat dataframe baru rata-rata jpm (jumlah penduduk miskin) untuk dibuat grafiknya
dictjpm = {"jumlah_penduduk_miskin": pd.Series([a,b,c,d]),
           "tahun": pd.Series([2007,2008,2009,2010])}
dfjpm = pd.DataFrame(dictjpm)
dfjpm.plot(kind = 'line',x = "tahun",y = "jumlah_penduduk_miskin", title = "Jumlah Penduduk Miskin di Indonesia Thn 2007-2010")
plt.show()

#rata-rata garis kemiskinan di Indonesia tahun 2007-2010(setiap tahun)
e = (df2007["garis_kemiskinan"].mean()) #175089
f = (df2008["garis_kemiskinan"].mean()) #198447
g = (df2009["garis_kemiskinan"].mean()) #223211
h = (df2010["garis_kemiskinan"].mean()) #245546
#membuat dataframe baru rata-rata gk (garis kemiskinan) untuk dibuat grafiknya
dictgk = {"garis_kemiskinan": pd.Series([e,f,g,h]),
          "tahun": pd.Series([2007,2008,2009,2010])}
dfgk = pd.DataFrame(dictgk)
dfgk.plot(kind = 'line',x = "tahun",y = "garis_kemiskinan", title = "Rata-rata Garis Kemiskinan di Indonesia Thn 2007-2010")
plt.show()

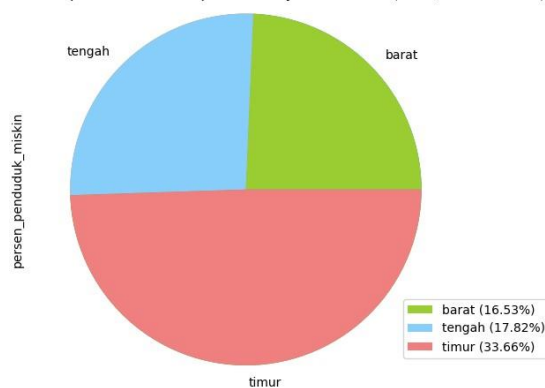
#rata-rata Poverty Gap Index (indeks_pl)di indonesia tahun 2007-2010 (setiap tahun)
i = (df2007["indeks_pl"].mean()) #3.228
j = (df2008["indeks_pl"].mean()) #4.003
k = (df2009["indeks_pl"].mean()) #2.853
l = (df2010["indeks_pl"].mean()) #2.813
#membuat dataframe baru rata-rata indeks_pl(poverty gap index) untuk dibuat grafiknya
dictpl = {"indeks_pl": pd.Series([i,j,k,l]),
          "tahun": pd.Series([2007,2008,2009,2010])}
dfpl = pd.DataFrame(dictpl)
dfpl.plot(kind = 'line',x = "tahun",y = "indeks_pl", title = "Rata-rata Poverty Gap Index di Indonesia Thn 2007-2010")
plt.show()

```

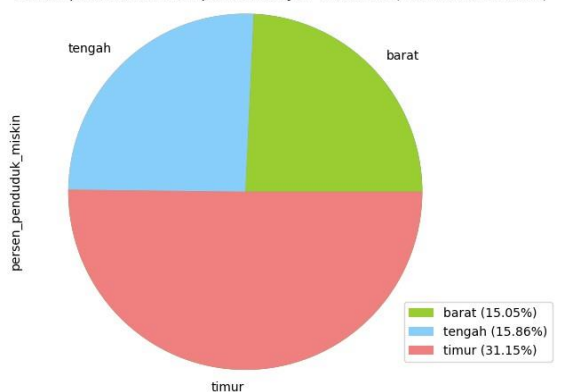
3. Whole-Part Relationship/ Hubungan Keseluruhan-Bagian

Pada hubungan keseluruhan bagian, kami mengambil 4 pie chart. Masing-masing pie chart memvisualisasikan berapa persen penduduk miskin yang berada di daerah barat, timur, atau tengah Indonesia. Tentu saja 4 pie chart ini dibuat dengan tahun yang berbeda (2007-2010). Perlu diingat bahwa pie chart ini menunjukkan bagian dari jumlah penduduk miskin saja, karena tidak mungkin Indonesia memiliki persen penduduk miskin total mencapai 100 persen.

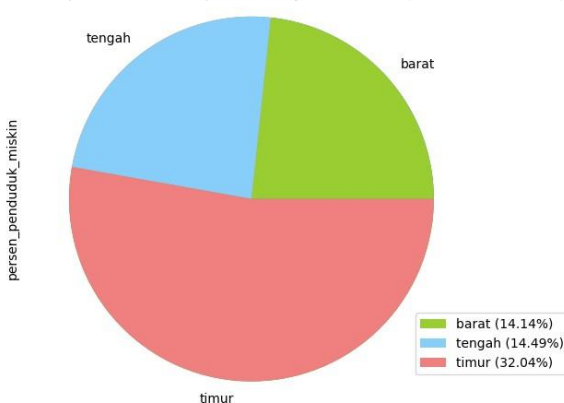
Persentase penduduk miskin pada 3 wilayah Indonesia (2007, total 68.02%)



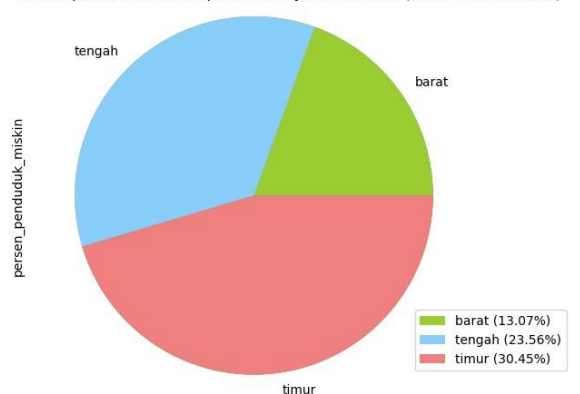
Persentase penduduk miskin pada 3 wilayah Indonesia (2008, total 62.06%)



Persentase penduduk miskin pada 3 wilayah Indonesia (2009, total 60.67%)



Persentase penduduk miskin pada 3 wilayah Indonesia (2010, total 67.08%)



Insight yang kami dapatkan dari keempat pie chart diatas adalah Indonesia bagian timur memiliki persen penduduk miskin yang paling tinggi dibandingkan Indonesia tengah dan Indonesia barat. Meskipun tahun berganti dari 2007-2010, Indonesia bagian timur yang persen penduduk miskinnya selalu yang paling terbesar dan tidak ada perubahan yang signifikan (fluktuatif turun naik). Berbeda dengan Indonesia bagian barat yang selalu paling kecil persen penduduk miskinnya, dan monoton turun dari tahun 2007-2010. Menurut kami, Indonesia timur memiliki persen penduduk miskin yang tinggi karena barang-barang yang dijual disana lebih mahal karena operasional transportasi dari Jawa (pusat produksi) ke Indonesia bagian timur. Sementara daya beli masyarakat untuk membeli barang-barang tidak memenuhi dan hasilnya Indonesia timur memiliki persen penduduk miskin yang tinggi seperti yang digambarkan pie chart diatas

Source Code:

Kode provinsi dibatasi untuk menentukan dataframe Indonesia barat, tengah dan timur

```
dfbarat = df.loc[(df["kode_provinsi"] < 51) | (df["kode_provinsi"] == 62) | (df["kode_provinsi"] == 61)]
dftengah = df.loc[(df["kode_provinsi"] >= 51 & (df["kode_provinsi"] <= 76)) & (df["kode_provinsi"] != 62) & (df["kode_provinsi"] != 61)]
dftimur = df.loc[df["kode_provinsi"] > 76]
```

Untuk tahun 2007:

```
# 2007
dfbarat2007 = dfbarat.loc[df["tahun"] == 2007]
dftengah2007 = dftengah.loc[df["tahun"] == 2007]
dftimur2007 = dftimur.loc[df["tahun"] == 2007]
a = dfbarat2007["persen_penduduk_miskin"].mean()
b = dftengah2007["persen_penduduk_miskin"].mean()
c = dftimur2007["persen_penduduk_miskin"].mean()
#persen_penduduk_miskin
#barat          16.537626
#tengah          17.824394
#timur           33.662558
#total           68.024578

dictppm = {"persen_penduduk_miskin" : pd.Series([a,b,c], index = ["barat","tengah","timur"])}
dfppm = pd.DataFrame(dictppm)
print(dfppm)
dfppm.plot(kind = "pie",subplots = True,title = "Persen penduduk miskin pada 3 wilayah Indonesia (2007, total 68.02%)")

labels = [r'barat (16.53%)', r'tengah (17.82%)',
r'timur (33.66%)]
sizes = [24.32, 26.2, 49.48]
colors = ['yellowgreen', 'lightskyblue', 'lightcoral']
patches, texts = plt.pie(sizes, colors=colors, startangle=0)
plt.legend(patches,labels, bbox_to_anchor=(0.85,0.05), loc="lower left")
plt.axis('equal')
plt.tight_layout()
plt.show()
```

Untuk tahun 2008:

```
#2008
dfbarat2008 = dfbarat.loc[df["tahun"] == 2008]
dftengah2008 = dftengah.loc[df["tahun"] == 2008]
dftimur2008 = dftimur.loc[df["tahun"] == 2008]
d = dfbarat2008["persen_penduduk_miskin"].mean()
e = dftengah2008["persen_penduduk_miskin"].mean()
f = dftimur2008["persen_penduduk_miskin"].mean()

#          persen_penduduk_miskin
#barat          15.058029
#tengah          15.866818
#timur           31.150444
#total           62.075291

dictppm8 = {"persen_penduduk_miskin" : pd.Series([d,e,f], index = ["barat","tengah","timur"])}
dfppm8 = pd.DataFrame(dictppm8)
print(dfppm8)
dfppm8.plot(kind = "pie",subplots = True,title = "Persen penduduk miskin pada 3 wilayah Indonesia (2008, total 62.06%)")

labels = [r'barat (15.05%)', r'tengah (15.86%)',
r'timur (31.15%)]
sizes = [24.25, 25.56, 50.18]
colors = ['yellowgreen', 'lightskyblue', 'lightcoral']
patches, texts = plt.pie(sizes, colors=colors, startangle=0)
plt.legend(patches, labels, bbox_to_anchor=(0.85,0.05), loc="lower left")
plt.axis('equal')
plt.tight_layout()
plt.show()
```

Untuk tahun 2009:

```
#2009
dfbarat2009 = dfbarat.loc[df["tahun"] == 2009]
dftengah2009 = dftengah.loc[df["tahun"] == 2009]
dftimur2009 = dftimur.loc[df["tahun"] == 2009]
g = dfbarat2009["persen_penduduk_miskin"].mean()
h = dftengah2009["persen_penduduk_miskin"].mean()
i = dftimur2009["persen_penduduk_miskin"].mean()

dictppm9 = {"persen_penduduk_miskin" : pd.Series([g,h,i], index = ["barat","tengah","timur"])}
dfppm9 = pd.DataFrame(dictppm9)
print(dfppm9)
dfppm9.plot(kind = "pie",subplots = True,title = "Persen penduduk miskin pada 3 wilayah Indonesia (2009, total 60.67%)")
#      persen_penduduk_miskin
#barat      14.144240
#tengah      14.496194
#timur      32.044528
#total      60.67
labels = [r'barat (14.14%)', r'tengah (14.49%)',
r'timur (32.04%)]
sizes = [23.31, 23.88, 52.81]
colors = ['yellowgreen', 'lightskyblue', 'lightcoral']
patches, texts = plt.pie(sizes, colors=colors, startangle=0)
plt.legend(patches, labels, bbox_to_anchor=(0.85,0.05), loc="lower left")
plt.axis('equal')
plt.tight_layout()
plt.show()
```

Untuk tahun 2010:

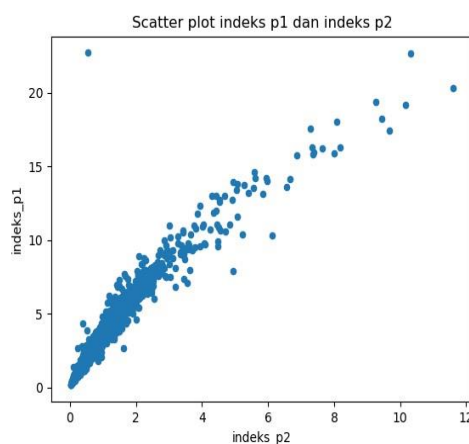
```
#2010
dfbarat2010 = dfbarat.loc[df["tahun"] == 2010]
dftengah2010 = dftengah.loc[df["tahun"] == 2010]
dftimur2010 = dftimur.loc[df["tahun"] == 2010]
j = dfbarat2010["persen_penduduk_miskin"].mean()
k = dftengah2010["persen_penduduk_miskin"].mean()
l = dftimur2010["persen_penduduk_miskin"].mean()

dictppm10 = {"persen_penduduk_miskin" : pd.Series([j,k,l], index = ["barat","tengah","timur"])}
dfppm10 = pd.DataFrame(dictppm10)
print(dfppm10)
dfppm10.plot(kind = "pie",subplots = True,title = "Persen penduduk miskin pada 3 wilayah Indonesia (2010, total 67.08%)")
#      persen_penduduk_miskin
#barat      13.074815
#tengah      23.565000
#timur      30.457000
#total      67.08
labels = [r'barat (13.07%)', r'tengah (23.56%)',
r'timur (30.45%)]
sizes = [19.48, 35.12, 45.39]
colors = ['yellowgreen', 'lightskyblue', 'lightcoral']
patches, texts = plt.pie(sizes, colors=colors, startangle=0)
plt.legend(patches, labels, bbox_to_anchor=(0.85,0.05), loc="lower left")
plt.axis('equal')
plt.tight_layout()
plt.show()
```

4. Plotting Relationship

Untuk Plotting relationship, kami mengambil 6 scatter plot yang paling relevan untuk divisualisasikan.

- indeks_p1 dengan indeks_p2 (koef korelasi :0.9479857205698)



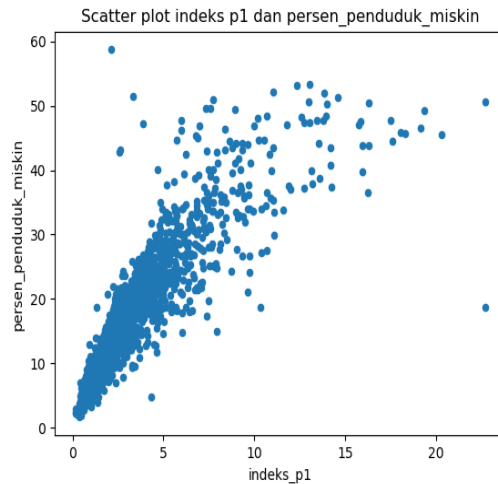
Insight :

Indeks p1 sangat berkorelasi lurus dengan indeks p2, karena berdasarkan referensi yang kami cari, perumusan indeks p2 diturunkan dari indeks p1. Perlu diingat bahwa indeks p2 tidak kami gambarkan di plot ini karena indeks p2 tidak jarang sekali dipakai dan indeks p2 lebih sulit untuk diimplementasikan.

Source code:

```
#Nilai Korelasi antara indeks p1 dan indeks p2
print(df["indeks_p2"].corr(df["indeks_p1"]))
df.plot.scatter(x = "indeks_p2", y = "indeks_p1", title = "Scatter plot indeks p1 dan indeks p2")
plt.show()
```

- indeks_p1 dan persen_penduduk_miskin (koef korelasi:0.8669342652637)



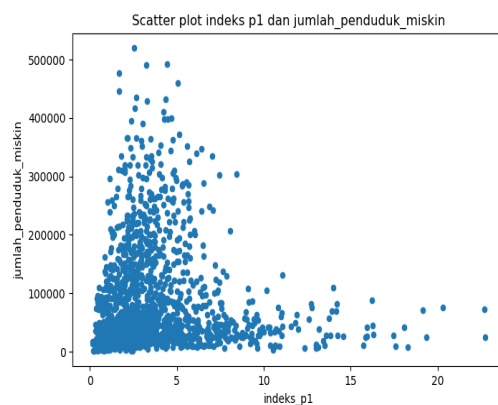
Insight:

Indeks p1 dengan persen_penduduk miskin setiap kabupaten sangat berkorelasi lurus. Hal ini cukup jelas apabila kita perhatikan jika indeks p1 semakin kecil, semakin besar potensi ekonomi untuk pengentasan kemiskinan. Yang berarti persen penduduk miskin harus kecil agar potensi tersebut besar.

Source Code:

```
#nilai korelasi antara indeks p1 dengan persen_penduduk_miskin
print(df["indeks_p1"].corr(df["persen_penduduk_miskin"]))
df.plot.scatter(x = "indeks_p1", y = "persen_penduduk_miskin", title = "Scatter plot indeks p1 dan persen_penduduk_miskin")
plt.show()
```

- indeks_p1 dengan jumlah_penduduk_miskin(koef korelasi: 0.0801582924054)



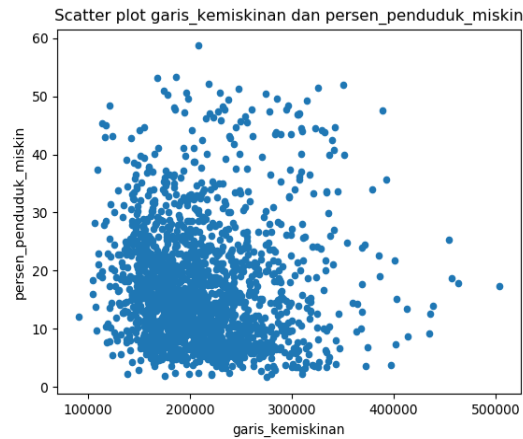
Insight:

Jumlah penduduk miskin di setiap kabupaten sangat tidak berkorelasi dengan indeks p1.

Source Code:

```
#nilai korelasi antara indeks p1 dengan jumlah_penduduk_miskin
print(df["indeks_p1"].corr(df["jumlah_penduduk_miskin"]))
df.plot.scatter(x = "indeks_p1", y = "jumlah_penduduk_miskin", title = "Scatter plot indeks p1 dan jumlah_penduduk_miskin")
plt.show()
```

- garis_kemiskinan dengan persen_penduduk_miskin(koef korelasi :0.0357986876499)

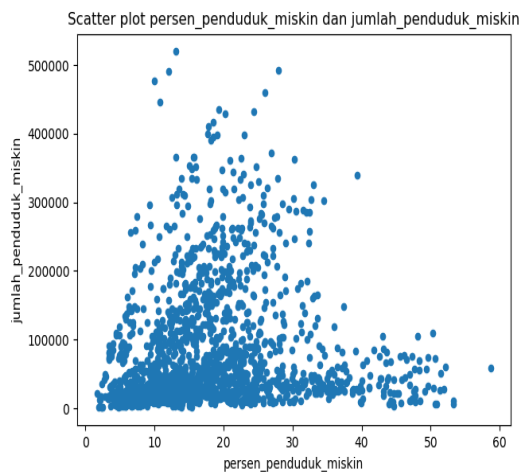


Insight:
Garis_kemiskinan dengan
persen_penduduk_miskin sangat
tidak berkorelasi (terbalik).

Source Code:

```
#nilai korelasi antara garis_kemiskinan dengan persen_penduduk_miskin
print(df["garis_kemiskinan"].corr(df["persen_penduduk_miskin"]))
df.plot.scatter(x = "garis_kemiskinan", y = "persen_penduduk_miskin", title = "Scatter plot garis_kemiskinan dan persen_penduduk_miskin")
plt.show()
```

- persen_penduduk_miskin dan jumlah_penduduk_miskin(koef korelasi:0.1682279326211)

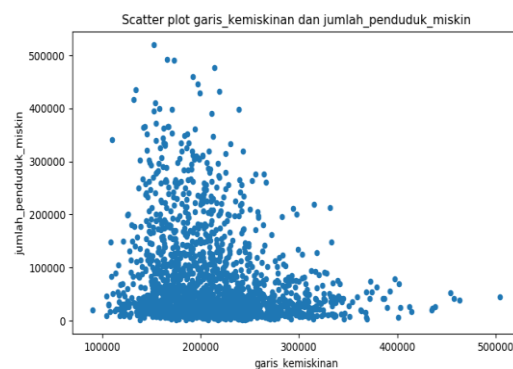


Insight:
Persen_penduduk_miskin dengan
jumlah_penduduk_miskin tidak
berkorelasi karena menurut kami
jumlah penduduk miskin memiliki
jumlah yang bervariasi di setiap
daerahnya, begitu juga dengan
komposisi persen penduduk miskin
di setiap daerahnya

Source Code:

```
#nilai korelasi antara persen_penduduk_miskin dengan jumlah_penduduk_miskin
print(df["persen_penduduk_miskin"].corr(df["jumlah_penduduk_miskin"]))
df.plot.scatter(x = "persen_penduduk_miskin", y = "jumlah_penduduk_miskin", title = "Scatter plot persen_penduduk_miskin dan jumlah_penduduk_miskin")
plt.show()
```

- garis_kemiskinan dengan jumlah_penduduk_miskin(koef korelasi :-0.1897423461805)

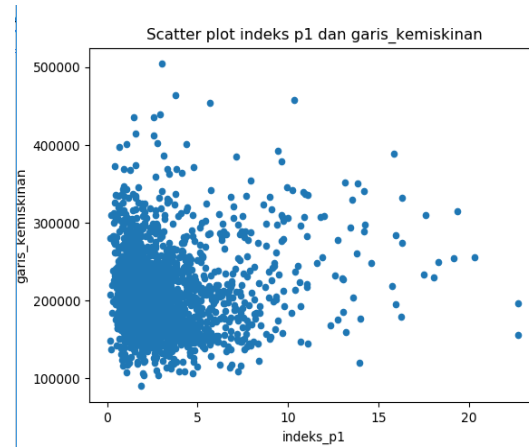


Insight:
Garis_kemiskinan dengan
jumlah_penduduk_miskin tidak
berkorelasi (terbalik) karena
menurut kami, tingkat sumber daya
alam dan sumber daya manusia di
setiap daerah bervariasi.

Source Code:

```
#nilai korelasi antara garis_kemiskinan dengan jumlah_penduduk_miskin
print(df["garis_kemiskinan"].corr(df["jumlah_penduduk_miskin"]))
df.plot.scatter(x = "garis_kemiskinan", y = "jumlah_penduduk_miskin", title = "Scatter plot garis_kemiskinan dan jumlah_penduduk_miskin")
plt.show()
```

- indeks_p1 dengan garis_kemiskinan (koef korelasi :0.0515064261464)



Insight :

Indeks p1 / Indeks kedalaman kemiskinan sangat tidak berkorelasi dengan garis_kemiskinan, karena menurut kami garis kemiskinan bergantung pada sumber daya alam dan sumber daya manusia yang berada pada daerah tersebut, karena garis kemiskinan bervariasi dan indeks p1 juga bervariasi, maka keduanya tidak saling bergantung/berkorelasi

Source Code:

```
#nilai korelasi antara indeks p1 dengan garis_kemiskinan
print(df["indeks_p1"].corr(df["garis_kemiskinan"]))
df.plot.scatter(x = "indeks_p1", y = "garis_kemiskinan", title = "Scatter plot indeks p1 dan garis_kemiskinan")
plt.show()
```

BAB 6

KEKOTORAN DATA DAN MENGATASINYA

Data yang kami ambil dari data.go.id memiliki beberapa kesalahan meskipun dapat dihitung jari. Ada beberapa data yang memiliki nilai yang kosong. Kami memilih untuk menghapus data yang kosong tersebut karena data kosong tersebut mengganggu perhitungan nilai ekstremum. Cara kami menentukan banyaknya data yang kosong tersebut dan cara membersihkannya adalah dengan sintaks sebagai berikut:

```
#Meload data file menjadi dataframe dengan variabel df
dfkotor = pd.read_csv("processed-penduduk-miskin-indeks-kemiskinan-2007-2010.csv")
print(dfkotor.isnull().sum())
df = dfkotor.dropna() # membersihkan data kosong
print(df.isnull().sum())
```

Fungsi `isnull().sum()` menentukan banyaknya data yang kosong disetiap atribut, maka, sintaks `"print(dfkotor.isnull().sum())"` akan keluar output berupa:

```
kode_provinsi      0
nama_provinsi      0
kode_kabkota       0
nama_kabkota       0
tahun              0
jumlah_penduduk_miskin    2
persen_penduduk_miskin    2
indeks_p1          2
indeks_p2          2
garis_kemiskinan     2
dtype: int64
```

Angka 2 pada setiap atribut diatas, menunjukkan adanya 2 data pada atribut tersebut yang memiliki data yang kosong, setelah di "cuci" oleh fungsi `.dropna()`, dan kami membuat data yang bersih tersebut menjadi variabel `df`, maka apabila kita gunakan kembali `isnull().sum()`, outputnya menjadi:

```
--
kode_provinsi      0
nama_provinsi      0
kode_kabkota       0
nama_kabkota       0
tahun              0
jumlah_penduduk_miskin    0
persen_penduduk_miskin    0
indeks_p1          0
indeks_p2          0
garis_kemiskinan     0
dtype: int64
```

Ini menunjukkan bahwa data sudah bersih dan tidak ada nilai data yang kosong.

Langkah yang sama digunakan untuk membersihkan data pertanian, namun untuk keefektifan laporan, kami gabungkan pembahasan kekotoran data pada bab ini.

Pembagian Tugas

1. Andhika Rahadian 16518123
Mengerjakan : Laporan dan Powerpoint Visualisasi, Korelasi dari Data Kemiskinan. Menyunting, Menyatukan Powerpoint & Laporan Data Kemiskinan dan Data Pertanian
2. Faris Kautsar 16518327
Mengerjakan: Laporan dan Powerpoint Statistika, Deskripsi, Sampel, Makna Atribut dari Data Kemiskinan
3. Raras Pradnya Pramudita 16518021
Mengerjakan: Laporan dan Powerpoint Statistika, Deskripsi, Sampel, Makna Atribut dari Data Pertanian
4. Stefanus Gusega Gunawan 16518225
Mengerjakan : Laporan dan Powerpoint Visualisasi, Korelasi dari Data Pertanian