|  |  |  |
| --- | --- | --- |
| **Nama:**  Adriansyah Maulana Putra  **NIM:** 064002200046 | C:\Users\RPL-SI 02\Pictures\288px-Trisakti_Logo.svg.png | **MODUL 3**  **Nama Dosen:**  **Dedy Sugiarto** |
| **Hari/Tanggal:**  **Hari,** Senin 17 Juli 2023 | **Praktikum Statistika** | **Nama Asisten Labratorium:**   1. **Elen Fadilla Estri**   **064002000008**   1. **Rukhy Zaifa Aduhalim**   **064002000041** |

**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, klasterisasi, dan sebagainya) serta grafik. R, sebagaimana S, dirancang sebagai bahasa komputer sebenarnya, dan mengizinkan penggunanya 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.

1. **Alat dan Bahan**

Hardware : Laptop/PC

Software : R Studio

1. **Elemen Kompetensi**
   1. Latihan pertama – Merge Data

1. Jalankan souce 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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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

|  |
| --- |
| RStudio |

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

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

Output:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

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:

|  |
| --- |
| RStudio |

* 1. 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\_Adrian = read.delim("clipboard")  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

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

|  |
| --- |
| #subset data kolom tinggi  data\_Adrian <- data\_Adrian[data\_Adrian$Tinggi.Badan==157,]  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

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

|  |
| --- |
| #ubah nilai suatu variable untuk kolom tinggi>160  data\_Adrian$Tinggi.Badan <- ifelse(data\_Adrian$Tinggi.Badan > 160, c("tinggi"), c("pendek"))  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

4. Menambah Variabel

* Buat variabel kolom baru dengan nama “Jurusan” dengan isi baris datanya adalah “Infor22”
* Buat varibel kolom baru kedua dengan nama “Fakultas” dengan isi baris datanya adalah “FTI”

|  |
| --- |
| #menambah variabel  data\_Adrian$Jurusan <- ("INFOR22")  View(data\_Adrian)  data\_Adrian$Fakultas <- ("FTI")  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

5. Delete Variabel 🡪 Hapus kolom Fakultas

|  |
| --- |
| data\_Adrian$Fakultas <- NULL  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

6. Merge Data Frame

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

|  |
| --- |
| #merge data frame  data\_Adrian.kolom1dan2 <- data\_Adrian[,1:2]  View(data\_Adrian.kolom1dan2)  data\_Adrian.kolom3dan4 <- data\_Adrian[,3:4]  View(data\_Adrian.kolom3dan4)  data\_Adrian.kolom1sd4 <- cbind(data\_Adrian.kolom1dan2, data\_Adrian.kolom3dan4)  View(data\_Adrian.kolom1sd4) |

Output:

|  |
| --- |
|  |

7. Merge Baris

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

|  |
| --- |
| data\_Adrian.sort<-  data\_Adrian[order(data\_Adrian$Waktu.Perjalanan),]  View(data\_Adrian) |

Output:

|  |
| --- |
|  |

8. Sort Data Frame 🡪 Lakukan sort berdasarkan waktu perjalanan

|  |
| --- |
| data\_Adrian.baris1sd5 <- data\_Adrian.kolom1sd4[1:5,]  View(data\_Adrian.baris1sd5)  data\_Adrian.baris25sd30 <- data\_Adrian.kolom1sd4[25:30,]  View(data\_Adrian.baris25sd30)  data\_Adrian.baris1sd30<-rbind(data\_Adrian.baris1sd5, data\_Adrian.baris25sd30)  View(data\_Adrian.baris1sd30) |

Output:

|  |
| --- |
|  |

1. **File Praktikum**

Github Repository:

|  |
| --- |
|  |

1. **Soal Latihan**

Soal:

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

Jawaban:  
1. Pengelolaan data pada data frame memiliki beberapa kegunaan, antara lain:

-Pembersihan data: Data frame sering kali memuat data yang tidak lengkap, duplikat, atau tidak relevan. Pengelolaan data pada data frame memungkinkan Anda untuk membersihkan dan memfilter data agar lebih berkualitas.

-Transformasi data: Anda dapat menggunakan pengelolaan data pada data frame untuk melakukan transformasi data seperti mengubah format tanggal, melakukan pemotongan atau penggabungan kolom, atau mengubah tipe data.

-Analisis data: Pengelolaan data pada data frame memungkinkan Anda untuk mengambil, menghitung, dan menganalisis data dengan mudah. Anda dapat menggunakan fungsi dan metode yang tersedia pada data frame untuk menjalankan analisis statistik, agregasi data, atau perhitungan lainnya.

-Visualisasi data: Setelah data dikumpulkan dan diproses, pengelolaan data pada data frame memungkinkan Anda untuk menggambarkan data secara visual melalui grafik atau plot. Hal ini memudahkan pemahaman dan interpretasi data secara visual.

2.Ada beberapa kasus di mana data perlu dihapus, antara lain:

-Data yang tidak relevan: Jika data tidak relevan dengan tujuan analisis atau proyek yang sedang Anda kerjakan, maka data tersebut dapat dihapus untuk menghindari kebingungan atau kesalahan analisis.

-Data yang rusak atau tidak lengkap: Jika data mengandung kesalahan, kesenjangan, atau kecacatan yang signifikan, maka data tersebut mungkin perlu dihapus. Data yang rusak atau tidak lengkap dapat mengganggu analisis dan menghasilkan kesimpulan yang tidak akurat.

-Data yang duplikat: Duplikasi data dapat menyebabkan bias dalam analisis dan menghasilkan informasi yang tidak valid. Oleh karena itu, data duplikat perlu dihapus agar hasil analisis akurat dan tidak terdistorsi.

-Data yang melanggar privasi atau kebijakan: Dalam beberapa kasus, data perlu dihapus jika melanggar privasi individu atau kebijakan penggunaan data. Misalnya, jika data berisi informasi pribadi yang tidak sah atau sensitif, maka data tersebut harus dihapus sesuai dengan aturan privasi yang berlaku.

1. **Kesimpulan**
   1. Dalam pengerjaan praktikum Statistika,kali ini kita diperintah untuk mengelola data pada data frame dan beberapa kegunaannya
   2. Kita juga dapat mengetahui data frame juga memliki beberapa kegunaan,antara lain yaitu: pembersihan data,transformasi data,analisis data d

Data perlu dihapus dalam beberapa kasus, seperti data yang tidak relevan,data yang rusak atau tidak lengkap,data duplikat

1. **Cek List (✔)**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| **Selesai** | **Tidak Selesai** |
| **1.** | Latihan Pertama | **V** |  |
| **2.** | Latihan Kedua | **V** |  |

1. **Formulir Umpan Balik**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| **1.** | Latihan Pertama | 25 Menit | Menarik |
| **2.** | Latihan Kedua | 40 Menit | Menarik |

Keterangan:

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