R Notebook

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data <- read.csv("crime_data.csv")
head(data, 5)
```

```
##         DR_NO      Date.Rptd       DATE.OCC TIME.OCC AREA   AREA.NAME
## 1 241711715   8/1/2024 0:00   8/1/2024 0:00     1319   17  Devonshire
## 2 231014031  9/21/2023 0:00  9/15/2023 0:00     1930   10 West Valley
## 3 231010808  6/27/2023 0:00  6/26/2023 0:00     1230   10 West Valley
## 4 211410441  4/25/2021 0:00  4/25/2021 0:00     2330   14     Pacific
## 5 211114569 10/25/2021 0:00 10/25/2021 0:00     1455   11   Northeast
##   Rpt.Dist.No Part.1.2 Crm.Cd                     Crm.Cd.Desc
## 1        1791        1    440 THEFT PLAIN - PETTY ($950 & UNDER)
## 2        1011        2    354                 THEFT OF IDENTITY
## 3        1015        2    354                 THEFT OF IDENTITY
## 4        1488        2    626  INTIMATE PARTNER - SIMPLE ASSAULT
## 5        1123        1    210                           ROBBERY
##            Mocodes Vict.Age Vict.Sex Vict.Descent Premis.Cd
## 1        0344 0394       25        M            O       501
## 2        1822 0930       23        F            W       501
## 3        1822 0928       37        F            O       501
## 4     0913 0400 0448       25        F            B       503
## 5 1309 0945 0334 0325        0        X            X       412
##                            Premis.Desc Weapon_Used_Cd
## 1              SINGLE FAMILY DWELLING             NA
## 2              SINGLE FAMILY DWELLING             NA
## 3              SINGLE FAMILY DWELLING             NA
## 4                              HOTEL            400
## 5 ELECTRONICS STORE (IE:RADIO SHACK, ETC.)          200
##                               Weapon.Desc Status Status.Desc Crm.Cd.1
## 1                                            IC Invest Cont      440
## 2                                            IC Invest Cont      354
## 3                                            IC Invest Cont      354
```

```
## 4 STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)     IC Invest Cont      626
## 5                 KNIFE WITH BLADE 6INCHES OR LESS     IC Invest Cont      210
##   Crm.Cd.2 Crm.Cd.3 Crm.Cd.4                              LOCATION
## 1       NA       NA       NA  8300    KELVIN                        AV
## 2       NA       NA       NA 18900    CANTLAY                       ST
## 3       NA       NA       NA  7300    ENFIELD                       AV
## 4       NA       NA       NA  5800 W  CENTURY                       BL
## 5       NA       NA       NA  2900    LOS FELIZ                     BL
##   Cross.Street     LAT      LON
## 1              34.2200 -118.5863
## 2              34.2023 -118.5458
## 3              34.2033 -118.5241
## 4              33.9456 -118.3835
## 5               0.0000    0.0000
```

```
#colSums(is.na(data))
```

```
missing_threshold <- nrow(data) * 0.5
cols_to_drop <- names(which(colSums(is.na(data)) > missing_threshold))
colSums(is.na(data))
```

```
##           DR_NO        Date.Rptd        DATE.OCC        TIME.OCC            AREA
##               0                0               0               0               0
##       AREA.NAME      Rpt.Dist.No        Part.1.2          Crm.Cd     Crm.Cd.Desc
##               0                0               0               0               0
##          Mocodes         Vict.Age        Vict.Sex    Vict.Descent       Premis.Cd
##               0                0               0               0               0
##     Premis.Desc  Weapon_Used_Cd     Weapon.Desc          Status     Status.Desc
##               0            33654               0               0               0
##        Crm.Cd.1         Crm.Cd.2        Crm.Cd.3        Crm.Cd.4        LOCATION
##               2            46448           49885           49995               0
##    Cross.Street             LAT             LON
##               0                0               0
```

```
clean_data <- data %>% select(-all_of(cols_to_drop))
```

```
glimpse(clean_data)
```

```
## Rows: 50,000
## Columns: 24
## $ DR_NO       <int> 241711715, 231014031, 231010808, 211410441, 211114569, 22~
## $ Date.Rptd   <chr> "8/1/2024 0:00", "9/21/2023 0:00", "6/27/2023 0:00", "4/2~
## $ DATE.OCC    <chr> "8/1/2024 0:00", "9/15/2023 0:00", "6/26/2023 0:00", "4/2~
## $ TIME.OCC    <int> 1319, 1930, 1230, 2330, 1455, 2239, 1430, 1930, 1750, 900~
## $ AREA        <int> 17, 10, 10, 14, 11, 14, 1, 12, 18, 7, 2, 13, 20, 7, 2, 4,~
## $ AREA.NAME   <chr> "Devonshire", "West Valley", "West Valley", "Pacific", "N~
## $ Rpt.Dist.No <int> 1791, 1011, 1015, 1488, 1123, 1432, 185, 1256, 1844, 721,~
## $ Part.1.2    <int> 1, 2, 2, 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, 1, 1, 1, 1, 2, 1, ~
## $ Crm.Cd      <int> 440, 354, 354, 626, 210, 230, 624, 310, 624, 354, 420, 76~
## $ Crm.Cd.Desc <chr> "THEFT PLAIN - PETTY ($950 & UNDER)", "THEFT OF IDENTITY"~
## $ Mocodes     <chr> "0344 0394", "1822 0930", "1822 0928", "0913 0400 0448", ~
## $ Vict.Age    <int> 25, 23, 37, 25, 0, 27, 25, 0, 53, 29, 0, 62, 44, 0, 41, 2~
## $ Vict.Sex    <chr> "M", "F", "F", "F", "X", "F", "F", "X", "F", "F", "", "M"~
```

```
## $ Vict.Descent <chr> "O", "W", "O", "B", "X", "H", "H", "X", "H", "W", "", "B"~
## $ Premis.Cd    <int> 501, 501, 501, 503, 412, 103, 103, 203, 501, 502, 108, 50~
## $ Premis.Desc  <chr> "SINGLE FAMILY DWELLING", "SINGLE FAMILY DWELLING", "SING~
## $ Weapon.Desc  <chr> "", "", "", "STRONG-ARM (HANDS, FIST, FEET OR BODILY FORC~
## $ Status       <chr> "IC", "IC", "IC", "IC", "IC", "AO", "IC", "IC", "IC", "IC~
## $ Status.Desc  <chr> "Invest Cont", "Invest Cont", "Invest Cont", "Invest Cont~
## $ Crm.Cd.1     <int> 440, 354, 354, 626, 210, 230, 624, 310, 624, 354, 420, 76~
## $ LOCATION     <chr> "8300    KELVIN                          AV", "18900    CANT~
## $ Cross.Street <chr> "", "", "", "", "", "", "", "", "", "", "", "", "HOOVER",~
## $ LAT          <dbl> 34.2200, 34.2023, 34.2033, 33.9456, 0.0000, 33.9947, 34.0~
## $ LON          <dbl> -118.5863, -118.5458, -118.5241, -118.3835, 0.0000, -118.~
```

```r
clean_data <- clean_data %>%
  mutate(
    DATE.OCC = as.Date(DATE.OCC, format = "%m/%d/%Y"),
    Year = as.numeric(format(DATE.OCC, "%Y")),
    Month = as.numeric(format(DATE.OCC, "%m")),
    Day = as.numeric(format(DATE.OCC, "%d")),
    Hour = as.numeric(substr(TIME.OCC, 1, 2))
  )
```

```r
burglary_2023 <- clean_data %>%
  filter(Year == 2023, Crm.Cd.Desc == "BURGLARY")

head(burglary_2023)
```

```
##       DR_NO      Date.Rptd   DATE.OCC TIME.OCC AREA  AREA.NAME Rpt.Dist.No
## 1 231107877  4/15/2023 0:00 2023-01-15     500   11  Northeast        1151
## 2 231912840  8/15/2023 0:00 2023-08-14    2200   19    Mission        1962
## 3 230813484  8/19/2023 0:00 2023-08-19     510    8    West LA         857
## 4 230126836 12/20/2023 0:00 2023-12-10    1200    1    Central         154
## 5 231506351  2/26/2023 0:00 2023-02-22    1230   15 N Hollywood       1562
## 6 231609277  6/17/2023 0:00 2023-06-12    2000   16   Foothill        1667
##   Part.1.2 Crm.Cd Crm.Cd.Desc         Mocodes Vict.Age Vict.Sex
## 1        1    310    BURGLARY    0531 1822 0451       53        F
## 2        1    310    BURGLARY    0314 0344 1402       45        F
## 3        1    310    BURGLARY 0358 1609 0321 1822        0        X
## 4        1    310    BURGLARY             344       34        M
## 5        1    310    BURGLARY       1609 0344       60        F
## 6        1    310    BURGLARY       1202 0344       80        F
##   Vict.Descent Premis.Cd          Premis.Desc Weapon.Desc Status Status.Desc
## 1            H       504        OTHER RESIDENCE                IC Invest Cont
## 2            B       507  CONDOMINIUM/TOWNHOUSE               IC Invest Cont
## 3            X       210   RESTAURANT/FAST FOOD               IC Invest Cont
## 4            B       221         PUBLIC STORAGE               IC Invest Cont
## 5            W       501 SINGLE FAMILY DWELLING               IC Invest Cont
## 6            W       501 SINGLE FAMILY DWELLING               IC Invest Cont
##   Crm.Cd.1                              LOCATION Cross.Street     LAT
## 1      310 5000 W  SUNSET                              BL         34.0981
## 2      310 15000   CORE                                LN         34.2424
## 3      310 9400 W  PICO                                BL         34.0553
## 4      310  100 E  6TH                                 ST         34.0460
## 5      310 4400    BABCOCK                             AV         34.1504
```
```

```
## 6       310  9400    LA TUNA CANYON                RD                34.2429
##         LON Year Month Day Hour
## 1 -118.2983 2023     1  15   50
## 2 -118.4596 2023     8  14   22
## 3 -118.3943 2023     8  19   51
## 4 -118.2493 2023    12  10   12
## 5 -118.4063 2023     2  22   12
## 6 -118.3417 2023     6  12   20
```

```r
crime_stats <- clean_data %>%
  group_by(AREA.NAME) %>%
  summarise(Total_Crimes = n(), Avg_Victim_Age = mean(Vict.Age, na.rm = TRUE)) %>%
  arrange(desc(Total_Crimes))  # Sort by total crimes

print(crime_stats)
```

```
## # A tibble: 21 x 3
##    AREA.NAME   Total_Crimes Avg_Victim_Age
##    <chr>              <int>          <dbl>
##  1 Central             3517           28.5
##  2 77th Street         3115           29.2
##  3 Pacific             3004           29.0
##  4 Southwest           2831           29.0
##  5 Hollywood           2556           28.3
##  6 Newton              2537           25.1
##  7 Olympic             2523           27.7
##  8 N Hollywood         2503           28.3
##  9 Southeast           2459           28.4
## 10 Rampart             2388           26.3
## # i 11 more rows
```

```r
monthly_crimes <- clean_data %>%
  group_by(Month) %>%
  summarise(Total_Crimes = n())

print(monthly_crimes)
```

```
## # A tibble: 12 x 2
##    Month Total_Crimes
##    <dbl>        <int>
##  1     1         4578
##  2     2         4290
##  3     3         4361
##  4     4         4189
##  5     5         4088
##  6     6         4058
##  7     7         4179
##  8     8         4147
##  9     9         4054
## 10    10         4226
## 11    11         3948
## 12    12         3882
```

```r
weapon_crimes <- clean_data %>%
  filter(Weapon.Desc != "") %>%
  summarise(Total_Weapon_Crimes = n())

print(weapon_crimes)
```

```
##   Total_Weapon_Crimes
## 1               16346
```

```r
premis_crimes <- clean_data %>%
  group_by(Premis.Desc) %>%
  summarise(Total_Crimes = n()) %>%
  arrange(desc(Total_Crimes))

print(premis_crimes)
```

```
## # A tibble: 267 x 2
##    Premis.Desc                                Total_Crimes
##    <chr>                                             <int>
##  1 STREET                                            13061
##  2 SINGLE FAMILY DWELLING                             8144
##  3 MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)       5947
##  4 PARKING LOT                                        3426
##  5 OTHER BUSINESS                                     2350
##  6 SIDEWALK                                           2050
##  7 VEHICLE, PASSENGER/TRUCK                           1520
##  8 GARAGE/CARPORT                                      979
##  9 DRIVEWAY                                            819
## 10 DEPARTMENT STORE                                    747
## # i 257 more rows
```

```r
clean_data <- clean_data %>%
  mutate(
    Severity_Score = case_when(
      !is.na(Weapon.Desc) ~ 5,
      Crm.Cd.Desc == "BURGLARY" ~ 3,
      TRUE ~ 1
    )
  )

area_severity <- clean_data %>%
  group_by(AREA.NAME) %>%
  summarise(Total_Severity_Score = sum(Severity_Score, na.rm = TRUE)) %>%
  arrange(desc(Total_Severity_Score))

print(area_severity)
```

```
## # A tibble: 21 x 2
##    AREA.NAME   Total_Severity_Score
##    <chr>                      <dbl>
##  1 Central                    17585
##  2 77th Street                15575
```

```
##  3 Pacific                15020
##  4 Southwest              14155
##  5 Hollywood              12780
##  6 Newton                 12685
##  7 Olympic                12615
##  8 N Hollywood            12515
##  9 Southeast              12295
## 10 Rampart                11940
## # i 11 more rows
```