

# Assignment-1

Team-2

2023-02-09

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr  1.0.1
## v tibble  3.1.8      v dplyr  1.1.0
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 1.0.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(dplyr)
library(ggplot2)
library(readxl)
```

```
Spotify <- read_excel("E:/Piyush Files/SEM 2/Intro to Analytics Systems/Assignment 1/archive/Spotify.xlsx")
str(Spotify)
```

```
## tibble [49 x 9] (S3: tbl_df/tbl/data.frame)
##  $ Position      : num [1:49] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Artist Name    : chr [1:49] "abc" "Juice WRLD" "Lil Uzi Vert" "J. Cole" ...
##  $ Song Name      : chr [1:49] "Sunflower SpiderMan: Into the SpiderVerse" NA "XO TOUR Llif3" ...
##  $ Days           : num [1:49] 1506 1673 1853 2547 1223 ...
##  $ Top 10 (xTimes) : num [1:49] 302 178 212 6 186 4 233 44 133 2 ...
##  $ Peak Position   : num [1:49] 1 1 1 7 1 8 1 2 1 5 ...
##  $ Peak Position (xTimes): chr [1:49] "(x29)" "(x20)" "(x4)" "0" ...
##  $ Peak Streams    : num [1:49] 2118242 2127668 1660502 659366 2905678 ...
##  $ Total Streams   : num [1:49] 8.83e+08 8.65e+08 7.81e+08 7.35e+08 7.19e+08 ...
```

```
names(Spotify)
```

```
## [1] "Position"      "Artist Name"    "Song Name"
## [4] "Days"          "Top 10 (xTimes)" "Peak Position"
## [7] "Peak Position (xTimes)" "Peak Streams"    "Total Streams"
```

```
head(Spotify,n=15)
```

```
## # A tibble: 15 x 9
##   Position 'Artist Name' Song ~1 Days Top 1~2 Peak ~3 Peak ~4 Peak ~5 Total~6
##   <dbl> <chr>          <chr>   <dbl>   <dbl>   <dbl> <chr>   <dbl>   <dbl>
```

```
## 1      1 abc      Sunflo~ 1506      302      1 (x29) 2118242 8.83e8
## 2      2 Juice WRLD <NA>      1673      178      1 (x20) 2127668 8.65e8
## 3      3 Lil Uzi Vert XO TOU~ 1853      212      1 (x4) 1660502 7.81e8
## 4      4 J. Cole      No Rol~ 2547      6      7 0      659366 7.35e8
## 5      5 Post Malone rockst~ 1223      186      1 (x124) 2905678 7.19e8
## 6      6 Travis Scott gooseb~ 1995      4      8 0      977275 6.73e8
## 7      7 The Weeknd Blindi~ 1100      233      1 (x11) 2355059 6.44e8
## 8      8 XXXTENTACION Jocely~ 1673      44      2 (x1) 3175206 6.24e8
## 9      9 XXXTENTACION SAD!      1217      133      1 (x6) 4437612 6.20e8
## 10     10 Juice WRLD All Gi~ 1681      2      5 0      1239152 6.14e8
## 11     11 Kendrick Lamar HUMBLE. 1175      152      1 (x67) 4060034 6.06e8
## 12     12 Post Malone Circles 1172      163      1 (x38) 3441328 5.99e8
## 13     13 Travis Scott SICKO ~ 1046      190      1 (x19) 4129691 5.87e8
## 14     14 Lil Baby      Drip T~ 1453      122      1 (x6) 1759294 5.83e8
## 15     15 Post Malone Congra~ 1215      136      5 0      1263208 5.46e8
## # ... with abbreviated variable names 1: 'Song Name', 2: 'Top 10 (xTimes)',
## # 3: 'Peak Position', 4: 'Peak Position (xTimes)', 5: 'Peak Streams',
## # 6: 'Total Streams'
```

```
myfunc<- function(x)
{
print(Spotify[ , c("Artist Name", "Song Name")],n=5)
}
myfunc()
```

```
## # A tibble: 49 x 2
##   'Artist Name' 'Song Name'
##   <chr>         <chr>
## 1 abc          Sunflower SpiderMan: Into the SpiderVerse
## 2 Juice WRLD    <NA>
## 3 Lil Uzi Vert XO TOUR Llif3
## 4 J. Cole       No Role Modelz
## 5 Post Malone   rockstar
## # ... with 44 more rows
```

```
Samplenew = as.data.frame(filter(Spotify,Spotify$"Peak Position" == 5))
print(Samplenew)
```

```
##   Position  Artist Name      Song Name Days Top 10 (xTimes)
## 1      10    Juice WRLD All Girls Are The Same 1681      2
## 2      15    Post Malone Congratulations 1215      136
## 3      21 XXXTENTACION      Fuck Love 1406      9
## 4      31 Lewis Capaldi      Someone You Loved 952      40
##   Peak Position Peak Position (xTimes) Peak Streams Total Streams
## 1           5           0      1239152      613872384
## 2           5           0      1263208      546036924
## 3           5           0      2438715      473417295
## 4           5           0      1118953      410725370
```

```
print(cbind(Spotify$"Song Name",Spotify$"Peak Position"))
```

```
##           [,1]           [,2]
```

```
## [1,] "Sunflower SpiderMan: Into the SpiderVerse" "1"
## [2,] NA "1"
## [3,] "XO TOUR Llif3" "1"
## [4,] "No Role Modelz" "7"
## [5,] "rockstar" "1"
## [6,] "goosebumps" "8"
## [7,] "Blinding Lights" "1"
## [8,] "Jocelyn Flores" "2"
## [9,] "SAD!" "1"
## [10,] "All Girls Are The Same" "5"
## [11,] "HUMBLE." "1"
## [12,] "Circles" "1"
## [13,] "SICKO MODE" "1"
## [14,] "Drip Too Hard (Lil Baby & Gunna)" "1"
## [15,] "Congratulations" "5"
## [16,] "I Fall Apart" "2"
## [17,] "Heat Waves" "1"
## [18,] "God's Plan" "1"
## [19,] "The Box" "1"
## [20,] "MIDDLE CHILD" "1"
## [21,] "Fuck Love" "5"
## [22,] "Better Now" "1"
## [23,] "One Dance" "1"
## [24,] "good 4 u" "1"
## [25,] "Robbery" "2"
## [26,] "lovely" "18"
## [27,] "SLOW DANCING IN THE DARK" "26"
## [28,] "Moonlight" "2"
## [29,] "Ric Flair Drip (& Metro Boomin)" "4"
## [30,] "Stay" "1"
## [31,] "Someone You Loved" "5"
## [32,] "Watermelon Sugar" "3"
## [33,] "Wow." "1"
## [34,] "Location" "9"
## [35,] "Sweater Weather" "12"
## [36,] "Yes Indeed" "2"
## [37,] "Closer" "1"
## [38,] "Psycho" "1"
## [39,] "The Hills" "1"
## [40,] "Levitating" "3"
## [41,] "bad guy" "1"
## [42,] "Whiskey Glasses" "33"
## [43,] "drivers license" "1"
## [44,] "Kiss Me More" "1"
## [45,] "ROCKSTAR" "1"
## [46,] "INDUSTRY BABY" "1"
## [47,] "Starboy" "1"
## [48,] "Redbone" "8"
## [49,] "Blueberry Faygo" "2"
```

```
na.omit(Spotify)
```

```
## # A tibble: 48 x 9
```

```
##   Position 'Artist Name' Song ~1 Days Top 1~2 Peak ~3 Peak ~4 Peak ~5 Total~6
```

```
##      <dbl> <chr>          <chr> <dbl> <dbl> <dbl> <chr>      <dbl> <dbl>
## 1      1 abc             Sunflo~ 1506   302      1 (x29)  2118242  8.83e8
## 2      3 Lil Uzi Vert    XO TOU~ 1853   212      1 (x4)   1660502  7.81e8
## 3      4 J. Cole        No Rol~ 2547    6       7 0      659366   7.35e8
## 4      5 Post Malone    rockst~ 1223   186      1 (x124) 2905678  7.19e8
## 5      6 Travis Scott   gooseb~ 1995    4       8 0      977275   6.73e8
## 6      7 The Weeknd     Blindi~ 1100   233      1 (x11)  2355059  6.44e8
## 7      8 XXXTENTACION   Jocely~ 1673    44       2 (x1)   3175206  6.24e8
## 8      9 XXXTENTACION   SAD!    1217   133      1 (x6)   4437612  6.20e8
## 9     10 Juice WRLD     All Gi~ 1681    2       5 0      1239152  6.14e8
## 10     11 Kendrick Lamar HUMBLe. 1175   152      1 (x67)  4060034  6.06e8
## # ... with 38 more rows, and abbreviated variable names 1: 'Song Name',
## # 2: 'Top 10 (xTimes)', 3: 'Peak Position', 4: 'Peak Position (xTimes)',
## # 5: 'Peak Streams', 6: 'Total Streams'
```

```
Spotify$"Artist Name"[duplicated(Spotify$"Artist Name")]
```

```
## [1] "XXXTENTACION" "Juice WRLD" "Post Malone" "Travis Scott"
## [5] "Post Malone" "Post Malone" "J. Cole" "XXXTENTACION"
## [9] "Post Malone" "Drake" "Juice WRLD" "XXXTENTACION"
## [13] "Post Malone" "Lil Baby" "Post Malone" "The Weeknd"
## [17] "Billie Eilish" "Olivia Rodrigo" "The Weeknd"
```

```
print(Spotify%>%distinct(Spotify$"Artist Name", .keep_all = TRUE))
```

```
## # A tibble: 30 x 10
##   Posit-1 Artis-2 Song ~3 Days Top 1~4 Peak ~5 Peak ~6 Peak ~7 Total~8 Spoti-9
##   <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <chr>
## 1      1 abc Sunflo~ 1506   302      1 (x29)  2118242  8.83e8 abc
## 2      2 Juice ~ <NA> 1673   178      1 (x20)  2127668  8.65e8 Juice ~
## 3      3 Lil Uz~ XO TOU~ 1853   212      1 (x4)   1660502  7.81e8 Lil Uz~
## 4      4 J. Cole No Rol~ 2547    6       7 0      659366   7.35e8 J. Cole
## 5      5 Post M~ rockst~ 1223   186      1 (x124) 2905678  7.19e8 Post M~
## 6      6 Travis~ gooseb~ 1995    4       8 0      977275   6.73e8 Travis~
## 7      7 The We~ Blindi~ 1100   233      1 (x11)  2355059  6.44e8 The We~
## 8      8 XXXTEN~ Jocely~ 1673    44       2 (x1)   3175206  6.24e8 XXXTEN~
## 9      11 Kendri~ HUMBLe. 1175   152      1 (x67)  4060034  6.06e8 Kendri~
## 10     14 Lil Ba~ Drip T~ 1453   122      1 (x6)   1759294  5.83e8 Lil Ba~
## # ... with 20 more rows, and abbreviated variable names 1: Position,
## # 2: 'Artist Name', 3: 'Song Name', 4: 'Top 10 (xTimes)', 5: 'Peak Position',
## # 6: 'Peak Position (xTimes)', 7: 'Peak Streams', 8: 'Total Streams',
## # 9: 'Spotify$Artist Name'
```

```
Spotify%>%arrange(desc(Spotify$"Position"))
```

```
## # A tibble: 49 x 9
##   Position 'Artist Name' Song ~1 Days Top 1~2 Peak ~3 Peak ~4 Peak ~5 Total~6
##   <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <dbl>
## 1      49 Lil Mosey Bluebe~ 603   152      2 (x12)  1428849  3.55e8
## 2      48 Childish Gamb~ Redbone 1075   42      8 0      950460   3.64e8
## 3      47 The Weeknd Starboy 785   141      1 (x29)  2070596  3.69e8
## 4      46 Lil Nas X INDUST~ 472   267      1 (x42)  2492612  3.73e8
```

```
## 5      45 DaBaby          ROCKST~  549    164      1 (x65)  1975668  3.73e8
## 6      44 Doja Cat        Kiss M~  550    151      1 (x11)  1846647  3.77e8
## 7      43 Olivia Rodrigo driver~  613     99      1 (x46)  6145601  3.81e8
## 8      42 Morgan Wallen  Whiske~ 1324     0     33 0      447546  3.85e8
## 9      41 Billie Eilish  bad guy  652    108      1 (x20)  2931099  3.86e8
## 10     40 Dua Lipa        Levita~  738    113      3 (x3)   1284080  3.88e8
## # ... with 39 more rows, and abbreviated variable names 1: 'Song Name',
## # 2: 'Top 10 (xTimes)', 3: 'Peak Position', 4: 'Peak Position (xTimes)',
## # 5: 'Peak Streams', 6: 'Total Streams'
```

```
Spotify%>%arrange(desc(Days))
```

```
## # A tibble: 49 x 9
##   Position 'Artist Name' Song N~1 Days Top 1~2 Peak ~3 Peak ~4 Peak ~5 Total~6
##   <dbl> <chr>          <chr> <dbl> <dbl> <dbl> <chr> <dbl> <dbl>
## 1      4 J. Cole        No Role~ 2547     6     7 0      659366  7.35e8
## 2      6 Travis Scott  goosebu~ 1995     4     8 0      977275  6.73e8
## 3      3 Lil Uzi Vert  XO TOUR~ 1853    212     1 (x4)  1660502  7.81e8
## 4     10 Juice WRLD    All Gir~ 1681     2     5 0      1239152  6.14e8
## 5      2 Juice WRLD    <NA>     1673    178     1 (x20)  2127668  8.65e8
## 6      8 XXXTENTACION Jocelyn~ 1673     44     2 (x1)  3175206  6.24e8
## 7      1 abc           Sunflow~ 1506    302     1 (x29)  2118242  8.83e8
## 8     27 Joji          SLOW DA~ 1496     0    26 0      661274  4.37e8
## 9     14 Lil Baby      Drip To~ 1453    122     1 (x6)  1759294  5.83e8
## 10     21 XXXTENTACION Fuck Lo~ 1406     9     5 0      2438715  4.73e8
## # ... with 39 more rows, and abbreviated variable names 1: 'Song Name',
## # 2: 'Top 10 (xTimes)', 3: 'Peak Position', 4: 'Peak Position (xTimes)',
## # 5: 'Peak Streams', 6: 'Total Streams'
```

```
Spotify%>%rename(Years = Days, ID = Position)
```

```
## # A tibble: 49 x 9
##   ID 'Artist Name' 'Song Name' Years Top 1~1 Peak ~2 Peak ~3 Peak ~4 Total~5
##   <dbl> <chr>          <chr> <dbl> <dbl> <dbl> <chr> <dbl> <dbl>
## 1      1 abc           Sunflower ~ 1506    302     1 (x29)  2118242  8.83e8
## 2      2 Juice WRLD    <NA>     1673    178     1 (x20)  2127668  8.65e8
## 3      3 Lil Uzi Vert  XO TOUR Ll~ 1853    212     1 (x4)  1660502  7.81e8
## 4      4 J. Cole        No Role Mo~ 2547     6     7 0      659366  7.35e8
## 5      5 Post Malone   rockstar   1223    186     1 (x124)  2905678  7.19e8
## 6      6 Travis Scott  goosebumps 1995     4     8 0      977275  6.73e8
## 7      7 The Weeknd     Blinding L~ 1100    233     1 (x11)  2355059  6.44e8
## 8      8 XXXTENTACION  Jocelyn Fl~ 1673     44     2 (x1)  3175206  6.24e8
## 9      9 XXXTENTACION SAD!       1217    133     1 (x6)  4437612  6.20e8
## 10     10 Juice WRLD    All Girls ~ 1681     2     5 0      1239152  6.14e8
## # ... with 39 more rows, and abbreviated variable names 1: 'Top 10 (xTimes)',
## # 2: 'Peak Position', 3: 'Peak Position (xTimes)', 4: 'Peak Streams',
## # 5: 'Total Streams'
```

```
Spotify%>%mutate(Dividend = Days*2)
```

```
## # A tibble: 49 x 10
##   Posit~1 Artis~2 Song ~3 Days Top 1~4 Peak ~5 Peak ~6 Peak ~7 Total~8 Divid~9
```

```
##      <dbl> <chr>   <chr>   <dbl>   <dbl>   <dbl> <chr>       <dbl>   <dbl>   <dbl>
## 1      1      abc      Sunflo~ 1506     302      1 (x29)    2118242  8.83e8   3012
## 2      2      Juice ~ <NA>    1673     178      1 (x20)    2127668  8.65e8   3346
## 3      3      Lil Uz~ XO TOU~ 1853     212      1 (x4)     1660502  7.81e8   3706
## 4      4      J. Cole No Rol~ 2547       6      7 0        659366  7.35e8   5094
## 5      5      Post M~ rockst~ 1223     186      1 (x124)   2905678  7.19e8   2446
## 6      6      Travis~ gooseb~ 1995       4      8 0        977275  6.73e8   3990
## 7      7      The We~ Blindi~ 1100     233      1 (x11)    2355059  6.44e8   2200
## 8      8      XXXTEN~ Jocely~ 1673      44      2 (x1)     3175206  6.24e8   3346
## 9      9      XXXTEN~ SAD!    1217     133      1 (x6)     4437612  6.20e8   2434
## 10     10     Juice ~ All Gi~ 1681       2      5 0        1239152  6.14e8   3362
## # ... with 39 more rows, and abbreviated variable names 1: Position,
## # 2: 'Artist Name', 3: 'Song Name', 4: 'Top 10 (xTimes)', 5: 'Peak Position',
## # 6: 'Peak Position (xTimes)', 7: 'Peak Streams', 8: 'Total Streams',
## # 9: Dividend
```

```
set.seed(1234)
Spotify%>%sample_n(5,replace = FALSE)
```

```
## # A tibble: 5 x 9
##   Position 'Artist Name'   Song ~1 Days Top 1~2 Peak ~3 Peak ~4 Peak ~5 Total~6
##     <dbl> <chr>         <chr>   <dbl>   <dbl>   <dbl> <chr>       <dbl>   <dbl>
## 1      28 XXXTENTACION   Moonli~  923      57      2 (x2)     3167290  4.32e8
## 2      16 Post Malone     I Fall~ 1278     132     2 (x14)   1275320  5.43e8
## 3      22 Post Malone     Better~  862      72      1 (x8)     3467168  4.64e8
## 4      37 The Chainsmoke~ Closer   696     158     1 (x70)   2504045  3.98e8
## 5      44 Doja Cat        Kiss M~  550     151     1 (x11)   1846647  3.77e8
## # ... with abbreviated variable names 1: 'Song Name', 2: 'Top 10 (xTimes)',
## # 3: 'Peak Position', 4: 'Peak Position (xTimes)', 5: 'Peak Streams',
## # 6: 'Total Streams'
```

```
summary(Spotify)
```

```
##      Position Artist Name      Song Name      Days
## Min.   : 1   Length:49      Length:49      Min.   : 472
## 1st Qu.:13   Class :character Class :character 1st Qu.: 699
## Median :25   Mode  :character Mode  :character Median :1050
## Mean   :25
## 3rd Qu.:37
## Max.   :49
## Top 10 (xTimes) Peak Position Peak Position (xTimes) Peak Streams
## Min.   : 0   Min.   : 1.000 Length:49      Min.   : 447546
## 1st Qu.: 43   1st Qu.: 1.000 Class :character 1st Qu.:1239152
## Median :122   Median : 1.000 Mode  :character Median :2070596
## Mean   :118   Mean   : 3.898      Mean   :2274634
## 3rd Qu.:178   3rd Qu.: 4.000      3rd Qu.:3167290
## Max.   :302   Max.   :33.000      Max.   :6145601
## Total Streams
## Min.   :355016723
## 1st Qu.:397735618
## Median :443773199
## Mean   :500062960
## 3rd Qu.:586638599
## Max.   :883369738
```

```
mean(Spotify$"Days")
```

```
## [1] 1070.939
```

```
median(Spotify$"Days")
```

```
## [1] 1050
```

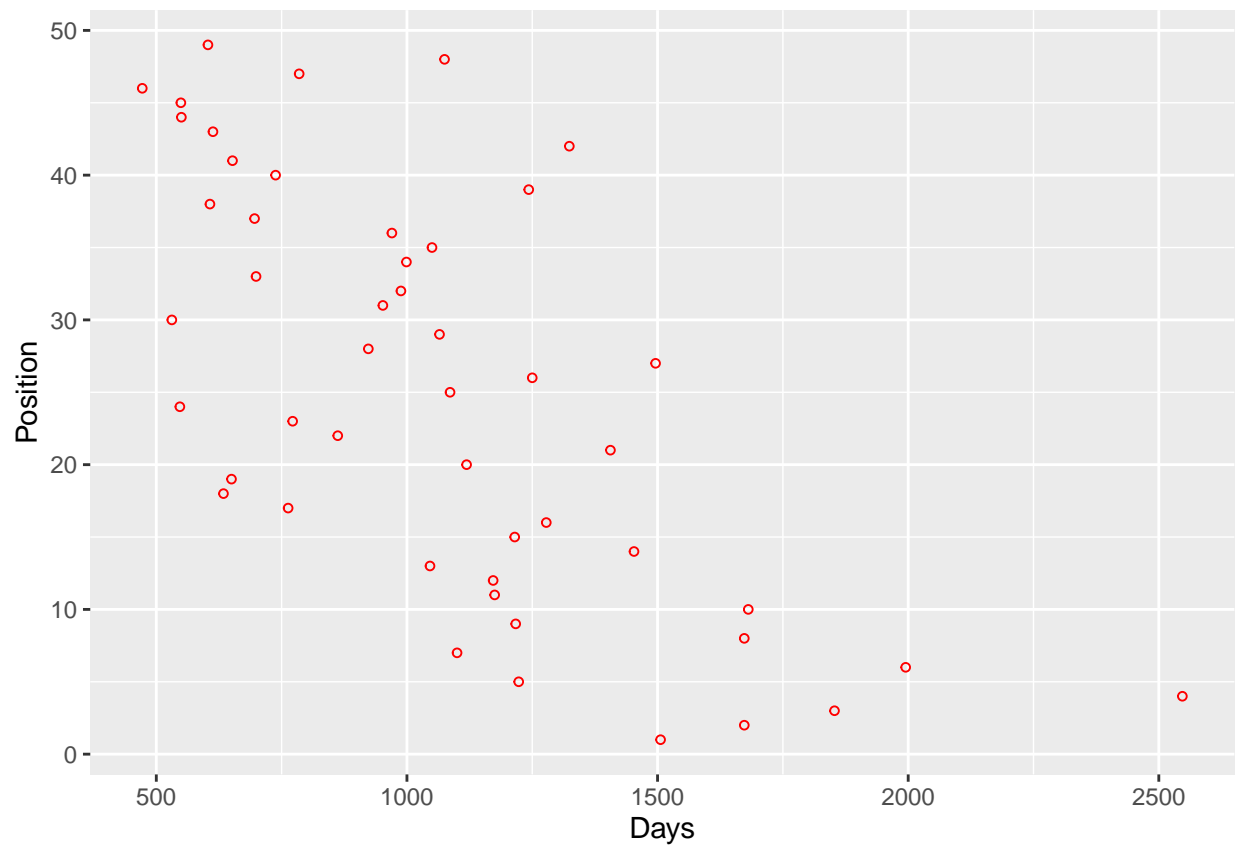
```
mode(Spotify$"Days")
```

```
## [1] "numeric"
```

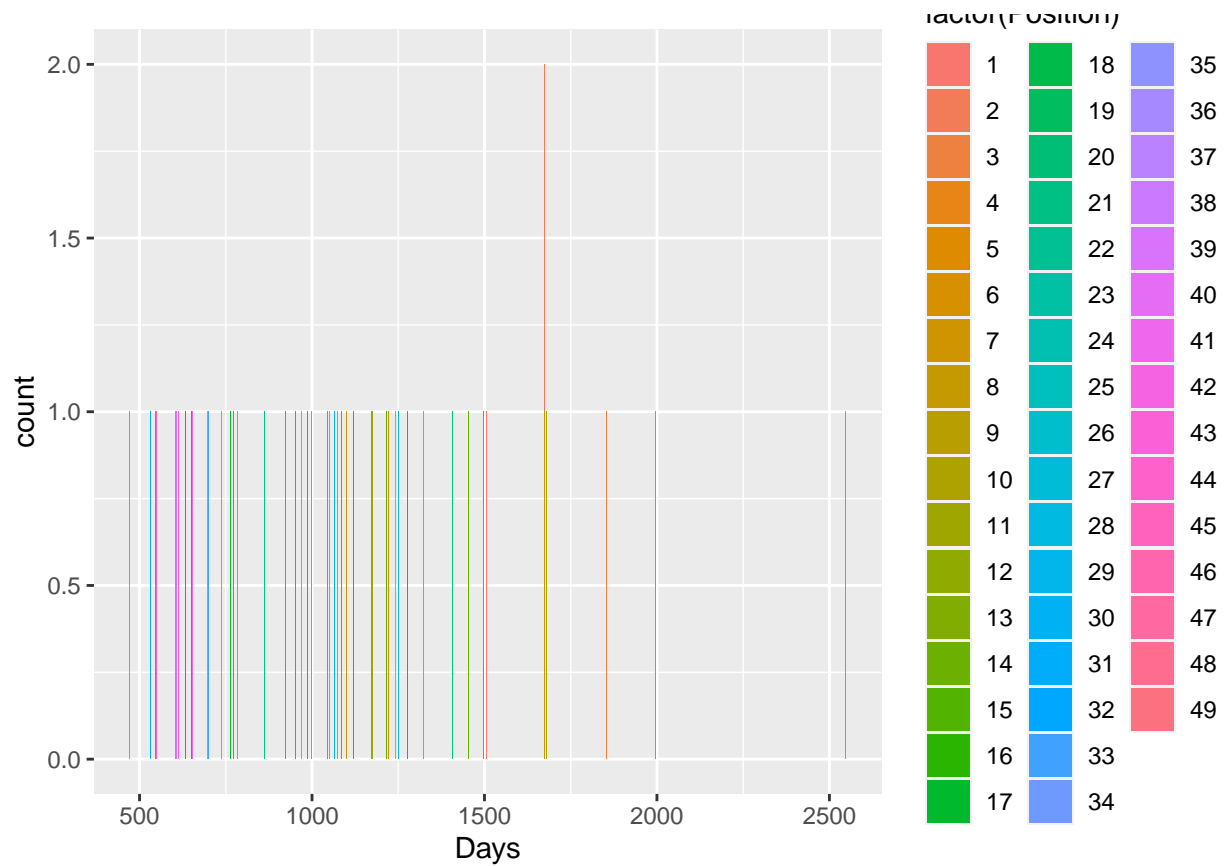
```
range(Spotify$"Days")
```

```
## [1] 472 2547
```

```
ggplot(Spotify, aes(x = Days, y = Position)) + geom_point(size=1.2, color = "Red", shape=21)
```



```
ggplot(data = Spotify, aes(x = Days, fill = factor(Position)))+geom_bar()
```



```
x = Spotify$Days
y = Spotify$Position
corrnew = cor(x,y,method = 'pearson')
corrnew
```

```
## [1] -0.6704169
```