data.table

R'm Balım Peteğim

31 08 2020

Contents

Introduction to data.table Package

`data.table` package is used for the data manupilation processes. It is accepted as the fastest package for this purpose. Even if R is consired as bad for high volume data, this package makes all works faster than any other package. When it is compared with *dply* or *pandas* package of Python, it is the winner all the time. Also, most of other packages in R are built on data.table package.

```
pti <- c("data.table")
pti <- pti[!(pti %in% installed.packages())]
if(length(pti)>0){
    install.packages(pti)
}

library(data.table)
```

Examples of data.table

fread function

Also we can use `fread()` for reading the data from internet or local. This function works like `read.csv()` function but in a more efficient way.

```
file <- if(file.exists("turkey_car_market.csv")) {
  "turkey_car_market.csv"
} else {
  url('https://raw.githubusercontent.com/pjournal/boun01g-data-mine-r-s/gh-pages/data.table/turkey_car_
}
car_market = fread(file)
```

data.table object

As other data manupilation packages, data.table has its own data type which is *data.table*. We can create an example data.table object like this:

```
datatable = data.table(student_no = c(1, 2, 3, 4, 5),
                       names = c("A", "B", "C", "D", "E"),
                       city = c("Istanbul", "Adana", "Eskişehir", "Antalya", "Hakkari"))
datatable
```

```
##    student_no names      city
## 1:          1     A  Istanbul
## 2:          2     B     Adana
## 3:          3     C Eskişehir
## 4:          4     D   Antalya
## 5:          5     E   Hakkari
```

We can use all functions in base R with data.table object like:

```
nrow(car_market)
```

```
## [1] 9044
```

```
ncol(car_market)
```

```
## [1] 15
```

```
summary(car_market)
```

```
##  Ä°lan Tarihi          Marka           Arac Tip Grubu       Arac Tip
##  Length:9044        Length:9044        Length:9044        Length:9044
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
```

```
##
##    Model YÄ±l   YakÄ±t Turu           Vites               CCM
##  Min.   :1959   Length:9044        Length:9044        Length:9044
##  1st Qu.:2010   Class :character   Class :character   Class :character
##  Median :2014   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2012
##  3rd Qu.:2017
##  Max.   :2020
##  Beygir Gucu            Renk            Kasa Tipi            Kimden
##  Length:9044        Length:9044        Length:9044        Length:9044
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##     Durum                 Km              Fiyat
##  Length:9044        Min.   :      0   Min.   :   5500
##  Class :character   1st Qu.:  68000   1st Qu.:  60938
##  Mode  :character   Median : 125000   Median :  95500
##                     Mean   : 135813   Mean   : 148070
##                     3rd Qu.: 194130   3rd Qu.: 153562
##                     Max.   :1850000   Max.   :5086500
```

```
str(car_market)
```

```
## Classes 'data.table' and 'data.frame':   9044 obs. of  15 variables:
##  $ Ä°lan Tarihi  : chr  "27/05/2020" "16/06/2020" "14/06/2020" "11/06/2020" ...
##  $ Marka         : chr  "Jaguar" "Acura" "Acura" "Acura" ...
##  $ Arac Tip Grubu: chr  "XF" "CL" "CL" "CL" ...
##  $ Arac Tip      : chr  "2.0 D Prestige Plus" "-" "2.2" "-" ...
##  $ Model YÄ±l    : num  2017 2015 1994 2013 2010 ...
##  $ YakÄ±t Turu   : chr  "Dizel" "Dizel" "Benzin/LPG" "Dizel" ...
##  $ Vites         : chr  "Otomatik Vites" "YarÄ± Otomatik Vites" "DÃ¼z Vites" "DÃ¼z Vites" ...
##  $ CCM           : chr  "1801-2000 cc" "1301-1600 cc" "1301-1600 cc" "1301-1600 cc" ...
##  $ Beygir Gucu   : chr  "176-200 BG" "101-125 BG" "101-125 BG" "76-100 BG" ...
##  $ Renk          : chr  "Lacivert" "Mavi" "Turkuaz" "Kahverengi" ...
##  $ Kasa Tipi     : chr  "Hatchback 5 KapÄ±" "Sedan" "Sedan" "Sedan" ...
##  $ Kimden        : chr  "Galeriden" "Sahibinden" "Sahibinden" "Sahibinden" ...
##  $ Durum         : chr  "2. El" "2. El" "2. El" "2. El" ...
##  $ Km            : int  26100 127000 175000 325 207000 320000 183 79000 15662 31971 ...
##  $ Fiyat         : int  634500 151500 19750 52000 148750 42500 170000 82500 189000 189000 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
head(car_market)
```

```
##    Ä°lan Tarihi  Marka Arac Tip Grubu            Arac Tip Model YÄ±l
## 1:   27/05/2020 Jaguar             XF 2.0 D Prestige Plus       2017
## 2:   16/06/2020  Acura             CL                   -       2015
## 3:   14/06/2020  Acura             CL                 2.2       1994
## 4:   11/06/2020  Acura             CL                   -       2013
## 5:   11/06/2020  Acura             CL                 2.2       2010
## 6:   04/06/2020  Acura             CL                 2.2       1999
##    YakÄ±t Turu               Vites          CCM Beygir Gucu       Renk
```

```
## 1:       Dizel       Otomatik Vites 1801-2000 cc  176-200 BG   Lacivert
## 2:       Dizel YarÄ± Otomatik Vites 1301-1600 cc  101-125 BG       Mavi
## 3:  Benzin/LPG           DÃ¼z Vites 1301-1600 cc  101-125 BG    Turkuaz
## 4:       Dizel           DÃ¼z Vites 1301-1600 cc   76-100 BG Kahverengi
## 5:       Dizel       Otomatik Vites 1801-2000 cc  151-175 BG      Beyaz
## 6:       Dizel           DÃ¼z Vites 1801-2000 cc  101-125 BG   Lacivert
##            Kasa Tipi     Kimden Durum     Km  Fiyat
## 1: Hatchback 5 KapÄ±  Galeriden 2. El  26100 634500
## 2:             Sedan Sahibinden 2. El 127000 151500
## 3:             Sedan Sahibinden 2. El 175000  19750
## 4:             Sedan Sahibinden 2. El    325  52000
## 5:             Sedan Sahibinden 2. El 207000 148750
## 6:          Roadster  Galeriden 2. El 320000  42500
```

DT[i, j, by]

The most important consept of data.table package is the usage of data.table object. When we want to filter the data, we need to give that information in `i` parameter. If we want to select some of columns (or all of them), we need to use the `j` parameter. If we want to group the data with a column / columns, we need to indicate it in the `by` parameter.

Filtering

For example, if we want to get all records whose prices(`Fiyat` column) are less than 50000 Turkish Liras, we can use this code:

```
car_market[Fiyat < 50000]
```

```
##       Ä°lan Tarihi  Marka Arac Tip Grubu       Arac Tip Model YÄ±l YakÄ±t Turu
##    1:   14/06/2020  Acura             CL            2.2       1994  Benzin/LPG
##    2:   04/06/2020  Acura             CL            2.2       1999       Dizel
##    3:   24/05/2020  Acura             CL              -       2001  Benzin/LPG
##    4:   27/04/2020  Acura             CL              -       2004       Dizel
##    5:   27/04/2020  Acura             CL              -       2004  Benzin/LPG
##   ---
## 1514:   06/06/2020 TofaÅŸ         Å\236ahin            1.6       1994  Benzin/LPG
## 1515:   05/06/2020 TofaÅŸ         Å\236ahin Å\236ahin 5 vites       1992  Benzin/LPG
## 1516:   05/06/2020 TofaÅŸ         DoÄŸan            SLX       1993  Benzin/LPG
## 1517:   04/06/2020 TofaÅŸ         Kartal Kartal 5 Vites       1991  Benzin/LPG
## 1518:   04/06/2020 TofaÅŸ         Kartal Kartal 5 Vites       1996  Benzin/LPG
##            Vites          CCM Beygir Gucu       Renk         Kasa Tipi
##    1: DÃ¼z Vites 1301-1600 cc  101-125 BG    Turkuaz             Sedan
##    2: DÃ¼z Vites 1801-2000 cc  101-125 BG   Lacivert          Roadster
##    3: DÃ¼z Vites 1301-1600 cc  101-125 BG      Siyah Hatchback 5 KapÄ±
##    4: DÃ¼z Vites 1301-1600 cc   76-100 BG      Beyaz Hatchback 5 KapÄ±
##    5: DÃ¼z Vites 1301-1600 cc  101-125 BG   Lacivert             Sedan
##   ---
## 1514: DÃ¼z Vites 1301-1600 cc    51-75 BG      Beyaz             Sedan
## 1515: DÃ¼z Vites 1301-1600 cc   76-100 BG        Bej             Sedan
## 1516: DÃ¼z Vites 1301-1600 cc   76-100 BG     YeÅŸil             Sedan
## 1517: DÃ¼z Vites 1601-1800 cc   76-100 BG KÄ±rmÄ±zÄ± Hatchback 5 KapÄ±
## 1518: DÃ¼z Vites 1301-1600 cc    51-75 BG      Beyaz             Sedan
```

```
##           Kimden Durum     Km Fiyat
##    1: Sahibinden 2. El 175000 19750
##    2:  Galeriden 2. El 320000 42500
##    3: Sahibinden 2. El 252000 36750
##    4:  Galeriden 2. El 230000 32000
##    5: Sahibinden 2. El  18000 34500
##   ---
## 1514:  Galeriden 2. El 121212 12250
## 1515: Sahibinden 2. El  54000 16300
## 1516: Sahibinden 2. El 135000 19000
## 1517:  Galeriden 2. El 250000 11000
## 1518:  Galeriden 2. El     11 15750
```

```
#car_market[car_market$Fiyat < 50000,] in base R
```

Also we can give more conditions with & or | operators.

```
car_market[Fiyat < 50000 & Marka == 'Acura']
```

```
##    Ä°lan Tarihi Marka Arac Tip Grubu Arac Tip Model YÄ±l YakÄ±t Turu      Vites
## 1:   14/06/2020 Acura             CL      2.2       1994  Benzin/LPG DÃ¼z Vites
## 2:   04/06/2020 Acura             CL      2.2       1999       Dizel DÃ¼z Vites
## 3:   24/05/2020 Acura             CL        -       2001  Benzin/LPG DÃ¼z Vites
## 4:   27/04/2020 Acura             CL        -       2004       Dizel DÃ¼z Vites
## 5:   27/04/2020 Acura             CL        -       2004  Benzin/LPG DÃ¼z Vites
## 6:   15/04/2020 Acura             CL        -       2009       Dizel DÃ¼z Vites
## 7:   15/04/2020 Acura             CL        -       2009       Dizel DÃ¼z Vites
##                 CCM Beygir Gucu     Renk          Kasa Tipi     Kimden Durum
## 1:     1301-1600 cc  101-125 BG  Turkuaz              Sedan Sahibinden 2. El
## 2:     1801-2000 cc  101-125 BG Lacivert           Roadster  Galeriden 2. El
## 3:     1301-1600 cc  101-125 BG    Siyah Hatchback 5 KapÄ± Sahibinden 2. El
## 4:     1301-1600 cc   76-100 BG    Beyaz Hatchback 5 KapÄ±  Galeriden 2. El
## 5:     1301-1600 cc  101-125 BG Lacivert              Sedan Sahibinden 2. El
## 6: 1300 cc ve altÄ±   76-100 BG    Bordo              Sedan Sahibinden 2. El
## 7: 1300 cc ve altÄ±   76-100 BG    Bordo              Sedan Sahibinden 2. El
##        Km Fiyat
## 1: 175000 19750
## 2: 320000 42500
## 3: 252000 36750
## 4: 230000 32000
## 5:  18000 34500
## 6: 150000 34000
## 7: 150000 34000
```

```
#car_market[car_market$Fiyat < 50000 & car_market$Marka == 'Acura',] in base R
```

If we want to get all rows except this condition we can use ! sign.

```
car_market[!(Fiyat < 50000 & Marka == 'Acura')]
```

```
##       Ä°lan Tarihi  Marka Arac Tip Grubu            Arac Tip Model YÄ±l
```

```
##    1:   27/05/2020 Jaguar             XF 2.0 D Prestige Plus       2017
##    2:   16/06/2020  Acura             CL                     -       2015
##    3:   11/06/2020  Acura             CL                     -       2013
##    4:   11/06/2020  Acura             CL                   2.2       2010
##    5:   03/06/2020  Acura             CL                   2.2       2014
##   ---
## 9033:   06/06/2020 TofaÅŸ         Å\236ahin                 1.6       1994
## 9034:   05/06/2020 TofaÅŸ         Å\236ahin      Å\236ahin 5 vites       1992
## 9035:   05/06/2020 TofaÅŸ         DoÄŸan                  SLX       1993
## 9036:   04/06/2020 TofaÅŸ         Kartal      Kartal 5 Vites       1991
## 9037:   04/06/2020 TofaÅŸ         Kartal      Kartal 5 Vites       1996
##       YakÄ±t Turu               Vites          CCM Beygir Gucu       Renk
##    1:       Dizel       Otomatik Vites 1801-2000 cc  176-200 BG   Lacivert
##    2:       Dizel YarÄ± Otomatik Vites 1301-1600 cc  101-125 BG       Mavi
##    3:       Dizel           DÃ¼z Vites 1301-1600 cc   76-100 BG Kahverengi
##    4:       Dizel       Otomatik Vites 1801-2000 cc  151-175 BG      Beyaz
##    5:       Dizel       Otomatik Vites 1301-1600 cc  101-125 BG KÄ±rmÄ±zÄ±
##   ---
## 9033:  Benzin/LPG           DÃ¼z Vites 1301-1600 cc    51-75 BG      Beyaz
## 9034:  Benzin/LPG           DÃ¼z Vites 1301-1600 cc   76-100 BG        Bej
## 9035:  Benzin/LPG           DÃ¼z Vites 1301-1600 cc   76-100 BG     YeÅŸil
## 9036:  Benzin/LPG           DÃ¼z Vites 1601-1800 cc   76-100 BG KÄ±rmÄ±zÄ±
## 9037:  Benzin/LPG           DÃ¼z Vites 1301-1600 cc    51-75 BG      Beyaz
##               Kasa Tipi     Kimden Durum     Km  Fiyat
##    1: Hatchback 5 KapÄ±  Galeriden 2. El  26100 634500
##    2:             Sedan Sahibinden 2. El 127000 151500
##    3:             Sedan Sahibinden 2. El    325  52000
##    4:             Sedan Sahibinden 2. El 207000 148750
##    5:             Sedan  Galeriden 2. El    183 170000
##   ---
## 9033:             Sedan  Galeriden 2. El 121212  12250
## 9034:             Sedan Sahibinden 2. El  54000  16300
## 9035:             Sedan Sahibinden 2. El 135000  19000
## 9036: Hatchback 5 KapÄ±  Galeriden 2. El 250000  11000
## 9037:             Sedan  Galeriden 2. El     11  15750
```

Selecting

For example, we need to get the head of the colours (`Renk` column). To do so, we can use these notations:

```
cols = c("Renk")
head(car_market[, Renk])
```

```
## [1] "Lacivert"   "Mavi"       "Turkuaz"    "Kahverengi" "Beyaz"
## [6] "Lacivert"
```

```
head(car_market[, .(Renk)])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
```

```
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

```
head(car_market[, list(Renk)])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

```
head(car_market[, c("Renk"), with = FALSE])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

```
head(car_market[, 10, with = FALSE])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

```
head(car_market[, cols, with = FALSE])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

```
head(car_market[, ..cols])
```

```
##          Renk
## 1:   Lacivert
## 2:       Mavi
## 3:    Turkuaz
## 4: Kahverengi
## 5:      Beyaz
## 6:   Lacivert
```

We used the `with` argument. If we assign *FALSE* to that argument, it uses the character as the column name. `.()` and `list()` is used for selecting the columns. As you can see, the first one returns a vector, not a data.table object. With that usage, we can not use data.table operations like:

```
nrow(head(car_market[, Renk]))
```

```
## NULL
```

We can select multiple columns with these usages:

```
cols = c("Renk", "Fiyat")
head(car_market[, .(Renk, Fiyat)])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

```
head(car_market[, list(Renk, Fiyat)])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

```
head(car_market[, c("Renk", "Fiyat"), with = FALSE])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

```
head(car_market[, c(10, 15), with = FALSE])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

```
head(car_market[, cols, with = FALSE])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

```
head(car_market[, ..cols])
```

```
##          Renk  Fiyat
## 1:   Lacivert 634500
## 2:       Mavi 151500
## 3:    Turkuaz  19750
## 4: Kahverengi  52000
## 5:      Beyaz 148750
## 6:   Lacivert  42500
```

We can select multiple rows with using slicing or excluding the columns with ! sign. For example:

```
head(car_market[, Marka:Fiyat])
```

```
##     Marka Arac Tip Grubu            Arac Tip Model YÄ±l YakÄ±t Turu
## 1: Jaguar             XF 2.0 D Prestige Plus       2017       Dizel
## 2:  Acura             CL                   -       2015       Dizel
## 3:  Acura             CL                 2.2       1994  Benzin/LPG
## 4:  Acura             CL                   -       2013       Dizel
## 5:  Acura             CL                 2.2       2010       Dizel
## 6:  Acura             CL                 2.2       1999       Dizel
##                    Vites          CCM Beygir Gucu       Renk         Kasa Tipi
## 1:       Otomatik Vites 1801-2000 cc  176-200 BG   Lacivert Hatchback 5 KapÄ±
## 2: YarÄ± Otomatik Vites 1301-1600 cc  101-125 BG       Mavi             Sedan
## 3:           DÃ¼z Vites 1301-1600 cc  101-125 BG    Turkuaz             Sedan
## 4:           DÃ¼z Vites 1301-1600 cc   76-100 BG Kahverengi             Sedan
## 5:       Otomatik Vites 1801-2000 cc  151-175 BG      Beyaz             Sedan
## 6:           DÃ¼z Vites 1801-2000 cc  101-125 BG   Lacivert          Roadster
##        Kimden Durum     Km  Fiyat
## 1:  Galeriden 2. El  26100 634500
## 2: Sahibinden 2. El 127000 151500
## 3: Sahibinden 2. El 175000  19750
## 4: Sahibinden 2. El    325  52000
## 5: Sahibinden 2. El 207000 148750
## 6:  Galeriden 2. El 320000  42500
```

```
head(car_market[, !c("Ä°lan Tarihi"), with = FALSE])
```

```
##     Marka Arac Tip Grubu            Arac Tip Model YÄ±l YakÄ±t Turu
## 1: Jaguar             XF 2.0 D Prestige Plus       2017       Dizel
## 2:  Acura             CL                   -       2015       Dizel
```

```
## 3:  Acura             CL                2.2       1994  Benzin/LPG
## 4:  Acura             CL                  -       2013       Dizel
## 5:  Acura             CL                2.2       2010       Dizel
## 6:  Acura             CL                2.2       1999       Dizel
##                   Vites          CCM Beygir Gucu       Renk         Kasa Tipi
## 1:       Otomatik Vites 1801-2000 cc  176-200 BG   Lacivert Hatchback 5 KapÄ±
## 2: YarÄ± Otomatik Vites 1301-1600 cc  101-125 BG       Mavi             Sedan
## 3:           DÃ¼z Vites 1301-1600 cc  101-125 BG    Turkuaz             Sedan
## 4:           DÃ¼z Vites 1301-1600 cc   76-100 BG Kahverengi             Sedan
## 5:       Otomatik Vites 1801-2000 cc  151-175 BG      Beyaz             Sedan
## 6:           DÃ¼z Vites 1801-2000 cc  101-125 BG   Lacivert          Roadster
##        Kimden Durum     Km  Fiyat
## 1:  Galeriden 2. El  26100 634500
## 2: Sahibinden 2. El 127000 151500
## 3: Sahibinden 2. El 175000  19750
## 4: Sahibinden 2. El    325  52000
## 5: Sahibinden 2. El 207000 148750
## 6:  Galeriden 2. El 320000  42500
```

In selecting the columns we can use some operations like `%like%`. The usage of this operator is like this:

```
car_market[, names(car_market) %like% 'Arac', with = FALSE]
```

```
##       Arac Tip Grubu            Arac Tip
##    1:             XF 2.0 D Prestige Plus
##    2:             CL                   -
##    3:             CL                 2.2
##    4:             CL                   -
##    5:             CL                 2.2
##   ---
## 9040:         Å\236ahin                 1.6
## 9041:         Å\236ahin      Å\236ahin 5 vites
## 9042:         DoÄŸan                 SLX
## 9043:         Kartal      Kartal 5 Vites
## 9044:         Kartal      Kartal 5 Vites
```

We can rename the columns with `setnames` function like:

```
setnames(car_market, c("Ä°lan Tarihi"), c("Date"))
setnames(car_market, c("Model YÄ±l", "YakÄ±t Turu"), c("Model_Year", "Fuel_Type"))
head(car_market)
```

```
##          Date  Marka Arac Tip Grubu            Arac Tip Model_Year  Fuel_Type
## 1: 27/05/2020 Jaguar             XF 2.0 D Prestige Plus       2017      Dizel
## 2: 16/06/2020  Acura             CL                   -       2015      Dizel
## 3: 14/06/2020  Acura             CL                 2.2       1994 Benzin/LPG
## 4: 11/06/2020  Acura             CL                   -       2013      Dizel
## 5: 11/06/2020  Acura             CL                 2.2       2010      Dizel
## 6: 04/06/2020  Acura             CL                 2.2       1999      Dizel
##                   Vites          CCM Beygir Gucu       Renk         Kasa Tipi
## 1:       Otomatik Vites 1801-2000 cc  176-200 BG   Lacivert Hatchback 5 KapÄ±
## 2: YarÄ± Otomatik Vites 1301-1600 cc  101-125 BG       Mavi             Sedan
```

```
## 3:           DÃ¼z Vites 1301-1600 cc  101-125 BG    Turkuaz             Sedan
## 4:           DÃ¼z Vites 1301-1600 cc   76-100 BG Kahverengi             Sedan
## 5:       Otomatik Vites 1801-2000 cc  151-175 BG      Beyaz             Sedan
## 6:           DÃ¼z Vites 1801-2000 cc  101-125 BG   Lacivert          Roadster
##        Kimden Durum     Km  Fiyat
## 1:  Galeriden 2. El  26100 634500
## 2: Sahibinden 2. El 127000 151500
## 3: Sahibinden 2. El 175000  19750
## 4: Sahibinden 2. El    325  52000
## 5: Sahibinden 2. El 207000 148750
## 6:  Galeriden 2. El 320000  42500
```

We can sort the data with `setorder` function like:

```
head(setorder(car_market, Date))
```

```
##          Date     Marka Arac Tip Grubu                   Arac Tip Model_Year
## 1: 01/04/2020     Acura             CL                     DiÄŸer       2004
## 2: 01/04/2020      Audi             A4                   2.0 TDI       2014
## 3: 01/04/2020      Audi       S Serisi        S5 3.0 TFSI Quattro       2010
## 4: 01/04/2020       BMW       1 Serisi                     1.18i       2008
## 5: 01/04/2020       BMW       3 Serisi 3.20i ED EfficientDynamics       2014
## 6: 01/04/2020 Chevrolet        Captiva                   2.0 D LT       2007
##     Fuel_Type                Vites          CCM     Beygir Gucu  Renk
## 1: Benzin/LPG           DÃ¼z Vites 1301-1600 cc 100 BG ve altÄ±   Gri
## 2:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Beyaz
## 3:     Benzin YarÄ± Otomatik Vites 2501-3000 cc      326-350 BG Siyah
## 4: Benzin/LPG       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
## 5:     Benzin       Otomatik Vites 1301-1600 cc      176-200 BG Siyah
## 6:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
##            Kasa Tipi     Kimden Durum     Km  Fiyat
## 1:             Sedan Sahibinden 2. El 170000  55000
## 2:             Sedan  Galeriden 2. El 160000 176500
## 3:      Spor / Coupe  Galeriden 2. El 160000 350000
## 4: Hatchback 5 KapÄ± Sahibinden 2. El 162000  82000
## 5:             Sedan  Galeriden 2. El 160000 192500
## 6:      Arazi AracÄ± Sahibinden 2. El 192500  80000
```

It orders the data ascending by default. We can use - sign for descending and can order the data with multiple columns like:

```
head(setorder(car_market, Date, -Marka))
```

```
##          Date   Marka Arac Tip Grubu                    Arac Tip Model_Year
## 1: 01/04/2020   Skoda          Rapid                1.2 Ambition       2015
## 2: 01/04/2020   Skoda        Octavia             1.6 TDI Optimal       2016
## 3: 01/04/2020    Seat          Ibiza                       1.6\n       1999
## 4: 01/04/2020    Seat          Ibiza 1.4 TSI Sport Coupe Cupra\n       2013
## 5: 01/04/2020 Peugeot           3008          1.6 BlueHDi Allure       2015
## 6: 01/04/2020    Opel          Astra                      1.6 16V       2004
##     Fuel_Type                Vites              CCM     Beygir Gucu       Renk
## 1: Benzin/LPG           DÃ¼z Vites 1300 cc ve altÄ± 100 BG ve altÄ±      Beyaz
```

```
## 2:      Dizel YarÄ± Otomatik Vites     1601-1800 cc      101-125 BG      Beyaz
## 3: Benzin/LPG       Otomatik Vites       Bilmiyorum      Bilmiyorum      SarÄ±
## 4:     Benzin       Otomatik Vites     1301-1600 cc      176-200 BG KÄ±rmÄ±zÄ±
## 5:      Dizel       Otomatik Vites     1301-1600 cc      101-125 BG      Beyaz
## 6: Benzin/LPG           DÃ¼z Vites     1301-1600 cc       76-100 BG       Mavi
##            Kasa Tipi     Kimden Durum     Km  Fiyat
## 1:             Sedan Sahibinden 2. El  98000  92000
## 2: Hatchback 5 KapÄ± Sahibinden 2. El  77000 144900
## 3: Hatchback 5 KapÄ± Sahibinden 2. El 145000  32900
## 4: Hatchback 3 KapÄ±  Galeriden 2. El 116000 109000
## 5:         Crossover Sahibinden 2. El  97000 166000
## 6:             Sedan Sahibinden 2. El    179  65000
```

Grouping

We can group the data with using the `by` parameter. We can get average prices for each brand like this:

```
car_market[, .(mean_price= mean(Fiyat)), by = Marka]
```

```
##          Marka mean_price
##  1:      Skoda  128163.50
##  2:       Seat  120630.11
##  3:    Peugeot  107926.51
##  4:       Opel   88531.83
##  5:     Nissan  131742.66
##  6:   Mercedes  316946.25
##  7:        Kia  117616.85
##  8:    Hyundai   98060.93
##  9:       Ford   94689.59
## 10:       Fiat   68076.59
## 11:  Chevrolet   99175.19
## 12:        BMW  294943.74
## 13:       Audi  259749.37
## 14:      Acura  123534.52
## 15:    Renault   85985.26
## 16:    Porsche  966249.82
## 17:       Mini  150258.06
## 18: Land Rover  594363.74
## 19:       Jeep  198877.36
## 20:      Honda  106489.06
## 21:      Dacia   81834.67
## 22:    Citroen   79485.17
## 23:      Mazda   64183.33
## 24: Alfa Romeo  110961.76
## 25: Mitsubishi   77779.69
## 26:       Lada   22143.75
## 27:     TofaÅŸ   17234.44
## 28:   Chrysler   94126.92
## 29:      Chery   25500.00
## 30:      Rover   34125.00
## 31:     Jaguar  531438.33
## 32:   Infiniti  228390.00
## 33:      Geely   43250.00
```

```
## 34: Volkswagen  123500.00
## 35:   Maserati  526740.67
## 36:      Isuzu  119788.89
##          Marka mean_price
```

We can summarize the data for more than one column with respect to more than one column.

```
car_market[, .(mean = mean(Fiyat), max = max(Fiyat)), by = c("Marka", "Arac Tip")]
```

```
##         Marka                         Arac Tip     mean    max
##    1:   Skoda                     1.2 Ambition  70500.0  92000
##    2:   Skoda                1.6 TDI Optimal 144287.5 146500
##    3:    Seat                          1.6\n  32900.0  32900
##    4:    Seat    1.4 TSI Sport Coupe Cupra\n 109000.0 109000
##    5: Peugeot             1.6 BlueHDi Allure 180150.0 252500
##   ---
## 1902:     BMW                320d Technology 200000.0 200000
## 1903:    Fiat                   1.3 M.jet EL  34500.0  34500
## 1904: Renault                            GTL  14000.0  14000
## 1905: Peugeot 1.6 HDi Premium Pack 112 bg A. 118000.0 118000
## 1906: Hyundai                  1.6 CRDi Team  80000.0  80000
```

We can get total number of rows with `.N` like:

```
car_market[, .N, by = Marka]
```

```
##          Marka    N
##  1:      Skoda  259
##  2:       Seat  151
##  3:    Peugeot  421
##  4:       Opel  647
##  5:     Nissan  229
##  6:   Mercedes  613
##  7:        Kia  166
##  8:    Hyundai  639
##  9:       Ford  602
## 10:       Fiat  654
## 11:  Chevrolet  127
## 12:        BMW  616
## 13:       Audi  430
## 14:      Acura   42
## 15:    Renault 2079
## 16:    Porsche   56
## 17:       Mini   31
## 18: Land Rover  182
## 19:       Jeep   91
## 20:      Honda  263
## 21:      Dacia  323
## 22:    Citroen  199
## 23:      Mazda   24
## 24: Alfa Romeo   17
## 25: Mitsubishi   32
```

```
## 26:       Lada   16
## 27:     TofaÅŸ   45
## 28:   Chrysler   13
## 29:      Chery    1
## 30:      Rover    4
## 31:     Jaguar   30
## 32:   Infiniti   10
## 33:      Geely    2
## 34: Volkswagen   18
## 35:   Maserati    3
## 36:      Isuzu    9
##          Marka    N
```

Index Operators

If we use the index of the data, we can perform better processes with respect to time. To set a column to the index we can use `setkey()` function like:

```
setkey(car_market, Date)
key(car_market)
```

```
## [1] "Date"
```

```
head(car_market)
```

```
##          Date   Marka Arac Tip Grubu                    Arac Tip Model_Year
## 1: 01/04/2020   Skoda          Rapid                1.2 Ambition       2015
## 2: 01/04/2020   Skoda        Octavia             1.6 TDI Optimal       2016
## 3: 01/04/2020    Seat          Ibiza                       1.6\n       1999
## 4: 01/04/2020    Seat          Ibiza 1.4 TSI Sport Coupe Cupra\n       2013
## 5: 01/04/2020 Peugeot           3008          1.6 BlueHDi Allure       2015
## 6: 01/04/2020    Opel          Astra                      1.6 16V       2004
##     Fuel_Type                Vites              CCM     Beygir Gucu       Renk
## 1: Benzin/LPG           DÃ¼z Vites 1300 cc ve altÄ± 100 BG ve altÄ±      Beyaz
## 2:      Dizel YarÄ± Otomatik Vites     1601-1800 cc      101-125 BG      Beyaz
## 3: Benzin/LPG       Otomatik Vites       Bilmiyorum      Bilmiyorum      SarÄ±
## 4:     Benzin       Otomatik Vites     1301-1600 cc      176-200 BG KÄ±rmÄ±zÄ±
## 5:      Dizel       Otomatik Vites     1301-1600 cc      101-125 BG      Beyaz
## 6: Benzin/LPG           DÃ¼z Vites     1301-1600 cc       76-100 BG       Mavi
##            Kasa Tipi     Kimden Durum     Km  Fiyat
## 1:             Sedan Sahibinden 2. El  98000  92000
## 2: Hatchback 5 KapÄ± Sahibinden 2. El  77000 144900
## 3: Hatchback 5 KapÄ± Sahibinden 2. El 145000  32900
## 4: Hatchback 3 KapÄ±  Galeriden 2. El 116000 109000
## 5:         Crossover Sahibinden 2. El  97000 166000
## 6:             Sedan Sahibinden 2. El    179  65000
```

As you can see, the data is sorted after by key setting it. Alse we can give multiple index like:

```
setkey(car_market, Date, Marka)
key(car_market)
```

```
## [1] "Date"  "Marka"
```

We can group the data by the key(s) with `bykey` argument like:

```
head(car_market[, .(mean_price= mean(Fiyat)), by = Date])
```

```
##          Date mean_price
## 1: 01/04/2020  120929.03
## 2: 01/05/2020  224372.28
## 3: 01/06/2020  108768.84
## 4: 02/04/2020   88376.47
## 5: 02/05/2020   94832.72
## 6: 02/06/2020   91292.09
```

Creating Columns

Adding Columns

We can create columns in data.table package. To do so we need to use `:=` operator like:

```
head(car_market[, fpk := Fiyat / Km])
```

```
##          Date     Marka Arac Tip Grubu                   Arac Tip Model_Year
## 1: 01/04/2020     Acura             CL                     DiÄŸer       2004
## 2: 01/04/2020      Audi             A4                    2.0 TDI       2014
## 3: 01/04/2020      Audi       S Serisi        S5 3.0 TFSI Quattro       2010
## 4: 01/04/2020       BMW       1 Serisi                     1.18i       2008
## 5: 01/04/2020       BMW       3 Serisi 3.20i ED EfficientDynamics       2014
## 6: 01/04/2020 Chevrolet        Captiva                   2.0 D LT       2007
##     Fuel_Type                Vites          CCM     Beygir Gucu  Renk
## 1: Benzin/LPG           DÃ¼z Vites 1301-1600 cc 100 BG ve altÄ±   Gri
## 2:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Beyaz
## 3:     Benzin YarÄ± Otomatik Vites 2501-3000 cc      326-350 BG Siyah
## 4: Benzin/LPG       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
## 5:     Benzin       Otomatik Vites 1301-1600 cc      176-200 BG Siyah
## 6:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
##            Kasa Tipi     Kimden Durum     Km  Fiyat       fpk
## 1:             Sedan Sahibinden 2. El 170000  55000 0.3235294
## 2:             Sedan  Galeriden 2. El 160000 176500 1.1031250
## 3:      Spor / Coupe  Galeriden 2. El 160000 350000 2.1875000
## 4: Hatchback 5 KapÄ± Sahibinden 2. El 162000  82000 0.5061728
## 5:             Sedan  Galeriden 2. El 160000 192500 1.2031250
## 6:      Arazi AracÄ± Sahibinden 2. El 192500  80000 0.4155844
```

We can also create multiple columns.

```
head(car_market[, c("kpf", "full_vehicle_type") := list(Km / Fiyat, paste(‘Arac Tip Grubu‘, ‘Arac Tip‘))
```

```
##          Date     Marka Arac Tip Grubu                   Arac Tip Model_Year
## 1: 01/04/2020     Acura             CL                     DiÄŸer       2004
## 2: 01/04/2020      Audi             A4                    2.0 TDI       2014
```

```
## 3: 01/04/2020      Audi       S Serisi        S5 3.0 TFSI Quattro       2010
## 4: 01/04/2020       BMW       1 Serisi                      1.18i       2008
## 5: 01/04/2020       BMW       3 Serisi 3.20i ED EfficientDynamics       2014
## 6: 01/04/2020 Chevrolet        Captiva                   2.0 D LT       2007
##     Fuel_Type                Vites          CCM     Beygir Gucu  Renk
## 1: Benzin/LPG          DÃ¼z Vites 1301-1600 cc 100 BG ve altÄ±   Gri
## 2:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Beyaz
## 3:     Benzin YarÄ± Otomatik Vites 2501-3000 cc      326-350 BG Siyah
## 4: Benzin/LPG       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
## 5:     Benzin       Otomatik Vites 1301-1600 cc      176-200 BG Siyah
## 6:      Dizel       Otomatik Vites 1801-2000 cc      126-150 BG Siyah
##            Kasa Tipi     Kimden Durum     Km  Fiyat       fpk       kpf
## 1:             Sedan Sahibinden 2. El 170000  55000 0.3235294 3.0909091
## 2:             Sedan  Galeriden 2. El 160000 176500 1.1031250 0.9065156
## 3:      Spor / Coupe  Galeriden 2. El 160000 350000 2.1875000 0.4571429
## 4: Hatchback 5 KapÄ± Sahibinden 2. El 162000  82000 0.5061728 1.9756098
## 5:             Sedan  Galeriden 2. El 160000 192500 1.2031250 0.8311688
## 6:      Arazi AracÄ± Sahibinden 2. El 192500  80000 0.4155844 2.4062500
##                      full_vehicle_type
## 1:                          CL DiÄŸer
## 2:                         A4 2.0 TDI
## 3:        S Serisi S5 3.0 TFSI Quattro
## 4:                       1 Serisi 1.18i
## 5: 3 Serisi 3.20i ED EfficientDynamics
## 6:                     Captiva 2.0 D LT
```

We can write sub queries like in SQL. We can show it from the code above:

```
car_market[, c("kpf", "full_vehicle_type") := list(Km / Fiyat, paste(‘Arac Tip Grubu‘, ‘Arac Tip‘))][,
```

```
##              kpf                    full_vehicle_type
##    1:  3.0909091                          CL DiÄŸer
##    2:  0.9065156                         A4 2.0 TDI
##    3:  0.4571429        S Serisi S5 3.0 TFSI Quattro
##    4:  1.9756098                       1 Serisi 1.18i
##    5:  0.8311688 3 Serisi 3.20i ED EfficientDynamics
##   ---
## 9040:  1.6525424 5008 1.6 HDi Premium Pack 112 bg A.
## 9041: 20.7142857                                5 GTL
## 9042:  2.5411765               Fluence 1.5 dCi Icon
## 9043:  0.7680000          Latitude 1.5 dCi Executive
## 9044:  0.3669725                      Megane 1.6 Joy
```

Aggregation

For the creation of new columns, we can use aggregate functions like min, max etc.

```
car_market[, .(mean = mean(Fiyat, na.rm = TRUE),
median = median(Fiyat, na.rm = TRUE),
min = min(Fiyat, na.rm = TRUE),
max = max(Fiyat, na.rm = TRUE))]
```

```
##        mean median  min     max
## 1: 148069.6  95500 5500 5086500
```

We can summarize multiple columns like:

```
car_market[, .(mean(Fiyat), mean(Km))]
```

```
##          V1       V2
## 1: 148069.6 135812.8
```

If we need to apply aggregation for more columns, we can use `.SD` and `.SDcols`. .SD means 'Subset of Data'. If we don't use .SDcols parameter, it includes all columns. If we use .SDcols, it applies that process for these columns.

```
car_market[, lapply(.SD, mean), .SDcols = c("Fiyat", "Km")]
```

```
##       Fiyat       Km
## 1: 148069.6 135812.8
```

Also, we can have multiple summarize function in the same line like:

```
car_market[, sapply(.SD, function(x) c(mean=mean(x), median=median(x)))]
```

```
##        Date Marka Arac Tip Grubu Arac Tip Model_Year Fuel_Type Vites CCM
## mean     NA    NA             NA       NA   2011.824        NA    NA  NA
## median   NA    NA             NA       NA   2014.000        NA    NA  NA
##        Beygir Gucu Renk Kasa Tipi Kimden Durum       Km    Fiyat       fpk
## mean            NA   NA        NA     NA    NA 135812.8 148069.6       Inf
## median          NA   NA        NA     NA    NA 125000.0  95500.0 0.8744919
##             kpf full_vehicle_type
## mean   2.237604                NA
## median 1.143522                NA
```