CC3201

Bases de Datos

Primavera 2025

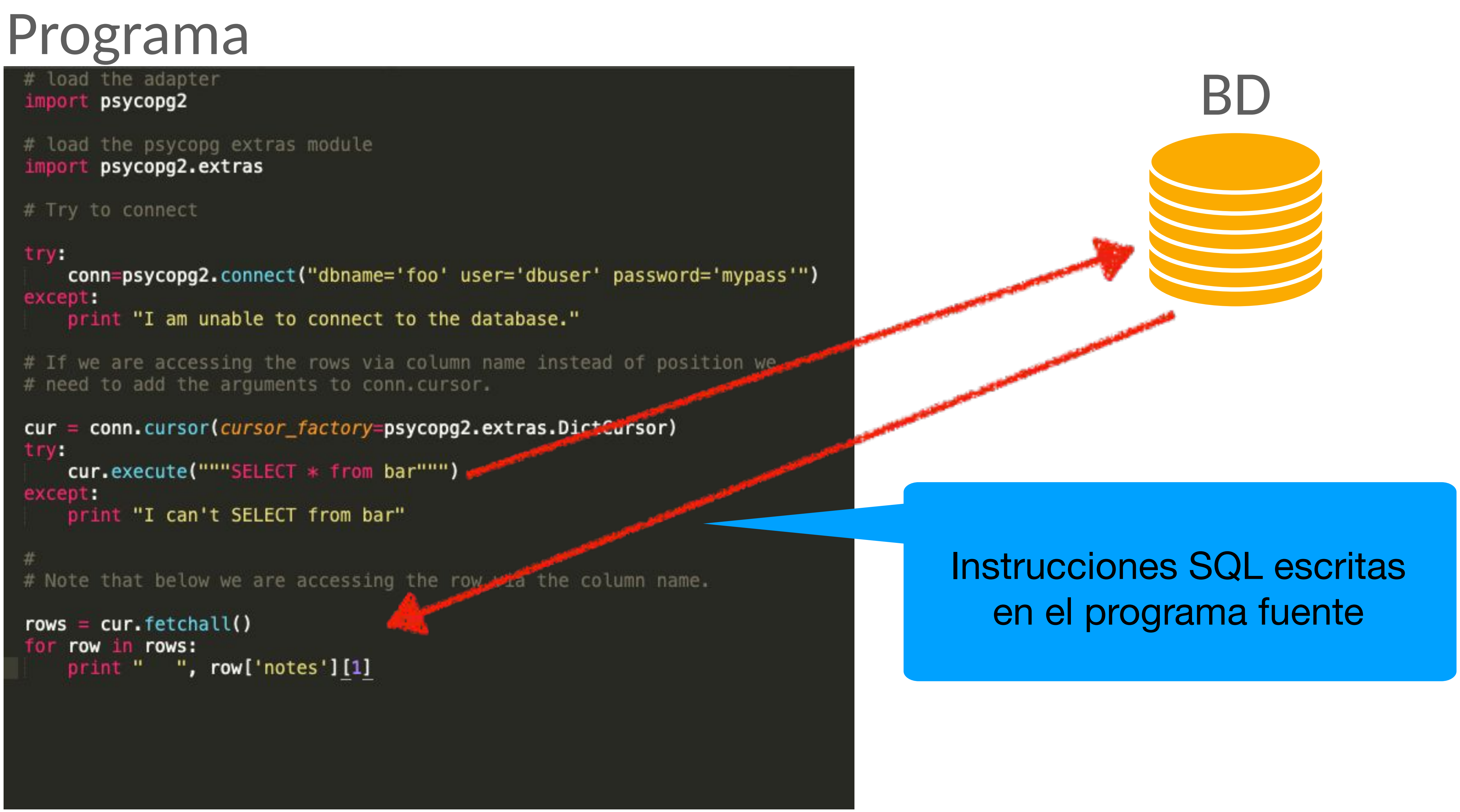
Clase 9: Acceso Programático

{Claudio Gutierrez, Cristián Salazar,
Eduardo Godoy}

Acceso Programático

Capítulo 6 | Ramakrishnan / Gehrke

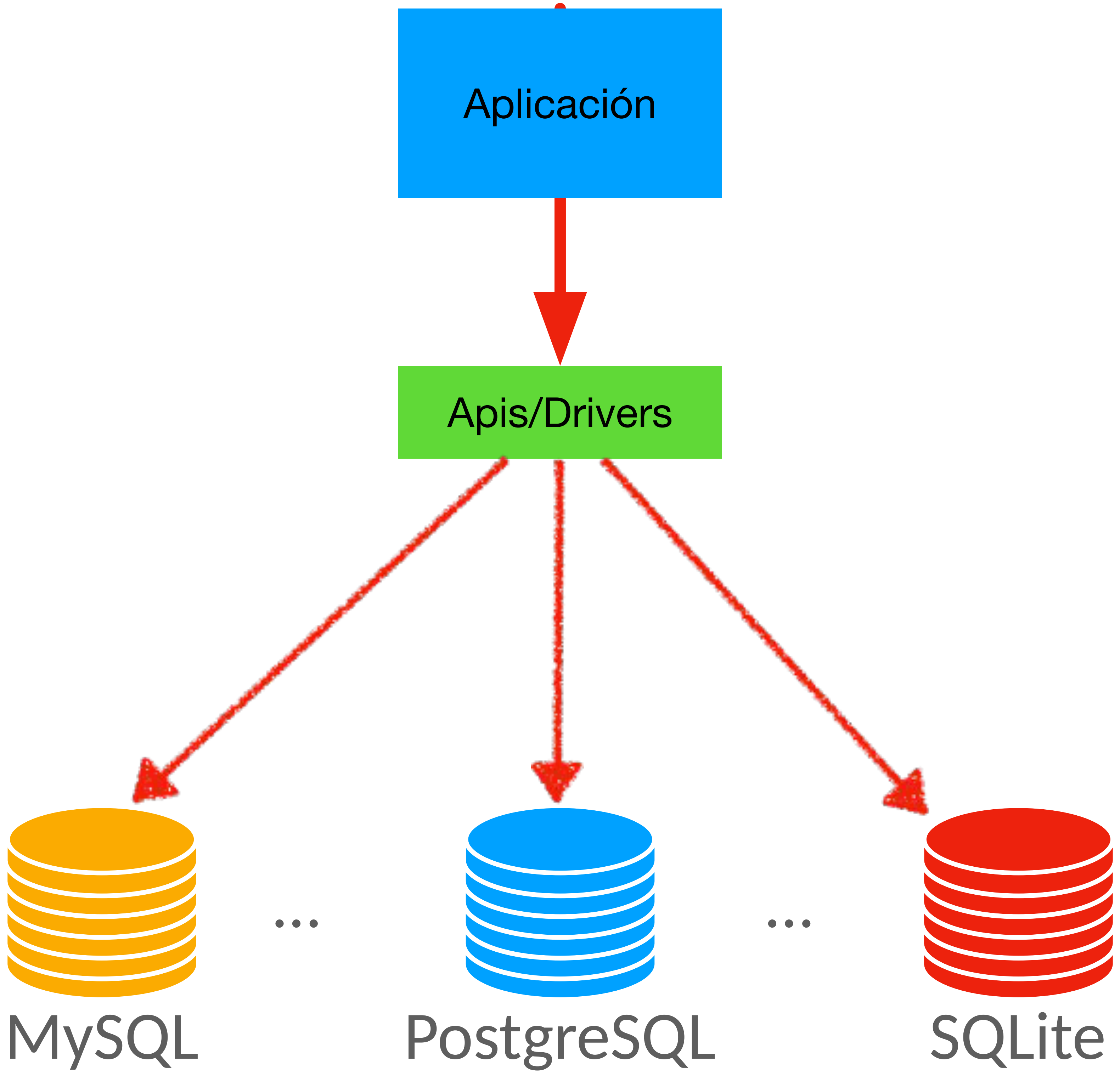
SQL incorporado



```
# load the adapter
import psycopg2

# load the psycopg extras module
import psycopg2.extras

# Try to connect

try:
    conn=psycopg2.connect("dbname='foo' user='dbuser' password='mypass'")
except:
    print "I am unable to connect to the database."

# If we are accessing the rows via column name instead of position we
# need to add the arguments to conn.cursor.

cur = conn.cursor(cursor_factory=psycopg2.extras.DictCursor)
try:
    cur.execute("""SELECT * from bar""")
except:
    print "I can't SELECT from bar"

#
# Note that below we are accessing the row via the column name.

rows = cur.fetchall()
for row in rows:
    print "   ", row['notes'][1]
```

Instrucciones SQL escritas en el programa fuente

SQL incorporado



Muchos sabores

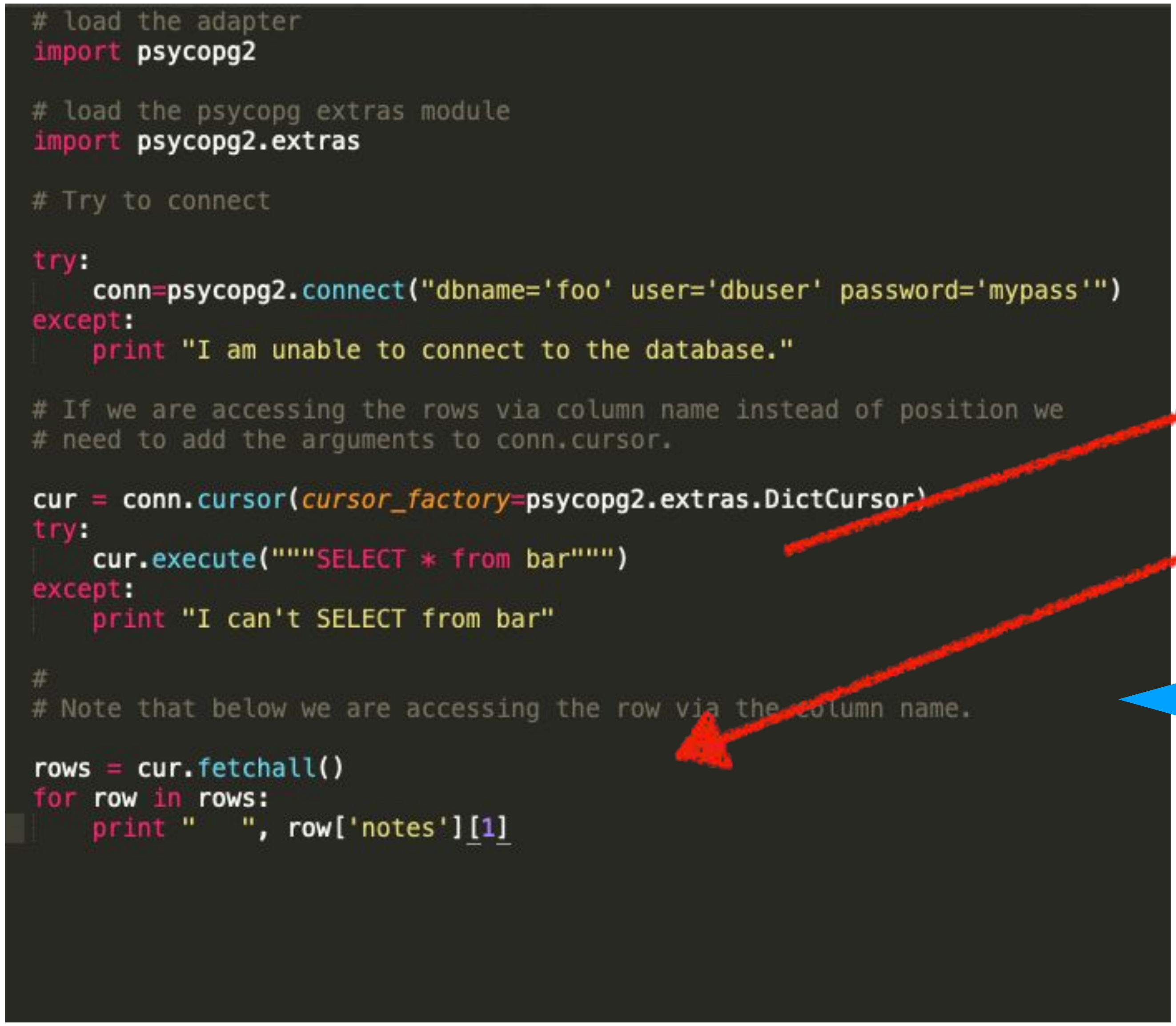
python™
Java
php
JavaScript
C++
Ruby
Scala
node JS™
C#
PostgreSQL

¿Cómo nos conectamos a la bd?

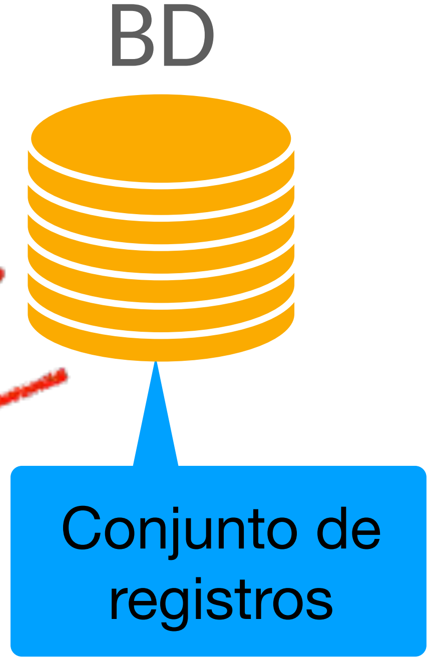
```
import psycopg2

conn = psycopg2.connect(host="cc3201.dcc.uchile.cl",
                        database="cc3201",
                        user=“cc3201",
                        password="foo")
```

¿Cómo realizamos una consulta?

Programa

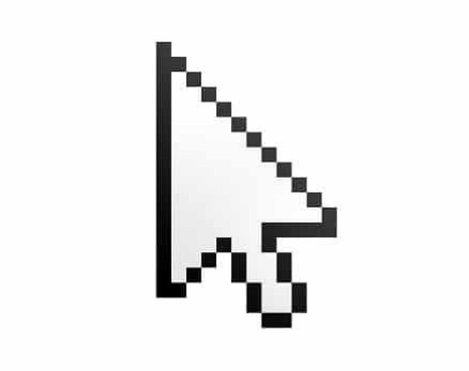


```
# load the adapter
import psycopg2

# load the psycopg extras module
import psycopg2.extras

# Try to connect

try:
    conn=psycopg2.connect("dbname='foo' user='dbuser' password='mypass'")
except:
    print "I am unable to connect to the database."

# If we are accessing the rows via column name instead of position we
# need to add the arguments to conn.cursor.

cur = conn.cursor(cursor_factory=psycopg2.extras.DictCursor)
try:
    cur.execute("""SELECT * from bar""")
except:
    print "I can't SELECT from bar"

#
# Note that below we are accessing the row via the column name.

rows = cur.fetchall()
for row in rows:
    print "   ", row['notes'][1]
```



??

Cursores



Abstracción para recorrer conjuntos de datos

```
cursor = conn.cursor()
```

Puntero a algún dato

¿Cómo realizamos una consulta?

```
cursor.execute(“SELECT * FROM Sailors”)
for row in cursor:
    print row
```

```
(22,’dustin’,7, 45.0)
(31, ‘lubber’, 8, 55.0)
(58, ‘rusty’, 10, 35.0)
```

→

sid	sname	rating	age
22	dustin	7	45.0
31	lubber	8	55.5
58	rusty	10	35.0

¿Cómo realizamos una consulta?

```
cursor.execute(“SELECT * FROM Sailors”)
row = cursor.fetchone()
while row:
    print row
    row = cursor.fetchone()
```

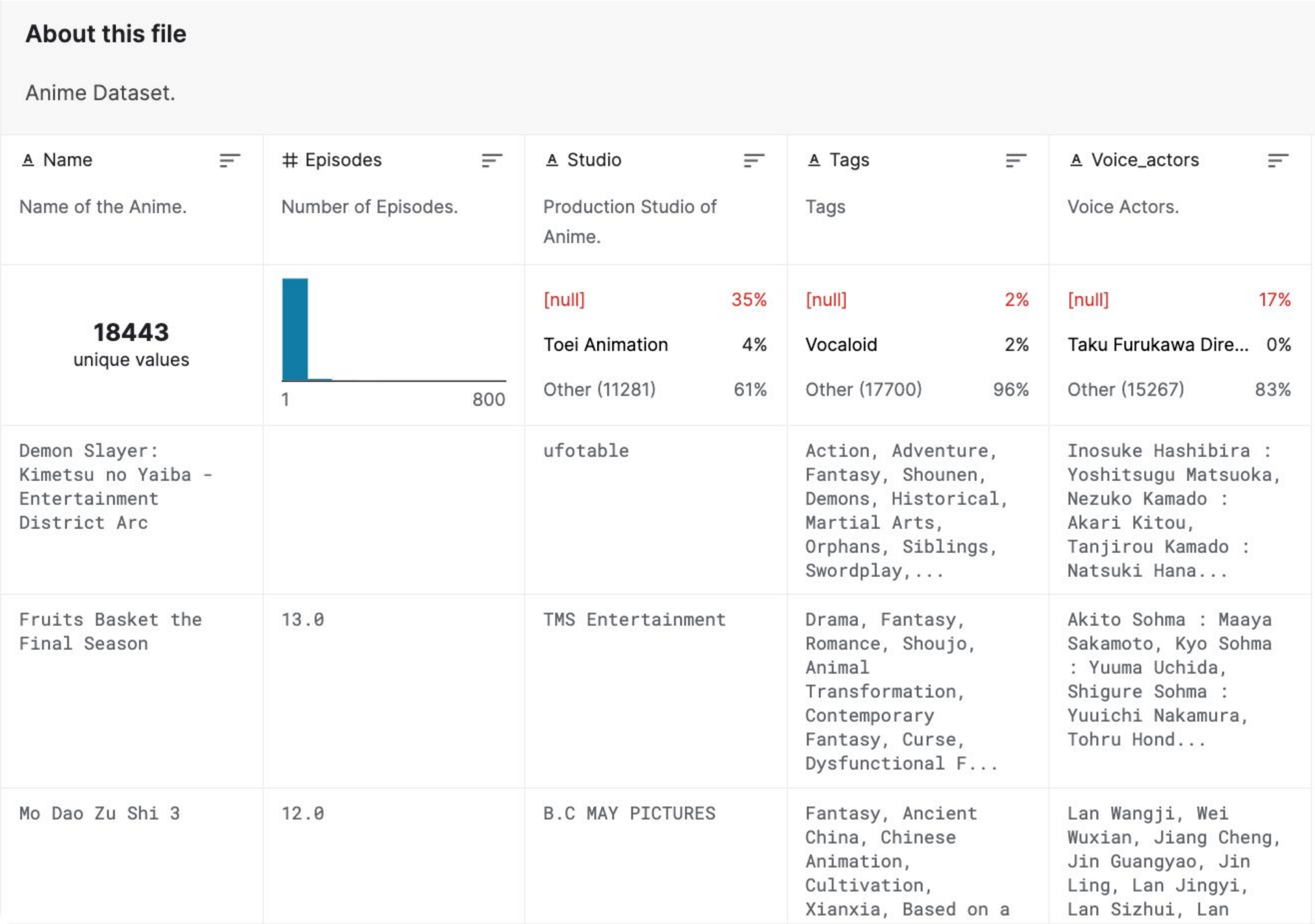
```
cursor.execute(“SELECT * FROM Sailors”)
rows = cursor.fetchall()
for row in rows:
    print row
```

¿Cómo realizamos una actualización?

```
cursor.execute(“UPDATE …”)
```

Importar datos

Una base de datos de anime

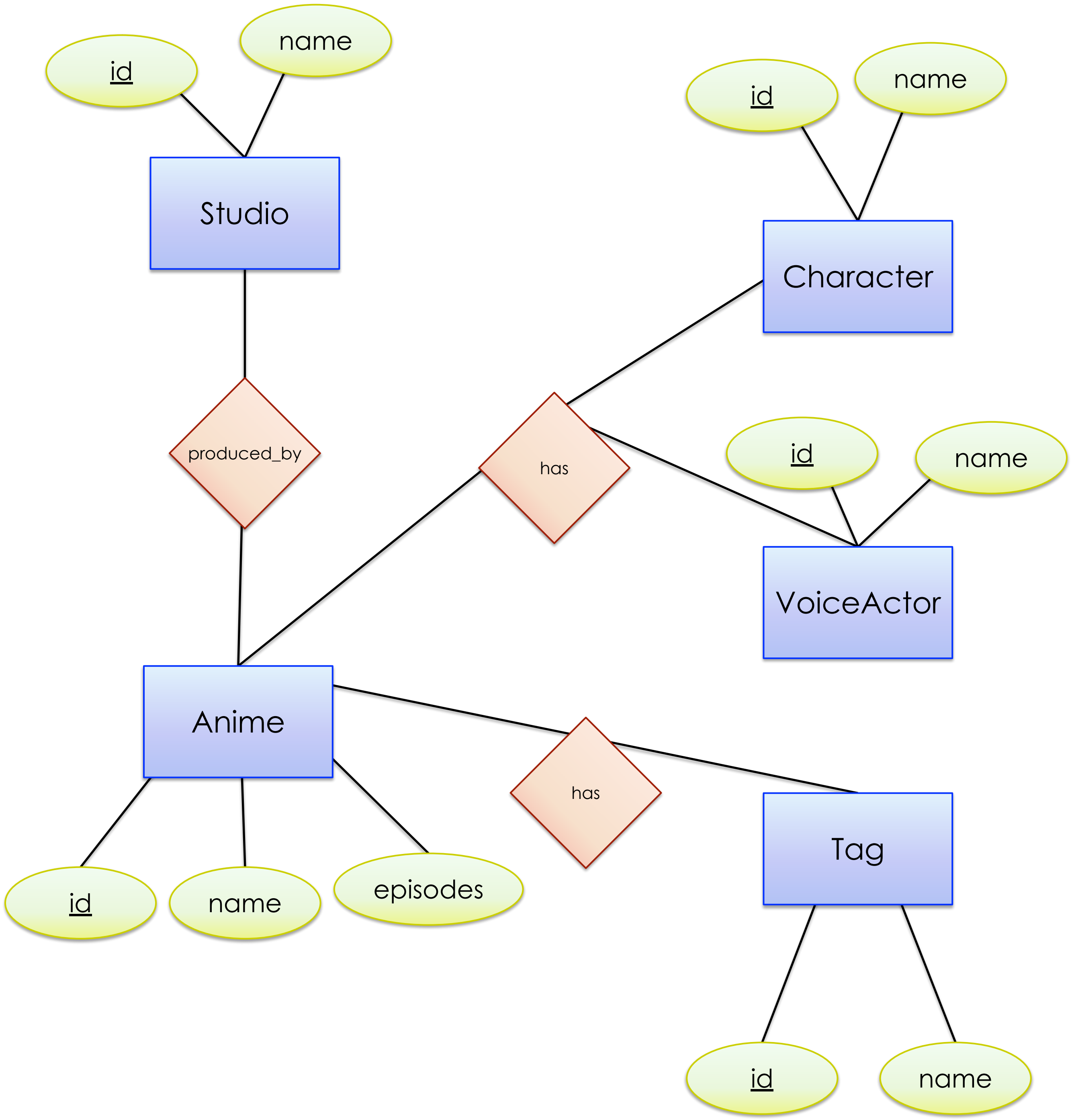


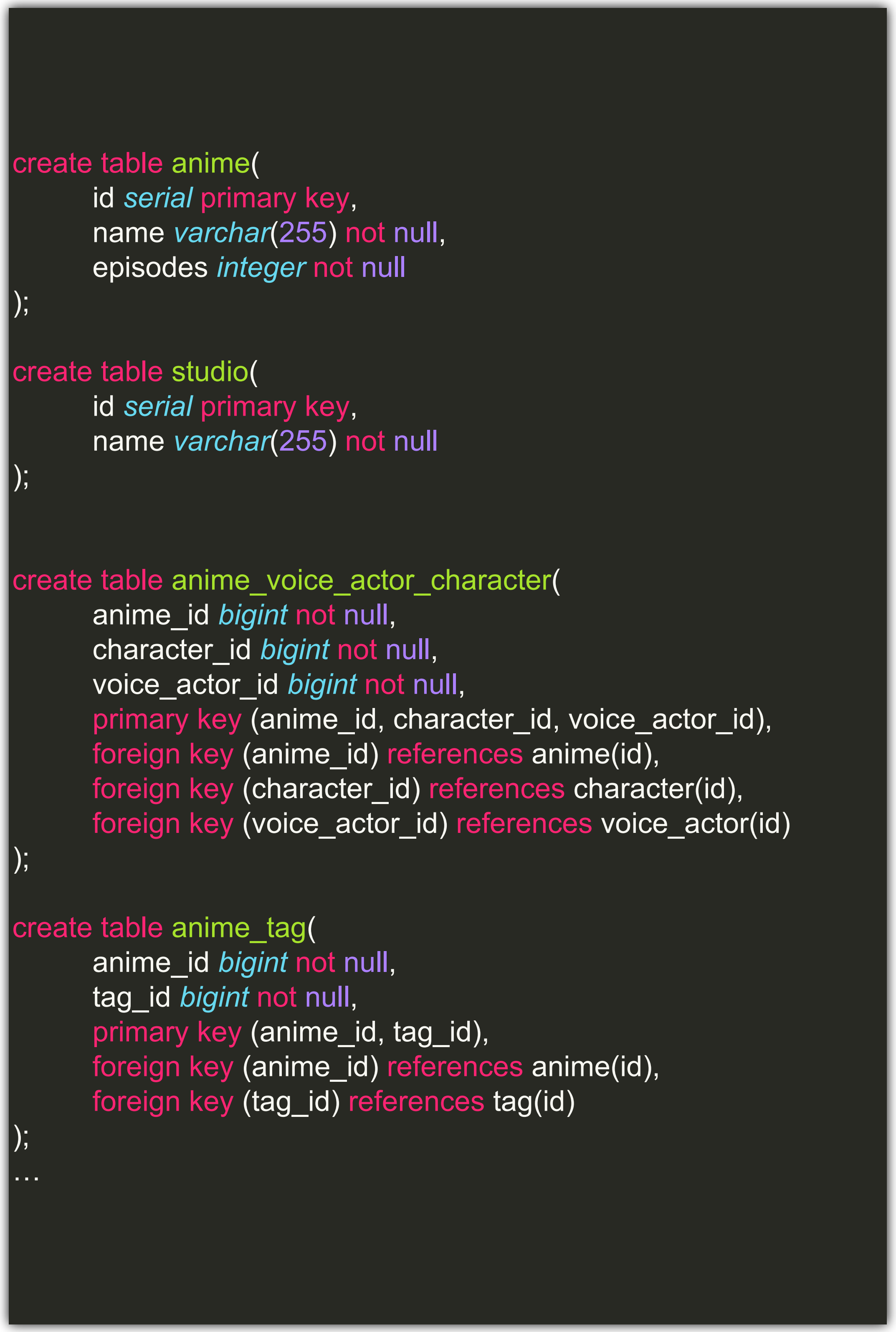
About this file

Anime Dataset.

Name	# Episodes	Studio	Tags	Voice_actors
Name of the Anime.	Number of Episodes.	Production Studio of Anime.	Tags	Voice Actors.
18443 unique values	1 800	[null] 35% Toei Animation 4% Other (11281) 61%	[null] 2% Vocaloid 2% Other (17700) 96%	[null] 17% Taku Furukawa Dire... 0% Other (15267) 83%
Demon Slayer: Kimetsu no Yaiba - Entertainment District Arc		ufotable	Action, Adventure, Fantasy, Shounen, Demons, Historical, Martial Arts, Orphans, Siblings, Swordplay,...	Inosuke Hashibira : Yoshitsugu Matsuoka, Nezuko Kamado : Akari Kitou, Tanjirou Kamado : Natsuki Hana...
Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...
Mo Dao Zu Shi 3	12.0	B.C MAY PICTURES	Fantasy, Ancient China, Chinese Animation, Cultivation, Xianxia, Based on a	Lan Wangji, Wei Wuxian, Jiang Cheng, Jin Guangyao, Jin Ling, Lan Jingyi, Lan Sizhui, Lan

Una base de datos de anime



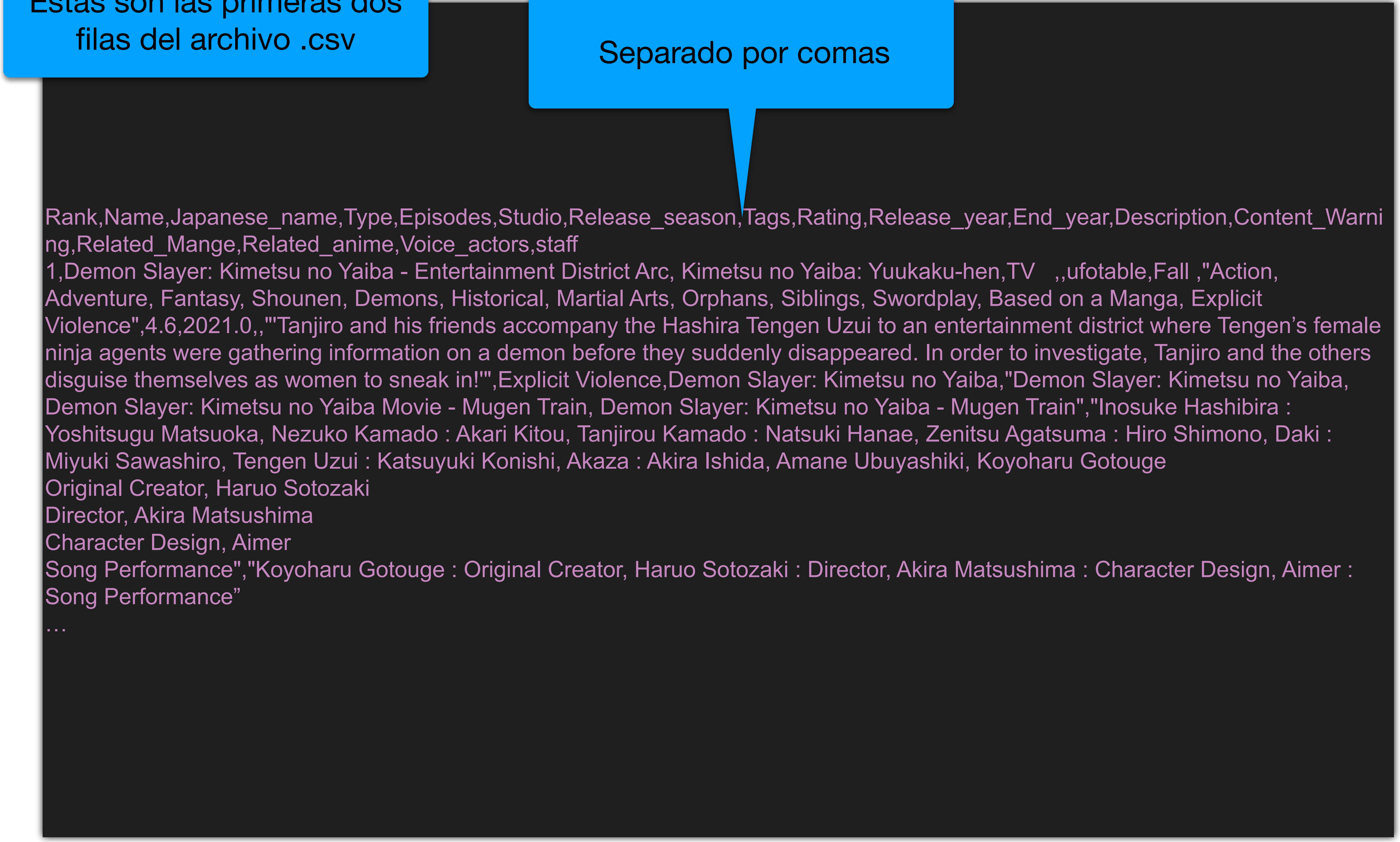


```
create table anime(
        id serial primary key,
        name varchar(255) not null,
        episodes integer not null
);

create table studio(
        id serial primary key,
        name varchar(255) not null
);


create table anime_voice_actor_character(
        anime_id bigint not null,
        character_id bigint not null,
        voice_actor_id bigint not null,
        primary key (anime_id, character_id, voice_actor_id),
        foreign key (anime_id) references anime(id),
        foreign key (character_id) references character(id),
        foreign key (voice_actor_id) references voice_actor(id)
);

create table anime_tag(
        anime_id bigint not null,
        tag_id bigint not null,
        primary key (anime_id, tag_id),
        foreign key (anime_id) references anime(id),
        foreign key (tag_id) references tag(id)
);
…
```

Leyendo el csv

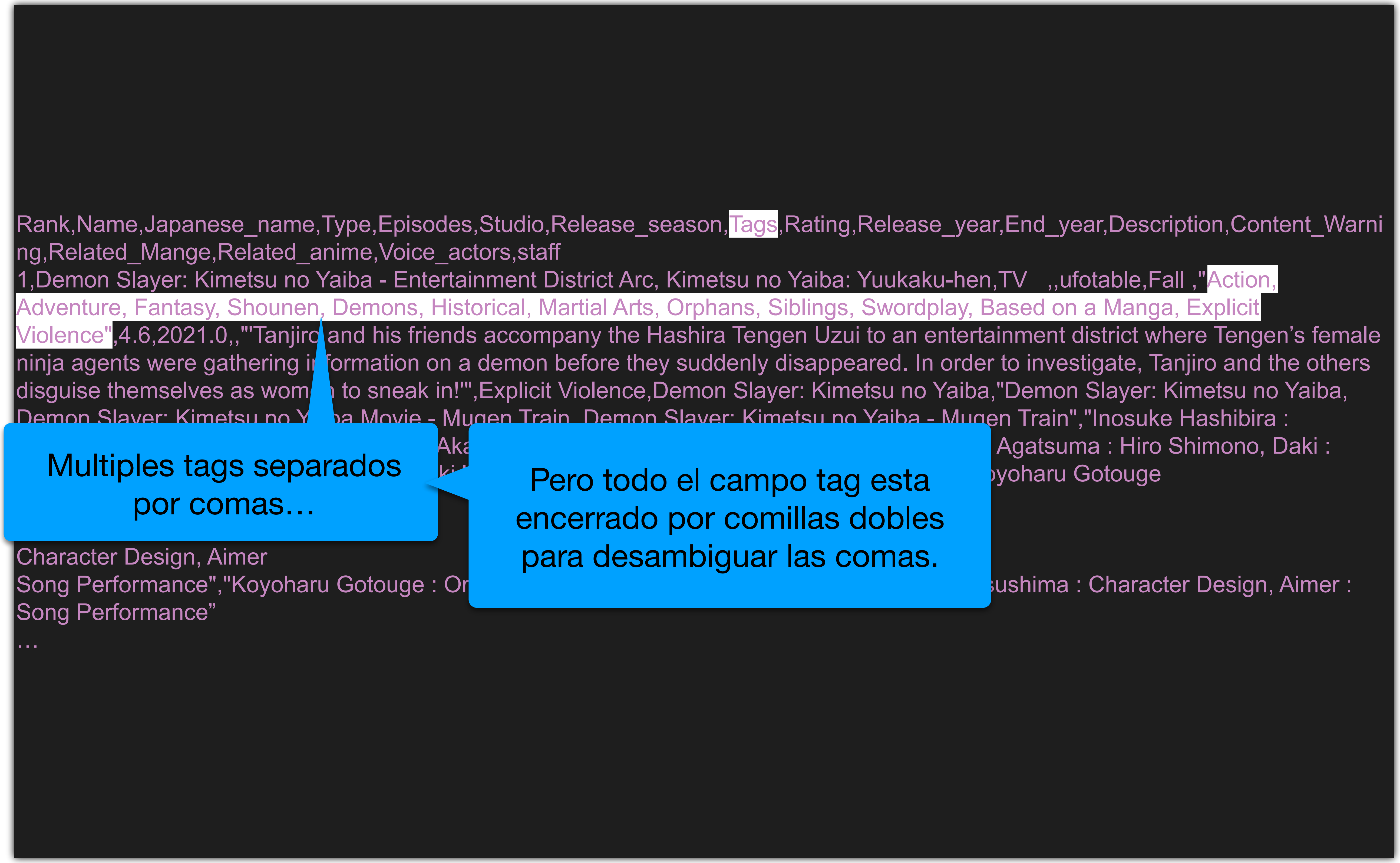
Estas son las primeras dos filas del archivo .csv

Separado por comas



```
Rank,Name,Japanese_name,Type,Episodes,Studio,Release_season,Tags,Rating,Release_year,End_year,Description,Content_Warni
ng,Related_Mange,Related_anime,Voice_actors,staff
1,Demon Slayer: Kimetsu no Yaiba - Entertainment District Arc, Kimetsu no Yaiba: Yuukaku-hen,TV    ,,ufotable,Fall ,"Action,
Adventure, Fantasy, Shounen, Demons, Historical, Martial Arts, Orphans, Siblings, Swordplay, Based on a Manga, Explicit
Violence",4.6,2021.0,,"'Tanjiro and his friends accompany the Hashira Tengen Uzui to an entertainment district where Tengen’s female
ninja agents were gathering information on a demon before they suddenly disappeared. In order to investigate, Tanjiro and the others
disguise themselves as women to sneak in!'",Explicit Violence,Demon Slayer: Kimetsu no Yaiba,"Demon Slayer: Kimetsu no Yaiba,
Demon Slayer: Kimetsu no Yaiba Movie - Mugen Train, Demon Slayer: Kimetsu no Yaiba - Mugen Train","Inosuke Hashibira :
Yoshitsugu Matsuoka, Nezuko Kamado : Akari Kitou, Tanjirou Kamado : Natsuki Hanae, Zenitsu Agatsuma : Hiro Shimono, Daki :
Miyuki Sawashiro, Tengen Uzui : Katsuyuki Konishi, Akaza : Akira Ishida, Amane Ubuyashiki, Koyoharu Gotouge
Original Creator, Haruo Sotozaki
Director, Akira Matsushima
Character Design, Aimer
Song Performance","Koyoharu Gotouge : Original Creator, Haruo Sotozaki : Director, Akira Matsushima : Character Design, Aimer :
Song Performance”
…
```

Leyendo el csv

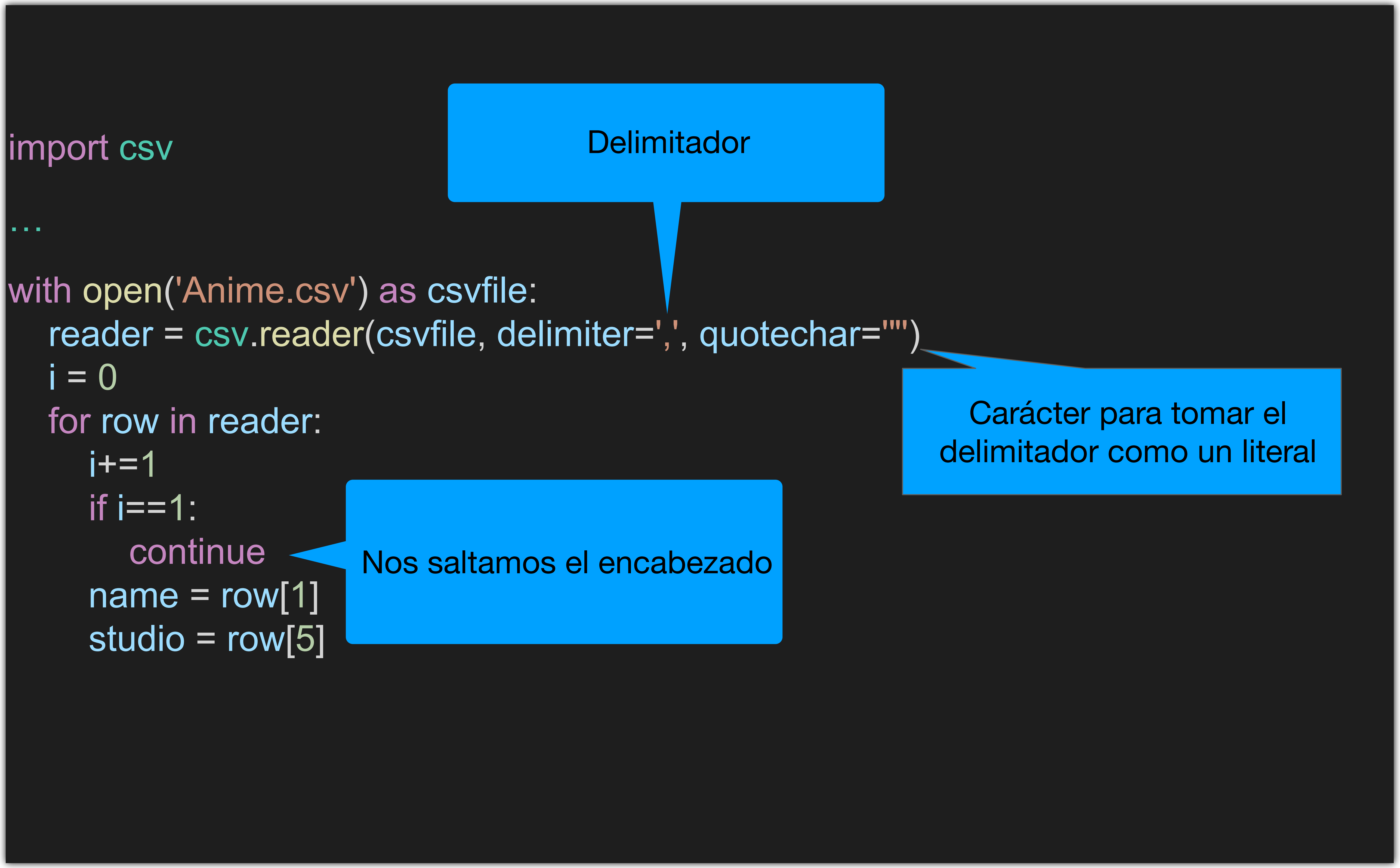


```
Rank,Name,Japanese_name,Type,Episodes,Studio,Release_season,Tags,Rating,Release_year,End_year,Description,Content_Warning,Related_Mange,Related_anime,Voice_actors,staff
1,Demon Slayer: Kimetsu no Yaiba - Entertainment District Arc, Kimetsu no Yaiba: Yuukaku-hen,TV   ,,ufotable,Fall ,"Action, Adventure, Fantasy, Shounen, Demons, Historical, Martial Arts, Orphans, Siblings, Swordplay, Based on a Manga, Explicit Violence",4.6,2021.0,,"'Tanjiro and his friends accompany the Hashira Tengen Uzui to an entertainment district where Tengen’s female ninja agents were gathering information on a demon before they suddenly disappeared. In order to investigate, Tanjiro and the others disguise themselves as woman to sneak in!'",Explicit Violence,Demon Slayer: Kimetsu no Yaiba,"Demon Slayer: Kimetsu no Yaiba, Demon Slayer: Kimetsu no Yaiba Movie - Mugen Train, Demon Slayer: Kimetsu no Yaiba - Mugen Train","Inosuke Hashibira :
Aka                                                      Agatsuma : Hiro Shimono, Daki :
ki                                                       oyoharu Gotouge

Character Design, Aimer
Song Performance","Koyoharu Gotouge : Or                    sushima : Character Design, Aimer :
Song Performance”
…
```

Multiples tags separados por comas…

Pero todo el campo tag esta encerrado por comillas dobles para desambiguar las comas.

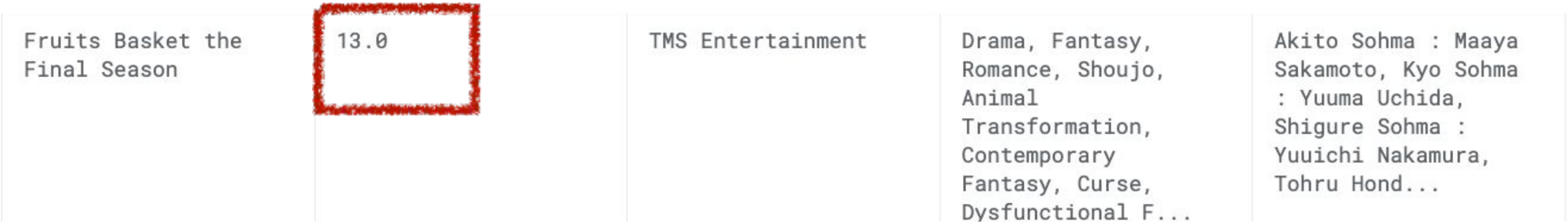
Leyendo el csv



```
import csv

…

with open('Anime.csv') as csvfile:
    reader = csv.reader(csvfile, delimiter=',', quotechar='"')
    i = 0
    for row in reader:
        i+=1
        if i==1:
            continue
        name = row[1]
        studio = row[5]
```

Delimitador
Carácter para tomar el delimitador como un literal
Nos saltamos el encabezado

Parseando el número de episodios



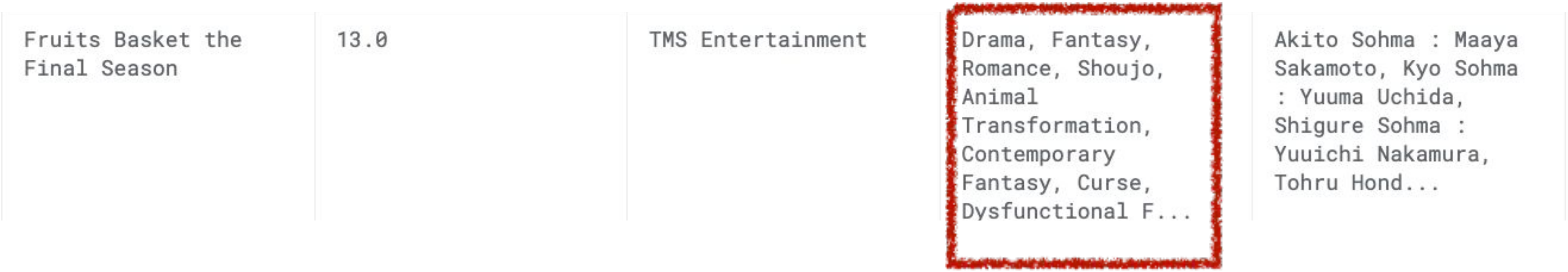
Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...

El número de episodios puede ser nulo

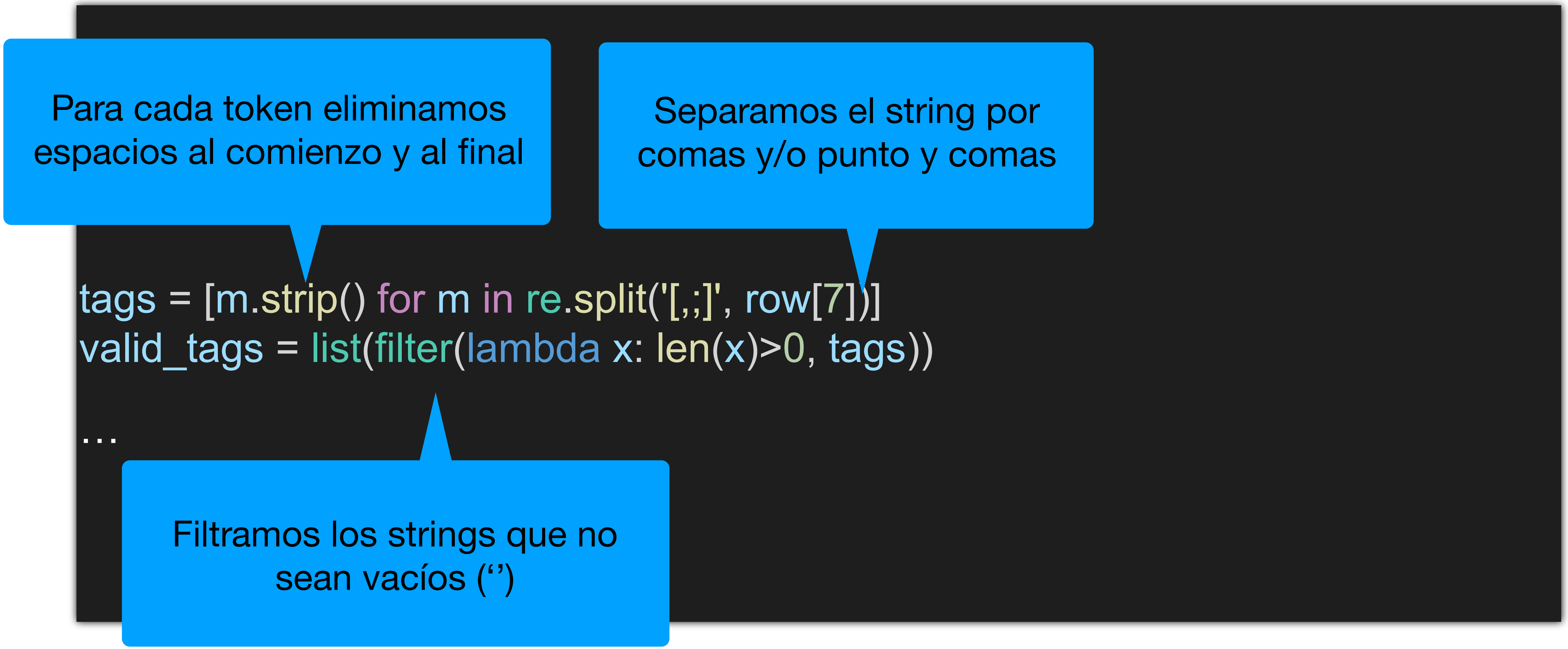
```
...

episodes = int(row[4].split('.')[0]) if row[4] else None

...
```

“13.0” lo transformamos a 13

Parseando los tags



Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...



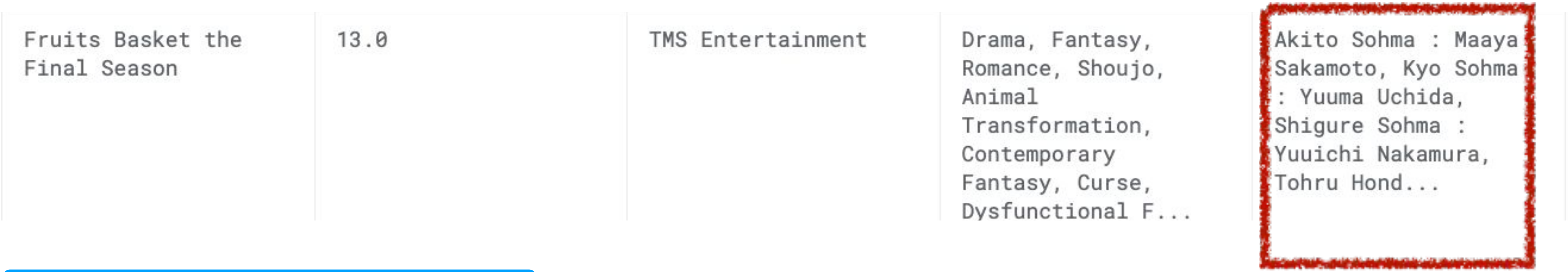
```
tags = [m.strip() for m in re.split('[,;]', row[7])]
valid_tags = list(filter(lambda x: len(x)>0, tags))

...
```

Para cada token eliminamos espacios al comienzo y al final

Separamos el string por comas y/o punto y comas

Filtramos los strings que no sean vacíos (‘’)

Parseando los actores y personajes



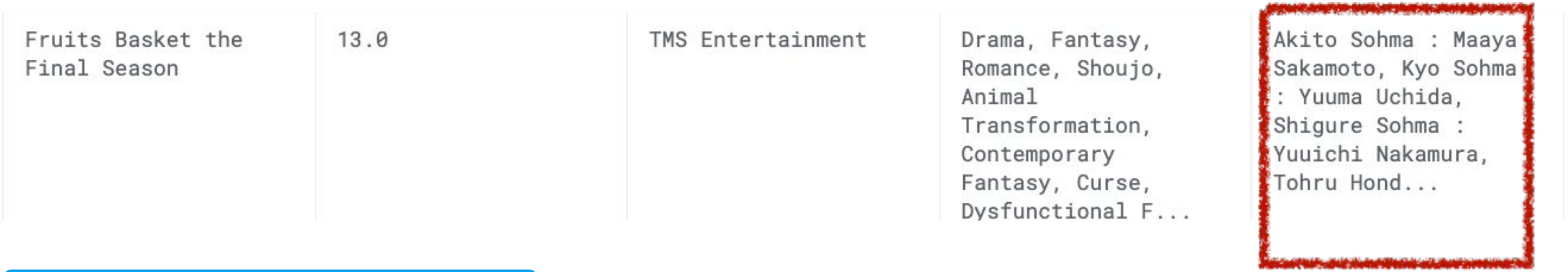
Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...

Este es un ejemplo del campo "voice actors"

"Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Honda : Manaka Iwami, Yuki Sohma : Nobunaga Shimazaki, Arisa Uotani : Atsumi Tanezaki, Hatsuharu Sohma : Makoto Furukawa, Isuzu Sohma : Aki Toyosaki, Natsuki Takaya
Original Creator, Yoshihide Ibata
Director & Episode Director & Storyboard, Taku Kishimoto
Screenplay & Series Composition, Masaru Yokoyama
Music, Masaru Shindou
Character Design & Chief Animation Director, Baek-Ryun Chae
Photography Director, Youko Koyama
Art Director, Mika Sugawara
Color Design"

!!!Hay mucha basura en este campo!!!!

Parseando los actores y personajes



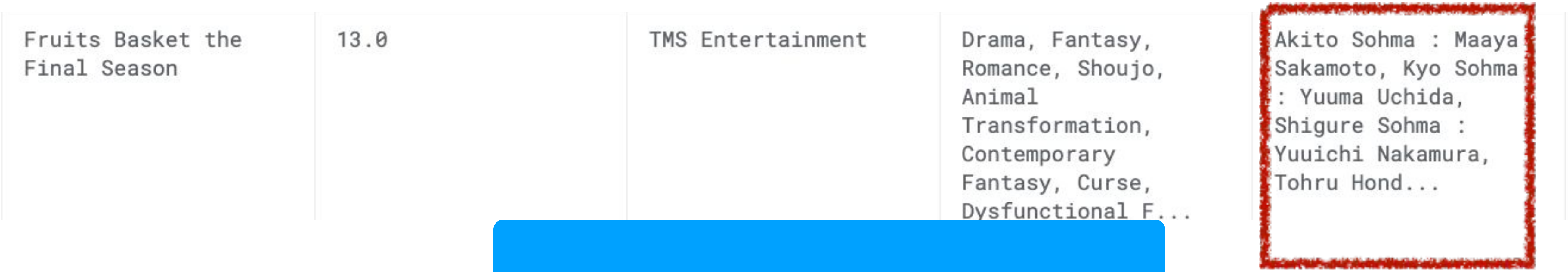
Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...

Este es un ejemplo del campo “voice actors”

```
"Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Honda : Manaka Iwami, Yuki
Sohma : Nobunaga Shimazaki, Arisa Uotani : Atsumi Tanezaki, Hatsuharu Sohma : Makoto Furukawa, Isuzu Sohma : Aki Toyosaki, Natsuki
Takaya
Original Creator, Yoshihide Ibata
Director & Episode Director & Storyboard, Taku Kishimoto
Screenplay & Series Composition, Masaru Yokoyama
Music, Masaru Shindou
Character Design & Chief Animation Director, Baek-Ryun Chae
Photography Director, Youko Koyama
Art Director, Mika Sugawara
Color Design"
```

Filtramos los elementos que tengan un “:” al medio usando la expresión regular “([^\:]+)\:([^\,]+)”

Parseando los actores y personajes



Fruits Basket the Final Season	13.0	TMS Entertainment	Drama, Fantasy, Romance, Shoujo, Animal Transformation, Contemporary Fantasy, Curse, Dysfunctional F...	Akito Sohma : Maaya Sakamoto, Kyo Sohma : Yuuma Uchida, Shigure Sohma : Yuuichi Nakamura, Tohru Hond...

Filtramos los elementos que tengan un “:” al medio usando la expresión regular “([^\:]+)\:([^\,]+)”

```
…

voice_actors = [m.strip() for m in re.split('[,;]', row[15])
]
voice_actors_ok = list(filter(lambda x:
        re.match(‘([^\:]+)\:([^\,]+)', x), voice_actors))
valid_voice_actors = [
    [m.split(':')[0].strip(), m.split(':')[1].strip()]
    for m in voice_actors_ok]

…
```

Generamos una lista de tuplas (personaje, actor)

Insertando los objetos

Creamos una función auxiliar para buscar el elemento por nombre y devolver el id…

```
…

def findOrInsert(table, name):
    cur.execute("select id from "+table+" where name=%s limit 1", [name])
    r = cur.fetchone()
    if(r):
        return r[0]
    else:
        cur.execute("insert into "+table+" (name) values (%s) returning id", [name])
        return cur.fetchone()[0]

…
```

… o insertarlo y devolver el nuevo id

Insertando los objetos

Insertamos estudios, tags, voice actors y personajes, guardando los ids respectivos.

```
…

studio_id = findOrInsert('studio', studio.strip())

tags_id = []
for tag in valid_tags:
    tags_id.append(findOrInsert('tag', tag))

vacs_id = []
for vac in valid_voice_actors:
    vacs_id.append([findOrInsert('character', vac[0]), findOrInsert('voice_actor', vac[1])])


…
```

Insertando los objetos

Este puede ser un proceso de prueba y error por lo que no queremos insertar un objeto más de una vez.

Primero buscamos si ya lo insertamos antes (asumiendo que no hay dos animes con el mismo nombre).

```
…

cur.execute("select id from anime where name=%s limit 1", [name])
r = cur.fetchone()
anime_id = None
if(r):
    anime_id = r[0]
else:
    cur.execute("insert into anime (name, episodes) values (%s, %s) returning id", [name, episodes])
    anime_id = cur.fetchone()[0]


…
```

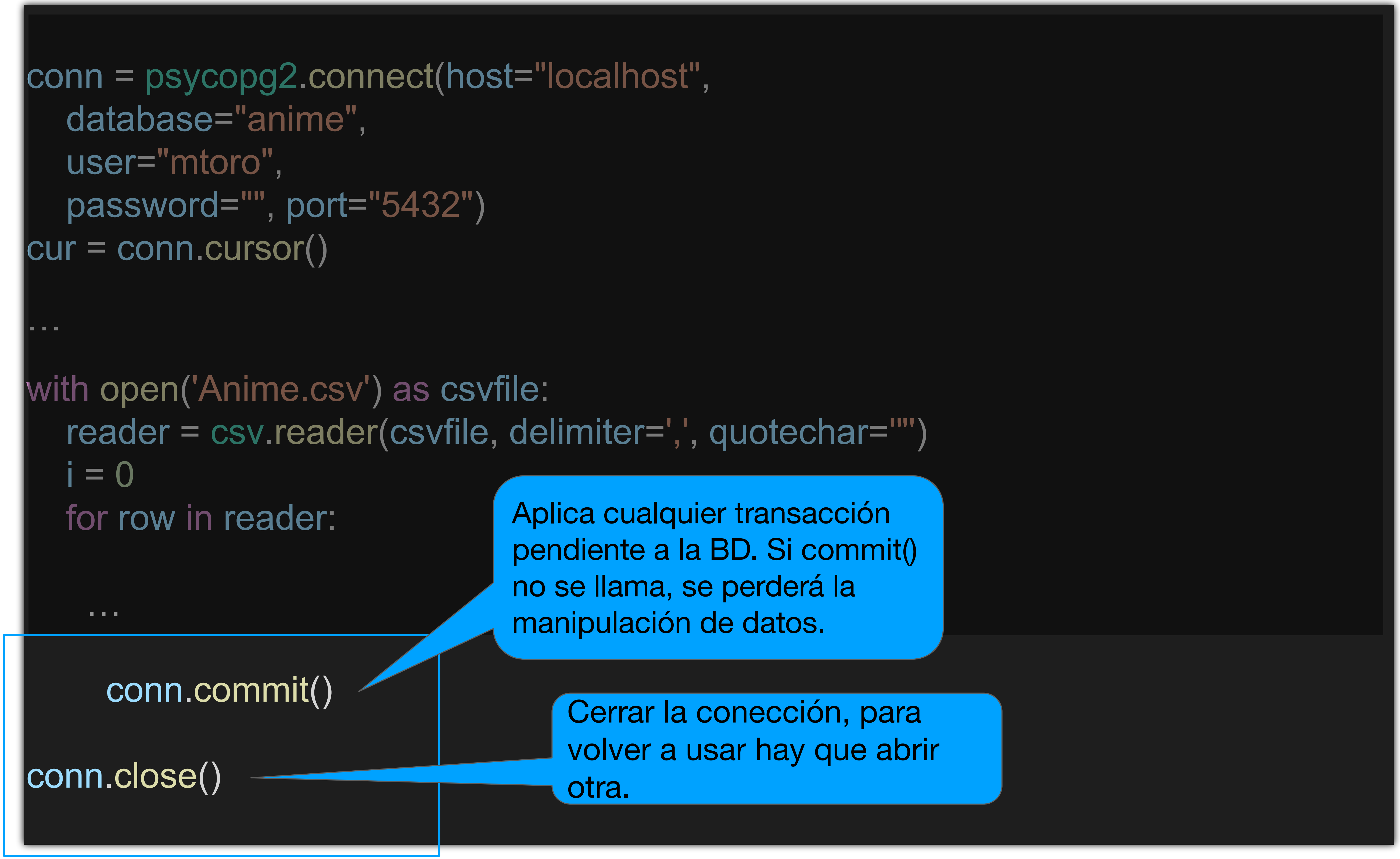
Si ya existe obtenemos su id, de lo contrario lo insertamos y obtenemos su id.

Insertando los objetos

Insertamos las filas de las relaciones entre tablas preguntando siempre si es que no está en la base de datos.

```
…

if(anime_id):
    for tag_id in tags_id:
        cur.execute("select * from anime_tag where (anime_id, tag_id) = (%s, %s) limit 1", [anime_id, tag_id])
        if(not cur.fetchone()):
            cur.execute("insert into anime_tag (anime_id, tag_id) values (%s, %s)", [anime_id, tag_id])

    for vac_id in vacs_id:
        cur.execute("select * from anime_voice_actor_character where (anime_id, voice_actor_id, character_id) = (%s, %s, %s)",
[anime_id, vac_id[1], vac_id[0]])
        if(not cur.fetchone()):
            cur.execute("insert into anime_voice_actor_character (anime_id, voice_actor_id, character_id) values (%s, %s, %s)",
[anime_id, vac_id[1], vac_id[0]])





…
```

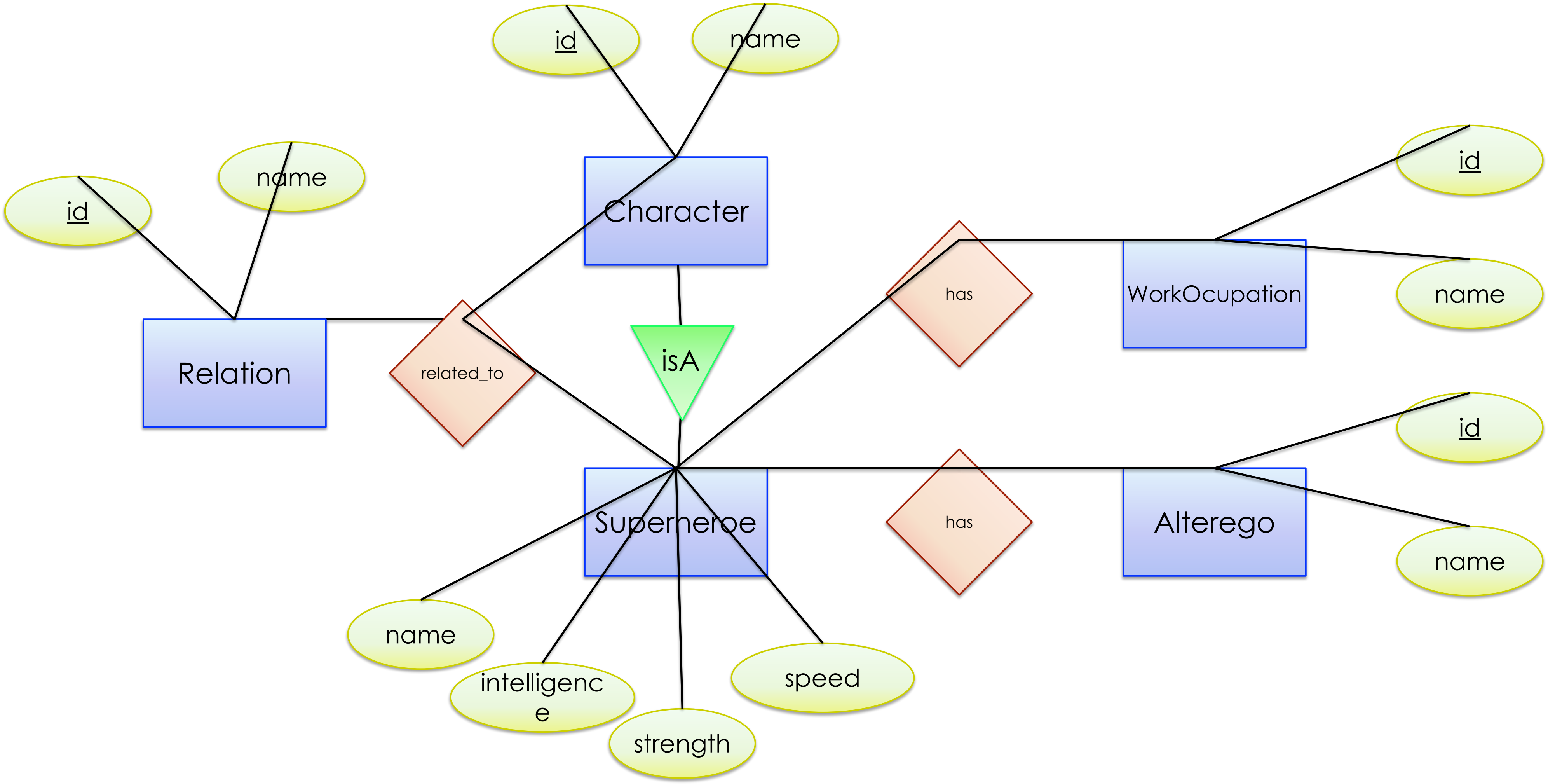
Confirmando los cambios y cerrando la conexión



```
conn = psycopg2.connect(host="localhost",
    database="anime",
    user="mtoro",
    password="", port="5432")
cur = conn.cursor()

…

with open('Anime.csv') as csvfile:
    reader = csv.reader(csvfile, delimiter=',', quotechar='"')
    i = 0
    for row in reader:

        …

        conn.commit()

conn.close()
```

Aplica cualquier transacción pendiente a la BD. Si commit() no se llama, se perderá la manipulación de datos.

Cerrar la conección, para volver a usar hay que abrir otra.


id
name
Character
id
name
Relation
related_to
isA
has
WorkOcupation
id
name
id
Superheroe
has
Alterego
name
name
intelligenc
e
strength
speed

Preguntas?