05 Description Clustering

Fabian Blasch

05/30/2022

1 Load Data

```r
# source AUX
source("././../Misc/Auxilliary.R")

# packages
get.package(c("bizdays", "lubridate"))

# load data
auctions <- readRDS("./../../Data/Bid Tab RDS/Bid_Tabs_Cleaned.RDS")
```

2 Final Checks before Mining

```r
# check for descriptions that are shorter than 50 characters after cleaning
lapply(auctions, \(x){

  # over project ID
  lapply(x, \(y){

    # Description length
    Dlen <- y$Text["Contract Description"] |> nchar()

    # required min char count
    dlen_check <- Dlen > 50

    # return
    list("Description" = Dlen,
         "Dlcheck" = dlen_check)

  })

}) -> filter_dat

# pull index

lapply(filter_dat, \(x){

  sapply(x, \(x){

    # extr
    x[["Dlcheck"]]
```

```
  })

}) -> ind

# files that will be removed
(!unlist(ind)) |> sum()
```

```
## [1] 4
```

```
# subset
Map(\(au, ind) au[ind], auctions, ind) -> auctions
```

3