

# wi\_sem\_team\_14\_plotting

Chris

7/11/2021

Short summary of the raw dataset

```
summary(data_event_log)
```

```
##      CASE_ID          ACTIVITY          TIMESTAMP
## Length:178078      Length:178078      Min.   :2013-05-22 10:39:39
## Class :character    Class :character    1st Qu.:2018-06-11 09:41:52
## Mode  :character    Mode  :character    Median :2018-10-31 10:17:36
##                                         Mean  :2018-10-16 14:14:51
##                                         3rd Qu.:2019-02-23 10:12:57
##                                         Max.   :2019-06-28 08:39:30
## REPAIR_IN_TIME_5D  DEVICETYPE          SERVICEPOINT
## Min.   :0.000      Length:178078      Length:178078
## 1st Qu.:0.000      Class :character    Class :character
## Median :0.000      Mode  :character    Mode  :character
## Mean    :0.326
## 3rd Qu.:1.000
## Max.    :1.000
```

```
head(data_event_log)
```

```
## # A tibble: 6 x 6
##   CASE_ID ACTIVITY  TIMESTAMP          REPAIR_IN_TIME_~ DEVICETYPE SERVICEPOINT
##   <chr>   <chr>    <dtm>          <dbl> <chr>      <chr>
## 1 Case10  Creation  2018-01-02 13:39:47      0 AB52      E
## 2 Case10  Letter    2018-01-05 00:00:00      0 AB52      E
## 3 Case10  DeviceRe~ 2018-01-05 16:45:34      0 AB52      E
## 4 Case10  StockEnt~ 2018-01-17 00:00:00      0 AB52      E
## 5 Case10  InDelive~ 2018-01-17 00:00:00      0 AB52      E
## 6 Case10  NoteWork~ 2018-01-17 07:37:19      0 AB52      E
```

Wertebereich für interessante Spalten ausgeben

```
unique(data_event_log$ACTIVITY)
```

```
## [1] "Creation"      "Letter"        "DeviceReceived" "StockEntry"
## [5] "InDelivery"    "NoteWorkshop"  "Completed"      "NoteHotline"
## [9] "StatusRequest" "Transmission"  "Approved"       "FreeticketCust"
## [13] "FreeticketComp"
```

```
unique(data_event_log$DEVICETYPE)
```

```
## [1] "AB52" "AB41" "AB47" "AB22" "AB49" "AB62" "AB29" "AB63" "AB20" "AB53"
## [11] "AB50" "AB44" "AB45" "AB36" "AB61" "AB16" "AB34" "AB25" "AB40" "AB8"
## [21] "AC68" "AB38" "AB65" "AB60" "AB31" "AB27" "AB10" "AB19" "AB59" "AB21"
## [31] "AB56" "AB26" "AB55" "AB9"  "AB58" "AB39" "AB14" "AB43" "AB24" "A07"
```

```
## [41] "AB57" "AB23" "AB28" "AB64" "AB32" "AB15" "AB30" "AF3" "AB33" "AG5"
## [51] "AB12" "AB51" "AB54" "AB18" "AB17" "AB35" "AB46" "AB37" "AB48" NA
## [61] "AB42" "AG4" "AB66" "AB67" "AB13"
```

```
unique(data_event_log$SERVICEPOINT)
```

```
## [1] "E" "G" "J" "L" NA "C" "H" "I" "K" "D" "B" "A"
```

```
unique(data_event_log$REPAIR_IN_TIME_5D)
```

```
## [1] 0 1
```

## Data cleaning

Some data exploration

```
## [1] "Number of datapoint in the clean dataset:"
```

```
## [1] 161553
```

```
## [1] "Number of unique case IDs:"
```

```
## [1] 21931
```

## Some data modification

- creating column DATE (timestamps without the time information)
- creating column WEEKDAY (not sure if we need this as a column in the dataset, can just compute it insitro when needed)

```
# writing the modified df to csv with relative path to the folder "data"
write.csv(df_clean_logs, "../data/modified_logs.csv")
```

## Basic univariate plotting

our quantitative variables are :

- NONE ??

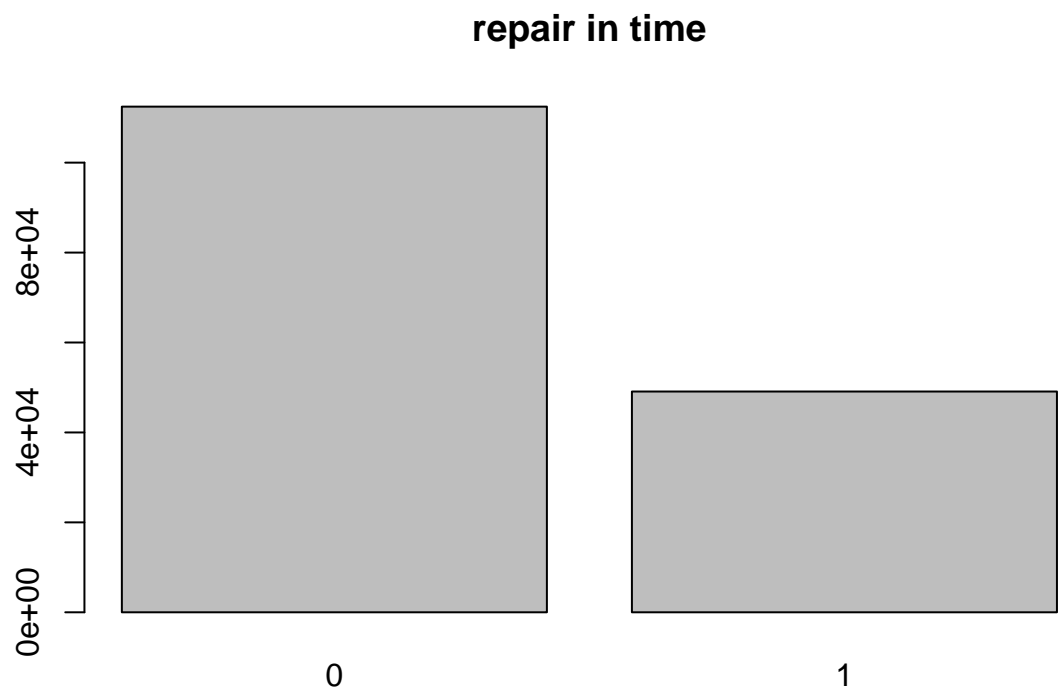
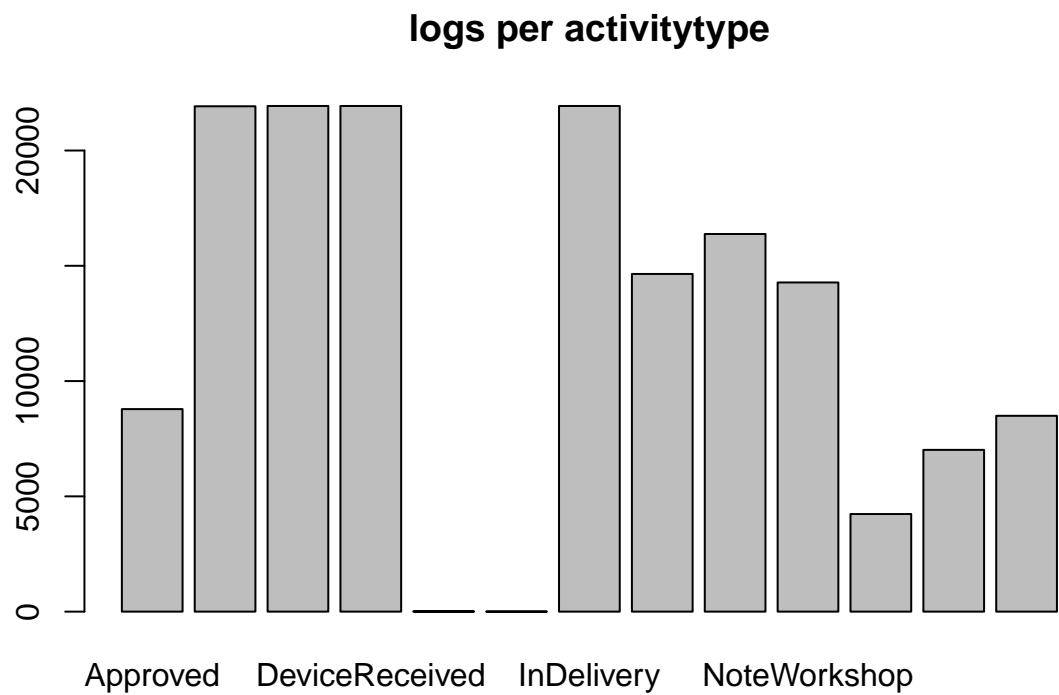
our qualitative variables are:

- CASE\_ID (not sure tbh, bc this is part of the “primary key” of the dataset entities) {string} <- maybe convert to integer for easier processing
- ACTIVITY {string}
- SERVICEPOINT {char}
- DEVICETYPE {string}
- REPAIR\_IN\_TIME {double} <- maybe convert to boolean for easier processing

neither?:

- TIMESTAMP {double - feels more like a string tho}

Plotting the frequency of our qualitative variables:



### logs per devicetype

