

▼ Evidencia 2 (Avance Final)

Integrantes del Equipo:

-Celic Gabriel Hernández Archundia 2877240

-Diego Alfredo López Malerva 2999206

-Ivan Gutierrez Gomez 2877087

-Diego Iván García Zarza 2897709

```
%%capture
```

```
!pip install rpy2==3.5.1
```

```
%load_ext rpy2.ipython
```

```
from google.colab import drive  
drive.mount('/content/drive')
```

```
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 masked from 'package:base':  
  filter, lag
```

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: The following objects are masked from 'package:stats':  
  intersect, setdiff, setequal, union
```

%%R

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

1. La columna "salary_in_usd" muestra el salario anual en dólares de cada puesto de trabajo. Si un dolar equivale a 17.81 pesos mexicanos, elabore una nueva columna que muestre la conversión a salario mensual en pesos y muestre el top 10 de los títulos profesionales que tienen el mejor sueldo en ciencia de datos.

%%R

```
datos %>%
```

```
select(job_title, salary_in_usd) %>%
```

```
mutate(salario_mx = (17.81*salary_in_usd)/12) %>%
```

```
select(job_title, salario_mx) %>%
```

```
arrange( desc(salario_mx)) %>%
```

```
head(n= 10)
```

	job_title	salario_mx
1	Research Scientist	667875.0
2	Data Analyst	639626.9
3	AI Scientist	629040.3
4	Applied Machine Learning Scientist	627802.5
5	Principal Data Scientist	617413.3
6	Data Scientist	611476.7
7	Data Analytics Lead	601087.5
8	Data Analyst	571404.2
9	Applied Data Scientist	563983.3
10	Data Architect	558165.4

2. Suponga que desea aplicar este año para un puesto de los presentados en este conjunto de datos. Si un nivel de experiencia es "EN" y quiere obtener el sueldo más alto posible, ¿Qué título le conviene más obtener? Obtenga el salario promedio de cada puesto cuyo año sea el año actual para encontrar la respuesta y muéstrela junto con su salario mensual en pesos mexicanos.

%%R

```
datos %>%
```

```
select(work_year, experience_level, job_title, salary_in_usd, company_size) %>%
```

```
filter(work_year == 2023)%>%
```

```
filter(experience_level == "EN") %>%
```

```
group_by(job_title) %>%
```

```
summarize(salario.en.pesos = mean((salary_in_usd)*17.81)/12) %>%
```

```

arrange(desc(salario.en.pesos)) %>%
print(n= 10)

# A tibble: 19 × 2
  job_title          salario.en.pesos
  <chr>              <dbl>
1 Computer Vision Engineer 326517.
2 Applied Scientist      248385.
3 Machine Learning Engineer 229364.
4 Research Scientist      213720
5 Deep Learning Engineer  200362.
6 BI Developer           192942.
7 Big Data Engineer       192942.
8 Research Engineer       192942.
9 AI Developer            177730.
10 Data Quality Analyst    148417.
# i 9 more rows
# i Use `print(n = ...)` to see more rows

```

Le conviene más ser Computer Vision Engineer

3. Existen registro de trabajadores que radican en México. Averigüe qué título profesional, nivel de experiencia y salario mensual en pesos mexicanos tienen y ordene los registros de mayor a menor por su salario.

Nota: La columna para saber en donde radica un trabajador es "employee_residence"

```

%%R
datos %>%
filter(employee_residence == "MX") %>%
select(job_title, experience_level, salary_in_usd) %>%
mutate(salary_in_mx_mensual = ((salary_in_usd)*17.81)/12) %>%
select(-salary_in_usd) %>%
arrange(desc(salary_in_mx_mensual)) %>%
head(10)

```

	job_title	experience_level	salary_in_mx_mensual
1	Machine Learning Engineer	SE	378462.50
2	Machine Learning Engineer	SE	274570.83
3	Data Science Manager	SE	222625.00
4	Data Science Manager	SE	133575.00
5	Data Engineer	MI	106860.00
6	Data Engineer	MI	89050.00
7	Machine Learning Engineer	MI	53430.00
8	Data Engineer	SE	49735.91
9	Data Scientist	MI	44525.00
10	BI Analyst	EN	17810.00

4. Para alguien con experiencia "EN" que desea buscar un puesto de tiempo completo y ganar lo más posible, ¿le conviene trabajar de manera remota o presencial?

```
%%R
datos %>%
select(experience_level, employment_type, remote_ratio, salary_in_usd) %>%
filter(experience_level == "EN" & employment_type == "FT") %>%
group_by(remote_ratio) %>%
summarize(salario = mean(salary_in_usd))

# A tibble: 3 × 2
  remote_ratio salario
    <int>     <dbl>
1         0 104582.
2        50  65344
3       100  67016.
```

Es más conveniente trabajar de forma presencial que remota si es que el objetivo es ganar más dinero, tomando en cuenta los requerimientos mencionados en la pregunta

5. Cantidad de registros donde el país de residencia de la compañía y del trabajador son diferentes

```
%%R
datos %>%
filter(employee_residence != company_location)%>%
count()

      n
1  96
```

6. Salario mensual promedio en pesos mexicanos para un trabajador con experiencia "EN" en una empresa de tamaño. S, L y M de acuerdo con los datos del año 2022

```
%%R
datos %>%
select(work_year, experience_level, salary_in_usd, company_size) %>%
```

```
filter(work_year == 2022)%>%
filter(experience_level == "EN") %>%
group_by(company_size) %>%
summarize(salario.en.pesos = mean((salary_in_usd)*17.81)/12)
```

```
# A tibble: 3 × 2
  company_size salario.en.pesos
  <chr>         <dbl>
1 L             105853.
2 M             127823.
3 S              83360.
```

7. Top 10 Cantidad de trabajadores de cada país registrado que trabajan en una compañía ubicada en Estados Unidos (ordenados de mayor a menor)

```
%%R
datos %>%
filter(company_location == "US")%>%
group_by(employee_residence)%>%
count()%>%
arrange(desc(n))%>%
head(10)
```

```
# A tibble: 10 × 2
# Groups:   employee_residence [10]
  employee_residence    n
  <chr>              <int>
1 US                 2999
2 IN                   7
3 BR                   3
4 ES                   3
5 AR                   2
6 BO                   2
7 PT                   2
8 RU                   2
9 BE                   1
10 BG                  1
```

8. ¿Cuál es la equivalencia en promedio de cada moneda utilizada para la conversión en dólares?

```
%%R
datos %>%
group_by(salary_currency) %>%
summarize(prom = mean(salary), prom_usd = mean(salary_in_usd)) %>%
summarize(curr = salary_currency, promtot = prom / prom_usd)
```

```
# A tibble: 20 × 2
  curr  promtot
  <chr>   <dbl>
1 AUD     1.43
2 BRL     5.27
3 CAD     1.29
4 CHF     0.946
5 CLP    759.
6 CZK     23.4
7 DKK      6.41
8 EUR     0.918
9 GBP     0.809
10 HKD     7.84
11 HUF    319.
12 ILS      3.54
13 INR     77.0
14 JPY    109.
15 MXN     21.5
16 PLN      4.19
17 SGD      1.35
18 THB     34.0
19 TRY      8.92
20 USD      1
```

✓ 0 s se ejecutó 18:40

