

# Proyecto IMB (SQL)

## Paso 1. Importar librerías

```
In [ ]: import sqlite3
import pandas as pd
```

## Paso 2. Cargar base de datos

```
In [ ]: conn = sqlite3.connect('./movies.sqlite')
```

## Paso 3. Conocer el nombre de las tablas que contiene la base de datos

```
In [ ]: tablas = pd.read_sql("SELECT name FROM sqlite_master WHERE type = 'table'", conn)
print(tablas)
```

```
      name
0      movies
1  sqlite_sequence
2      directors
```

## Paso 4. Ejecutando tareas

### Paso 4.1

**Pregunta 1:** ¿Puede obtener todos los datos sobre películas?

```
In [ ]: consulta_1 = pd.read_sql("SELECT * FROM movies", conn)
consulta_1.head()
```

Out[ ]:		id	original_title	budget	popularity	release_date	revenue	title	vote_average	vote_cou
0	43597		Avatar	237000000	150	2009-12-10	2787965087	Avatar	7.2	118
1	43598		Pirates of the Caribbean: At World's End	300000000	139	2007-05-19	961000000	Pirates of the Caribbean: At World's End	6.9	45
2	43599		Spectre	245000000	107	2015-10-26	880674609	Spectre	6.3	44
3	43600		The Dark Knight Rises	250000000	112	2012-07-16	1084939099	The Dark Knight Rises	7.6	91
4	43601		John Carter	260000000	43	2012-03-07	284139100	John Carter	6.1	21

## Paso 4.2

**Pregunta 2:** ¿Cómo obtienes todos los datos sobre los directores?

```
In [ ]: consulta_2 = pd.read_sql("SELECT * FROM directors", conn)
        consulta_2.head()
```

Out[ ]:

	name	id	gender	uid	department
0	James Cameron	4762	2	2710	Directing
1	Gore Verbinski	4763	2	1704	Directing
2	Sam Mendes	4764	2	39	Directing
3	Christopher Nolan	4765	2	525	Directing
4	Andrew Stanton	4766	2	7	Directing

## Paso 4.3

**Pregunta 3:** Compruebe cuántas películas hay en iMDB.

```
In [ ]: consulta_3 = pd.read_sql("""SELECT COUNT(*) AS 'cantidad de peliculas'
                                FROM movies""", conn)
consulta_3
```

```
Out[ ]: cantidad de peliculas
0                4773
```

## Paso 4.4

**Pregunta 4:** Encuentra estos 3 directores: James Cameron; Luc Besson; John Woo

```
In [ ]: consulta_4 = pd.read_sql("""SELECT * FROM directors
                                WHERE name IN ('James Cameron',
                                                'Luc Besson',
                                                'John Woo')""", conn)
consulta_4
```

```
Out[ ]:
```

	name	id	gender	uid	department
0	James Cameron	4762	2	2710	Directing
1	John Woo	4893	2	11401	Directing
2	Luc Besson	4949	2	59	Directing

## Paso 4.5

**Pregunta 5:** Encuentre todos los directores cuyo nombre comience con Steven.

```
In [ ]: consulta_5 = pd.read_sql("""SELECT * FROM directors
                                WHERE name LIKE 'Steve%'""", conn)
consulta_5
```

Out[ ]:

	name	id	gender	uid	department
0	Steven Spielberg	4799	2	488	Directing
1	Steve Hickner	4852	2	44113	Directing
2	Steven Soderbergh	4909	2	1884	Directing
3	Steve Martino	4943	2	71729	Directing
4	Steven Brill	5013	2	32593	Directing
5	Steve Carr	5048	2	52112	Directing
6	Steven Zaillian	5117	2	2260	Directing
7	Steve Antin	5177	2	42745	Directing
8	Steve Oedekerk	5190	2	4489	Directing
9	Steven Quale	5216	2	93214	Directing
10	Steven Seagal	5221	2	23880	Directing
11	Steve Miner	5362	2	58712	Directing
12	Steve Pink	5374	2	3227	Directing
13	Steven E. de Souza	5390	2	1726	Directing
14	Steve Boyum	5500	2	72258	Directing
15	Steve Bendelack	5564	2	16356	Directing
16	Steve Barron	5581	2	18308	Directing
17	Steve McQueen	5657	2	72757	Directing
18	Steve Trenbirth	5662	0	142007	Directing
19	Steve Beck	5672	2	57521	Directing
20	Steve Rash	5688	2	54590	Directing
21	Steven Shainberg	5803	2	67795	Directing
22	Steve Gomer	5878	2	104507	Directing
23	Steve James	6209	2	122449	Directing
24	Steve Carver	6376	2	102627	Directing
25	Steven R. Monroe	6713	2	88039	Directing
26	Steve Buscemi	6763	2	884	Directing
27	Steve Taylor	6781	2	440517	Directing

## Paso 4.6

**Pregunta 6:** Contar el número de directoras

```
In [ ]: consulta_6 = pd.read_sql("""SELECT COUNT(*) AS 'num. directoras'
                                FROM directors
```

```
WHERE gender = 1""", conn)

consulta_6
```

```
Out[ ]:  num. directoras
0          150
```

## Paso 4.7

**Pregunta 7:** ¿Encontrar el nombre de las 10 primeras mujeres directoras?

```
In [ ]: consulta_7 = pd.read_sql("""SELECT name FROM directors
WHERE gender = 1
LIMIT 10""", conn)

consulta_7
```

```
Out[ ]:  name
0  Brenda Chapman
1  Lilly Wachowski
2  Jennifer Yuh Nelson
3  Kathryn Bigelow
4  Nancy Meyers
5  Jill Culton
6  Mimi Leder
7  Vicky Jenson
8  Betty Thomas
9  Angelina Jolie
```

## Paso 4.8

**Pregunta 8:** ¿Cuáles son las 3 películas más populares?

```
In [ ]: consulta_8 = pd.read_sql("""SELECT original_title, popularity
FROM movies
ORDER BY popularity DESC
LIMIT 3""", conn)

consulta_8
```

```
Out[ ]:  original_title  popularity
0  Minions             875
1  Interstellar        724
2  Deadpool            514
```

## Paso 4.9

**Pregunta 9:** ¿Cuáles son las 3 películas más rentables?

```
In [ ]: consulta_9 = pd.read_sql("""SELECT original_title, revenue
                                FROM movies
                                ORDER BY revenue
                                DESC LIMIT 3""", conn)

consulta_9
```

```
Out[ ]:   original_title  revenue
0      Avatar  2787965087
1      Titanic  1845034188
2  The Avengers  1519557910
```

## Paso 4.10

**Pregunta 10:** ¿Cuál es el voto promedio más otorgado desde el 1 de enero de 2000?

```
In [ ]: consulta_10 = pd.read_sql("""SELECT vote_average
                                   FROM movies
                                   WHERE release_date > '2000-01-01'
                                   ORDER BY vote_average DESC
                                   LIMIT 1""", conn)

consulta_10
```

```
Out[ ]:   vote_average
0          9.5
```

## Paso 4.11

**Pregunta 11:** ¿Qué película(s) fue(n) dirigida(s) por Brenda Chapman?

```
In [ ]: consulta_11 = pd.read_sql("""SELECT m.original_title
                                   FROM movies m
                                   JOIN directors d ON m.director_id = d.id
                                   WHERE d.name = 'Brenda Chapman'
                                   """, conn)

consulta_11
```

```
Out[ ]:   original_title
0      Brave
```

## Paso 4.12

**Pregunta 12:** ¿Quién fue el director que hizo más películas?

```
In [ ]: consulta_12 = pd.read_sql("""SELECT d.name, COUNT(1)
                                   FROM directors d
```

```
        JOIN movies m ON m.director_id = d.id
        GROUP BY m.director_id
        ORDER BY COUNT(1) DESC
        LIMIT 1
        """, conn)

consulta_12
```

```
Out[ ]:
      name  COUNT(1)
0  Steven Spielberg      27
```

## Paso 4.13

**Pregunta 13:** ¿Quién es el director más financiable?

```
In [ ]: consulta_13 = pd.read_sql("""SELECT d.name, m.revenue
        FROM directors d
        JOIN movies m ON d.id = m.director_id
        ORDER BY revenue DESC
        LIMIT 1
        """, conn)

consulta_13
```

```
Out[ ]:
      name  revenue
0  James Cameron  2787965087
```