

Exercise

Pivot Tables in Google Sheets

Summary

1. Module: Understanding Data¹
2. Objective: Learn about summarising data using pivot tables in Google Sheets
3. Time Allotment: 2 hours

Steps

Conceptual Exercise

Let's start off by looking at the data table we prepared:

Participant's Name	Organisation	Type of Organisation (Media, CSO, Academic, Think Tank, Political)	Number of rows in your dataset	Number of columns in your dataset
Maung Than Chaung	MIMU	CSO	3000	10
Ma Thay Mhyin	7Day	Media	300	20
...

Let's say we want to find out how many participants were from CSOs. In the previous exercise, we looked at how Google Sheets' filter function can be used to filter just for CSO participants and then we use COUNT to calculate how many participants are from CSOs. Now let's say we want to

¹ This lesson was adapted from the World Bank's Introduction to Data Literacy training manual by Eva Constantaras, and adapted by Yan Naung Oak, Open Development Cambodia and Open Development Initiative, and is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/). For full terms of use, see [here](#).

repeat this for Media, Think Tanks and other types of organisations. If we filter for each type one at a time, it can get very tedious and time consuming.

Wouldn't it be nice if there was a to automatically calculate the number of participants from each type of organisation, so that we can get the results in a table like below:

Type of Organisation	Count of participants
Media	5
CSO	8
Think Tank	4

The answer is “Yes!”. We can use the pivot table function in Google Sheets and other spreadsheet software to get these kinds of useful summary calculations and much more.

For example, instead of counting the number of participants, we can automatically calculate the sum of rows of data in the datasets for each type of organisation:

Type of Organisation	Total number of rows
Media	5,000
CSO	30,000
Think Tank	9,000

Pivot tables are one of the most useful features for data analysis.

First attempt to practice using the Pivot Table with this [dataset of conference attendees](#).

- 1. Count of “participants” from each country**
- 2. Count of attendance by country in August 2019**



Main Exercise

Next, let's move on to doing some data analysis on a familiar dataset. In this exercise, we will be analysing data from the [Hydropower Dams in Cambodia dataset](#) from Open Development Cambodia. We have prepared a more ready to use version of this dataset and it has been uploaded to [Google Sheets here](#). **We will be looking at the questions below:**

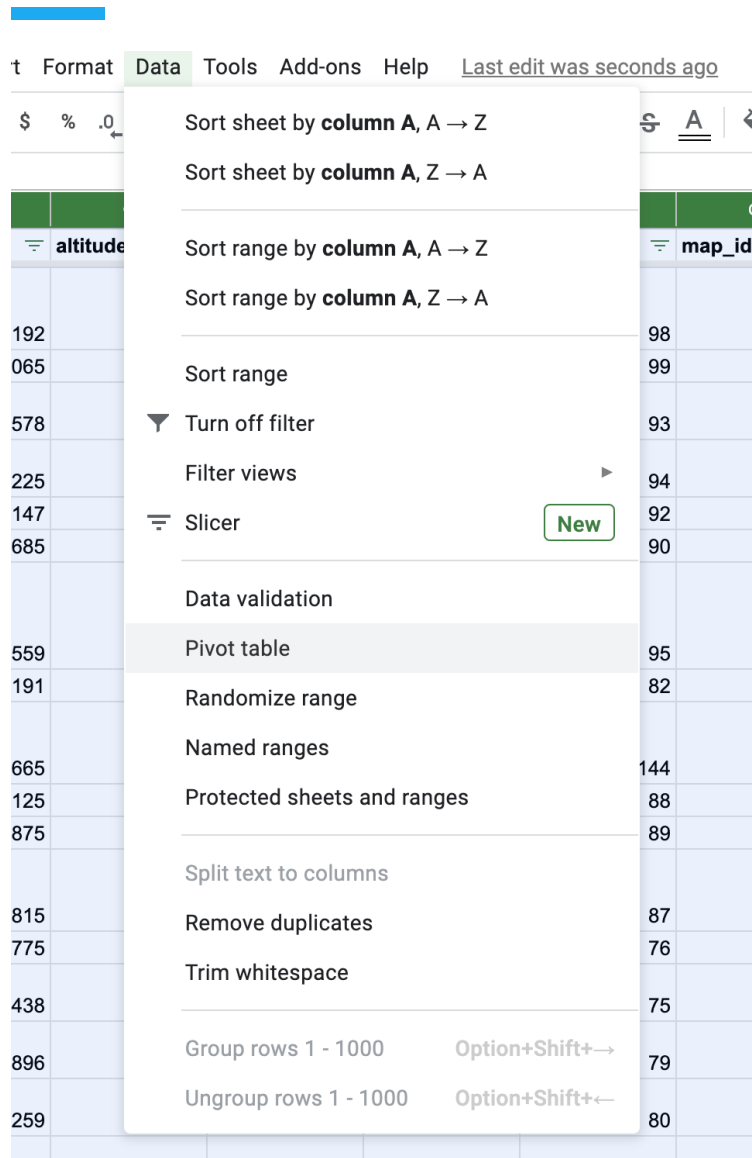
- 1. For each type of status (operational, potential site, under study, etc), calculate their total combined capacity.**
- 2. What is the total capacity for small scale projects?**
- 3. What is the median capacity for all the projects?**
- 4. What is the total capacity for large scale projects that are invested by Korea?**
- 5. What province has the highest number of potential sites?**
- 6. What province will have the largest total capacity if all the projects are completed?**
- 7. What is the total capacity of projects that are operational by 2013 in Koh Kong province?**

Steps to Follow

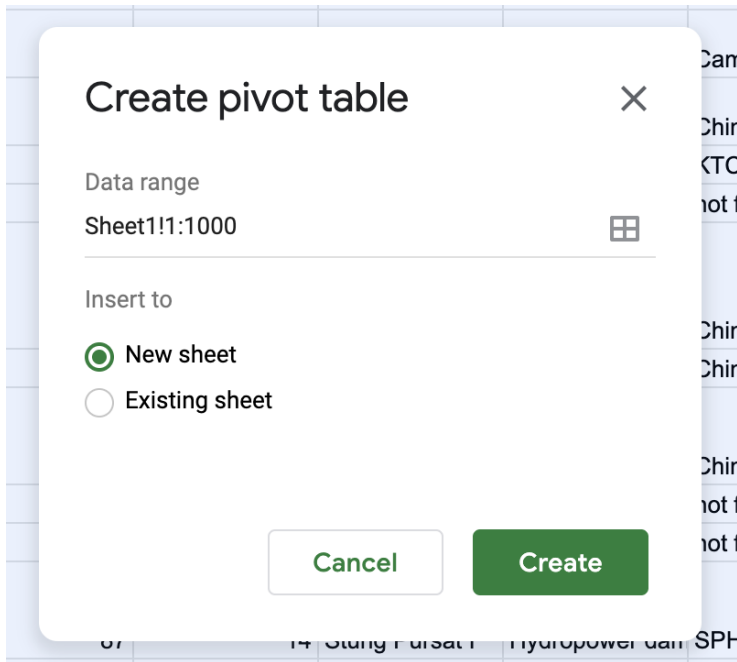
Click on the top left corner of the spreadsheet table in between column A and row 1. This will select the entire sheet.

	A	B	C	D	E	F	G	H	I	J	K
1	latitude	longitude	altitude	geometry	id	ogc_fid	map_id	name	purpose	investor	country
2	13.22406233	104.3677898		Point	Hydropower_dam	108		35 Stung Chikreng	Hydropower dam	Not found	Not found
3	14.02163799	106.8506775		Point	Hydropower_dam	76		3 Lower Sesan III	Hydropower dam	KTC (Korea)	Korea
4	11.99625418	103.4421819		Point	Hydropower_dam	91		18 Upper Stung Rus	Hydropower dam	KTC (Korea)	Korea
5	11.92681814	103.3244147		Point	Hydropower_dam	92		19 Middle Stung Ru	Hydropower dam	KTC (Korea)	Korea
6	13.22264503	104.7661404		Point	Hydropower_dam	107		34 Stung Staung	Hydropower dam	Not found	Not found
7	14.43170516	107.3565642		Point	Hydropower_dam	74		1 Prek Liang II	Hydropower dam	KTC (Korea)	Korea
8	14.31413064	107.2891438		Point	Hydropower_dam	75		2 Prek Liang I	Hydropower dam	KTC (Korea)	Korea
9	13.46665272	106.4005428		Point	Hydropower_dam	78		5 Lower Srepok II	Hydropower dam	China Huandian	Not found
10	13.3851599	107.0421896		Point	Hydropower_dam	79		6 Lower Srepok III	Hydropower dam	Huadian Corpore	China
11	13.33719024	107.4651259		Point	Hydropower_dam	80		7 Lower Srepok IV	Hydropower dam	Huadian Corpore	China
12	13.57526937	105.9848048		Point	Hydropower_dam	81		8 Stung Treng	Hydropower dam	Open Joint Stock Viet nam	
13	12.76849795	105.947191		Point	Hydropower_dam	82		9 Sambor	Hydropower dam	China Southern I	China
14	12.11342973	106.2590354		Point	Hydropower_dam	83		10 Prek Chhlong II	Hydropower dam	Not found	Singapore
15	13.37570371	105.2453209		Point	Hydropower_dam	84		11 Stung Sen	Hydropower dam	Royal Group (Ca	Cambodia
16	12.81697702	102.9044498		Point	Hydropower_dam	85		12 Stung Battambar	Hydropower dam	KTC (Korea)	Korea
17	12.4424646	102.9086224		Point	Hydropower_dam	86		13 Stung Battambar	Hydropower dam	KTC (Korea)	Korea
18	12.20898554	102.7650125		Point	Hydropower_dam	88		15 Stung Meteuk I	Hydropower dam	not found	Not found
19	12.02768875	102.863875		Point	Hydropower_dam	89		16 Stung Meteuk II	Hydropower dam	not found	Not found
20	11.84261578	102.8912685		Point	Hydropower_dam	90		17 Stung Meteuk III	Hydropower dam	not found	Not found

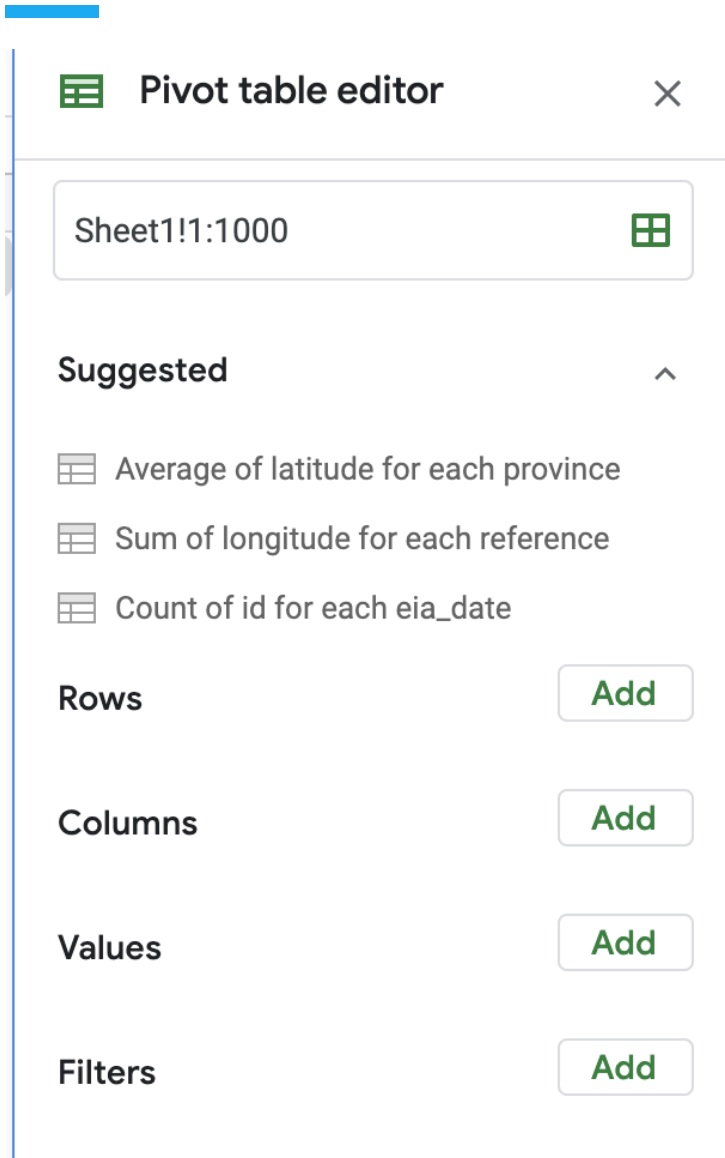
Now, select Data -> Pivot table from the top menu bar.



Next click on “Create” in the pop-up that appears to create a new pivot table in a separate sheet.



On the new pivot table sheet, there is a panel on the right that lets you choose rows and columns and filters to summarize by.



The first question asks “For each type of status (operational, potential site, under study, etc), calculate their total combined capacity.”

In order to answer this, we want to summarise the data according to the “status” column. In the “Rows” section of the panel on the right, click on “Add”. Choose “status”, and you will see that the row labels in the pivot table will be populated with the various values of the status column.

The image shows a spreadsheet with a pivot table and its corresponding editor. The pivot table is located in the range A1:C6. The data is as follows:

	A	B	C
1	status		
2			
3	Operational		
4	Potential site		
5	Under study		
6	Grand Total		

The Pivot table editor on the right shows the following settings:

- Source: Sheet1!1:1000
- Suggested: (dropdown)
- Rows: Add
- Columns: Add
- Values: Add
- Filters: Add

The 'status' field in the Rows section is highlighted with a blue box. Its settings are shown in a sub-dialog:

- Order: Ascending
- Sort by: status
- Show totals:

Next, in the values section, click “Add” and choose “capacity”. By default, the “summarize by” will be SUM, which means for each cell in the pivot table, it will sum the capacity values of all the rows in the original data associated with the appropriate value for “status”, as shown below.

The screenshot shows a spreadsheet with a pivot table and its configuration in the Pivot table editor. The pivot table is located in the range B1:C6 and has the following data:

	status	SUM of capacity
1		
2		708
3	Operational	1449.1
4	Potential site	1134
5	Under study	5391
6	Grand Total	8682.1

The Pivot table editor is open on the right side of the screen. It shows the following configuration:

- Source:** Sheet1!1:1000
- Rows:** status
- Columns:** (empty)
- Values:** capacity
- Filters:** (empty)

The configuration for the 'status' row field is as follows:

- Order:** Ascending
- Sort by:** status
- Show totals:**

The configuration for the 'capacity' value field is as follows:

- Summarize by:** SUM
- Show as:** Default

The next question asks “What is the total capacity for small scale projects?”

Similar to the previous question, we go back to our dataset, create a new pivot table, and choose “size_scale” for rows, and “capacity” for values.

	A	B	C
1	size_scale	SUM of capacity	
2		1058	
3	Large scale	6931.1	
4	Medium scale	422	
5	Not found	120	
6	Small scale	151	
7	Grand Total	8682.1	
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			

Pivot table editor

Sheet1!1:1000

Suggested

Rows Add

size_scale

Order: Ascending | Sort by: size_scale

Show totals

Columns Add

Values Add

capacity

Summarize by: SUM | Show as: Default

Filters Add

The next question asks “The next question asks “What is the median capacity for all the projects?”

Here, we do not need to choose any rows in the Pivot table editor panel on the right. We can go straight to add Values. We choose “capacity” in the “Values” section and change the “Summarize by” to MEDIAN.

The screenshot shows a Google Sheets interface with a Pivot table editor. The spreadsheet data is as follows:

	A	B	C
1	MEDIAN of capacity		
2		15	
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

The Pivot table editor on the right is configured with the following settings:

- Source: Sheet1!1:1000
- Suggested:
 - Rows: Add
 - Columns: Add
 - Values: Add
- Expanded Values section:
 - Field: capacity
 - Summarize by: MEDIAN
 - Show as: Default
- Filters: Add

The next question asks “What is the total capacity for large scale projects that are invested by Korea?”

In this case, we want to choose both Rows and Columns to summarize with across two variables. Choose “size_scale” for Rows, “country” for columns, and “capacity” for Values.

	A	B	C	D	E	F	G	H	I
1	SUM of capacity	country							
2	size_scale		Cambodia	China	Korea	Not found	Singapore	Viet nam	Grand Total
3		350							350
4	Large scale			3984.1	564	1255		1128	6931.1
5	Medium scale		38	35	156	168	25		422
6	Not found			120					120
7	Small scale					151			151
8	Grand Total	350	38	4139.1	720	1574	25	1128	7974.1

The next question asks “What province has the highest number of potential sites?”

Here we choose “province” for Rows, and “capacity” for Values, and change the “Summarize by” to COUNTA

	A	B
1	province	COUNTA of id
2		0
3	Battambang	5
4	Kampong Speu	4
5	Kampong Thom	2
6	Kampot	3
7	Koh Kong	11
8	Kratie	8
9	Mondul Kiri	11
10	Oddar Meanchey	1
11	Preah Sihanouk	1
12	Preah Vihear	1
13	Pursat	8
14	Ratanak Kiri	10
15	Siem Reap	3
16	Stung Treng	5
17	Grand Total	73
18		
19		
20		
21		
22		
23		
24		
25		

Pivot table editor

Sheet1!1:1000

Suggested

Rows Add

province

Order: Ascending Sort by: province

Show totals

Columns Add

Values Add

id

Summarize by: COUNTA Show as: Default

Filters Add

The next question asks “What province will have the largest total capacity if all the projects are completed?”

Here we choose “province” for Rows, and “capacity” for Values, and change the “Summarize by” to SUM

The screenshot shows a spreadsheet with a pivot table and its corresponding Pivot Table Editor. The pivot table summarizes the capacity of projects by province. The editor shows the following configuration:

- Source:** Sheet1!1:1000
- Rows:** province
- Columns:** (None)
- Values:** capacity, Summarize by: SUM, Show as: Default
- Filters:** (None)
- Order:** Ascending
- Sort by:** province
- Show totals:** Checked

	A	B
1	province	SUM of capacity
2		0
3	Battambang	87
4	Kampong Speu	35
5	Kampong Thom	9
6	Kampot	223.1
7	Koh Kong	976
8	Kratie	2667
9	Mondul Kiri	104
10	Oddar Meanchey	7
11	Preah Sihanouk	2
12	Preah Vihear	38
13	Pursat	726
14	Ratanak Kiri	1253
15	Siem Reap	7
16	Stung Treng	1840
17	Grand Total	7974.1
18		
19		
20		
21		
22		
23		
24		
25		

The next question asks “What is the total capacity of projects that are operational by 2013 in Koh Kong province?”

Again we have to choose both Rows and Columns here to summarize with across two variables. Choose “province” for Rows, “year” for columns, and “capacity” for Values.

province	2012	2013	2015	2016	2017	2018	2019	Grand Total
Battambang	0				24		36	
Kampong Speu	27							
Kampong Thom	35							
Kampot	9							
Koh Kong	29	194.1						22
Kratie	109	18	584		108	32	125	
Mondul Kiri	67							2600
Oddar Meanchey	104							
Preah Sihanouk	7							
Preah Vihear	2			38				
Pursat					97			
Ratanak Kiri	629				128	375	330	1
Siem Reap	420							
Stung Treng	7							
Stung Treng	90				400		1350	1
Grand Total	1535	212.1	584	38	636	528	1805	2636

Further Practice

Now we will practice the same skills for a [dataset of private schools in Cambodia](#).

Do the following:

1. How many schools of each kind are there in each province?
2. Which provinces have schools with the highest median number of students?
3. Which provinces have schools with the highest median percentage of female students?
4. Which type of school has the highest median percentage of female staff?
5. Which province has schools with the lowest median student to staff ratio?

