

UNIVERSIDAD
COMPLUTENSE
DE MADRID



Evaluación Final

Base de datos NoSQL

Autor: Bryan Manuel Frías Rey

Introducción

Este conjunto de datos que seleccioné fue extraído de la plataforma kaggle y contiene información relacionada al trabajo en el sector de los datos. A continuación se muestra una descripción general de las columnas:

_id: Un identificador único generado automáticamente para cada registro.

work_year: Año asociado a los datos.

experience_level: Nivel de experiencia del empleado.

employment_type: Tipo de empleo.

job_title: Título del trabajo.

salary: Salario en la moneda especificada.

salary_currency: Moneda del salario.

salary_in_usd: Salario convertido a dólares estadounidenses.

employee_residence: País de residencia del empleado.

remote_ratio: Porcentaje de trabajo remoto.

company_location: Ubicación de la empresa.

company_size: Tamaño de la empresa.

Aquí se ve representada información detallada sobre salarios, condiciones de trabajo y ubicación para distintos roles y niveles de experiencia en el campo de la tecnología y análisis de datos. Este tipo de conjunto de datos es común en análisis de recursos humanos y puede utilizarse para evaluar tendencias salariales.

El autor del dataset que postea en kaggle explica la proveniencia de los datos:

“Este conjunto de datos se recuperó de la página <https://ai-jobs.net/salaries/download/>

Este sitio recopila información salarial anónima de profesionales de todo el mundo en el ámbito de la IA, el ML y la ciencia de datos, y la pone a disposición del público para que cualquiera pueda utilizarla, compartirla y jugar con ella.

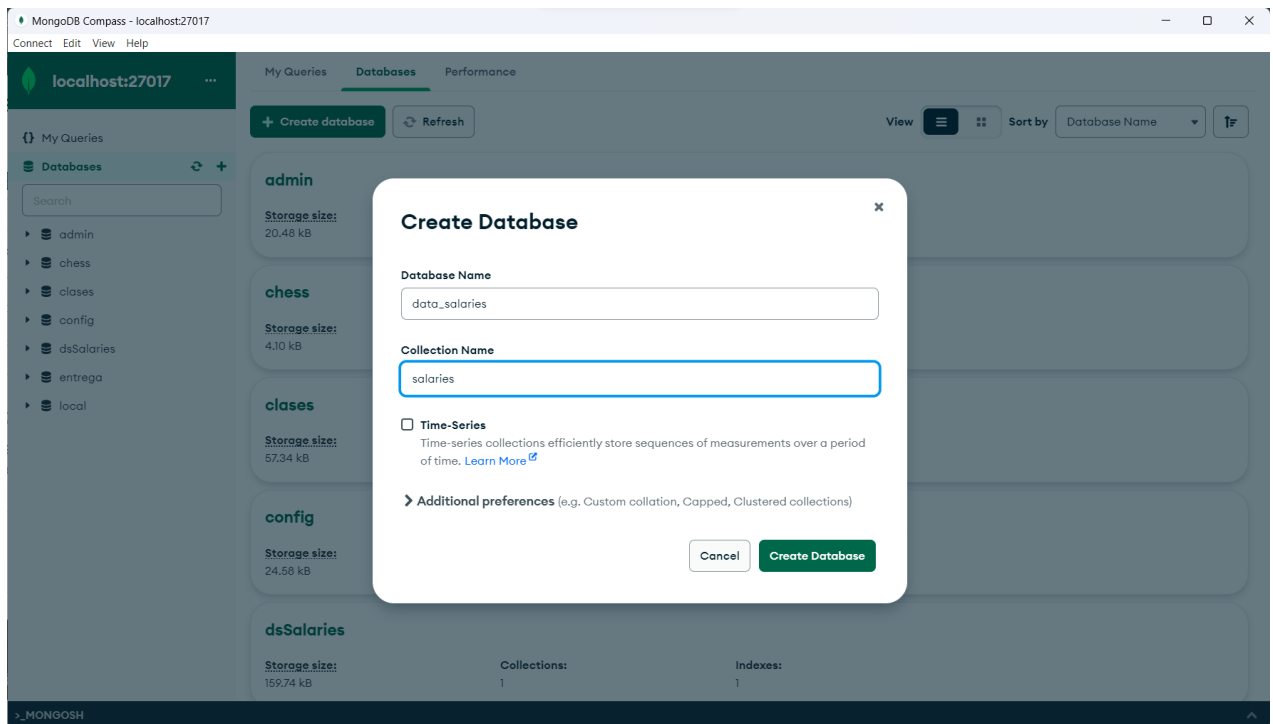
El objetivo principal es disponer de datos que puedan proporcionar una mejor orientación con respecto a lo que se paga en todo el mundo. Así, los novatos, los profesionales con experiencia, los responsables de contratación, los reclutadores y también los fundadores de startups o las personas que desean cambiar de carrera pueden tomar decisiones mejor informadas.”

1. Cargar / Importar dataset

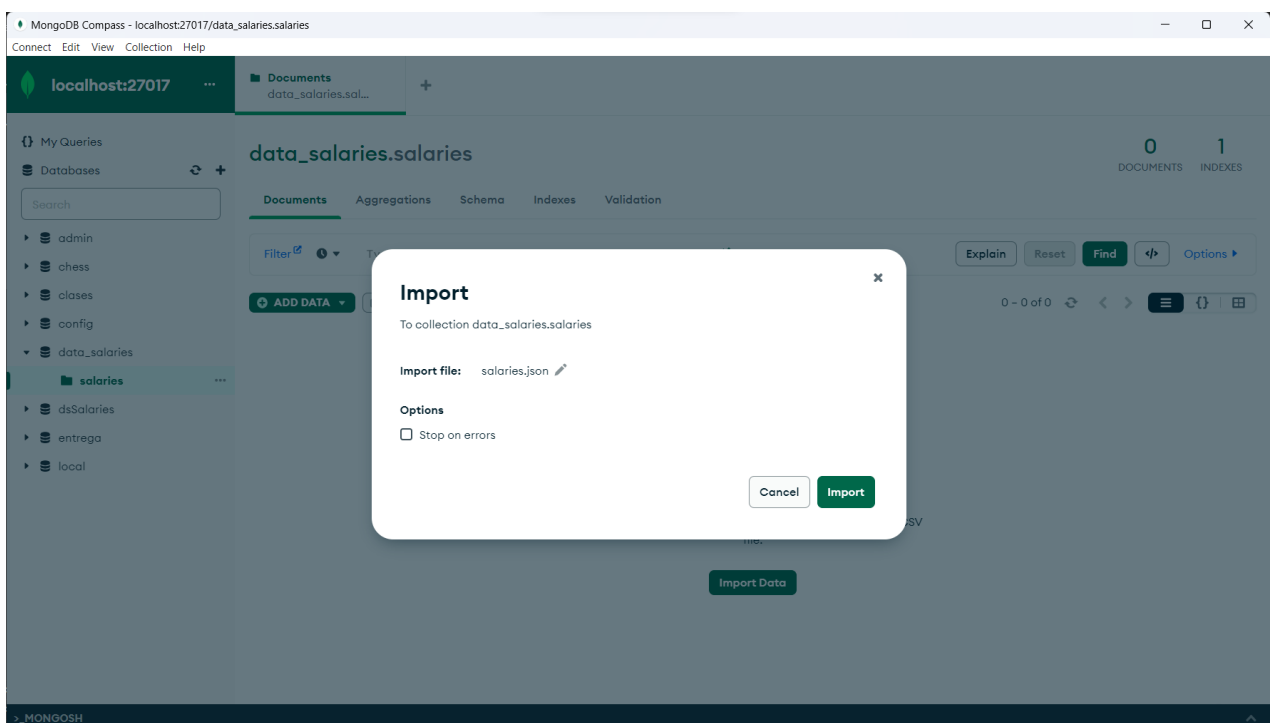
Primero que todo nos disponemos a iniciar el servicio de MongoDB en nuestro computador. Esto lo haremos ejecutando la terminal de windows como administrador y lanzando la siguiente línea de código:

```
net start mongod
```

Luego continuaremos abriendo Mongo DB Compass para importar la base de datos que queremos analizar en esta ocasión.



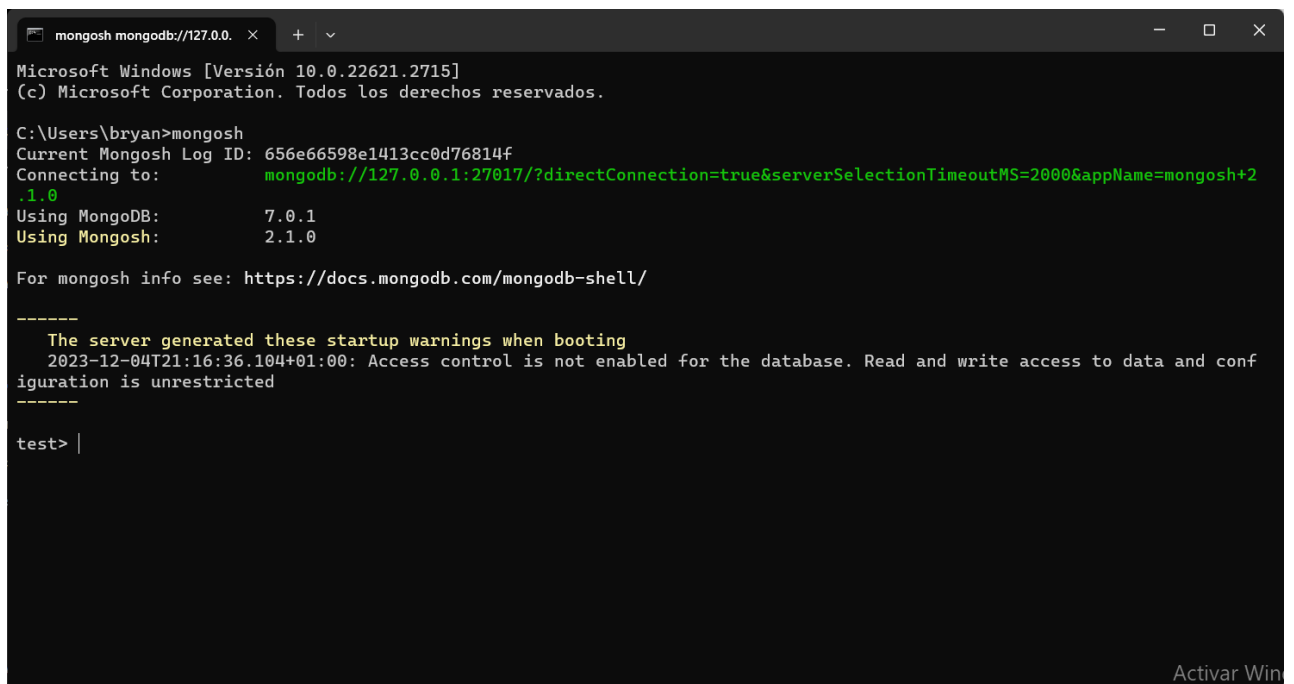
Se le asignó un nombre a la base de datos y luego un nombre a la colección que íbamos a trabajar en cuestión.



Como último paso de esta etapa le damos al botón de *Import* y con esto ya tenemos nuestra mesa de trabajo lista para iniciar con la exploración.

2. Ejercicios sobre actualización, proyección y filtrado

Para ponernos en contexto y simular lo más cercano posible a un entorno de trabajo real vamos realizar todas las consultas desde la terminal. Para esto ejecutamos el comando *mongosh* dentro de la terminal y tenemos acceso a *MongoDB Shell* que es la interfaz que nos comunica con MongoDB.



```
Microsoft Windows [Versión 10.0.22621.2715]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\bryan>mongosh
Current Mongosh Log ID: 656e66598e1413cc0d76814f
Connecting to:  mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+2.1.0
Using MongoDB: 7.0.1
Using Mongosh: 2.1.0

For mongosh info see: https://docs.mongodb.com/mongodb-shell/

-----
The server generated these startup warnings when booting
2023-12-04T21:16:36.104+01:00: Access control is not enabled for the database. Read and write access to data and configuration is unrestricted
-----

test> |
```

Una vez en este punto, tenemos que dirigirnos a la base de datos que importamos. Para esto es necesario primero pedirle a MongoDB que nos muestre las bases de datos que tiene alojadas, luego seleccionamos la de nuestro interés y finalmente mostramos 5 resultados para explorar los diferentes campos que se encuentran en cada documento.

```
show dbs
use data_salaries
db.salaries.find().limit(5)
```

Gracias a esta última consulta podemos notar varias cosas. Una de las más importantes es que todos los campos son de tipo string incluyendo los campos con valores numéricos (excepto *remote_ratio*), por lo que nos veremos en la obligación de modificar a enteros si queremos utilizar consultas que incluyan filtrado a través de esos campos. Otro punto interesante es descifrar las siglas dentro *experience_level* y *employment_type*. Utilizando comandos como *distinct*, *count* y *length* exploraremos más a fondo nuestra data como podemos observar a continuación.

```
mongosh mongodb://127.0.0.1:27020/
test> show dbs
admin          40.00 KiB
chess          8.00 KiB
clases         112.00 KiB
config         108.00 KiB
data_salaries  472.00 KiB
dsSalaries     208.00 KiB
entrega        156.00 KiB
local          72.00 KiB
test> use data_salaries
switched to db data_salaries
data_salaries> db.salaries.find().limit(5)
[
  {
    _id: ObjectId('656e5a253c89b74e177a1833'),
    work_year: '2023',
    experience_level: 'EX',
    employment_type: 'FT',
    job_title: 'Data Science Director',
    salary: '212000',
    salary_currency: 'USD',
    salary_in_usd: '212000',
    employee_residence: 'US',
    remote_ratio: 0,
    company_location: 'US',
    company_size: 'M'
  },
  {
    _id: ObjectId('656e5a253c89b74e177a1834'),
```

```
db.salaries.find().count()
db.salaries.distinct("experience_level")
db.salaries.distinct("employment_type")
db.salaries.distinct("company_location")
db.salaries.distinct("company_location").length
db.salaries.distinct("job_title").length
db.salaries.distinct("job_title")
```

```
data_salaries> db.salaries.find().count()
8805
data_salaries> db.salaries.distinct("experience_level")
[ 'EN', 'EX', 'MI', 'SE' ]
data_salaries> db.salaries.distinct("employment_type")
[ 'CT', 'FL', 'FT', 'PT' ]
data_salaries> db.salaries.distinct("company_location")
[
  'AD', 'AE', 'AM', 'AR', 'AS', 'AT', 'AU', 'BA',
  'BE', 'BR', 'BS', 'CA', 'CF', 'CH', 'CL', 'CN',
  'CO', 'CZ', 'DE', 'DK', 'DZ', 'EC', 'EE', 'EG',
  'ES', 'FI', 'FR', 'GB', 'GH', 'GI', 'GR', 'HK',
  'HN', 'HR', 'HU', 'ID', 'IE', 'IL', 'IN', 'IQ',
  'IR', 'IT', 'JP', 'KE', 'KR', 'LT', 'LU', 'LV',
  'MD', 'MT', 'MU', 'MX', 'MY', 'NG', 'NL', 'NO',
  'NZ', 'PH', 'PK', 'PL', 'PR', 'PT', 'QA', 'RO',
  'RU', 'SA', 'SE', 'SG', 'SI', 'TH', 'TR', 'UA',
  'US', 'ZA'
]
data_salaries> db.salaries.distinct("company_location").length
74
data_salaries> db.salaries.distinct("job_title").length
124
data_salaries> db.salaries.distinct("job_title")
[
  'AI Architect',
  'AI Developer',
  'AI Engineer',
  'AI Programmer',
  'AI Research Engineer',
  'AI Scientist',
  'AWS Data Architect',
  'Analytics Engineer',
  'Analytics Engineering Manager',
  'Applied Data Scientist',
```

Con estos resultados podemos arribar a las siguientes conclusiones según nuestro dataset:

- Tenemos un dataset de 8805 documentos
- Hay 74 países donde se encuentran empresas relacionadas con los datos
- Existen 124 títulos de trabajos diferentes en el mundo Data Science

experience_level:

Contract (CT): Empleo por una duración o proyecto específico.

Freelance (FL): Trabajador autónomo contratado para proyectos o tareas específicas.

Full Time (FT): Empleo tradicional a tiempo completo.

Part Time (PT): Empleo por menos horas que el estándar a tiempo completo.

employment_type:

Entry-Level (EN): Nivel de entrada

Experienced (EX): Experimentado

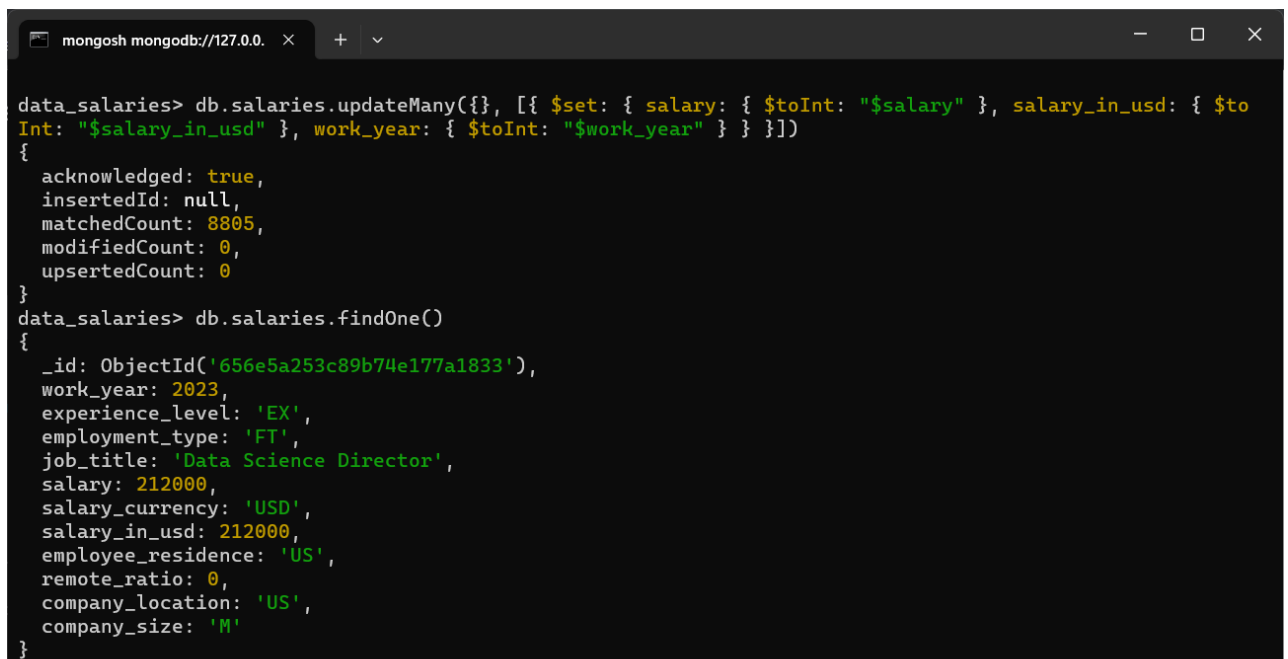
Mid-Level (MI): Nivel medio

Senior-Level (SE): Nivel senior

Lo siguiente que haremos será convertir los datos numéricos tipo string a tipo integer mediante una operación *"update"* seguido de un *findOne* para visualizar nuestro cambio:

```
db.salaries.updateMany({}, [{ $set: { salary: { $toInt: "$salary" }, salary_in_usd: { $toInt: "$salary_in_usd" }, work_year: { $toInt: "$work_year" } } }])
```

```
db.salaries.findOne()
```



```
mongosh mongodb://127.0.0.1:27027/ > data_salaries> db.salaries.updateMany({}, [{ $set: { salary: { $toInt: "$salary" }, salary_in_usd: { $toInt: "$salary_in_usd" }, work_year: { $toInt: "$work_year" } } }])
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 8805,
  modifiedCount: 0,
  upsertedCount: 0
}
data_salaries> db.salaries.findOne()
{
  _id: ObjectId('656e5a253c89b74e177a1833'),
  work_year: 2023,
  experience_level: 'EX',
  employment_type: 'FT',
  job_title: 'Data Science Director',
  salary: 212000,
  salary_currency: 'USD',
  salary_in_usd: 212000,
  employee_residence: 'US',
  remote_ratio: 0,
  company_location: 'US',
  company_size: 'M'
}
```

En la siguiente consulta queremos mostrar los 5 trabajos full remote y full time peor pagados, pero sólo proyectando tres campos de interés (el campo *_id* es necesario ocultarlo a la fuerza):

```
db.salaries.find({ employment_type: 'FT', remote_ratio: 100, experience_level: 'EN' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: 1 }).limit(5)
```

```
mongosh mongodb://127.0.0.1:27020 > use salaries
salaries> db.salaries.find({ employment_type: 'FT', remote_ratio: 100, experience_level: 'EN' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: 1 }).limit(5)
[
  {
    job_title: 'Machine Learning Developer',
    salary_in_usd: 15000,
    company_location: 'TH'
  },
  {
    job_title: 'Data Scientist',
    salary_in_usd: 16000,
    company_location: 'EC'
  },
  {
    job_title: 'Big Data Engineer',
    salary_in_usd: 16228,
    company_location: 'IN'
  },
  {
    job_title: 'Research Engineer',
    salary_in_usd: 16455,
    company_location: 'RU'
  },
  {
    job_title: 'Data Engineer',
    salary_in_usd: 17025,
    company_location: 'IN'
  }
]
salaries> |
```

Podemos notar que el salario peor pagado a tiempo completo y full remote, con el nivel mínimo de experiencia, es a un “Machine Learning Developer” y es en Tailandia con la cifra de 15,000 USD al año.

Ahora nos vamos a disponer a analizar los mejores salarios con estas mismas condiciones. Utilizamos la misma consulta pero ligeramente modificada (`{ salary_in_usd: -1 }`)

```
db.salaries.find({ employment_type: 'FT', remote_ratio: 100, experience_level: 'EN' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: -1 }).limit(5)
```

Podemos concluir que todos los salarios son iguales o superiores a 200,000 USD y además de los 5 primeros, 4 provienen de compañías en Estados Unidos.

Por último me gustaría saber cuál es el comportamiento solamente en España. Obviamente quitamos la restricción de que sea solamente full remote.

```
db.salaries.find({ employment_type: 'FT', experience_level: 'EN', company_location: 'ES' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: -1 })
```

```
mongosh mongodb://127.0.0.1:27017/
data_salaries> db.salaries.find({ employment_type: 'FT', remote_ratio: 100, experience_level: 'EN' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: -1 }).limit(5)
[
  {
    job_title: 'Research Scientist',
    salary_in_usd: 252990,
    company_location: 'US'
  },
  {
    job_title: 'Data Integration Specialist',
    salary_in_usd: 240000,
    company_location: 'US'
  },
  {
    job_title: 'Machine Learning Scientist',
    salary_in_usd: 225000,
    company_location: 'US'
  },
  {
    job_title: 'AI Developer',
    salary_in_usd: 215936,
    company_location: 'DE'
  },
  {
    job_title: 'Research Scientist',
    salary_in_usd: 200000,
    company_location: 'US'
  }
]
```

```
mongosh mongodb://127.0.0.1:27017/
data_salaries> db.salaries.find({ employment_type: 'FT', experience_level: 'EN', company_location: 'ES' }, { salary_in_usd: 1, job_title: 1, company_location: 1, _id: 0 }).sort({ salary_in_usd: -1 })
[
  {
    job_title: 'Data Engineer',
    salary_in_usd: 41027,
    company_location: 'ES'
  },
  {
    job_title: 'AI Scientist',
    salary_in_usd: 31520,
    company_location: 'ES'
  },
  {
    job_title: 'Data Scientist',
    salary_in_usd: 31520,
    company_location: 'ES'
  },
  {
    job_title: 'Data Science Consultant',
    salary_in_usd: 27317,
    company_location: 'ES'
  },
  {
    job_title: 'Business Data Analyst',
    salary_in_usd: 21593,
    company_location: 'ES'
  }
]
data_salaries> |
```

El rango de salario va desde 21,593 USD hasta 41,027 USD para un nivel de experiencia mínimo en España.

3. Ejercicios sobre pipeline de agregación

Usando el pipeline de agregación ahora podemos obtener consultas un poco más complicadas. Por ejemplo podemos agrupar, sacar promedios y hacer otras operaciones.

Agrupar por año y calcular el promedio de salarios

```
db.salaries.aggregate([ { $group: { _id: "$work_year", avgSalary: { $avg: "$salary_in_usd" } } }, { $sort: { _id: 1 } } ])
```

```
data_salaries> db.salaries.aggregate([ { $group: { _id: "$work_year", avgSalary: { $avg: "$salary_in_usd" } } }, { $sort: { _id: 1 } } ] )
[
  { _id: 2020, avgSalary: 102250.86666666667 },
  { _id: 2021, avgSalary: 99922.07339449541 },
  { _id: 2022, avgSalary: 134449.13809812235 },
  { _id: 2023, avgSalary: 155131.43206997085 }
]
```

Año 2020: El salario promedio es de aproximadamente 102,250.87 USD.

Año 2021: El salario promedio es de alrededor de 99,922.07 USD.

Año 2022: El salario promedio ha aumentado significativamente a unos 134,449.14 USD.

Año 2023: Hasta el momento, el salario promedio es de aproximadamente 155,131.43 USD.

Se puede palpar un aumento de los salarios conforme avanzan los años.

Calcular el rango salarial por nivel de experiencia

```
db.salaries.aggregate([ { $group: { _id: "$experience_level", maxSalary: { $max: "$salary_in_usd" }, minSalary: { $min: "$salary_in_usd" } } } ])
```

```
data_salaries> db.salaries.aggregate([ { $group: { _id: "$experience_level", maxSalary: { $max: "$salary_in_usd" }, minSalary: { $min: "$salary_in_usd" } } } ] )
[
  { _id: 'EX', maxSalary: 416000, minSalary: 15000 },
  { _id: 'MI', maxSalary: 450000, minSalary: 15000 },
  { _id: 'SE', maxSalary: 417937, minSalary: 15809 },
  { _id: 'EN', maxSalary: 615201, minSalary: 15000 }
]
```

Notamos que hay un valor anormal, pues con nivel de experiencia “entry level” tiene mayor salario que el resto, por lo que hacemos un borrado de este documento y volvemos a ejecutar la línea anterior para ver el cambio.

```
db.salaries.deleteOne({ experience_level: 'EN', salary_in_usd: 615201 })
```

```
data_salaries> db.salaries.deleteOne({ experience_level: 'EN', salary_in_usd: 615201 })
{ acknowledged: true, deletedCount: 1 }
```

```
data_salaries> db.salaries.aggregate([ { $group: { _id: "$experience_level", maxSalary: { $max: "$salary_in_usd" }, minSalary: { $min: "$salary_in_usd" } } } ] )
[
  { _id: 'SE', maxSalary: 417937, minSalary: 15809 },
  { _id: 'EN', maxSalary: 281700, minSalary: 15000 },
  { _id: 'EX', maxSalary: 416000, minSalary: 15000 },
  { _id: 'MI', maxSalary: 450000, minSalary: 15000 }
]
```

Sólo hay diferencia notable en el salario máximo en el nivel mínimo de experiencia. Los otros 3 niveles se encuentran bastante parejos. Incluso podemos notar que el salario máximo en mid level es mayor que el resto.

Promedio de salarios por país de compañía llevando una cuenta

```
db.salaries.aggregate([ { $group: { _id: "$company_location", avgSalary: { $avg: "$salary_in_usd" }, count: { $sum: 1 } } }, { $sort: { avgSalary: -1 } } ])
```

```
data_salaries> db.salaries.aggregate([ { $group: { _id: "$company_location", avgSalary: { $avg: "$salary_in_usd"
}, count: { $sum: 1 } } }, { $sort: { avgSalary: -1 } } ] )
[
  { _id: 'QA', avgSalary: 300000, count: 1 },
  { _id: 'IL', avgSalary: 217332, count: 3 },
  { _id: 'PR', avgSalary: 167500, count: 4 },
  { _id: 'US', avgSalary: 158329.92106652586, count: 7576 },
  { _id: 'CA', avgSalary: 141255.38048780488, count: 205 },
  { _id: 'SA', avgSalary: 134999, count: 2 },
  { _id: 'AU', avgSalary: 132282.70833333334, count: 24 },
  { _id: 'NZ', avgSalary: 125000, count: 1 },
  { _id: 'UA', avgSalary: 121333.33333333333, count: 6 },
  { _id: 'BA', avgSalary: 120000, count: 1 },
  { _id: 'IE', avgSalary: 115188.28571428571, count: 7 },
  { _id: 'JP', avgSalary: 110821.625, count: 8 },
  { _id: 'GB', avgSalary: 105983.15801886792, count: 424 },
  { _id: 'CN', avgSalary: 100000, count: 1 },
  { _id: 'MU', avgSalary: 100000, count: 1 },
  { _id: 'IQ', avgSalary: 100000, count: 1 },
  { _id: 'AE', avgSalary: 100000, count: 3 },
  { _id: 'DZ', avgSalary: 100000, count: 1 },
  { _id: 'IR', avgSalary: 100000, count: 1 },
  { _id: 'SE', avgSalary: 98791.33333333333, count: 3 }
]
```

QA (Qatar): Salario promedio de 300,000 USD con 1 empleado.

IL (Israel): Salario promedio de 217,332 USD con 3 empleados.

PR (Puerto Rico): Salario promedio de 167,500 USD con 4 empleados.

US (Estados Unidos): Salario promedio de aproximadamente 158,329.92 USD con 7,583 empleados.

CA (Canadá): Salario promedio de aproximadamente 141,255.38 USD con 205 empleados.

SA (Arabia Saudita): Salario promedio de 134,999 USD con 2 empleados.

AU (Australia): Salario promedio de aproximadamente 132,282.71 USD con 24 empleados.

NZ (Nueva Zelanda): Salario promedio de 125,000 USD con 1 empleado.

UA (Ucrania): Salario promedio de aproximadamente 121,333.33 USD con 6 empleados.

BA (Bosnia y Herzegovina): Salario promedio de 120,000 USD con 1 empleado.

IE (Irlanda): Salario promedio de aproximadamente 115,188.29 USD con 7 empleados.

JP (Japón): Salario promedio de aproximadamente 110,821.63 USD con 8 empleados.

GB (Reino Unido): Salario promedio de aproximadamente 105,983.16 USD con 424 empleados.

CN (China): Salario promedio de 100,000 USD con 1 empleado.

MU (Mauricio): Salario promedio de 100,000 USD con 1 empleado.

IQ (Irak): Salario promedio de 100,000 USD con 1 empleado.

AE (Emiratos Árabes Unidos): Salario promedio de 100,000 USD con 3 empleados.

DZ (Argelia): Salario promedio de 100,000 USD con 1 empleado.

IR (Irán): Salario promedio de 100,000 USD con 1 empleado.

SE (Suecia): Salario promedio de aproximadamente 98,791.33 USD con 3 empleados.

Concluimos que América del Norte (US y CA) tiene una presencia significativa con salarios promedio de alrededor de 158,329.92 USD en EE. UU. y 141,255.38 USD en Canadá con la mayor concentración de empleados.

Filtrar por nivel de experiencia y calcular la suma total de salarios

```
db.salaries.aggregate([ { $group: { _id: "$experience_level", totalSalary: { $sum: "$salary_in_usd" } } } ] )
```

```
data_salaries> db.salaries.aggregate([ { $group: { _id: "$experience_level", totalSalary: { $sum: "$salary_in_usd"
" } } } ] )
[
  { _id: 'SE', totalSalary: 1025728771 },
  { _id: 'EN', totalSalary: 40417524 },
  { _id: 'EX', totalSalary: 50855154 },
  { _id: 'MI', totalSalary: 198627529 }
]
```

Estos resultados representan la suma total de los salarios en USD para cada nivel de experiencia en tu conjunto de datos. Esta información nos hace comprender la distribución relativa de los salarios en cada categoría de experiencia:

Para 'MI' (Mid-Level), el total de salarios es de 198,627,529 USD.

Para 'SE' (Senior), el total de salarios es de 1,025,728,771 USD.
Para 'EN' (Entry-Level), el total de salarios es de 40,417,524 USD.
Para 'EX' (Executive), el total de salarios es de 50,855,154 USD.

Obtener el promedio de salario para un nivel de experiencia mínimo, full remote y full time

```
db.salaries.aggregate([ { $match: { experience_level: 'EN', employment_type: 'FT', remote_ratio: 100 } }, { $group: { _id: null, avgSalary: { $avg: "$salary_in_usd" } } } ] )
```

```
data_salaries> db.salaries.aggregate([ { $match: { experience_level: 'EN', employment_type: 'FT', remote_ratio: 100 } }, { $group: { _id: null, avgSalary: { $avg: "$salary_in_usd" } } } ] )
[ { _id: null, avgSalary: 75783.19553072625 } ]
```

El promedio de salario anual para un nivel de experiencia mínimo, full remote y full time es de **75,783.19 USD**

Obtener la distribución de salarios por rango

```
db.salaries.aggregate([ { $bucket: { groupBy: "$salary_in_usd", boundaries: [0, 20000, 50000, 100000, 150000, 200000, 300000, 400000, 500000], output: { count: { $sum: 1 } } } } ] )
```

```
data_salaries> db.salaries.aggregate([ { $bucket: { groupBy: "$salary_in_usd", boundaries: [0, 20000, 50000, 100000, 150000, 200000, 300000, 400000, 500000], output: { count: { $sum: 1 } } } } ] )
[
  { _id: 0, count: 33 },
  { _id: 20000, count: 313 },
  { _id: 50000, count: 1535 },
  { _id: 100000, count: 2862 },
  { _id: 150000, count: 2238 },
  { _id: 200000, count: 1612 },
  { _id: 300000, count: 202 },
  { _id: 400000, count: 9 }
]
```

Estos resultados indican la distribución de salarios según los compartimentos definidos. Aquí está la interpretación:

- Compartimento 0: Hay 33 empleados con salarios menores a 20,000.
- Compartimento 20,000: Hay 313 empleados con salarios entre 20,000 y 50,000.
- Compartimento 50,000: Hay 1,535 empleados con salarios entre 50,000 y 100,000.
- Compartimento 100,000: Hay 2,862 empleados con salarios entre 100,000 y 150,000.
- Compartimento 150,000: Hay 2,238 empleados con salarios entre 150,000 y 200,000.
- Compartimento 200,000: Hay 1,612 empleados con salarios entre 200,000 y 300,000.
- Compartimento 300,000: Hay 202 empleados con salarios entre 300,000 y 400,000.
- Compartimento 400,000: Hay 9 empleados con salarios entre 400,000 y 500,000.

La mayor cantidad de empleados tiene un salario anual entre 100,000 y 200,000 USD.

Calcular el porcentaje de salario en relación con el salario máximo por nivel de experiencia

```
db.salaries.aggregate([ { $group: { _id: "$experience_level", avgSalary: { $avg: "$salary_in_usd" }, maxSalary: { $max: "$salary_in_usd" } } }, { $project: { _id: 1, avgSalary: 1, maxSalary: 1, percentageOfMax: { $multiply: [ { $divide: ["$avgSalary", "$maxSalary"] }, 100 ] } } } ] )
```

```

data_salaries> db.salaries.aggregate([{$group: { _id: "$experience_level", avgSalary: { $avg: "$salary_in_usd" }
, maxSalary: { $max: "$salary_in_usd" } } }, { $project: { _id: 1, avgSalary: 1, maxSalary: 1, percentageOfMax: {
$multiply: [{ $divide: ["$avgSalary", "$maxSalary"] }, 100] } } }])
[
  {
    _id: 'MI',
    avgSalary: 114681.0213625866,
    maxSalary: 450000,
    percentageOfMax: 25.48467141390813
  },
  {
    _id: 'SE',
    avgSalary: 161889.01057449495,
    maxSalary: 417937,
    percentageOfMax: 38.73526645750315
  },
  {
    _id: 'EN',
    avgSalary: 86547.16059957174,
    maxSalary: 281700,
    percentageOfMax: 30.723166702013398
  },
  {
    _id: 'EX',
    avgSalary: 189052.61710037175,
    maxSalary: 416000,
    percentageOfMax: 45.44534064912782
  }
]

```

Estos resultados proporcionan una visión detallada de la distribución de salarios para cada nivel de experiencia, mostrando el salario promedio, el salario máximo y el porcentaje del salario promedio en relación con el salario máximo. Este análisis puede ser útil para entender cómo se comparan los salarios promedio con los salarios máximos en diferentes niveles de experiencia.

El salario promedio se mantiene siempre por debajo del 50% del salario máximo en todos los niveles de experiencia, acercándose más a este valor en el nivel de experiencia "EX".

4. Ejercicios sobre inserción

Como los ejercicios de inserción los considero puramente académicos y no me aportan valor al análisis realizado, los dejé para el final.

Inserción de un documento

```
db.salaries.insertOne({ work_year: '2023', experience_level: 'MI', employment_type: 'FT', job_title: 'Data Analyst', salary: '75000', salary_currency: 'USD', salary_in_usd: '75000', employee_residence: 'US', remote_ratio: 0, company_location: 'US', company_size: 'M'})
```

```
data_salaries> db.salaries.insertOne({ work_year: '2023', experience_level: 'MI', employment_type: 'FT', job_title: 'Data Analyst', salary: '75000', salary_currency: 'USD', salary_in_usd: '75000', employee_residence: 'US', remote_ratio: 0, company_location: 'US', company_size: 'M'})
{
  acknowledged: true,
  insertedId: ObjectId('657aeeb28e1413cc0d768150')
}
```

Insertar varios documentos al mismo tiempo

```
db.salaries.insertMany([
  {work_year:'2023',experience_level:'MI',employment_type:'FT',job_title:'Data Analyst',salary:'75000',salary_currency:'USD',salary_in_usd:'75000',employee_residence:'US',remote_ratio:0,company_location:'US',company_size:'M'},
  {work_year:'2023',experience_level:'SE',employment_type:'FT',job_title:'Machine Learning Engineer',salary:'100000',salary_currency:'USD',salary_in_usd:'100000',employee_residence:'US',remote_ratio:1,company_location:'US',company_size:'L'},
  {work_year:'2023',experience_level:'EN',employment_type:'FT',job_title:'Software Developer',salary:'85000',salary_currency:'USD',salary_in_usd:'85000',employee_residence:'US',remote_ratio:1,company_location:'US',company_size:'M'}])
```

```
data_salaries> db.salaries.insertMany([
  {work_year:'2023',experience_level:'MI',employment_type:'FT',job_title:'Data Analyst',salary:'75000',salary_currency:'USD',salary_in_usd:'75000',employee_residence:'US',remote_ratio:0,company_location:'US',company_size:'M'},
  {work_year:'2023',experience_level:'SE',employment_type:'FT',job_title:'Machine Learning Engineer',salary:'100000',salary_currency:'USD',salary_in_usd:'100000',employee_residence:'US',remote_ratio:1,company_location:'US',company_size:'L'},
  {work_year:'2023',experience_level:'EN',employment_type:'FT',job_title:'Software Developer',salary:'85000',salary_currency:'USD',salary_in_usd:'85000',employee_residence:'US',remote_ratio:1,company_location:'US',company_size:'M'}])
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId('657aef198e1413cc0d768151'),
    '1': ObjectId('657aef198e1413cc0d768152'),
    '2': ObjectId('657aef198e1413cc0d768153')
  }
}
```

Conclusiones

Resumen de los Resultados y Aspectos Destacados:

1. General:

- Existen 8805 documentos en nuestra colección.
- Hay 74 países dónde se encuentran empresas relacionadas con los datos.
- Existen 124 títulos de trabajos diferentes en el mundo Data Science.

La presencia de empresas relacionadas con datos en 74 países podría estar vinculada a la creciente demanda global de profesionales en Data Science. Empresas de diversos sectores y regiones reconocen la importancia de datos y análisis en sus operaciones.

2. Salarios Full Remote y Full Time para Entry-Level:

- Peor Pagado: Machine Learning Developer en Tailandia con 15,000 USD al año.
- Mejores Pagados: Todos superan los 200,000 USD, con 4 de 5 en compañías de Estados Unidos.

Las disparidades en los salarios Entry-Level pueden estar influenciadas por factores como el costo de vida en diferentes países, la demanda local de profesionales en Data Science y la madurez del mercado en esas regiones.

3. Salarios en España para Entry-Level y Full Time:

- Rango de salario: Desde 21,593 USD hasta 41,027 USD.

Los salarios En España dejan mucho que desear con respecto a los de Estados Unidos. La variación que se puede notar dentro de el propio país no es pequeña y esto podría reflejar diferencias en la demanda local, la presencia de empresas internacionales, y las condiciones económicas del país.

4. Análisis Temporal:

- Salario promedio ha aumentado de 102,250.87 USD (2020) a 155,131.43 USD (2023).
- Tendencia de aumento de salarios a lo largo de los años.

El aumento en el salario promedio a lo largo del tiempo podría indicar un crecimiento continuo en la demanda de profesionales en Data Science. La creciente importancia de la analítica de datos en diversas industrias podría contribuir a este aumento.

5. Distribución Geográfica de Salarios:

- Mayor concentración en América del Norte (EE. UU. y Canadá).
- Salario promedio en EE. UU.: Aproximadamente 158,329.92 USD.
- Salario promedio en Canadá: Aproximadamente 141,255.38 USD.

La concentración de oportunidades en América del Norte puede estar relacionada con la presencia de importantes empresas tecnológicas y financieras en la región. La demanda sostenida de expertos en datos en estos sectores es justificación segura para el impulso de los salarios.

6. Rango Salarial por Experiencia:

- Existe una anomalía en el salario máximo para Entry-Level, corregida eliminando el dato anormal.

7. Promedio de Salarios por País:

- Qatar (QA) tiene el salario promedio más alto: 300,000 USD con 1 empleado.
- Irlanda (IE) y Japón (JP) también tienen salarios promedio significativos.

Los salarios más altos en Qatar, Irlanda y Japón podrían estar relacionados con la presencia de industrias

específicas, como el sector financiero en Irlanda y la industria tecnológica en Japón. En Qatar, la alta remuneración podría ser atribuible al desarrollo económico y la demanda de talento especializado.

8. Distribución de Salarios por Rango:

- Mayor cantidad de empleados en el rango de 100,000 a 200,000 USD anuales.

La concentración en el rango de 100,000 a 200,000 USD sugiere que muchas empresas están dispuestas a pagar salarios competitivos para atraer y retener talento en Data Science, especialmente aquellos con cierta experiencia.

9. Porcentaje de Salario Promedio en Relación con el Salario Máximo:

- El salario promedio se mantiene generalmente por debajo del 50% del salario máximo en todos los niveles de experiencia.

El hecho de que el salario promedio generalmente se mantenga por debajo del 50% del salario máximo podría indicar diferencias significativas en la compensación dentro de los niveles de experiencia. Factores como la negociación salarial y las políticas internas de las empresas podrían contribuir a esta brecha.

Las mayores oportunidades de empleo con mejores salarios incluso para niveles de experiencia bajos se encuentran en América del Norte y los salarios de entrada en España dejan bastante que desear en comparación con los de Estados Unidos.

Es importante señalar que las variaciones pueden estar influenciadas por una amplia gama de factores y no solo los comentados anteriormente.