



PHÂN TÍCH DỮ LIỆU TRẢ LỜI CÂU HỎI

Nhóm biên soạn:

1. Lê Ngọc Thành
2. Nguyễn Ngọc Thảo
3. Phạm Trọng Nghĩa
4. Nguyễn Thái Vũ
5. Trương Tấn Khoa

Năm 2022



1 CHUẨN BỊ DỮ LIỆU ĐỂ PHÂN TÍCH



Phân tích

Phân tích là quá trình sử dụng để hiểu dữ liệu được thu thập.

Mục tiêu của phân tích là xác định các xu hướng và mối quan hệ trong dữ liệu từ đó bạn có thể trả lời chính xác câu hỏi của bạn

4 giai đoạn phân tích:

- Tổ chức dữ liệu
- Định dạng và điều chỉnh dữ liệu
- Lấy thêm thông tin từ những người khác
- Chuyển đổi dữ liệu



Phân tích

- **Tổ chức dữ liệu:** xác định tập dữ liệu mà bạn có thể phân tích để đưa ra quyết định
- Định dạng và điều chỉnh dữ liệu
- Lấy thêm thông tin từ những người khác
- Chuyển đổi dữ liệu



Phân tích

- Tổ chức dữ liệu
- **Định dạng và điều chỉnh dữ liệu:** định dạng dữ liệu theo cách dễ đọc, điều chỉnh dữ liệu để dễ lọc và sắp xếp.
- Lấy thêm thông tin từ những người khác
- Chuyển đổi dữ liệu



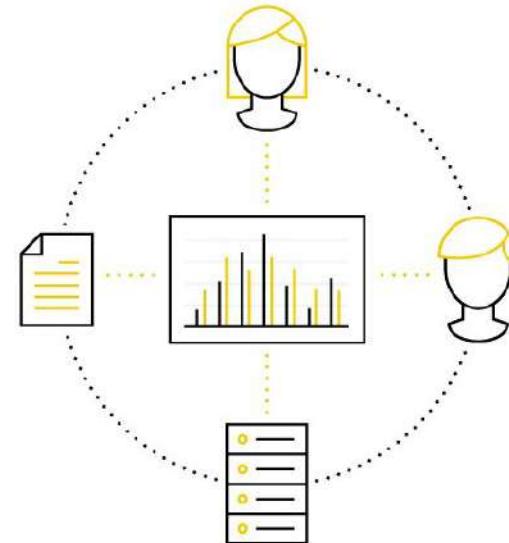
Phân tích

- Tổ chức dữ liệu
- Định dạng và điều chỉnh dữ liệu
- **Lấy thêm thông tin từ những người khác:**
lấy ý kiến từ những người khác cung cấp cho bạn những góc nhìn mới mà bạn chưa biết
- Chuyển đổi dữ liệu

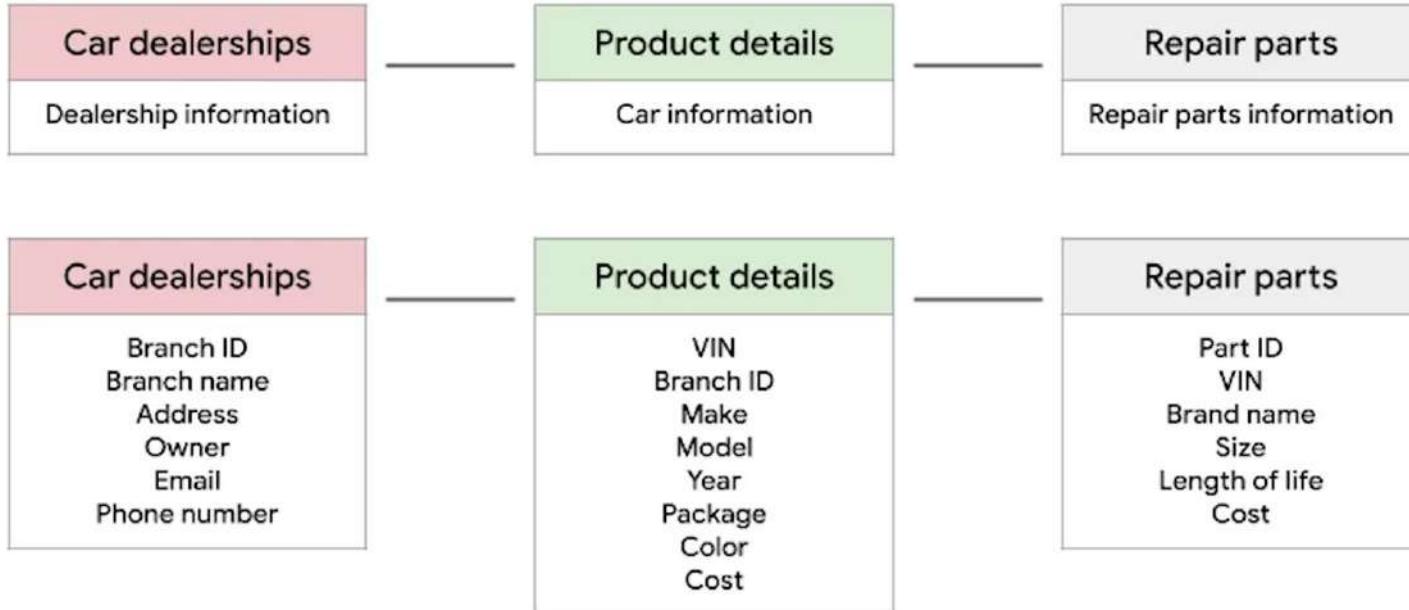


Phân tích

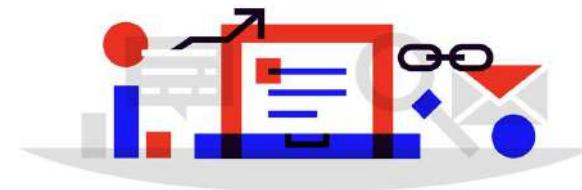
- Tổ chức dữ liệu
- Định dạng và điều chỉnh dữ liệu
- Lấy thêm thông tin từ những người khác
- **Chuyển đổi dữ liệu:** xác định các mối quan hệ và mẫu giữa dữ liệu và thực hiện các phép tính dựa trên dữ liệu bạn có.



Tổ chức dữ liệu



Tổ chức dữ liệu



Tổ chức dữ liệu

- Làm cho mọi thứ trông có trật tự
- Tìm kiếm và định vị dữ liệu nhanh chóng và dễ dàng

Cách tổ chức dữ liệu

- Sắp xếp
- Lọc

Sắp xếp

Sắp xếp (sorting): tổ chức dữ liệu thành một thứ tự có ý nghĩa để dễ hiểu, phân tích và hình dung hơn

Ví dụ: sắp xếp các mặt hàng khi bạn đang mua sắm trực tuyến:

- Giá thấp nhất đến cao nhất
- Theo thứ tự bảng chữ cái
- Mới nhất đến cũ nhất
- Gần nhất đến xa nhất



Lọc

- **Lọc (filtering):** chỉ hiển thị dữ liệu đáp ứng một tiêu chí cụ thể trong khi ẩn phần dữ liệu còn lại.
- Lọc để thu hẹp lượng dữ liệu bạn cần xem xét
- Lọc bằng SQL: mệnh đề WHERE

```
SELECT
*
FROM
    movie_data.movies
WHERE
    Genre = 'Comedy'
```

Sắp xếp dữ liệu trong bảng tính

Sắp xếp trong bảng tính:

- Sắp xếp theo thứ tự tăng dần hoặc giảm dần
- Sử dụng số hoặc chữ

Sort sheet: sắp xếp theo điều kiện của một cột, nhưng dữ liệu ở các cột liên quan cũng được sắp xếp theo

Sort range: không giữ thông tin giữa các hàng. Khi bạn dùng sort range, bạn giới hạn sự sắp xếp của mình trong một tập hợp ô được xác định

2 phương pháp sort: dùng **menu** và dùng **hàm sort**



Sort menu

Bộ dữ liệu movie

	A	B	C	D	E	F	G	H	I	J	K	
1	Movie Title	Release Date	Wikipedia URL	Genre (1)	Genre (2)	Director (1)	Director (2)	Cast (1)	Cast (2)	Cast (3)	Cast (4)	Cast (5)
2	10 Cloverfield Lane	2016-03-08	https://en.wikipedia.org	Thriller	Horror	Dan Trachtenberg		Mary Elizabeth Winstead	John Goodman	John Gallagher		
3	13 Hours: The Secret Soldiers of Benghazi	2016-01-15	https://en.wikipedia.org	Action	Thriller	Michael Bay		James Badge Dale	John Krasinski	Toby Stephens	Pablo Schreiber	Max Martini
4	2 Guns	2013-08-02	https://en.wikipedia.org	Action	Crime	Baltasar Kormákur		Mark Wahlberg	Denzel Washington	Paula Patton	Bill Paxton	Edward Norton
5	21 Jump Street	2012-03-16	https://en.wikipedia.org	Comedy	Action	Phil Lord	Chris Miller	Jonah Hill	Channing Tatum	Ice Cube	Brie Larson	Rob Riggle
6	22 Jump Street	2014-08-04	https://en.wikipedia.org	Action	Comedy	Phil Lord	Chris Miller	Channing Tatum	Jonah Hill	Ice Cube		
7	300: Rise of an Empire	2014-03-04	https://en.wikipedia.org	Action	Fantasy	Noam Murro		Rodrigo Santoro	Eva Green	Sullivan Stapleton	Hans Matheson	Lena Headey
8	42	2013-04-12	https://en.wikipedia.org	Biography	Drama	Brian Helgeland		Harrison Ford	Chadwick Boseman	Christopher Meloni	Ryan Merriman	Andre Holland
9	71	2014-02-07	https://en.wikipedia.org	Action	Drama	Yann Demange		Jack O'Connell	Richard Dormer	Sean Harris	Sam Reid	
10	90 Minutes in Heaven	2015-09-11	https://en.wikipedia.org	Drama		Michael Polish		Hayden Christensen	Kate Bosworth	Dwight Yoakam	Michael W. Smith	
11	A Good Day to Die Hard	2013-02-14	https://en.wikipedia.org	Action	Thriller	John Moore		Bruce Willis	Jai Courtney	Sebastian Koch	Yuliya Snigir	Radivoje Lukic
12	A Haunted House	2013-01-11	https://en.wikipedia.org	Comedy	Horror	Michael Tiddes		Marlon Wayans	Nick Swardson	David Koehner	Essence Atkins	Cedric the Entertainer
13	A Haunted House 2	2014-04-18	https://en.wikipedia.org	Comedy		Michael Tiddes		Marlon Wayans	Jaime Pressly	Essence Atkins	Cedric the Entertainer	Gabriel Iglesias
14	A Long Way Down	2014-02-10	https://en.wikipedia.org	Comedy		Pascal Chaumeil		Toni Collette	Pierce Brosnan	Aaron Paul	Imogen Poots	
15	A Most Violent Year	2014-11-06	https://en.wikipedia.org	Action	Crime	J. C. Chandor		Oscar Isaac	Jessica Chastain	Alessandro Nivola	David Oyelowo	Albert Elías
16	A Most Wanted Man	2014-01-19	https://en.wikipedia.org	Thriller		Anton Corbijn		Philip Seymour Hoffman	Rachel McAdams	Robin Wright	Willem Dafoe	Grigoriy Dobrygin
17	A Thousand Words	2012-03-09	https://en.wikipedia.org	Comedy		Brian Robbins		Eddie Murphy	Kerry Washington	Cliff Curtis	Clark Duke	Allison Janney
18	A Walk Among the Tombstones	2014-09-19	https://en.wikipedia.org	Crime	Drama	Scott Frank		Liam Neeson	Dan Stevens	Ruth Wilson	Boyd Holbrook	
19	A Walk in the Woods	2015-01-23	https://en.wikipedia.org	Adventure	Comedy	Ken Kwapis		Robert Redford	Nick Nolte	Emma Thompson		
20	Abraham Lincoln: Vampire Hunter	2012-06-22	https://en.wikipedia.org	Horror	Action	Timur Bekmambetov		Benjamin Walker	Dominic Cooper	Anthony Mackie	Mary Elizabeth Winstead	Rufus Wainwright
21	After Earth	2013-05-31	https://en.wikipedia.org	Sci-Fi		M. Night Shyamalan		Jaden Smith	Will Smith	Isabelle Fuhrman	Zoe Kravitz	
22	Ain't Them Bodies Saints	2013-08-16	https://en.wikipedia.org	Drama		David Lowery		Casey Affleck	Rooney Mara	Ben Foster	Rami Malek	Keith C. Williams

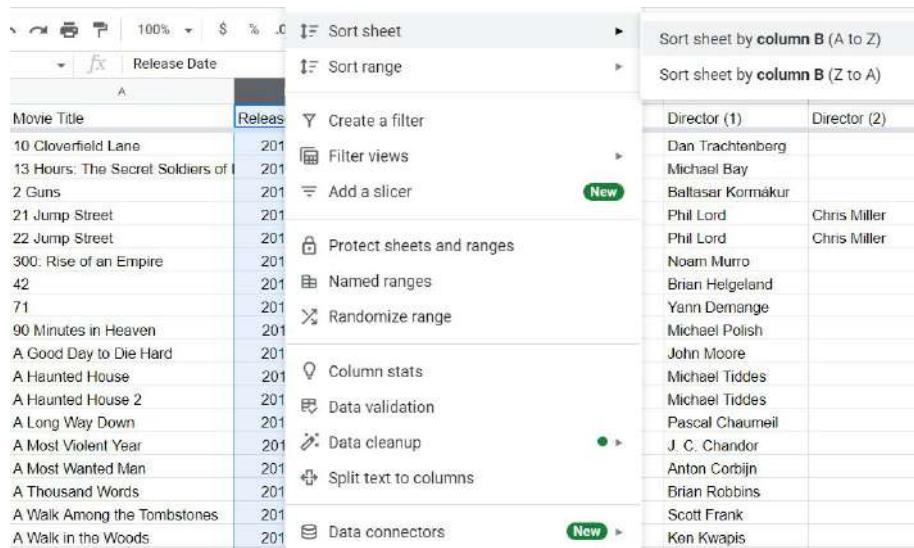
Sort menu

- Sắp xếp các bộ phim theo ‘Realese date’
- B1: Chọn cột ‘Realese date’

	Movie Title	Release Date	Wikipedia URL
1	10 Cloverfield Lane	2016-03-08	https://en.wikipedia.org/wiki/10_Cloverfield_Lane
2	13 Hours: The Secret Soldiers of Benghazi	2016-01-15	https://en.wikipedia.org/wiki/13_Hours:_The_Secret_Soldiers_of_Benghazi
3	2 Guns	2013-08-02	https://en.wikipedia.org/wiki/2_Guns
4	21 Jump Street	2012-03-16	https://en.wikipedia.org/wiki/21_Jump_Street
5	22 Jump Street	2014-06-04	https://en.wikipedia.org/wiki/22_Jump_Street
6	300: Rise of an Empire	2014-03-04	https://en.wikipedia.org/wiki/300:_Rise_of_an_Empire
7	42	2013-04-12	https://en.wikipedia.org/wiki/42_(2013_film)
8	71	2014-02-07	https://en.wikipedia.org/wiki/71
9	90 Minutes in Heaven	2015-09-11	https://en.wikipedia.org/wiki/90_Minutes_in_Heaven
10	A Good Day to Die Hard	2013-02-14	https://en.wikipedia.org/wiki/A_Good_Day_to_Die_Hard
11	A Haunted House	2013-01-11	https://en.wikipedia.org/wiki/A_Haunted_House
12	A Haunted House 2	2014-04-18	https://en.wikipedia.org/wiki/A_Haunted_House_2
13	A Long Way Down	2014-02-10	https://en.wikipedia.org/wiki/A_Long_Way_Down
14	A Most Violent Year	2014-11-06	https://en.wikipedia.org/wiki/A_Most_Violent_Year
15	A Most Wanted Man	2014-01-19	https://en.wikipedia.org/wiki/A_Most_Wanted_Man
16	A Thousand Words	2012-03-09	https://en.wikipedia.org/wiki/A_Thousand_Words
17	A Walk Among the Tombstones	2014-09-19	https://en.wikipedia.org/wiki/A_Walk_Among_the_Tombstones
18	A Walk in the Woods	2015-01-23	https://en.wikipedia.org/wiki/A_Walk_in_the_Woods
19	Abraham Lincoln: Vampire Hunter	2012-06-22	https://en.wikipedia.org/wiki/Abraham_Lincoln:_Vampire_Hunter
20	After Earth	2013-05-31	https://en.wikipedia.org/wiki/After_Earth
21	Ain't Them Bodies Saints	2013-08-16	https://en.wikipedia.org/wiki/Ain%27t_Them_Bodies_Saints
22	Alex Cross	2012-10-19	https://en.wikipedia.org/wiki/Alex_Cross
23	Aloha	2015-05-27	https://en.wikipedia.org/wiki/Aloha_(2015_film)
24	Alvin and the Chipmunks: The Road Chip	2015-12-18	https://en.wikipedia.org/wiki/Alvin_and_the_Chipmunks:_The_Road_Chip
25	Alvin and the Chipmunks: The Road Chip	2015-12-18	https://en.wikipedia.org/wiki/Alvin_and_the_Chipmunks:_The_Road_Chip

Sort menu

- B2: Chọn menu Data -> Sort sheet -> (A to Z) hay (Z to A)



The screenshot shows a Google Sheets interface with a table of movie titles and release years. A context menu is open over the table, specifically the 'Sort sheet' option under the 'Data' menu. The menu includes options like 'Sort range', 'Create a filter', 'Filter views', 'Add a slicer', 'Protect sheets and ranges', 'Named ranges', 'Randomize range', 'Column stats', 'Data validation', 'Data cleanup', 'Split text to columns', and 'Data connectors'. To the right of the menu, there are two columns of names: 'Director (1)' and 'Director (2)', with names like Dan Trachtenberg, Michael Bay, Baltasar Kormákur, etc., listed.

Movie Title	Release Date
10 Cloverfield Lane	2016
13 Hours: The Secret Soldiers of Benghazi	2016
2 Guns	2016
21 Jump Street	2013
22 Jump Street	2014
300: Rise of an Empire	2014
42	2013
71	2013
90 Minutes in Heaven	2013
A Good Day to Die Hard	2013
A Haunted House	2013
A Haunted House 2	2013
A Long Way Down	2013
A Most Violent Year	2014
A Most Wanted Man	2014
A Thousand Words	2013
A Walk Among the Tombstones	2014
A Walk in the Woods	2015

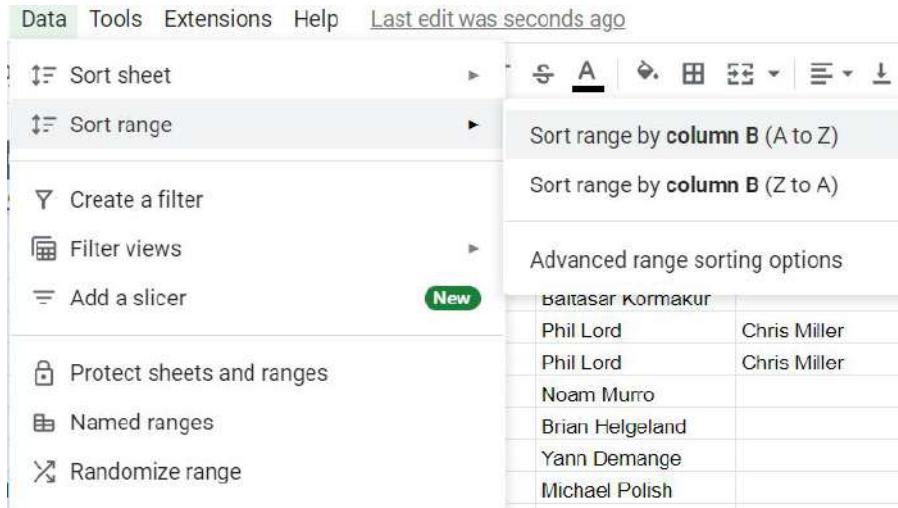
Director (1)	Director (2)
Dan Trachtenberg	
Michael Bay	
Baltasar Kormákur	
Phil Lord	Chris Miller
Phil Lord	Chris Miller
Noam Murro	
Brian Helgeland	
Yann Demange	
Michael Polish	
John Moore	
Michael Tiddes	
Michael Tiddes	
Pascal Chaumeil	
J. C. Chandor	
Anton Corbijn	
Brian Robbins	
Scott Frank	
Ken Kwapis	

Sort menu

Kết quả

	Movie Title	Release Date	Wikipedia URL	Genre (1)
2	The Devil Inside	2012-01-06	https://en.wikipedia.org	Horror
3	Red Tails	2012-01-20	https://en.wikipedia.org	Drama
4	Declaration of War	2012-01-27	https://en.wikipedia.org	Romance
5	Man on a Ledge	2012-01-27	https://en.wikipedia.org	Action
6	One for the Money	2012-01-27	https://en.wikipedia.org	Crime
7	The Grey	2012-01-27	https://en.wikipedia.org	Action
8	Chronicle	2012-02-03	https://en.wikipedia.org	Sci-Fi
9	The Woman in Black	2012-02-03	https://en.wikipedia.org	Horror
10	Journey 2: The Mysterious Island	2012-02-10	https://en.wikipedia.org	Adventure
11	Safe House	2012-02-10	https://en.wikipedia.org	Action
12	Ghost Rider: Spirit of Vengeance	2012-02-17	https://en.wikipedia.org	Action
13	The Secret World of Arrietty	2012-02-17	https://en.wikipedia.org	Adventure
14	This Means War	2012-02-17	https://en.wikipedia.org	Action
15	Good Deeds	2012-02-24	https://en.wikipedia.org	Romance
16	Wanderlust	2012-02-24	https://en.wikipedia.org	Comedy
17	Dr. Seuss' The Lorax	2012-03-02	https://en.wikipedia.org	Family
18	Project X	2012-03-02	https://en.wikipedia.org	Comedy
19	A Thousand Words	2012-03-09	https://en.wikipedia.org	Comedy
20	Friends with Kids	2012-03-09	https://en.wikipedia.org	Comedy
21	Salmon Fishing in the Yemen	2012-03-09	https://en.wikipedia.org	Drama
22	21 Jump Street	2012-03-16	https://en.wikipedia.org	Comedy

Sort menu



	Movie Title	Release Date	Wikipedia URL
1	10 Cloverfield Lane	2012-01-06	https://en.wikipedia.org/w/index.php?title=10_Cloverfield_Lane&oldid=5880111
2	13 Hours: The Secret Soldiers of Benghazi	2012-01-20	https://en.wikipedia.org/w/index.php?title=13_Hours:_The_Secret_Soldiers_of_Benghazi&oldid=5880111
3	2 Guns	2012-01-27	https://en.wikipedia.org/w/index.php?title=2_Guns_(2012_film)&oldid=5880111
4	21 Jump Street	2012-01-27	https://en.wikipedia.org/w/index.php?title=21_Jump_Street&oldid=5880111
5	22 Jump Street	2012-01-27	https://en.wikipedia.org/w/index.php?title=22_Jump_Street&oldid=5880111
6	300: Rise of an Empire	2012-01-27	https://en.wikipedia.org/w/index.php?title=300:_Rise_of_an_Empire&oldid=5880111
7	42	2012-02-03	https://en.wikipedia.org/w/index.php?title=42_(2012_film)&oldid=5880111
8	71	2012-02-03	https://en.wikipedia.org/w/index.php?title=71_(2012_film)&oldid=5880111
9	90 Minutes in Heaven	2012-02-10	https://en.wikipedia.org/w/index.php?title=90_Minutes_in_Heaven&oldid=5880111
10	A Good Day to Die Hard	2012-02-10	https://en.wikipedia.org/w/index.php?title=A_Good_Day_to_Die_Hard&oldid=5880111
11	A Haunted House	2012-02-17	https://en.wikipedia.org/w/index.php?title=A_Haunted_House_(2012_film)&oldid=5880111
12	A Haunted House 2	2012-02-17	https://en.wikipedia.org/w/index.php?title=A_Haunted_House_2_(2013_film)&oldid=5880111
13	A Long Way Down	2012-02-17	https://en.wikipedia.org/w/index.php?title=A_Long_Way_Down_(2014_film)&oldid=5880111
14	A Most Violent Year	2012-02-24	https://en.wikipedia.org/w/index.php?title=A_Most_Violent_Year_(2014_film)&oldid=5880111
15	A Most Wanted Man	2012-02-24	https://en.wikipedia.org/w/index.php?title=A_Most_Wanted_Man_(2014_film)&oldid=5880111
16	A Thousand Words	2012-03-02	https://en.wikipedia.org/w/index.php?title=A_Thousand_Words_(2014_film)&oldid=5880111
17	A Walk Among the Tombstones	2012-03-02	https://en.wikipedia.org/w/index.php?title=A_Walk_Among_the_Tombstones_(2014_film)&oldid=5880111
18	A Walk in the Woods	2012-03-09	https://en.wikipedia.org/w/index.php?title=A_Walk_in_the_Woods_(2015_film)&oldid=5880111
19	Abraham Lincoln: Vampire Hunter	2012-03-09	https://en.wikipedia.org/w/index.php?title=Abraham_Lincoln:_Vampire_Hunter&oldid=5880111
20	After Earth	2012-03-09	https://en.wikipedia.org/w/index.php?title=After_Earth_(2013_film)&oldid=5880111
21	Ain't Them Bodies Saints	2012-03-16	https://en.wikipedia.org/w/index.php?title=Ain%27t_Them_Bodies_Saints_(2013_film)&oldid=5880111

Hàm sort

Bộ dữ liệu [Party Plan](#)

	A	B	C	D	E
1	Guest Name	Table Number	Dietary Restriction	Sent Invitation	Row Index
2	Jack	2	Vegetarian	No	5
3	Omar	1	None	Yes	4
4	Aida	3	None	Yes	3
5	Nancy	3	None	No	2
6	Jianyu	2	Vegetarian	Yes	1
7					
8					
9	Guest Name	Table Number	Dietary Restriction	Sent Invitation	Formula
10	Omar	1	None	Yes	=SORT(A2:D6, 2, TRUE)
11	Jack	2	Vegetarian	No	
12	Jianyu	2	Vegetarian	Yes	
13	Aida	3	None	Yes	
14	Nancy	3	None	No	
15					

Hàm sort

Muốn sắp xếp tăng dần theo 'Table Number'

Guest Name	Table Number	Dietary Restriction	Sent Invitation	Row Index
Jack	2	Vegetarian	No	5
Omar	1	None	Yes	4
Aida	3	None	Yes	3
Nancy	3	None	No	2
Jianyu	2	Vegetarian	Yes	1

```
=SORT(A2:D6, 2, TRUE)
```

```
SORT(range, sort_column, [isAscending],  
[sort_column2, ...], [isAscending2, ...])
```

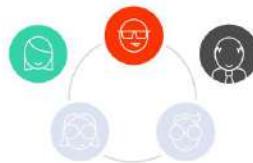
Các tham số:

- Range: các ô cần sắp xếp
- Sort_column: muốn sắp xếp theo điều kiện cột nào.
- isAscending: TRUE nếu xếp tăng dần. FALSE cho giảm dần

Hàm sort

Kết quả

Guest Name	Table Number	Dietary Restriction	Sent Invitation	Row Index				
Jack	2	Vegetarian	No	5	Omar	1	None	Yes
Omar	1	None	Yes	4	Jack	2	Vegetarian	No
Aida	3	None	Yes	3	Jianyu	2	Vegetarian	Yes
Nancy	3	None	No	2	Aida	3	None	Yes
Jianyu	2	Vegetarian	Yes	1	Nancy	3	None	No



Sắp xếp trên nhiều điều kiện

- Sắp xếp theo ‘Sent Invitation’
- Nếu trùng, xếp theo Guess Name.
- B1: chọn các ô cần sắp xếp

	A	B	C	D
1	Guest Name	Table Number	Dietary Restriction	Sent Invitation
2	Jack	2	Vegetarian	No
3	Omar	1	None	Yes
4	Aida	3	None	Yes
5	Nancy	3	None	No
6	Jianyu	2	Vegetarian	Yes

Sắp xếp trên nhiều điều kiện

B2: Chọn Menu Data -> Sort range -> Advanced range sorting option

The screenshot shows a Google Sheets spreadsheet with a table of guest names and table numbers. The 'Data' menu is open, with 'Sort range' selected. A sub-menu for 'Advanced range sorting options' is displayed, containing two items: 'Sort range by column A (A to Z)' and 'Sort range by column A (Z to A)'. A green 'New' button is located at the bottom right of this sub-menu.

Guest Name	Table Number	R
Jack	2	V
Omar	1	V
Aida	3	V
Nancy	3	V
Jianyu	2	V

Sắp xếp trên nhiều điều kiện

B3: Chọn các option như trên ảnh. Rồi nhấn sort

Sort range from A1 to D6 ×

Data has header row

Sort by

Sent Invitation ▼

A → Z

Z → A

then by

Guest Name ▼

A → Z

Z → A



Add another sort column

Cancel

Sort

Sắp xếp trên nhiều điều kiện

Kết quả

Guest Name	Table Number	Dietary Restriction	Sent Invitation	Row Index
Jack	2	Vegetarian	No	5
Nancy	3	None	No	4
Aida	3	None	Yes	3
Jianyu	2	Vegetarian	Yes	2
Omar	1	None	Yes	1



Câu truy vấn sắp xếp trong SQL

ORDER BY: câu lệnh để sắp xếp kết quả

The screenshot shows the BigQuery UI. At the top, there's a header with 'Explorer' and a '+ ADD DATA' button. Below the header is a search bar with a magnifying glass icon and the placeholder 'Type to search'. To the right of the search bar is a help icon (a question mark inside a circle). The main area is titled 'Viewing pinned projects.' and lists a single project: 'testbigquery-353006'. Underneath this project, there's a table named 'movie_data' with a 'movies' sub-table selected. The 'movies' table has a blue background. To the right of each table name are three vertical dots, likely for more options.

```
SELECT *
FROM `movie_data.movies`
ORDER BY Release_Date
```

Câu truy vấn sắp xếp trong SQL

Kết quả truy vấn

Query results

 SAV

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Movie_Title		Release_Date	Wikipedia_URL
1	The Devil Inside		2012-01-06	https://en.wikipedia.org/wiki/The_Devil_Inside_(film)
2	Red Tails		2012-01-20	https://en.wikipedia.org/wiki/Red_Tails
3	Man on a Ledge		2012-01-27	https://en.wikipedia.org/wiki/Man_on_a_Ledge
4	The Grey		2012-01-27	https://en.wikipedia.org/wiki/The_Grey_(film)
5	Declaration of War		2012-01-27	https://en.wikipedia.org/wiki/Declaration_of_War_(film)
6	One for the Money		2012-01-27	https://en.wikipedia.org/wiki/One_for_the_Money_(film)
7	Chronicle		2012-02-03	https://en.wikipedia.org/wiki/Chronicle_(film)

Câu truy vấn sắp xếp trong SQL

SELECT *
FROM `movie_data.movies`
ORDER BY Release_Date DESC

Query results SAVE RESULTS

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Movie_Title		Release_Date	Wikipedia_URL
1	Mechanic: Resurrection		2016-08-26	https://en.wikipedia.org/wiki/Mechanic:_Resurrection
2	Don't Breathe		2016-08-26	https://en.wikipedia.org/wiki/Don%2527t_Breathe_(2016_film)
3	Hands of Stone		2016-08-26	https://en.wikipedia.org/wiki/Hands_of_Stone
4	Kubo and the Two Strings		2016-08-19	https://en.wikipedia.org/wiki/Kubo_and_the_Two_Strings
5	Ben-Hur		2016-08-19	https://en.wikipedia.org/wiki/Ben-Hur_(2016_film)
6	War Dogs		2016-08-19	https://en.wikipedia.org/wiki/War_Dogs_(2016_film)
7	Suicide Squad		2016-08-05	https://en.wikipedia.org/wiki/Suicide_Squad_(film)
8	Nine Lives		2016-08-05	https://en.wikipedia.org/wiki/Nine_Lives_(2016_film)

Lọc với mệnh đề WHERE

Yêu cầu:

- Chỉ quan tâm đến phim hài (Comedy)
- Vẫn sắp xếp giảm dần theo ngày phát hành

Thêm mệnh đề WHERE vào câu truy vấn.

```
SELECT *
FROM `movie_data.movies`
WHERE Genre_1_ = 'Comedy'
ORDER BY Release_Date DESC
```

Lọc với mệnh đề WHERE

Kết quả

Query results

 SAVE RESULTS ▾

 EXPLORE DATA ▾

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		
Row	Movie_Title		Release_Date	Wikipedia_URL	Genre_1_	Genre_2_
1	Nine Lives		2016-08-05	https://en.wikipedia.org/wiki/Nine_Lives_(2016_film)	Comedy	null
2	Bad Moms		2016-07-29	https://en.wikipedia.org/wiki/Bad_Moms	Comedy	null
3	Ghostbusters		2016-07-15	https://en.wikipedia.org/wiki/Ghostbusters_(2016_film)	Comedy	Sci-Fi
4	Popstar: Never Stop Never Stopping		2016-06-03	https://en.wikipedia.org/wiki/Popstar:_Never_Stop_Never_Stopping	Comedy	null
5	Neighbors 2: Sorority Rising		2016-05-20	https://en.wikipedia.org/wiki/Neighbors_2:_Sorority_Rising	Comedy	null
6	The Angry Birds Movie		2016-05-05	https://en.wikipedia.org/wiki/The_Angry_Birds_Movie	Comedy	null
7	Keanu		2016-04-29	https://en.wikipedia.org/wiki/Keanu_(film)	Comedy	null
8	Barbershop: The Next Cut		2016-04-15	https://en.wikipedia.org/wiki/Barbershop:_The_Next_Cut	Comedy	null

Lọc với mệnh đề WHERE

Yêu cầu:

- Chỉ quan tâm đến phim hài (Comedy)
- Vẫn sắp xếp giảm dần theo ngày phát hành
- Doanh thu trên 30M

Thêm mệnh đề AND vào câu truy vấn.

```
SELECT *
FROM `movie_data.movies`
WHERE Genre__1_ = 'Comedy'
AND Box_Office_Revenue____ > 30000000
ORDER BY Release_Date DESC
```

Lọc với mệnh đề WHERE

Kết quả

Query results

SAVE RESULTS ▾

EXPLORE DATA ▾

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		
Row	Cast_2_		Cast_3_	Cast_4_	Cast_5_	Budget__Box_Office_Revenue__
1	Christina Applegate		Kristen Bell	Jada Pinkett Smith	Kathryn Hahn	20000000.0 124200000.0
2	Kristen Wiig		Leslie Jones	Kate McKinnon	Chris Hemsworth	144000000.0 217500000.0
3	Zac Efron		Rose Byrne	Chloë Grace Moretz	Dave Franco	35000000.0 107900000.0
4	Josh Gad		Danny McBride	Bill Hader	Maya Rudolph	73000000.0 346900000.0
5	Cedric the Entertainer		Nicki Minaj	Regina Hall	Eve	20000000.0 55000000.0
6	Julia Roberts		Kate Hudson	Timothy Olyphant	Jason Sudeikis	25000000.0 43800000.0
-



2 ĐỊNH DẠNG VÀ ĐIỀU CHỈNH DỮ LIỆU



Định dạng dữ liệu

Định dạng dữ liệu: biến đổi dữ liệu từ dạng này sang dạng khác.

Ví dụ: 0.24 □ 24%, chuỗi “22-2-2022” sang dạng ngày tháng

Vấn đề:

- Sắp xếp các bộ phim theo giờ gian
- Nếu ngày phát hành được lưu ở dạng chuỗi. Không thể sắp xếp được.
- Phải chuyển sang dạng ngày tháng (Datetime)



Dữ liệu định dạng sai

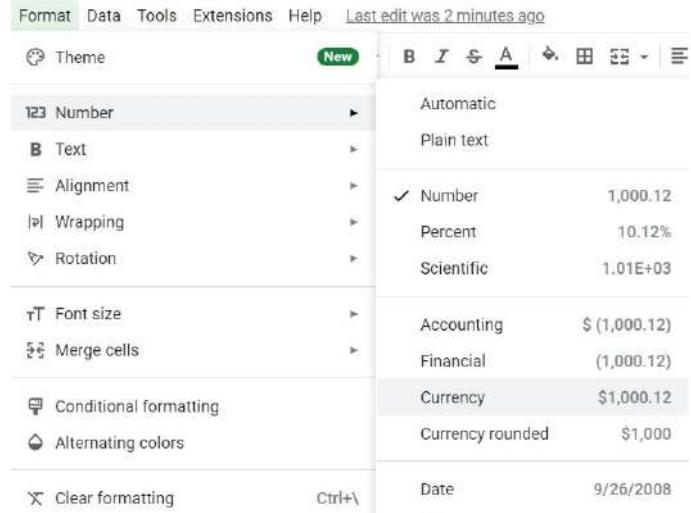
- Dẫn đến sai lầm
- Tốn thời gian để sửa
- Ảnh hưởng đến khả năng ra quyết định



Định dạng dữ liệu trong bảng tính

- Mở tập dữ liệu Movie trong Google Sheet
- Định dạng tiền tệ

M	N
Budget (\$)	Box Office Revenue (\$)
15,000,000.00	108,300,000.00
45,000,000.00	69,400,000.00
61,000,000.00	131,900,000.00
55,000,000.00	201,500,000.00
84,500,000.00	331,300,000.00
110,000,000.00	337,600,000.00
40,000,000.00	97,500,000.00
8,100,000.00	2,900,000.00
5,000,000.00	4,800,000.00
92,000,000.00	304,700,000.00
2,500,000.00	60,100,000.00
4,000,000.00	24,000,000.00
22,700,000.00	7,100,000.00



M	N
Budget (\$)	Box Office Revenue
\$15,000,000.00	\$108,300,000.00
\$45,000,000.00	\$69,400,000.00
\$61,000,000.00	\$131,900,000.00
\$55,000,000.00	\$201,500,000.00
\$84,500,000.00	\$331,300,000.00
\$110,000,000.00	\$337,600,000.00
\$40,000,000.00	\$97,500,000.00
\$8,100,000.00	\$2,900,000.00
\$5,000,000.00	\$4,800,000.00
\$92,000,000.00	\$304,700,000.00
\$2,500,000.00	\$60,100,000.00
\$4,000,000.00	\$24,000,000.00
\$22,700,000.00	\$7,100,000.00

Chuyển đổi dữ liệu

Tập dữ liệu thời tiết

Timestamp	Temperature	Precipitation Total	Wind Speed	Wind Direction [10 m]
2020-08-20T00:00:00	69.994736	0	8.5248	89.99999
2020-08-20T01:00:00	67.51074	0	8.573356	100.61965
2020-08-20T02:00:00	65.20673	0	8.08909	107.241455
2020-08-20T03:00:00	62.722736	0	8.014169	111.44774
2020-08-20T04:00:00	60.760735	0	7.5066204	117.474434
2020-08-20T05:00:00	58.096737	0	7.168091	131.98721
2020-08-20T06:00:00	59.842735	0	5.8668513	129.47246
2020-08-20T07:00:00	60.814735	0	4.9553266	143.74615
2020-08-20T08:00:00	66.26874	0	5.1725526	168.11134
2020-08-20T09:00:00	72.83874	0	6.6600003	196.26021
2020-08-20T10:00:00	77.78874	0	7.2713056	208.44293
2020-08-20T11:00:00	82.59474	0	8.726379	211.26373
2020-08-20T12:00:00	85.45673	0	11.973191	212.27565
2020-08-20T13:00:00	87.05874	0	15.81438	212.61925
2020-08-20T14:00:00	88.13873	0	16.551136	213.17851
2020-08-20T15:00:00	88.94873	0	16.357048	217.05653
2020-08-20T16:00:00	89.36273	0	15.823352	225
2020-08-20T17:00:00	89.32674	0	15.611131	233.32565
2020-08-20T18:00:00	88.69673	0	15.861428	240.85193
2020-08-20T19:00:00	87.22073	0	15.34057	249.67686

Hàm CONVERT

Hàm CONVERT: chuyển dữ liệu từ đơn vị đo lường này sang đơn vị khác

Timestamp	Temperature	Precipitation Total	Wind Speed	Wind Direction [10 m]	
2020-08-20T00:00:00	69.994736	0	8.5248	89.99999	$21.10818667 \times$ <code>=CONVERT(B2, "F", "C")</code>
2020-08-20T01:00:00	67.51074	0	8.673356	100.61965	<code>CONVERT(value, start_unit, end_unit)</code>
2020-08-20T02:00:00	65.20673	0	8.08909	107.241455	
2020-08-20T03:00:00	62.722736	0	8.014169	111.44774	
2020-08-20T04:00:00	60.760735	0	7.5066204	117.474434	

Tham số:

- Value: giá trị cần đổi.
- Start_unit: Đơn vị đo lường gốc
- End_unit: Đơn vị đo lường đích

Xác thực dữ liệu

Xác thực dữ liệu (data validation): cho phép kiểm soát những gì có thể và không thể nhập vào trang tính.

Các kỹ thuật xác thực dữ liệu:

- Thêm danh sách thả xuống (Drop-down list)
- Tạo checkbox
- Bảo vệ dữ liệu có cấu trúc và công thức



Drop-down list

Bộ dữ liệu: Project spreadsheet

Giá trị cho cột Status:

- "Not Yet Started"
- "In Progress"
- "Ready"

	A	B	C	D
1	Task	Asignee	Status	Review By This Date
2	Review C5M3 plan	Tony		2020-03-13
3	Map out next week's goals	Ximena		2020-03-19
4	Review weekly invoices	Sally		2020-04-02
5	Send out stakeholder emails	Erica		2020-04-10
6	Follow up on customer emails	Ayanna		2020-04-17
7	Prepare presentation for VP	Hallie		2020-04-24
8	Schedule next team meeting	Tony		2020-04-27
9	Map out next week's goals	Ximena		2020-05-18
10	Review weekly invoices	Sally		2020-05-25
11	Send out stakeholder emails	Erica		2020-04-24
12	Set next week's team agenda	Ayanna		2020-04-27
13	Vet all the datasets to be used	Hallie		2020-05-18
14	Schedule next team meeting	Tony		2020-05-25
15	Map out next week's goals	Ximena		2020-04-24
16	Review weekly invoices	Sally		2020-04-27
17	Schedule next team meeting	Erica		2020-05-18
18	Map out next week's goals	Ayanna		2020-05-25
19	Review weekly invoices	Hallie		2020-04-24
20	Schedule next team meeting	Tony		2020-04-27
21	Map out next week's goals	Ximena		2020-05-18
22	Review weekly invoices	Sally		2020-05-25

Drop-down list

- Chọn cột Status
- Chọn Data -> Data Validation

Project Spreadsheet - for Data Validation

File Edit View Insert Format Data Tools Extensions Help Last edit was 4 hours ago

100% Status

A B

Task	Asignee
Review C5M3 plan	Tony
Map out next week's goals	Ximena
Review weekly invoices	Sally
Send out stakeholder emails	Erica
Follow up on customer emails	Ayanna
Prepare presentation for VP	Hallie
Schedule next team meeting	Tony
Map out next week's goals	Ximena
Review weekly invoices	Sally
Send out stakeholder emails	Erica
Set next week's team agenda	Ayanna
Vet all the datasets to be used	Hallie
Schedule next team meeting	Tony
Map out next week's goals	Ximena
Review weekly invoices	Sally
Schedule next team meeting	Erica
Map out next week's goals	Ayanna
Review weekly invoices	Hallie
Schedule next team meeting	Tony

⋮ fx

Sort sheet

Sort range

Create a filter

Filter views

Add a slicer New

Protect sheets and ranges

Named ranges

Randomize range

Column stats

Data validation

Data cleanup

Split text to columns

Data connectors New

Drop-down list

Data validation

X

Cell range:

'Sheet 1'!C1:C100



Criteria:

List from a range

Enter a range or formula



Tip: Use absolute references (e.g. =\$A\$1:\$B\$1) to lock rows & columns.

Show dropdown list in cell

On invalid data: Show warning Reject input

Appearance: Show validation help text:

Cancel

Remove validation

Save

Drop-down list

Data validation

Cell range: 'Sheet 1'!C1:C100

Criteria: List of items ▾ Not Yet Started, In progress, Ready

Show dropdown list in cell

On invalid data: Show warning Reject input

Appearance: Show validation help text:

	A	B	C
1	Task	Asignee	Not Yet Started
2	Review C5M3 plan	Tony	In progress
3	Map out next week's goals	Ximena	Ready
4	Review weekly invoices	Sally	
5	Send out stakeholder emails	Erica	
6	Follow up on customer emails	Ayanna	
7	Prepare presentation for VP	Hallie	
8	Schedule next team meeting	Tony	
9	Map out next week's goals	Ximena	
10	Review weekly invoices	Sally	
11	Send out stakeholder emails	Erica	
12	Set next week's team agenda	Ayanna	
13	Vet all the datasets to be used	Hallie	
14	Schedule next team meeting	Tony	
15	Map out next week's goals	Ximena	
16	Review weekly invoices	Sally	
17	Schedule next team meeting	Erica	
18	Map out next week's goals	Ayanna	

Checkbox

Yêu cầu: tạo checkbox giúp chúng ta biết nếu nhiệm vụ đã được phê duyệt hay chưa

The screenshot shows a Google Sheets interface with a data validation dialog box overlaid. The spreadsheet has columns C, D, and E. Column C is labeled 'Status' and column D is labeled 'Review By This Date'. The data in column D includes dates like 2020-03-13, 2020-03-19, 2020-04-02, 2020-04-10, 2020-04-17, 2020-04-24, 2020-04-27, 2020-05-18, 2020-05-25, 2020-04-24, 2020-04-27, 2020-05-18, 2020-05-25, 2020-04-24, 2020-04-27, 2020-05-18, 2020-05-25, and 2020-04-24.

The Data validation dialog box contains the following settings:

- Cell range:** 'Sheet 1'!E1:E100
- Criteria:** List from a range (selected)
- Show dropdown list in cell** (checkbox checked)
- On invalid data:** Show warning (radio button selected)
- Appearance:** Show validation help text (checkbox unchecked)

At the bottom of the dialog box are three buttons: Cancel, Remove validation, and Save (green).

Checkbox

Data validation

X

Cell range:

'Sheet 1'!E1:E100

Criteria:

Checkbox

Use custom cell values

Checked: Approved

Unchecked: Not Approved

On invalid data: Show warning Reject input

Appearance: Show validation help text:

Checkbox

Kết quả

	A	B	C	D	E
1	Task	Asignee	Status	Review By This Date	<input type="checkbox"/>
2	Review C5M3 plan	Tony		2020-03-13	<input type="checkbox"/>
3	Map out next week's goals	Ximena		2020-03-19	<input type="checkbox"/>
4	Review weekly invoices	Sally		2020-04-02	<input checked="" type="checkbox"/>
5	Send out stakeholder emails	Erica		2020-04-10	<input type="checkbox"/>
6	Follow up on customer emails	Ayanna		2020-04-17	<input type="checkbox"/>
7	Prepare presentation for VP	Hallie		2020-04-24	<input type="checkbox"/>
8	Schedule next team meeting	Tony		2020-04-27	<input type="checkbox"/>
9	Map out next week's goals	Ximena		2020-05-18	<input type="checkbox"/>
10	Review weekly invoices	Sally		2020-05-25	<input type="checkbox"/>
11	Send out stakeholder emails	Erica		2020-04-24	<input checked="" type="checkbox"/>
12	Set next week's team agenda	Ayanna		2020-04-27	<input type="checkbox"/>
13	Vet all the datasets to be used	Hallie		2020-05-18	<input type="checkbox"/>
14	Schedule next team meeting	Tony		2020-05-25	<input type="checkbox"/>
15	Map out next week's goals	Ximena		2020-04-24	<input type="checkbox"/>
16	Review weekly invoices	Sally		2020-04-27	<input type="checkbox"/>
17	Schedule next team meeting	Erica		2020-05-18	<input type="checkbox"/>
18	Map out next week's goals	Ayanna		2020-05-25	<input type="checkbox"/>

Bảo vệ dữ liệu có cấu trúc và công thức

Data validation

Cell range: 'Sheet 1'!E1:E100

Criteria: Checkbox

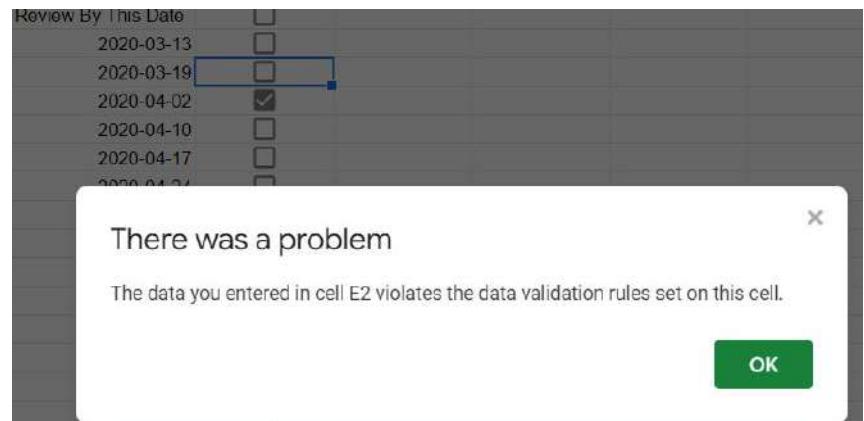
Use custom cell values

Checked: Approved

Unchecked: Not Approved

On invalid data: Show warning Reject input

Appearance: Show validation help text:



Định dạng có điều kiện

- **Định dạng có điều kiện (conditional formatting):** một công cụ bảng tính thay đổi giao diện của các ô khi các giá trị đáp ứng các điều kiện cụ thể.
- Bộ dữ liệu: Project Spreadsheet cho định dạng có điều kiện.

	A	B	C	D	E
1	Task	Asignee	Status	Review By This Date	Review
2	Review C5M3 plan	Tony	Not Yet Started ▾	2020-03-13	<input type="checkbox"/>
3	Map out next week's goals	Ximena	Not Yet Started ▾	2020-03-19	<input type="checkbox"/>
4	Review weekly invoices	Sally	Not Yet Started ▾	2020-04-02	<input type="checkbox"/>
5	Send out stakeholder emails	Erica	Not Yet Started ▾	2020-04-10	<input type="checkbox"/>
6	Follow up on customer emails	Ayanna	Not Yet Started ▾	2020-04-17	<input type="checkbox"/>
7	Prepare presentation for VP	Hallie	Not Yet Started ▾	2020-04-24	<input type="checkbox"/>
8	Schedule next team meeting	Tony	Not Yet Started ▾	2020-04-27	<input type="checkbox"/>
9	Map out next week's goals	Ximena	Not Yet Started ▾	2020-05-18	<input type="checkbox"/>
10	Review weekly invoices	Sally	Not Yet Started ▾	2020-05-25	<input type="checkbox"/>

Định dạng có điều kiện

Chọn cột, sau đó chọn Format -> Conditional Formatting

B	C	D
Asignee	Status	Review By This Date
Tony	Not Yet Started ▾	2020-03-13
Ximena	Not Yet Started ▾	2020-03-19
Sally	Not Yet Started ▾	2020-04-02
Erica	Not Yet Started ▾	2020-04-10
Ayanna	Not Yet Started ▾	2020-04-17
Hallie	Not Yet Started ▾	2020-04-24
Tony	Not Yet Started ▾	2020-04-27
Ximena	Not Yet Started ▾	2020-05-18
Sally	Not Yet Started ▾	2020-05-25
Erica	Not Yet Started ▾	2020-04-24
Ayanna	Not Yet Started ▾	2020-04-27
Hallie	Not Yet Started ▾	2020-05-18
Tony	Not Yet Started ▾	2020-05-25

Format Data Tools Extensions Help Las

Conditional format rules X

Theme New

Number

Text

Alignment

Wrapping

Rotation

Font size

Merge cells

Conditional formatting (highlighted)

Alternating colors

Clear formatting Ctrl+V

Single color Color scale

Apply to range C1:C1000

Format rules

Format cells if... Is not empty

Formatting style Default

B I U S A More

Cancel Done

Định dạng có điều kiện

Chọn như hình

The screenshot shows a Google Sheets interface with a conditional formatting rule applied to the range C1:C1000. The rule is set to format cells where the value in column C is "Not Yet Started". The "Formatting style" is set to "Default". The dialog includes tabs for "Single color" and "Color scale", and a section for applying the rule to a range.

C	D	E	F	G	H
Status	Review By This Date	Review	Conditional format rules X		
Not Yet Started	2020-03-13	<input type="checkbox"/>			
Not Yet Started	2020-03-19	<input type="checkbox"/>			
Not Yet Started	2020-04-02	<input type="checkbox"/>			
Not Yet Started	2020-04-10	<input type="checkbox"/>			
Not Yet Started	2020-04-17	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			
Not Yet Started	2020-05-25	<input type="checkbox"/>			
Not Yet Started	2020-04-24	<input type="checkbox"/>			
Not Yet Started	2020-04-27	<input type="checkbox"/>			
Not Yet Started	2020-05-18	<input type="checkbox"/>			

Định dạng có điều kiện

Single color

Color scale

Apply to range

C1:C1000

Format rules if...

Text is exactly

In Progress

Formatting style

Custom

B I U & A Font color

Cancel Done

+ Add another rule

Single color

Color scale

Apply to range

C1:C1000

Format rules if...

Text is exactly

Ready

Formatting style

Custom

B I U & A Font color

Cancel Done

+ Add another rule

Conditional format rules

X

Format	Condition	Range
Green	Text is exactly "Not Yet Started"	C1:C1000
Red	Text is exactly "In Progress"	C1:C1000
Blue	Text is exactly "Ready"	C1:C1000

+ Add another rule

Định dạng có điều kiện

Kết quả

Asignee	Status	Review By This Date	Review
Tony	Not Yet Started ▾	2020-03-13	<input type="checkbox"/>
Ximena	Ready ▾	2020-03-19	<input type="checkbox"/>
Sally	Not Yet Started ▾	2020-04-02	<input type="checkbox"/>
Erica	Not Yet Started ▾	2020-04-10	<input type="checkbox"/>
Ayanna	In Progress ▾	2020-04-17	<input type="checkbox"/>
Hallie	Not Yet Started ▾	2020-04-24	<input type="checkbox"/>
Tony	In Progress ▾	2020-04-27	<input type="checkbox"/>
Ximena	Not Yet Started ▾	2020-05-18	<input type="checkbox"/>
Sally	Ready ▾	2020-05-25	<input type="checkbox"/>
Erica	Not Yet Started ▾	2020-04-24	<input type="checkbox"/>
Ayanna	Not Yet Started ▾	2020-04-27	<input type="checkbox"/>

Hàm CONCAT trong SQL

- **CONCAT:** nối chuỗi từ nhiều nguồn
- Ví dụ:

```
1 SELECT
2   usertype,
3   CONCAT(start_station_name, "_to_", end_station_name) AS route,
4   count(*) AS num_trip
5 FROM `bigquery-public-data.new_york_citibike.citibike_trips`
6 GROUP BY
7   start_station_name, end_station_name, usertype
8 LIMIT 10
```

LIMIT 10: chỉ lấy 10 hàng đầu tiên.

Hàm CONCAT trong SQL

Kết quả

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	usertype	route		num_trip
1	Subscriber	E 45 St & 3 Ave_to_E 80 St & 2 Ave		90
2	Subscriber	Mercer St & Bleecker St_to_State St		76
3	Subscriber	Clinton St & Joralemon St_to_Emerson Pl & Myrtle Ave		37
4	Subscriber	5 Ave & E 88 St_to_E 68 St & 3 Ave		36
5	Subscriber	Queens Plaza North & Crescent St_to_Allen St & Stanton St		40
6	Subscriber	Greene Ave & Throop Ave_to_Kingston Ave & Herkimer St		47

Một số hàm xử lý chuỗi trong bảng tính

- **LEN**: chiều dài của chuỗi
- **LEFT**: bên trái chuỗi
- **RIGHT**: bên phải chuỗi
- **FIND**: tìm kiếm một ký tự trong chuỗi



Một số hàm xử lý chuỗi trong bảng tính

Bộ dữ liệu City Bike Trip

	A	B	C	D	E	F	G	H	I	J
1	tripduration		starttime	stoptime	start station id	start station name	start station latitu	start station	end station id	end station na
2	364		2017-09-01 00:02:01	2017-09-01 00:08:05	3183	Exchange Place	40.7162469	-74.0334588	3276	Marin Light Rai
3	357		2017-09-01 00:08:12	2017-09-01 00:14:09	3187	Warren St	40.7211236	-74.0380509	3199	Newport Pkwy
4	432		2017-09-01 0:10:02	2017-09-01 00:17:24	3195	Sip Ave	40.73074263	-74.0637838	3280	Astor Place
5	934		2017-09-01 00:10:11	2017-09-01 00:25:46	3272	Jersey & 3rd	40.72333159	-74.0459525	3207	Oakland Ave
6	932		2017-09-01 00:10:16	2017-09-01 00:25:48	3272	Jersey & 3rd	40.72333159	-74.0459525	3207	Oakland Ave
7	414		2017-09-01 00:15:32	2017-09-01 00:22:26	3186	Grove St PATH	40.71958612	-74.0431174	3480	WS Don't Use
8	625		2017-09-01 00:15:32	2017-09-01 00:25:58	3194	McGinley Square	40.72533993	-74.0676221	3194	McGinley Squa
9	178		2017-09-01 00:36:20	2017-09-01 00:39:19	3183	Exchange Place	40.7162469	-74.0334588	3214	Essex Light Rail
10	557		2017-09-01 00:38:42	2017-09-01 00:48:00	3183	Exchange Place	40.7162469	-74.0334588	3199	Newport Pkwy
11	220		2017-09-01 00:40:11	2017-09-01 00:43:52	3187	Warren St	40.7211236	-74.0380509	3270	Jersey & 6th St
12	153		2017-09-01 01:24:28	2017-09-01 01:27:02	3272	Jersey & 3rd	40.72333159	-74.0459525	3278	Monmouth and
13	809		2017-09-01 02:05:11	2017-09-01 02:18:41	3206	Hilltop	40.7311689	-74.0575736	3281	Leonard Gord

Một số hàm xử lý chuỗi trong bảng tính

A	B	C
tripduration	19 ×	
364	=LEN(C2)	2017-09-01 00:02:01
357		
432	LEN(text)	
934		
222		

Một số hàm xử lý chuỗi trong bảng tính

A	B	C	D
tripduration	2017-09-01 ×	starttime	s
364		19 2017-09-01 00:02:01	
357	=LEFT(C3,10)	2017-09-01 00:08:12	x
432			
934	LEFT(string, [number_of_characters])		▼
932			⋮

Một số hàm xử lý chuỗi trong bảng tính

A	B	C	D
tripduration		starttime	s
364	0:10:02 × 1	19	2017-09-01 00:02:01
357			2017-09-01 00:08:12
432	=RIGHT(C4, 8)		2017-09-01 0:10:02
934			
932	RIGHT(string, [number_of_characters])		
414			

Một số hàm xử lý chuỗi trong bảng tính

A	B	C	D
tripduration		starttime	\$1
364		19	2017-09-01 00:02:01
357	2017-09-01		2017-09-01 00:08:12
432	11 × 02		2017-09-01 0:10:02
934	=FIND(" ",C5)		2017-09-01 00:10:11
932			
414	FIND(search_for, text_to_search,		▼
625	[starting_at])		
179			



3 TỔNG HỢP DỮ LIỆU ĐỂ PHÂN TÍCH



Tổng hợp dữ liệu

Tổng hợp: thu thập hoặc tập hợp nhiều phần riêng biệt thành một tổng thể

Tổng hợp dữ liệu: quá trình thu thập dữ liệu từ nhiều nguồn để kết hợp thành một tập hợp tóm tắt duy nhất.

Lợi ích của tổng hợp dữ liệu:

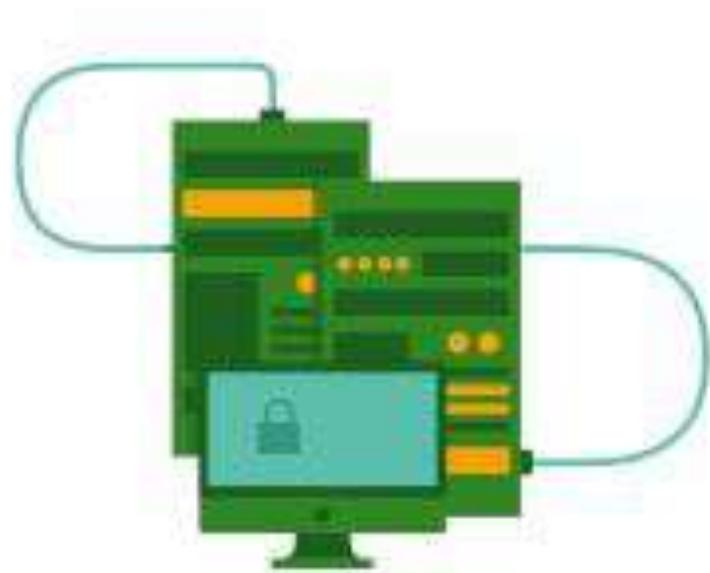
- Xác định xu hướng
- Thực hiện so sánh
- Đạt được kiến thức hữu ích



Tổng hợp dữ liệu

Dữ liệu được tổng hợp trong một khoảng thời gian nhất định để cung cấp số liệu thống kê:

- Trung bình (average)
- Tối thiểu (min)
- Tối đa (max)
- Tổng (sum)



Tổng hợp dữ liệu

Biến đổi kiểu dữ liệu:

- Ngày được định dạng dưới dạng số
- Số được biểu thị dưới dạng chuỗi

Định dạng thiếu nhất quán: hàm VLOOKUP
sẽ trả về lỗi.



Biến đổi chuỗi thành số

Sử dụng bộ dữ liệu

	A	B
1	Value stored as plain text	Numeric value
2	123	
3	456	
4	789	

Cột A là số lưu dưới dạng chuỗi

	A	B
1	Value stored as plain text	Numeric value
2	123	
3	456	
4	789	
5	=sum(A2:A4)	

Biến đổi chuỗi thành số

VALUE: hàm chuyển đổi một chuỗi văn bản thể hiện số thành một giá trị số

	A	B
1	Value stored as plain text	123 × Numeric value
2	123	=VALUE(A2)
3	456	
4	789	
5	0	

	A	B	C
1	Value stored as plain text	Numeric value	
2	123	123	
3	456	456	
4	789	789	
5	0	=SUM(B2:B4)	
6			
7		SUM(value1, [value2, ...])	
8			

Hàm TRIM

TRIM: hàm tự động xóa bất kỳ khoảng trắng thừa nào được thêm vào ô

	A	B
1		
2	Large Animal	

	A	B
1		
2	Large Animal	$=\text{TRIM}(A2)$

Xóa mục trùng lặp

Xóa mục trùng lặp (Remove duplicate)

một công cụ tự động tìm kiếm và loại bỏ các hàng trùng lặp khỏi bảng tính

A screenshot of a Google Sheets spreadsheet. The spreadsheet contains two columns of address data. A context menu is open over the data, with the 'Remove duplicates' option highlighted at the bottom. Other visible options in the menu include sorting by column H (A-Z or Z-A), filtering, and pivot tables.

Remove duplicates

1 duplicate row found and removed.

70 unique rows remain.

OK

Tổng hợp dữ liệu - VLOOKUP

VLOOKUP (Vertical Lookup)

một hàm tìm kiếm một giá trị nhất định trong một cột để trả về một phần thông tin tương ứng

=VLOOKUP(103, A2:B26, 2, FALSE)

Ví Dụ:

- 103: giá trị cần tìm
- A2:B26 phạm vi tìm kiếm
- 2: cột 2
- FALSE: tìm chính xác

VLOOKUP

Employee #	Rate of Pay
FT12578	\$25.00
FT12579	\$13.00
FT12580	\$42.00
FT12581	\$25.00

+ ⌂ Employee Rates ▾ Employee Hours ▾

Employee #	Hours Worked
FT12578	20
FT12579	20
FT12580	20
FT12581	20

+ ⌂ Employee Rates ▾ Employee Hours ▾

VLOOKUP

Employee #	Hours Worked
FT12578	2 =?=vlookup(A2, 'Employee Rates'!\$A\$2:\$B\$5, 2, FALSE)
FT12579	20
FT12580	20
FT12581	20

- Hàm VLOOKUP để tìm kiếm mức lương từ tab 'Employee Rates' và thêm nó vào tab 'Employee Hours'
- Lưu ý: \$A\$2: \$B\$5: tham chiếu tuyệt đối

VLOOKUP

Employee #	Hours Worked	Pay Rate	
FT12578	20		\$25.00
FT12579	20		\$13.00
FT12580	20		\$42.00
FT12581	20		\$25.00

Các câu hỏi để xử lý sự cố

- Nên **ưu tiên** những **vấn đề** như thế nào?
- Hỏi trong **một câu** duy nhất, vấn đề mình đang gặp phải?
- Những **tài nguyên** nào có thể giúp giải quyết vấn đề?
- Làm cách nào để **ngăn chặn** vấn đề xảy ra trong tương lai?



Một số giới hạn của VLOOKUP

- Hàm VLOOKUP chỉ trả về kết quả **phù hợp đầu tiên**
- VLOOKUP chỉ có thể trả về một giá trị từ **dữ liệu ở bên phải**. Nó không thể nhìn sang trái.
- Nhớ dùng **tham chiếu tuyệt đối (absolute reference)**: tham chiếu được khóa để các hàng và cột không thay đổi khi sao chép.
- Nơi tham chiếu đến bị thay đổi.
 - Data -> Protected sheet and range.
- Lựa chọn tham số TRUE/FALSE phù hợp.



Câu lệnh JOINS trong SQL

JOIN: mệnh đề SQL được sử dụng để kết hợp các hàng từ hai hoặc nhiều bảng dựa trên một cột có liên quan.

Có 4 loại JOIN:

- Inner
- Left
- Right
- Outer

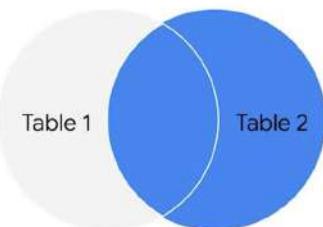
INNER JOIN



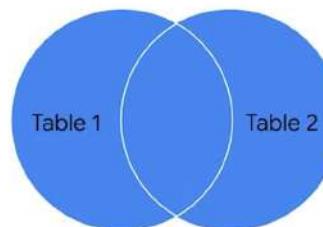
LEFT JOIN



RIGHT JOIN

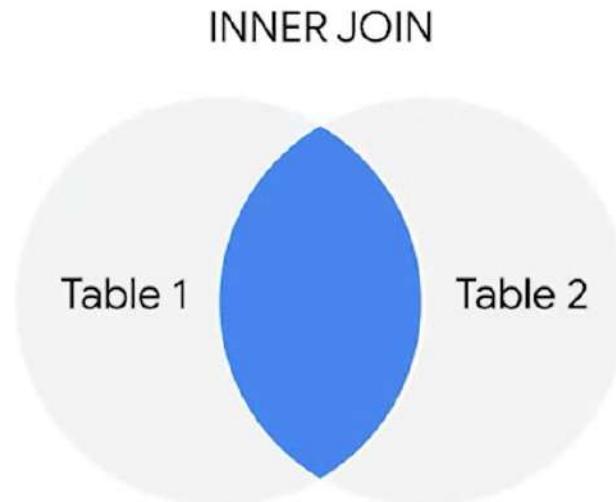


FULL OUTER JOIN



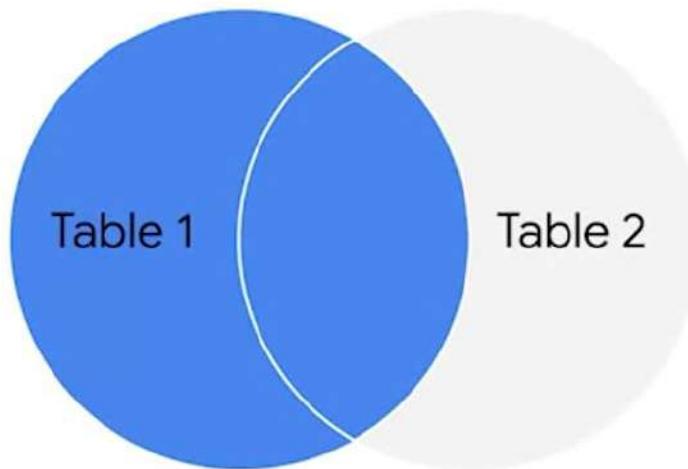
INNER JOIN

INNER JOIN: một hàm trả về các bản ghi có giá trị phù hợp trong cả hai bảng



LEFT JOIN

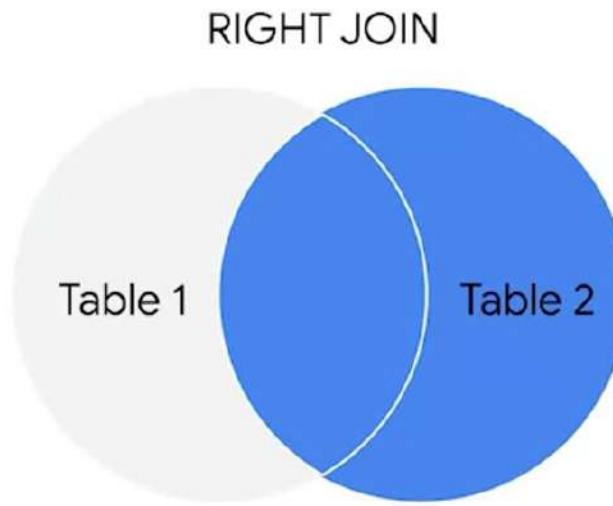
LEFT JOIN: hàm trả về tất cả các bản ghi từ bảng **bên trái** và chỉ các bản ghi phù hợp từ bảng **bên phải**.



```
SELECT
*
FROM
    tableA
LEFT JOIN
    tableB
ON
    keyA = keyB
```

RIGHT JOIN

RIGHT JOIN: hàm trả về tất cả các bản ghi từ bảng **bên phải** và chỉ các bản ghi phù hợp từ bảng **bên trái**.

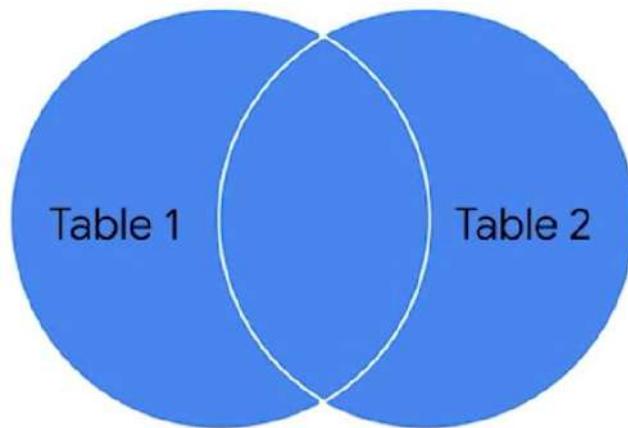


```
SELECT
*
FROM
    tableB
RIGHT JOIN
    tableA
ON
    keyA = keyB
```

OUTER JOIN

OUTER JOIN: kết hợp *RIGHT* và *LEFT JOIN* để trả về tất cả các bản ghi phù hợp trong cả hai bảng

FULL OUTER JOIN



INNER JOIN

Bộ dữ liệu employee_data

Row	name	department_id	role
1	Dave Smith	1	Product Marketing Manager
2	Scott Tanner	1	Director of Demand Gen
3	Margaret Lane	1	VP of Marketing
4	Julie Jones	2	Software Engineer
5	Ted Connors	2	Software Engineer
6	Mary Martin	5	Receptionist

Row	name	department_id
1	Marketing	1
2	Engineering	2
3	Accounting	3
4	Sales	4

INNER JOIN

```
1 SELECT
2   employees.name AS employee_name,
3   employees.role AS employee_role,
4   departments.name AS department_name
5
6 FROM testbigquery-353006.employee_data.employees
7 INNER JOIN
8   testbigquery-353006.employee_data.departments ON
9   employees.department_ID = departments.department_ID
```

INNER JOIN

Row	employee_name	employee_role	department_name
1	Dave Smith	Product Marketing Manager	Marketing
2	Scott Tanner	Director of Demand Gen	Marketing
3	Margaret Lane	VP of Marketing	Marketing
4	Julie Jones	Software Engineer	Engineering
5	Ted Connors	Software Engineer	Engineering

LEFT JOIN

```
1 SELECT
2   employees.name AS employee_name,
3   employees.role AS employee_role,
4   departments.name AS department_name
5
6 FROM testbigquery-353006.employee_data.employees
7 LEFT JOIN
8   testbigquery-353006.employee_data.departments ON
9   employees.department_ID = departments.department_ID
```

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	employee_name	employee_role	department_name	
1	Dave Smith	Product Marketing Manager	Marketing	
2	Scott Tanner	Director of Demand Gen	Marketing	
3	Margaret Lane	VP of Marketing	Marketing	
4	Julie Jones	Software Engineer	Engineering	
5	Ted Connors	Software Engineer	Engineering	
6	Mary Martin	Receptionist	null	

RIGHT JOIN

```
1 SELECT
2   employees.name AS employee_name,
3   employees.role AS employee_role,
4   departments.name AS department_name
5
6 FROM testbigquery-353006.employee_data.employees
7 RIGHT JOIN
8   testbigquery-353006.employee_data.departments ON
9   employees.department_ID = departments.department_ID
```

Row	employee_name	employee_role	department_na
1	Dave Smith	Product Marketing Manager	Marketing
2	Scott Tanner	Director of Demand Gen	Marketing
3	Margaret Lane	VP of Marketing	Marketing
4	Julie Jones	Software Engineer	Engineering
5	Ted Connors	Software Engineer	Engineering
6	<i>null</i>	<i>null</i>	Accounting
7	<i>null</i>	<i>null</i>	Sales

OUTER JOIN

```
1 SELECT
2   employees.name AS employee_name,
3   employees.role AS employee_role,
4   departments.name AS department_name
5
6 FROM testbigquery-353006.employee_data.employees
7 FULL OUTER JOIN
8   testbigquery-353006.employee_data.departments ON
9   employees.department_ID = departments.department_ID
```

Row	employee_name	employee_role	department_name
1	Julie Jones	Software Engineer	Engineering
2	Ted Connors	Software Engineer	Engineering
3	Dave Smith	Product Marketing Manager	Marketing
4	Scott Tanner	Director of Demand Gen	Marketing
5	Margaret Lane	VP of Marketing	Marketing
6	Mary Martin	Receptionist	null
7	null	null	Sales
8	null	null	Accounting

Hàm COUNT

- **COUNT (bảng tính)**: đếm tổng số giá trị số trong một phạm vi cụ thể
- **COUNT (SQL)**: câu truy vấn trả về số hàng trong một phạm vi được chỉ định
- **COUNT DISTINCT**: câu truy vấn chỉ trả về các **giá trị riêng biệt** trong phạm vi đó.
- Dùng để trả lời cho câu hỏi: **bao nhiêu?**



Hàm COUNT

Bộ dữ liệu warehouse_orders: gồm hai bảng warehouse và order

Row	order_id	customer_id	warehouse_id	order_date	shipper_date
1	789	3731	8118	2019-01-01	2019-01-04
2	790	3486	8118	2019-01-01	2019-01-04
3	791	2623	8118	2019-01-01	2019-01-04
4	792	9869	8118	2019-01-01	2019-01-04
5	793	6866	8118	2019-01-01	2019-01-04

Row	warehouse_id	warehouse_alias	maximum_capacity	employee_total	state
1	1543	Somerset Fulfillment Center	210	14	KY
2	2270	Bowling Green Warehouse	280	13	KY
3	9080	Frankfort Fulfillment Center	235	5	KY
4	2666	Lansing Fulfillment Center	290	16	MI
5	3961	Lansing Storage Warehouse	740	22	MI
6	8118	Ann Arbor Fulfillment Center	780	17	MI

Hàm COUNT

```
1  SELECT
2    orders.*,
3    warehouses.warehouse_alias,
4    warehouses.state
5  FROM warehouse_orders.Orders orders
6  JOIN
7    warehouse_orders.Warehouse warehouses
8  ON
9    orders.warehouse_id = warehouses.warehouse_id
```

Row	order_id	customer_id	warehouse_id	order_date	shipper_date	warehouse_alias	state
1	789	3731	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI
2	790	3486	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI
3	791	2623	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI
4	792	9869	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI
5	793	6866	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI
6	794	8055	8118	2019-01-01	2019-01-04	Ann Arbor Fulfillment Center	MI

Hàm COUNT

Bộ dữ liệu warehouse_orders: gồm hai bảng warehouse và order

```
1 SELECT
2   COUNT(warehouses.state) AS num_state
3   FROM warehouse_orders.Orders orders
4   JOIN
5     warehouse_orders.Warehouse warehouses
6   ON
7     orders.warehouse_id = warehouses.warehouse_id
```

JOB INFORMATION		RESULTS
Row	num_state	
1	9999	

Hàm COUNT

Câu hỏi: có bao nhiêu trạng thái (state) trong dữ liệu về đơn hàng (order)?

```
1 SELECT
2 | COUNT (DISTINCT warehouses.state) AS num_state
3 FROM warehouse_orders.Orders orders
4 JOIN
5 | warehouse_orders.Warehouse warehouses
6 ON
7 | orders.warehouse_id = warehouses.warehouse_id
```

JOB INFORMATION

RESULTS

Row	num_state
1	3

Hàm COUNT

```
1 SELECT
2   warehouses.state AS state,
3   COUNT (DISTINCT order_id) AS num_order
4 FROM warehouse_orders.Orders orders
5 JOIN
6   warehouse_orders.Warehouse warehouses
7 ON
8   orders.warehouse_id = warehouses.warehouse_id
9 GROUP BY
10 warehouses.state
```

Row	state	num_order
1	MI	6205
2	KY	1048
3	TN	2746

Câu truy vấn con

- **Truy vấn con:** một truy vấn SQL được lồng vào bên trong một truy vấn lớn hơn
- **Truy vấn ngoài (outer query) hoặc lựa chọn ngoài (outer select):** Câu lệnh chứa truy vấn con
- **Truy vấn trong (inner query) hoặc lựa chọn trong (inner select):** câu truy vấn con
- Truy vấn bên trong **thực thi trước** để kết quả có thể được chuyển cho truy vấn bên ngoài để sử dụng



Câu truy vấn con

Bộ dữ liệu: *new_york_citibike* trong *bigquery-public-data*

Row	station_id	name	short_name	latitude
1442	259	South St & Whitehall St	4846.01	40.70122128
1443	3686	Gansevoort St & Hudson St	6072.14	40.739448201304036
1444	4037	Roosevelt Island Tramway	6599.01	40.757284
1445	4474	10 Ave & W 207 St	8613.09	40.864632

Row	tripduration	starttime	stoptime	start_station_id
1	432	2013-09-16T19:22:43	2013-09-16T19:29:55	509
2	1186	2015-12-30T13:02:38	2015-12-30T13:22:25	280
3	799	2017-09-02T16:27:37	2017-09-02T16:40:57	335
4	238	2017-11-15T06:57:09	2017-11-15T07:01:08	146

Câu truy vấn con

```
1 SELECT
2   station_id,
3   num_bikes_available,
4   (SELECT
5     | avg(num_bikes_available)
6   FROM `bigquery-public-data.new_york_citibike.citibike_stations` AS avg_num_bikes_available
7   FROM `bigquery-public-data.new_york_citibike.citibike_stations`
```

Row	station_id	num_bikes_available	avg_num_bikes_available
1445	4474	35	14.527146464646446
1446	3382	35	14.527146464646446
1447	380	35	14.527146464646446
1448	254	35	14.527146464646446
1449	347	35	14.527146464646446

Câu truy vấn con

- **HAVING:** cho phép bạn *thêm bộ lọc* vào câu truy vấn của mình thay vì bảng nền khi bạn đang làm việc với các hàm tổng hợp.
- **CASE:** trả về các bản ghi *theo điều kiện xác định* bằng cách đưa các câu lệnh if/then vào truy vấn của mình.
- **Bí danh (Aliasing):** đặt tên tạm cho một bảng hoặc cột trong truy vấn của để giúp đọc và viết dễ dàng hơn.
- Bộ dữ liệu **warehouse_orders**



Câu truy vấn con

```
FROM warehouse_orders.Orders orders
JOIN
| warehouse_orders.Warehouse warehouses
ON
| orders.warehouse_id = warehouses.warehouse_id
GROUP BY
| warehouses.warehouse_id,
| warehouse_name
```

Câu truy vấn con

```
1 SELECT
2   warehouses.warehouse_id,
3   CONCAT(warehouses.state, ' : ',warehouses.warehouse_alias) AS warehouse_name,
4   COUNT(orders.order_id) AS number_of_order
5 FROM warehouse_orders.Orders orders
6 JOIN
7   warehouse_orders.Warehouse warehouses
8 ON
9   orders.warehouse_id = warehouses.warehouse_id
10 GROUP BY
11   warehouses.warehouse_id,
12   warehouse_name
```

Câu truy vấn con

```
1 SELECT
2     warehouses.warehouse_id,
3     CONCAT(warehouses.state, ' : ',warehouses.warehouse_alias) AS warehouse_name,
4     COUNT(orders.order_id) AS number_of_order,
5     (SELECT
6         COUNT(*)
7         FROM warehouse_orders.Orders orders) AS total_orders
8     FROM warehouse_orders.Orders orders
9     JOIN
10    warehouse_orders.Warehouse warehouses
11    ON
12    orders.warehouse_id = warehouses.warehouse_id
13    GROUP BY
14    warehouses.warehouse_id,
15    warehouse_name
```

Câu truy vấn con

```
1 SELECT
2     warehouses.warehouse_id,
3     CONCAT(warehouses.state, ':' ,warehouses.warehouse_alias) AS warehouse_name,
4     COUNT(orders.order_id) AS number_of_order,
5     (SELECT
6         COUNT(*)
7         FROM warehouse_orders.Orders orders) AS total_orders,
8     CASE
9         WHEN COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) <= 0.20
10        THEN "Fullfilled 0-20% of Orders"
11        WHEN COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) > 0.20
12        AND COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) <= 0.60
13        THEN "Fullfilled 21-60% of Orders"
14        ELSE "Fullfilled more than 60% of Orders"
15    END AS fulfillment_summary
```

Câu truy vấn con

```
16 FROM warehouse_orders.Orders orders
17 JOIN
18   warehouse_orders.Warehouse warehouses
19 ON
20   orders.warehouse_id = warehouses.warehouse_id
21 GROUP BY
22   warehouses.warehouse_id,
23   warehouse_name
24 HAVING
25   COUNT(orders.order_id) > 0
```

Câu truy vấn con

```
1  SELECT
2    warehouses.warehouse_id,
3    CONCAT(warehouses.state, ' : ',warehouses.warehouse_alias) AS warehouse_name,
4    COUNT(orders.order_id) AS number_of_order,
5    (SELECT
6      COUNT(*)
7      FROM warehouse_orders.Orders orders) AS total_orders,
8    CASE
9      WHEN COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) <= 0.20
10     THEN "Fullfilled 0-20% of Orders"
11     WHEN COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) > 0.20
12     AND COUNT(orders.order_id)/(SELECT COUNT(*) FROM warehouse_orders.Orders orders ) <= 0.60
13     THEN "Fullfilled 21-60% of Orders"
14     ELSE "Fullfilled more than 60% of Orders"
15   END AS fulfillment_summary
16  FROM warehouse_orders.Orders orders
17  JOIN
18    warehouse_orders.Warehouse warehouses
19  ON
20    orders.warehouse_id = warehouses.warehouse_id
21  GROUP BY
22    warehouses.warehouse_id,
23    warehouse_name
24  HAVING
25    COUNT(orders.order_id) > 0
```

Câu truy vấn con

Kết quả

Row	warehouse_id	warehouse_name	number_of_order	total_orders	fulfillment_summary
1	8118	MI: Ann Arbor Fulfillment Center	3027	9999	Fullfilled 21-60% of Orders
2	2666	MI: Lansing Fulfillment Center	3178	9999	Fullfilled 21-60% of Orders
3	9080	KY: Frankfort Fulfillment Center	500	9999	Fullfilled 0-20% of Orders
4	1543	KY: Somerset Fulfillment Center	548	9999	Fullfilled 0-20% of Orders
5	6509	TN: Memphis Fulfillment Center	2403	9999	Fullfilled 21-60% of Orders
6	4338	TN: Knoxville Fulfillment Center	343	9999	Fullfilled 0-20% of Orders



4 TÍNH TOÁN TRÊN DỮ LIỆU



Tính toán trên dữ liệu

- Công thức tính cơ bản
- Công thức có điều kiện sử dụng hàm IF
- Hàm SUMPRODUCT.
- Bảng tổng hợp (pivot table) để tổ chức tính toán.
- Truy vấn và tính toán trong SQL.
- Bảng tạm trong SQL



Công thức tính cơ bản

Doanh thu tháng	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12
2011	\$ 286.00	\$ 719.00	\$ 513.00	\$ 718.00	\$ 929.00	\$ 510.00	\$ 964.00	\$ 878.00	\$ 104.00	\$ 783.00	\$ 581.00	\$ 594.00
2012	\$ 750.00	\$ 680.00	\$ 775.00	\$ 706.00	\$ 335.00	\$ 322.00	\$ 409.00	\$ 702.00	\$ 867.00	\$ 861.00	\$ 464.00	\$ 536.00
2013	\$ 777.00	\$ 713.00	\$ 200.00	\$ 125.00	\$ 122.00	\$ 296.00	\$ 978.00	\$ 233.00	\$ 720.00	\$ 224.00	\$ 660.00	\$ 218.00
2014	\$ 471.00	\$ 635.00	\$ 337.00	\$ 807.00	\$ 407.00	\$ 517.00	\$ 556.00	\$ 774.00	\$ 984.00	\$ 333.00	\$ 796.00	\$ 501.00
2015	\$ 625.00	\$ 243.00	\$ 112.00	\$ 894.00	\$ 319.00	\$ 724.00	\$ 288.00	\$ 476.00	\$ 673.00	\$ 753.00	\$ 336.00	\$ 591.00
2016	\$ 311.00	\$ 669.00	\$ 608.00	\$ 961.00	\$ 995.00	\$ 111.00	\$ 446.00	\$ 524.00	\$ 386.00	\$ 338.00	\$ 455.00	\$ 796.00
2017	\$ 239.00	\$ 472.00	\$ 392.00	\$ 784.00	\$ 546.00	\$ 770.00	\$ 775.00	\$ 182.00	\$ 937.00	\$ 295.00	\$ 695.00	\$ 139.00
2018	\$ 298.00	\$ 186.00	\$ 186.00	\$ 164.00	\$ 196.00	\$ 269.00	\$ 384.00	\$ 580.00	\$ 595.00	\$ 489.00	\$ 740.00	\$ 783.00
2019	\$ 901.00	\$ 495.00	\$ 716.00	\$ 533.00	\$ 327.00	\$ 382.00	\$ 384.00	\$ 847.00	\$ 161.00	\$ 849.00	\$ 730.00	\$ 211.00
2020	\$ 153.00	\$ 109.00	\$ 495.00	\$ 941.00	\$ 103.00	\$ 609.00	\$ 353.00	\$ 551.00	\$ 507.00	\$ 739.00	\$ 523.00	\$ 339.00

Công thức tính cơ bản

Doanh thu	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12	\$ 7,579.00 × m
2011	\$ 286.00	\$ 719.00	\$ 513.00	\$ 718.00	\$ 929.00	\$ 510.00	\$ 964.00	\$ 878.00	\$ 104.00	\$ 783.00	\$ 581.00	\$ 594.00	=sum(B20:M20)
2012	\$ 750.00	\$ 680.00	\$ 775.00	\$ 706.00	\$ 335.00	\$ 322.00	\$ 409.00	\$ 702.00	\$ 867.00	\$ 861.00	\$ 464.00	\$ 536.00	
2013	\$ 777.00	\$ 713.00	\$ 200.00	\$ 125.00	\$ 122.00	\$ 296.00	\$ 978.00	\$ 233.00	\$ 720.00	\$ 224.00	\$ 660.00	\$ 218.00	

Tháng 10	Tháng 11	Tháng 12	Doanh thu năm
\$ 783.00	\$ 581.00	\$ 594.00	\$ 7,579.00
\$ 861.00	\$ 464.00	\$ 536.00	
\$ 224.00	\$ 660.00	\$ 218.00	
\$ 333.00	\$ 796.00	\$ 501.00	
\$ 753.00	\$ 336.00	\$ 591.00	
\$ 338.00	\$ 455.00	\$ 796.00	
\$ 295.00	\$ 695.00	\$ 139.00	
\$ 489.00	\$ 740.00	\$ 783.00	
\$ 849.00	\$ 730.00	\$ 211.00	
\$ 739.00	\$ 523.00	\$ 339.00	

Công thức tính cơ bản

Doanh thu năm	Tăng trưởng hàng năm
\$4,870.00	\$396.00 ×
\$5,266.00	=N21 - N20
\$5,422.00	
\$6,034.00	
\$6,226.00	
\$6,536.00	
\$6,600.00	
\$7,118.00	
\$7,407.00	
\$7,579.00	

Doanh thu năm	Tăng trưởng hàng năm	% tăng
\$4,870.00		0.08131416838 ×
\$5,266.00	\$396.00	=D21 / N20
\$5,422.00	\$156.00	
\$6,034.00	\$612.00	
\$6,226.00	\$192.00	
\$6,536.00	\$310.00	
\$6,600.00	\$64.00	
\$7,118.00	\$518.00	
\$7,407.00	\$289.00	
\$7,579.00	\$172.00	

Format Data Tools Extensions Help Last edit was seconds ago

New **B** *I* A

Theme

Number

Automatic Plain text
Number 1,000.12 10.12% 1.01E+03 \$(1,000.12) \$(1,000.12) \$1,000.12 \$1,000.00 08/13/2016 00:08:13 08/13/2016 00:08:13 00:08:13

Text

Alignment

Wrapping

Rotation

Font size

Merge cells

Conditional formatting

Alternating colors

Clear formatting Ctrl+K

0.08131416838

0.08131416838

0.02962400304

0.1128734784

0.03181988843

0.0497911962

0.009791921065

0.07848484848

0.0406012925

0.02322127717

Công thức tính cơ bản

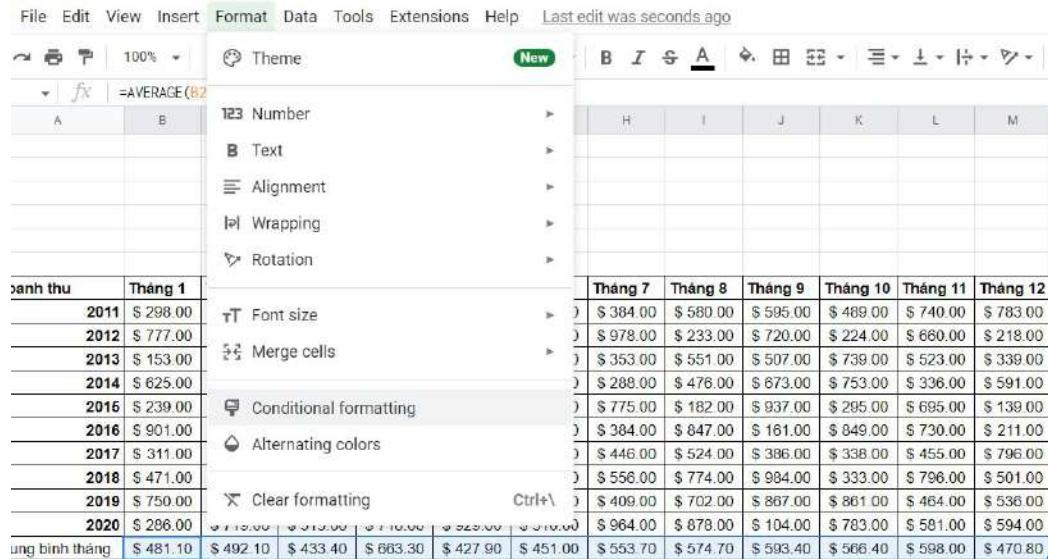
Doanh thu năm	Tăng trưởng hàng năm	% tăng
\$4,870.00		
\$5,266.00	\$396.00	8.13%
\$5,422.00	\$156.00	2.96%
\$6,034.00	\$612.00	11.29%
\$6,226.00	\$192.00	3.18%
\$6,536.00	\$310.00	4.98%
\$6,600.00	\$64.00	0.98%
\$7,118.00	\$518.00	7.85%
\$7,407.00	\$289.00	4.06%
\$7,579.00	\$172.00	2.32%

Công thức tính cơ bản

17		AVEDEV	
18			
19	Doanh thu	AVERAGE(B20:B29)	\$ 481.10
20	2011	AVERAGE	
21	2012	AVERAGEA	
22	2013	DAVERAGE	
23	2014	AVERAGEIF	
24	2015	AVERAGEIFS	
25	2016		
26	2017		
27	2018	AVERAGE.WEIGHTED	
28	2019	Tab to accept. ↑ ↓ to navigate ⏮	
29	2020	\$ 286.00 \$ 719.00 \$ 513.00 \$ 718.00 \$ 929.00	
30	Trung bình tháng	=AVERAGE(B20:B29)	

Định dạng có điều kiện

Định dạng có điều kiện: công cụ trong bảng tính cho phép thay đổi bề ngoài của một ô khi giá trị thỏa một điều kiện nào đó



The screenshot shows a Google Sheets interface with a table of annual revenue data from 2011 to 2020. The 'Format' menu is open, and the 'Conditional formatting' option is highlighted. The table has columns for 'Năm thu' (Year) and 'Tháng 1' (Month 1), followed by months from Tháng 7 to Tháng 12.

Năm thu	Tháng 1	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12
2011	\$ 298.00	\$ 384.00	\$ 580.00	\$ 595.00	\$ 489.00	\$ 740.00	\$ 783.00
2012	\$ 777.00	\$ 978.00	\$ 233.00	\$ 720.00	\$ 224.00	\$ 660.00	\$ 218.00
2013	\$ 153.00	\$ 353.00	\$ 551.00	\$ 507.00	\$ 739.00	\$ 523.00	\$ 339.00
2014	\$ 625.00	\$ 288.00	\$ 476.00	\$ 673.00	\$ 753.00	\$ 336.00	\$ 591.00
2015	\$ 239.00	\$ 775.00	\$ 182.00	\$ 937.00	\$ 295.00	\$ 685.00	\$ 139.00
2016	\$ 901.00	\$ 384.00	\$ 847.00	\$ 161.00	\$ 849.00	\$ 730.00	\$ 211.00
2017	\$ 311.00	\$ 446.00	\$ 524.00	\$ 386.00	\$ 338.00	\$ 455.00	\$ 796.00
2018	\$ 471.00	\$ 556.00	\$ 774.00	\$ 984.00	\$ 333.00	\$ 796.00	\$ 501.00
2019	\$ 750.00	\$ 409.00	\$ 702.00	\$ 867.00	\$ 861.00	\$ 464.00	\$ 536.00
2020	\$ 286.00	\$ 964.00	\$ 878.00	\$ 104.00	\$ 783.00	\$ 581.00	\$ 594.00
Trung bình tháng	\$ 481.10	\$ 492.10	\$ 433.40	\$ 863.30	\$ 427.90	\$ 451.00	\$ 553.70

Định dạng có điều kiện

Định dạng có điều kiện: công cụ trong bảng tính cho phép thay đổi bề ngoài

The screenshot shows a Google Sheets interface with a conditional format rules dialog open. The formula bar shows '=AVERAGE(B20:B29)'. The range B30:M30 is selected for the rule. The 'Color scale' option is chosen. The 'Default' midpoint is set at \$211.00. The 'Min value' is set at \$139.00 and the 'Max value' is set at \$579.00. The table below shows monthly sales data from 2011 to 2020.

Doanh thu	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12	Doanh thu năm	Tăng trưởng
2011	\$ 288.00	\$ 186.00	\$ 106.00	\$ 164.00	\$ 166.00	\$ 266.00	\$ 384.00	\$ 595.00	\$ 488.00	\$ 740.00	\$ 783.00	\$ 4.870.00		
2012	\$ 777.00	\$ 713.00	\$ 200.00	\$ 125.00	\$ 122.00	\$ 296.00	\$ 978.00	\$ 233.00	\$ 720.00	\$ 224.00	\$ 660.00	\$ 218.00	\$ 5.266.00	
2013	\$ 153.00	\$ 109.00	\$ 495.00	\$ 941.00	\$ 103.00	\$ 609.00	\$ 353.00	\$ 551.00	\$ 507.00	\$ 739.00	\$ 523.00	\$ 339.00	\$ 5.422.00	
2014	\$ 625.00	\$ 243.00	\$ 112.00	\$ 894.00	\$ 319.00	\$ 724.00	\$ 288.00	\$ 476.00	\$ 673.00	\$ 753.00	\$ 336.00	\$ 591.00	\$ 6.034.00	
2015	\$ 239.00	\$ 472.00	\$ 392.00	\$ 784.00	\$ 546.00	\$ 770.00	\$ 775.00	\$ 182.00	\$ 937.00	\$ 295.00	\$ 695.00	\$ 139.00	\$ 6.226.00	
2016	\$ 901.00	\$ 495.00	\$ 716.00	\$ 533.00	\$ 327.00	\$ 382.00	\$ 384.00	\$ 847.00	\$ 161.00	\$ 849.00	\$ 730.00	\$ 211.00	\$ 6.538.00	
2017	\$ 311.00	\$ 669.00	\$ 608.00	\$ 981.00	\$ 995.00	\$ 111.00	\$ 446.00	\$ 524.00	\$ 386.00	\$ 338.00	\$ 455.00	\$ 796.00	\$ 6.600.00	
2018	\$ 471.00	\$ 635.00	\$ 337.00	\$ 807.00	\$ 407.00	\$ 517.00	\$ 556.00	\$ 774.00	\$ 984.00	\$ 333.00	\$ 796.00	\$ 501.00	\$ 7.118.00	
2019	\$ 750.00	\$ 680.00	\$ 775.00	\$ 706.00	\$ 335.00	\$ 322.00	\$ 499.00	\$ 702.00	\$ 867.00	\$ 861.00	\$ 464.00	\$ 536.00	\$ 7.407.00	
2020	\$ 286.00	\$ 719.00	\$ 513.00	\$ 718.00	\$ 929.00	\$ 510.00	\$ 964.00	\$ 878.00	\$ 104.00	\$ 783.00	\$ 581.00	\$ 504.00	\$ 7.579.00	
Trung bình tháng	\$ 481.10	\$ 492.10	\$ 433.40	\$ 663.30	\$ 427.90	\$ 451.00	\$ 583.70	\$ 574.70	\$ 593.40	\$ 566.40	\$ 598.00	\$ 470.80		

Định dạng có điều kiện

The screenshot shows a Google Sheets interface with a data table and a conditional formatting rules panel.

Data Table:

Doanh thu	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12	Doanh thu năm	Tăng trưởng
2011	\$ 298.00	\$ 186.00	\$ 186.00	\$ 164.00	\$ 196.00	\$ 269.00	\$ 384.00	\$ 580.00	\$ 595.00	\$ 489.00	\$ 740.00	\$ 783.00	\$ 4,870.00	
2012	\$ 777.00	\$ 713.00	\$ 200.00	\$ 125.00	\$ 122.00	\$ 298.00	\$ 978.00	\$ 233.00	\$ 720.00	\$ 224.00	\$ 660.00	\$ 218.00	\$ 5,268.00	
2013	\$ 153.00	\$ 109.00	\$ 495.00	\$ 941.00	\$ 103.00	\$ 609.00	\$ 353.00	\$ 551.00	\$ 507.00	\$ 739.00	\$ 523.00	\$ 339.00	\$ 5,422.00	
2014	\$ 625.00	\$ 243.00	\$ 112.00	\$ 894.00	\$ 319.00	\$ 724.00	\$ 288.00	\$ 476.00	\$ 673.00	\$ 753.00	\$ 336.00	\$ 591.00	\$ 6,034.00	
2015	\$ 239.00	\$ 472.00	\$ 392.00	\$ 784.00	\$ 546.00	\$ 770.00	\$ 775.00	\$ 182.00	\$ 937.00	\$ 295.00	\$ 695.00	\$ 139.00	\$ 6,226.00	
2016	\$ 901.00	\$ 495.00	\$ 716.00	\$ 533.00	\$ 327.00	\$ 382.00	\$ 384.00	\$ 847.00	\$ 181.00	\$ 849.00	\$ 730.00	\$ 211.00	\$ 6,538.00	
2017	\$ 311.00	\$ 669.00	\$ 608.00	\$ 961.00	\$ 995.00	\$ 111.00	\$ 446.00	\$ 524.00	\$ 386.00	\$ 338.00	\$ 455.00	\$ 796.00	\$ 6,600.00	
2018	\$ 471.00	\$ 635.00	\$ 337.00	\$ 807.00	\$ 407.00	\$ 517.00	\$ 556.00	\$ 774.00	\$ 984.00	\$ 333.00	\$ 796.00	\$ 501.00	\$ 7,118.00	
2019	\$ 750.00	\$ 680.00	\$ 775.00	\$ 706.00	\$ 335.00	\$ 322.00	\$ 409.00	\$ 702.00	\$ 867.00	\$ 861.00	\$ 464.00	\$ 536.00	\$ 7,407.00	
2020	\$ 286.00	\$ 719.00	\$ 513.00	\$ 718.00	\$ 929.00	\$ 510.00	\$ 964.00	\$ 878.00	\$ 104.00	\$ 783.00	\$ 581.00	\$ 594.00	\$ 7,579.00	
Tổng bình tháng	\$ 481.10	\$ 492.10	\$ 433.40	\$ 663.30	\$ 427.90	\$ 451.00	\$ 553.70	\$ 574.70	\$ 593.40	\$ 566.40	\$ 598.00	\$ 470.80		

Conditional Format Rules Panel:

- Color scale** is selected.
- Apply to range:** B30:M30
- Format rules:** Default (green gradient).
- Minpoint:** None.
- Midpoint:** None.
- Maxpoint:** None.
- Buttons:** Cancel, Done.

Định dạng có điều kiện

Doanh thu	Tháng 1	Tháng 2
2011	\$ 298.00	\$ 186.00
2012	\$ 777.00	\$ 713.00
2013	\$ 153.00	\$ 109.00
2014	\$ 625.00	\$ 243.00
2015	\$ 239.00	\$ 472.00
2016	\$ 901.00	\$ 495.00
2017	\$ 311.00	\$ 669.00
2018	\$ 471.00	\$ 635.00
2019	\$ 750.00	\$ 680.00
2020	\$ 286.00	\$ 719.00
Trung bình tháng	\$ 481.10	\$ 492.10
	\$ 153.00 ×	
Doanh thu thấp nhất tháng	=MIN(B20:B29)	
Doanh thu cao nhất tháng		

Doanh thu	Tháng 1	Tháng 2
2011	\$ 298.00	\$ 186.00
2012	\$ 777.00	\$ 713.00
2013	\$ 153.00	\$ 109.00
2014	\$ 625.00	\$ 243.00
2015	\$ 239.00	\$ 472.00
2016	\$ 901.00	\$ 495.00
2017	\$ 311.00	\$ 669.00
2018	\$ 471.00	\$ 635.00
2019	\$ 750.00	\$ 680.00
2020	\$ 286.00	\$ 719.00
Trung bình tháng	\$ 481.10	\$ 492.10
	\$ 901.00 ×	
Doanh thu thấp nhất tháng	=MIN(B20:B29)	
Doanh thu cao nhất tháng	=MAX(B20:B29)	

Hàm COUNTIF và SUMIF

- Hàm **COUNTIF** trả về số lượng ô khớp với một giá trị xác định.
- Công thức: = COUNTIF(dãy_ô, “giá trị”)
- Hàm **SUMIF**: cộng các ô dựa trên điều kiện.
- Công thức: = SUMIF(dãy_ô, điều_kiện, [Dãy_tổng])



Ví dụ

Một số câu hỏi:

- Bao nhiêu giao dịch chỉ chứa 1 món hàng?
- Bao nhiêu giao dịch chứa nhiều hơn 1 món hàng ?
- Tổng doanh thu của các giao dịch chỉ chứa 1 món hàng ?
- Tổng doanh thu của các giao dịch chứa nhiều hơn 1 món hàng ?
- Doanh thu trung bình của các giao dịch chỉ chứa 1 món hàng ?
- Doanh thu trung bình các giao dịch chứa nhiều hơn 1 món hàng?

Mã giao dịch	Số lượng	Doanh thu
121	1	\$ 400.00
122	1	\$ 278.00
123	1	\$ 145.00
124	1	\$ 151.00
125	3	\$ 275.00
126	2	\$ 144.00
127	1	\$ 374.00
128	1	\$ 289.00
129	1	\$ 175.00
130	3	\$ 310.00
131	2	\$ 167.00
132	4	\$ 357.00
133	1	\$ 133.00
134	1	\$ 205.00
135	2	\$ 154.00
136	5	\$ 186.00
137	1	\$ 108.00
138	1	\$ 104.00
139	1	\$ 359.00
140	1	\$ 272.00
141	2	\$ 117.00
142	1	\$ 260.00
143	3	\$ 276.00
144	5	\$ 196.00

Bảng tóm tắt

Bảng tóm tắt: chứa các thông tin thống kê về dữ liệu

Số lượng	Đếm	Tổng doanh thu	Doanh thu TB
1			
>1			

Bảng tóm tắt

Mã giao dịch	Số lượng	Doanh thu
121	1	\$ 400.00
122	1	\$ 278.00
123	1	\$ 145.00
124	1	\$ 151.00
125	3	\$ 275.00
126	2	\$ 144.00
127	1	\$ 374.00
128	1	\$ 289.00
129	1	\$ 175.00
130	3	\$ 310.00
131	2	\$ 167.00
132	4	\$ 357.00
133	1	\$ 133.00
134	1	\$ 205.00
135	2	\$ 154.00
136	5	\$ 186.00
137	1	\$ 108.00
138	1	\$ 104.00
139	1	\$ 359.00
140	1	\$ 272.00
141	2	\$ 117.00
142	1	\$ 260.00
143	3	\$ 276.00
144	5	\$ 196.00

Số lượng	Tổng doanh thu	Doanh thu TB
1	=COUNTIF(B2:B25, "=1")	
>1	COUNTIF(range, criterion)	

Bảng tóm tắt

Mã giao dịch	Số lượng	Doanh thu				
121	1	\$ 400.00				
122	1	\$ 278.00				
123	1	\$ 145.00				
124	1	\$ 151.00				
125	3	\$ 275.00				
126	2	\$ 144.00				
127	1	\$ 374.00				
128	1	\$ 289.00				
129	1	\$ 175.00				
130	3	\$ 310.00				
131	2	\$ 167.00				
132	4	\$ 357.00				
133	1	\$ 133.00				
134	1	\$ 205.00				
135	2	\$ 154.00				
136	5	\$ 186.00				
137	1	\$ 108.00				
138	1	\$ 104.00				
139	1	\$ 359.00				
140	1	\$ 272.00				
141	2	\$ 117.00				
142	1	\$ 260.00				
143	3	\$ 276.00				
144	5	\$ 196.00				

Số lượng	Đếm	Tổng doanh thu	Doanh thu TB
1	10×14		
>1	=COUNTIF(B2:B25, ">1")		

COUNTIF(range, criterion)

Bảng tóm tắt

Mã giao dịch	Số lượng	Doanh thu		
121	1	\$ 400.00		
122	1	\$ 278.00		
123	1	\$ 145.00		
124	1	\$ 151.00		
125	3	\$ 275.00		
126	2	\$ 144.00		
127	1	\$ 374.00		
128	1	\$ 289.00		
129	1	\$ 175.00		
130	3	\$ 310.00		
131	2	\$ 167.00		
132	4	\$ 357.00		
133	1	\$ 133.00		
134	1	\$ 205.00		
135	2	\$ 154.00		
136	5	\$ 186.00		
137	1	\$ 108.00		
138	1	\$ 104.00		
139	1	\$ 359.00		
140	1	\$ 272.00		
141	2	\$ 117.00		
142	1	\$ 260.00		
143	3	\$ 276.00		
144	5	\$ 196.00		

Số lượng	Đếm	3253 ×	Doanh thu	Doanh thu TB
1	14	=SUMIF(B2:B25,"=1",C2:C25)		
>1	10			

Bảng tóm tắt

Mã giao dịch	Số lượng	Doanh thu
121	1	\$ 400.00
122	1	\$ 278.00
123	1	\$ 145.00
124	1	\$ 151.00
125	3	\$ 275.00
126	2	\$ 144.00
127	1	\$ 374.00
128	1	\$ 289.00
129	1	\$ 175.00
130	3	\$ 310.00
131	2	\$ 167.00
132	4	\$ 357.00
133	1	\$ 133.00
134	1	\$ 205.00
135	2	\$ 154.00
136	5	\$ 186.00
137	1	\$ 108.00
138	1	\$ 104.00
139	1	\$ 359.00
140	1	\$ 272.00
141	2	\$ 117.00
142	1	\$ 260.00
143	3	\$ 276.00
144	5	\$ 196.00

Số lượng	Đếm	Tổng doanh thu	Doanh thu TB
1	14	2182	3253
>1	10	=SUMIF(B2:B25, ">1", C2:C25)	

SUMIF(range, criterion, [sum_range])

Bảng tóm tắt

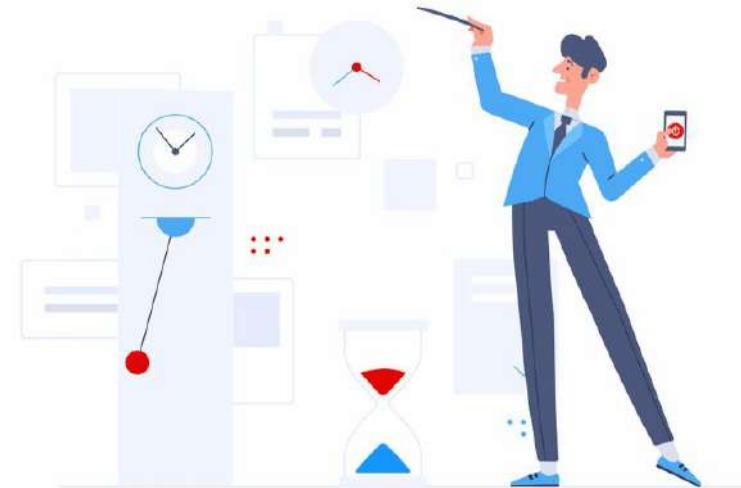
Số lượng	Đếm	Tổng doanh thu	232.3571429 ×
1	14	3253	=G8/F8
>1	10	2182	

Số lượng	Đếm	Tổng doanh thu	Doanh thu TB
1	14	3253	232.3571429
>1	10	2182	=G9/F9

Tab G9/F9 218.2

Hàm SUMPRODUCT

- **SUMPRODUCT** là một hàm nhân các mảng và trả về tổng của các tích đó.
- Công thức: =sumproduct(mảng 1, [mảng 2],...)
- **Mảng** là tập hợp giá trị trong các ô



Hàm SUMPRODUCT

Mã sản phẩm	Số lượng	Đơn giá	Biên lợi nhuận		
200	20	\$1.25	20%		Tổng doanh thu
201	30	\$5.00	25%		Tổng lợi nhuận
202	15	\$4.25	20%		
203	5	\$3.00	30%		
204	100	\$0.50	15%		

- Tổng doanh thu = số lượng 1 * Đơn giá 1 + số lượng 2 * Đơn giá 2 + ...
- Tổng lợi nhuận = số lượng 1 * Đơn giá 1 * Biên lợi nhuận 1 + số lượng 2 * Đơn giá 2 * Biên lợi nhuận 2 + ...

Hàm SUMPRODUCT

Mã sản phẩm	Số lượng	Đơn giá	Biên lợi nhuận
200	20	\$1.25	20%
201	30	\$5.00	25%
202	15	\$4.25	20%
203	5	\$3.00	30%
204	100	\$0.50	15%

303.75 ×
=SUMPRODUCT(B2:B6,C2:C6)

Tổng doanh thu
Tổng lợi nhuận

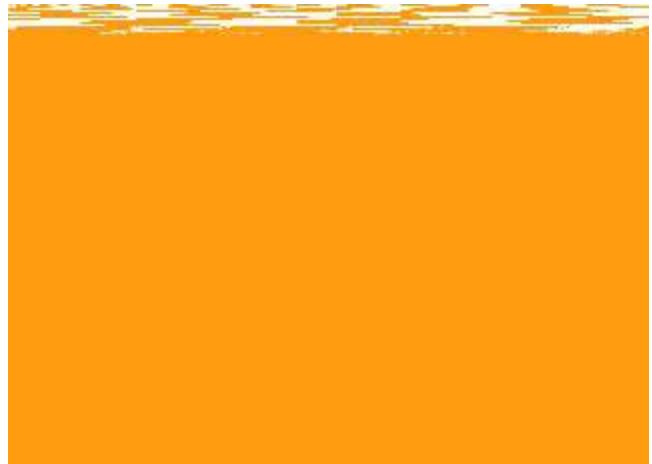
Mã sản phẩm	Số lượng	Đơn giá	Biên lợi nhuận
200	20	\$1.25	20%
201	30	\$5.00	25%
202	15	\$4.25	20%
203	5	\$3.00	30%
204	100	\$0.50	15%

67.25 × 303.75
=SUMPRODUCT(B2:B6,C2:C6,D2:D6)

Tổng doanh thu
Tổng lợi nhuận

Pivot table

- **Pivot table (bảng tổng hợp):** Bảng cho phép xem dữ liệu theo nhiều cách để tìm ra các thông tin hữu ích và xu hướng
- Bảng tổng hợp làm sạch và tổ chức dữ liệu: sắp xếp và nhóm dữ liệu
 - Tính tổng và trung bình của một cột



Pivot table

Movie Dataset

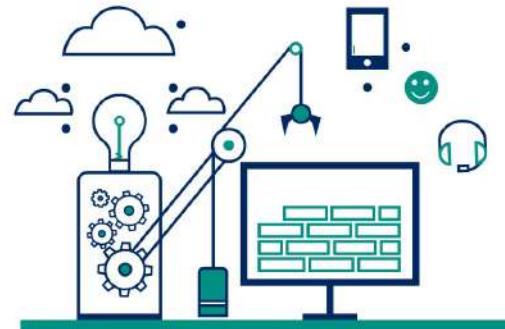
File Edit View Insert Format Data Tools Extensions Help Last edit was 2 minutes ago

P14

	A	B	I	J	K	L	M	N
1	Movie Title	Release Date	Cast (2)	Cast (3)	Cast (4)	Cast (5)	Budget (\$)	Box Office Revenue (\$)
2	The Devil Inside	2012-01-06	Fernanda Andrade	Simon Quarterman	Evan Helmuth		\$1,000,000.00	\$101,800,000.00
3	Red Tails	2012-01-20	Nate Parker	David Oyelowo	Daniela Ruah	Terrence Howard	\$58,000,000.00	\$50,400,000.00
4	Declaration of War	2012-01-27	(ator)				\$1,500,000.00	\$6,500,000.00
5	Man on a Ledge	2012-01-27	Ed Harris	Elizabeth Banks	Jamie Bell	Kyra Sedgwick	\$42,000,000.00	\$46,200,000.00
6	One for the Money	2012-01-27	Debbie Reynolds	Jason O'Mara	Daniel Sunjata	Sherri Shepherd	\$40,000,000.00	\$36,900,000.00
7	The Grey	2012-01-27	Frank Grillo	Dermot Mulroney	Dallas Roberts		\$25,000,000.00	\$77,300,000.00
8	Chronicle	2012-02-03	Michael B. Jordan	Alex Russell	Michael Kelly		\$12,000,000.00	\$126,600,000.00
9	The Woman in Black	2012-02-03	Ciarán Hinds	Janet McTeer	Liz White	Alisa Khezenova	\$15,000,000.00	\$128,500,000.00
10	Journey 2: The Mysterious Island	2012-02-10	Michael Caine	Josh Hutcherson	Vanessa Hudgens	Luis Guzmán	\$79,000,000.00	\$335,300,000.00
11	Safe House	2012-02-10	Ryan Reynolds	Sam Shepard	Robert Patrick	Liam Cunningham	\$85,000,000.00	\$208,100,000.00
12	Ghost Rider: Spirit of Vengeance	2012-02-17	Idris Elba	Fergus Riordan	Ciarán Hinds	Violante Placido	\$57,000,000.00	\$132,600,000.00
13	The Secret World of Arrietty	2012-02-17	Amy Poehler	Carol Burnett	Bridgit Mendler	David Henrie	\$23,000,000.00	\$145,600,000.00
14	This Means War	2012-02-17	Chris Pine	Reese Witherspoon	Chelsea Handler	Til Schweiger	\$65,000,000.00	\$156,500,000.00
15	Good Deeds	2012-02-24	Thandie Newton	Rebecca Romijn	Brian J. White	Jamie Kennedy	\$14,000,000.00	\$35,600,000.00
16	Wanderlust	2012-02-24	Jennifer Aniston	Malin Akerman	Justin Theroux	Ken Marino	\$35,000,000.00	\$24,200,000.00
17	Dr. Seuss' The Lorax	2012-03-02	Zac Efron	Taylor Swift	Ed Helms	Rob Riggle	\$70,000,000.00	\$348,800,000.00
18	Project X	2012-03-02	Oliver Cooper	Jonathan Daniel Brow Kirby Bliss Blanton		Thomas Mann	\$12,000,000.00	\$102,700,000.00
19	A Thousand Words	2012-03-09	Kerry Washington	Cliff Curtis	Clark Duke	Allison Janney	\$40,000,000.00	\$22,000,000.00
20	Friends with Kids	2012-03-09	Jennifer Westfeldt	Jon Hamm	Kristen Wiig	Maya Rudolph	\$10,000,000.00	\$12,200,000.00
21	Salmon Fishing in the Yemen	2012-03-09	Emily Blunt	Kristin Scott Thomas	Amr Waked		\$14,400,000.00	\$34,600,000.00
22	21 Jump Street	2012-03-16	Channing Tatum	Ice Cube	Brie Larson	Rob Riggle	\$55,000,000.00	\$201,500,000.00

Các bước phân tích

- Doanh thu hàng năm
- Xây dựng pivot table để hiển thị doanh thu hàng năm
- Doanh thu trung bình của phim trong từng năm
- Tìm ra một số xu hướng



Tạo Pivot Table

Insert -> Pivot Table

Create pivot table X

Data range
`'Movie Data'!A1:O509` [grid icon]

Insert to

New sheet Existing sheet

Cancel Create

Tạo Pivot Table

The screenshot shows the Microsoft Excel ribbon at the top with the 'Data' tab selected. Below the ribbon, a PivotTable is being created on a worksheet titled 'Movie Data'. The PivotTable structure is visible with columns A through J labeled across the top and rows 1 through 509 labeled down the left side. The 'Columns' button is highlighted in the PivotTable Fields pane, which is located on the right side of the screen. This pane also displays a 'Suggested' section with three items: 'Count of Movie Title for each Genre (2)', 'Average of Budget (\$) for each Genre (1)', and 'Number of unique Cast (1) for each Genre (2)'. Each suggested item has an 'Add' button next to it. The overall interface is clean and modern, typical of a recent version of Excel.

Tạo Pivot Table

The screenshot shows a Microsoft Excel interface with a sidebar on the left containing filter categories: Suggested, Rows, Columns, Values, and Filters. The main area displays two tables and a context menu.

Table 1: Movie Title

Movie Title
Release Date
Wikipedia URL
Genre (1)
Genre (2)
Director (1)
Director (2)
Cast (1)
Cast (2)
Cast (3)
Cast (4)
Cast (5)
Budget (\$)
Box Office Revenue (\$)
Column O

Table 2: Release Date

Release Date
2012-01-06
2012-01-20
2012-01-27
2012-02-03
2012-02-10
2012-02-17
2012-02-24
2012-03-02
2012-03-09
2012-03-16
2012-03-23
2012-03-30
2012-04-06
2012-04-13
2012-04-20
2012-04-27
2012-05-04
2012-05-11
2012-05-16
2012-05-18
2012-05-25
2012-06-01
2012-06-08

Context Menu (Open at 2012-02-10):

- Cut
- Copy
- Paste
- Paste special
- + Insert 1 column left
- Create pivot date group
 - Day of the week
 - Day of the year
 - Day of the month
 - Day-Month
 - Month
 - Quarter
 - Year
 - Year-Month
 - Year-Quarter
 - Year-Month-Day
- Get link to this cell
- Comment
- Insert note
- Convert to people chip

Table 3: Release Date - Year

Release Date - Year
2012
2013
2014
2015
2016
Grand Total

Tạo Pivot Table

The screenshot shows a Pivot Table interface with the following data:

Year	Box Office Revenue (\$)
2012	\$18,078,040,000.00
2013	\$13,672,800,000.00
2014	\$20,013,420,000.00
2015	\$13,521,310,000.00
2016	\$11,921,900,000.00
Grand Total	\$77,207,470,000.00

The interface includes various filters and settings on the left and right sides.

Tạo Pivot Table

The image shows a screenshot of a Pivot Table configuration dialog in Google Sheets. The dialog is open over a table of movie data. On the left, there's a sidebar with various aggregation functions: SUM, COUNTA, COUNT, COUNTUNIQUE, AVERAGE, MAX, MIN, MEDIAN, PRODUCT, STDEV, STDEVP, VAR, and VARP. Each function has an 'Add' button to its right. In the center, there's a 'Sort by' dropdown menu set to 'Release Date - ...'. At the bottom, there are 'Show as' dropdowns set to 'Default' and a 'Filters' section with an 'Add' button.

Tạo Pivot Table

Tương tự: chọn Add (Values) -> Box Office Revenue -> Chuyển sang AVERAGE

The screenshot shows a Pivot Table interface with the following data:

Release Date	SUM of Box Office Revenue (\$)	AVERAGE of Box Office Revenue (\$)
2012	\$18,078,040,000.00	\$170,547,547.17
2013	\$13,672,800,000.00	\$160,856,470.59
2014	\$20,013,420,000.00	\$168,180,000.00
2015	\$13,521,310,000.00	\$109,042,822.58
2016	\$11,921,900,000.00	\$161,106,756.76
Grand Total	\$77,207,470,000.00	\$151,983,208.66

The interface includes several configuration panels on the right side:

- Order:** Ascending, Sort by Release Date.
- Show totals:** Checked.
- Columns:** Add button.
- Values as:** Columns, Add button.
 - Box Office Revenue (\$):** Summarize by SUM, Show as Default.
 - Box Office Revenue (\$):** Summarize by AVERAGE, Show as Default.
- Filters:** Add button.

Tạo Pivot Table

Release Date - \	SUM of Box Office Revenue (\$)	AVERAGE of Box Office Revenue	COUNT of Box Off
2012	\$18,078,040,000.00	\$170,547,547.17	106
2013	\$13,672,800,000.00	\$160,856,470.59	85
2014	\$20,013,420,000.00	\$168,180,000.00	119
2015	\$13,521,310,000.00	\$109,042,822.58	124
2016	\$11,921,900,000.00	\$161,106,756.76	74
Grand Total	\$77,207,470,000.00	\$151,983,208.66	508

Một số nhận xét:

- Năm 2015 có tổng doanh thu xếp thứ 2 từ dưới đếm lên mặc dù có rất nhiều bộ phim
- Phim năm 2015 không có chất lượng như các năm khác?

Tạo Pivot Table

Release Date -	SUM of Box Office Revenue	AVERAGE of Box Office	COUNT of Box Offi
2012	\$18,078,040,000.00	\$170,547,547.17	106
2013	\$13,672,800,000.00	\$160,856,470.59	85
2014	\$20,013,420,000.00	\$168,180,000.00	119
2015	\$13,521,310,000.00	\$109,042,822.58	124
2016	\$11,921,900,000.00	\$161,106,756.76	74
Grand Total	\$77,207,470,000.00	\$151,983,208.66	508

Release Date -	SUM <\$10M	AVERAGE < \$10M	COUNT <\$10M
2012	\$18,078,040,000.00	\$170,547,547.17	106
2013	\$13,672,800,000.00	\$160,856,470.59	85
2014	\$20,013,420,000.00	\$168,180,000.00	119
2015	\$13,521,310,000.00	\$109,042,822.58	124
2016	\$11,921,900,000.00	\$161,106,756.76	74
Grand Total	\$77,207,470,000.00	\$151,983,208.66	508

Tạo Pivot Table

Pivot table editor

Show totals

Columns

Values as: Columns

Box Office Revenue (\$)(SUM)

Summarize by: SUM Show as: Default

Box Office Revenue (\$)(AVERAGE)

Summarize by: AVERAGE Show as: Default

Box Office Revenue (\$)(COUNT)

Summarize by: COUNT Show as: Default

Filters

Movie Title
Release Date
Wikipedia URL
Genre (1)
Genre (2)
Director (1)
Director (2)
Cast (1)
Cast (2)
Cast (3)
Cast (4)
Cast (5)
Budget (\$)
Box Office Revenue (\$)
Column O

Values as: Columns

Box Office Revenue (\$)(SUM)

Summarize by: SUM Show as: Default

Box Office Revenue (\$)(AVERAGE)

Summarize by: AVERAGE Show as: Default

Box Office Revenue (\$)(COUNT)

Summarize by: COUNT Show as: Default

Filters

Box Office Revenue (\$)

Status: Showing all items

Tạo Pivot Table

- ▶ Filter by condition
- ▼ Filter by values

Select all - Clear

🔍

- ✓ \$1,000,000.00
- ✓ \$1,100,000.00
- ✓ \$1,600,000.00
- ✓ \$1,700,000.00

CancelOK

Values as: Columns Add

Box Office Revenue (\$)(SUM <)

Summarize by Show as

SUM Default

Filter by condition

Less than

10000000

Filter by values

Select all - Clear

✓ \$1,000,000.00

✓ \$1,100,000.00

✓ \$1,600,000.00

✓ \$1,700,000.00

CancelOK

Tạo Pivot Table

Release Date - \ SUM of Box Office Revenue AVERAGE of Box Office COUNT of Box Office

Year	Sum of Box Office Revenue	Average of Box Office	Count of Box Office
2012	\$18,078,040,000.00	\$170,547,547.17	106
2013	\$13,672,800,000.00	\$160,856,470.59	85
2014	\$20,013,420,000.00	\$168,180,000.00	119
2015	\$13,521,310,000.00	\$109,042,822.58	124
2016	\$11,921,900,000.00	\$161,106,756.76	74
Grand Total	\$77,207,470,000.00	\$151,983,208.66	508

Release Date - \ SUM <\$10M AVERAGE < \$10M COUNT <\$10M

Year	Sum <\$10M	Average < \$10M	Count <\$10M
2012	\$67,940,000.00	\$6,794,000.00	10
2013	\$26,400,000.00	\$3,300,000.00	8
2014	\$46,920,000.00	\$4,265,454.55	11
2015	\$93,040,000.00	\$4,652,000.00	20
2016	\$41,100,000.00	\$5,137,500.00	8
Grand Total	\$275,400,000.00	\$4,831,578.95	57

Values as: Columns Add

Box Office Revenue (\$) (SUM < X)
Summarize by Show as
SUM Default

Box Office Revenue (\$) (AVERA X)
Summarize by Show as
AVERAGE Default

Box Office Revenue (\$) (COUNT X)
Summarize by Show as
COUNT Default

Filters Add

Box Office Revenue (\$)
Status
Value is less than 10000000

Tạo Pivot Table

Movie Data!A1:O509

Suggested Rows Columns Values as:

Release Date - Year

Order Ascending Sort by Release Date

Show totals

Genre (1)

Director (1)

Director (2)

Cast (1)

Cast (2)

Cast (3)

Cast (4)

Cast (5)

Budget (\$)

Box Office Revenue (\$)

Column O

AVERAGE of Box Office F COUNT of Box Offi

	AVERAGE	COUNT	Calculated Field
Movie Title	\$170,547,547.17	106	
Movie Title	\$160,856,470.59	85	
Movie Title	\$168,180,000.00	119	
Movie Title	\$109,042,822.58	124	
Movie Title	\$161,106,756.76	74	
	\$151,983,208.66	508	
	AVERAGE < \$10M	COUNT <\$10M	Calculated Field
Genre (2)	\$6,794,000.00	10	\$6,794,000.00
Director (1)	\$3,300,000.00	8	\$3,300,000.00
Director (2)	\$4,265,454.55	11	\$4,265,454.55
Cast (1)	\$4,652,000.00	20	\$4,652,000.00
Cast (2)	\$5,137,500.00	8	\$5,137,500.00
	\$4,831,578.95	57	4831578.947

Box Office Revenue (\$) (SUM <)

Summarize by SUM Show as Default

Box Office Revenue (\$) (AVERA <)

Summarize by AVERAGE Show as Default

Box Office Revenue (\$) (COUN <)

Summarize by COUNT Show as Default

Calculated Field 1

Formula: =SUM(Box Office Revenue (\$))/COUNT(Box

Summarize by Custom Show as Default

Tạo Pivot Table

COUNT <\$10M	Calculated Field
10	\$6,794,000.00
8	\$3,300,000.00
11	\$4,265,454.55
20	\$4,652,000.00
8	\$5,137,500.00
57	4831578.947

0.09433962264 × ie
=D10/D2

COUNT <\$10M	Calculated Field	% of total movie
10	\$6,794,000.00	9.43%
8	\$3,300,000.00	9.41%
11	\$4,265,454.55	9.24%
20	\$4,652,000.00	16.13%
8	\$5,137,500.00	10.81%
57	4831578.947	

Tính toán với SQL

- **Toán tử (operator)** là một ký hiệu cho hoạt động hoặc phép tính sẽ được thực hiện trong một công thức
- Bốn toán tử cơ bản: + - * /



Cộng hai cột với SQL

```
SELECT  
    columnA,  
    columnB,  
    columnA + columnB AS columnX  
FROM  
    table_name
```

<u>Column A</u>	<u>Column B</u>	<u>Column X</u>
A1	B1	Sum of A1 + B1
A2	B2	Sum of A2 + B2
A3	B3	Sum of A3 + B3

Cộng hai cột với SQL

```
SELECT  
    columnA,  
    columnB,  
    columnC,  
    (columnA + columnB) * columnC AS columnX  
FROM  
    table_name
```

<u>Column A</u>	<u>Column B</u>	<u>Column C</u>	<u>Column X</u>
A1	B1	C1	(A1 + B1)*C1
A2	B2	C2	(A2 + B2)*C2
A3	B3	C3	(A3 + B3)*C3

Modulo và Hàm

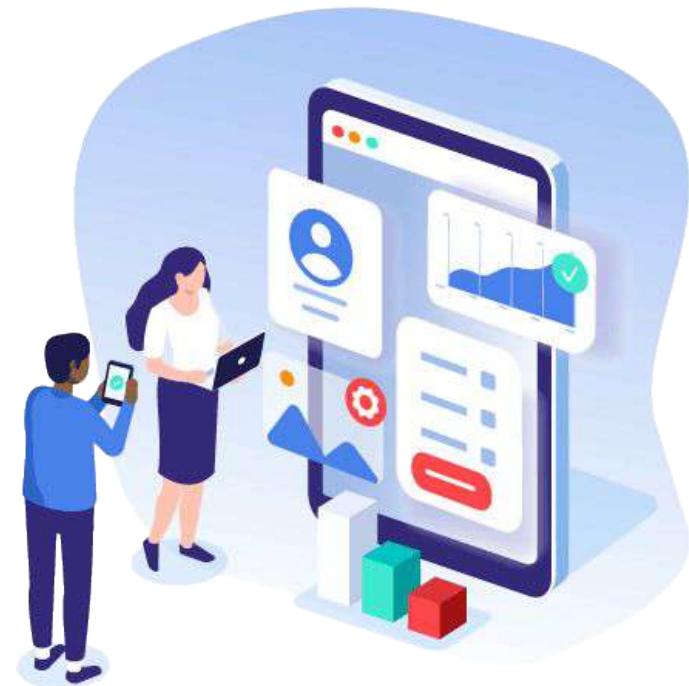
Toán tử modulo (%): trả về phần dư khi thực hiện phép chia.

Hàm trong bảng tính:

- SUM
- AVERAGE

Hàm trong SQL:

- SUM
- AVG



GROUP BY

GROUP BY: lệnh nhóm các hàng có cùng giá trị từ một bảng thành các hàng tóm tắt

SELECT FROM WHERE GROUP BY



GROUP BY

Q new york X ?

citibike_trips

QUERY SHARE COPY SNAPSHOT DELETE EXPORT

Found 5 results. [Narrow search to pinned projects.](#)

SCHEMA DETAILS PREVIEW

Row	tripduration	starttime	stoptime	start_station_id	start_station_name	start_station_latitude	start_station_longitude	end_stati
1	432	2013-09-16T19:22:43	2013-09-16T19:29:55	509	9 Ave & W 22 St	40.7454973	-74.00197139	442
2	1186	2015-12-30T13:02:38	2015-12-30T13:22:25	280	E 10 St & 5 Ave	40.73331967	-73.99510132	254
3	799	2017-09-02T16:27:37	2017-09-02T16:40:57	335	Washington Pl & Broadway	40.72903917	-73.99404649	540
4	238	2017-11-15T06:57:09	2017-11-15T07:01:08	146	Hudson St & Reade St	40.71625008	-74.0091059	387
5	668	2013-11-07T15:12:07	2013-11-07T15:23:15	529	W 42 St & 8 Ave	40.7575699	-73.99098507	352
6	593	2013-08-25T13:47:24	2013-08-25T13:57:17	470	W 20 St & 8 Ave	40.74345335	-74.00004031	252
7	414	2018-05-29T16:33:26.488000	2018-05-29T16:40:21.206000	3158	W 63 St & Broadway	40.77163851	-73.98261428	3167
8	1643	2014-02-07T20:24:02	2014-02-07T20:51:25	519	Pershing Square N	40.75188406	-73.97770164	147
9	474	2017-10-06T12:17:06	2017-10-06T12:25:00	470	W 20 St & 8 Ave	40.74345335	-74.00004031	496
10	2277	2014-09-11T05:46:58	2014-09-11T06:24:55	487	E 20 St & FDR Drive	40.73314259	-73.97573881	487
11	614	2018-05-10T08:26:57.182000	2018-05-10T08:13:12.168000	291	Madison St & Montgomery St	40.713126	-73.984844	3489
12	1122	2017-08-16T07:40:27	2017-08-16T07:59:10	3260	Mercer St & Bleeker St	40.7270633483063	-73.996621370315538	518
13	838	2014-11-03T14:25:11	2014-11-03T14:39:09	457	Broadway & W 58 St	40.76695317	-73.98169333	318
14	318	2017-09-22T06:44:27	2017-09-22T06:49:45	3255	8 Ave & W 31 St	40.7505853470215	-73.9946848154068	458

GROUP BY

```
SELECT
    EXTRACT(Year FROM starttime) as year,
    COUNT(*) as number_of_rides
FROM `bigquery-public-data.new_york_citibike.citibike_trips`
GROUP BY
    year
ORDER BY
    year
```

GROUP BY

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	year	number_of_rides		
1	<i>null</i>	5828994		
2	2013	5037185		
3	2014	8081216		
4	2015	9937969		
5	2016	10262649		
6	2017	14118669		
7	2018	5671033		

GROUP BY

```
SELECT
    EXTRACT(Year FROM STARTTIME) as year,
    COUNT(*) as number_of_rides
FROM `bigquery-public-data.new_york_citibike.citibike_trips`
GROUP BY
    year
ORDER BY
    year DESC
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	year	number_of_rides		
1	2018	5671033		
2	2017	14118669		
3	2016	10262649		
4	2015	9937969		
5	2014	8081216		
6	2013	5037185		
7	null	5828994		

Quá trình xác thực dữ liệu

Quá trình xác thực dữ liệu: kiểm tra và kiểm tra lại chất lượng dữ liệu của bạn để bảo đảm tính *hoàn chỉnh, chính xác, an toàn và nhất quán*

```
SELECT  
    Date,  
    Region,  
    Small_Bags,  
    Large_Bags,  
    XLarge_Bags,  
    Total_Bags,  
    Small_Bags + Large_Bags + XLarge_Bags AS Total_Bags_Calc  
FROM avocado_data.avocado_prices
```

Total_Bags	Total_Bags_Calc
8696.87	8696.87
66354.31	66354.31
144911.1	144911.1
23561.1	23561.1
86012.86	86012.86
55236.68	55236.68

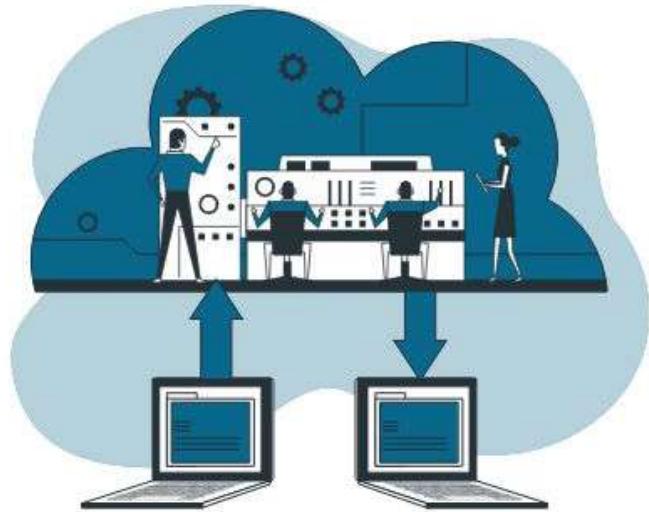
Bảng tạm

Bảng tạm: bảng cơ sở dữ liệu được tạo và tồn tại tạm thời trên máy chủ cơ sở dữ liệu

Tự động bị xóa khi **kết thúc phiên** (session) cơ sở dữ liệu SQL của mình

Cách tạo bảng tạm:

- Mệnh đề **WITH**
- Câu lệnh **SELECT INTO**
- Câu lệnh **CREATE TABLE**



Mệnh đề WITH

VD: bảng tạm *trips_over_1_hr* với từ khóa WITH

```
WITH trips_over_1_hr AS (
    SELECT *
    FROM
        bigquery-public-data.new_york.citibike_trips
    WHERE
        tripduration >= 60
)
```

Mệnh đề WITH

Dùng bảng tạm

```
SELECT  
    COUNT(*) AS cnt  
FROM  
    trips_over_1_hr
```

Bảng tạm với SELECT INTO

```
SELECT  
    *  
INTO  
    AfricaSales  
FROM  
    GlobalSales  
WHERE  
    Region = "Africa"
```

Tạo bảng lâu dài

Tạo bảng với CREATE TABLE. Cho phép nhiều người truy cập vào bảng mới

```
CREATE TABLE AfricaSales AS  
(  
    SELECT *  
    FROM GlobalSales  
    WHERE Region = "Africa"  
)
```



TỔNG KẾT



Tổng kết

- Trình bày được cách tổ chức dữ liệu để phân tích
- Mô tả được quy trình để định dạng và điều chỉnh dữ liệu
- Trình bày được cách để tổng hợp dữ liệu trong bảng tính và bằng cách sử dụng SQL
- Sử dụng được công thức và hàm trong bảng tính để tính toán dữ liệu
- Thực hiện được tính toán bằng cách sử dụng các câu truy vấn SQL



THANK YOU

