Nombre: Celic Gabriel Hernández Archundia

Matrícula: 2877240

```
%%capture
!pip install rpy2==3.5.1
```

```
%load_ext rpy2.ipython
```

```
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
```

```
Mounted at /content/drive
```

```
%%R
library(readr)
library(magrittr)
library(dplyr)
```

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]:
Attaching package: ‘dplyr’


WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: The following objects are mas

    filter, lag


WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: The following objects are mas

    intersect, setdiff, setequal, union
```

```
%%R
datos <- read.csv("drive/MyDrive/Tecmilenio/Big Data/movies.csv")
```

9. Top 10 de las compañías con más películas producidas

```
%%R
colnames(datos)
```

```
 [1] "name"     "rating"   "genre"    "year"     "released" "score"
 [7] "votes"    "director" "writer"   "star"     "country"  "budget"
[13] "gross"    "company"  "runtime"
```

```
%%R
datos %>%
group_by(company) %>%
summarize(películas_producidas = n()) %>%
arrange(desc(películas_producidas)) %>%
print(n=10)
```

```
# A tibble: 2,386 × 2
   company                   películas_producidas
   <chr>                                    <int>
 1 Universal Pictures                         377
 2 Warner Bros.                               334
 3 Columbia Pictures                          332
 4 Paramount Pictures                         320
 5 Twentieth Century Fox                      240
 6 New Line Cinema                            174
 7 Touchstone Pictures                        132
 8 Metro-Goldwyn-Mayer (MGM)                  125
 9 Walt Disney Pictures                       123
10 TriStar Pictures                            94
# ℹ 2,376 more rows
# ℹ Use `print(n = ...)` to see more rows
```

10. Película con el menor y película la mayor duración

```
%%R
datos %>%
select(name, runtime) %>%
arrange(desc(runtime)) %>%
head(n=1) -> mayorDuracion

datos %>%
select(name, runtime) %>%
filter(!is.na(runtime)) %>%
arrange(desc(runtime)) %>%
tail(n=1) -> menorDuracion

c(mayorDuracion,menorDuracion)
```

```
$name
[1] "The Best of Youth"

$runtime
[1] 366

$name
```

```
[1] "The Business of Show Business"

$runtime
[1] 55
```

11. Top 3 de escritores con más trabajos en el siglo XX

```
%%R
colnames(datos)
```

```
 [1] "name"     "rating"   "genre"    "year"     "released" "score"
 [7] "votes"    "director" "writer"   "star"     "country"  "budget"
[13] "gross"    "company"  "runtime"
```

```
%%R
datos %>%
filter(year <= 2000) %>%
group_by(writer) %>%
summarize(trabajos = n()) %>%
arrange(desc(trabajos)) %>%
head(3)
```

```
# A tibble: 3 × 2
  writer       trabajos
  <chr>           <int>
1 Stephen King       26
2 John Hughes        24
3 Woody Allen        21
```

12. Top 3 de directores con más trabajos en tu año de nacimiento

```
%%R
datos %>%
filter(year == 2002) %>%
group_by(director) %>%
summarize(trabajos = n()) %>%
arrange(desc(trabajos)) %>%
head(3)
```

```
# A tibble: 3 × 2
  director          trabajos
  <chr>                <int>
1 Barry Sonnenfeld         2
2 Charles Stone III        2
3 Joel Schumacher          2
```

13. Top 10 de los mejores géneros de películas basados en su Score

```
%%R
datos %>%
group_by(genre) %>%
summarize(puntuacion = mean(score)) %>%
arrange(desc(puntuacion)) %>%
print(n=10)
```

```
# A tibble: 19 × 2
   genre     puntuacion
   <chr>          <dbl>
 1 History         8.3
 2 Musical         8.05
 3 Music           7.2
 4 Biography       7.03
 5 Animation       6.77
 6 Crime           6.67
 7 Mystery         6.66
 8 Romance         6.41
 9 Family          6.36
10 Adventure       6.29
# ℹ 9 more rows
# ℹ Use `print(n = ...)` to see more rows
```

14. ¿Cuántas películas se estrenaron en tu fecha de cumpleaños?

```
%%R
datos %>%
filter(grepl("October 2", released)) %>%
count()
```

```
      n
1 264
```

15. Cantidad de películas donde el escritor y el director son la misma persona

```
%%R
datos %>%
filter(writer == director) %>%
count()
```

```
      n
1 2164
```

16. Cantidad de películas donde el protagonista y el director son la misma persona

```
%%R
datos %>%
filter(star == director) %>%
count()
```

```
  n
1 205
```

17. Top 10 de directores que han sido escritores durante más películas

```
%%R
datos %>%
filter(director == writer) %>%
group_by(director) %>%
summarize(peliculas = n()) %>%
arrange(desc(peliculas)) %>%
print(n=10)
```

```
# A tibble: 1,233 × 2
   director           peliculas
   <chr>                  <int>
 1 Woody Allen               37
 2 Pedro Almodóvar           13
 3 Jim Jarmusch              11
 4 Lars von Trier            11
 5 M. Night Shyamalan        11
 6 Kevin Smith               10
 7 Richard Linklater         10
 8 Tyler Perry               10
 9 Luc Besson                 9
10 Quentin Tarantino          9
# ℹ 1,223 more rows
# ℹ Use `print(n = ...)` to see more rows
```

18. Actor con más películas protagonizadas de tu género favorito

```
%%R
datos %>%
group_by(genre) %>%
summarize(cantidad = n()) %>%
print(n=10)
```

```
# A tibble: 19 × 2
   genre     cantidad
   <chr>        <int>
 1 Action        1705
 2 Adventure      427
 3 Animation      338
 4 Biography      443
 5 Comedy        2245
 6 Crime          551
 7 Drama         1518
 8 Family          11
 9 Fantasy         44
10 History          1
# ℹ 9 more rows
# ℹ Use `print(n = ...)` to see more rows
```

```
%%R
datos %>%
filter(genre == "Action") %>%
group_by(star) %>%
summarize(peliculas = n()) %>%
arrange(desc(peliculas)) %>%
print(n=3)
```

```
# A tibble: 757 × 2
  star                  peliculas
  <chr>                     <int>
1 Sylvester Stallone           25
2 Arnold Schwarzenegger        23
3 Jackie Chan                  22
# ℹ 754 more rows
# ℹ Use `print(n = ...)` to see more rows
```