

Module 1- Data Cleaning - v + x lookup INDEX & MATCH

XLOOKUP

VS

VLOOKUP

S.No.	Name	Math	Language	History	Physics
1	Alex	45	62	100	99
2	Aron	30	90	66	72
3	Aubrey	99	70	51	80
4	Calista	82	52	74	66
5	Chase	92	96	90	58
6	Claris	38	43	52	97
7	Curt	59	76	72	50
8	Delaine	50	37	30	31
9	Enoch	34	35	51	36
10	Florene	45	77	32	87

Aron	
Subject	Marks
Language	90
Math	30
History	66
Physics	72

movies

financials

Sholay

Inception

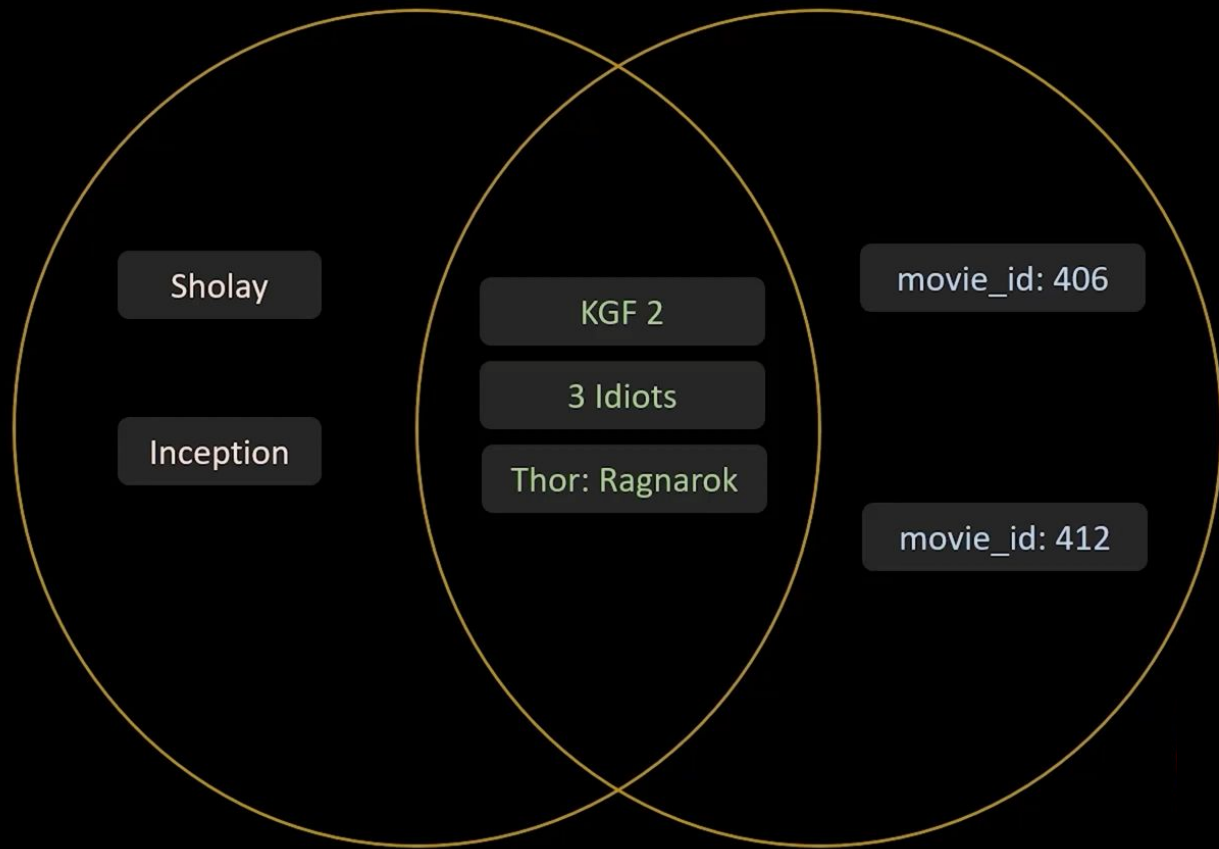
KGF 2

3 Idiots

Thor: Ragnarok

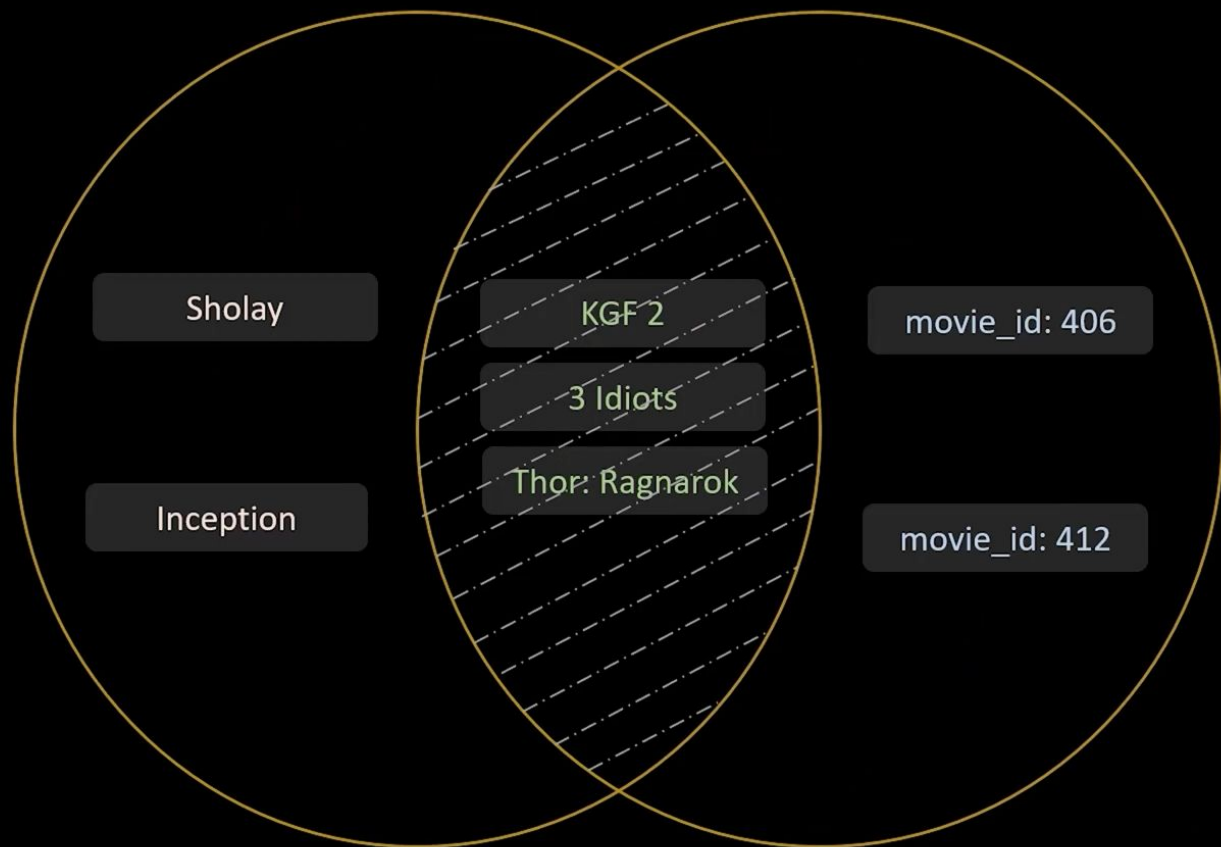
movie_id: 406

movie_id: 412



movies

financials



INNER JOIN

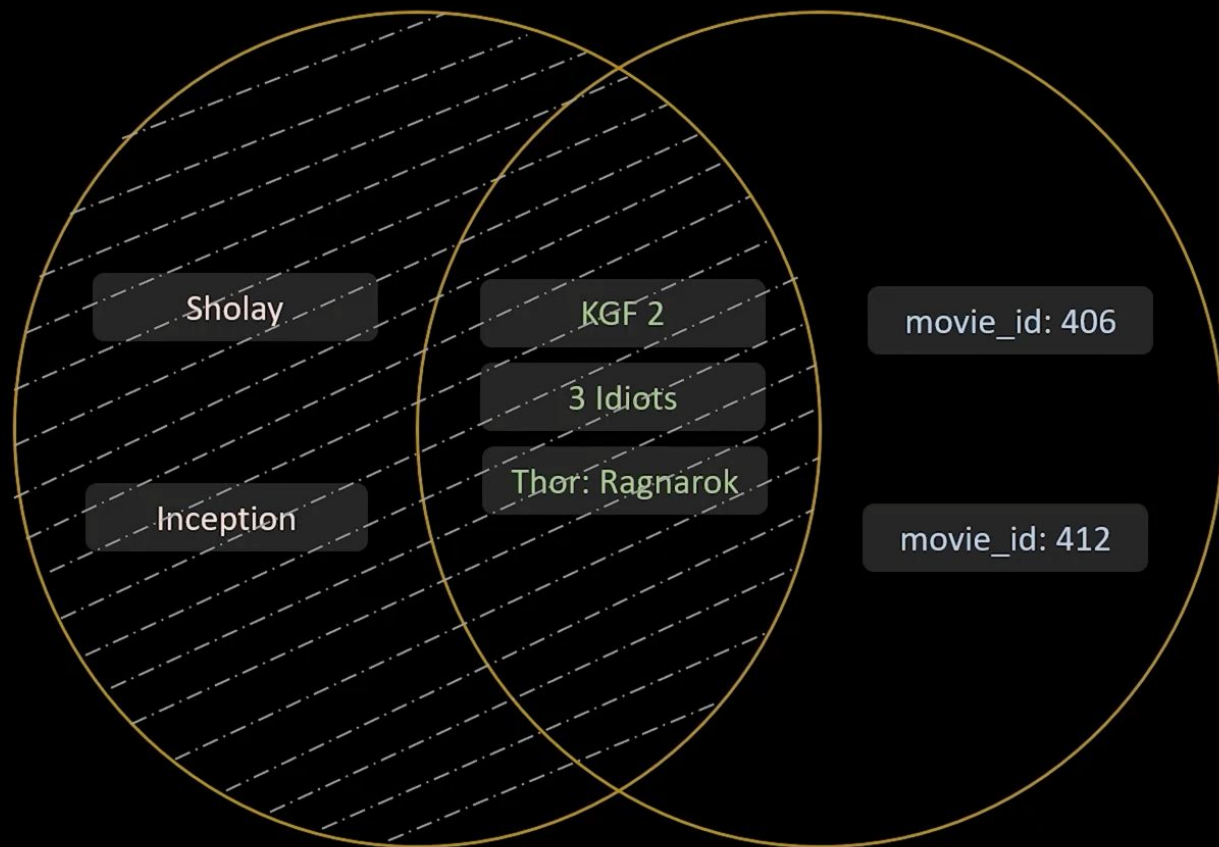
KGF 2

3 Idiots

Thor: Ragnarok

movies

financials



LEFT JOIN

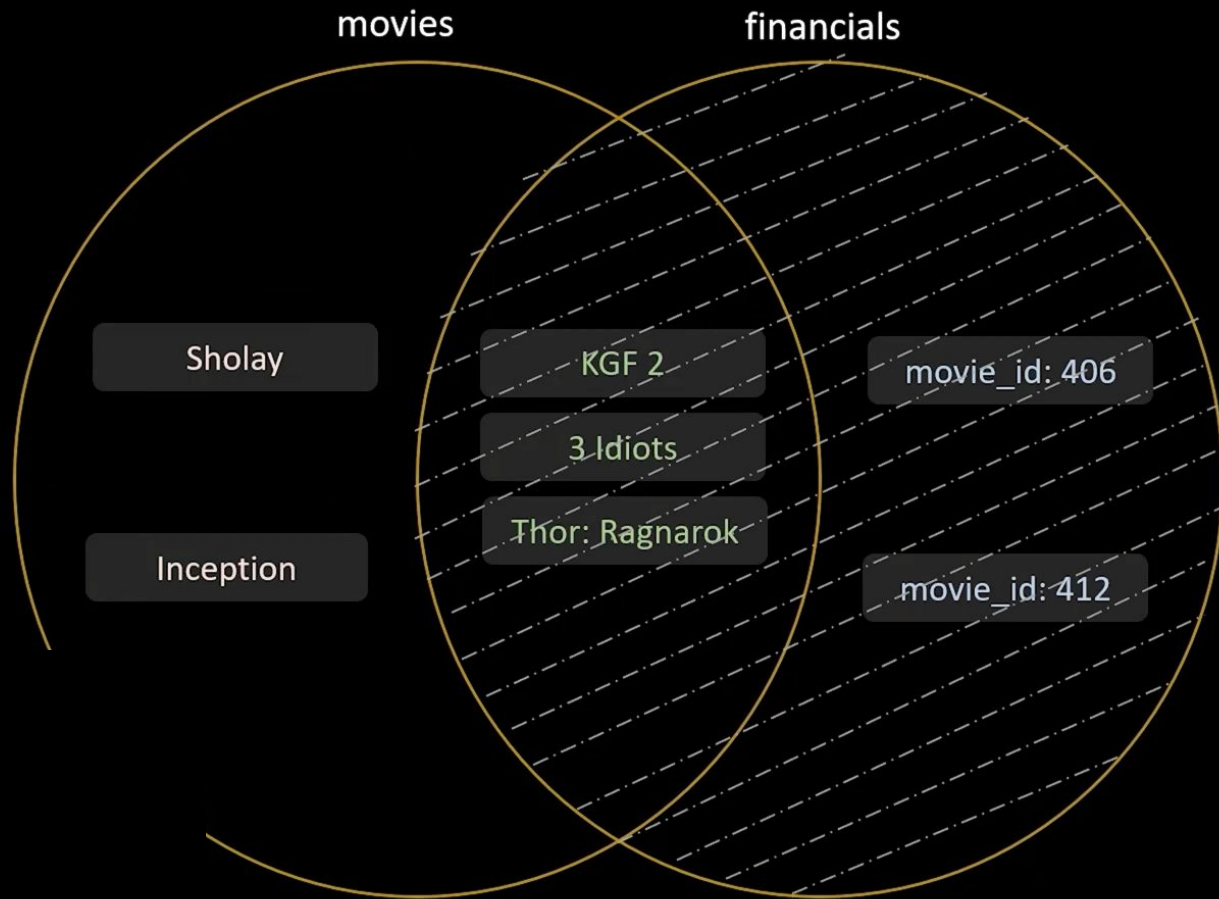
Sholay

Inception

KGF 2

3 Idiots

Thor: Ragnarok



RIGHT JOIN

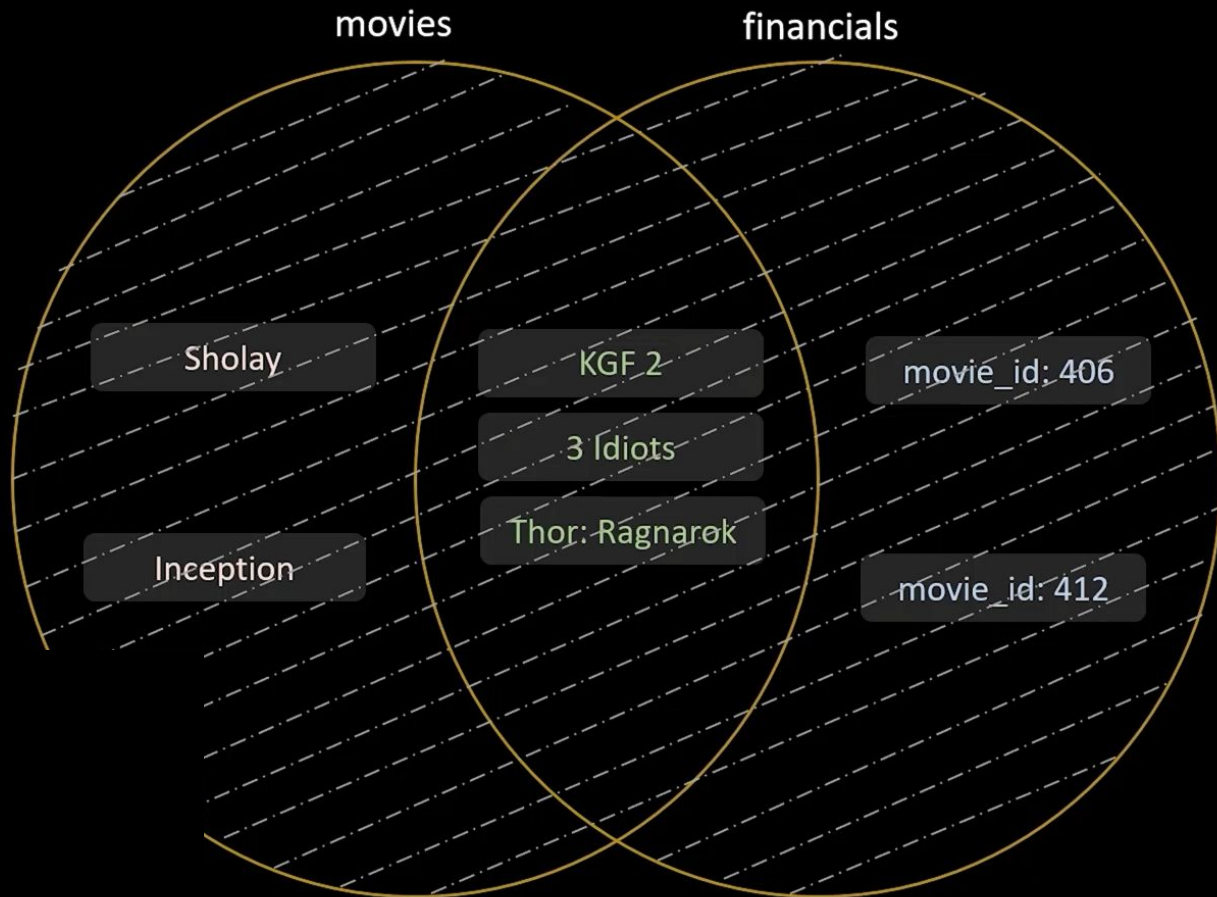
KGF 2

3 Idiots

Thor: Ragnarok

movie_id: 406

movie_id: 412



FULL JOIN

Sholay

Inception

KGF 2

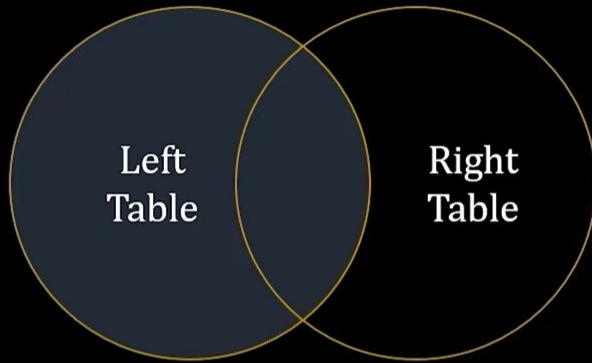
3 Idiots

Thor: Ragnarok

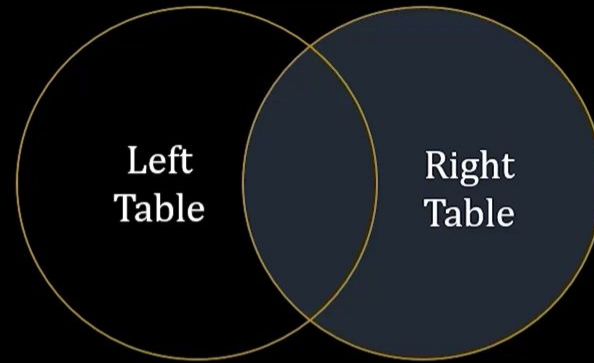
movie_id: 406

movie_id: 412

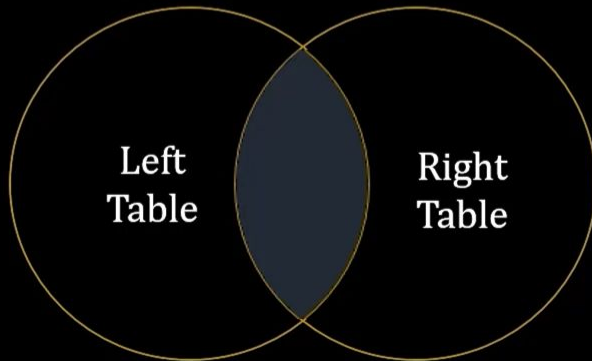
LEFT JOIN



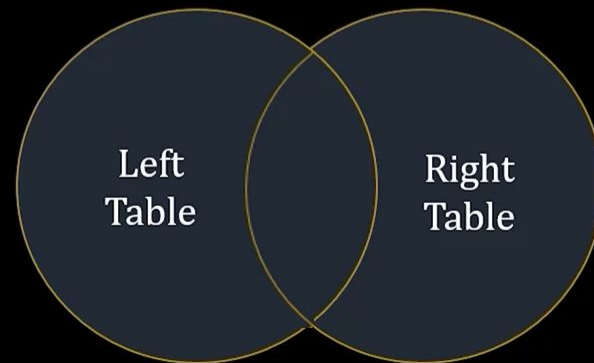
RIGHT JOIN



INNER JOIN



FULL JOIN



movies

movie_id	title	...
101	KGF2	...
102	Doctor Strange Multiverse	...
220	Bahubali 3	...
221	THOR 10	...

movie_actor

movie_id	actor_id
101	50
101	51
102	52
102	53

actors

actor_id	name	...
50	Yash	...
51	Sanjay Dutt	...
52	Benedict C	...
53	Elizabeth O	...

financials

movie_id	budget	revenue	...
101	1 billion INR	12.5 billion INR	...
102	200 million USD	954 million USD	...

languages

language_id	language
1	Hindi
2	Telugu
3	Kannada
5	English

Takeaways

- **Data cleaning** is an important step in the data analysis process because it ensures that the data is accurate and reliable.
- The **TRIM()** function eliminates extra spaces in the text, including those at the start and end.
- **Conditional Formatting** helps to identify and highlight duplicate data in a range of cells.
- The **"Text to Columns"** feature helps to split a single cell containing text into multiple cells based on a specified delimiter, such as a comma, semicolon, or space.

Takeaways

- **VLOOKUP** function is used to look up a value in a table by searching for a corresponding value in the leftmost column of another table.
- Syntax of the **VLOOKUP()** function is: **VLOOKUP (lookup_value, table_array, col_index_num, [range_lookup])**
- **Some limitations of VLookup:**
 - It searches only the first column of a reference table for a matching value.
 - Adding a new column to the reference table can cause errors by shifting column indexes and changing the referenced columns.

Takeaways

- **INDEX()** and **MATCH()** are two powerful Excel functions that are often used together to search for and retrieve data from a table or range.
- The syntax of the **INDEX()** function is: **INDEX(array, row_num, [column_num])**.
- The syntax of the **MATCH()** function is:
MATCH(lookup_value, lookup_array, [match_type]).

Takeaways

- The syntax of the **XLOOKUP()** function is: **XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode])**.
- Unlike **VLOOKUP**, which only searches in the leftmost column of a table, **XLOOKUP** is capable of searching in any column.