ML\_daphnia.R

US16120

2019-09-26

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.6.1

## Registered S3 methods overwritten by 'ggplot2':  
## method from   
## [.quosures rlang  
## c.quosures rlang  
## print.quosures rlang

## Registered S3 method overwritten by 'rvest':  
## method from  
## read\_xml.response xml2

## -- Attaching packages ------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.1 v purrr 0.3.2   
## v tibble 2.1.1 v dplyr 0.8.0.1  
## v tidyr 0.8.3 v stringr 1.4.0   
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ---------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readxl)  
library(janitor)

##   
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(here)

## Warning: package 'here' was built under R version 3.6.1

## here() starts at C:/Users/us16120/Projects/repos/odsc\_west\_2019

library(dplyr)  
library(reticulate)

## Warning: package 'reticulate' was built under R version 3.6.1

path\_to\_data <- here('data', 'Khan\_Chemosphere\_229\_8.xlsx')  
  
df <-  
 readxl::read\_xlsx(path\_to\_data, sheet = 'Daphnia Dataset') %>%  
 data.frame() %>%  
 janitor::clean\_names()

## New names:  
## \* `` -> ...4

names(df)

## [1] "id\_biocides" "cas\_number" "canonical\_smiles"   
## [4] "x4" "exp\_log\_ec50\_mmol\_l" "geom\_mean\_value\_mmol"  
## [7] "mol\_l" "p\_ec50\_mol\_l\_daphnia"

df <- df %>%  
 select(cas\_number, canonical\_smiles, p\_ec50\_mol\_l\_daphnia) %>%  
 rename('CAS' = 'cas\_number',  
 'SMILES' = 'canonical\_smiles',  
 'pEC50' = 'p\_ec50\_mol\_l\_daphnia')