

EDA MinneMUDAC

Al Ashir Intisar

2025-03-13

```
# to read in .xlsx files
library(readxl)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# library for dataset summary functions
library(skimr)
#library for join functions
library(dplyr)
# library for writing/saving excel file
library(openxlsx)
# ggplot library for plots
library(ggplot2)

# reading in the training data
data <- read_excel("~/MUDAC/minneMUDAC2025/Data/Training.xlsx")

# filtering in only date related variables.
date_variables_df <- data %>%
  select("Big ID", "Little ID", "Match ID 18Char", "Big Acceptance Date", "Big Approved Date", "Big Day)

# checking how many rows have big acceptance date in it
date_variables_df %>%
  filter(!is.na(`Big Acceptance Date`)) %>%
  summarise(count = n())

## # A tibble: 1 x 1
##   count
##   <int>
## 1 14442

# Checking unique big ID to see how many unique bigs are there in the dataset
date_variables_df %>%
  distinct(`Big ID`) %>%
  summarise(count = n())
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1  3080
```

```
date_variables_df %>%
  filter(!is.na(`Big Acceptance Date`)) %>%
  distinct(`Big ID`, .keep_all = TRUE) %>%
  summarise(count = n())
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1  1273
```

```
date_variables_df %>%
  select("Big Re-Enroll", "Big ID") %>%
  filter(!is.na(`Big Re-Enroll`)) %>%
  distinct(`Big ID`, .keep_all = TRUE) %>%
  group_by(`Big Re-Enroll`) %>%
  summarise(count = n())
```

```
## # A tibble: 2 x 2
##   `Big Re-Enroll` count
##           <dbl> <int>
## 1             0  1261
## 2             1    13
```