

SERIOUS SQL LIVE

WEEK 5: 29TH JAN

BY DANNY MA



AGENDA:

- Intro (5 mins)
- Marketing Analytics Case Study Brief (15 mins)
- Exploring the Data (30 mins)
- Understanding SQL Requirements (30 mins)

WHAT WE'VE LEARNED SO FAR

SELECT & SORT

- Basic SELECT ✓
- LIMIT result rows ✓
- ORDER BY ASC/DESC ✓
- Multi column sort ✓

RECORD COUNTS & DISTINCT VALUES

- Column Aliases
 - DISTINCT * and column(s) ✓
 - COUNT DISTINCT
 - GROUP BY basics
 - Percentage column
- window
functions

WHAT WE'VE LEARNED SO FAR

SUMMARY STATISTICS

- Mean, Median, Mode
- Min, Max, Range
- Standard Deviation
- Variance 
- Percentiles

- The Empirical Rule
- Distribution Skew
- ROUND function with CAST
- Implementing math in SQL

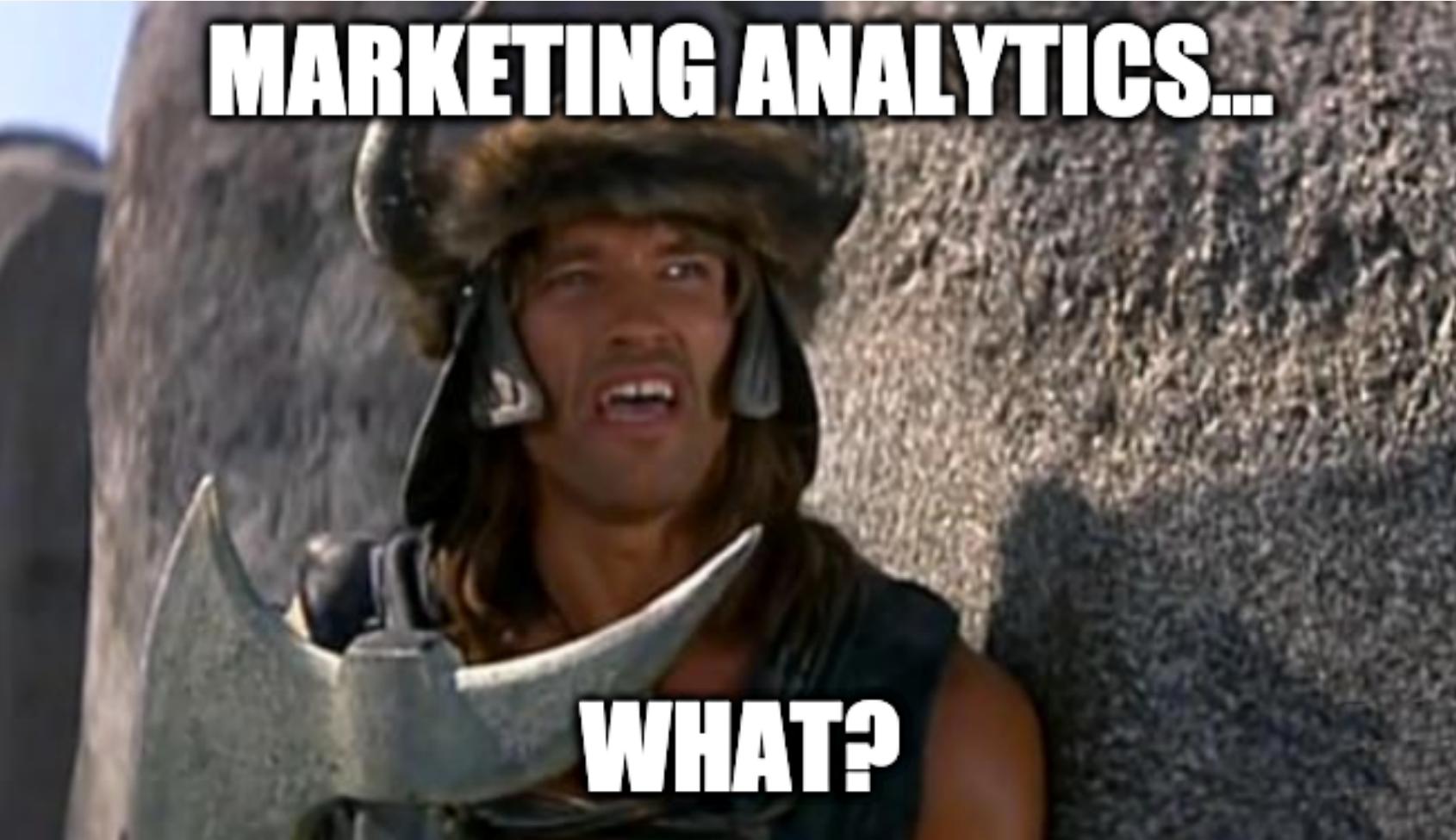
central tendency

WHAT WE'VE LEARNED SO FAR

Distribution Functions

- Cumulative distribution function
- NTILE vs CUME_DIST for percentiles
 - 1st
 - 100th
- Removing outliers with temp table
- Cumulative distribution chart
- WIDTH_BUCKET for bucketing data
- Frequency distribution chart

MARKETING ANALYTICS CASE STUDY



MARKETING ANALYTICS...

WHAT?



YOU GET AN EMAIL

AND YOU GET AN EMAIL

YOU ALL GET AN EMAIL

AND YOU GET AN EMAIL

REAL CAMPAIGN ANALYST PROJECT

5 W's

WHO?

- DVD Rentals Co customers
- Marketing Analytics team

WHAT?

- Fill an email template w/ data points
- Explore available data

WHERE?

- PostgreSQL database
- Cloud / On-prem / local

WHEN?

- timelines
- handover or transition points
- Value: \$\$\$
- purpose → import

WHY?

DVD RENTAL CO

Personalized recommendations from our very own team of film aficionados!



TRAVEL • ↗

You've watched 10 Travel films, that's 4 more than the DVD Rental Co average and puts you in the top 10% of Travel Gurus!

Your expertly chosen recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]



SCI-FI • -

You've watched 6 Sci-Fi films, making up 25% of your entire viewing history!

Your hand-picked recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]



RUFUS PAWS • •

You've watched 5 films featuring Rufus Paws! Here are some other films Rufus stars in that might interest you!

[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]

REQUIREMENT #1

The screenshot shows a user interface for movie recommendations. On the left, there's a photo of a woman in a hat and sunglasses looking at a camera. A green circle with the number '1' is overlaid on the image. To the right, there are two sections. The top section is titled 'TRAVEL' in a red oval, with a red arrow pointing to it from handwritten text above. It contains the text: 'You've watched 10 Travel films, that's 4 more than the DVD Rental Co average and puts you in the top 10% of Travel Gurus!'. Below this, under 'Your expertly chosen recommendations:', there are three placeholder links: '[Film #1 title goes here]', '[Film #2 title goes here]', and '[Film #3 title goes here]'. The bottom section is titled 'SCI-FI' in a red oval, with a red arrow pointing to it from handwritten text above. It contains the text: 'You've watched 6 Sci-Fi films, making up 25% of your entire viewing history!'. Below this, under 'Your hand-picked recommendations:', there are three placeholder links: '[Film #1 title goes here]', '[Film #2 title goes here]', and '[Film #3 title goes here]'. There is also a small image of a futuristic city skyline on the right.

Top 2 Movie categories for each customer

Select
customer_id,
category,
count (#) as dvd-rental_count
from ???
group by

ranking

REQUIREMENT #2



TRAVEL

You've watched 10 Travel films, that's 4 more than the Daily average and puts you in the top 10% of Travel Gurus!

Your expertly chosen recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]



SCI-FI 2

You've watched 10 Sci-Fi films, making up 25% of your total viewing.

Your hand-picked recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]

flag out users w/ no recommendations

Film recommendations for top 2 categories

- Max 3 films per category
- Must not have watched recommended films before

#1

not seen
most watched

REQUIREMENT #3

DVD RENTAL CO

Personalized recommendations from our very own team of film aficionados!

3

TRAVEL

You've watched 10 Travel films, that's 4 more than the DVD Rental Co average and puts you in the top 10% of Travel Gurus!

Your expertly curated recommendations:

[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]

SCI-FI

You've watched 6 Sci-Fi films, making up 25% of your entire viewing history!

Your hand-picked recommendations:

[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]

Individual customer Top category insights

- Number of films watched — count
- Comparison to DVD Rental Co average — avg ~ groupby curt, category
- Top X% ranking (percentile / bucketing curve-dist)

REQUIREMENT #4

DVD RENTAL CO

Personalized recommendations from our very own team of film aficionados!



TRAVEL

You've watched 10 Travel films, that's 4 more than the DVD Rental Co average and puts you in the top 10% of Travel Gurus!

Your expertly chosen recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]



SCI-FI

4

You've watched 6 Sci-Fi films, making up 25% of your entire viewing history!

Your hand-picked recommendations:
[Film #1 title goes here]
[Film #2 title goes here]
[Film #3 title goes here]

Individual customer 2nd category insights

integer floor division



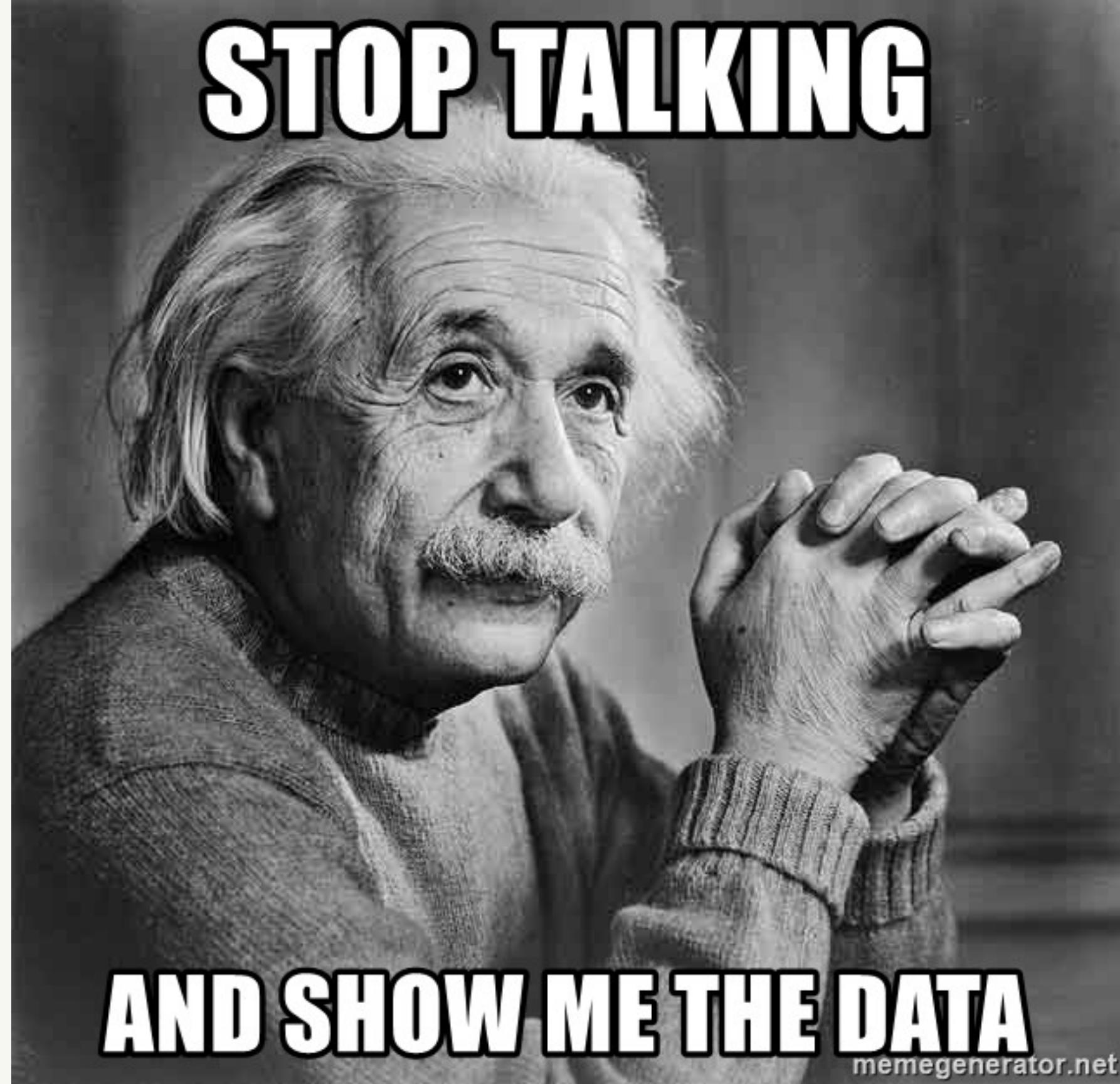
- **Number of films watched** *2nd category* *Overall count of films*
- **Percentage of viewing history**

REQUIREMENT #5



Favorite actor insights & recommendations

- Number of films watched — group by actor
- Max 3 film recommendations — count (cutover) — group by film
- Previously watched & top 2 category recommendations must not be included



STOP TALKING

AND SHOW ME THE DATA

memegenerator.net

ENTITY RELATIONSHIP DIAGRAM



1. dvd_rentals.rental

rental	1
rental_id	integer
rental_date	timestamp
inventory_id	integer
customer_id	smallint
return_date	timestamp
staff_id	smallint
last_update	timestamp

`SELECT *`
`FROM dvd_rentals.rental`
`WHERE customer_id = 130;`

DVD

HR analytics

rental_id	rental_date	inventory_id	customer_id	return_date	staff_id	last_update
1	2005-05-24 22:53:30	367	130	2005-05-26 22:04:30	1	2006-02-15 21:30:53
746	2005-05-29 09:25:10	4272	130	2005-06-02 04:20:10	2	2006-02-15 21:30:53
1630	2005-06-16 07:55:01	2413	130	2005-06-19 06:38:01	1	2006-02-15 21:30:53
1864	2005-06-17 01:39:47	1815	130	2005-06-24 19:39:47	2	2006-02-15 21:30:53
2163	2005-06-17 23:46:16	2600	130	2005-06-22 22:48:16	2	2006-02-15 21:30:53
2292	2005-06-18 07:37:48	173	130	2005-06-20 02:45:48	2	2006-02-15 21:30:53
2535	2005-06-19 01:39:04	901	130	2005-06-28 01:33:04	2	2006-02-15 21:30:53
2982	2005-06-20 08:38:29	2574	130	2005-06-28 13:21:29	1	2006-02-15 21:30:53
4339	2005-07-07 18:41:42	3215	130	2005-07-08 13:00:42	1	2006-02-15 21:30:53
4485	2005-07-08 01:07:54	2614	130	2005-07-16 03:19:54	2	2006-02-15 21:30:53
6353	2005-07-11 20:48:56	699	130	2005-07-21 00:11:56	1	2006-02-15 21:30:53
7181	2005-07-27 08:14:34	2788	130	2005-07-28 03:09:34	1	2006-02-15 21:30:53

2. dvd_rentals.inventory

rental	
rental_id	integer
rental_date	timestamp
inventory_id	integer
customer_id	smallint
return_date	timestamp
staff_id	smallint
last_update	timestamp

Inventory	
inventory_id	integer
film_id	smallint
store_id	smallint
last_update	timestamp

ID for each physical DVD

inventory_id	<u>film_id</u>	store_id	last_update
1	1	1	2006-02-15 05:09:17
2	1	1	2006-02-15 05:09:17
3	1	1	2006-02-15 05:09:17
4	1	1	2006-02-15 05:09:17
5	1	2	2006-02-15 05:09:17
6	1	2	2006-02-15 05:09:17
7	1	2	2006-02-15 05:09:17
8	1	2	2006-02-15 05:09:17

```
SELECT *
FROM dvd_rentals.inventory
WHERE film_id = 1;
```

3. dvd_rentals.film

The diagram illustrates the structure of the `dvd_rentals.film` database, specifically focusing on the `rental`, `Inventory`, and `film` tables.

Tables:

- rental**: Contains columns for `rental_id` (integer), `rental_date` (timestamp), `inventory_id` (integer), `customer_id` (smallint), `return_date` (timestamp), `staff_id` (smallint), and `last_update` (timestamp).
- Inventory**: Contains columns for `inventory_id` (integer), `film_id` (smallint), `store_id` (smallint), and `last_update` (timestamp).
- film**: Contains columns for `film_id` (integer), `title` (character varying(255)), `description` (text), `release_year` (year), `language_id` (smallint), `original_language_id` (smallint), `rental_duration` (smallint), `rental_rate` (numeric(4, 2)), `length` (smallint), `replacement_cost` (numeric(5, 2)), `rating` (character varying(5)), `last_update` (timestamp), `special_features` (text), and `fulltext` (tsvector).

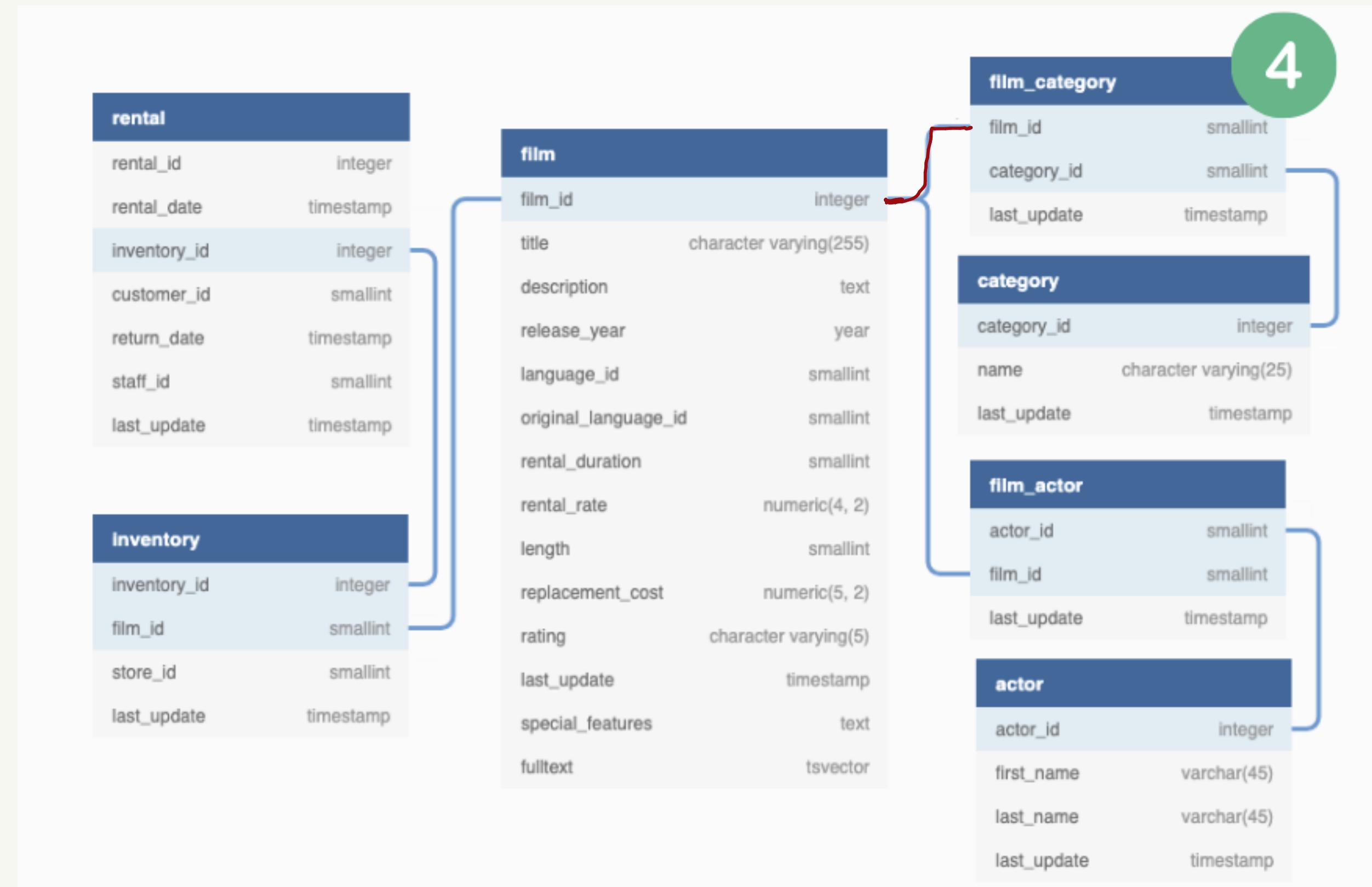
Relationships:

- 1**: A red bracket connects the `inventory_id` column in the `rental` table to the `film_id` column in the `film` table.
- 2**: A red bracket connects the `film_id` column in the `Inventory` table to the `film_id` column in the `film` table.
- 3**: A red bracket connects the `film_id` column in the `rental` table to the `film_id` column in the `film` table.

identify duplicates

film_id	title	description					
1	ACADEMY DINOSAUR	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies					
2	ACE GOLDFINGER	A Astounding Epistle of a Database Administrator And a Explorer who must Find a Car in Ancient China					
3	ADAPTATION HOLES	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory					
4	AFFAIR PREJUDICE	A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank					
5	AFRICAN EGG	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Forensic Psychologist in The Gulf of Mexico					
release_year	language_id	original_language_id	rental_duration	rental_rate	length	replacement_cost	rating
2006	1	null	6	0.99	86	20.99	PG
2006	1	null	3	4.99	48	12.99	G
2006	1	null	7	2.99	50	18.99	NC-17
2006	1	null	5	2.99	117	26.99	G
2006	1	null	6	2.99	130	22.99	G
last_update	special_features	fulltext					
2006-02-15 05:03:42.000	["Deleted Scenes", "Behind the Scenes"]	'academi':1 'battl':15 'canadian':20 'dinosaur':2 'drama':5 'epic':4 'feminist':8 'mad':11 'must':14 'rocki':21 'scientist':12 'teacher':17					
2006-02-15 05:03:42.000	["Trailers", "Deleted Scenes"]	'ace':1 'administr':9 'ancient':19 'astound':4 'car':17 'china':20 'databas':8 'epistl':5 'explor':12 'find':15 'goldfing':2 'must':14					
2006-02-15 05:03:42.000	["Trailers", "Deleted Scenes"]	'adapt':1 'astound':4 'baloon':19 'car':11 'factori':20 'hole':2 'lumberjack':8,16 'must':13 'reflect':5 'sink':14					
2006-02-15 05:03:42.000	["Commentaries", "Behind the Scenes"]	'affair':1 'chase':14 'documentari':5 'fanci':4 'frisbe':8 'lumberjack':11 'monkey':16 'must':13 'prejudic':2 'shark':19 'tank':20					
2006-02-15 05:03:42.000	["Deleted Scenes"]	'african':1 'chef':11 'dentist':14 'documentari':7 'egg':2 'fast':5 'fast-pac':4 'forens':19 'gulf':23 'mexico':25 'must':16 'pace':6 'pastri':10 'psychologist':20 'pursu':17					

4. dvd_rentals.film_category



group by count
— deduplication

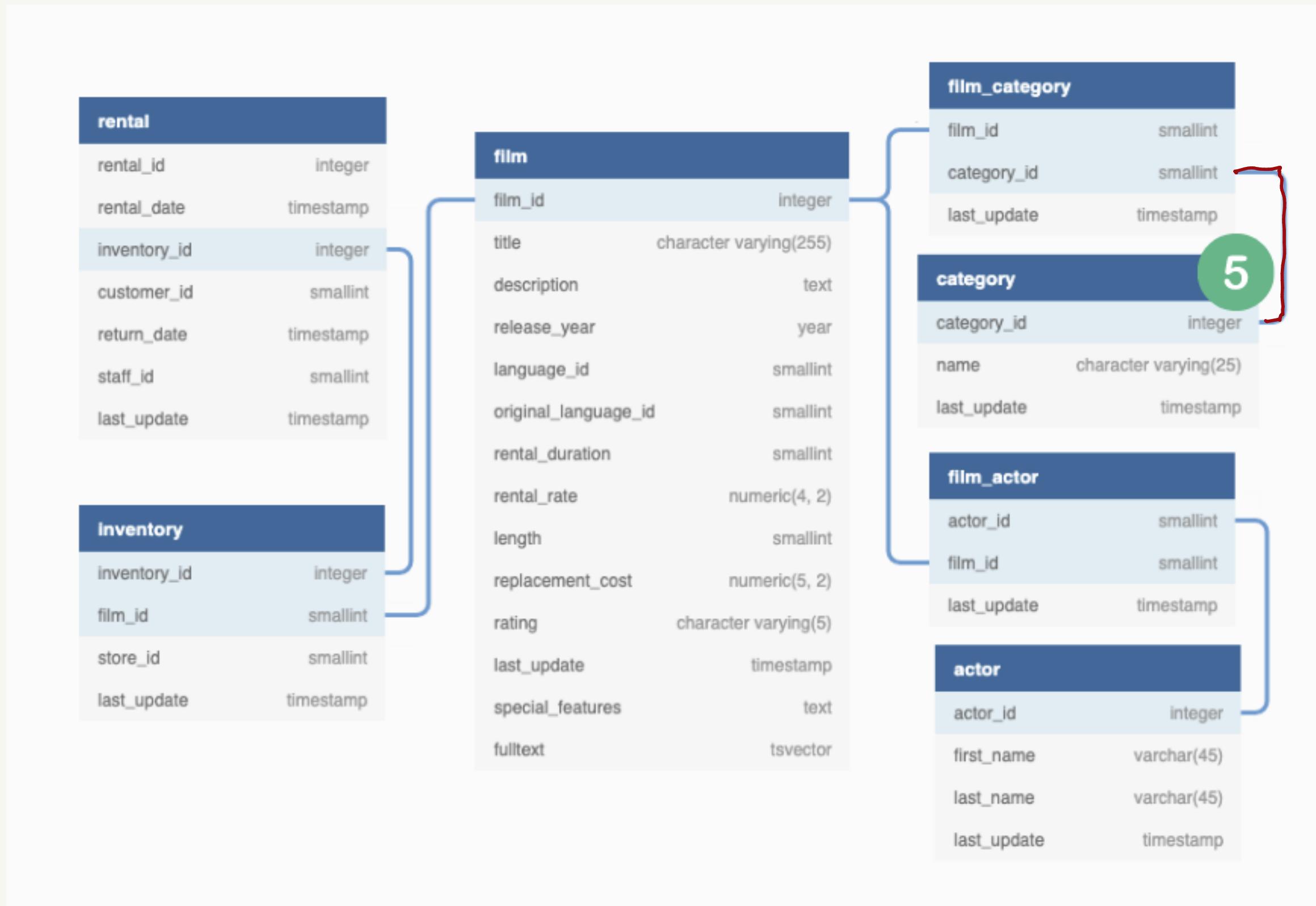
```

SELECT *
FROM dvd_rentals.film_category
LIMIT 5;

```

	film_id	category_id	last_update
1	1	6	2006-02-15 05:07:09
2	2	11	2006-02-15 05:07:09
3	3	6	2006-02-15 05:07:09
4	4	11	2006-02-15 05:07:09
5	5	8	2006-02-15 05:07:09

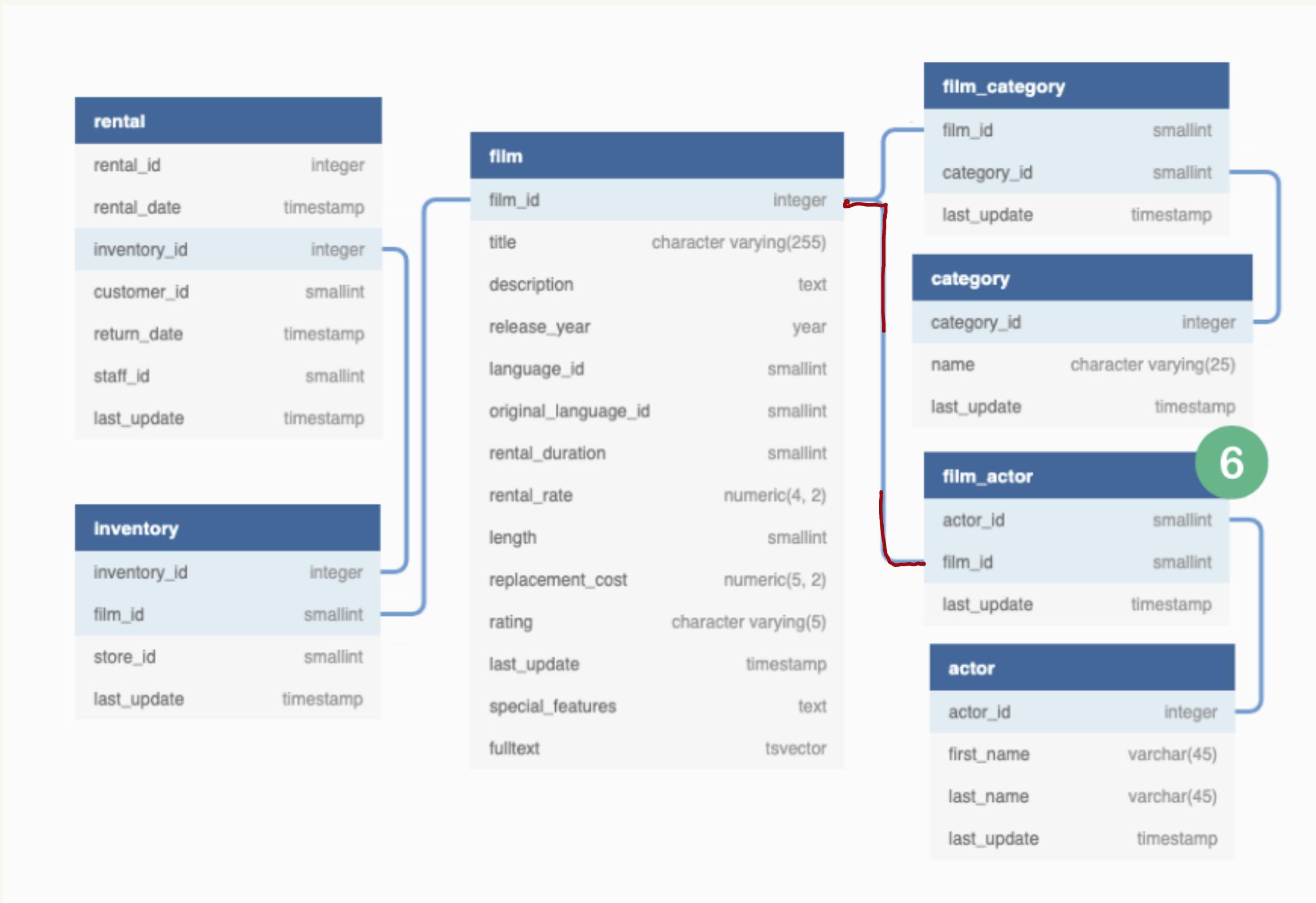
5. dvd_rentals.category



```
SELECT *
FROM dvd_rentals.category
LIMIT 5;
```

	category_id	name	last_update
1	Action	2006-02-15 04:46:27	
2	Animation	2006-02-15 04:46:27	
3	Children	2006-02-15 04:46:27	
4	Classics	2006-02-15 04:46:27	
5	Comedy	2006-02-15 04:46:27	

6. dvd_rentals.film_actor



```

SELECT *
FROM dvd_rentals.film_actor
WHERE actor_id = 1;
    
```

*many to many
relationships*

actor_id	film_id	last_update
1	1	2006-02-15 05:05:03
1	23	2006-02-15 05:05:03
1	25	2006-02-15 05:05:03
1	106	2006-02-15 05:05:03
1	140	2006-02-15 05:05:03
1	166	2006-02-15 05:05:03
1	277	2006-02-15 05:05:03
1	361	2006-02-15 05:05:03

7. dvd_rentals.actor



group by counts

```
SELECT *
FROM dvd_rentals.actor
LIMIT 5;
```

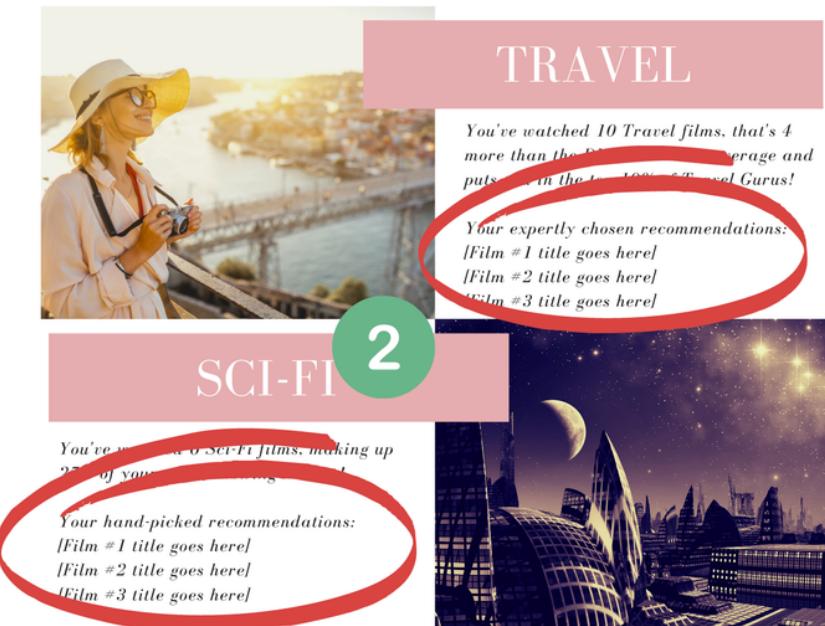
one to one mapping

	actor_id	first_name	last_name	last_update
1 .	PENELOPE	GUINNESS		2006-02-15 04:34:33
2	NICK	WAHLBERG		2006-02-15 04:34:33
3	ED	CHASE		2006-02-15 04:34:33
4	JENNIFER	DAVIS		2006-02-15 04:34:33
5	JOHNNY	LOLLOBRIGIDA		2006-02-15 04:34:33

Requirement #1: Top 2 Movie categories for each customer

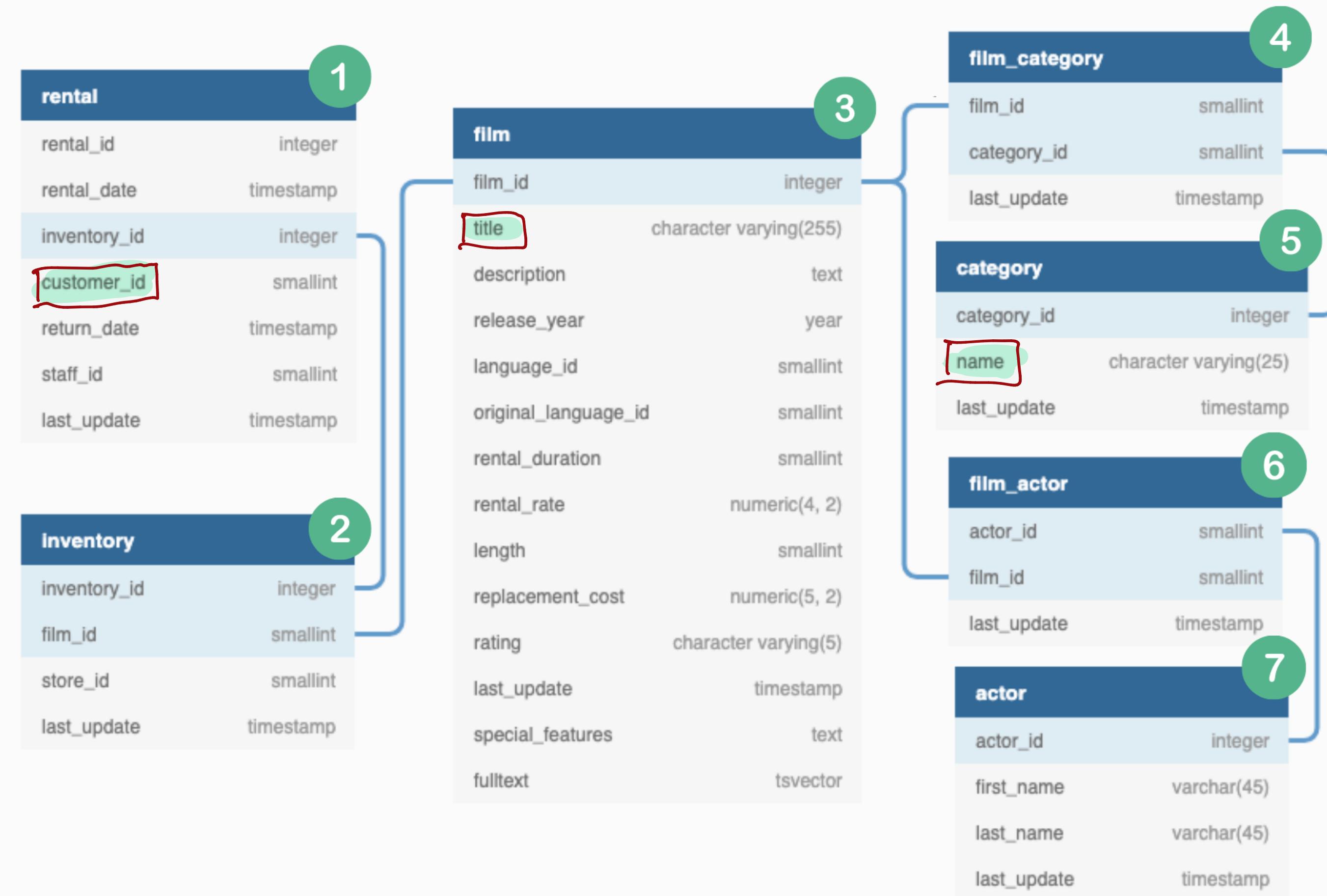


Requirement #2: Film Recommendations



Film recommendations for top 2 categories

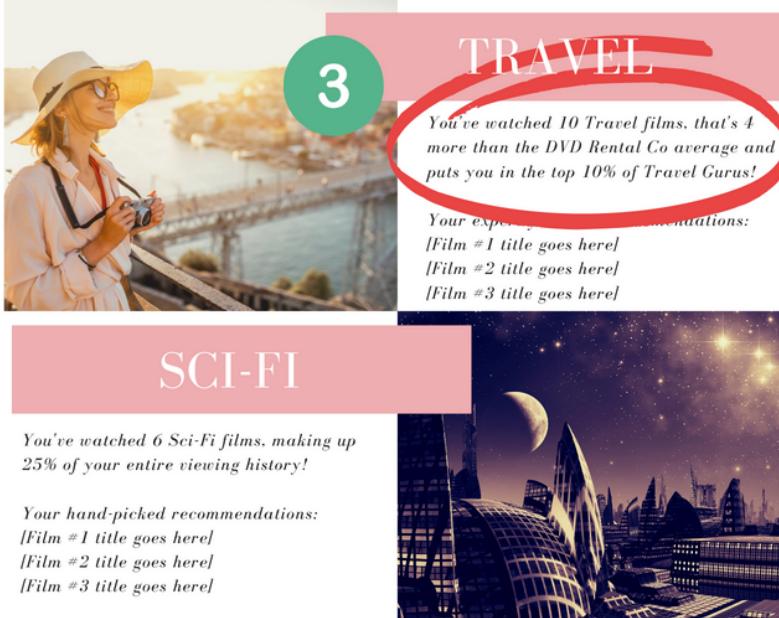
- Max 3 films per category
- Must not have watched recommended films before



Requirement #3: Top Category Insights

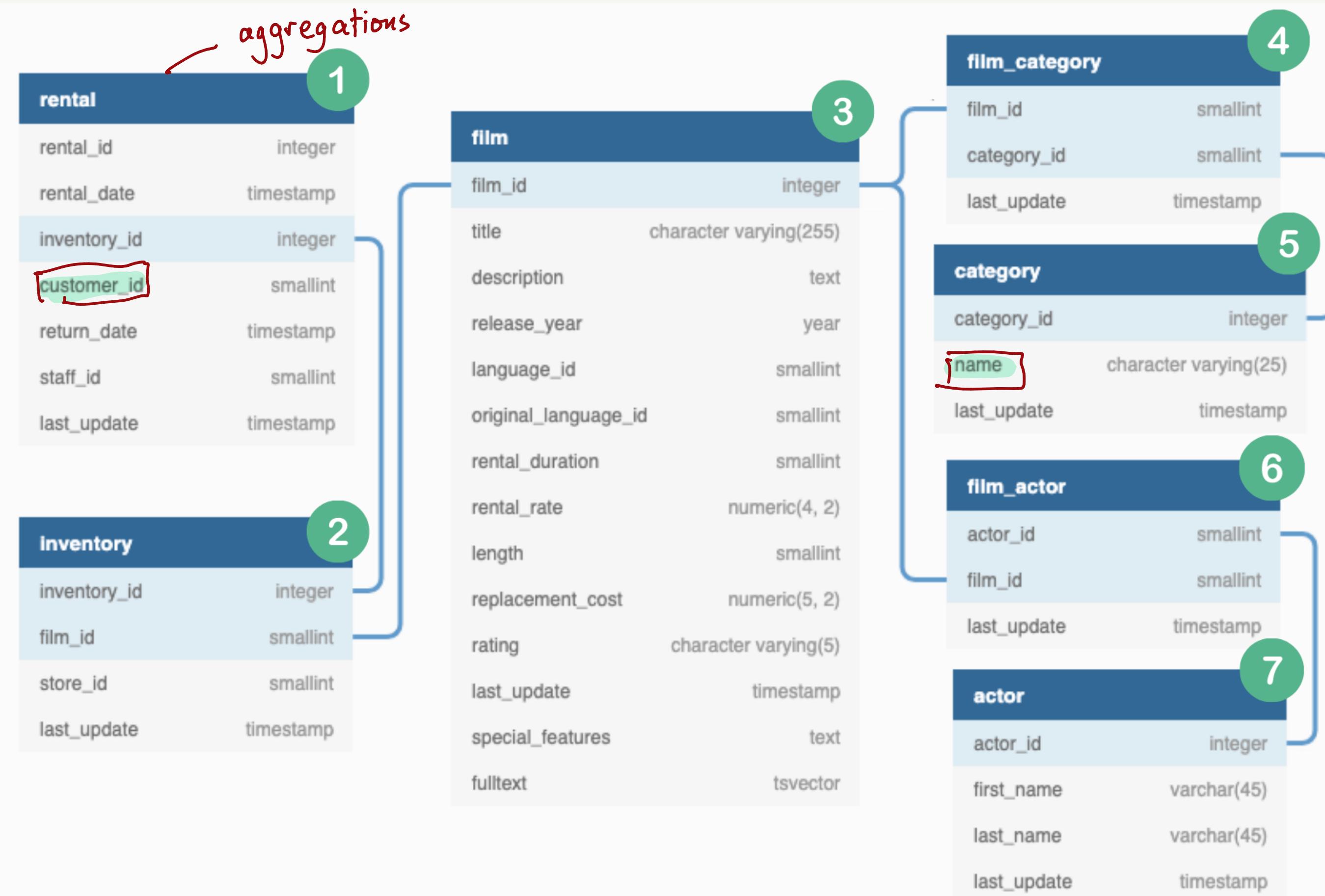
DVD RENTAL CO

Personalized recommendations from our very own team of film aficionados!



Individual customer Top category insights

- Number of films watched
- Comparison to DVD Rental Co average
- Top X% ranking



Requirement #4: 2nd Category Insights

DVD RENTAL CO

Personalized recommendations from our very own team of film aficionados!



TRAVEL

You've watched 10 Travel films, that's 4 more than the DVD Rental Co average and puts you in the top 10% of Travel Gurus!

Your expertly chosen recommendations:
/Film #1 title goes here!
/Film #2 title goes here!
/Film #3 title goes here!

4

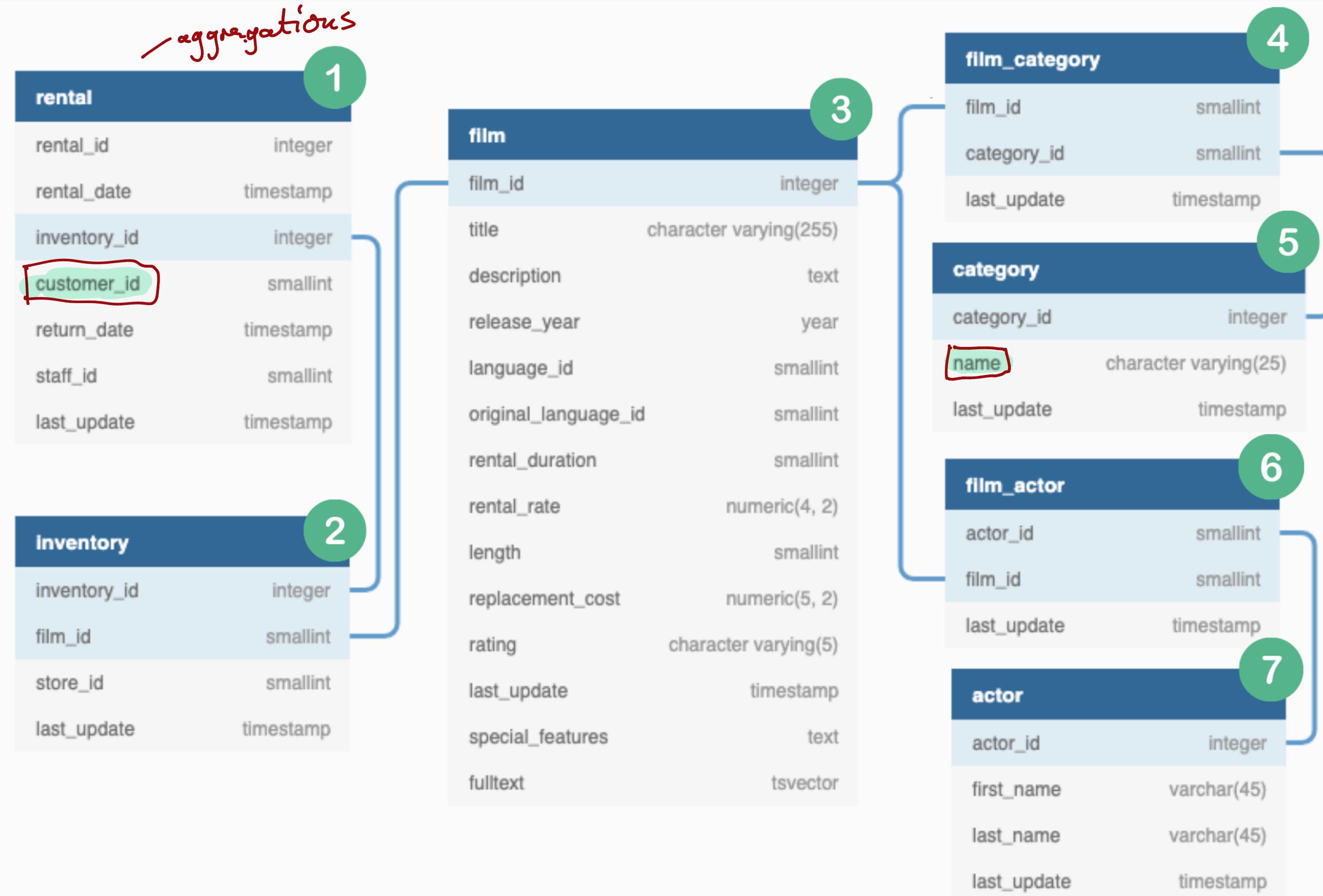


4

Individual customer 2nd category insights

- Number of films watched
- Percentage of viewing history

→ aggregations



Requirement #5: Favorite Actor Insights



RUFUS PAWS

You've watched 5 films featuring Rufus Paws! Here are some other films Rufus stars in that might interest you!

/Film #1 title goes here!
/Film #2 title goes here!
/Film #3 title goes here!

Favorite actor insights & recommendations

- Number of films watched
- Max 3 film recommendations
- Previously watched & top 2 category recommendations must not be included



TUTORIAL SUMMARY

- Marketing Analytics Case Study
- 5 W's (who/what/where/when/why)
- Campaign/business requirements
- Data exploration & ERD analysis
- Matching requirements to data