

## Fuente de los datos:

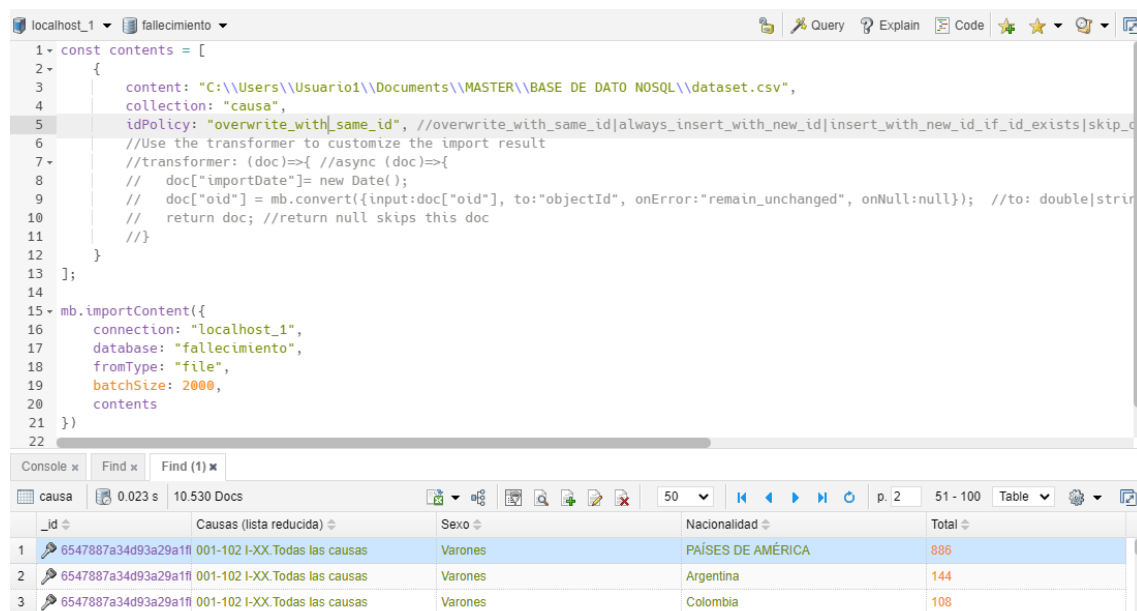
<https://datos.gob.es/en/catalogo/ea0010587-defunciones-por-causas-lista-reducida-sexo-y-nacionalidad-ecm-identificador-api-t15-p417-a2009-l0-01002b-px1>

## Descripción:

Se trata de una base de datos que clasifica las causas de defunción por género y por nacionalidad.

Contiene un total de 21.060 documentos.

## Importación de la data:



The screenshot shows the MongoDB Compass interface. The top pane displays a JavaScript script for importing data from a CSV file into a MongoDB collection. The script uses the `mb.importContent` function with the following configuration:

- `connection`: "localhost\_1"
- `database`: "fallecimiento"
- `fromType`: "file"
- `batchSize`: 2000

The bottom pane shows a table of data with the following columns: `_id`, `Causas (lista reducida)`, `Sexo`, `Nacionalidad`, and `Total`. The table contains three rows of data:

	_id	Causas (lista reducida)	Sexo	Nacionalidad	Total
1	6547887a34d93a29a1f1	001-102 I-XX.Todas las causas	Varones	PAÍSES DE AMÉRICA	886
2	6547887a34d93a29a1f1	001-102 I-XX.Todas las causas	Varones	Argentina	144
3	6547887a34d93a29a1f1	001-102 I-XX.Todas las causas	Varones	Colombia	108

## Estructura de los datos:

Los documentos presentan 5 campos cada uno, estos campos son:

- Id único para cada uno de los documentos.
- Causa de fallecimiento: este valor se repite para cada valor de la variable género y para cada valor de la variable nacionalidad.
- Sexo: el género de estudio, se divide en "Varones", "Mujeres", "Ambos"
- Nacionalidad: con un valor diferenciado para cada continente (además de los propios países).
- Total: la cuantía de cada causa en función de su género.

causa	0.030 s	21.060 Docs		50		p. 1	1 - 50	Tr
Key	Value	Type						
(1) 6547887a34d93a29a1fbce05	{ } (5 fields)	Document						
_id	6547887a34d93a29a1fbce05	ObjectId						
Causas (lista reducida)	001-102 I-XX.Todas las causas	String						
Sexo	Ambos sexos	String						
Nacionalidad	Total de defunciones	String						
Total	371,478.0	Double						
(2) 6547887a34d93a29a1fbce06	{ } (5 fields)	Document						
(3) 6547887a34d93a29a1fbce07	{ } (5 fields)	Document						
(4) 6547887a34d93a29a1fbce08	{ } (5 fields)	Document						
(5) 6547887a34d93a29a1fbce09	{ } (5 fields)	Document						
(6) 6547887a34d93a29a1fbce0a	{ } (5 fields)	Document						
(7) 6547887a34d93a29a1fbce0b	{ } (5 fields)	Document						
(8) 6547887a34d93a29a1fbce0c	{ } (5 fields)	Document						

## Análisis de datos:

El objetivo es analizar la prevalencia entre hombres y mujeres ante ciertas causas de defunción, nos vamos a centrar en la categoría de otras causas ("090-102 XX.Causas externas de mortalidad"), por lo tanto, vamos a seleccionar en la data todas las causas externas (sin contar aquella que muestra el total de todas las causas externas de mortalidad), ya que queremos estudiar la prevalencia en las causas de defunción de manera individualizada.

*fase1 = { \$match: { "Causas (lista reducida)": { \$in: [*

*"091 Otros accidentes de transporte",*

*"092 Caídas accidentales",*

*"093 Ahogamiento, sumersión y sofocación accidentales",*

*"094 Accidentes por fuego, humo y sustancias calientes",*

*"095 Envenenamiento accidental por psicofármacos y drogas de abuso",*

*"096 Otros envenenamientos accidentales",*

*"097 Otros accidentes",*

*"098 Suicidio y lesiones autoinfligidas",*

*"099 Agresiones (Homicidios)",*

*"100 Eventos de intención no determinada",*

*"101 Complicaciones de la atención médica y quirúrgica",*

*"102 Otras causas externas y sus efectos tardíos" ] } }*

*etapas = [fase1]*

*db.causa.aggregate(etapas)*

```

24
25 - fase1 = { $match: { "Causas (lista reducida)": {$in:[
26     "091 Otros accidentes de transporte",
27     "092 Caídas accidentales",
28     "093 Ahogamiento, sumersión y sofocación accidentales",
29     "094 Accidentes por fuego, humo y sustancias calientes",
30     "095 Envenenamiento accidental por psicofármacos y drogas de abuso",
31     "096 Otros envenenamientos accidentales",
32     "097 Otros accidentes",
33 ]}} }

```

causa 0.050 s 2160 Docs

_id	Causas (lista reducida)	Sexo	Nacionalidad	Total
1	097 Otros accidentes	Ambos sexos	Noruega	1
2	097 Otros accidentes	Ambos sexos	Países Bajos	8
3	097 Otros accidentes	Ambos sexos	Portugal	12
4	097 Otros accidentes	Ambos sexos	Reino Unido	17
5	097 Otros accidentes	Ambos sexos	Rumanía	20
6	097 Otros accidentes	Ambos sexos	Suecia	1
7	097 Otros accidentes	Ambos sexos	Suiza	2
8	097 Otros accidentes	Ambos sexos	Resto de Europa	22
9	097 Otros accidentes	Ambos sexos	PAÍSES DE ÁFRICA	32
10	097 Otros accidentes	Ambos sexos	Marruecos	22

Como podemos ver en esta tabla, el análisis que podemos hacer sobre la prevalencia de la causa de defunción por género es bastante deficiente, ya que no tenemos los datos ordenados por países y género, por tanto, para realizar el análisis tendremos que reorganizar la data para poder ver en una misma fila la cantidad de defunciones por género en función de una causa en un país determinado.

Para ello primero vamos a seleccionar los países que queremos observar (vamos a quitar todos aquellos valores dentro de la variable “Nacionalidad” que no corresponden a un país, por ejemplo, los continentes).

Puesto que queremos ver la prevalencia por sexo también vamos a realizar una llamada a la data únicamente para el sexo masculino y femenino (eliminando así los valores para ambos sexos).

Una vez depurados los datos en función de lo que queremos estudiar, vamos a agrupar los datos en función de la nacionalidad y la causa de defunción, y, una vez agrupado vamos a generar un condicional para generar dos columnas (una por cada género) que nos proporcione el sumatorio de las defunciones para esa causa en ese país.

Tengo en cuenta que al agrupar los datos por nacionalidad y causa de defunción el total de los valores se va a duplicar, por tanto, para corregir este error dividimos los datos entre dos.

Generaremos una columna con la suma de los dos géneros para obtener el total de defunciones por causa y por país.

Eliminamos todas aquellas filas que no tengan valor ni para hombres ni para mujeres (causas por nacionalidad donde el total sea cero).

Finalmente ordenamos de manera descendente.

### Selección de los países a estudiar:

```
fase2 = {$match: {
"Nacionalidad":{$in:["Alemania","Bélgica","Dinamarca","España","Finlandia","Francia",
"Italia","Noruega", "Países Bajos","Portugal","Reino
```

***"Italia", "Noruega", "Países Bajos", "Portugal", "Reino***

*Unido", "Rumanía", "Suecia", "Suiza", "Marruecos", "Argentina",  
"Colombia", "Cuba", "Ecuador", "Estados Unidos de América"]}]}}*

**Selección del género a analizar:**

*fase3 = {\$match:{"Sexo":{\$in:["Mujeres", "Varones"]}}}*

**Reorganización de la data para analizar de forma más eficiente los datos:**

*fase4 = {\$group: {\_id: {"Nacionalidad": "\$Nacionalidad",  
"Causas": "\$Causas (lista reducida)"},  
"Varones":{\$sum: { \$cond: [{ \$eq: ["\$Sexo", 'Varones'] }, '\$Total', 0] }},  
"Mujeres":{\$sum: { \$cond: [{ \$eq: ["\$Sexo", "Mujeres"] }, '\$Total', 0] }}}}*

**Corrección de los datos:**

*fase5 = {\$project: {"\_id": 0,  
"Nacionalidad": "\$\_id.Nacionalidad",  
"Causas": "\$\_id.Causas",  
"Varones": { \$divide: [ "\$Varones", 2] },  
"Mujeres": { \$divide: [ "\$Mujeres", 2] }}}}*

**Obtención del total:**

*fase6 = {\$project:{"Nacionalidad": 1,  
"Causas": 1,  
"Varones": 1,  
"Mujeres": 1,  
"Total": { \$add: ["\$Varones", "\$Mujeres"] }}}}*

**Eliminar filas con "Total" = 0**

*fase7 = {\$match:{Total: {\$ne:0}}}*

### Organización de la data:

```
Fase8 = {$sort:{"Total":-1}}
```

```
etapas = [fase1,fase2,fase3,fase4,fase5,fase6,fase7,fase8]
```

```
db.causa.aggregate(etapas)
```

```
55     "Total": { $add: [ "$Varones", "$Mujeres" ] }}
56 fase7 = {$match:{Total: {$ne:0}}}
57 fase8 = {$sort:{"Total":-1}}
58 etapas = [fase1,fase2,fase3,fase4,fase5,fase6,fase7,fase8]
59 db.causa.aggregate(etapas)
60
61
```

causa				
Nacionalidad	Causas	Varones	Mujeres	Total
1 España	092 Caídas accidentales	839.0	718.0	1557.0 (1.6K)
2 España	097 Otros accidentes	1,318.0	687.0	688,318.0
3 España	098 Suicidio y lesiones autoinfligidas	2,314.0	683.0	685,314.0
4 España	095 Envenenamiento accidental por psicofá	368.0	87.0	455.0
5 España	101 Complicaciones de la atención médica	177.0	188.0	365.0
6 España	091 Otros accidentes de transporte	280.0	70.0	350.0
7 España	099 Agresiones (Homicidios)	164.0	94.0	258.0
8 España	096 Otros envenenamientos accidentales	152.0	57.0	209.0
9 España	094 Accidentes por fuego, humo y sustancia	111.0	72.0	183.0
10 España	100 Eventos de intención no determinada	61.0	16.0	77.0
11 Alemania	093 Ahogamiento, sumersión y sofocación a	33.0	8.0	41.0
12 Reino Unido	090 Suicidio y lesiones autoinfligidas	37.0	9.0	46.0

Como se puede ver, reorganizando la data en función de los objetivos que buscamos, esta se muestra con mayor claridad, ofreciendo un análisis más eficiente.

Como cabría esperar, la población con nacionalidad española se encuentra en los puestos más altos en cuanto a número de defunciones se refiere, se trata de un resultado esperable, ya que la proporción de población con nacionalidad española frente al resto de nacionalidades suponemos claramente superior en España.

### Estudio de las causas mayoritarias de defunción por género:

Estudiamos a las causas de defunción más prevalentes por género, para ello tenemos que agrupar los datos únicamente por causas (sin tener en cuenta los países):

```
fase9 = { $group: { _id: "$Causas", Mujeres: { $sum: "$Mujeres" }, Varones: { $sum: "$Varones" }, Total: { $sum: "$Total" } } }
```

### Varones:

Para estudiar las causas de defunción mayoritarias entre los hombres lo que tenemos que hacer es ordenar la data de mayor a menor en función de la cuantía de casos por hombres.

```
fase10 = {$sort: {"Varones":-1}}
```

```
etapas = [fase1,fase2,fase3,fase4,fase5,fase6,fase7,fase9,fase10]
```

## db.causa.aggregate(etapas)

	_id	Mujeres	Varones	Total
1	092 Caídas accidentales	366.0	445.0	811.0
2	095 Envenenamiento accidental por psicofármacos y drogas	47.0	197.5	244.5
3	091 Otros accidentes de transporte	37.5	155.0	192.5
4	099 Agresiones (Homicidios)	56.0	109.5	165.5
5	101 Complicaciones de la atención médica y quirúrgica	97.0	93.0	190.0
6	096 Otros envenenamientos accidentales	30.0	83.5	113.5
7	098 Suicidio y lesiones autoinfligidas	358.5	68,157.0	426,657.0
8	097 Otros accidentes	353.0	64,659.0	417,659.0
9	093 Ahogamiento, sumersión y sofocación accidentales	17,06.0	61,1665.0	78,2265.0
10	094 Accidentes por fuego, humo y sustancias calientes	37.0	58.0	95.0
11	100 Eventos de intención no determinada	10.0	33.5	43.5
12	102 Otras causas externas y sus efectos tardíos	5.5	9.0	14.5

Como podemos ver en los hombres es más común los fallecimientos derivados de situaciones de riesgo, lo cual es compatible con el resultado de varios estudios realizados sobre la percepción del riesgo en función del género, teniendo las personas de género masculino una percepción menor del riesgo que el que presentan las personas de género femenino.

### Mujeres:

Para estudiar las causas de defunción mayoritarias entre los hombres lo que tenemos que hacer es ordenar la data de mayor a menor en función de la cuantía de casos por hombres.

*fase11 = {\$sort: {"Mujeres":-1}}*

etapas = [fase1,fase2,fase3,fase4,fase5,fase6,fase7,fase9,fase11]

## db.causa.aggregate(etapas)

	_id	Mujeres	Varones	Total
1	092 Caídas accidentales	366.0	445.0	811.0
2	098 Suicidio y lesiones autoinfligidas	358.5	68,157.0	426,657.0
3	097 Otros accidentes	353.0	64,659.0	417,659.0
4	101 Complicaciones de la atención médica y quirúrgica	97.0	93.0	190.0
5	099 Agresiones (Homicidios)	56.0	109.5	165.5
6	095 Envenenamiento accidental por psicofármacos y drogas de abuso	47.0	197.5	244.5
7	091 Otros accidentes de transporte	37.5	155.0	192.5
8	094 Accidentes por fuego, humo y sustancias calientes	37.0	58.0	95.0
9	096 Otros envenenamientos accidentales	30.0	83.5	113.5
10	093 Ahogamiento, sumersión y sofocación accidentales	17,06.0	61,1665.0	78,2265.0
11	100 Eventos de intención no determinada	10.0	33.5	43.5
12	102 Otras causas externas y sus efectos tardíos	5.5	9.0	14.5

Como podemos ver en el resultado sobre las causas menos repetidas entre las mujeres, se encuentran aquellas en las que existe un riesgo perceptible, lo que coincide con el análisis y los resultados que hemos obtenido entre los hombres. Sin embargo, observamos un dato bastante sorprendente, ya que la segunda causa de suicidio entre las mujeres es el “Suicidio y lesiones autoinfligidas”, siendo la cuantía de esta mucho mayor a la cuantía encontrada en el género masculino, lo cual contradice a la mayoría de los estudios sobre la materia, los cuales reflejan un mayor índice de suicidios entre la población masculina derivada de una sensación de obligación en el cuidado económico de su familia y unos objetivos laborales, que, al no verse cumplidos provocan sensación de frustración y fracaso.

## Estudio de las causas mayoritarias de defunción por nacionalidad (sin contabilizar España):

Otro estudio que podemos realizar es, eliminando España de la búsqueda, ver qué Nacionalidad tiene un mayor índice de mortalidad y la causa de esta, para ello debemos volver a tener en cuenta las causas de fallecimiento:

```
Fase12 = {$match: {  
  "Nacionalidad":{$in:["Alemania", "Bélgica", "Dinamarca", "España", "Finlandia", "Francia",  
    "Italia", "Noruega", "Países Bajos", "Portugal", "Reino  
    Unido", "Rumanía", "Suecia", "Suiza", "Marruecos", "Argentina",  
    "Colombia", "Cuba", "Ecuador", "Estados Unidos de América"]}}}
```

Ordenamos de mayor a menor el total:

```
Fase8 = {$sort: {"Total":-1}}
```

```
etapas = [fase1,fase12,fase3,fase4,fase5,fase6,fase7,fase8]
```

```
db.causa.aggregate(etapas)
```

	Nacionalidad	Causas	Varones	Mujeres	Total
1	Alemania	093 Ahogamiento, sumersión y sofocación accidentales	16,5,0	4,0	20,5,0
2	Reino Unido	098 Suicidio y lesiones autoinfligidas	13,5,0	4,0	17,5,0
3	Reino Unido	093 Ahogamiento, sumersión y sofocación accidentales	11,5,0	2,5,0	14,0
4	Alemania	098 Suicidio y lesiones autoinfligidas	9,5,0	2,0	11,5,0
5	Marruecos	093 Ahogamiento, sumersión y sofocación accidentales	11,0	0,5,0	11,5,0
6	Marruecos	097 Otros accidentes	9,5,0	1,5,0	11,0
7	Rumanía	097 Otros accidentes	10,0	0	10,0
8	Marruecos	098 Suicidio y lesiones autoinfligidas	8,5,0	1,0	9,5,0
9	Rumanía	098 Suicidio y lesiones autoinfligidas	6,0	3,0	9,0
10	Ecuador	097 Otros accidentes	8,5,0	0,5,0	9,0
11	Reino Unido	097 Otros accidentes	7,5,0	1,0	8,5,0
12	Reino Unido	092 Caídas accidentales	6,0	2,0	8,0

El resultado traslada que los alemanes y los ingleses presentan más número de defunciones que el resto de las nacionalidades, además las causas son “Ahogamiento, sumersión y sofocación accidentales” y “Suicidio y lesiones autoinfligidas”, lo que nos podría orientar hacia una población que fallece en España durante el turismo.

Además, en este caso vemos que estas causas se encuentran mayoritariamente entre la población masculina, lo cual, está acorde con una falta sensación de seguridad que comentamos en el análisis del género masculino.

Comprobamos que estos resultados representan la realidad con respecto a la nacionalidad con más fallecimientos en España, para ello generamos una tabla con únicamente las nacionalidades, y vemos qué nacionalidad tiene mayor número de fallecimientos.

```
fase14 = {$group: {_id:"$Nacionalidad", Mujeres:{$sum:"$Mujeres"},
Varones:{$sum:"$Varones"}, Total:{$sum:"$Total"}}}
```

```
etapas = [fase1,fase12,fase3,fase4,fase5,fase6,fase7,fase14,fase8]
```

```
db.causa.aggregate(etapas)
```

_id	Mujeres	Varones	Total
1 Reino Unido	14,5.0	51,5.0	66.0
2 Alemania	11,5.0	40,5.0	52.0
3 Marruecos	6,5.0	44,5.0	51.0
4 Rumania	7.0	26,5.0	33,5.0
5 Ecuador	4,5.0	25.0	29,5.0
6 Francia	7.0	16,5.0	23,5.0
7 Portugal	4.0	18,5.0	22,5.0
8 Colombia	3,5.0	12,5.0	16.0
9 Italia	1,5.0	13.0	14,5.0
10 Bélgica	2,5.0	8.0	10,5.0

Comprobamos que los resultados son los mismos que los obtenidos cuando filtrábamos tanto por causa como por nacionalidad, siendo la población procedente de Inglaterra y Alemania las nacionalidades con mayor número de defunciones.

## Conclusiones:

Una vez analizada la data, escogiendo los principales grupos de análisis, causa de fallecimiento en función de género y causa y nacionalidad mayoritaria en fallecimientos, obtenemos:

- Una cantidad mayor de fallecimientos por motivo de una sensación de falsa seguridad entre los hombres, lo cual concuerda con los estudios realizados por diferentes instituciones (como por ejemplo FAD juventud).
- Una cantidad de fallecimientos por suicidio superior entre las mujeres, lo cual no concuerda con diferentes estudios sobre la materia. Para poder estudiar el motivo de este resultado necesitaríamos saber cómo se ha obtenido la data (ej: periodo de estudio).
- Un mayor número de fallecimientos entre alemanes e ingleses con respecto al resto de nacionalidades, en particular por causas relacionadas con el primer punto de la conclusión (ya que mayoritariamente este grupo se compone de hombres).