ica4_shuangyu_zhao

shuangyu_zhao

2023-01-19

```
complaint_his <- read.csv("/Users/apple/Desktop/STT811 appl_stat_model/data/Complaint_History.csv")
customer_his <- read.csv("/Users/apple/Desktop/STT811 appl_stat_model/data/Customer_History.csv")
head(complaint_his)
```

```
##   CustomerID ComplaintYear
## 1        209          2007
## 2        427          2015
## 3         24          2009
## 4        334          2019
## 5        353          2014
## 6         65          2010
```

```
head(customer_his)
```

```
##   Customer First_Year Last_Year Year_Born Male_Female
## 1        1       2007      2017      1961           F
## 2        2       2007      2016      1953           F
## 3        3       2011      2020      1957           F
## 4        4       2006      2014      1988           M
## 5        5       2005      2018      1965           F
## 6        6       2009      2019      1979           F
```

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/Resources/
## library/tcltk/libs//tcltk.so'' had status 1
```

```
## Loading required package: RSQLite
```

(a) Summarizing the number of complaints by customer and year.

```
sum_complain <- sqldf("SELECT CustomerID, ComplaintYear, COUNT(*) AS complaint_times
                       FROM complaint_his
                       GROUP BY CustomerID, ComplaintYear")
head(sum_complain)
```

```
##   CustomerID ComplaintYear complaint_times
## 1          1          2007               1
## 2          1          2011               1
## 3          2          2009               1
## 4          3          2011               1
## 5          3          2016               1
## 6          3          2018               1
```

(b) Add a column for predictive year

```
sum_complain_pre <- sqldf("SELECT *, ComplaintYear+1 AS predict_year
                           FROM sum_complain")
head(sum_complain_pre)
```

```
##   CustomerID ComplaintYear complaint_times predict_year
## 1          1          2007               1         2008
## 2          1          2011               1         2012
## 3          2          2009               1         2010
## 4          3          2011               1         2012
## 5          3          2016               1         2017
## 6          3          2018               1         2019
```

(c) Add a column which includes all complaints made by the customer on or before the year field (not
    predictive year).

```
sum_complain_pre_before <- sqldf("SELECT ta2.ComplaintYear AS year, ta2.complaint_times AS new_complaint
                                  FROM sum_complain_pre AS ta1
                                  INNER JOIN sum_complain_pre AS ta2
                                  ON ta1.CustomerID = ta2.CustomerID
                                  WHERE ta1.ComplaintYear >= year")
head(sum_complain_pre_before)
```

```
##   year new_complaint_times CustomerID ComplaintYear complaint_times
## 1 2007                   1          1          2007               1
## 2 2007                   1          1          2011               1
## 3 2011                   1          1          2011               1
## 4 2009                   1          2          2009               1
## 5 2011                   1          3          2011               1
## 6 2011                   1          3          2016               1
##   predict_year
## 1         2008
## 2         2012
## 3         2012
## 4         2010
## 5         2012
## 6         2017
```

```
sum_complain_pre_before <- sqldf("SELECT ta.CustomerID, ta.ComplaintYear, ta.complaint_times, ta.predict
                                  SUM(new_complaint_times) AS past_times
                                  FROM sum_complain_pre_before AS ta
                                  GROUP BY CustomerID, ComplaintYear")

head(sum_complain_pre_before)
```

```
##   CustomerID ComplaintYear complaint_times predict_year past_times
## 1          1          2007               1         2008          1
## 2          1          2011               1         2012          2
## 3          2          2009               1         2010          1
## 4          3          2011               1         2012          1
## 5          3          2016               1         2017          2
## 6          3          2018               1         2019          3
```