

SERIOUS SQL LIVE

WEEK 7: 12TH FEB

BY DANNY MA



AGENDA:

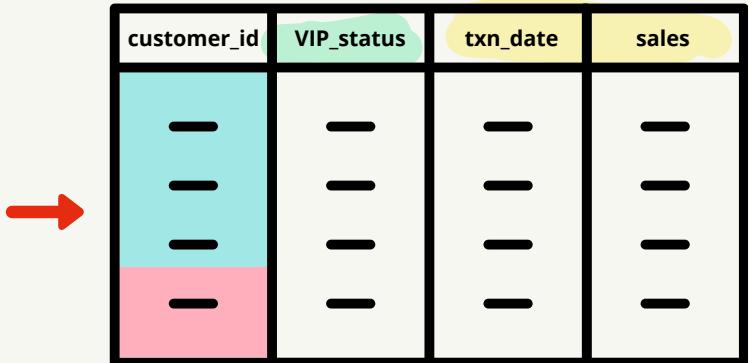
- Table Joins Recap (10 mins)
- Case Study Revision (10 mins)
- Applied Table Joins (40 mins)

INNER JOIN

left

customer_id	VIP_status
-	-
-	-
-	-
-	-

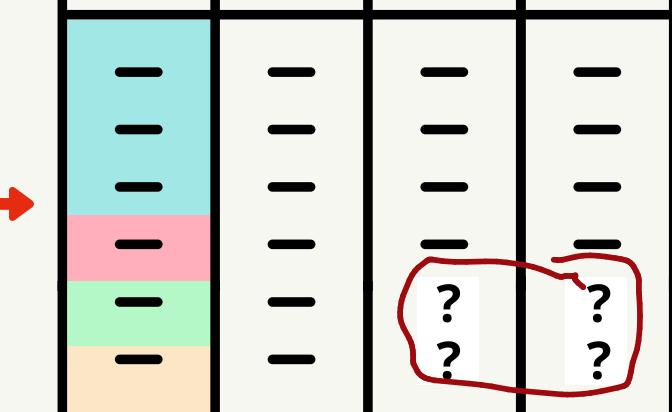
right



LEFT JOIN

1

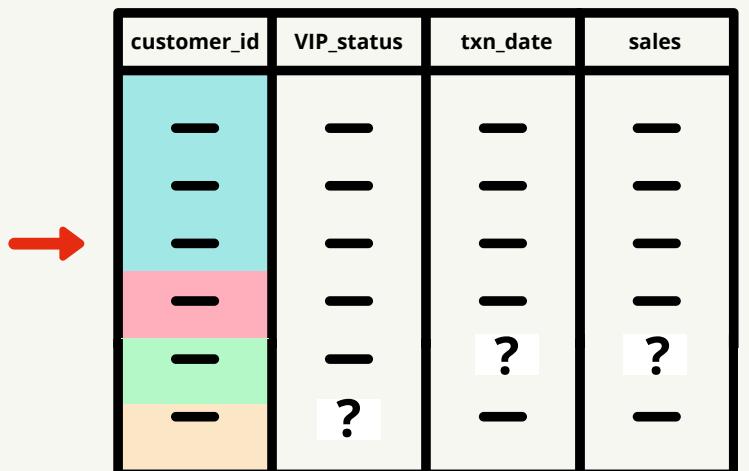
customer_id	VIP_status
-	-
-	-
-	-
-	-



FULL JOIN

A hand-drawn diagram showing two red, oval-shaped objects representing kidneys. The word "left" is written above the left kidney, and the word "right" is written above the right kidney.

customer_id	VIP_status
-	-
-	-
-	-



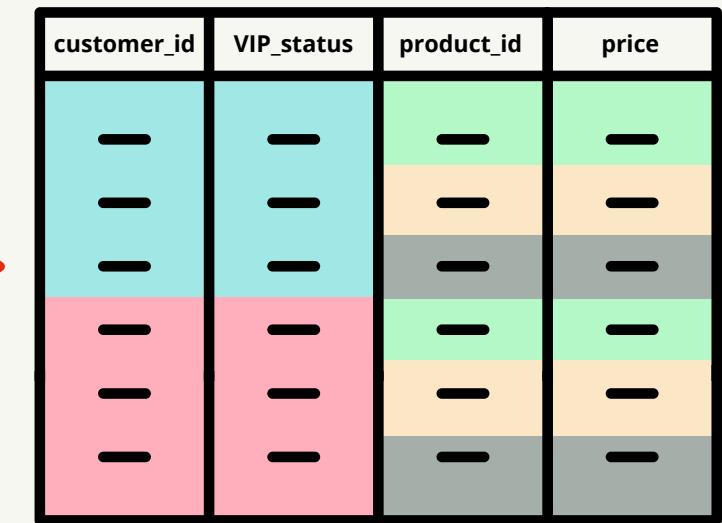
CROSS JOIN

customer_id	VIP_status
-	-
-	-



A diagram showing a table with two columns: "product_id" and "price". The table has four rows. The first row (top) has a light green background color. The second row has a light orange background color. The third row has a light grey background color. The fourth row (bottom) has a light green background color. A large red 'X' mark is placed to the left of the table, indicating that the structure shown is incorrect or invalid.

product_id	price
-	-
-	-
-	-

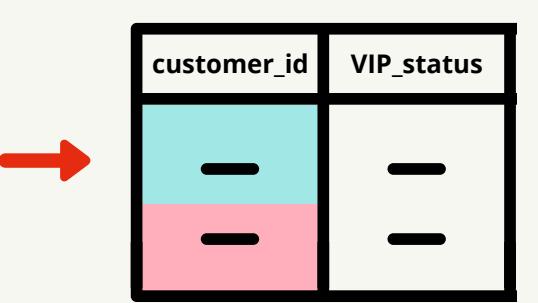


LEFT SEMI JOIN

right - look up

where exists

customer_id	VIP_status
-	-
-	-
-	-
-	-
-	-

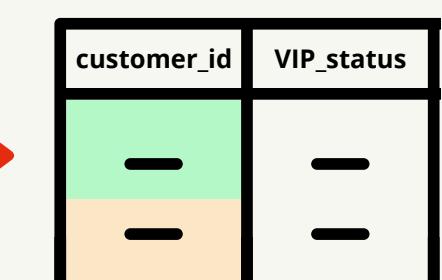


ANTI JOIN

right

customer_id	VIP_status
1	Yes
2	No
3	Yes
4	No

customer_id	txn_date
—	—
—	—
—	—
—	—
—	—



MARKETING ANALYTICS CASE STUDY

CASE STUDY OVERVIEW

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



1

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]



Top 2 Movie categories for each customer

Group by
Count
join tables

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]

Film recommendations for top 2 categories

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

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
- Comparison to DVD Rental Co average
- Top X% ranking

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]

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

• Number of films watched

• Percentage of viewing history

REQUIREMENT #5



Favorite actor insights & recommendations

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



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]



2 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]



- category_name : The name of the top 2 ranking categories
- rental_count : How many total films have they watched in this category
- average_comparison : How many more films has the customer watched compared to the average DVD Rental Co customer
- percentile : How does the customer rank in terms of the top X% compared to all other customers in this film category?
- category_percentage : What proportion of total films watched does this category make up?

customer_id	category_ranking	category_name	rental_count	average_comparison	percentile	category_percentage
1	1	Classics	6	4	2	19
1	2	Comedy	5	4	1	16
2	1	Sports	5	3	7	19
2	2	Classics	4	2	11	15
3	1	Action	4	2	14	15
3	2	Animation	3	1	39	12
4	1	Horror	3	2	22	14
4	2	Travel	2	1	57	9

RENTAL COUNT

final output

customer_id	category_name	rental_count
1	Classics	6
1	Comedy	5
2	Sports	5
2	Classics	4
3	Action	4
3	Animation	3
4	Horror	3
4	Travel	2

customer_id	category_name	rental_count
1	Classics	6
1	Comedy	5
1	Drama	4
1	Sports	2
1	Action	2
1	Music	2
1	Animation	2
1	New	2
1	Sci-Fi	2
1	Travel	1
1	Documentary	1
1	Family	1
1	Foreign	1
1	Games	1

customer_id	title	category_name
1	MINDS TRUMAN	Action
1	WOMEN DORADO	Action
1	DOORS PRESIDENT	Animation
1	BIKINI BORROWERS	Animation
1	PATIENT SISTER	Classics
1	MUSKeteers WAIT	Classics
1	DETECTIVE VISION	Classics
1	FROST HEAD	Classics
1	JEEPERS WEDDING	Classics
1	PATIENT SISTER	Classics
1	FIREBALL PHILADELPHIA	Comedy
1	CLOSER BANG	Comedy
1	FIREBALL PHILADELPHIA	Comedy
1	SNATCH SLIPPER	Comedy
1	FERRIS MOTHER	Comedy
1	ADAPTATION HOLES	Documentary
1	RACER EGG	Drama
1	SAVANNAH TOWN	Drama
1	DALMATIONS SWEDEN	Drama
1	LUCK OPUS	Drama
1	FINDING ANACONDA	Family
1	USUAL UNTOUCHABLES	Foreign
1	FIRE WOLVES	Games
1	CONFIDENTIAL INTERVIEW	Music
1	YOUTH KICK	Music
1	AMISTAD MIDSUMMER	New
1	JUMANJI BLADE	New
1	UNFORGIVEN ZOOLANDER	Sci-Fi
1	ATTACKS HATE	Sci-Fi
1	SATURDAY LAMBS	Sports
1	TALENTED HOMICIDE	Sports
1	EXPECATIONS NATURAL	Travel

Join Journey Part	Start	End	Foreign Key
Part 1	rental	inventory	inventory_id
Part 2	inventory	film	film_id
Part 3	film	film_category	film_id
Part 4	film_category	category	category_id

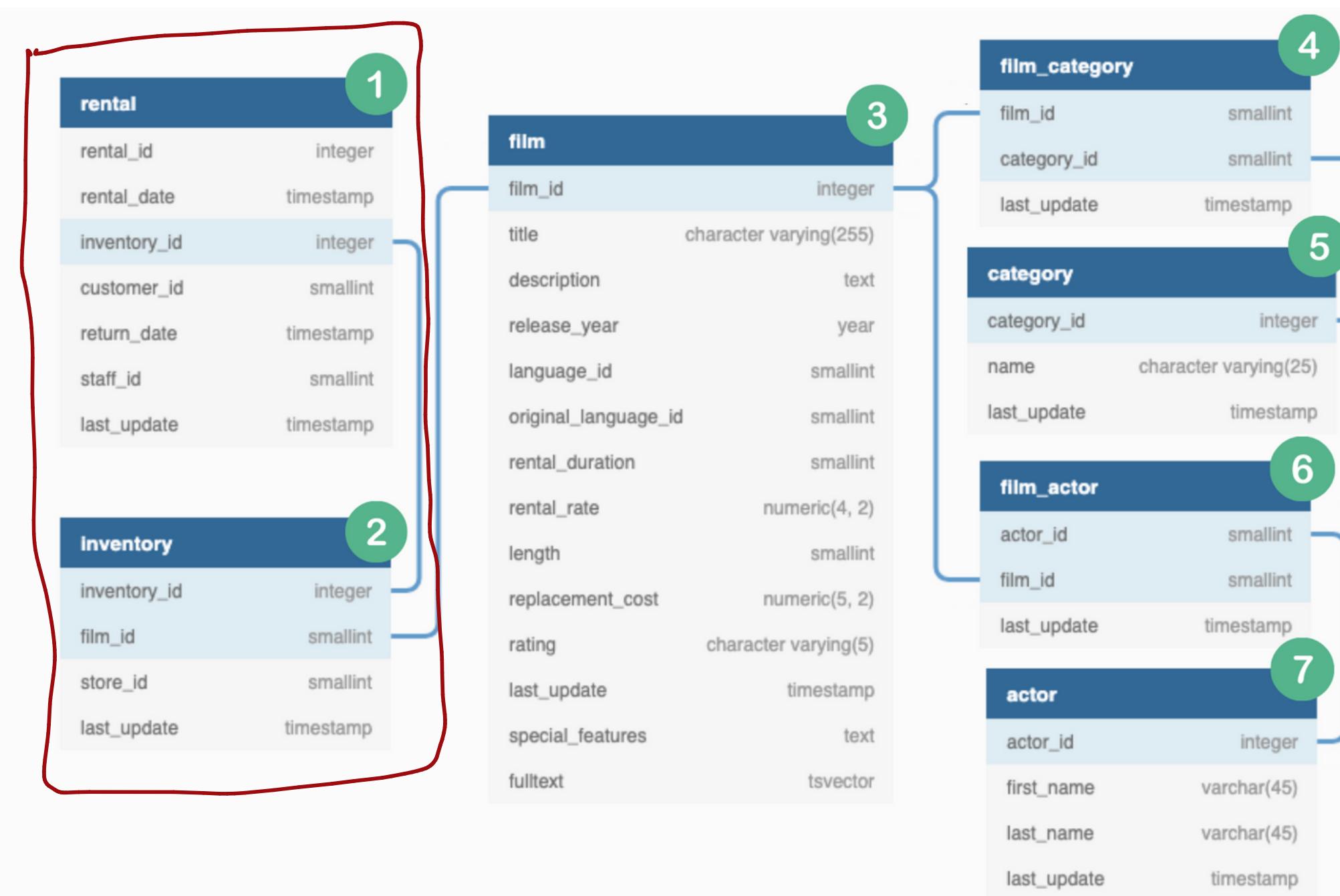


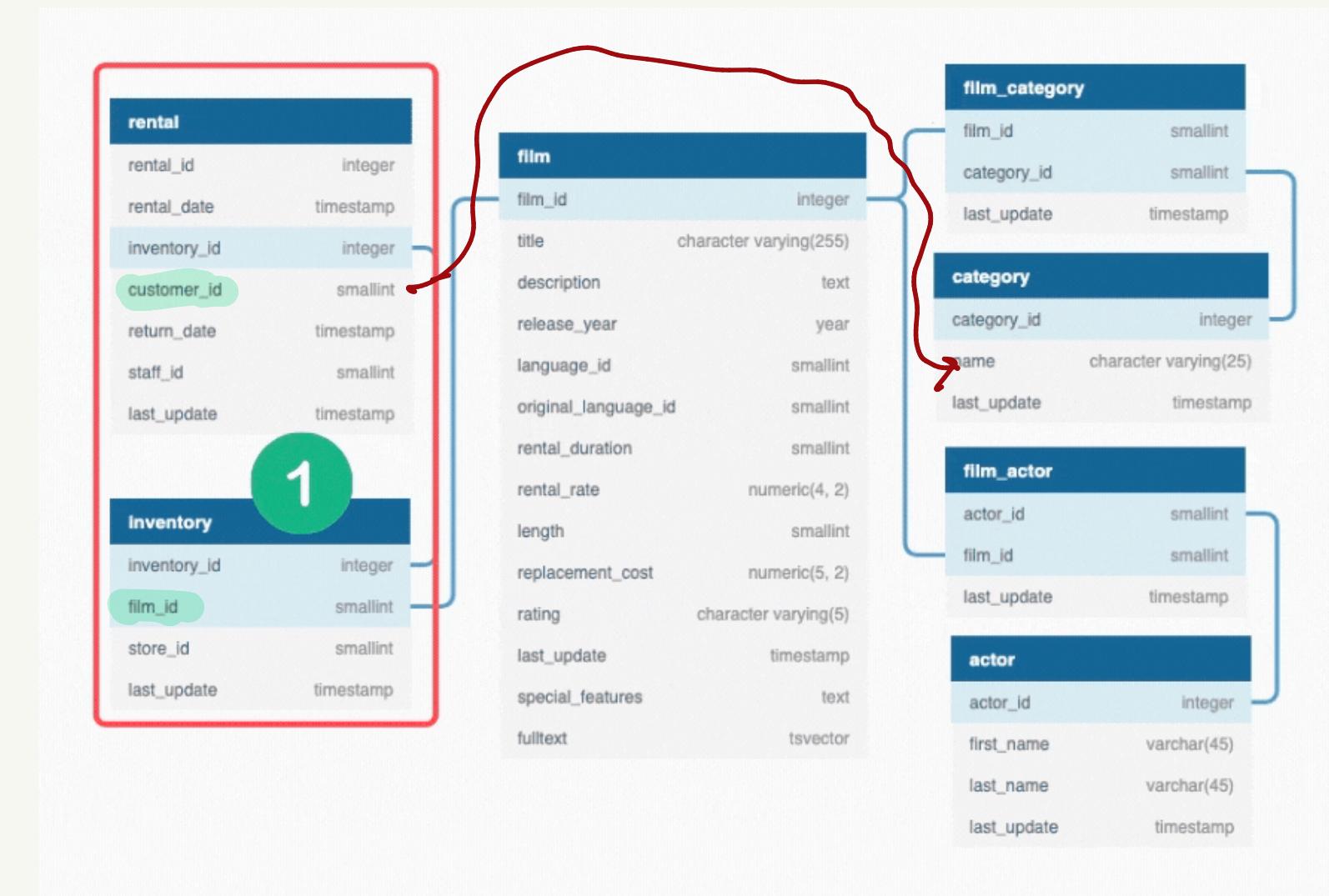
TABLE JOIN CHECKLIST

1. What is the purpose of joining these tables together?
2. What is the distribution of foreign key values for each table?
3. What is the overlap of key values?



Find
your
purpose

We need to keep all of the ~~customer~~ rental records from *dvd_rentals.rental* and match up each record with its equivalent *film_id* value from the *dvd_rentals.inventory* table.





**Using data to
confirm hypotheses**

**Using hypotheses
to confirm our data**

DATA HYPOTHESES

Relationships
within
table

- 1:1

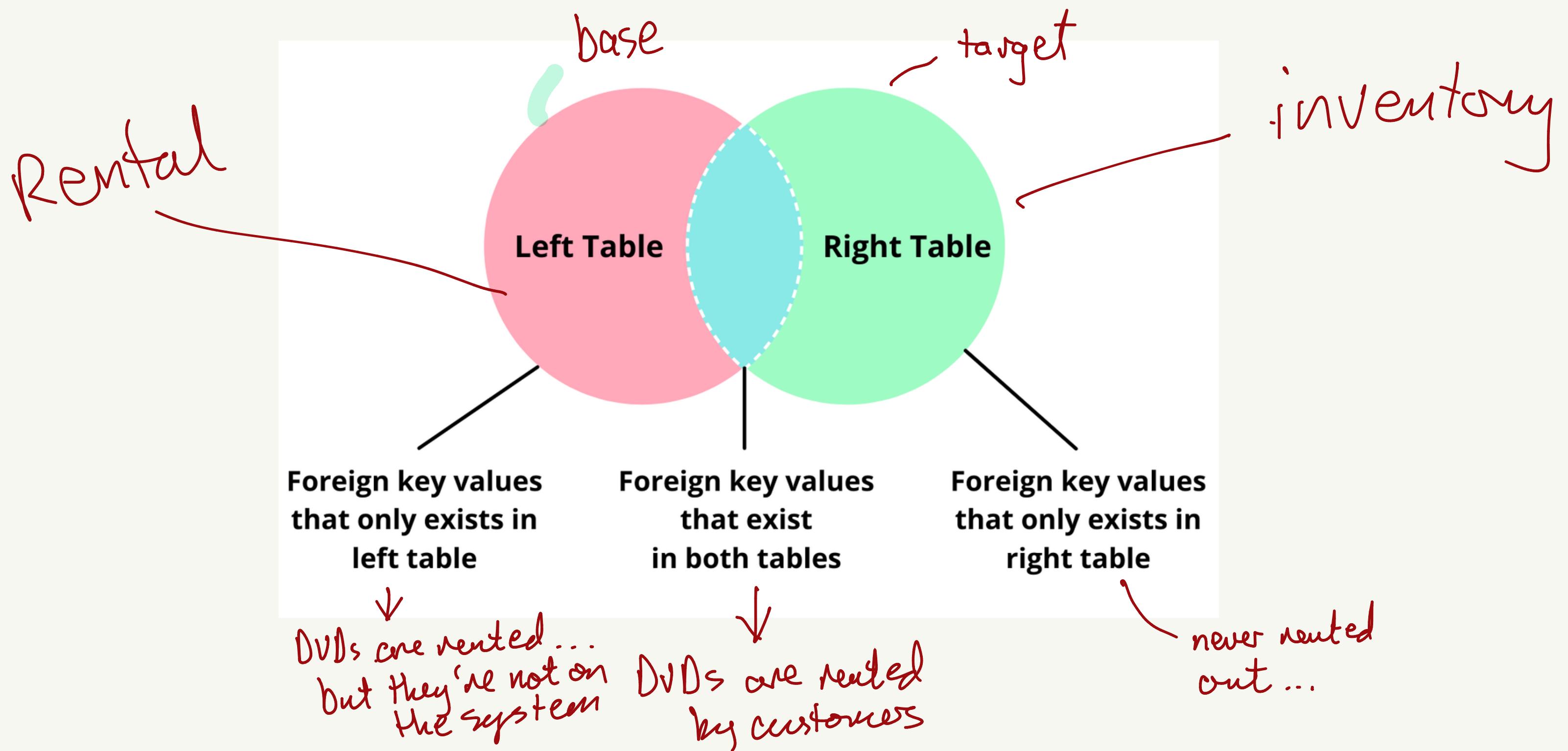
- 1:MANY

- MANY:1

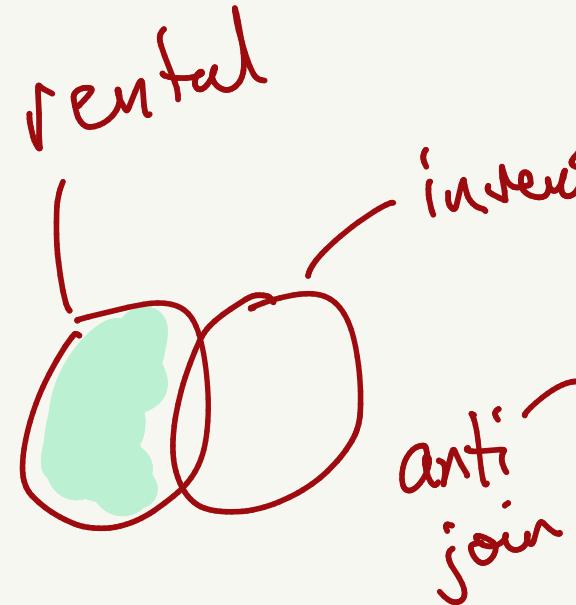
- MANY:MANY

- | 1: The number of unique *inventory_id* records will be equal in both *dvd_rentals.rental* and *dvd_rentals.inventory* tables ✓
- | 2: There are multiple records per unique *inventory_id* in the *dvd_rentals.rental* table
- | 3: There will be multiple *inventory_id* records per unique *film_id* value in the *dvd_rentals.inventory* table

TABLE OVERLAP ANALYSIS



LEFT ONLY

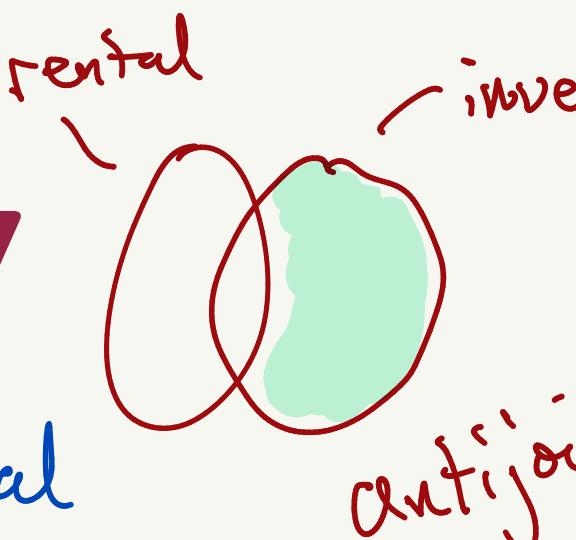


SELECT

```
COUNT(DISTINCT rental.inventory_id)
FROM dvd_rentals.rental — left
WHERE NOT EXISTS (
    SELECT inventory_id
    FROM dvd_rentals.inventory
    WHERE rental.inventory_id = inventory.inventory_id
);
```

right

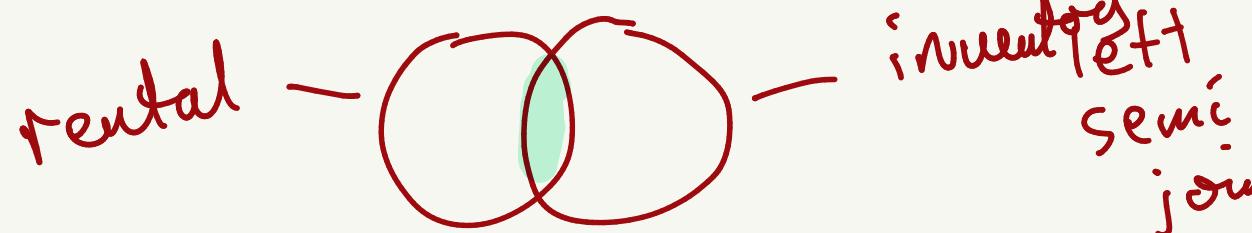
RIGHT ONLY



SELECT

```
COUNT(DISTINCT inventory.inventory_id)
FROM dvd_rentals.inventory — right
WHERE NOT EXISTS (
    SELECT inventory_id
    FROM dvd_rentals.rental
    WHERE rental.inventory_id = inventory.inventory_id
);
```

EXISTS IN BOTH



SELECT

```
COUNT(DISTINCT rental.inventory_id)
FROM dvd_rentals.rental
WHERE EXISTS (
    SELECT inventory_id
    FROM dvd_rentals.inventory
    WHERE rental.inventory_id = inventory.inventory_id
);
```

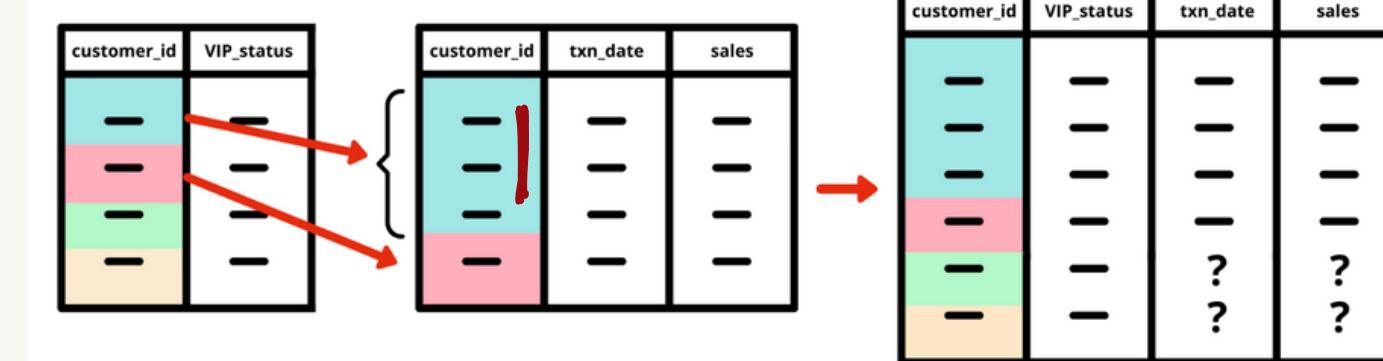
WHICH JOIN SHOULD I USE?

- Purpose
- Distribution
- Overlap — 

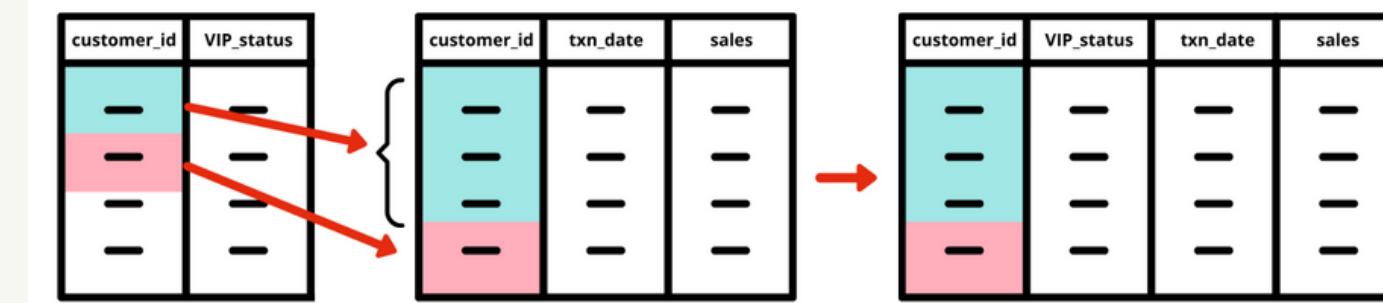
What output looks like

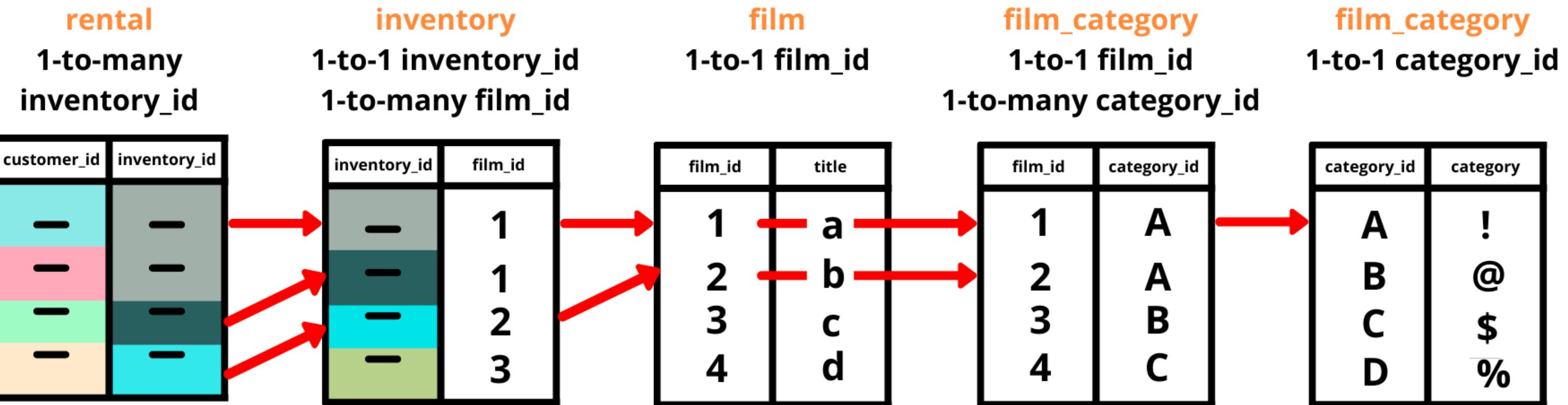
1:1
1:many
many:many

Left Join



Inner Join





customer_id	film_id	title	category
-	1	a	!
-	1	a	!
-	1	a	!
-	2	b	!

```

DROP TABLE IF EXISTS complete_joint_dataset;
CREATE TEMP TABLE complete_joint_dataset AS
SELECT
    rental.customer_id,
    inventory.film_id,
    film.title,
    film_category.category_id,
    category.name AS category_name
FROM dvd_rentals.rental
INNER JOIN dvd_rentals.inventory
    ON rental.inventory_id = inventory.inventory_id
INNER JOIN dvd_rentals.film
    ON inventory.film_id = film.film_id
INNER JOIN dvd_rentals.film_category
    ON film.film_id = film_category.film_id
INNER JOIN dvd_rentals.category
    ON film_category.category_id = category.category_id;

SELECT * FROM complete_joint_dataset limit 10;

```

customer_id	film_id	title	category_id	category
130	80	BLANKET BEVERLY	8	Family
459	333	FREAKY POCUS	12	Music
408	373	GRADUATE LORD	3	Children
333	535	LOVE SUICIDES	11	Horror
222	450	IDOLS SNATCHERS	3	Children
549	613	MYSTIC TRUMAN	5	Comedy
269	870	SWARM GOLD	11	Horror
239	510	LAWLESS VISION	2	Animation
126	565	MATRIX SNOWMAN	9	Foreign
399	396	HANGING DEEP	7	Drama

SUMMARY

1. Table Join Checklist (purpose, distribution & overlap)
2. Data Hypotheses (1:1 vs 1:many vs many:1 vs many:many)
3. Table Overlap Analysis