


<p>Nama: Dave Ryano F.M</p> <p>NIM: 064002300039</p>	 <p>Praktikum Statistika & Probabilitas</p>	<p>MODUL 3</p> <p>Nama Dosen: Drs. Joko Riyono, M.Si.</p>
<p>Hari/Tanggal: Sabtu, 30 Maret 2024</p>		<p>Nama Asisten Labratorium: 1. Adzriel Yusak Noah (064.20.11) 2. Muhammad Fahmi (064.22.36)</p>

Pengelolaan Data pada Data Frame

1. Teori Singkat

R (juga dikenal sebagai GNU S) adalah bahasa pemrograman dan perangkat lunak untuk analisis statistika dan grafik. R dibuat oleh Ross Ihaka dan Robert Gentleman di Universitas Auckland, Selandia Baru, dan kini dikembangkan oleh R Development Core Team, di mana Chambers merupakan anggotanya. R dinamakan sebagian setelah nama dua pembuatnya (Robert Gentleman dan Ross Ihaka), dan sebagian sebagian dari permainan nama dari S.

Bahasa R kini menjadi standar de facto di antara statistikawan untuk pengembangan perangkat lunak statistika, serta digunakan secara luas untuk pengembangan perangkat lunak statistika dan analisis data. R merupakan bagian dari proyek GNU. Kode sumbernya tersedia secara bebas di bawah Lisensi Publik Umum GNU, dan versi biner prekompilasinya tersedia untuk berbagai sistem operasi. R menggunakan antarmuka baris perintah, meski beberapa antarmuka pengguna grafik juga tersedia.

R menyediakan berbagai teknik statistika (permodelan linier dan nonlinier, uji statistik klasik, analisis deret waktu, klasifikasi, klusterisasi, dan sebagainya) serta grafik. R, sebagaimana S, dirancang sebagai bahasa komputer sebenarnya, dan mengizinkan penggunaannya untuk menambah fungsi tambahan dengan mendefinisikan fungsi baru. Kekuatan besar dari R yang

lain adalah fasilitas grafiknya, yang menghasilkan grafik dengan kualitas publikasi yang dapat memuat simbol matematika. R memiliki format dokumentasi seperti LaTeX, yang digunakan untuk menyediakan dokumentasi yang lengkap, baik secara daring (dalam berbagai format) maupun secara cetakan.

RStudio merupakan integrated development environment (IDE) khusus bagi bahasa pemrograman R. Software ini menyediakan R console, code editor dengan syntax highlighting, code completion dan direct execution, environment, history, connections, dan fitur-fitur tambahan lainnya seperti file manager, packages manager, help, plot viewer, hingga project versioning menggunakan git. RStudio sebenarnya memiliki dua versi, yaitu open source (gratis) dan commercial edition (berbayar). RStudio juga tidak hanya terbatas dalam bentuk aplikasi dekstop, melainkan terdapat versi RStudio Server, yaitu RStudio yang dapat diakses melalui browser yang terhubung dengan suatu jaringan komputer. Untuk saat ini, versi RStudio yang akan dijelaskan hanyalah RStudio open source berbasis dekstop saja.

2. Alat dan Bahan

Hardware : Laptop/PC

Software : R Studio



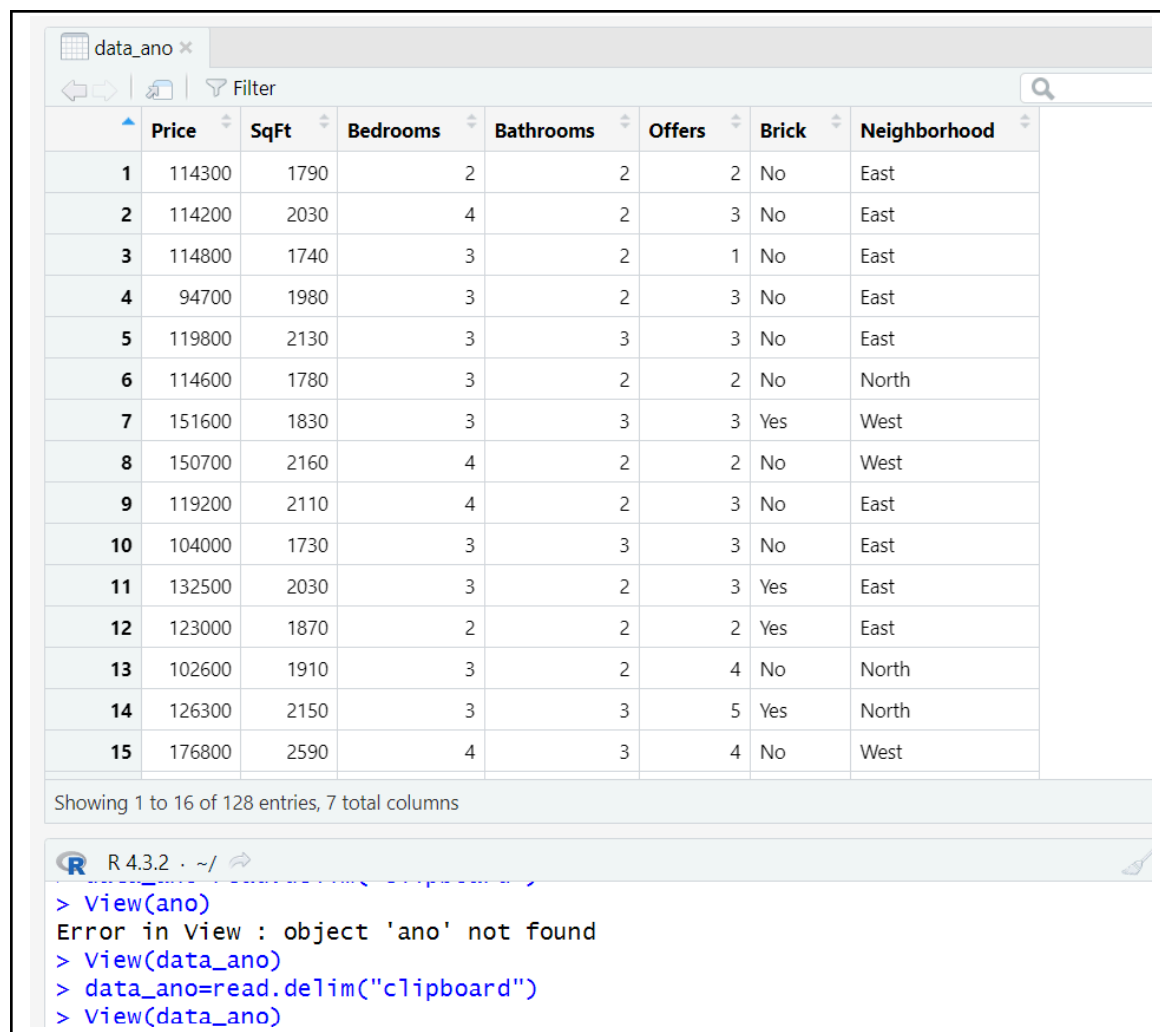
3. Elemen Kompetensi

a. Latihan pertama – Merge Data

1. Jalankan source code berikut. Ganti nama variable (seperti data_nama) menjadi variable dengan nama kalian masing - masing. Data yang digunakan adalah data houseprice.csv. Melakukan Read CSV dengan cara menginput data houseprice, sesuaikan dengan lokasi direktori dimana kalian menyimpan file csvnya.

```
data_nama = read.delim("clipboard")  
View(data_nama)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	2	2	No	East
2	114200	2030	4	2	3	No	East
3	114800	1740	3	2	1	No	East
4	94700	1980	3	2	3	No	East
5	119800	2130	3	3	3	No	East
6	114600	1780	3	2	2	No	North
7	151600	1830	3	3	3	Yes	West
8	150700	2160	4	2	2	No	West
9	119200	2110	4	2	3	No	East
10	104000	1730	3	3	3	No	East
11	132500	2030	3	2	3	Yes	East
12	123000	1870	2	2	2	Yes	East
13	102600	1910	3	2	4	No	North
14	126300	2150	3	3	5	Yes	North
15	176800	2590	4	3	4	No	West

Showing 1 to 16 of 128 entries, 7 total columns

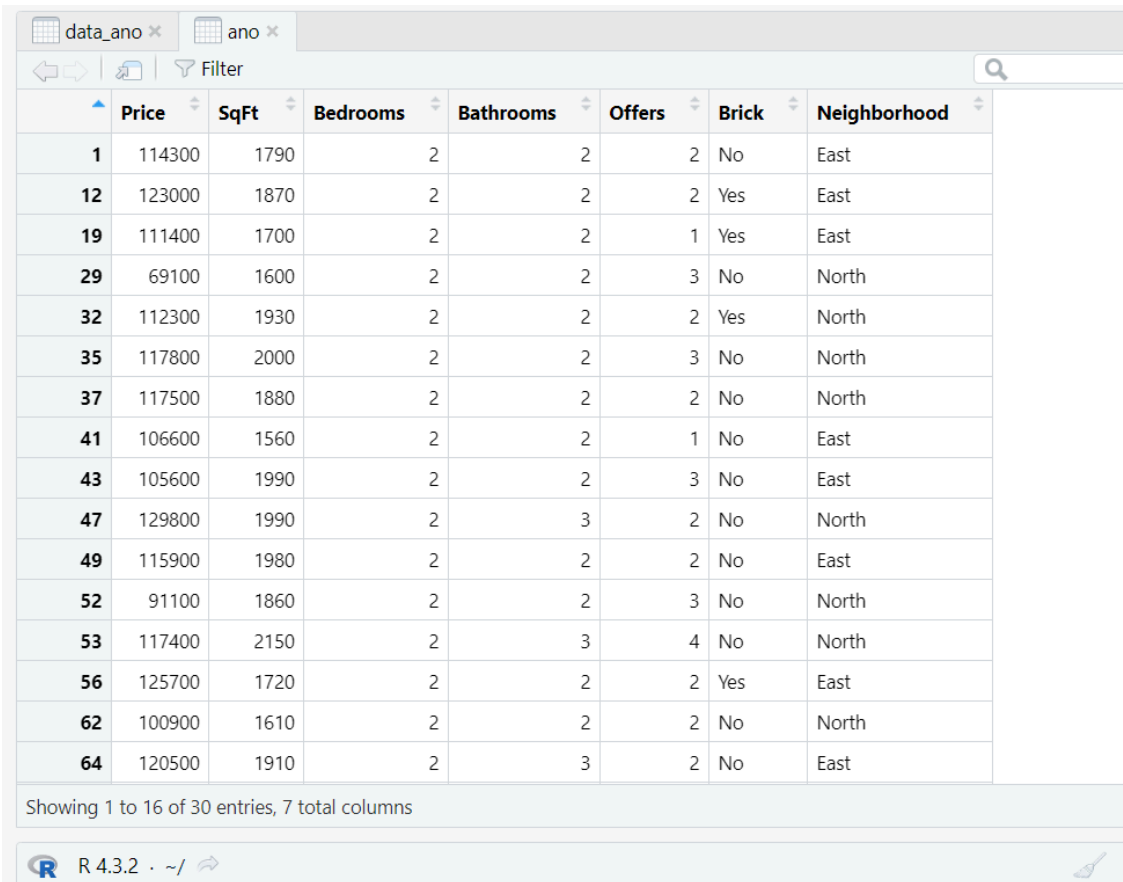
```
R 4.3.2 · ~/ > View(ano)  
Error in View : object 'ano' not found  
> View(data_ano)  
> data_ano=read.delim("clipboard")  
> View(data_ano)
```



2. Lakukan subset data dengan cara sub set data frame khusus yang memiliki nilai variabel Bedrooms =2

```
nama <- data_nama[data_nama$Bedrooms==2,]
View(nama)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	2	2	No	East
12	123000	1870	2	2	2	Yes	East
19	111400	1700	2	2	1	Yes	East
29	69100	1600	2	2	3	No	North
32	112300	1930	2	2	2	Yes	North
35	117800	2000	2	2	3	No	North
37	117500	1880	2	2	2	No	North
41	106600	1560	2	2	1	No	East
43	105600	1990	2	2	3	No	East
47	129800	1990	2	3	2	No	North
49	115900	1980	2	2	2	No	East
52	91100	1860	2	2	3	No	North
53	117400	2150	2	3	4	No	North
56	125700	1720	2	2	2	Yes	East
62	100900	1610	2	2	2	No	North
64	120500	1910	2	3	2	No	East

Showing 1 to 16 of 30 entries, 7 total columns

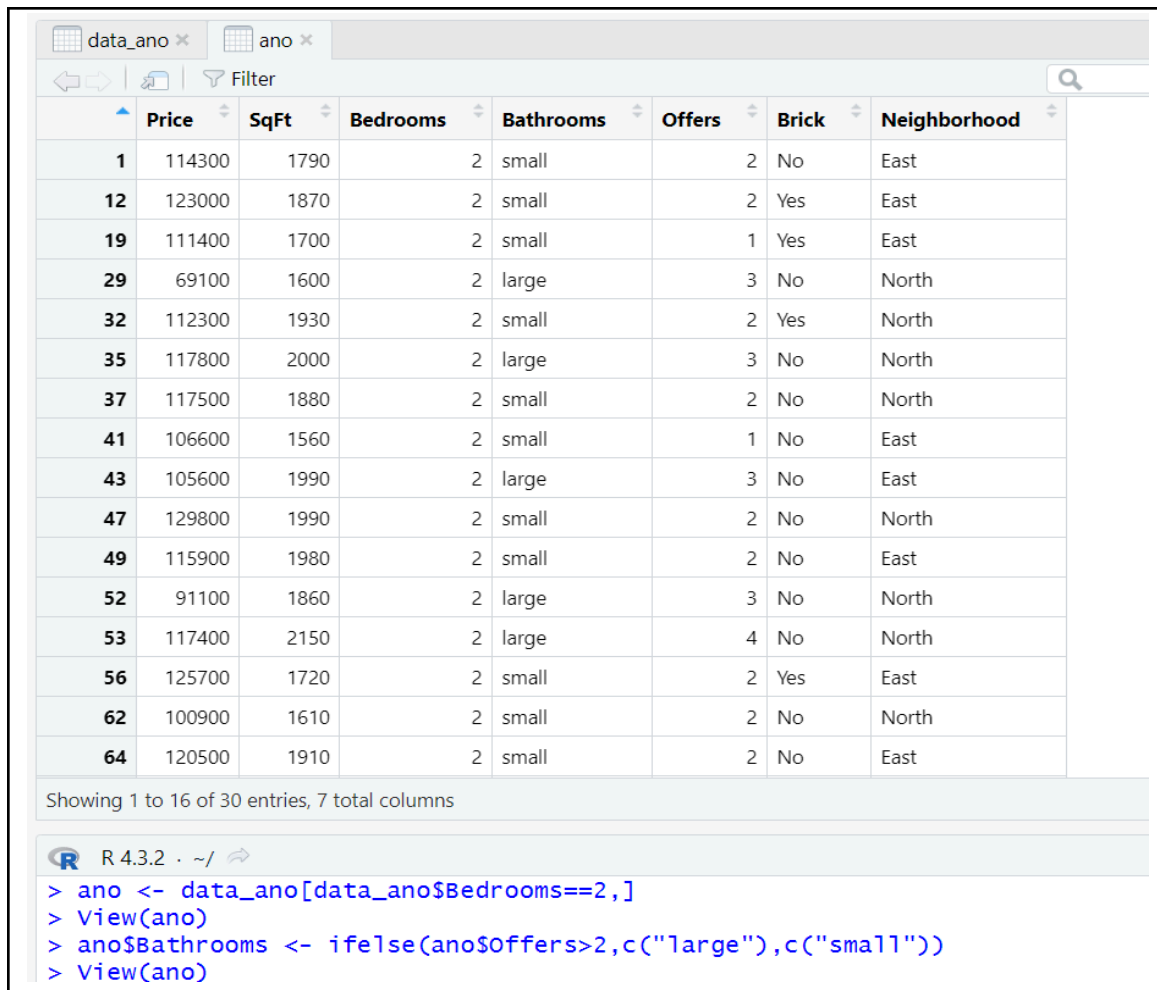
```
R 4.3.2 · ~/
> data_ano=read.delim("clipboard")
> View(data_ano)
> ano <- data_ano[data_ano$Bedrooms==2,]
> View(ano)
```



3. Selanjutnya ialah mengubah nama variabel. Berikut adalah tahapan untuk mengubah nilai dalam kolom Bathrooms dalam kondisi jika Bathrooms >2 , maka akan diganti dengan nilai large, jika tidak maka small.

```
nama$Bathrooms <- ifelse(nama$Bathrooms > 2, c("large"), c("small"))  
View(nama)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	small	2	No	East
12	123000	1870	2	small	2	Yes	East
19	111400	1700	2	small	1	Yes	East
29	69100	1600	2	large	3	No	North
32	112300	1930	2	small	2	Yes	North
35	117800	2000	2	large	3	No	North
37	117500	1880	2	small	2	No	North
41	106600	1560	2	small	1	No	East
43	105600	1990	2	large	3	No	East
47	129800	1990	2	small	2	No	North
49	115900	1980	2	small	2	No	East
52	91100	1860	2	large	3	No	North
53	117400	2150	2	large	4	No	North
56	125700	1720	2	small	2	Yes	East
62	100900	1610	2	small	2	No	North
64	120500	1910	2	small	2	No	East

Showing 1 to 16 of 30 entries, 7 total columns

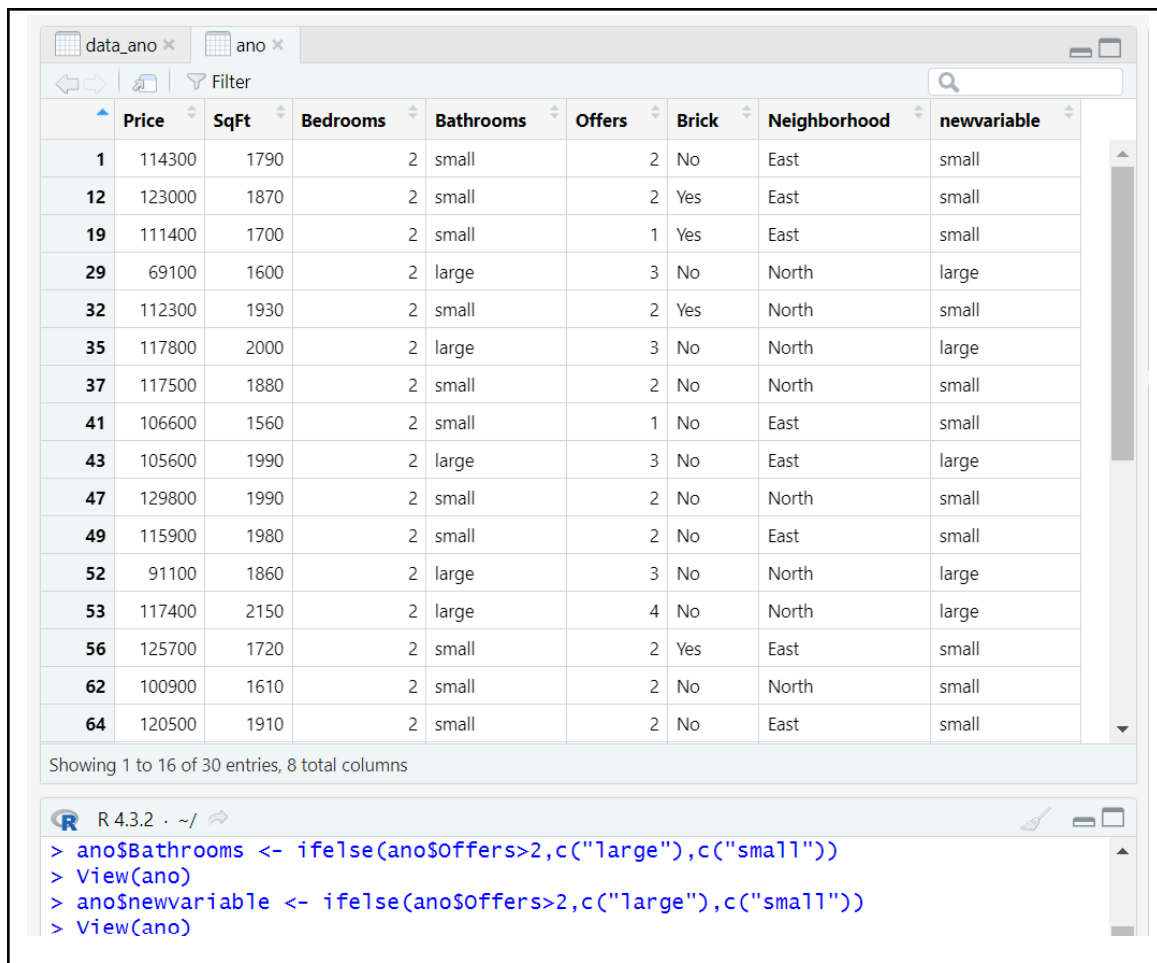
```
R 4.3.2 · ~/  
> ano <- data_ano[data_ano$Bedrooms==2,]  
> View(ano)  
> ano$Bathrooms <- ifelse(ano$Offers>2,c("large"),c("small"))  
> View(ano)
```



4. Lalu langkah selanjutnya adalah menambah variabel. Tahap 1 □ Membuat variable baru dari dataku1 dengan nilai sesuai kondisi ifelse yang ditentukan. Sesuaikan nama variable baru dengan nama praktikan.

```
nama$newvariable <- ifelse(nama$Offers > 2, c("large"), c("small"))  
View(nama)
```

Output



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood	newvariable
1	114300	1790	2	small	2	No	East	small
12	123000	1870	2	small	2	Yes	East	small
19	111400	1700	2	small	1	Yes	East	small
29	69100	1600	2	large	3	No	North	large
32	112300	1930	2	small	2	Yes	North	small
35	117800	2000	2	large	3	No	North	large
37	117500	1880	2	small	2	No	North	small
41	106600	1560	2	small	1	No	East	small
43	105600	1990	2	large	3	No	East	large
47	129800	1990	2	small	2	No	North	small
49	115900	1980	2	small	2	No	East	small
52	91100	1860	2	large	3	No	North	large
53	117400	2150	2	large	4	No	North	large
56	125700	1720	2	small	2	Yes	East	small
62	100900	1610	2	small	2	No	North	small
64	120500	1910	2	small	2	No	East	small

Showing 1 to 16 of 30 entries, 8 total columns

```
R 4.3.2 · ~/> ano$Bathrooms <- ifelse(ano$Offers>2,c("large"),c("small"))  
> View(ano)  
> ano$newvariable <- ifelse(ano$Offers>2,c("large"),c("small"))  
> View(ano)
```



Tahap 2 □ Mengubah isi baris data dari kolom baru yang telah dibuat

```
nama$newvariable <- nama$Price/nama$SqFt
View(nama)
```

Output:

	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood	newvariable
1	114300	1790	2	small	2	No	East	TRUE
12	123000	1870	2	small	2	Yes	East	TRUE
19	111400	1700	2	small	1	Yes	East	TRUE
29	69100	1600	2	large	3	No	North	TRUE
32	112300	1930	2	small	2	Yes	North	TRUE
35	117800	2000	2	large	3	No	North	TRUE
37	117500	1880	2	small	2	No	North	TRUE
41	106600	1560	2	small	1	No	East	TRUE
43	105600	1990	2	large	3	No	East	TRUE
47	129800	1990	2	small	2	No	North	TRUE
49	115900	1980	2	small	2	No	East	TRUE
52	91100	1860	2	large	3	No	North	TRUE
53	117400	2150	2	large	4	No	North	TRUE
56	125700	1720	2	small	2	Yes	East	TRUE
62	100900	1610	2	small	2	No	North	TRUE
64	120500	1910	2	small	2	No	East	TRUE

Showing 1 to 16 of 30 entries, 8 total columns

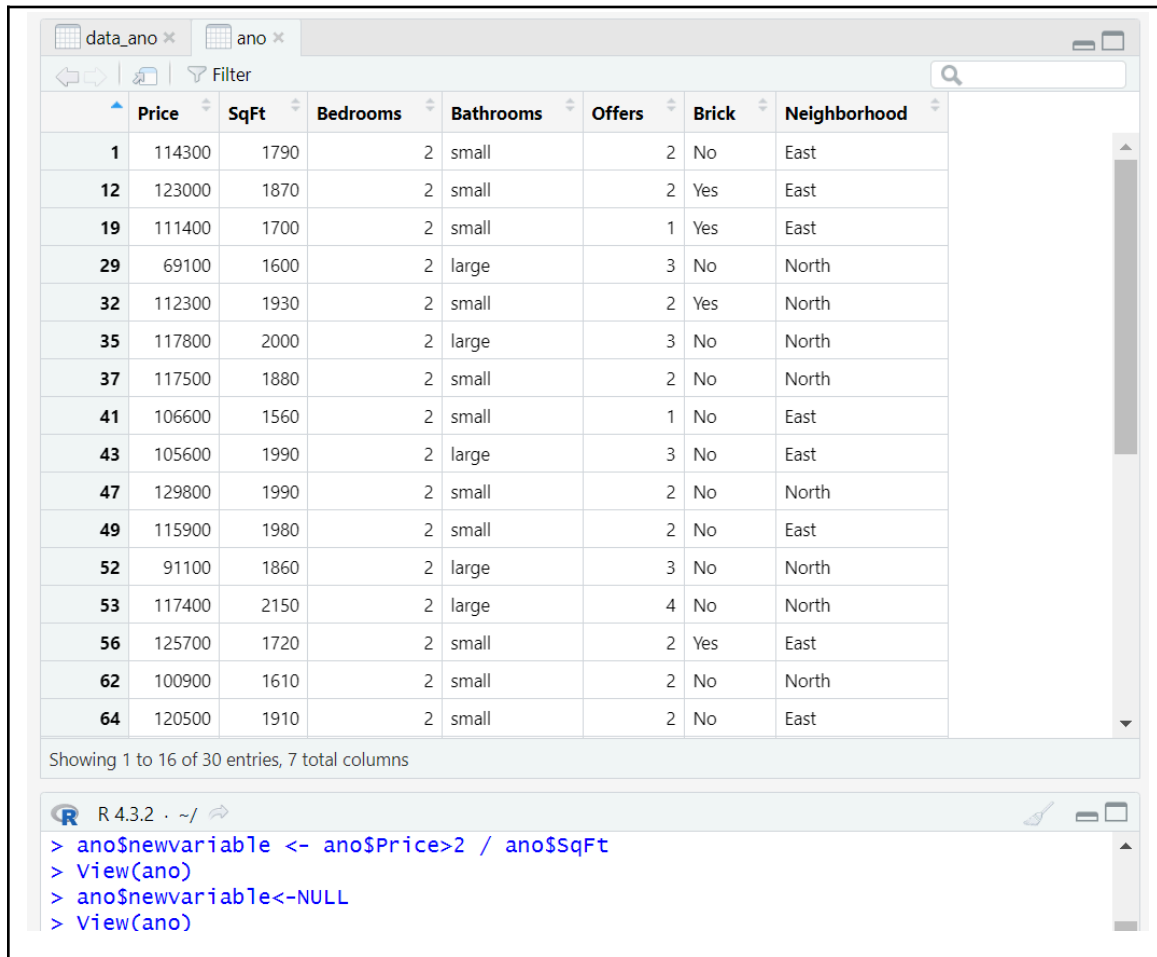
```
R 4.3.2 · ~/
Error in `<-data.frame`(`*tmp*`, newvariable, value = logical(0)) :
  replacement has 0 rows, data has 30
> ano$newvariable <- ano$Price>2 / ano$SqFt
> View(ano)
```



5. Delete Variabel. Selain bisa menambah, kita juga bisa menghapus variable. Dalam percobaan ini kita akan menghapus variable yang baru saja kita buat.

```
nama$newvariable<-NULL
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	small	2	No	East
12	123000	1870	2	small	2	Yes	East
19	111400	1700	2	small	1	Yes	East
29	69100	1600	2	large	3	No	North
32	112300	1930	2	small	2	Yes	North
35	117800	2000	2	large	3	No	North
37	117500	1880	2	small	2	No	North
41	106600	1560	2	small	1	No	East
43	105600	1990	2	large	3	No	East
47	129800	1990	2	small	2	No	North
49	115900	1980	2	small	2	No	East
52	91100	1860	2	large	3	No	North
53	117400	2150	2	large	4	No	North
56	125700	1720	2	small	2	Yes	East
62	100900	1610	2	small	2	No	North
64	120500	1910	2	small	2	No	East

Showing 1 to 16 of 30 entries, 7 total columns

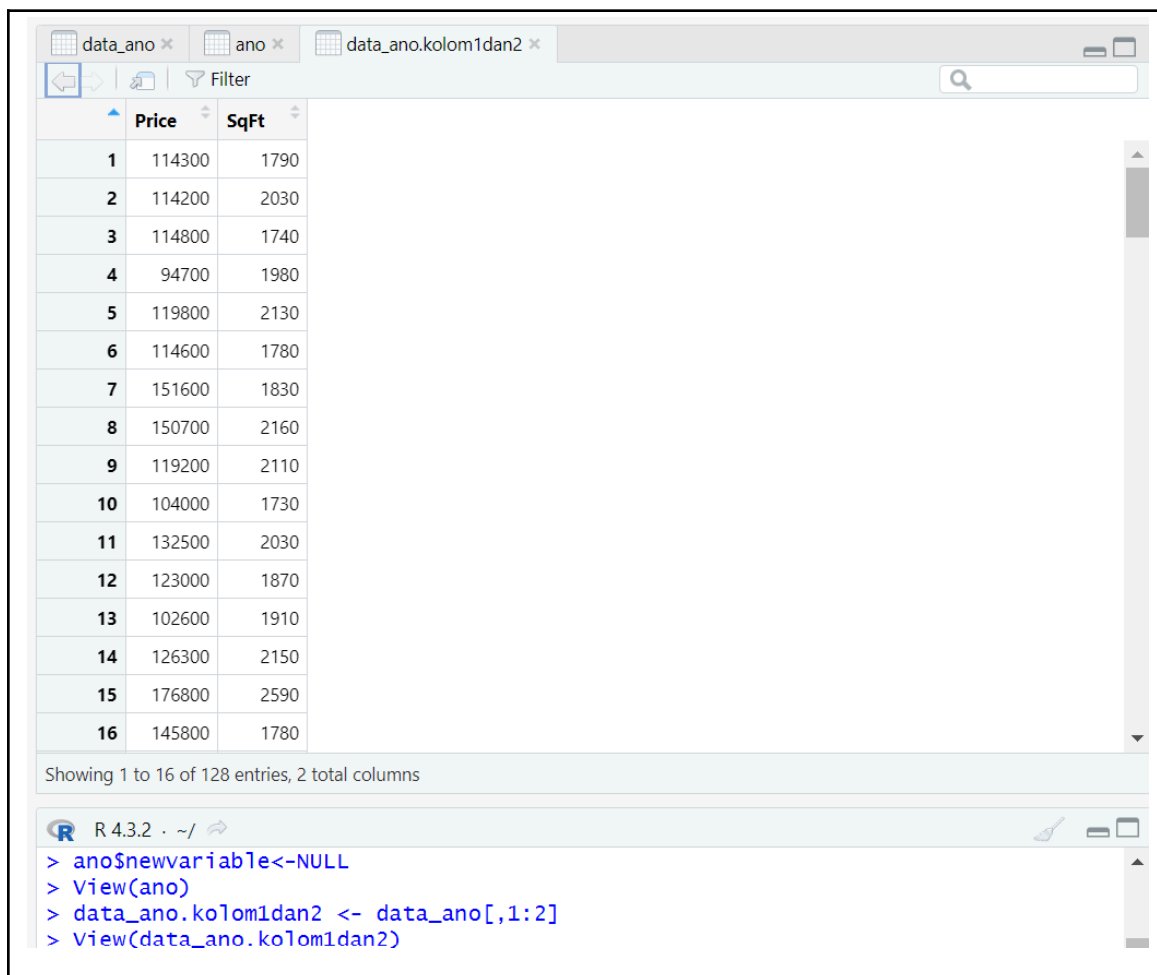
```
R 4.3.2 . ~/
> ano$newvariable <- ano$Price>2 / ano$SqFt
> View(ano)
> ano$newvariable<-NULL
> View(ano)
```



6. Merge Data Frame. Tahap 1 □ Merge kolom Artinya memisahkan data sesuai kolom yang diinginkan lalu kemudian digabungkan.

```
data_nama.kolom1dan2 <- data_nama[,1:2]  
View(data_nama.kolom1dan2)
```

Output:



	Price	SqFt
1	114300	1790
2	114200	2030
3	114800	1740
4	94700	1980
5	119800	2130
6	114600	1780
7	151600	1830
8	150700	2160
9	119200	2110
10	104000	1730
11	132500	2030
12	123000	1870
13	102600	1910
14	126300	2150
15	176800	2590
16	145800	1780

Showing 1 to 16 of 128 entries, 2 total columns

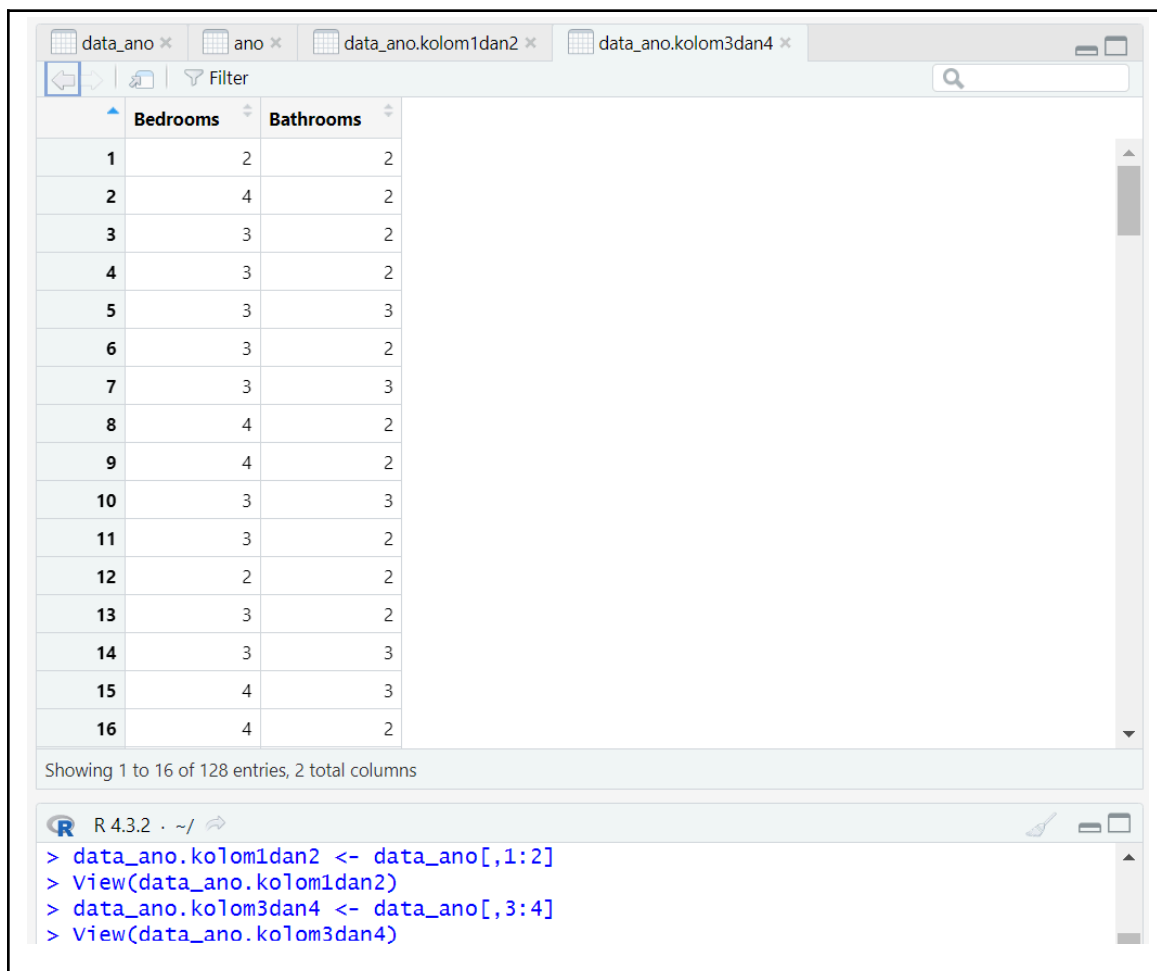
```
R 4.3.2 . ~/  
> ano$newvariable<-NULL  
> View(ano)  
> data_ano.kolom1dan2 <- data_ano[,1:2]  
> View(data_ano.kolom1dan2)
```



Tahap 2 □ Merge kolom Artinya memisahkan data sesuai kolom yang diinginkan lalu kemudian digabungkan.

```
data_elen.kolom3dan4 <- data_elen[,3:4]
View(data_elen.kolom3dan4
```

Output:



	Bedrooms	Bathrooms
1	2	2
2	4	2
3	3	2
4	3	2
5	3	3
6	3	2
7	3	3
8	4	2
9	4	2
10	3	3
11	3	2
12	2	2
13	3	2
14	3	3
15	4	3
16	4	2

Showing 1 to 16 of 128 entries, 2 total columns

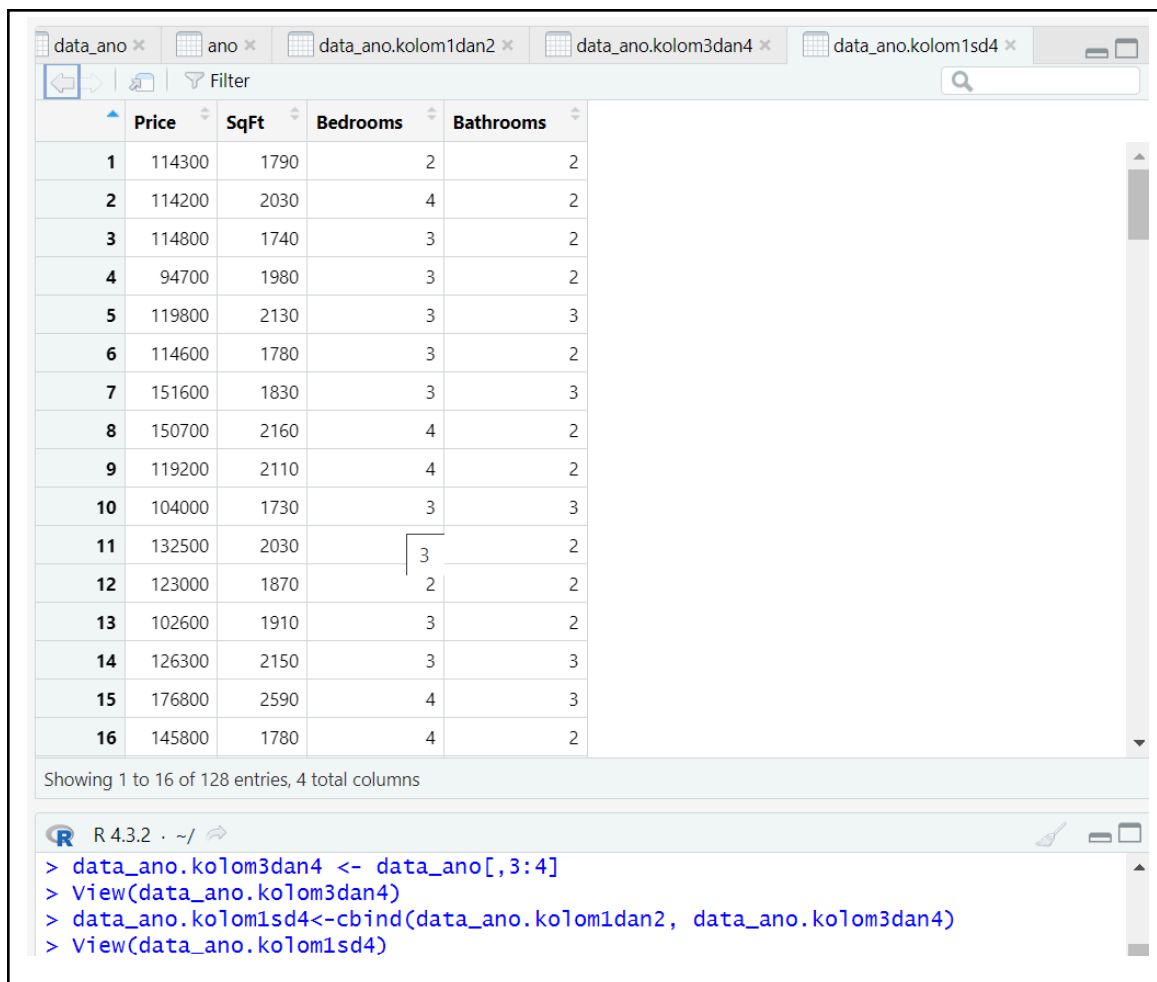
```
R 4.3.2 · ~/
> data_ano.kolom1dan2 <- data_ano[,1:2]
> View(data_ano.kolom1dan2)
> data_ano.kolom3dan4 <- data_ano[,3:4]
> View(data_ano.kolom3dan4)
```



Tahap 3 □ Merge kolom Artinya memisahkan data sesuai kolom yang diinginkan lalu kemudian digabungkan.

```
data_nama.kolom1sd4<-cbind(data_nama.kolom1dan2, data_nama.kolom3dan4)
View(data_nama.kolom1sd4)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms
1	114300	1790	2	2
2	114200	2030	4	2
3	114800	1740	3	2
4	94700	1980	3	2
5	119800	2130	3	3
6	114600	1780	3	2
7	151600	1830	3	3
8	150700	2160	4	2
9	119200	2110	4	2
10	104000	1730	3	3
11	132500	2030	3	2
12	123000	1870	2	2
13	102600	1910	3	2
14	126300	2150	3	3
15	176800	2590	4	3
16	145800	1780	4	2

Showing 1 to 16 of 128 entries, 4 total columns

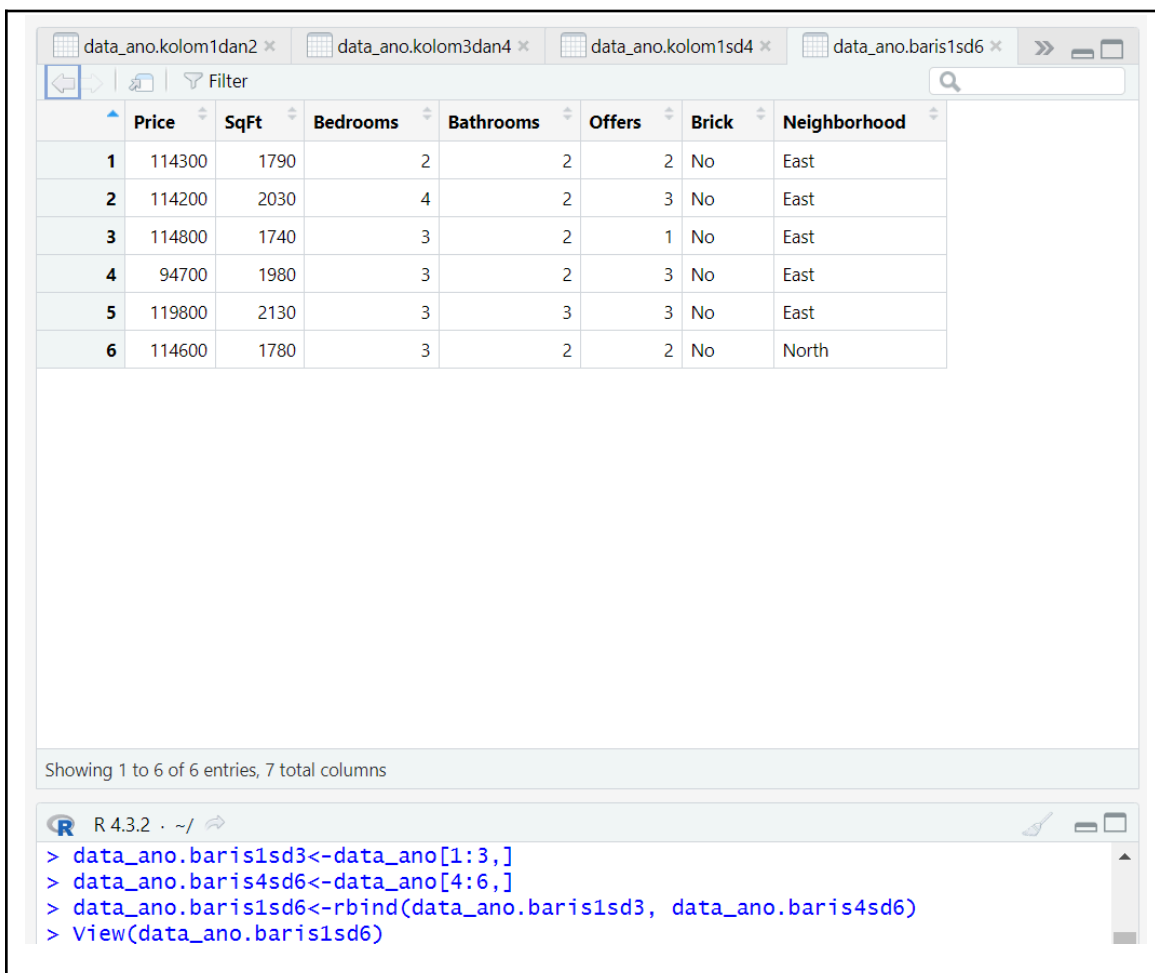
```
R 4.3.2 · ~/
> data_ano.kolom3dan4 <- data_ano[,3:4]
> View(data_ano.kolom3dan4)
> data_ano.kolom1sd4<-cbind(data_ano.kolom1dan2, data_ano.kolom3dan4)
> View(data_ano.kolom1sd4)
```



7. Merge Baris artinya memisahkan data sesuai baris yang diinginkan dengan menggunakan range baris. Lalu kemudian digabungkan.

```
data_nama.baris1sd3 <- data_nama[1:3,]
data_nama.baris4sd6 <- data_nama[4:6,]
data_nama.baris1sd6 <- rbind(data_nama.baris1sd3, data_nama.baris4sd6)
View(data_nama.baris1sd6)
```

Output:



	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
1	114300	1790	2	2	2	No	East
2	114200	2030	4	2	3	No	East
3	114800	1740	3	2	1	No	East
4	94700	1980	3	2	3	No	East
5	119800	2130	3	3	3	No	East
6	114600	1780	3	2	2	No	North

Showing 1 to 6 of 6 entries, 7 total columns

```
R 4.3.2 ~ /
> data_ano.baris1sd3<-data_ano[1:3,]
> data_ano.baris4sd6<-data_ano[4:6,]
> data_ano.baris1sd6<-rbind(data_ano.baris1sd3, data_ano.baris4sd6)
> View(data_ano.baris1sd6)
```

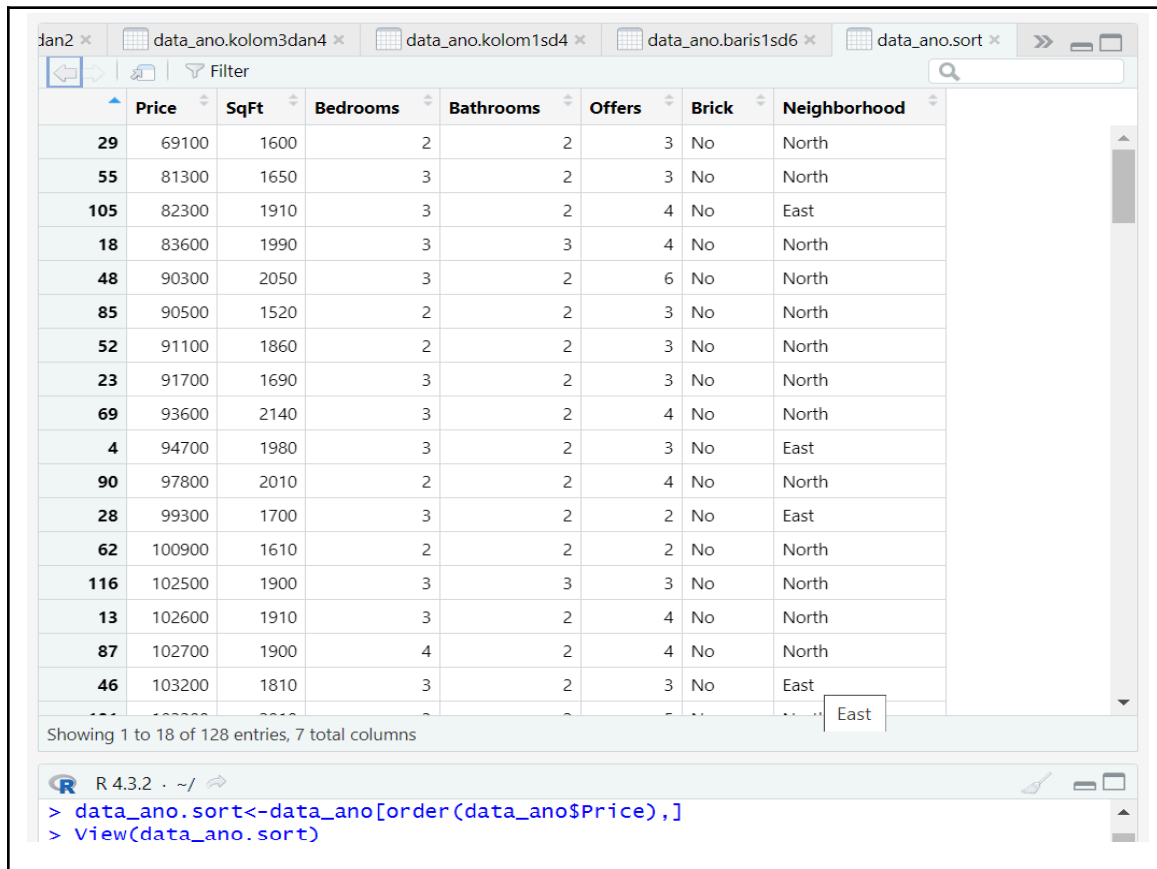


8. Sort data frame. Apa yang terjadi dengan data setelah di sort?

Jawaban: ?

```
data_nama.sort<-data_nama[order(data_nama$Price),]
View(data_nama.sort)
```

Output:



Showing 1 to 18 of 128 entries, 7 total columns

```
R 4.3.2 ~ /
> data_ano.sort<-data_ano[order(data_ano$Price),]
> View(data_ano.sort)
```

	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
29	69100	1600	2	2	3	No	North
55	81300	1650	3	2	3	No	North
105	82300	1910	3	2	4	No	East
18	83600	1990	3	3	4	No	North
48	90300	2050	3	2	6	No	North
85	90500	1520	2	2	3	No	North
52	91100	1860	2	2	3	No	North
23	91700	1690	3	2	3	No	North
69	93600	2140	3	2	4	No	North
4	94700	1980	3	2	3	No	East
90	97800	2010	2	2	4	No	North
28	99300	1700	3	2	2	No	East
62	100900	1610	2	2	2	No	North
116	102500	1900	3	3	3	No	North
13	102600	1910	3	2	4	No	North
87	102700	1900	4	2	4	No	North
46	103200	1810	3	2	3	No	East



b. Latihan Kedua – Tugas

Gunakan dataset pada tugas 1 yang telah ditambah lagi datanya dengan 10 mhs TIF/SI. Sehingga total baris data pada file tersebut berjumlah 30. Ulangi kembali perintah-perintah di atas dan sesuaikan dengan data anda. Dan lampirkan Screen Capture untuk tiap poin yang ada.

1. Read CSV

```
> data_ano = read.delim("clipboard")
> View(data_ano)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal
1	Putra	laki - laki	2023	170	150	Bekasi
2	Rafiq	laki - laki	2023	170	30	Kampung Tengah
3	Faiz Abyan Heryanto	laki - laki	2023	165	45	Tanjung Priuk,
4	Michael	laki - laki	2023	167	15	Jakarta Utara
5	partahi daniel tua limbong	laki - laki	2023	173	45	jakarta utara
6	hafidz ramadhan	laki - laki	2023	169	15	Grogol
7	Francisco	laki - laki	2023	170	30	daan mogort
8	Muhammad Rayyan Naufal	laki - laki	2023	171	50	Jakarta Timur
9	Yustianas Rombon	laki - laki	2023	173	30	Jakarta Utara
10	Zahwa Nur Azkia Putri	perempuan	2023	154	15	grogol
11	Dave F.M	laki - laki	2023	168	60	Tangerang
12	Calista Azzahra	perempuan	2023	160	150	Bekasi
13	Fairuz	perempuan	2023	156	45	Tanjung Priok
14	Gwen Alaina Marela	perempuan	2023	152	17	Gambir
15	Michael Briant	laki - laki	2023	167	10	Jl. Tawakal
16	ariel febrio	laki - laki	2023	170	35	kelapa gading
17	Rasyid Ridho Muhammad Noor Rafi	laki - laki	2023	170	45	Cengkareng



2. Subset Data □ Gunakan Kolom Tinggi Badan untuk bagian ini

```
> ano<-data_ano[data_ano$tinggi.badan>100,]  
> View(ano)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal
1	Putra	laki - laki	2023	170	150	Bekasi
2	Rafiq	laki - laki	2023	170	30	Kampung Tengah
3	Faiz Abyan Heryanto	laki - laki	2023	165	45	Tanjung Priuk,
4	Michael	laki - laki	2023	167	15	Jakarta Utara
5	partahi daniel tua limbong	laki - laki	2023	173	45	jakarta utara
6	hafidz ramadhan	laki - laki	2023	169	15	Grogol
7	Francisco	laki - laki	2023	170	30	daan mogort
8	Muhammad Rayyan Naufal	laki - laki	2023	171	50	Jakarta Timur
9	Yustianas Rombon	laki - laki	2023	173	30	Jakarta Utara
10	Zahwa Nur Azkia Putri	perempuan	2023	154	15	grogol
11	Dave F.M	laki - laki	2023	168	60	Tangerang
12	Calista Azzahra	perempuan	2023	160	150	Bekasi
13	Fairuz	perempuan	2023	156	45	Tanjung Priok
14	Gwen Alaina Marela	perempuan	2023	152	17	Gambir
15	Michael Briant	laki - laki	2023	167	10	Jl. Tawakal
16	ariel febrio	laki - laki	2023	170	35	kelapa gading



3. Mengubah nilai suatu variabel □ Ubah isi kolom tinggi badan untuk tinggi > 160 menjadi “Tinggi” dan jika bukan berarti “Pendek”

```
> ano$tinggi.badan<-ifelse(ano$tinggi.badan>160,c("tinggi"),c("pendek"))
> View(ano)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal
1	Putra	laki - laki	2023	tinggi	150	Bekasi
2	Rafiq	laki - laki	2023	tinggi	30	Kampung Tengah
3	Faiz Abyan Heryanto	laki - laki	2023	tinggi	45	Tanjung Priuk,
4	Michael	laki - laki	2023	tinggi	15	Jakarta Utara
5	partahi daniel tua limbong	laki - laki	2023	tinggi	45	jakarta utara
6	hafidz ramadhan	laki - laki	2023	tinggi	15	Grogol
7	Francisco	laki - laki	2023	tinggi	30	daan mogort
8	Muhammad Rayyan Naufal	laki - laki	2023	tinggi	50	Jakarta Timur
9	Yustianas Rombon	laki - laki	2023	tinggi	30	Jakarta Utara
10	Zahwa Nur Azkia Putri	perempuan	2023	pendek	15	grogol
11	Dave F.M	laki - laki	2023	tinggi	60	Tangerang
12	Calista Azzahra	perempuan	2023	pendek	150	Bekasi
13	Fairuz	perempuan	2023	pendek	45	Tanjung Priok
14	Gwen Alaina Marela	perempuan	2023	pendek	17	Gambir
15	Michael Briant	laki - laki	2023	tinggi	10	Jl. Tawakal
16	ariel febrio	laki - laki	2023	tinggi	35	kelapa gading



4. Menambah Variabel

- Buat variabel kolom baru dengan nama “Jurusan” dengan isi baris datanya adalah “Infor20”
- Buat variabel kolom baru kedua dengan nama “Fakultas” dengan isi baris datanya adalah “FTI”

```
> ano$Jurusan<-"Infor23"
> ano$fakultas<-"FTI"
> View(ano)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal	Jurusan	fakultas
1	Putra	laki - laki	2023	tinggi	150	Bekasi	Infor23	FTI
2	Rafiq	laki - laki	2023	tinggi	30	Kampung Tengah	Infor23	FTI
3	Faiz Abyan Heryanto	laki - laki	2023	tinggi	45	Tanjung Priuk,	Infor23	FTI
4	Michael	laki - laki	2023	tinggi	15	Jakarta Utara	Infor23	FTI
5	partahi daniel tua limbong	laki - laki	2023	tinggi	45	jakarta utara	Infor23	FTI
6	hafidz ramadhan	laki - laki	2023	tinggi	15	Grogol	Infor23	FTI
7	Francisco	laki - laki	2023	tinggi	30	daan mogort	Infor23	FTI
8	Muhammad Rayyan Naufal	laki - laki	2023	tinggi	50	Jakarta Timur	Infor23	FTI
9	Yustianas Rombon	laki - laki	2023	tinggi	30	Jakarta Utara	Infor23	FTI
10	Zahwa Nur Azkia Putri	perempuan	2023	pendek	15	grogol	Infor23	FTI
11	Dave F.M	laki - laki	2023	tinggi	60	Tangerang	Infor23	FTI
12	Calista Azzahra	perempuan	2023	pendek	150	Bekasi	Infor23	FTI
13	Fairuz	perempuan	2023	pendek	45	Tanjung Priok	Infor23	FTI
14	Gwen Alaina Marela	perempuan	2023	pendek	17	Gambir	Infor23	FTI
15	Michael Briant	laki - laki	2023	tinggi	10	Jl. Tawakal	Infor23	FTI



5. Delete Variabel □ Hapus kolom Fakultas

```
> ano$fakultas<-NULL
>
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal	Jurusan
1	Putra	laki - laki	2023	tinggi	150	Bekasi	Infor23
2	Rafiq	laki - laki	2023	tinggi	30	Kampung Tengah	Infor23
3	Faiz Abyan Heryanto	laki - laki	2023	tinggi	45	Tanjung Priuk,	Infor23
4	Michael	laki - laki	2023	tinggi	15	Jakarta Utara	Infor23
5	partahi daniel tua limbong	laki - laki	2023	tinggi	45	jakarta utara	Infor23
6	hafidz ramadhan	laki - laki	2023	tinggi	15	Grogol	Infor23
7	Francisco	laki - laki	2023	tinggi	30	daan mogort	Infor23
8	Muhammad Rayyan Naufal	laki - laki	2023	tinggi	50	Jakarta Timur	Infor23
9	Yustianas Rombon	laki - laki	2023	tinggi	30	Jakarta Utara	Infor23
10	Zahwa Nur Azkia Putri	perempuan	2023	pendek	15	grogol	Infor23
11	Dave F.M	laki - laki	2023	tinggi	60	Tangerang	Infor23
12	Calista Azzahra	perempuan	2023	pendek	150	Bekasi	Infor23
13	Fairuz	perempuan	2023	pendek	45	Tanjung Priok	Infor23
14	Gwen Alaina Marela	perempuan	2023	pendek	17	Gambir	Infor23
15	Michael Briant	laki - laki	2023	tinggi	10	Jl. Tawakal	Infor23
16	ariel febrio	laki - laki	2023	tinggi	35	kelapa gading	Infor23



6. Merge Data Frame

- Gabung kolom Nama dan Gender
- Gabung kolom Angkatan dan Tinggi Badan
- Gabungkan 2 variabel kolom gabungan diatas

```
> data_ano.kolom1dan2<-data_ano[,1:2]
> View(data_ano.kolom1dan2)

> data_ano.kolom3dan4<-data_ano[,3:4]
> View(data_ano.kolom3dan4)

> data_ano.kolom1sd4<-cbind(data_ano.kolom1dan2, data_ano.kolom3dan4)
> View(data_ano.kolom1sd4)
```

Output:

	nama.lengkap	gender
1	Putra	laki - laki
2	Rafiq	laki - laki
3	Faiz Abyan Heryanto	laki - laki
4	Michael	laki - laki
5	partahi daniel tua limbong	laki - laki
6	hafidz ramadhan	laki - laki
7	Francisco	laki - laki
8	Muhammad Rayyan Naufal	laki - laki
9	Yustianas Rombon	laki - laki
10	Zahwa Nur Azkia Putri	perempuan
11	Dave F.M	laki - laki
12	Calista Azzahra	perempuan
13	Fairuz	perempuan
14	Gwen Alaina Marela	perempuan
15	Michael Briant	laki - laki
16	ariel febrio	laki - laki
17	Rasyid Ridho Muhammad Noor Rafi	laki - laki
18	Monic	perempuan
19	Mara Sutan Arrafi Siregar	laki - laki
20	Zulfadli Daniswara	laki - laki

Showing 1 to 20 of 35 entries, 2 total columns



	angkatan	tinggi.badan
1	2023	170
2	2023	170
3	2023	165
4	2023	167
5	2023	173
6	2023	169
7	2023	170
8	2023	171
9	2023	173
10	2023	154
11	2023	168
12	2023	160
13	2023	156
14	2023	152
15	2023	167
16	2023	170
17	2023	170
18	2023	160
19	2023	170
20	2023	170

	nama.lengkap	gender	angkatan	tinggi.badan
1	Putra	laki - laki	2023	170
2	Rafiq	laki - laki	2023	170
3	Faiz Abyan Heryanto	laki - laki	2023	165
4	Michael	laki - laki	2023	167
5	partahi daniel tua limbong	laki - laki	2023	173
6	hafidz ramadhan	laki - laki	2023	169
7	Francisco	laki - laki	2023	170
8	Muhammad Rayyan Naufal	laki - laki	2023	171
9	Yustianas Rombon	laki - laki	2023	173
10	Zahwa Nur Azkia Putri	perempuan	2023	154
11	Dave F.M	laki - laki	2023	168
12	Calista Azzahra	perempuan	2023	160
13	Fairuz	perempuan	2023	156
14	Gwen Alaina Marela	perempuan	2023	152
15	Michael Briant	laki - laki	2023	167
16	ariel febrio	laki - laki	2023	170
17	Rasyid Ridho Muhammad Noor Rafi	laki - laki	2023	170
18	Monic	perempuan	2023	160
19	Mara Sutan Arrafi Siregar	laki - laki	2023	170
20	Zulfadli Daniswara	laki - laki	2023	170

7. Merge Baris

- Buat gabungan baris 1-5
- Buat gabungan baris 25-30
- Gabungkan 2 variabel diatas

```
> data_ano.baris1sd5<-data_ano[1:5,]
> data_ano.baris25sd30<-data_ano[25:30,]
> data_ano.baris1sd30<-rbind(data_ano.baris1sd5, data_ano.baris25sd30)
> View(data_ano.baris1sd30)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal
1	Putra	laki - laki	2023	170	150	Bekasi
2	Rafiq	laki - laki	2023	170	30	Kampung Tengah
3	Faiz Abyan Heryanto	laki - laki	2023	165	45	Tanjung Priuk,
4	Michael	laki - laki	2023	167	15	Jakarta Utara
5	partahi daniel tua limbong	laki - laki	2023	173	45	jakarta utara
25	Noval Susanto	laki - laki	2023	177	60	Depok
26	Fadhil Ibrahim Wira Dharna	laki - laki	2023	169	60	Cipadu
27	Kevin Rizky Pradana	laki - laki	2023	173	60	Duren Sawit
28	Dewanto Maulana Sukaro Putra	laki - laki	2023	172	40	Tanjung Priok
29	Arik Dias Putra	laki - laki	2023	177	60	Tangerang
30	Muhammad Hakim	laki - laki	2023	170	10	Tamansari



8. Sort Data Frame □ Lakukan sort berdasarkan waktu perjalanan

```
> data_ano.sort<-data_ano[order(data_ano$waktu.perjalanan...menit),]  
> View(data_ano.sort)
```

Output:

	nama.lengkap	gender	angkatan	tinggi.badan	waktu.perjalanan...menit	wilayah.tinggal
15	Michael Briant	laki - laki	2023	167	10	Jl. Tawakal
30	Muhammad Hakim	laki - laki	2023	170	10	Tamansari
33	Putri	perempuan	2023	155	10	Muwardi
34	April	perempuan	2023	155	10	Muwardi
35	Naya	perempuan	2023	165	10	Tawakal
31	Andrew Hendrian Nathanael	laki - laki	2023	172	13	tomang
4	Michael	laki - laki	2023	167	15	Jakarta Utara
6	hafidz ramadhan	laki - laki	2023	169	15	Grogol
10	Zahwa Nur Azkia Putri	perempuan	2023	154	15	grogol
24	rafael gala herlambang	laki - laki	2023	175	15	gunung sahari
14	Gwen Alaina Marela	perempuan	2023	152	17	Gambir
2	Rafiq	laki - laki	2023	170	30	Kampung Tengah
7	Francisco	laki - laki	2023	170	30	daan mogort
9	Yustianas Rombon	laki - laki	2023	173	30	Jakarta Utara
23	Faiz Firdaus Priyanto	laki - laki	2023	165	30	Jakarta
16	ariel fehrin	laki - laki	2023	170	35	kelana nading



4. File Praktikum

Github Repository:

5. Soal Latihan

Soal:

1. Apa saja kegunaan pengelolaan data pada data frame?
2. Dalam kasus apakah data perlu dihapus?

Jawaban:

1. Pengelolaan data Frame biasanya dimanfaatkan untuk penyaringan data, transformasi data jugasebagai anlisis pendataan.
2. Biasanya data perlu dihapus ketika data yang dimasukkan salah dan banyak nilainya yang missing value atau bermasalah sehingga data perlu dibersihkan dan dihapus.

6. Kesimpulan

- a. Dalam pengerjaan praktikum Statistika, cukup mudah
- b. Kita juga dapat mengetahui bahwasannya pengimplementasian pengelolaan data pada frame yang kita lakukan di aplikasi R studio sangat memudahkan kita untuk memproses apa saja yang kita ingin perintah untuk mengubah data sesuai yang kita inginkan secara efisien dengan ini kita paham apa yang modul ini tuju dan maksud untuk mengoperasikan data frame ke R studio.



7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Pertama	✓	
2.	Latihan Kedua	✓	

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	10 Menit	Menarik
2.	Latihan Kedua	15 Menit	Menarik

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang



Distribusi Multinomial, Distribusi Hipergeometrik, dan Distribusi Poisson

1. Teori Singkat

Distribusi Multinomial, Distribusi Hipergeometrik, dan Distribusi Poisson, merupakan konsep penting dalam teori probabilitas dan statistika. Ketiga distribusi tersebut sangat penting dalam menganalisis data dan memodelkan berbagai permasalahan statistika dalam bidang matematika.

Distribusi Multinomial

Distribusi Multinomial sendiri merupakan distribusi probabilitas yang memodelkan peluang hasil dari percobaan dengan lebih dari dua kategori yang mungkin. Distribusi ini juga generalisasi dari Distribusi Binomial, yang hanya memiliki dua kategori atau hasil yang mungkin.

Distribusi Hipergeometrik

Distribusi Hipergeometrik merupakan distribusi yang biasanya digunakan menghitung peluang suatu kasus dalam situasi di mana kita mengambil sampel dari populasi terbatas tanpa penggantian. Distribusi ini juga sering digunakan dalam statistika untuk memodelkan situasi, di mana kita ingin mengetahui peluang suatu peristiwa yang terjadi berdasarkan sampel populasi yang memiliki dua kategori atau jenis yang berbeda.

Distribusi Poisson

Distribusi Poisson merupakan distribusi yang digunakan untuk memodelkan jumlah kejadian yang terjadi dalam suatu interval waktu atau ruang tertentu, jika kejadian tersebut terjadi secara acak dengan laju rata-rata yang diketahui dan independen dari waktu atau ruang.

2. Alat dan Bahan

Hardware : Laptop/PC
Software : Jupyter Notebook, Visual Studio Code,dll.



3. Elemen Kompetensi

a. Distribusi Multinomial

Running code berikut ini di dalam IDE masing-masing

```
In [2]: import numpy as np
        from scipy.stats import multinomial

        n_trials_nama = 10
        probabilities_nama = [0.2, 0.3, 0.5]

        multinom_dist = multinomial(n_trials_nama, probabilities_nama)

        sample = multinom_dist.rvs()

        print("sample hasil percobaan:")
        print(sample)

        pmf_sample = multinom_dist.pmf(sample)
        print("\nProbabilitas nassa dari sampel yang dihasilkan:", pmf_sample)

        sample hasil percobaan:
        [[3 2 5]]

        Probabilitas nassa dari sampel yang dihasilkan: [0.0567]
```

Output

```
sample hasil percobaan:
[[3 2 5]]

Probabilitas nassa dari sampel yang dihasilkan: [0.0567]
```



Penerapan pada soal cerita

Sebuah kantin menawarkan tiga jenis makanan: nasi goreng, mie goreng, dan ayam goreng. Dalam sebuah survei terhadap 20 pelanggan, didapatkan bahwa 8 pelanggan memilih nasi goreng, 6 pelanggan memilih mie goreng, dan 6 pelanggan memilih ayam goreng. Berapa peluang bahwa dalam 20 pelanggan, akan ada 8 yang memilih nasi goreng, 6 yang memilih mie goreng, dan 6 memilih ayam goreng?

Penyelesaian :

```
In [3]: from scipy.stats import multinomial
        n_trials = 20
        probabilities = [8/20, 6/20, 6/20]
        multinom_dist = multinomial(n_trials, probabilities)
        outcome = [8, 6, 6]
        pmf_outcome = multinom_dist.pmf(outcome)
        print("probabilities massa:", pmf_outcome)
```

Output

```
probabilities massa: 0.04053909860481753
```



b. Distribusi Hipergeometrik

Running kode berikut di IDE masing-masing

```
In [4]: import numpy as np
        from scipy.stats import hypergeom

        N = 50
        K = 20
        n = 1
        k = 5

        hypergeom_dist = hypergeom(N, K, n)

        pmf_value = hypergeom_dist.pmf(k)

        print("probabilitas massa", pmf_value)
```

Output

```
probabilitas massa 0.0
```



Penerapan pada soal cerita

Sebuah kotak berisi 30 bola, di mana 10 bola berwarna merah, 10 bola berwarna biru, dan 10 bola berwarna hijau. Jika 8 bola diambil secara acak tanpa pengembalian, berapa peluangnya bahwa tepat 3 bola berwarna merah, 2 bola berwarna biru, dan 3 bola berwarna hijau?

Penyelesaian

```
from scipy.stats import hypergeom

N = 30
K = 10
n = 8
k1 = 3
k2 = 2
k3 = 3

hypergeom_dist = hypergeom(N, K, n)

pmf_value = hypergeom_dist.pmf([k1, k2, k3])

print("probabilitas massa", pmf_value)
```

Output

```
probabilitas massa [0.31787183 0.29800484 0.31787183]
```



c. Distribusi Poisson

Running kode berikut di IDE masing-masing

```
import numpy as np
from scipy.stats import poisson

mu = 2

poisson_dist = poisson(mu)

sample = poisson_dist.rvs(size=10)

print("sampel hasil distribusi poisson")
print(sample)

pmf_sample = poisson_dist.pmf(sample)
print("\nProbabilitas massa dari sampel yang dihasilkan:", pmf_sample)
```

Output

```
sampel hasil distribusi poisson
[7 1 2 2 3 3 0 2 0 1]

Probabilitas massa dari sampel yang dihasilkan: [0.00343709 0.27067057 0.27067057 0.27067057 0.18044704 0.18044704
0.13533528 0.27067057 0.13533528 0.27067057]
```

Penerapan pada soal cerita

Di sebuah kantor pos, rata-rata 4 paket kiriman pos tiba setiap jam. Berapakah peluang bahwa dalam jam berikutnya akan tiba tepat 6 paket?

Penyelesaian

```
from scipy.stats import poisson

mu = 4

prob_tepat_6 = poisson.pmf(6, mu)

print("peluang bahwa akan tiba tepat 6 paket dalam jam berikutnya:", prob_tepat_6)
```

Output

```
peluang bahwa akan tiba tepat 6 paket dalam jam berikutnya: 0.10419563456702102
```



4. File Praktikum

5. Soal Latihan

1. Tuliskan hasil dari output dari setiap percobaan
 - a. Distribusi Multinomial
 - b. Distribusi Hipergeometrik
 - c. Distribusi Poisson

Jawaban

Distribusi Multinomial

1. sample hasil perobaan : $[[3, 2, 5]]$ dan Probabilitas massa dari ampel yang dihasilkan $[0.0567]$
2. Probabilitas massa : 0.04053909860481753

Distribusi Hipergeometrik

1. Probabilitas massa 0.0
2. Probabilitas massa $[0.31787183 \ 0.29800484 \ 0.31787183]$

Distribusi Poisson

1. sampel hasil distribusi poisson $[7 \ 1 \ 2 \ 2 \ 3 \ 3 \ 0 \ 2 \ 0 \ 1]$
Probabilitas massa dari sampel yang dihasilkan: $[0.00343709 \ 0.27067057 \ 0.27067057 \ 0.27067057 \ 0.18044704 \ 0.18044704 \ 0.13533528 \ 0.27067057 \ 0.13533528 \ 0.27067057]$.
2. peluang bahwa akan tiba tepat 6 paket dalam jam berikutnya: 0.10419563456702102

6. Kesimpulan

- a. Dalam pengerjaan praktikum ini, sangat mudah
- b. Kita juga dapat mengetahui bahwasannya probabilitas memiliki beberapa macam distribusi dan yang kita pelajari di modul ini yaitu distribusi multinomial,hipergeometrik, dan poisson yang dimana ketiga metode tersebut memiliki teori eksekusi yang berbeda namun memiliki tujuan yang sama yaitu menghitung suatu probabilitas.



7. Cek List

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Distribusi Multinomial	✓	
2.	Latihan Distribusi Hipergeometrik	✓	
3.	Latihan Distribusi Poisson	✓	

8. Form Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Distribusi Multinomial	10 Menit	Menarik
2.	Latihan Distribusi Hipergeometrik	5 Menit	Menarik
3.	Latihan Distribusi Poisson	5 Menit	Menarik

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

