01_make_data_.Rmd

Section 1

The NCAA March Madness dataset includes disccriptive variables, performance analytics, and public sentement regarding each team eligible for the yearly NCAA mens basketball tournament.

Section 2

The aim is to use K-mean clustering to show the distribution of each teams performance and group teams based on their perfoamcne leading up to the NCAA tournament.

Section 3

```
## install.packages("tidytuesdayR")
tuesdata <- tidytuesdayR::tt_load('2024-03-26')
```

```
## ---- Compiling #TidyTuesday Information for 2024-03-26 ----
## --- There are 2 files available ---
##
##
## ─ Downloading files ─────────────────────────────────────
##
##   1 of 2: "team-results.csv"
##   2 of 2: "public-picks.csv"
```

```
team_results <- tuesdata$`team-results`; team_results
```

```
## # A tibble: 236 × 20
##     TEAMID TEAM   PAKE PAKERANK   PASE PASERANK GAMES     W     L WINPERCENT   R64
##      <dbl> <chr> <dbl>    <dbl>  <dbl>    <dbl> <dbl> <dbl> <dbl>      <dbl> <dbl>
##  1       1 Abil…   0.7       45    0.7       52     3     1     2      0.333     2
##  2       2 Akron  -0.9      179   -1.1      187     4     0     4      0         4
##  3       3 Alab…  -2.1      211   -2.9      220    10     5     5      0.5       5
##  4       4 Alba…  -0.4      147   -0.3      138     3     0     3      0         3
##  5       6 Amer…  -0.5      160   -0.4      150     3     0     3      0         3
##  6       8 Ariz…  -1.7      206   -2.5      216    28    17    11      0.607    11
##  7       9 Ariz…  -2        209   -1.9      206     5     1     4      0.2       4
##  8      10 Arka…   4.3       11    3.5       16    18    11     7      0.611     7
##  9      11 Arka…   0         76    0         78     1     0     1      0         1
## 10      12 Aubu…   0.6       53    1.4       30    11     7     4      0.636     4
## # i 226 more rows
## # i 9 more variables: R32 <dbl>, S16 <dbl>, E8 <dbl>, F4 <dbl>, F2 <dbl>,
## #   CHAMP <dbl>, TOP2 <dbl>, F4PERCENT <chr>, CHAMPPERCENT <chr>
```

```
public_picks <- tuesdata$`public-picks`; public_picks
```

```
## # A tibble: 64 × 9
##     YEAR TEAMNO TEAM                   R64    R32    S16    E8     F4     FINALS
##    <dbl>  <dbl> <chr>                  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>
##  1  2024   1079 Akron                  7.16%  2.18%  0.63%  0.22%  0.10%  0.04%
##  2  2024   1078 Alabama                87.11% 55.73% 13.55% 7.31%  2.89%  1.10%
##  3  2024   1077 Arizona                95.19% 81.05% 52.02% 22.26% 10.05% 4.06%
##  4  2024   1076 Auburn                 90.17% 59.97% 15.93% 7.95%  3.64%  1.34%
##  5  2024   1075 Baylor                 92.34% 71.07% 33.68% 12.15% 4.35%  1.83%
##  6  2024   1073 BYU                    75.17% 21.70% 6.92%  1.64%  0.64%  0.32%
##  7  2024   1072 Clemson                51.46% 10.43% 3.16%  0.79%  0.30%  0.16%
##  8  2024   1071 Colgate                6.38%  1.89%  0.54%  0.20%  0.09%  0.04%
##  9  2024   1070 College of Charleston 11.57% 2.72%  0.68%  0.24%  0.10%  0.05%
## 10  2024   1069 Colorado               22.97% 3.66%  0.90%  0.35%  0.11%  0.05%
## # i 54 more rows
```

Section 4

```
dim(team_results)
```

```
## [1] 236  20
```

```
dim(public_picks)
```

```
## [1] 64  9
```