week-4-tasks

Viki-Misi

October 13, 2022

Install the `nycflights13` package and check it contents.

```
# install.packages('nycflights13')
library(nycflights13)
library( tidyverse )
```

Read the 3 diamonds data sets and show the head of that.

`diamonds.txt`

```
##   carat       cut color clarity depth table price    x    y    z
## 1  0.23     Ideal     E     SI2  61.5    55   326 3.95 3.98 2.43
## 2  0.21   Premium     E     SI1  59.8    61   326 3.89 3.84 2.31
## 3  0.23      Good     E     VS1  56.9    65   327 4.05 4.07 2.31
## 4  0.29   Premium     I     VS2  62.4    58   334 4.20 4.23 2.63
## 5  0.31      Good     J     SI2  63.3    58   335 4.34 4.35 2.75
## 6  0.24 Very Good     J    VVS2  62.8    57   336 3.94 3.96 2.48
```

`diamonds.csv`

```
##   carat       cut color clarity depth table price    x    y    z
## 1  0.23     Ideal     E     SI2  61.5    55   326 3.95 3.98 2.43
## 2  0.21   Premium     E     SI1  59.8    61   326 3.89 3.84 2.31
## 3  0.23      Good     E     VS1  56.9    65   327 4.05 4.07 2.31
## 4  0.29   Premium     I     VS2  62.4    58   334 4.20 4.23 2.63
## 5  0.31      Good     J     SI2  63.3    58   335 4.34 4.35 2.75
## 6  0.24 Very Good     J    VVS2  62.8    57   336 3.94 3.96 2.48
```

`diamonds_with_extra_lines.txt`

```
##   carat       cut color clarity depth table price    x    y    z
## 1  0.23     Ideal     E     SI2  61.5    55   326 3.95 3.98 2.43
## 2  0.21   Premium     E     SI1  59.8    61   326 3.89 3.84 2.31
## 3  0.23      Good     E     VS1  56.9    65   327 4.05 4.07 2.31
## 4  0.29   Premium     I     VS2  62.4    58   334 4.20 4.23 2.63
## 5  0.31      Good     J     SI2  63.3    58   335 4.34 4.35 2.75
## 6  0.24 Very Good     J    VVS2  62.8    57   336 3.94 3.96 2.48
```

Read the `my_csvfile.csv` from `https://github.com/CEU-Economics-and-Business/ECBS-5208-Coding-1-Business-Analytics/tree/mast` repository.

```
##   country city_actual rating_count center1label center2label neighbourhood
## 1 Austria      Vienna           36  City centre    Donauturm   17. Hernals
## 2 Austria      Vienna          189  City centre    Donauturm   17. Hernals
## 3 Austria      Vienna           53  City centre    Donauturm    Alsergrund
## 4 Austria      Vienna           55  City centre    Donauturm    Alsergrund
## 5 Austria      Vienna           33  City centre    Donauturm    Alsergrund
## 6 Austria      Vienna           25  City centre    Donauturm    Alsergrund
##   price   city stars ratingta ratingta_count scarce_room hotel_id offer
## 1    81 Vienna     4      4.5            216           1    21894     1
## 2    81 Vienna     4      3.5            708           0    21897     1
```

```
## 3    85 Vienna     4      3.5           629           0    21901     1
## 4    83 Vienna     3      4.0            52           0    21902     1
## 5    82 Vienna     4      3.5           219           1    21903     1
## 6   229 Vienna     5      4.5            27           1    21904     1
##      offer_cat year month weekend holiday distance distance_alter
## 1 15-50% offer 2017    11       0       0      2.7            4.4
## 2  1-15% offer 2017    11       0       0      1.7            3.8
## 3 15-50% offer 2017    11       0       0      1.4            2.5
## 4 15-50% offer 2017    11       0       0      1.7            2.5
## 5 15-50% offer 2017    11       0       0      1.2            2.8
## 6  1-15% offer 2017    11       0       0      0.9            3.0
##   accommodation_type nnights rating
## 1          Apartment       1    4.4
## 2              Hotel       1    3.9
## 3              Hotel       1    3.7
## 4              Hotel       1    4.0
## 5              Hotel       1    3.9
## 6          Apartment       1    4.8
```

Filter the `flights` data where departure delays was over 1000.

```
## # A tibble: 5 × 19
##    year month   day dep_time sched_dep…¹ dep_d…² arr_t…³ sched…⁴ arr_d…⁵ carrier
##   <int> <int> <int>    <int>       <int>   <dbl>   <int>   <int>   <dbl> <chr>
## 1  2013     1     9      641         900    1301    1242    1530    1272 HA
## 2  2013     1    10     1121        1635    1126    1239    1810    1109 MQ
## 3  2013     6    15     1432        1935    1137    1607    2120    1127 MQ
## 4  2013     7    22      845        1600    1005    1044    1815     989 MQ
## 5  2013     9    20     1139        1845    1014    1457    2210    1007 AA
## # … with 9 more variables: flight <int>, tailnum <chr>, origin <chr>,
## #   dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>,
## #   time_hour <dttm>, and abbreviated variable names ¹sched_dep_time,
## #   ²dep_delay, ³arr_time, ⁴sched_arr_time, ⁵arr_delay
```

Filter the `flights` data where departure delays was over 500 and destination was `MIA` and carrier either `EV, AA, US` .

```
## # A tibble: 5 × 19
##    year month   day dep_time sched_dep…¹ dep_d…² arr_t…³ sched…⁴ arr_d…⁵ carrier
##   <int> <int> <int>    <int>       <int>   <dbl>   <int>   <int>   <dbl> <chr>
## 1  2013    11    24     2301        1225     636     149    1535     614 AA
## 2  2013    12     5      756        1700     896    1058    2020     878 AA
## 3  2013    12    15      625        1925     660     933    2245     648 AA
## 4  2013    12    17      705        1700     845    1026    2020     846 AA
## 5  2013     7    21     1555         615     580    1955     910     645 AA
## # … with 9 more variables: flight <int>, tailnum <chr>, origin <chr>,
## #   dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>,
## #   time_hour <dttm>, and abbreviated variable names ¹sched_dep_time,
## #   ²dep_delay, ³arr_time, ⁴sched_arr_time, ⁵arr_delay
```

Merge the result with the airlines datasets.

```
##   carrier year month day dep_time sched_dep_time dep_delay arr_time
## 1      AA 2013    11  24     2301           1225       636      149
## 2      AA 2013    12   5      756           1700       896     1058
## 3      AA 2013    12  15      625           1925       660      933
## 4      AA 2013    12  17      705           1700       845     1026
## 5      AA 2013     7  21     1555            615       580     1955
##   sched_arr_time arr_delay flight tailnum origin dest air_time distance hour
## 1           1535       614   1697  N634AA    JFK  MIA      143     1089   12
## 2           2020       878    172  N5DMAA    EWR  MIA      149     1085   17
## 3           2245       648   2437  N635AA    LGA  MIA      165     1096   19
## 4           2020       846    172  N5EMAA    EWR  MIA      145     1085   17
## 5            910       645   1895  N3EMAA    EWR  MIA      177     1085    6
```

```
##   minute           time_hour                   name
## 1     25 2013-11-24 12:00:00 American Airlines Inc.
## 2      0 2013-12-05 17:00:00 American Airlines Inc.
## 3     25 2013-12-15 19:00:00 American Airlines Inc.
## 4      0 2013-12-17 17:00:00 American Airlines Inc.
## 5     15 2013-07-21 06:00:00 American Airlines Inc.
```

Write out the result into

- csv without rownames
- excel
- rds

Create a new column in the flights dataset of the first three column which will be a string like `2015-01-15`

```
## # A tibble: 6 × 4
##    year month   day my_date
##   <int> <int> <int> <chr>
## 1  2013     1     1 2013-01-01
## 2  2013     1     1 2013-01-01
## 3  2013     1     1 2013-01-01
## 4  2013     1     1 2013-01-01
## 5  2013     1     1 2013-01-01
## 6  2013     1     1 2013-01-01
```

Change it to date column

```
## # A tibble: 6 × 4
##    year month   day my_date
##   <int> <int> <int> <date>
## 1  2013     1     1 2013-01-01
## 2  2013     1     1 2013-01-01
## 3  2013     1     1 2013-01-01
## 4  2013     1     1 2013-01-01
## 5  2013     1     1 2013-01-01
## 6  2013     1     1 2013-01-01
```

Create a new column `delays` which will indicate the delays, sum of absulute values of two delays

```
## # A tibble: 6 × 6
##    year month   day dep_delay arr_delay delays
##   <int> <int> <int>     <dbl>     <dbl>  <dbl>
## 1  2013     1     1         2        11     13
## 2  2013     1     1         4        20     24
## 3  2013     1     1         2        33     35
## 4  2013     1     1        -1       -18     19
## 5  2013     1     1        -6       -25     31
## 6  2013     1     1        -4        12     16
```

What is the carrier name of the flights which had the most departure delay.

```
##                    name
## 1 Hawaiian Airlines Inc.
```