Renaissance: Задание 2

Georgie Shimanovsky

13 11 2017

Описание шагов по решению задания №2

Загрузка пакета “data.table” / данных из CSV файла текущей рабочей папки.

```
require(data.table)
```

```
csv2.path <- "data.csv" #datafile path
df.test2  <- read.csv2(csv2.path, stringsAsFactors = FALSE,
                       fileEncoding = "windows-1251")[, 1:3]
dt.test2  <- as.data.table(df.test2) #Data Frame to data.table
names(dt.test2) <- c("case_id", "part1", "part2") #Columns rename
dt.test2
```

```
##      case_id                          part1                          part2
##   1:       1     Беляев Владислав Аркадьевич   Ефимов Малик Константинович
##   2:       2         Давыдова Ануш Оскаровна     Никифорова Божена Львовна
##   3:       3         Белов Сергей Михайлович  Третьяков Никита Харитонович
##   4:       4         Власов Артём Михайлович Аксенова Вероника Кузьминична
##   5:       5          Яковлев Алан Макарович      Сысоев Елисей Тимофеевич
##  ---
## 257:     257         Михайлова Юлия Петровна  Колобова Милослава Натановна
## 258:     258         Киселев Денис Романович     Тимофеева Эльвира Юрьевна
## 259:     259 Щербаков Мстислав Станиславович    Сергеева Татьяна Даниловна
## 260:     260         Гущина Ирина Эльдаровна       Лукин Евгений Антонович
## 261:     261         Крылова Таисия Ждановна  Николаев Святослав Матвеевич
```

Подготовка (tidy) данных для анализа.

Группировка имен в один столбец, без потери информация о № участника.

```
test2.tidy <- data.table::melt(dt.test2, id = 1)
test2.tidy
```

```
##      case_id variable                        value
##   1:       1    part1  Беляев Владислав Аркадьевич
##   2:       2    part1      Давыдова Ануш Оскаровна
##   3:       3    part1      Белов Сергей Михайлович
##   4:       4    part1      Власов Артём Михайлович
##   5:       5    part1       Яковлев Алан Макарович
##  ---
## 518:     257    part2 Колобова Милослава Натановна
## 519:     258    part2    Тимофеева Эльвира Юрьевна
## 520:     259    part2   Сергеева Татьяна Даниловна
## 521:     260    part2      Лукин Евгений Антонович
## 522:     261    part2 Николаев Святослав Матвеевич
```

Новый столбец: сумма страховых случаев по каждому имени без учета № участника.

```
accid.num <- test2.tidy[, .(case_id, cases_ttl = .N), by = value][]
accid.num
```

```
##                             value case_id cases_ttl
##   1:  Беляев Владислав Аркадьевич       1         1
##   2:      Давыдова Ануш Оскаровна       2         1
##   3:      Белов Сергей Михайлович       3         1
##   4:      Власов Артём Михайлович       4         1
##   5:       Яковлев Алан Макарович       5         1
##  ---
## 518: Колобова Милослава Натановна     257         1
## 519:    Тимофеева Эльвира Юрьевна     258         1
## 520:   Сергеева Татьяна Даниловна     259         1
## 521:      Лукин Евгений Антонович     260         1
## 522: Николаев Святослав Матвеевич     261         1
```

Количество страховых случаев перенесено в названия новых столбцов, значения этих столбцов - соотносимое количество участников страхового случая. Данные в разрезе уникальных страховых случаев - строк.

```
cast.accid <- dcast(accid.num, case_id ~ cases_ttl, fun.aggregate = length)
cast.accid
```

```
##      case_id 1 2 3 5
##   1:       1 2 0 0 0
##   2:       2 2 0 0 0
##   3:       3 2 0 0 0
##   4:       4 2 0 0 0
##   5:       5 2 0 0 0
##  ---
## 257:     257 2 0 0 0
## 258:     258 2 0 0 0
## 259:     259 2 0 0 0
## 260:     260 2 0 0 0
## 261:     261 2 0 0 0
```

Анализ

Подозрения на мошенничество: присвоение классификации по страховым случаям. “Low”: Если оба участника имеют по одному страховому случаю.
“Medium”: Если только один из участников имеет более одного страхового случая.
“High”: Если оба участника имеют более одного страхового случая.

```
cast.accid[cast.accid$"1" == 2, "fraud" := "Low"]
cast.accid[cast.accid$"1" == 1, "fraud" := "Medium"]
cast.accid[cast.accid$"1" == 0, "fraud" := "High"]
```

Совмещение таблицы подозрительных случаев с таблицей данных об участниках.

```
# Merge fraud ranking with with total_cases
dt.full <- merge(accid.num, cast.accid, by = "case_id")
dt.full
```

```
##      case_id                       value cases_ttl 1 2 3 5 fraud
##   1:       1  Беляев Владислав Аркадьевич         1 2 0 0 0   Low
##   2:       1  Ефимов Малик Константинович         1 2 0 0 0   Low
```

```
##   3:       2      Давыдова Ануш Оскаровна         1 2 0 0 0   Low
##   4:       2    Никифорова Божена Львовна         1 2 0 0 0   Low
##   5:       3      Белов Сергей Михайлович         1 2 0 0 0   Low
##  ---
## 518:     259   Сергеева Татьяна Даниловна         1 2 0 0 0   Low
## 519:     260      Гущина Ирина Эльдаровна         1 2 0 0 0   Low
## 520:     260      Лукин Евгений Антонович         1 2 0 0 0   Low
## 521:     261      Крылова Таисия Ждановна         1 2 0 0 0   Low
## 522:     261 Николаев Святослав Матвеевич         1 2 0 0 0   Low
```

Результат

Выделение списка лиц подозреваемых в мошенничестве.

```
# List of names with high fraud suspicion
dt.suspicion <- dt.full[fraud == "High", .(name = unique(value))]
dt.suspicion
```

```
##                              name
## 1: Мухамадеев Александр Валерьевич
## 2:   Сенчукова Екатерина Семеновна
## 3:         Комин Сергей Николаевич
## 4:       Павлова Мария Геннадиевна
## 5:     Воробьев Иван Александрович
## 6:      Рогачев Антон Владимирович
## 7:     Коробов Вадим Александрович
```

Подозрительные случаи и лица в первоначальном формате.

```
dt.test2[unique(dt.full[fraud == "High"]$case_id)]
```

```
##    case_id                          part1                         part2
## 1:      50 Мухамадеев Александр Валерьевич Сенчукова Екатерина Семеновна
## 2:      63         Комин Сергей Николаевич     Павлова Мария Геннадиевна
## 3:      90     Воробьев Иван Александрович       Комин Сергей Николаевич
## 4:     116      Рогачев Антон Владимирович Сенчукова Екатерина Семеновна
## 5:     137 Мухамадеев Александр Валерьевич   Коробов Вадим Александрович
## 6:     143       Павлова Мария Геннадиевна   Воробьев Иван Александрович
## 7:     185     Коробов Вадим Александрович    Рогачев Антон Владимирович
```