```
Mounted at /content/gdrive
/content
```

Reporte final de "Los salarios"

Inteligencia Artificial Avanzada para la ciencia de datos (grupo 101)

Diego Solis Higuera

A00827847

18 de septiembre de 2022

```
/content/gdrive/MyDrive/Colab Notebooks/SEMESTRE_7/Modulo1/Salarios
ds_salaries.csv  Evidencia1.ipynb
```

Column	Description
work_year	The year the salary was paid.
experience_level	The experience level in the job during the year with the following possible values: EN Entry-level / Junior MI Mid-level / Intermediate SE Senior-level / Expert EX Executive-level / Director
employment_type	The type of employement for the role: PT Part-time FT Full-time CT Contract FL Freelance
job_title	The role worked in during the year.
salary	The total gross salary amount paid.
salary_currency	The currency of the salary paid as an ISO 4217 currency code.
salary*in*usd	The salary in USD (FX rate divided by avg. USD rate for the respective year via fxdata.foorilla.com).
employee_residence	Employee's primary country of residence in during the work year as an ISO 3166 country code.
remote_ratio	The overall amount of work done remotely, possible values are as follows: 0 No remote work (less than 20%) 50 Partially remote 100 Fully remote (more than 80%)
company_location	The country of the employer's main office or contracting branch as an ISO 3166 country code.
company_size	The average number of people that worked for the company during the year: S less than 50 employees (small) M 50 to 250 employees (medium) L more than 250 employees (large)

	id	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd
0	0	2020	MI	FT	Data Scientist	70000	EUR	79833
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024
3	3	2020	MI	FT	Product Data Analyst	20000	USD	20000
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000

Exploración de variables

Identificación de cantidad de datos y variables

```
id                  0
work_year           0
experience_level    0
employment_type     0
job_title           0
salary              0
salary_currency     0
salary_in_usd       0
employee_residence  0
remote_ratio        0
company_location    0
company_size        0
dtype: int64

(607, 12)
```

```
id                   int64
work_year            int64
experience_level    object
employment_type     object
job_title           object
salary               int64
salary_currency     object
salary_in_usd        int64
employee_residence  object
remote_ratio         int64
company_location    object
company_size        object
dtype: object
```

Exploración de la base de datos

Medidas estadísticas

Variables cuantitativas

Datos estadísticos generales

	work_year	salary	salary_in_usd	remote_ratio
count	607.000000	6.070000e+02	607.000000	607.00000
mean	2021.405272	3.240001e+05	112297.869852	70.92257
std	0.692133	1.544357e+06	70957.259411	40.70913
min	2020.000000	4.000000e+03	2859.000000	0.00000
25%	2021.000000	7.000000e+04	62726.000000	50.00000
50%	2022.000000	1.150000e+05	101570.000000	100.00000
75%	2022.000000	1.650000e+05	150000.000000	100.00000
max	2022.000000	3.040000e+07	600000.000000	100.00000

Moda de variables cuantitativas

```
work_year
 0    2022
dtype: int64
 ~~~~~~~

salary
 0     80000
1    100000
dtype: int64
 ~~~~~~~

salary_in_usd
 0    100000
dtype: int64
 ~~~~~~~

remote_ratio
 0    100
dtype: int64
 ~~~~~~~
```

Varianza

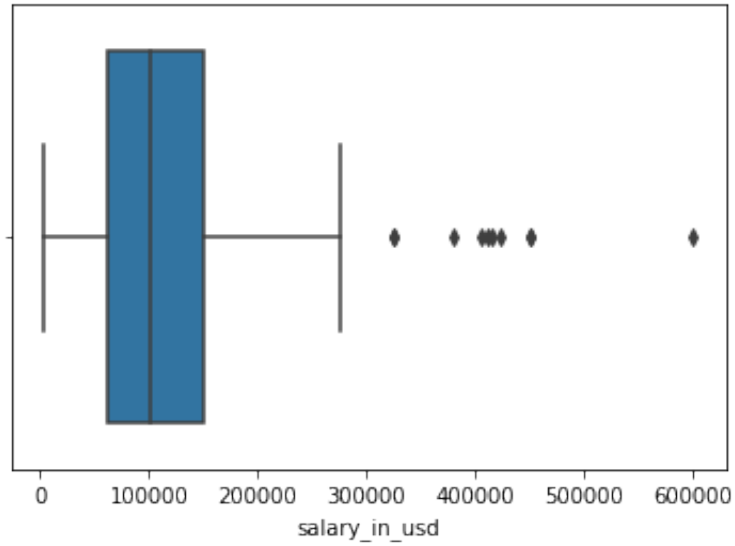
```
work_year        4.790481e-01
salary           2.385040e+12
salary_in_usd    5.034933e+09
remote_ratio     1.657233e+03
dtype: float64
```

Variables cualitativas

¿Cuál es el salario al que pueda aspirar un analista de datos?

Para poder responder esta pregunta, primero es importante mencionar que estaremos utilizando salario en usd, ya que este permite estandarizar este aspecto de nuestra base de datos, dándonos un resultado homogéneo. Dicho esto, podemos agrupar por título de trabajo y obtener las medias, lo que nos despliega una tabla con todos los nombres. Ahí podemos determinar que un **Analista de Datos** puede aspirar a un salario promedio de **92,893.06 USD** El salario más bajo es de **2,859.00 USD**, mientras que el más alto es de **600,000 USD**. Para hacer este análisis, fue necesario eliminar los datos atípicos, los cuales puede llegar a impactar de manera significativa los valores reales.

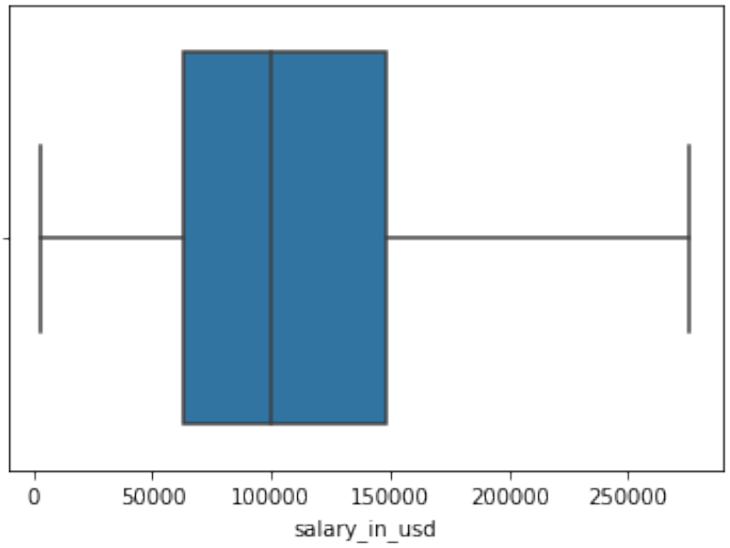
```
count       607.000000
mean     112297.869852
std       70957.259411
min        2859.000000
25%       62726.000000
50%      101570.000000
75%      150000.000000
max      600000.000000
Name: salary_in_usd, dtype: float64

<matplotlib.axes._subplots.AxesSubplot at 0x7f725a7137d0>
```



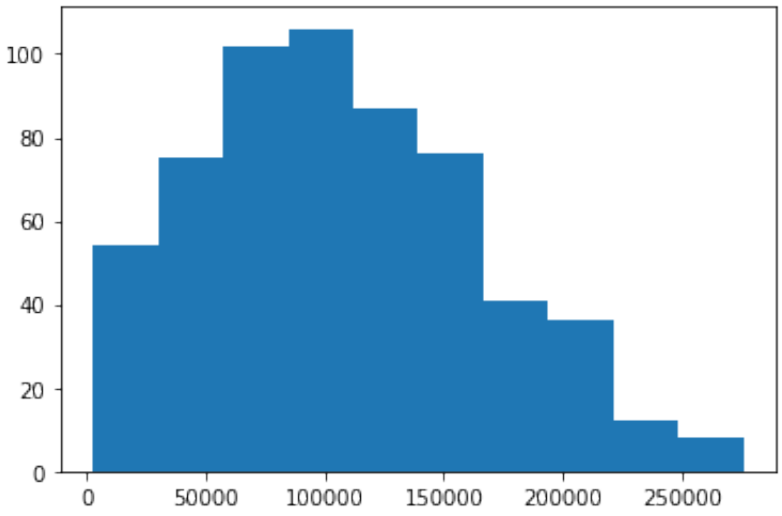
	id	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_
0	0	2020	MI	FT	Data Scientist	70000	EUR	798
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	2600
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	1090
3	3	2020	MI	FT	Product Data Analyst	20000	USD	200
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD	1500
...	...	...	...	...	...	...	...	
602	602	2022	SE	FT	Data Engineer	154000	USD	1540
603	603	2022	SE	FT	Data Engineer	126000	USD	1260
604	604	2022	SE	FT	Data Analyst	129000	USD	1290
605	605	2022	SE	FT	Data Analyst	150000	USD	1500
606	606	2022	MI	FT	AI Scientist	200000	USD	2000

597 rows × 12 columns

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f725a68c610>
```



```
(array([ 54.,  75., 102., 106.,  87.,  76.,  41.,  36.,  12.,   8.]),
 array([  2859. ,  30173.1,  57487.2,  84801.3, 112115.4, 139429.5,
        166743.6, 194057.7, 221371.8, 248685.9, 276000. ]),
 <a list of 10 Patch objects>)
```

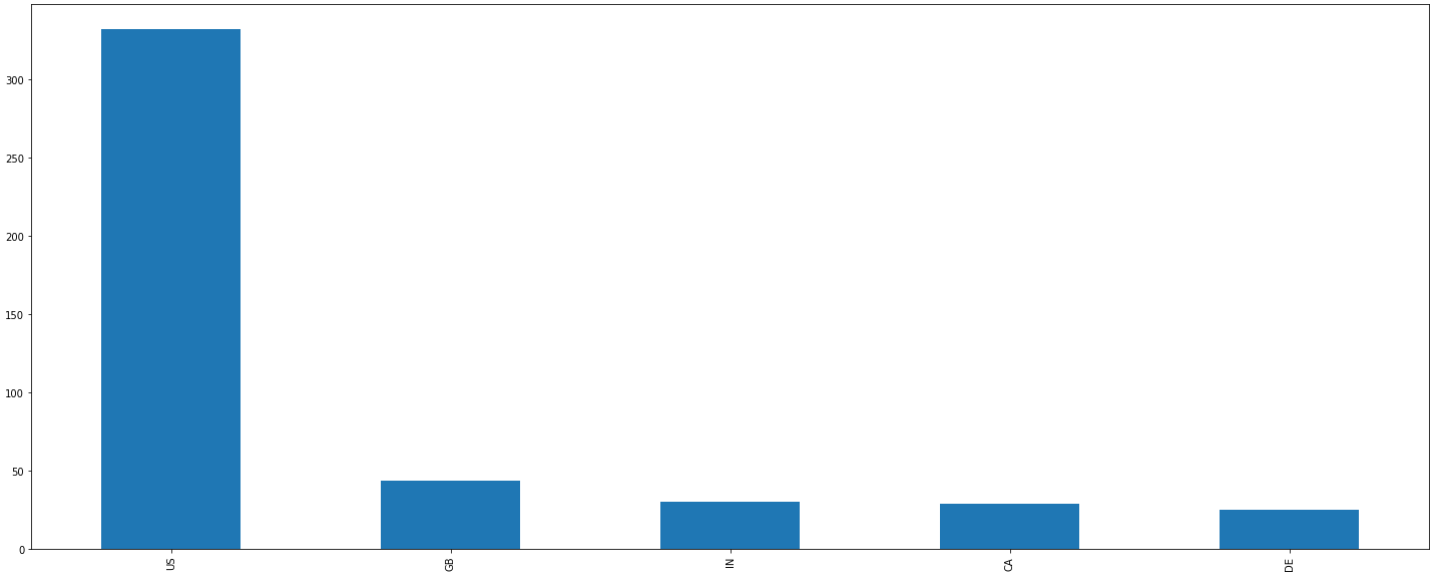


job_title	id	work_year	salary	salary_in_usd	remote_ratio
3D Computer Vision Researcher	77.000000	2021.000000	4.000000e+05	5409.000000	50.000000
AI Scientist	254.142857	2021.142857	2.905714e+05	66135.571429	78.571429
Analytics Engineer	458.250000	2022.000000	1.750000e+05	175000.000000	50.000000
Applied Data Scientist	351.600000	2021.600000	1.724000e+05	124568.750000	70.000000
Applied Machine Learning Scientist	321.000000	2021.500000	1.413500e+05	48425.000000	87.500000
BI Data Analyst	106.333333	2020.833333	1.902045e+06	74755.166667	66.666667
Big Data Architect	255.000000	2021.000000	1.250000e+05	99703.000000	50.000000
Big Data Engineer	123.125000	2020.625000	4.550000e+05	51974.000000	50.000000
Business Data Analyst	256.800000	2021.000000	3.550000e+05	76691.200000	90.000000
Cloud Data Engineer	122.000000	2021.000000	1.400000e+05	124647.000000	75.000000
Computer Vision Engineer	274.833333	2021.166667	8.350000e+04	44419.333333	58.333333
Computer Vision Software Engineer	235.666667	2021.333333	1.003333e+05	105248.666667	100.000000
Data Analyst	362.010309	2021.680412	9.660496e+04	92893.061856	75.257732
Data Analytics Engineer	216.750000	2021.250000	6.175000e+04	64799.250000	75.000000
Data Analytics Lead	523.000000	2022.000000	4.050000e+05	NaN	100.000000
Data Analytics Manager	366.285714	2021.571429	1.271343e+05	127134.285714	85.714286
Data Architect	390.636364	2021.727273	1.778739e+05	177873.909091	100.000000

Data Engineer	343.537879	2021.590909	1.792106e+05	111112.213740	75.000000
Data Engineering Manager	107.200000	2020.600000	1.197998e+05	123227.200000	70.000000
Data Science Consultant	138.000000	2020.714286	1.227143e+05	69420.714286	71.428571
Data Science Engineer	229.666667	2021.333333	8.450000e+04	75803.333333	83.333333
Data Science Manager	274.000000	2021.333333	1.062599e+06	158328.500000	83.333333
Data Scientist	314.832168	2021.391608	5.083472e+05	106048.309859	63.986014
Data Specialist	165.000000	2021.000000	1.650000e+05	165000.000000	100.000000
Director of Data Engineering	171.500000	2021.000000	1.412500e+05	156738.000000	100.000000
Director of Data Science	185.857143	2021.000000	1.932857e+05	173419.666667	42.857143
ETL Developer	373.500000	2022.000000	5.000000e+04	54957.000000	0.000000
Finance Data Analyst	183.000000	2021.000000	4.500000e+04	61896.000000	50.000000
Financial Data Analyst	279.000000	2021.500000	2.750000e+05	100000.000000	75.000000
Head of Data	302.200000	2021.400000	1.564000e+05	160162.600000	90.000000
Head of Data Science	270.250000	2021.500000	1.467188e+05	146718.750000	50.000000
Head of Machine Learning	384.000000	2022.000000	6.000000e+06	79039.000000	50.000000
Lead Data Analyst	64.333333	2020.666667	5.690000e+05	92203.000000	100.000000
Lead Data Engineer	145.500000	2020.833333	1.403333e+05	139724.500000	66.666667
Lead Data Scientist	53.000000	2020.333333	1.101667e+06	115190.000000	50.000000
Lead Machine Learning Engineer	457.000000	2022.000000	8.000000e+04	87932.000000	0.000000
ML Engineer	179.333333	2021.000000	2.676667e+06	117504.000000	83.333333
Machine Learning Developer	358.000000	2021.666667	1.000000e+05	85860.666667	83.333333
Machine Learning Engineer	288.585366	2021.317073	2.727179e+05	104880.146341	67.073171
Machine Learning Infrastructure Engineer	234.333333	2021.000000	9.733333e+04	101145.000000	50.000000
Machine Learning Manager	29.000000	2020.000000	1.570000e+05	117104.000000	50.000000
Machine Learning Scientist	248.000000	2021.250000	1.584125e+05	158412.500000	68.750000
Marketing Data Analyst	90.000000	2021.000000	7.500000e+04	88654.000000	100.000000
NLP Engineer	455.000000	2022.000000	2.400000e+05	37236.000000	50.000000
Principal Data Analyst	370.000000	2021.500000	1.225000e+05	122500.000000	100.000000
Principal Data Engineer	196.000000	2021.000000	3.283333e+05	192500.000000	100.000000
Principal Data Scientist	205.285714	2021.000000	2.067143e+05	181782.833333	85.714286
Product Data Analyst	12.000000	2020.000000	2.350000e+05	13036.000000	50.000000
Research Scientist	246.562500	2021.125000	1.104937e+05	86287.466667	53.125000
Staff Data Scientist	283.000000	2021.000000	1.050000e+05	105000.000000	100.000000

¿En qué países se ofrecen mejores salarios?

Para responder a esta pregunta, podemos observar primeramente que la mayor parte de nuestro datos provienen de Estados Unidos, lo que puede ser un indicador de que en este país se ofrezcan mejores salarios. Sin embargo, aplicaremos herramientas estadísticas para obtener esta respuesta.



Podemos concluir que los países que mejor pagan son Myanmar, Puerto Rico, Estados Unidos y Nueva Zelanda.

employee_residence	salary_in_usd
MY	200000.00000
PR	160000.00000
US	149194.11747
NZ	125000.00000
CH	122346.00000

¿Qué tipo de contrato (parcial, tiempo completo, etc) ofrece mejores salarios?

Podemos observar que el tipo de empleo que ofrece mejores contratos es el de **Contract Freelance**, con un sueldo promedio de **126,718.75 USD**.

employment_type	salary_in_usd
CT	126718.750000
FT	108722.326425
FL	48000.000000
PT	33070.500000