

LAPORAN
MATA KULIAH ANALISIS DATA EKSPLORATIF (A)
“GROUPING DATASET”



DISUSUN OLEH:

Reza Putri Angga (22083010006)

DOSEN PENGAMPU:

Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR
2023

STUDI KASUS DAN PEMBAHASAN

Grouping Dataset : Pivoting Table

Untuk melakukan *pivoting table* pada dataset “data_bahan_pokok.csv” perlu di lakukan beberapa langkah-langkah terlebih dahulu, di antaranya yakni :

- Melakukan **Load Dataset "data_bahan_pokok.csv" Untuk Melakukan Grouping Pada Dataset**

```
• Melakukan Load Dataset "data_bahan_pokok.csv" Untuk Melakukan Grouping Pada Dataset

In [1]: import pandas as pd

BahanPokokDataset = pd.read_csv("data_bahan_pokok.csv", sep = ";")
BahanPokokDataset

Out[1]:
```

	TANGGAL	BARANG	SUPLIER	BULAN	TAHUN	KUARTAL	JUMLAH	HARGA	TOTAL
0	01/01/15	Cabai	Bejo	January	2015	Kuartal 1	55	IDR30.550	IDR1.680.250
1	06/01/15	Kunyit	Bejo	January	2015	Kuartal 1	72	IDR12.964	IDR933.408
2	11/01/15	Bawang Merah	Painah	January	2015	Kuartal 1	81	IDR28.875	IDR2.338.875
3	18/01/15	Bawang Putih	Tarno	January	2015	Kuartal 1	146	IDR20.812	IDR3.038.552
4	22/01/15	Kencur	Bejo	January	2015	Kuartal 1	159	IDR9.785	IDR1.555.815

Pada kode script di atas, di lakukan proses *load* (pembacaan) dataset data bahan pokok dengan jenis file “.csv” menggunakan *library* pandas yang di permisalkan sebagai pd. Terdapat variabel BahanPokokDataset yang di pergunakan sebagai variabel untuk menyimpan *dataframe* dari dataset data bahan pokok dengan *sep* (pemisah) menggunakan “;”. Dalam dataset tersebut terdapat data sebanyak 216 baris dan 9 kolom.

- Menampilkan Nama-Nama Kolom Pada "data_bahan_pokok.csv"

```
• Menampilkan Nama-Nama Kolom Pada "data_bahan_pokok.csv"

In [2]: BahanPokokKolom = BahanPokokDataset.columns

print("Di Tampilkan Nama Kolom Dari Dataset Data Bahan Pokok :")
BahanPokokKolom

Di Tampilkan Nama Kolom Dari Dataset Data Bahan Pokok :

Out[2]: Index(['TANGGAL', 'BARANG', 'SUPLIER', 'BULAN', 'TAHUN', 'KUARTAL', 'JUMLAH',
              'HARGA', 'TOTAL'],
              dtype='object')
```

Pada kode script di atas di lakukan proses penampilan nama kolom yang terdapat dalam dataset data bahan pokok. Terdapat variabel BahanPokokKolom yang di pergunakan untuk menuliskan proses pencarian dan pembacaan nama kolom. Maka, akan di tampilkan bahwa dataset data bahan pokok memiliki sembilan kolom, yakni tanggal, barang, suplier, bulan, tahun, kuartal, jumlah, harga, dan total. Di mana data-data pada kolom ini nantinya akan di proses lebih lanjut untuk melakukan *grouping dataset* dan *pivoting table*.

Lalu, bisa di lanjutkan dengan menjawab beberapa pertanyaan mengenai pivoting table, di antaranya yakni :

Lakukanlah Pivoting Tabel Menggunakan Kode Script Pemrograman Python Pembelian Bahan Pokok Makanan Yang Di Kirimkan Oleh Supplier Pada Kuartal Tertentu Menjadi Tabel Berikut:

A. Pivoting Tabel Berdasarkan Jumlah Pembelian Bahan Pokok (Kg) Per Kuartal

A. Pivoting Tabel Berdasarkan Jumlah Pembelian Bahan Pokok (Kg) Per Kuartal

```
In [42]: #define function untuk jumlah pembelian bahan pokok
def pivot_jumlah_pembelian_bahan>NamaBarang, BahanPokokDataset):

    #filter data berdasarkan kolom "BARANG" yang berisi nama-nama barang
    FilterData = BahanPokokDataset[BahanPokokDataset["BARANG"] ==>NamaBarang]

    #pengubahan data yang di filter menjadi tabel pivot
    PivotTable = pd.pivot_table(FilterData,
                                values = "JUMLAH",
                                index = "KUARTAL",
                                columns = "TAHUN",
                                aggfunc = "sum",
                                fill_value = 0,
                                margins = True,
                                margins_name = "Grand Total")

    #menyeleksi baris tertentu pada tabel pivot kemudian menambahkan baris nama barang ke jumlah pembelian bahan pokok
    PivotTable = PivotTable.loc[["Kuartal 1", "Kuartal 2", "Kuartal 3", "Kuartal 4", "Grand Total"]]
    PivotTable.loc>NamaBarang] = PivotTable.loc["Grand Total"]

    #penghapusan/mengosongkan nama index dan kolom
    PivotTable.index.name = None
    PivotTable.columns.name = None

    #pengembalian nilai tabel pivot yang telah di filter dan di modifikasi
    return PivotTable.loc[>NamaBarang, "Kuartal 1", "Kuartal 2", "Kuartal 3", "Kuartal 4"]

#pencarian nama-nama barang di kolom "BARANG"
UnikBarang = BahanPokokDataset["BARANG"].unique()
SortedBarang = sorted(UnikBarang)

#penggabungan tabel pivot yang telah di buat
SortedPivotTable = pd.concat([pivot_jumlah_pembelian_bahan(Barang, BahanPokokDataset) for Barang in SortedBarang])
SortedPivotTable.rename_axis("Row Labels", axis = 1, inplace = True)
print("Di Tampilkan Pivoting Tabel Berdasarkan Jumlah Pembelian Bahan Pokok (Kg) Per Kuartal : ")
SortedPivotTable
```

Di Tampilkan Pivoting Tabel Berdasarkan Jumlah Pembelian Bahan Pokok (Kg) Per Kuartal :

```
Out[42]:
```

Row Labels	2015	2016	2017	Grand Total
Bawang Merah	1450	1278	1574	4302
Kuartal 1	406	296	408	1110
Kuartal 2	277	248	427	952
Kuartal 3	431	403	430	1264
Kuartal 4	336	331	309	976
Bawang Putih	1419	1659	1601	4679
Kuartal 1	313	534	355	1202
Kuartal 2	407	447	443	1297
Kuartal 3	361	332	506	1199
Kuartal 4	338	346	297	981
Cabai	1543	1568	1612	4723
Kuartal 1	309	272	221	802
Kuartal 2	364	473	431	1268
Kuartal 3	519	366	453	1338
Kuartal 4	351	457	507	1315
Jahe	1317	1677	1659	4653
Kuartal 1	276	515	435	1226
Kuartal 2	251	262	387	900
Kuartal 3	363	488	426	1277
Kuartal 4	427	412	411	1250
Kencur	1607	1489	1376	4472
Kuartal 1	527	376	340	1243
Kuartal 2	345	328	420	1093
Kuartal 3	370	470	269	1109
Kuartal 4	365	315	347	1027
Kunyit	1640	1446	1419	4505
Kuartal 1	374	284	398	1056
Kuartal 2	439	345	255	1039
Kuartal 3	355	368	350	1073
Kuartal 4	472	449	416	1337

Pada kode script di atas di lakukan proses pivoting tabel berdasarkan jumlah pembelian bahan pokok (kg) perkuartal. Dengan data yang di gunakan adalah data bahan pokok yang telah di *load* sebelumnya dan di simpan dalam variabel BahanPokokDataset. Untuk melakukan pivoting tabel perlu di lakukan pembuatan

function dengan nama `pivot_jumlah_pembelian_bahan` dan terdapat parameter `NamaBarang` (nama barang yang akan di analisis) dan `BahanPokokDataset` (dataset). Di lakukan proses *filter* data berdasarkan kolom barang dan akan menghasilkan *dataframe* baru yang berisi data untuk nama barang tertentu. Terdapat variabel `PivotTabel` yang di pergunakan untuk penulisan pembuatan tabel pivot dengan nilai yang di isi berdasarkan kolom jumlah, *index* di isi berdasarkan nilai kuartal (3 bulan), kolom yang di isi berdasarkan kolom tahun kemudian di lakukan proses penjumlahan jumlah pembelian perkuartal menggunakan *aggfunc* dan melakukan penambahan baris dan kolom total keseluruhan.

Kemudian, di lakukan pemilihan baris-baris tertentu dari tabel pivot, yakni kuartal 1, kuartal 2, kuartal 3, kuartal 4, dan *grand total*. Dan di lakukan proses penambahan baris dengan nama barang yang telah di filter sebelumnya ke dalam tabel pivot, hal ini di lakukan agar tabel pivot mencakup jumlah pembelian bahan pokok perkuartal untuk barang tertentu. Lalu, di lakukan proses pengambilan nama barang unik dari kolom barang untuk mendapatkan daftar semua barang dan di lakukan proses *sorted*. Di lanjutkan dengan proses penggabungan dari setiap tabel pivot menjadi satu tabel besar dengan nama *index* kolom di ubah menjadi *row labels*.

Secara singkat, dapat di jelaskan bahwa kode tersebut di gunakan untuk membuat tabel pivot berdasarkan kolom jumlah dari dataset terdapat suatu fungsi dengan dua argumen yang di pergunakan untuk memfilter dataset berdasarkan kolom barang (nama barang) kemudian di buat tabel pivot dengan jumlah kuantitas pembelian pertahun yang di spesifikasikan perkuartal kemudian memilih baris dan menambahkan baris untuk mendapatkan total jumlah barang tersebut. Kemudian fungsi tersebut melakukan pengembalian nilai dari tabel pivot yang telah di buat dan terdapat proses *looping* yang membuat tabel pivot di gabungkan untuk semua barang dalam dataset. Maka, akan di tampilkan hasil dari tabel pivot yang menunjukkan jumlah pembelian bahan pokok (kg) perkuartal dengan nama barang bawang merah, bawang putih, cabai, jahe, kencur, dan kunyit. Tujuan pembuatan tabel pivot ini adalah untuk mengidentifikasi dan mengetahui jumlah pembelian tiap-tiap nama barang yang ada pada dataset dengan waktu perkuartal (3 bulan).

B. Pivoting Tabel Berdasarkan Total Pembelian Bahan Pokok (Rupiah) Per Kuartal

B. Pivoting Tabel Berdasarkan Total Pembelian Bahan Pokok (Rupiah) Per Kuartal

```
In [38]: #define function untuk total pembelian bahan pokok
def pivot_total_pembelian_bahan>NamaBarang, BahanPokokDataset):

    #pengecekan tipe data di kolom "TOTAL" dan mengonversinya menjadi integer
    if not pd.api.types.is_numeric_dtype(BahanPokokDataset["TOTAL"]):

        #penghapusan teks "IDR" dan tanda "." dari kolom "TOTAL"
        BahanPokokDataset["TOTAL"] = BahanPokokDataset["TOTAL"].str.replace("IDR", "").str.replace(".", "").astype(int)

    #filter data berdasarkan kolom "BARANG" yang berisi nama-nama barang
    FilterData = BahanPokokDataset[BahanPokokDataset["BARANG"] ==>NamaBarang]

    #pengubahan data yang di filter menjadi tabel pivot
    PivotTabel = pd.pivot_table(FilterData,
                                values = "TOTAL",
                                index = "KUARTAL",
                                columns = "TAHUN",
                                aggfunc = "sum",
                                fill_value = 0,
                                margins = True,
                                margins_name = "Grand Total")

    #menyeleksi baris tertentu pada tabel pivot kemudian menambahkan baris nama barang ke total pembelian bahan pokok
    PivotTabel = PivotTabel.loc[["Kuartal 1", "Kuartal 2", "Kuartal 3", "Kuartal 4", "Grand Total"]]
    PivotTabel.loc[>NamaBarang] = PivotTabel.loc["Grand Total"]

    #penghapusan/mengosongkan nama index dan kolom
    PivotTabel.index.name = None
    PivotTabel.columns.name = None

    #pengembalian nilai tabel pivot yang telah di filter dan di modifikasi
    return PivotTabel.loc[["Kuartal 1", "Kuartal 2", "Kuartal 3", "Kuartal 4"]]

#pencarian nama-nama barang di kolom "BARANG"
UnikBarang = BahanPokokDataset["BARANG"].unique()
SortedBarang = sorted(UnikBarang)
```

```
#penggabungan tabel pivot yang telah di buat
SortedPivotTabel = pd.concat([pivot_total_pembelian_bahan(Barang, BahanPokokDataset) for Barang in SortedBarang])
SortedPivotTabel.rename_axis("Row Labels", axis = 1, inplace = True)
print("Di Tampilkan Pivoting Tabel Berdasarkan Total Pembelian Bahan Pokok (Kg) Per Kuartal : ")
SortedPivotTabel
```

Di Tampilkan Pivoting Tabel Berdasarkan Total Pembelian Bahan Pokok (Kg) Per Kuartal :

```
Out[38]:
```

Row Labels	2015	2016	2017	Grand Total
Bawang Merah	37823531	31731767	40941357	110496655
Kuartal 1	11037115	6820091	10456137	28313343
Kuartal 2	7596610	6251504	11075213	24923327
Kuartal 3	10800386	10587768	11697016	33085170
Kuartal 4	8389420	8072404	7712991	24174815
Bawang Putih	31078455	37670606	35263522	104012583
Kuartal 1	6812135	12662576	7962251	27436962
Kuartal 2	8932531	10457967	9784401	29174899
Kuartal 3	8082531	7110937	10945924	26139392
Kuartal 4	7251258	7439126	6570946	21261330
Cabai	45057879	44441269	46168221	135667369
Kuartal 1	9219854	7652138	5787840	22659832
Kuartal 2	10617522	13099592	12230465	35947579
Kuartal 3	15774604	9718410	14312881	39805895
Kuartal 4	9445899	13971129	13837035	37254063
Jahe	29984636	36550966	38324202	104859804
Kuartal 1	6204072	11010367	10148692	27363131
Kuartal 2	5775616	5880108	8595257	20250981
Kuartal 3	7806801	10696194	10448307	28951302
Kuartal 4	10198147	8964297	9131946	28294390

Kencur	15030665	13539469	12267295	40837429
Kuartal 1	4981422	3332834	2922221	11236477
Kuartal 2	3362034	3039520	3769896	10171450
Kuartal 3	3173993	4161334	2297363	9632690
Kuartal 4	3513216	3005781	3277815	9796812
Kunyit	19931524	18179929	18339347	56450800
Kuartal 1	4048734	3119536	4489786	11658056
Kuartal 2	5524089	4272801	3401235	13198125
Kuartal 3	4177633	5073847	4809108	14060588
Kuartal 4	6181068	5713745	5639218	17534031

Pada kode script di atas, di lakukan proses pivoting tabel berdasarkan total pembelian bahan pokok (kg) perkuartal. Dengan data yang di gunakan adalah data

bahan pokok yang telah di load sebelumnya dan di simpan dalam variabel *BahanPokokDataset*. Untuk melakukan pivoting tabel perlu di lakukan pembuatan *function* dengan nama *pivot_total_pembelian_bahan* dan terdapat parameter *NamaBarang* (nama barang yang akan di analisis) dan *BahanPokokDataset* (dataset). Kemudian di lakukan pengecekan dan perubahan tipe data di kolom total menjadi numerik dengan penghilangan teks “IDR” dan tanda “.” Kemudian di konversi menjadi tipe data integer.

Di lakukan proses *filter* data yang ada dalam kolom barang dan di simpan dalam argumen *NamaBarang* dan akan menghasilkan *dataframe* baru yang hanya berisi data untuk barang tertentu. Terdapat variabel *PivotTabel* yang di pergunakan untuk penulisan pembuatan tabel pivot dengan nilai yang di isi berdasarkan kolom total, *index* di isi berdasarkan nilai kuartal (3 bulan), kolom di isi berdasarkan kolom tahun kemudian di lakukan proses penjumlahan total pembelian bahan pokok perkuartal menggunakan *aggfunc* dan melakukan penambahan baris dan kolom total keseluruhan. Pembuatan tabel pivot tersebut menggambarkan total pembelian bahan pokok perkuartal dan pertahun.

Kemudian, di lakukan pemilihan baris tetentu dari tabel pivot, yakni kuartal 1, kuartal 2, kuartal 3, kuartal 4, dan *grand total*. Dan di lakukan penambahan baris dengan nama barang yang telah di filter ke tabel pivot agar terdapat total pembelian bahan pokok perkuartal untuk barang tertentu. Lalu, di lakukan proses pengambilan nama barang unik dari kolom barang untuk mendapatkan daftar semua barang dan di lakukan proses *sorted*. Di lanjutkan dengan proses penggabungan dari setiap tabel pivot menjadi satu tabel besar dengan nama *index* kolom di ubah menjadi *row labels*.

Secara singkat, dapat di jelaskan bahwa kode tersebut di gunakan untuk membuat tabel pivot berdasarkan kolom total dari dataset terdapat satu fungsi dengan dua argumen yang di pergunakan untuk melakukan pengecekan tipe data pada kolom total dan mengkonversinya ke dalam tipe data integer. Fungsi tersebut juga di pergunakan untuk memfilter dataset berdasarkan kolom barang (nama barang) dan membuat tabel pivot dengan jumlah total pembelian perkuartal dan pertahun kemudian di lakukan proses penambahan baris untuk total jumlah pembelian barang tersebut. Kemudian di lakukan proses pengembalian nilai dari tabel pivot yang telah di buat dan terdapat proses *looping* untuk membuat tabel pivot yang di gabungkan untuk semua barang dalam dataset. Maka, akan di tampilkan hasil dari tabel pivot yang menunjukkan total pembelian bahan pokok (kg) perkuartal dan pertahun dengan nama barang bawang merah, bawang putih, jahe, kencur, dan kunyit. Tujuan pembuatan tabel pivot ini adalah untuk mengidentifikasi dan mengetahui total pembelian tiap-tiap nama barang yang ada pada dataset dengan waktu perkuartal (3 bulan).