Data engineering assignment

MongoDB dataset code exports for the report

Kory

2024-03-24

# Mongo DB dataset exports

## [1] "number of missing values in entire dataset: 0"

## [1] "is student\_id a PK: FALSE"

## [1] "is class\_id a PK: FALSE"

## student\_id rows count type rows count score rows count   
## 400000 400000 400000   
## grade rows count class\_id rows count   
## 400000 400000

## student\_id unique rows count type unique rows count   
## 10000 3   
## score unique rows count grade unique rows count   
## 400000 5   
## class\_id unique rows count   
## 501

## [1] "the range of scores (difference between the min and max scores): 99.9994806250756"

## [1] "mean: 49.9107237106642"

## [1] "median: 49.9121492863691"

## [1] "maximum: 99.9997434060823"

## [1] "minimum: 0.000262781006732382"

## [1] "1st quartile: 24.9377978071595"

## [1] "3rd quartile: 74.798609800433"

## [1] "standard deviation: 28.8425037321742"

## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.  
## ℹ Please use `after\_stat(density)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.







