

# Evaluación Bases de Datos NoSQL

Santiago Torres Busquets

# Contenido

01

## Introducción

Contexto y enfoque del trabajo

02

## Dataset

Estructura y descripción

03

## Ejercicios

Inserción, actualización, proyección  
filtrado, agregación

04

## Conclusiones

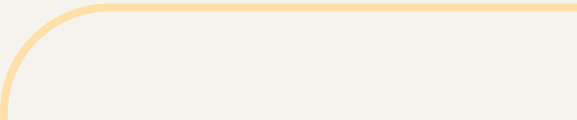
Análisis de resultados previos



01

# Introducción

El objetivo del presente trabajo consistió en seleccionar un dataset, importar las colecciones a MongoDB y realizar un análisis de la información en base a diversas consultas. En mi caso elegí una base de datos que contiene información histórica de los Juegos Olímpicos, tanto de verano como de invierno, desde sus comienzos en Atenas 1896 hasta Rio 2016.



La fuente de los datos puede encontrarse en:

<https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results>

Este dataset contiene registros de la participación de atletas en los JJOO modernos desde los inicios en el año 1896 en Atenas hasta Rio 2016.

El dataset contiene 2 ficheros:

- athletes\_events.csv: este fichero es el principal y contiene 271.116 registros y 15 campos
- noc\_regions.csv: este fichero es secundario y contiene 230 registros relacionando el nombre del NOC (National Olympic Committee, es decir el código de 3 letras del Comité Olímpico) con la región (país) al que representa y un campo de notas o comentarios.

La importación del dataset se realizó a través de **import** -> **JSON and CSV files** en NoSQLBooster

```
{
  "_id" : ObjectId("637367a4c38e976badfe4f95"),
  "ID" : 1262,
  "Name" : "Sandra Aguilar Navarro",
  "Sex" : "F",
  "Age" : 23,
  "Height" : 167,
  "Weight" : 50,
  "Team" : "Spain",
  "NOC" : "ESP",
  "Games" : "2016 Summer",
  "Year" : 2016,
  "Season" : "Summer",
  "City" : "Rio de Janeiro",
  "Sport" : "Rhythmic Gymnastics",
  "Event" : "Rhythmic Gymnastics Women's Group",
  "Medal" : "Silver"
},
```

Ejemplo de la estructura del fichero  
"athlete\_events" en JSON

db.athlete\_events.find()

db.noc\_regions.find()

```
{
  "_id" : ObjectId("6368e36e8dc843225197aa69"),
  "NOC" : "SCG",
  "region" : "Serbia",
  "notes" : "Serbia and Montenegro"
},
```

Ejemplo de la estructura del fichero "noc\_regions"  
en JSON

## 03

# Ejercicios y Queries

- El dataset presenta los campos sin información con el valor "NA". Reemplazo de campos Age, Height, Weight y Medal de valor "NA", con valor null

```
db.athlete_events.updateMany({Age: "NA"}, {$set: {Age: null}})
db.athlete_events.updateMany({Height: "NA"}, {$set: {Height: null}})
db.athlete_events.updateMany({Weight: "NA"}, {$set: {Weight: null}})
db.athlete_events.updateMany({Medal: "NA"}, {$set: {Medal: null}})
```

```
{
  "_id" : ObjectId("6376a06bdf6833c62d507e9d"),
  "ID" : 54,
  "Name" : "Mohamed Jamshid Abadi",
  "Sex" : "M",
  "Age" : "NA",
  "Height" : "NA",
  "Weight" : "NA",
  "Team" : "Iran",
  "NOC" : "IRI",
  "Games" : "1948 Summer",
  "Year" : 1948,
  "Season" : "Summer",
  "City" : "London",
  "Sport" : "Boxing",
  "Event" : "Boxing Men's Heavyweight",
  "Medal" : "NA"
}
```



```
{
  "_id" : ObjectId("637367a4c38e976badfe4740"),
  "ID" : 54,
  "Name" : "Mohamed Jamshid Abadi",
  "Sex" : "M",
  "Age" : null,
  "Height" : null,
  "Weight" : null,
  "Team" : "Iran",
  "NOC" : "IRI",
  "Games" : "1948 Summer",
  "Year" : 1948,
  "Season" : "Summer",
  "City" : "London",
  "Sport" : "Boxing",
  "Event" : "Boxing Men's Heavyweight",
  "Medal" : null
},
```

## 03

# Ejercicios y Queries

- Conteo de atletas/deportistas participantes de los JJOO desde 1896 a 2016. Utilizo el campo "ID" (propio, no ObjectId) y no el campo "Name" para agrupar los registros en caso de atletas que compartan el mismo nombre.

```
db.athlete_events.aggregate({$group: { _id: "$ID"}}).count("Atletas")
```

- Cantidad de participantes según sexo. F: Female, M: Male

```
db.athlete_events.aggregate({$match: {Sex: "F"}},{$group: {_id: "$ID"}}).count("Female Athletes")
db.athlete_events.aggregate({$match: {Sex: "M"}},{$group: {_id: "$ID"}}).count("Male Athletes")
```

- Número de ediciones de Juegos Olímpicos de **verano** desde 1896 a 2016

```
db.athlete_events.aggregate([{$match: {Season: "Summer"}}, { $group: { _id: "$Year" } }]).count("JJOO de Verano")
```

- Número de ediciones de Juegos Olímpicos de **verano** desde 1896 a 2016

```
db.athlete_events.aggregate([{$match: {Season: "Winter"}}, { $group: { _id: "$Year" } }]).count("JJOO de Invierno")
```

## Resultado de Queries



```
{
  "Atletas" : 135571
}
```

```
{
  "Female Athletes" : 33981
}
{
  "Male Athletes" : 101590
}
```

```
{
  "JJOO de Verano" : 29
}
```

```
{
  "JJOO de Invierno" : 22
}
```

# 03

## Ejercicios y Queries

- Promedio de edad** de los competidores olímpicos, en años.

```
db.athlete_events.aggregate({$group : {_id: "Age", edadPromedio:
{$avg: "$Age" }}}})
```



_id	edadPromedio
Age	25,5569

- Top 3 Atletas olímpico más jóvenes en participar.** Se excluye a los que tienen age: *null*.

```
db.athlete_events.find({ Age: {$ne: null }}).sort({ Age:1}).project({_id:0}).limit(3)
```

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
71691	Dimitrios Loundras	M	10	null	null	Ethnikos	GRE	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics	Bronze
40129	Luigina Giavotti	F	11	null	null	Italy	ITA	1928 Summer	1928	Summer	Amsterdam	Gymnastics	Gymnastics	Silver
37333	Carlos Bienvenido Fror	M	11	null	null	Spain	ESP	1992 Summer	1992	Summer	Barcelona	Rowing	Rowing Men	null

- Top 3 Atletas olímpicos más jóvenes en ganar una medalla.** Se excluye a los que tienen age: *null*.

```
db.athlete_events.find({ Age: {$ne: null}, Medal: {$ne: null}}).sort({ Age:1}) .project({_id:0}).limit(3)
```

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
71691	Dimitrios Loundras	M	10	null	null	Ethnikos	GRE	1896 Summer	1896	Summer	Athina	Gymnastics	Gymnastics	Bronze
40129	Luigina Giavotti	F	11	null	null	Italy	ITA	1928 Summer	1928	Summer	Amsterdam	Gymnastics	Gymnastics	Silver
74712	Carla Marangoni	F	12	null	null	Italy	ITA	1928 Summer	1928	Summer	Amsterdam	Gymnastics	Gymnastics	Silver



# 03

## Ejercicios y Queries

- **Top 3 Atletas olímpicos más viejos en participar.** Excluimos a los que tienen age: null.

Aclaración: son competidores en la categoría *pintura* o *escultura*, que eran competencias olímpicas en las primeras ediciones de los JJOO. Los mismos fueron premiados póstumamente, por eso las edades.

```
db.athlete_events.find({ Age: {$ne: null} }).sort({ Age:-1 }).project({_id:0}).limit(3)
```

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
128719	John Quincy Adams Williams	M	97	null	null	United States	USA	1928 Summer	1928	Summer	Amsterdam	Art Competitions	Art Competitions	null
49663	Winslow Homer	M	96	null	null	United States	USA	1932 Summer	1932	Summer	Los Angeles	Art Competitions	Art Competitions	null
31173	Thomas Cowperthwaite	M	88	null	null	United States	USA	1932 Summer	1932	Summer	Los Angeles	Art Competitions	Art Competitions	null

- **Top 3 Atletas olímpicos más viejos en ganar una medalla.**

```
db.athlete_events.find({ Age: {$ne: null}, Medal: {$ne: null}}).sort({ Age:-1 }).project({_id:0}).limit(3)
```

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
22984	John (Herbert) Crawford	M	73	null	null	Great Britain	GBR	1948 Summer	1948	Summer	London	Art Competitions	Art Competitions	Silver
117046	Oscar Gomer Swahn	M	72	null	null	Sweden	SWE	1920 Summer	1920	Summer	Antwerpen	Shooting	Shooting Men	Silver
30731	Jozu Dupon	M	72	null	null	Belgium	BEL	1936 Summer	1936	Summer	Berlin	Art Competitions	Art Competitions	Bronze

- Atletas con más **participaciones** (en nº de disciplinas) en JJOO.

```
db.athlete_events.aggregate({$group: { _id: "$Name", participaciones: {$sum: 1}}}
.sort({"participaciones": -1}).limit(3)
```

_id	participaciones
Robert Tait McKenzie	58
Heikki Ilmari Savolainen	39
Joseph "Josy" Stoffel	38

## 03

# Ejercicios y Queries

- **Top 10:** Ranking de **medallas por país**. Realicé *lookup* con la colección “noc\_events” para cambiar el código de 3 letras “NOC” (National Olympic Committee) por el nombre de la región/país.

```
var join = {$lookup:
  {
    from: 'noc_regions',
    localField: 'NOC',
    foreignField: 'NOC',
    as: 'OlympicComittee'
  }
}
var medalla14 = {$match: {Medal: {$ne: null}}}
var grupo14 = {$group: { _id:
"$OlympicComittee.region" ,total_medallas: {$sum:
1}}}}
var unwind = {$unwind: "$OlympicComittee"}
var sort = {$sort: {total_medallas: -1}}
db.athlete_events.aggregate([join,unwind,medalla14
,grupo14,sort]).limit(10)
```



_id ↕	total_medallas ↕
USA	5637 (5.6K)
Russia	3947 (3.9K)
Germany	3756 (3.8K)
UK	2068 (2.1K)
France	1777 (1.8K)
Italy	1637 (1.6K)
Sweden	1536 (1.5K)
Canada	1352 (1.4K)
Australia	1349 (1.3K)
Hungary	1135 (1.1K)

# 03

## Ejercicios y Queries

- Top 10 **Deportes** o disciplinas con más **atletas** a lo largo de la historia.
- Top 5 **Deportes** con **mayor altura promedio** -> En cm

```
db.athlete_events.aggregate({$group: { _id: "$Sport", conteo_atletas: {$sum: 1}}}).sort({conteo_atletas: -1}).limit(10)
```



	_id ↕	conteo_atletas ↕
1	Athletics	38.624 (38.6K)
2	Gymnastics	26.707 (26.7K)
3	Swimming	23.195 (23.2K)
4	Shooting	11.448 (11.4K)
5	Cycling	10.859 (10.9K)
6	Fencing	10.735 (10.7K)
7	Rowing	10.595 (10.6K)
8	Cross Country Skiing	9133 (9.1K)
9	Alpine Skiing	8829 (8.8K)
10	Wrestling	7154 (7.2K)

```
db.athlete_events.aggregate({$match: {Height: {$ne: null}}}, {$group: { _id: {Sport: "$Sport"}, altura: {$avg: "$Height"}}}).sort({"altura": -1}).limit(5)
```

	_id ↕	altura ↕
1	{ Sport : "Basketball" }	190,8699
2	{ Sport : "Volleyball" }	186,9948
3	{ Sport : "Beach Volleyb" }	186,145
4	{ Sport : "Water Polo" }	184,8346
5	{ Sport : "Rowing" }	184,1722

- Top 5 **Deportistas** mas **altos** en competir en JJOO. Se muestra nombre, equipo y deporte

```
db.athlete_events.aggregate({$match: {Height: {$ne: null}}}, {$group: { _id: {Name: "$Name", Team: "$Team", Sport: "$Sport"}, altura: {$max: "$Height"}}}).sort({"altura": -1}).limit(5)
```

(1)	{ Name : "Yao Ming", Team : "China", Sport : "Basketball" }	{ altura : 226 }
(2)	{ Name : "Tommy Loren Burleson", Team : "United States", Sport : "Basketball" }	{ altura : 223 }
(3)	{ Name : "Arvydas Romas Sabonis", Team : "Soviet Union", Sport : "Basketball" }	{ altura : 223 }
(4)	{ Name : "Arvydas Romas Sabonis", Team : "Lithuania", Sport : "Basketball" }	{ altura : 223 }
(5)	{ Name : "Zhang Zhaoxu", Team : "China", Sport : "Basketball" }	{ altura : 221 }

## 03

## Ejercicios y Queries

- Top 5 **Deportistas** mas **bajos** en competir en JJOO. -> En cm. Se muestra nombre, equipo y deporte

```
db.athlete_events.aggregate({$match:
{Height: {$ne: null}}, { $group: { _id:
{Name: "$Name",Team: "$Team",Sport:
"$Sport"} , altura: {$min:
"$Height"}}}).sort({"altura": 1}).limit(5)
```



(1) { Name : "Rosario Briones", Team : "Mexico", Sport : "Gymnastics" }	{ altura : 127 }
(2) { Name : "Lyton Levison Mphande", Team : "Malawi", Sport : "Boxing" }	{ altura : 127 }
(3) { Name : "Helman Palije", Team : "Malawi", Sport : "Boxing" }	{ altura : 128 }
(4) { Name : "Boston Simbeye", Team : "Malawi", Sport : "Boxing" }	{ altura : 130 }
(5) { Name : "Salvador Miranda", Team : "Nicaragua", Sport : "Boxing" }	{ altura : 130 }

- Top 5 **Deportes** con mayor **peso** en **promedio** -> En kg

```
db.athlete_events.aggregate({$match:
{Weight: {$ne: null}}, { $group: { _id:
{Sport: "$Sport",Event: "$Event"} , peso:
{$avg: "$Weight"}}}).sort({"peso": -
1}).limit(5)
```



(1) { Sport : "Weightlifting", Event : "Weightlifting Men's Super-Heavyweight" }	{ peso : 135.40862944162436 }
(2) { Sport : "Wrestling", Event : "Wrestling Men's Super-Heavyweight, Greco-" }	{ peso : 122.20320855614973 }
(3) { Sport : "Judo", Event : "Judo Men's Heavyweight" }	{ peso : 120.24754098360656 }
(4) { Sport : "Wrestling", Event : "Wrestling Men's Super-Heavyweight, Freest" }	{ peso : 118.75661375661376 }
(5) { Sport : "Athletics", Event : "Athletics Men's Shot Put" }	{ peso : 115.26989619377163 }

- Top 5 **Deportistas** mas **pesados** en competir en todos los JJOO -> En kg

```
db.athlete_events.aggregate({$match:
{Weight: {$ne: null}}, { $group: { _id:
{Name: "$Name", Team: "$Team", Sport:
"$Sport"} , peso: {$max:
"$Weight"}}}).sort({"peso": -1}).limit(5)
```



(1) { Name : "Ricardo Blas, Jr.", Team : "Guam", Sport : "Judo" }	{ peso : 214 }
(2) { Name : "Aytami Ruano Vega", Team : "Spain", Sport : "Judo" }	{ peso : 198 }
(3) { Name : "Marek Galiski", Team : "Poland", Sport : "Wrestling" }	{ peso : 190 }
(4) { Name : "Christopher J. \"Chris\" Taylor", Team : "United States", Sport : "Wrestling" }	{ peso : 182 }
(5) { Name : "Valentyn Rusliakov", Team : "Ukraine", Sport : "Judo" }	{ peso : 180 }

# 03 Ejercicios y Queries

- Top 5 **Deportistas** mas **livianos** en competir en todos los JJOO -> En kg

```
db.athlete_events.aggregate({$match:
{Weight: {$ne: null}}}, {$group: { _id:
{Name: "$Name", Team: "$Team", Sport:
"$Sport"} , peso: {$min:
"$Weight"}}}).sort({"peso": 1}).limit(5)
```



(1) { Name : "Choi Myong-Hui", Team : "North Korea", Sport : "Gymnastics" }	{ peso : 25 }
(2) { Name : "Kana Yamawaki", Team : "Japan", Sport : "Gymnastics" }	{ peso : 28 }
(3) { Name : "Albert Ferdinand \"Al\" Zerhusen", Team : "United States", Sport : "Football" }	{ peso : 28 }
(4) { Name : "Wang Xin (Ruoxue-)", Team : "China", Sport : "Diving" }	{ peso : 28 }
(5) { Name : "Anita Jokiel", Team : "Poland", Sport : "Gymnastics" }	{ peso : 28 }

- Conteo** de atletas en cada edición de JJOO, de mayor a menor

```
db.athlete_events.aggregate({$group: {_id:
{Año: "$Year", City: "$City"},CantidadAtletas:
{$sum: 1}}}).sort({"CantidadAtletas": -1})
```

(1) { "Año" : 2000, City : "Sydney" }	{ CantidadAtletas : 13821 }
(2) { "Año" : 1996, City : "Atlanta" }	{ CantidadAtletas : 13780 }
(3) { "Año" : 2016, City : "Rio de Janeiro" }	{ CantidadAtletas : 13688 }
(4) { "Año" : 2008, City : "Beijing" }	{ CantidadAtletas : 13602 }
(5) { "Año" : 2004, City : "Athina" }	{ CantidadAtletas : 13443 }
(6) { "Año" : 1992, City : "Barcelona" }	{ CantidadAtletas : 12977 }
(7) { "Año" : 2012, City : "London" }	{ CantidadAtletas : 12920 }
(8) { "Año" : 1988, City : "Seoul" }	{ CantidadAtletas : 12037 }
(9) { "Año" : 1972, City : "Munich" }	{ CantidadAtletas : 10304 }
(10) { "Año" : 1984, City : "Los Angeles" }	{ CantidadAtletas : 9454 }

- Ediciones de JJOO** con mayor participación de hombres y mujeres

```
db.athlete_events.aggregate({$
match: {Sex: "M"}}{$group:
{_id: "$Year",hombres: {$sum:
1}}).sort({"hombres": -1})
```

```
db.athlete_events.aggregate({$
match: {Sex: "F"}}{$group:
{_id: "$Year",mujeres: {$sum:
1}}).sort({"mujeres": -1})
```

(1) 1992	{ hombres : 11235 }	(1) 2016	{ mujeres : 6223 }
(2) 1988	{ hombres : 10453 }	(2) 2008	{ mujeres : 5816 }
(3) 1972	{ hombres : 9351 }	(3) 2012	{ mujeres : 5815 }
(4) 1996	{ hombres : 8772 }	(4) 2004	{ mujeres : 5546 }
(5) 1984	{ hombres : 8605 }	(5) 2000	{ mujeres : 5431 }
(6) 2000	{ hombres : 8390 }	(6) 1992	{ mujeres : 5178 }
(7) 1968	{ hombres : 8286 }	(7) 1996	{ mujeres : 5008 }
(8) 2004	{ hombres : 7897 }	(8) 1988	{ mujeres : 4223 }
(9) 1976	{ hombres : 7896 }	(9) 1984	{ mujeres : 2983 }
(10) 2008	{ hombres : 7786 }	(10) 1972	{ mujeres : 2608 }

# 03

## Ejercicios y Queries

- ¿Qué **ciudades** fueron sede de JJOO?

```
db.athlete_events.aggregate({$group: { _id: "$City"}}).sort("_id": 1)
```

	_id		_id		_id
1	Albertville	15	Innsbruck	28	Rio de Janeiro
2	Amsterdam	16	Lake Placid	29	Roma
3	Antwerpen	17	Lillehammer	30	Salt Lake City
4	Athina	18	London	31	Sankt Moritz
5	Atlanta	19	Los Angeles	32	Sapporo
6	Barcelona	20	Melbourne	33	Sarajevo
7	Beijing	21	Mexico City	34	Seoul
8	Berlin	22	Montreal	35	Sochi
9	Calgary	23	Moskva	36	Squaw Valley
10	Chamonix	24	Munich	37	St. Louis
11	Cortina d'Ampezzo	25	Nagano	38	Stockholm
12	Garmisch-Partenkirchen	26	Oslo	39	Sydney
13	Grenoble	27	Paris	40	Tokyo
14	Helsinki			41	Torino
				42	Vancouver

- Conteo de **deportes** y **eventos**

```
db.athlete_events.aggregate({$group: { _id: "$Sport"}}).count("Cantidad de Deportes")
```

```
{
  "Cantidad de Deportes" : 66
}
```

```
db.athlete_events.aggregate({$group: { _id: "$Event"}}).count("Cantidad de Disciplinas")
```

```
{
  "Cantidad de Disciplinas" : 765
}
```

# 03

## Ejercicios y Queries

- **Top 5 Atletas** con más **medallas** → **Doradas**

```
var match_gold = {$match: {Medal: "Gold"}}
var group_gold = {$group: {_id: "$Name", gold_medals: {$sum: 1}}}
db.athlete_events.aggregate(match_gold,group_gold).sort({gold_medals:-1}).limit(5)
```

- **Top 5 Atletas** con más **medallas** → **Plata**

```
var match_silver = {$match: {Medal: "Silver"}}
var group_silver = {$group: {_id: "$Name", silver_medals: {$sum: 1}}}
db.athlete_events.aggregate(match_silver,group_silver).sort({silver_medals:-1}).limit(5)
```

- **Top 5 Atletas** con más **medallas** → **Bronce**

```
var match_bronze = {$match: {Medal: "Bronze"}}
var group_bronze = {$group: {_id: "$Name", bronze_medals: {$sum: 1}}}
db.athlete_events.aggregate(match_bronze,group_bronze).sort({bronze_medals:-1}).limit(5)
```

	_id ↕	gold_medals ↕
1	Michael Fred Phelps, II	23
2	Raymond Clarence "Ray" Ewry	10
3	Mark Andrew Spitz	9
4	Frederick Carlton "Carl" Lewis	9
5	Paavo Johannes Nurmi	9

	_id ↕	silver_medals ↕
1	Shirley Frances Babashoff	6
2	Aleksandr Nikolayevich Dityatin	6
3	Yang Yang	6
4	Aleksandr Vladimirovich Popov	6
5	Mikhail Yakovlevich Voronin	6

	_id ↕	bronze_medals ↕
1	Franziska van Almsick	6
2	Merlene Joyce Ottey-Page	6
3	Harri Tapani Kirvesniemi	6
4	Heikki Ilmari Savolainen	6
5	Aleksey Yuryevich Nemov	6

# 03

## Ejercicios y Queries

- **Top 10 Países** con más **medallas** → **Doradas**

```
var join = {$lookup:
  {
    from: 'noc_regions',
    localField: 'NOC',
    foreignField: 'NOC',
    as: 'OlympicComittee'
  }
}
var unwind = {$unwind: "$OlympicComittee"}
var match23g = {$match: {Medal: {$eq: "Gold"}}}
var grupo23g = {$group: { _id: "$OlympicComittee.region" ,doradas: {$sum: 1}}}
var sort23g = {$sort: {doradas: -1}}
db.athlete_events.aggregate([join,unwind,match23g,grupo23g,sort23g]).limit(10)
```



	_id	doradas
1	USA	2638 (2.6K)
2	Russia	1599 (1.6K)
3	Germany	1301 (1.3K)
4	UK	678
5	Italy	575
6	France	501
7	Sweden	479
8	Canada	463
9	Hungary	432
10	Norway	378

- **Top 10 Países** con más **medallas** → **Plata**. Reutilizo en el código el lookup o join creado anteriormente.

```
var unwind = {$unwind: "$OlympicComittee"}
var match23s = {$match: {Medal: {$eq: "Silver"}}}
var grupo23s = {$group: { _id: "$OlympicComittee.region" ,plata: {$sum: 1}}}
var sort23s = {$sort: {plata: -1}}
db.athlete_events.aggregate([join,unwind,match23s,grupo23s,sort23s]).limit(10)
```



	_id	plata
1	USA	1641 (1.6K)
2	Germany	1195 (1.2K)
3	Russia	1170 (1.2K)
4	UK	739
5	France	610
6	Italy	531
7	Sweden	522
8	Australia	459
9	Canada	438
10	Norway	361



# 03

## Ejercicios y Queries

- **Top 5 Países** con más **medallas** → **Bronze**. Reutilizo el lookup/join creado anteriormente.

```
var unwind = {$unwind: "$OlympicCommittee"}
var match23b = {$match: {Medal: {$eq: "Bronze"}}}
var grupo23b = {$group: { _id: "$OlympicCommittee.region" ,bronze: {$sum: 1}}}
var sort23b = {$sort: {bronze: -1}}
db.athlete_events.aggregate([join,unwind,match23b,grupo23b,sort23b]).limit(10)
```



	_id	bronze
1	USA	1358 (1.4K)
2	Germany	1260 (1.3K)
3	Russia	1178 (1.2K)
4	France	666
5	UK	651
6	Sweden	535
7	Italy	531
8	Australia	522
9	Canada	451
10	Finland	432

- **Top 5 Países** con mayor **altura promedio** (en cm). Reutilizo el join creado anteriormente

```
var unwind = {$unwind: "$OlympicCommittee"}
var match24 = {$match: {Height: {$ne: null}}}
var grupo24 = {$group: { _id: "$OlympicCommittee.region" ,altura: {$avg: "$Height"}}}
var sort24 = {$sort: {altura: -1}}
db.athlete_events.aggregate([join,unwind,match24,grupo24,sort24]).limit(5)
```



	_id	altura
1	Croatia	183,6132
2	Montenegro	183,5638
3	Lithuania	182,0995
4	Iceland	181,2454
5	Serbia	180,8042

- **Top 5 Países** con mayor **peso promedio** (en kg). Reutilizo el join creado anteriormente

```
var unwind = {$unwind: "$OlympicCommittee"}
var match25 = {$match: {Weight: {$ne: null}}}
var grupo25 = {$group: { _id: "$OlympicCommittee.region" ,peso: {$avg: "$Weight"}}}
var sort25 = {$sort: {peso: -1}}
db.athlete_events.aggregate([join,unwind,match25,grupo25,sort25]).limit(5)
```



	_id	peso
1	Nauru	95,8462
2	American Samoa	87,9583
3	Montenegro	84,8298
4	Tonga	82,7556
5	Samoa	81,6442

# 03

## Ejercicios y Queries

- **Top 5 Países con menor altura promedio** (en cm). Reutilizo el join creado anteriormente

```
var unwind = {$unwind: "$OlympicCommittee"}
var match23b = {$match: {Medal: {$eq: "Bronze"}}}
var grupo23b = {$group: { _id: "$OlympicCommittee.region" ,bronze: {$sum: 1}}}
var sort23b = {$sort: {bronze: -1}}
db.athlete_events.aggregate([join,unwind,match23b,grupo23b,sort23b]).limit(10)
```



	_id ↕	altura ↕
1	Micronesia	161,24
2	Timor-Leste	161,4444
3	North Korea	161,558
4	Maldives	163,7619
5	Nepal	163,9059

- **Top 5 Países con menor peso promedio** (en kg). Reutilizo el join creado anteriormente

```
var unwind = {$unwind: "$OlympicCommittee"}
var match27 = {$match: {Weight: {$ne: null}}}}
var grupo27 = {$group: { _id: "$OlympicCommittee.region" ,peso: {$avg:
"$Weight"}}}}
var sort27 = {$sort: {peso: 1}}
db.athlete_events.aggregate([join,unwind,match27,grupo27,sort27]).limit(5)
```



	_id ↕	peso ↕
1	Maldives	54,9762
2	Timor-Leste	55,5
3	Eritrea	56,0541
4	Ethiopia	56,9407
5	North Korea	57,3145

## 04

# Conclusiones

- Se puede observar un aumento exponencial de participantes, tanto de hombres y mujeres. Las primeras ediciones de los JJOO convocaban a menos de mil participantes y las ultimas ediciones ya contaban con la participación de más de diez mil.
- Hasta 2016 (año que abarca este dataset), se celebraron 29 JJOO de verano y 22 de Invierno. Con 42 ciudades siendo sede de los mismos combinando juegos de invierno y verano. Pasando históricamente por un total de 135 mil atletas, a través de 66 disciplinas y 765 subdisciplinas
- La edad promedio histórica de los participantes se sitúa en los 25 años. Como casos excepcionales encontramos a deportistas llegando a ganar con edades como 10 años y hasta 72 años.
- Encontramos al basketball como el deporte con jugadores más altos en promedio, y con el jugador más alto de la historia de los JJOO.
- El deporte más practicado a lo largo de los años por la cantidad de subdisciplinas son el atletismo y la gimnasia, seguido por la natación
- El deportista con récord de mayor cantidad de medallas de oro es el estadounidense Michael Phelps y EEUU es el país con más medallas en total.
- En promedio, los países europeos son los que presentan mayor altura, y en Asia del este los de menor altura y menor peso también. Los deportistas más pesados en promedio se encuentran en islas de Oceanía



# Gracias

---

Master Big Data & Data Science

Santiago Torres Busquets

