

ANÁLISIS DE LOS SALARIOS DE DATA SCIENCE

Carlos Aldair Ortíz Suárez
Data-analysis-21-12
27/01/2022

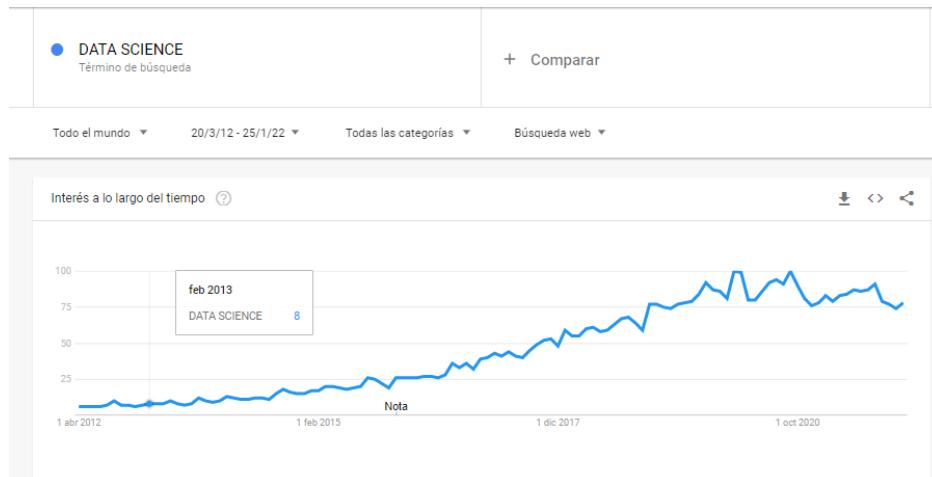
1. Introducción

Este dataset contiene la información de las posiciones de trabajo relacionadas con la ciencia de datos que se dan en la región de California EU, la adquisición de datos de este data se realizó el día 17 de Diciembre del 2021, contiene 3 tablas “trabajadas” para hacer más eficiente su análisis y una 4ta tabla que contiene los datos “crudos”, más adelante se detallarán las tablas.

Se realizará un análisis de dicha información para darle solución al planteamiento del problema usando herramientas de bases estructuradas y no estructuradas, se encuentra documentada la implementación de las herramientas.

2. Planteamiento del Problema

Durante los últimos años ha habido un auge en los temas relacionados al análisis de datos, ciencia de datos, inteligencia artificial, machine learning. Tal como se puede ver en la siguiente gráfico, tuvo un crecimiento importante su búsqueda a principios del año 2020, pero sigue siendo un tema con búsquedas recurrentes.



También se dice que dichos puestos son muy bien pagados, para nosotros que queremos incursionar en el mundo de la ciencia de datos es importe tener una referencia de que podemos aspirar y saber si son realidades o mitos.

Entonces se propone detallar:

- ¿Cuáles son las posiciones mejores pagadas?
- ¿Cuáles son las localidades donde mejor se paga?
- ¿Cuáles son las posiciones que se consideran dentro de estos datos?

Una de las finalidades de en analizar datos es no solo el hecho de resumir o visualizar información por lo que importa es tener este llamado a la acción proponiendo la respuesta final:

- ¿Vale la pena dedicarse a la Ciencia de Datos?

3. Definición de tablas y Diagrama

Los datos se adquirieron de las siguientes tablas:

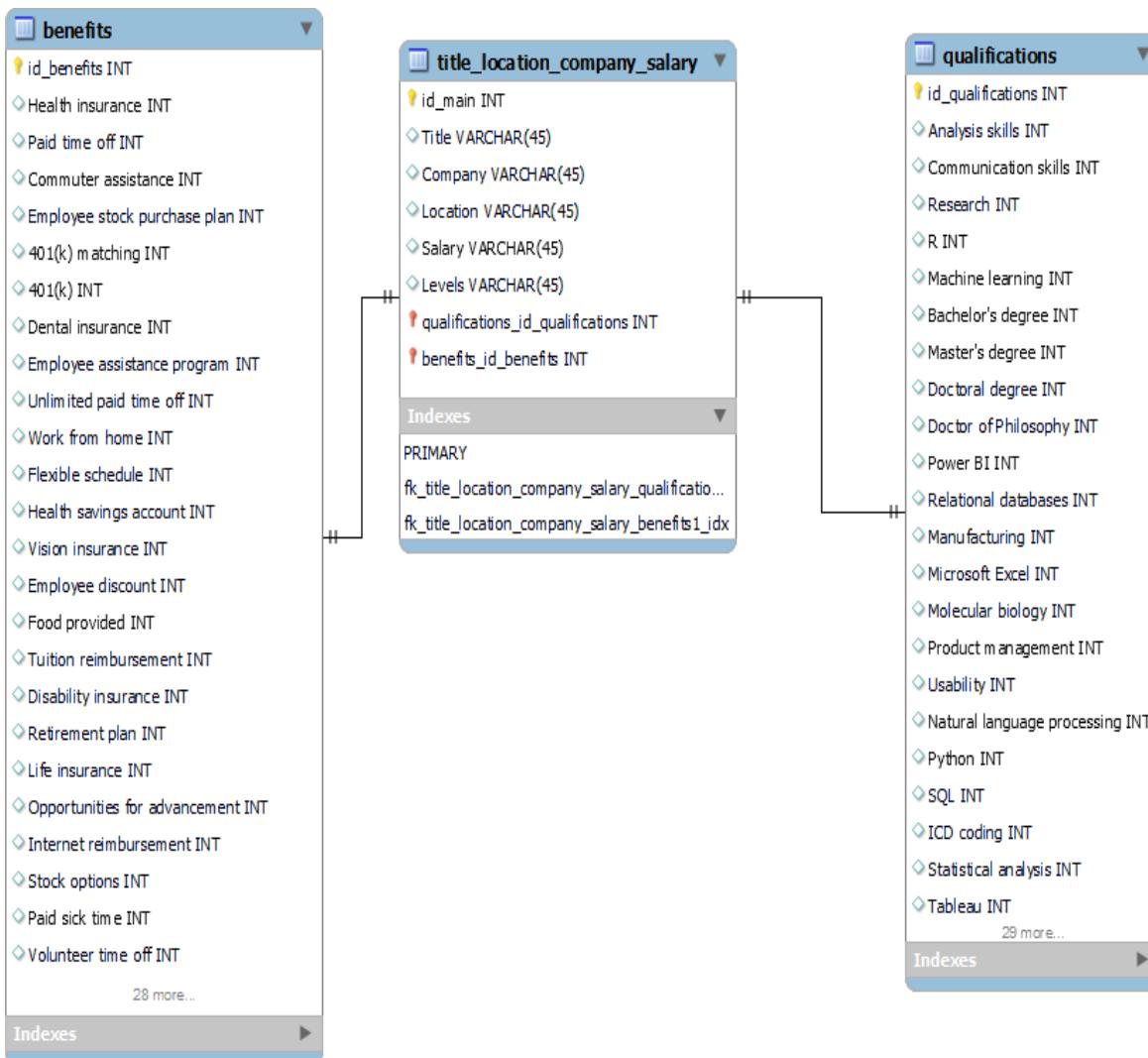
title_location_company_salary: Esta tabla contiene los títulos de las posiciones de ciencia de datos, el nombre de la compañía, la locación, el salario y el nivel o posición. Contiene datos de tipo int y varChar.

qualifications_corregidas: contiene las habilidades, para hacer más eficiente el análisis se usa 0 para "no" y 1 para sí. Contiene datos de tipo int y varChar.

benefits: contiene los beneficios que tienen por registro, para hacer más eficiente el análisis se usa 0 para "no" y 1 para sí. Contiene datos de tipo int y varChar.

Data-analysis-21-12

El análisis se realiza de acuerdo al siguiente diagrama de relaciones.



4. Creación de Base de Datos en MySQL Workbench

Tabla: **title_location_company_salary**

The screenshot shows the MySQL Workbench interface with the following details:

- File Menu:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Navigator:** Shows the schema `data_science_salaries` containing tables `benefits`, `qualifications_corregidas`, and `title_location_company_salary`.
- Query Editor:** Title: `Query 1`. SQL code:

```

1 • create database DATA_SCIENCE_SALARIES;
2 • USE data_science_salaries;
3 • select * from title_location_company_salary;
4 • select * from benefits;
5 • select * from qualifications_corregidas;
```
- Result Grid:** Displays the data for the `title_location_company_salary` table with columns: `id_main`, `Title`, `Company`, `Location`, `Salary`, and `Levels`. The data includes 17 rows of salary information for various Data Scientists across different companies like Numerodox, Cepheid, Verana Health, etc.

Tabla: **benefits**

The screenshot shows the MySQL Workbench interface with the following details:

- File Menu:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Navigator:** Shows the schema `data_science_salaries` containing tables `benefits`, `qualifications_corregidas`, and `title_location_company_salary`.
- Query Editor:** Title: `Query 1`. SQL code:

```

1 • create database DATA_SCIENCE_SALARIES;
2 • USE data_science_salaries;
3 • select * from title_location_company_salary;
4 • select * from benefits;
5 • select * from qualifications_corregidas;
```
- Result Grid:** Displays the data for the `benefits` table with columns: `id_benefits`, `Health insurance`, `Paid time off`, `Commuter assistance`, `Employee stock purchase plan`, `401(k) matching`, `401(k)`, `Dental insurance`, and `Employee ass program`. The data includes 15 rows of benefit information for employees.

Data-analysis-21-12

Tabla: **qualifications_corregidas**

The screenshot shows a database management interface with the following details:

- Toolbar:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Navigator:** Shows the schema 'data_science_salaries' with tables 'benefits', 'qualifications_corregidas', and 'title_location_comp' selected.
- Query Editor:** Displays the following SQL code:


```

1 •  create database DATA_SCIENCE_SALARIES;
2 •  USE DATA_SCIENCE_SALARIES;
3 •  select * from title_location_company_salary;
4 •  select * from benefits;
5 •  select * from qualifications_corregidas;
```
- Result Grid:** Shows the data for the 'qualifications_corregidas' table with 15 rows. The columns are: id_qualifications, Analysis skills, Communication skills, Research, R, Machine learning, Bachelors degree, Masters degree, Doctoral degree, and Doctor of Philosophy.
- Right Panel:** Includes tabs for Result Grid, Form Editor, Field Types, and Query Stats.
- Status Bar:** Shows 'Read Only' status.

5. Creación de Base de Datos en MongoDB

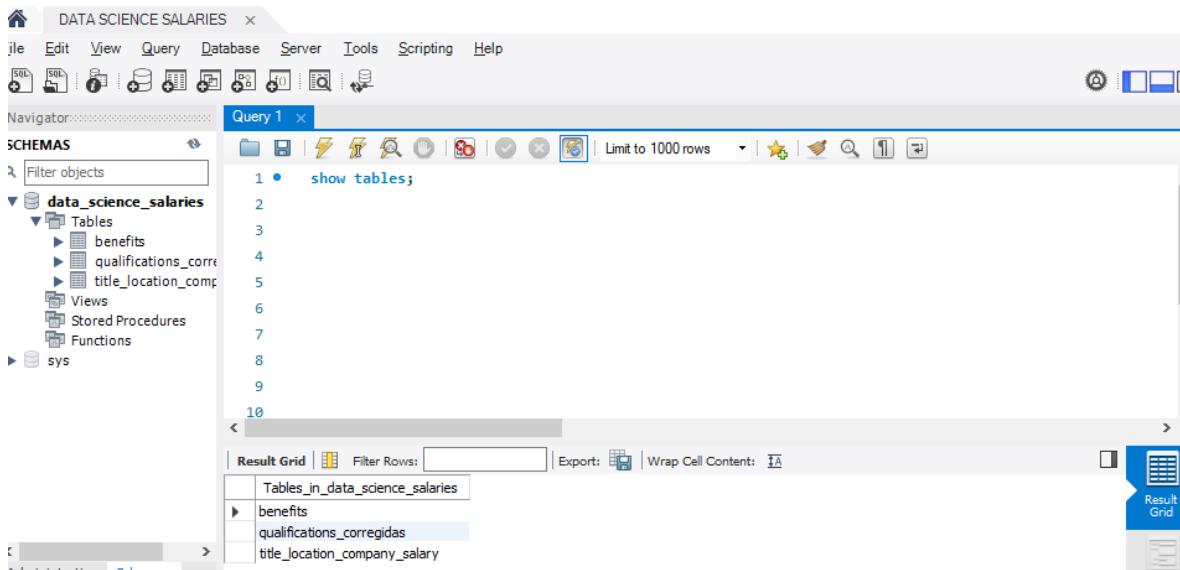
The screenshot shows the MongoDB Compass interface with the following details:

- Local** database selected.
- Collections** tab selected.
- Create collection** button.
- Sort by**: Collection Name dropdown.
- benefits** collection:
 - Storage size: 98.30 kB
 - Documents: 1.3 K
 - Avg. document size: 579.00 B
 - Indexes: 1
 - Total index size: 28.67 kB
- qualifications_corregidas** collection:
 - Storage size: 4.10 kB
 - Documents: 1.3 K
 - Avg. document size: 941.00 B
 - Indexes: 1
 - Total index size: 4.10 kB
- title_location_company_salary** collection:
 - Storage size: 65.54 kB
 - Documents: 1.3 K
 - Avg. document size: 145.00 B
 - Indexes: 1
 - Total index size: 28.67 kB
- Data_science_salaries** database expanded:
 - benefits
 - qualifications_corregidas
 - title_location_company_salary
 - admin
 - local
 - prueba
 - sample_airbnb

6. Consultas SQL

1. ¿Qué tablas contiene la base de datos?

```
show tables;
```

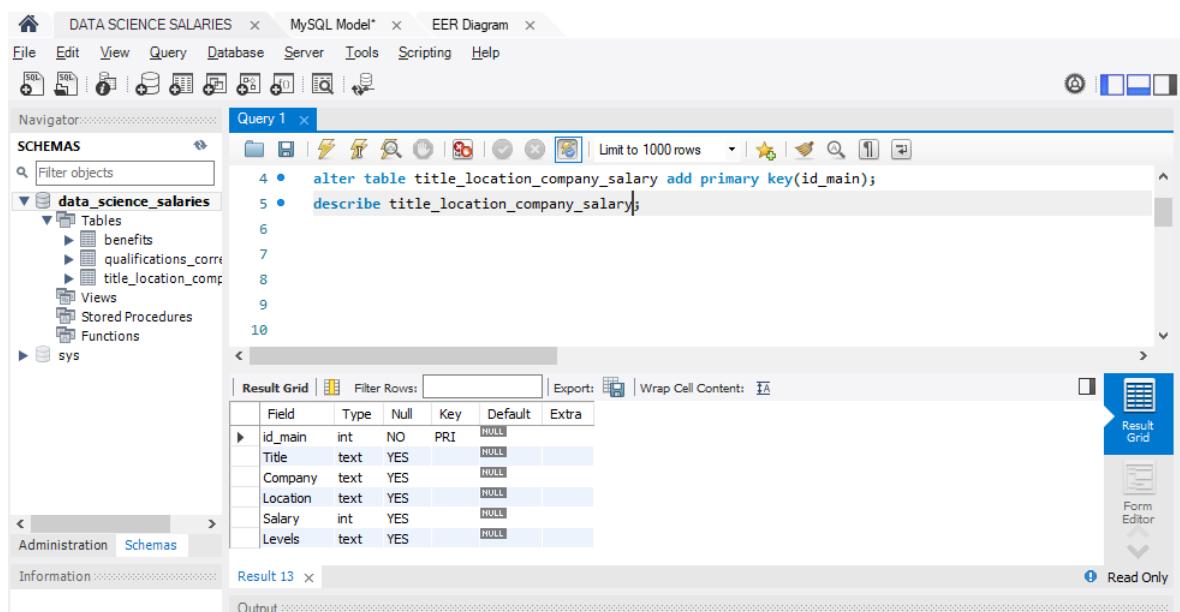


The screenshot shows the MySQL Workbench interface with the 'Query 1' tab active. The query 'show tables;' has been run, and the results are displayed in the 'Result Grid'. The results show three tables: 'benefits', 'qualifications_corregidas', and 'title_location_company_salary'.

```
1 • show tables;
2
3
4
5
6
7
8
9
10
Tables_in_data_science_salaries
▶ benefits
▶ qualifications_corregidas
▶ title_location_company_salary
```

2. Definir las llaves primarias de la tabla title_location_company_salary y verificar que el cambio se haya realizado

```
alter table title_location_company_salary add primary key(id_main);
describe title_location_company_salary;
```



The screenshot shows the MySQL Workbench interface with the 'Query 1' tab active. The queries 'alter table title_location_company_salary add primary key(id_main);' and 'describe title_location_company_salary;' have been run. The 'Result Grid' shows the table structure with the primary key defined on the 'id_main' field.

```
4 • alter table title_location_company_salary add primary key(id_main);
5 • describe title_location_company_salary;
6
7
8
9
10
Field Type Null Key Default Extra
▶ id_main int NO PRI NULL
Title text YES NULL
Company text YES NULL
Location text YES NULL
Salary int YES NULL
Levels text YES NULL
```

Data-analysis-21-12

3. ¿Cuánto le pagan a los 10 directores de Ciencia de Datos mejor pagados?

```
select Title, Salary from title_location_company_salary
where Title ="Director of Data Science"
order by Salary desc limit 10;
```

The screenshot shows the Oracle SQL Developer interface. The query window contains the following code:

```
1 • use data_science_salaries;
2 • alter table qualifications_corregidas add primary key(id_qualifications);
3 • describe qualifications_corregidas;
4 • alter table title_location_company_salary add primary key(id_main);
5 • describe title_location_company_salary;
6
7 • select Title, Salary from title_location_company_salary where Title ="Director of Data Science"
8 order by Salary desc limit 10;
9
10
```

The result grid displays the following data:

Title	Salary
Director of Data Science	434000
Director of Data Science	434000
Director of Data Science	434000
Director of Data Science	240000
Director of Data Science	240000
Director of Data Science	220000
Director of Data Science	220000
Director of Data Science	210000
Director of Data Science	210000

4. ¿Qué compañías hay en San Francisco dónde el salario sea mayor a 150000 mostrando primero las que pagan más?

```
select Company, Location from title_location_company_salary
where Location = "San Francisco, CA"
and salary >150000 order by Salary desc ;
```

The screenshot shows the Oracle SQL Developer interface. The query window contains the following code:

```
9     order by Salary desc limit 10;
10
11 • select Company, Location, Salary from title_location_company_salary
12   where Location = "San Francisco, CA"
13   and salary >150000
14   order by Salary desc;
15
```

The result grid displays the following data:

Company	Location	Salary
Indeed	San Francisco, CA	434000
Indeed	San Francisco, CA	434000
Indeed	San Francisco, CA	434000
Indeed	San Francisco, CA	324000
Harmhan	San Francisco, CA	240000
Harmhan	San Francisco, CA	240000
Wayfair	San Francisco, CA	240000
CyberCoders	San Francisco, CA	225000
Indeed	San Francisco, CA	223000
Indeed	San Francisco, CA	223000
Indeed	San Francisco, CA	223000
Harmhan	San Francisco, CA	220000
Uber	San Francisco, CA	220000
Ursus	San Francisco, CA	220000
Ursus	San Francisco, CA	220000

Data-analysis-21-12

5. ¿Hay trabajadores que tengan seguro dental, trabajo desde casa y seguro para mascotas?

```
select id_benefits, Health_insurance, Work_from_home,
Pet_insurance
from benefits
where Health_insurance = 1 and Work_from_home=1 and
Pet_insurance;
```

id_benefits	Health_insurance	Work_from_home	Pet_insurance
899	1	1	1

6. ¿Cuales son los 50 puestos que llevan "Science" o "Scientist" que ganan más?

```
select * from title_location_company_salary
where Title like '%cien%'
order by Salary desc limit 50;
```

id_main	Title	Company	Location	Salary	Levels
96	Data Science Manager	CyberCoders	San Francisco, CA	225000	Unknown
7	Data Scientist	Indeed	San Francisco, CA	223000	Sr.
26	Data Scientist	Indeed	San Francisco, CA	223000	Sr.
494	Data Scientist	Indeed	Sunnyvale, CA	223000	Sr.
355	Data Scientist	Indeed	Sunnyvale, CA	223000	Sr.
95	Data Scientist	Indeed	San Francisco, CA	223000	Sr.
301	Data Scientist	Indeed	Sunnyvale, CA	223000	Sr.
1112	Data Science Manager	Reddit		220000	Unknown
972	Head of Data Science	Parallel Bio	Los Angeles, CA	220000	Unknown
317	Data Scientist	Intuit	Mountain View, CA	220000	Principal
84	Director of Data Science	Sleep Number	San Jose, CA	220000	Unknown
722	Data Scientist	ISO	San Francisco, CA	220000	Sr.
165	Data Scientist	Ursus	San Francisco, CA	220000	Sr.
155	Data Scientist	Ursus	San Francisco, CA	220000	Sr.
19	Director of Data Science	Harnham	San Francisco, CA	220000	Unknown

Data-analysis-21-12

7. ¿Cuál compañía es la que paga menos, cuánto es lo que paga y en qué nivel?

```
select Title, Company, min(Salary), Levels
from title_location_company_salary
where Salary > 0 ;
```

Title	Company	min(Salary)	Levels
Data Scientist	Numerodox	10400	Jr.

8. ¿Cuál es el promedio de salarios en California sin contar los que no tienen datos?

```
select Title, Company, avg(Salary) as Avarage, Levels
from title_location_company_salary
where Salary > 0 ;
```

Title	Company	Avarage	Levels
Data Scientist	Numerodox	151220.0721	Jr.

Data-analysis-21-12

9. ¿Cuántos puestos tienen trabajo desde casa?

```
select count(Work_from_home) as "Total of work from home"
from benefits
where Work_from_home=1;
```

Total of work from home
30

10. ¿Cuáles son los puestos que han sido considerados para esta investigación?

```
select Title from title_location_company_salary
Group by Title;
```

Title
Data Scientist
Data Analyst
Machine Learning Scientist
Director of Data Science
Machine Learning Engineer
Data Engineer
Software Engineer
Data Science Manager
Data Architect
Applied Scientist
Statistician
Vice President of Data Sci...
Head of Data Science

Data-analysis-21-12

11. ¿Cuáles es el salario más alto y más bajo por posición? ordenados desde el más alto.

```
select Title, max(Salary) as "High", min(Salary) "Low"
from title_location_company_salary where Salary >0
Group by Title
Order by High desc;
```

The screenshot shows the MySQL Workbench interface with a query editor titled 'Query 1'. The code is identical to the one above. The results grid displays a table with three columns: 'Title', 'High' (maximum salary), and 'Low' (minimum salary). The data includes various roles like Director of Data Science, Data Engineer, Data Scientist, etc., with their respective salary ranges.

Title	High	Low
Director of Data Science	434000	130000
Data Engineer	350000	110000
Data Scientist	324000	65000
Software Engineer	311000	82000
Machine Learning Engineer	250000	78000
Data Science Manager	240000	90000
Machine Learning Scientist	230000	83000
Head of Data Science	220000	110000
Data Architect	210000	170000
Data Analyst	200000	10400
Applied Scientist	180000	110000
Vice President of Data Sci...	180000	130000
Statistician	110000	84000

12. ¿Cuáles son los trabajadores, posiciones y compañías ofrecen trabajo desde casa?

```
select id_main, Title, Company
from title_location_company_salary
where id_main in
    (select id_benefits
     from benefits where Work_from_home=1) ;
```

The screenshot shows the MySQL Workbench interface with a query editor titled 'Query 1'. The code is identical to the one above. The results grid displays a table with three columns: 'id_main', 'Title', and 'Company'. The data lists various employees along with their job titles and company names, filtered by those who work from home.

id_main	Title	Company
8	Data Scientist	Ezoic
28	Data Scientist	Hannham
39	Data Scientist	Ezoic
45	Machine Learning Engineer	Hannham
53	Data Science Manager	Hannham
66	Machine Learning Engineer	Hannham
77	Machine Learning Engineer	Hannham
94	Machine Learning Engineer	Hannham
118	Data Scientist	Afterpay Touch
151	Data Scientist	Gleed Sciences
222	Data Analyst	Kett Engineering
292	Machine Learning Engineer	Tubi
378	Machine Learning Engineer	Standard Cogn...
381	Data Analyst	Point Digital Fin...
430	Data Scientist	Freedom Finan...

Data-analysis-21-12

13. Si sé de Python, R y tengo un Bachelors degree, en ¿qué posiciones y cuánto podría ganar? De mayor a menor

```
from title_location_company_salary where id_main in(select
id_qualifications
from qualifications
where R=1 and Bachelors_degree=1 and Python =1)
order by Salary desc;
```

The screenshot shows a database management system (DBMS) interface. The top menu bar includes File, Edit, View, Query, Database, Server, Tools, Scripting, and Help. Below the menu is a toolbar with various icons. The left sidebar shows the 'Navigator' with the 'SCHEMAS' section expanded, revealing 'data_science_salaries' which contains 'Tables' like 'benefits', 'title_location_comp', 'qualifications', and 'Views'. The main area is titled 'Query 1' and contains the following SQL code:

```

60
61
62 •   select id_main, Title, Salary
63 ⚡ from title_location_company_salary where id_main in(select id_qualifications
64     from qualifications
65     where R=1 and Bachelors_degree=1 and Python =1)
66     order by Salary desc;
67
68
69
70
71

```

The result grid below displays the query results:

	id_main	Title	Salary
153	Machine Learning Engineer	170000	
899	Data Scientist	170000	
1164	Data Scientist	170000	
137	Data Scientist	160000	
288	Data Scientist	160000	
616	Data Scientist	160000	
671	Data Scientist	160000	
695	Software Engineer	160000	
379	Data Scientist	150000	
719	Machine Learning Engineer	150000	
846	Data Scientist	150000	
86	Data Scientist	140000	

14. Realiza la unión de la tabla dónde se pueda tener los datos de las habilidades de cada trabajador y la tabla dónde viene el nombre de la posición.

```
select * from title_location_company_salary
join qualifications
on title_location_company_salary.id_main =
qualifications.id_qualifications;
```

The screenshot shows a database management system interface with the following details:

- File Bar:** File, Edit, View, Query, Database, Server, Tools, Scripting, Help.
- Toolbar:** Includes icons for New, Open, Save, Print, Copy, Paste, etc.
- Navigator:** Shows the schema structure under the 'data_science_salaries' database, including Tables (benefits, title_location_comp, qualifications), Views, Stored Procedures, and Functions.
- Query Editor:** Titled 'Query 1', contains the following SQL code:


```
97
98 •   select * from title_location_company_salary ;
99 •   select * from qualifications;
100
101
102 •   select * from title_location_company_salary
103     join qualifications on title_location_company_salary.id_main = qualifications.id_qualifications;
104
105
```
- Result Grid:** Displays the results of the joined query. The columns are: id_main, Title, Company, Location, Salary, Levels, id_qualifications, Analysis_skills, and Communication. The data rows are:

	id_main	Title	Company	Location	Salary	Levels	id_qualifications	Analysis_skills	Communication
▶	0	Data Scientist	Numerdox	Sacramento, CA	110000	Jr.	0	1	1
	1	Data Analyst	Cepheid	Lodi, CA	120000	Unknown	1	0	0
	2	Data Scientist	Cepheid	Sunnyvale, CA	150000	Staff	2	0	0
	3	Data Scientist	Verana Health	San Francisco, CA	150000	Unknown	3	0	1
	4	Data Scientist	Tinder	San Francisco, CA	130000	Jr.	4	0	0
	5	Data Scientist	JPL/NASA	Pasadena, CA	100000	Jr.	5	0	0
	6	Data Scientist	IBM	San Jose, CA	120000	Unknown	6	0	1

Data-analysis-21-12

15. Realiza una tabla donde se puedan visualizar los datos de la tabla de puestos y datos ocupados con los beneficios.

```
select *from title_location_company_salary
right      join      benefits
title_location_company_salary.id_main
benefits.id_benefits;
```

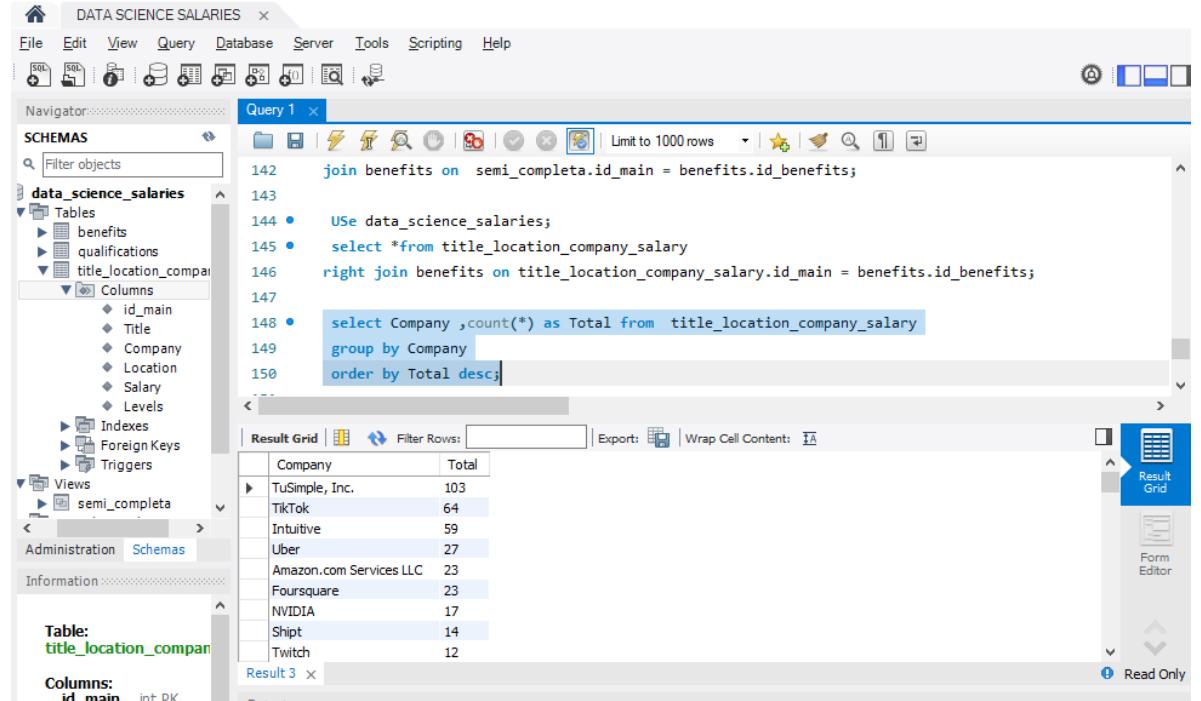
The screenshot shows the MySQL Workbench interface with the following details:

- Navigator:** Shows the schema `data_science_salaries` with tables `benefits`, `qualifications`, `title_location_comp`, and views `semi_completa`.
- Query Editor (Query 1):** Displays the SQL query from the previous code block.
- Result Grid:** Shows the output of the query, displaying 7 rows of data with columns: id_main, Title, Company, Location, Salary, Levels, id_benefits, Health_insurance, and Paid-time_off.

	id_main	Title	Company	Location	Salary	Levels	id_benefits	Health_insurance	Paid-time_off
▶	0	Data Scientist	Numerodox	Sacramento, CA	110000	Jr.	0	0	0
	1	Data Analyst	Cepheid	Lodi, CA	120000	Unknown	1	1	1
	2	Data Scientist	Cepheid	Sunnyvale, CA	150000	Staff	2	1	1
	3	Data Scientist	Verana Health	San Francisco, CA	150000	Unknown	3	0	0
	4	Data Scientist	Tinder	San Francisco, CA	130000	Jr.	4	1	0
	5	Data Scientist	JPL/NASA	Pasadena, CA	100000	Jr.	5	0	0
	6	Data Scientist	IBM	San Jose, CA	120000	Unknown	6	1	0

16. ¿Cuales son las empresas con más trabajadores en la rama de ciencia de datos?

```
select Company ,count(*) as Total from title_location_company_salary
group by Company
order by Total desc;
```



The screenshot shows the MySQL Workbench interface. The left sidebar displays the 'Navigator' with the 'SCHEMAS' section expanded, showing 'data_science_salaries' with its tables: 'benefits', 'qualifications', and 'title_location_company'. The 'Tables' section also lists 'semi_completa'. The 'Information' section shows the definition for 'Table: title_location_compan'.

The main area is titled 'Query 1' and contains the following SQL code:

```

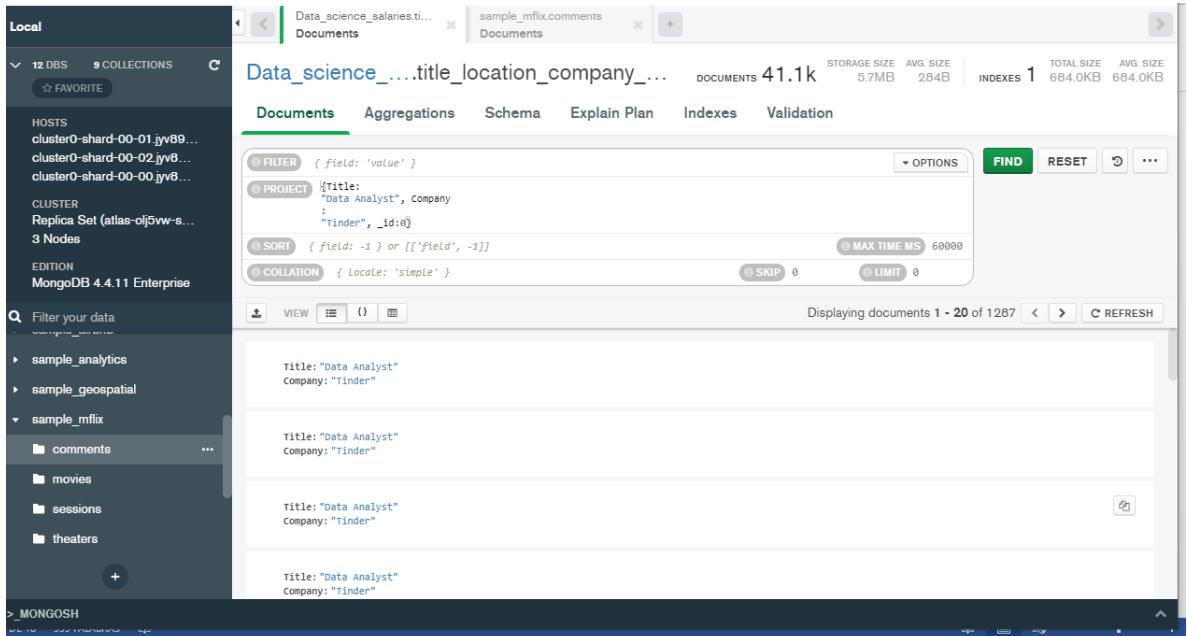
142     join benefits on semi_completa.id_main = benefits.id_benefits;
143
144 •   USE data_science_salaries;
145 •   select *from title_location_company_salary
146     right join benefits on title_location_company_salary.id_main = benefits.id_benefits;
147
148 •   select Company ,count(*) as Total from title_location_company_salary
149   group by Company
150   order by Total desc;
  
```

The results are displayed in a 'Result Grid' table:

Company	Total
TuSimple, Inc.	103
TikTok	64
Intuitive	59
Uber	27
Amazon.com Services LLC	23
Foursquare	23
NVIDIA	17
Shift	14
Twitch	12

7. Consultas Mongo

- Proyecta a los analistas de datos de la compañía Tinder
`{Title: "Data Analyst", Company : "Tinder", _id:0}`

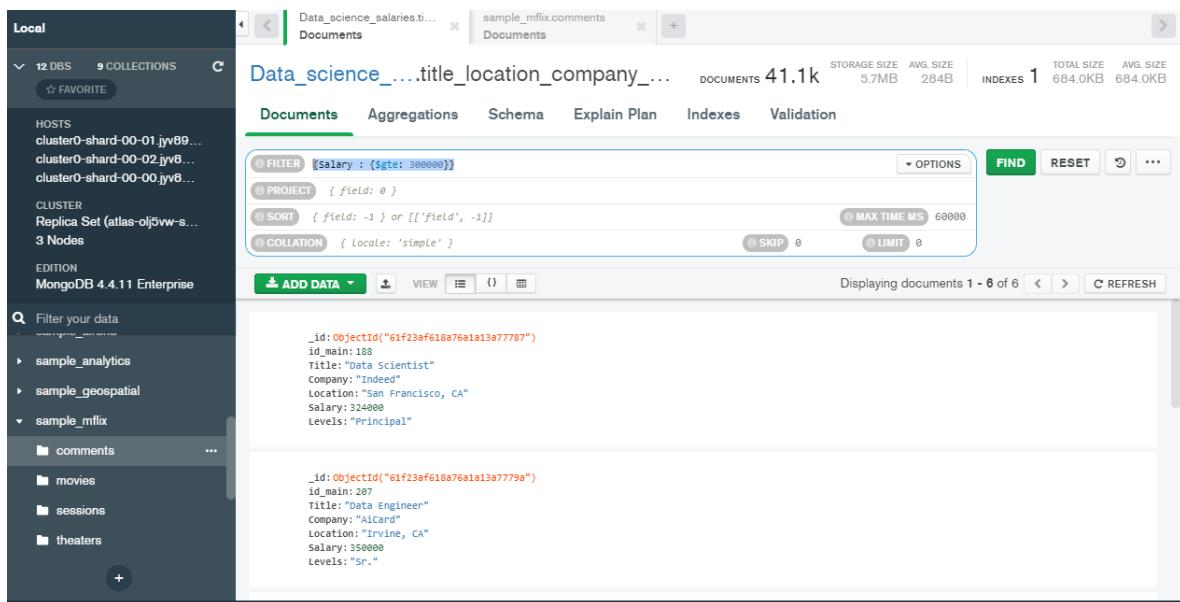


The screenshot shows the MongoDB Compass interface. On the left, the sidebar lists databases (Local, sample_analytics, sample_geospatial, sample_mflix) and collections (comments, movies, sessions, theaters). The main area shows two tabs: 'Data_science_salaries.ti...' and 'sample_mflix.comments'. The 'sample_mflix.comments' tab is active, displaying a table of documents. A query builder is open with the following parameters:

- FILTER:** `{Title: "Data Analyst", Company : "Tinder", _id:0}`
- OPTIONS:** MAX TIME MS: 60000
- COLLATION:** Locale: 'simple'

The results table shows four documents, each with the fields Title: "Data Analyst" and Company: "Tinder".

- Consultar aquellos trabajadores que ganan más de \$300000
`{Salary : {$gte: 300000}}`

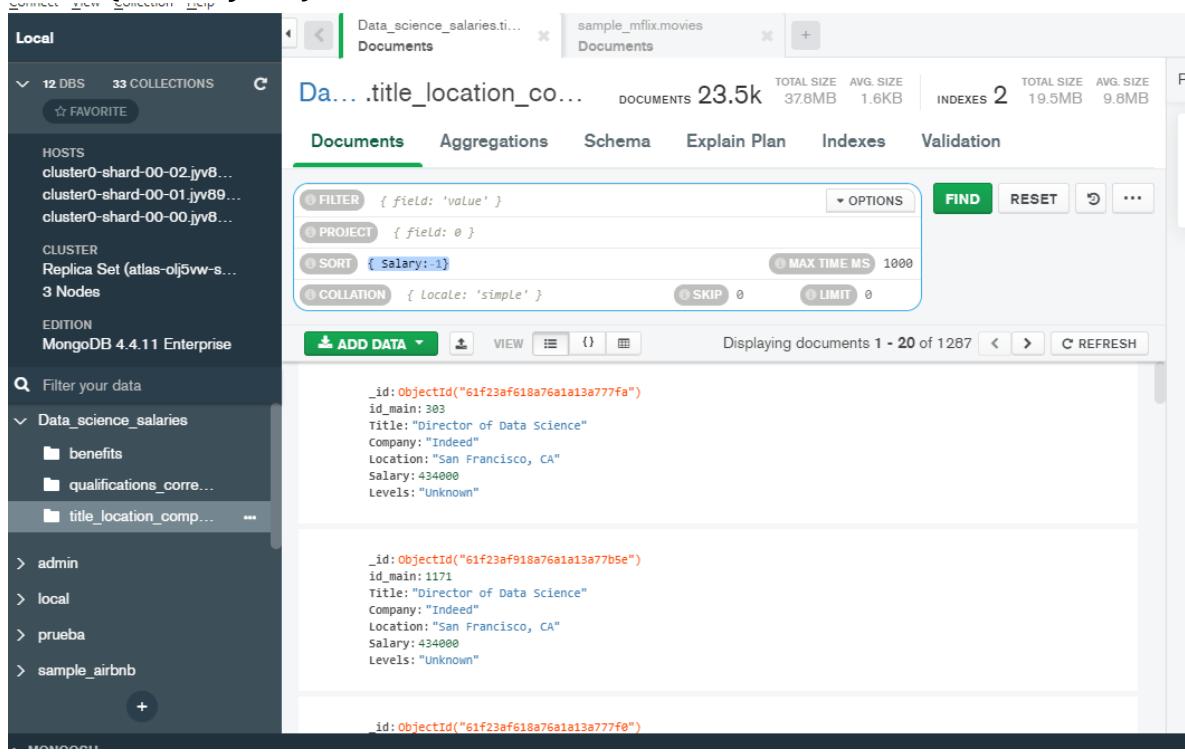


The screenshot shows the MongoDB Compass interface, similar to the previous one. The sidebar and tabs are identical. The query builder parameters are:

- FILTER:** `{Salary : {$gte: 300000}}`
- OPTIONS:** MAX TIME MS: 60000
- COLLATION:** Locale: 'simple'

The results table shows two documents, both with the field Salary: 324000. The first document also includes other fields like id_main, Title, Company, Location, and Levels.

3. Ordenar los salarios más altos con su puesto
 { Salary:-1}



The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays the local database with 12 DBs and 33 collections, including 'Data_science_salaries'. The main area shows two tabs: 'Data_science_salaries' and 'sample_mflix.movies'. The 'Data_science_salaries' tab is active, showing a table of documents. A query builder is overlaid on the interface, with the 'SORT' stage highlighted: { Salary:-1}. The results table shows three documents with salary information.

_id	id_main	Title	Company	Location	Salary	Levels
<code>ObjectId("61f23af618a76a1a13a777fa")</code>	303	Director of Data Science	Indeed	San Francisco, CA	434000	Unknown
<code>ObjectId("61f23af918a76a1a13a77b5e")</code>	1171	Director of Data Science	Indeed	San Francisco, CA	434000	Unknown
<code>ObjectId("61f23af618a76a1a13a777fa")</code>						

4. Muestra los registros que sean de las empresas Uber, IBM, Tinder.

Data-analysis-21-12

{Company: {\$in: ["Tinder", "IBM", "Uber"]}}

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays the local database with 12 DBs and 33 collections. The main area shows two tabs: 'Data_science_salaries.ti...' and 'sample_mflix.movies'. The 'Data_science_salaries' tab is active, showing a search query: {Company: {\$in: ["Tinder", "IBM", "Uber"]}}. The results pane displays two documents:

```

_id: ObjectId("61f23af618a76a1a13a776cf")
id_main: 4
Title: "Data Scientist"
Company: "Tinder"
Location: "San Francisco, CA"
Salary: 120000
Levels: "Jr."

```

```

_id: ObjectId("61f23af618a76a1a13a776d1")
id_main: 6
Title: "Data Scientist"
Company: "IBM"
Location: "San Jose, CA"
Salary: 120000
Levels: "Unknown"

```

5. ¿Cuánto le pagan a los 10 directores de Ciencia de Datos mejor pagados?

```
{Title : "Director of Data Science" }
{
  Title: 1,
  Salary: 1
}
{
  Salary: -1
}
```

The screenshot shows the MongoDB Compass interface with the following details:

- Left Sidebar:** Shows the local database with 12 DBs and 25 collections. The `Data_science_salaries` collection is selected.
- Top Bar:** Shows the collection name `Data_science_salaries.title_location_company_...`, document count **1.3k**, storage size **69.6KB**, avg. size **145B**, and index count **1**.
- Document View:** Displays three documents matching the query:
 - `_id: ObjectId("61f23af618a76a1a13a7771f")`, `Title: "Director of Data Science"`, `Salary: 220000`
 - `_id: ObjectId("61f23af618a76a1a13a77abf")`, `Title: "Director of Data Science"`, `Salary: 210000`
 - `_id: ObjectId("61f23af618a76a1a13a77985")`, `Title: "Director of Data Science"`, `Salary: 210000`

6. ¿Qué compañías hay en San Francisco dónde el salario sea mayor a 150000 mostrando primero las que pagan más?

```
{
  $and: [{Location : "San Francisco, CA"}, {Salary :{ "$gt" : 150000 } }
  ],
}
{
  Company: 1,
  Location: 1
}
{
  Salary: -1
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays the local database with 12 DBs and 25 collections. The 'title_location_company...' collection is selected. The main pane shows a table of document results with three entries. The first entry is:

```
_id:ObjectId("61f23af618a7ea1a13a777fa")
Company:"Indeed"
Location:"San Francisco, CA"
```

The second entry is:

```
_id:ObjectId("61f23af918a7ea1a13a777b5e")
Company:"Indeed"
Location:"San Francisco, CA"
```

The third entry is:

```
_id:ObjectId("61f23af618a7ea1a13a777f0")
Company:"Indeed"
Location:"San Francisco, CA"
```

The right sidebar shows the 'Past Queries' tab with the recent query displayed:

```
{
  $and: [
    {
      Location: "San Francisco, CA"
    },
    {
      Salary: {
        $gt: 150000
      }
    }
  ]
}
```

Data-analysis-21-12

7. ¿Cuál compañía es la que paga menos, cuánto es lo que paga y en qué nivel?

```
{
  Salary :{ "$gt": 0 }
}
{
  Title: 1,
  Company: 1,
  Salary: 1,
  Levels: 1
}
{Salary:1}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays the database structure with 'Data_science_salaries' selected. The main pane shows a query builder for the collection 'title_location_company'. The query consists of a FILTER stage (Salary > 0), a PROJECT stage (Title, Company, Salary, Levels), a SORT stage (Salary), and a LIMIT stage (1). The results pane displays one document:

```
_id: ObjectId("61f23af618a76a1a13a777a9")
Title: "Data Analyst"
Company: "Kett Engineering"
Salary: 16400
Levels: "Unknown"
```

8. ¿Cuántos puestos tienen trabajo desde casa?

```
{
  "Work_from_home" : 1
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar lists databases and collections, with 'Data_science_salaries' selected. The main area displays the 'benefits' collection. A filter is applied to find documents where 'Work_from_home' is 1. The results table shows one document with the following fields:

_id	id_benefits	Health_insurance	Paid_time_off	Commuter_assistance	Employee_stock_pur...	401(k)_matching	401(k)	Dental_insurance	Employee_assistance...
<code>ObjectId("61f23b0218a76a1a13a77c4a")</code>	118	0	0	0	0	0	0	0	0

8. Vista SQL

Crear una vista que contenga las tablas principales y la tabla de beneficios para futuras consultas.

The screenshot shows the MySQL Workbench interface. In the top navigation bar, the schema is set to 'DATA SCIENCE SALARIES'. The 'Navigator' pane on the left shows the 'data_science_salaries' schema with its tables ('benefits', 'title_location_comp', 'qualifications') and a view ('semi_completa'). The 'Query 1' editor pane contains the following SQL code:

```

134
135      create view semi_completa as
136          select * from title_location_company_salary
137          join qualifications on title_location_company_salary.id_main = qualifications.id_qualifications;
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153

```

Consulta a la tabla: Juntas las tres tablas en una sola tabla

```
select * from semi_completa
join benefits on semi_completa.id_main = benefits.id_benefits;
```

The screenshot shows the MySQL Workbench interface with the same schema and view setup as the previous screenshot. The 'Query 1' editor pane now contains the joined query:

```

137      join qualifications on title_location_company_salary.id_main = qualifications.id_qualifications;
138
139 •   select * from semi_completa;
140
141 •   select * from semi_completa
142     join benefits on semi_completa.id_main = benefits.id_benefits;
143
144
145

```

The 'Result Grid' pane at the bottom displays the resulting data from the joined query. The columns are: tics, iOS, Cooking, Attribution_modeling, Project_management, Customer_service, id_benefits, Health_insurance, Paid-time.

tics	iOS	Cooking	Attribution_modeling	Project_management	Customer_service	id_benefits	Health_insurance	Paid-time
0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1
0	0	0	0	0	2	1	1	1
0	0	0	0	0	3	0	0	0
0	0	0	0	0	4	1	0	0
0	0	0	0	0	5	0	0	0
0	0	0	0	0	6	1	0	0

9. Conclusiones

Ambas herramientas tienen características diferentes, de momento, me parece más visual trabajar con SQL, aunque MONGO DB tiene características que facilitan algunas tareas. Sobre todo la posibilidad de usar bases de datos no relacionales.

10. Referencias

Data Science Job Postings and Salaries. (2021, 22 diciembre). Kaggle.
<https://www.kaggle.com/michaelbryantds/california-salaries-in-data-science>