EXPLORE AI ACADEMY

Data aggregations and descriptive statistics

Pivot tables

Data overview

> We are a team of analysts hired to analyse the resource requirements of Covid-19 patients on life support in Kenya in the last quarter of the year 2020.

We need...

Summary statistics to help us understand the distributions per month, per resource type, and per gender.

The dataset

COVID-19 patients on life support in Kenya in the **last quarter of 2020.**

The dataset is from **openAFRICA**.

Data aggregations and descriptive statistics

Data overview

The distribution of Covid-19 patients on life support dataset contains 126 rows and the following columns:

A. Date and Month
The date and month of the observation.

B. Supplementary Oxygen
The number of patients who required supplementary oxygen resources.

C. Ventilatory Support
The number of patients who required ventilatory support resources.

D. New Covid-19 cases
The number of new Covid-19 cases recorded on that observation day.

The dataset

	A	B	C	D	E
1	Date	Month	Supplementary Oxygen	Ventilatory Support	New Covid-19 cases
2	10/18/2020	October			616
3	10/19/2020	October	39		685
4	10/20/2020	October	41		195
5	10/21/2020	October	41	27	571
6	10/22/2020	October	45	27	497
7	10/23/2020	October	59		1068
8	10/24/2020	October	62		631
9	10/26/2020	October	52		931
10	10/29/2020	October	40	18	1018
11	10/30/2020	October	60	26	761
12	11/01/2020	November	68	31	1395
13	11/02/2020	November	83	26	685
14	11/03/2020	November	107	25	724
15	11/04/2020	November	122	25	492
16	11/05/2020	November	118	26	1494
17	11/06/2020	November	116	27	1008
18	11/08/2020	November	102	23	1065
19	11/09/2020	November	123	21	719
20	11/10/2020	November	117	22	756
21	11/11/2020	November	120	18	1344
22	11/12/2020	November	124	20	1216
23	11/14/2020	November	120	23	1470
24	11/15/2020	November	126	26	1080
25	11/16/2020	November	130	27	972
26	11/17/2020	November	124	26	559
27	11/18/2020	November	115	27	925

Data aggregations and descriptive statistics

Consider the questions we want to investigate

Distribution of **Covid-19 patients** on life support in **Kenya**

How many patients required supplementary oxygen **per month per gender group**?

How many patients required ventilatory support **per month per gender group**?

How many new Covid-19 cases were recorded **per month per gender group**?

What is a pivot table?

> A tool that transforms data in order to highlight comparisons, patterns, and trends.

01. Summarises data by pivoting **rows to columns or columns to rows.**

02. **Expands** and **collapses** levels of data to focus and drill down to important details.

03. Allows for easier **subtotalling and** aggregation of numeric data.

04. Creates **categories** and **subcategories** of the data to create custom formulas and calculations.

05. Focuses on useful and intriguing subsets of data by **filtering, sorting, grouping, and conditionally formatting** the data.

06. Allows for easier **querying of large amounts of data** in a more user-friendly way.

What is a pivot table?

Here is an example of a pivot table.

A	B	C	D	E
Date	Month	Supplementary Oxygen	Ventilatory Support	New Covid-19 cases
10/18/2020	October			616
10/19/2020	October	39		685
10/20/2020	October	41		195
10/21/2020	October	41	27	571
10/22/2020	October	45	27	497
10/23/2020	October	59		1068
10/24/2020	October	62		631
10/26/2020	October	52		931
10/29/2020	October	40	18	1018
10/30/2020	October	60	26	761
11/01/2020	November	68	31	1395
11/02/2020	November	83	26	685
11/03/2020	November	107	25	724
11/04/2020	November	122	25	492
11/05/2020	November	118	26	1494
11/06/2020	November	116	27	1008

A data table

Month	SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19 cases
December	1966	713	11687
November	2839	706	23916
October	439	98	6973
Grand Total	**5244**	**1517**	**42576**

A pivot table of the same data

What is a pivot table?

Example

The pivot table in the example provides a summary of the data for each month.

It adds up the values for the columns `Supplementary Oxygen`, `Ventilatory Support`, and `New Covid-19 cases` to determine the total for each resource category per month.

For columns of interest, we can determine various descriptive statistics, such as the number of entries, average, minimum, and maximum.

	A	B	C	D	E
1	Date	Month	Supplementary Oxygen	Ventilatory Support	New Covid-19 cases
2	10/18/2020	October			616
3	10/19/2020	October	39		685
4	10/20/2020	October	41		195
5	10/21/2020	October	41	27	571
6	10/22/2020	October	45	27	497
7	10/23/2020	October	59		1068
8	10/24/2020	October	62		631
9	10/26/2020	October	52		931
10	10/29/2020	October	40	18	1018
11	10/30/2020	October	60	26	761
12	11/01/2020	November	68	31	1395
13	11/02/2020	November	83	26	685
14	11/03/2020	November	107	25	724
15	11/04/2020	November	122	25	492
16	11/05/2020	November	118	26	1494
17	11/06/2020	November	116	27	1008
18	11/08/2020	November	102	23	1065
19	11/09/2020	November	123	21	719
20	11/10/2020	November	117	22	756
21	11/11/2020	November	120	18	1344
22	11/12/2020	November	124	20	1216

Month	SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19 cases
December	1966	713	11687
November	2839	706	23916
October	439	98	6973
Grand Total	**5244**	**1517**	**42576**

Elements of a pivot table

There are three main elements when creating a pivot table.

01. Rows

The list with unique items that we'd like to group the pivot table by.

02. Values

The values to be included in the grouping and how we'd like to summarise them, for example, `SUM`, `COUNT`, `MIN`, `MAX`, etc.

03. Columns

The columns are used to further group the grouped values, i.e., create sub-categories.

2							
3		*Gender*	*Values*				
4		Female			Male		
5	*Month*	SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19 cases	SUM of Supplem	SUM of Ventilatc	SUM of New Co
6	December	1966	713	11687	1896	661	13589
7	November	2839	706	23916	1719	628	9961
8	October	439	98	6973	1033	320	6687
9	**Grand Total**	**5244**	**1517**	**42576**	**4648**	**1609**	**30237**
10							
11							

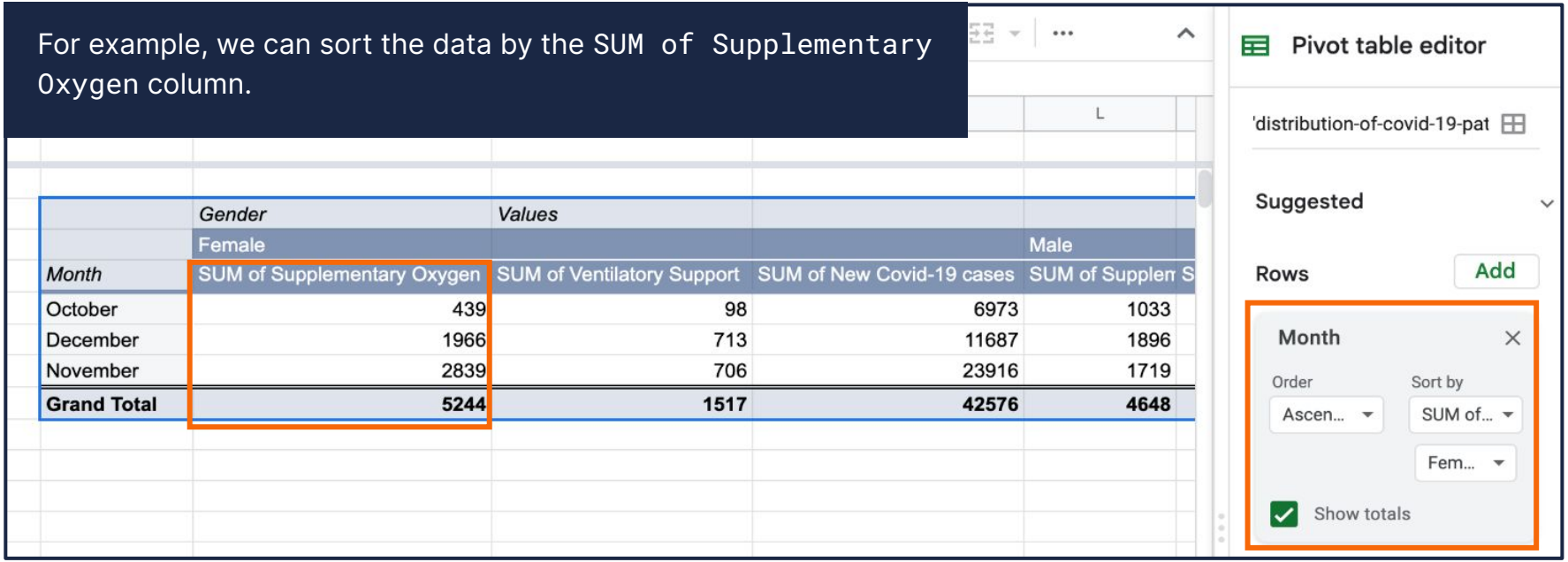
01. 02. 03.

Additional features

Sorting and ordering

Once a pivot table has been constructed, each **row field** may be used to **sort data**, either in **ascending or descending order**.

For example, we can sort the data by the `SUM of Supplementary Oxygen` column.



	Gender	*Values*		
	Female			Male
Month	SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19 cases	SUM of Supplen S
October	439	98	6973	1033
December	1966	713	11687	1896
November	2839	706	23916	1719
Grand Total	**5244**	**1517**	**42576**	**4648**

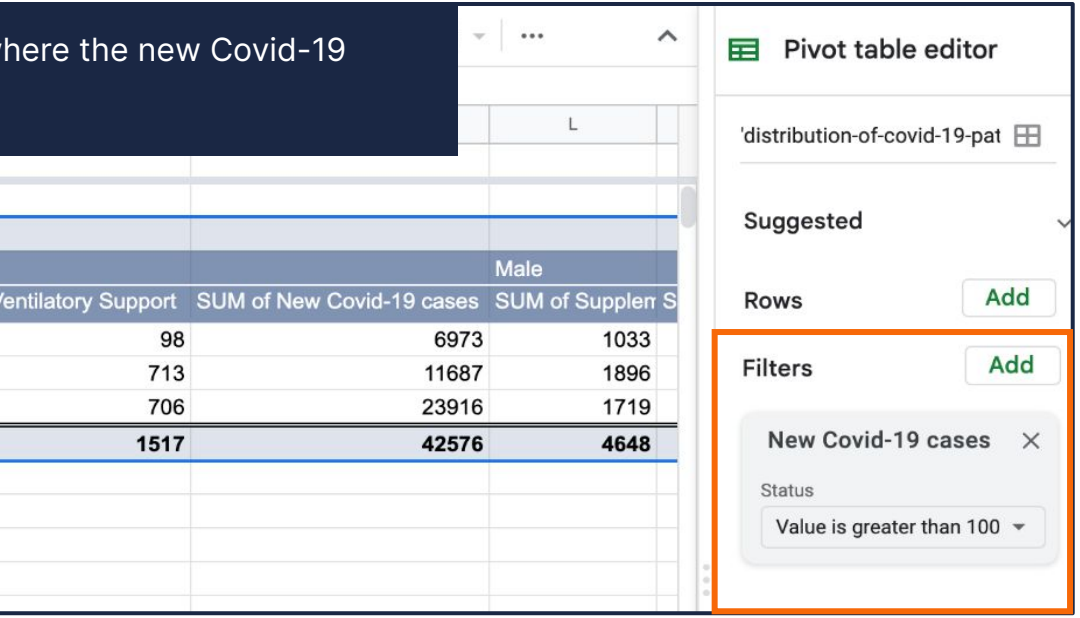
Additional features

Filtering

In the pivot table, we may choose **which data to include** in the display by using the **filter function**.

For example, we can filter for records where the new Covid-19 cases exceeded 100.

	Gender	Values		
	Female			Male
Month	SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19 cases	SUM of Supplem S
October	439	98	6973	1033
December	1966	713	11687	1896
November	2839	706	23916	1719
Grand Total	**5244**	**1517**	**42576**	**4648**

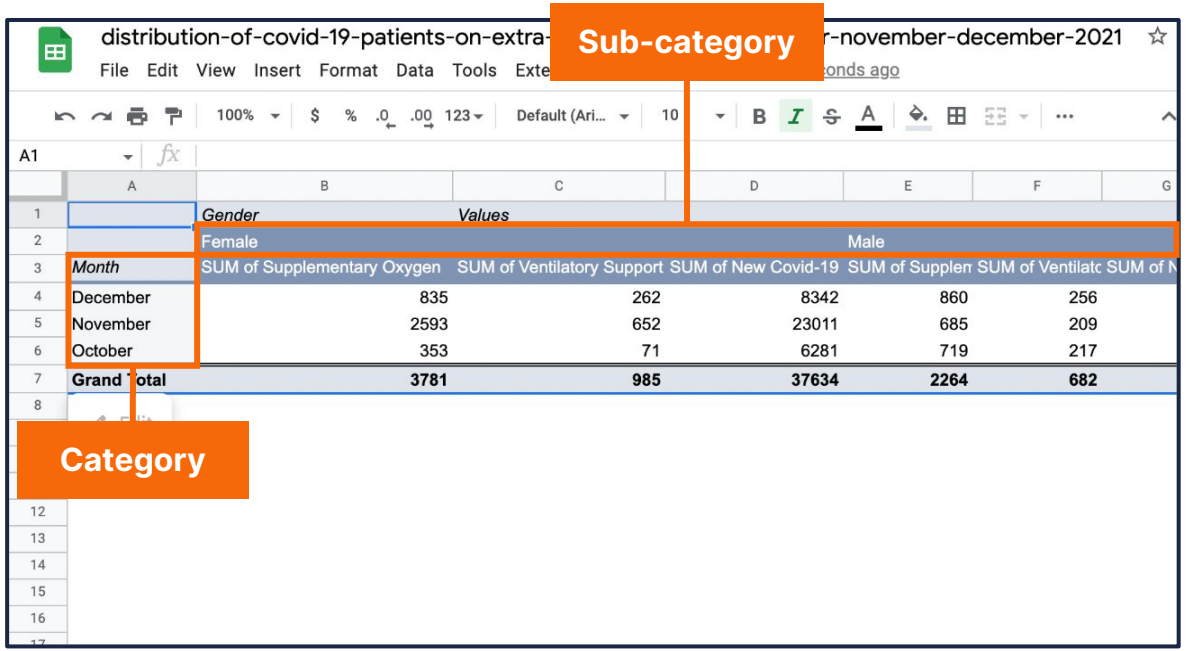


Additional features

Multiple categories

We can also create **multiple categories** in a pivot table, either **column or row sub-categories**.

As an example, we could group the data according to **both gender and month**, gender being the sub-category under the category of the month.



Additional features

Multiple categories

Creating sub-categories using **Columns**: the sub-categories are created as additional columns in the pivot table

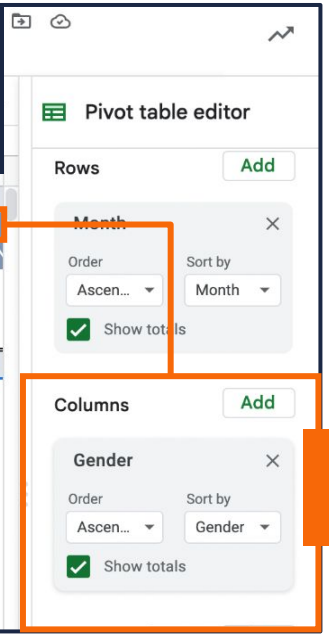
If we add a column under the Columns element it creates a sub-category, allowing our pivot table to display **all the possible combinations** of the two fields between the columns in the pivot table.

	Gender	Values				
		Female			Male	
Month		SUM of Supplementary Oxygen	SUM of Ventilatory Support	SUM of New Covid-19	SUM of Supplem	SUM of Ventilatc SUM of N
December		835	262	8342	860	256
November		2593	652	23011	685	209
October		353	71	6281	719	217
Grand Total		**3781**	**985**	**37634**	**2264**	**682**

Edit

Supplementary Oxygen Female

Supplementary Oxygen Male

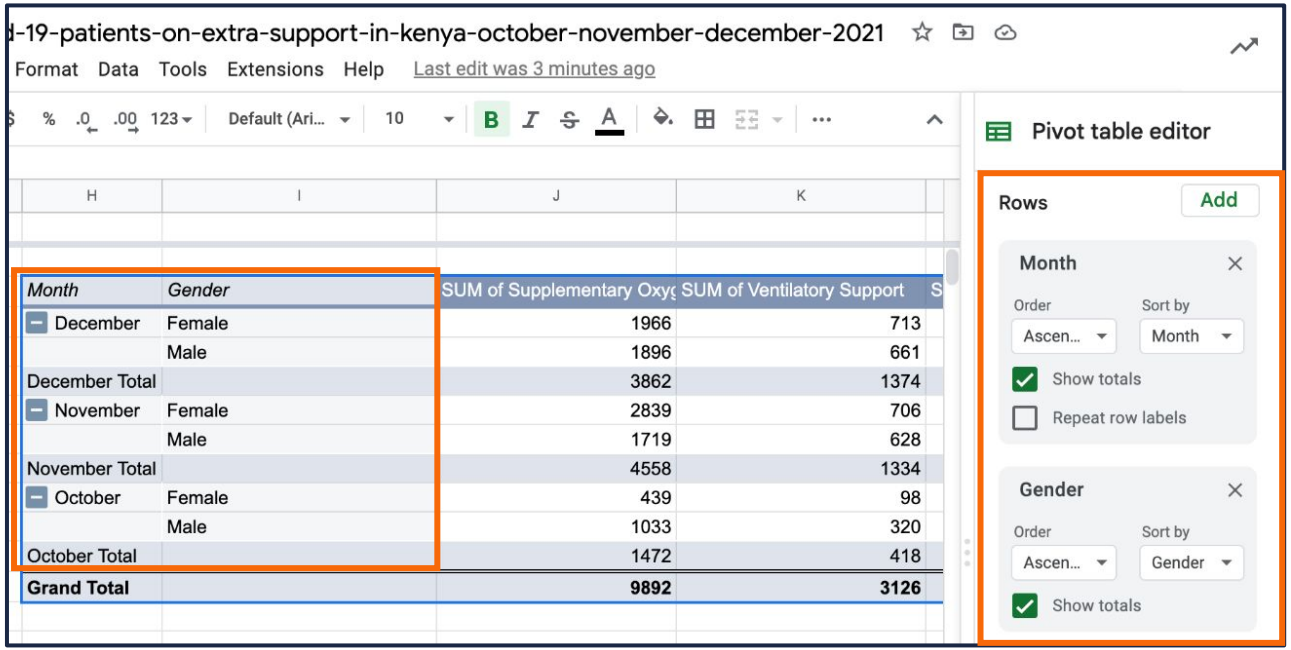


Column sub-category

Additional features

Multiple categories

Creating sub-categories using **Rows**: the sub-categories are created as additional rows in the pivot table



Month	Gender	SUM of Supplementary Oxyg	SUM of Ventilatory Support	S
December	Female	1966	713	
	Male	1896	661	
December Total		3862	1374	
November	Female	2839	706	
	Male	1719	628	
November Total		4558	1334	
October	Female	439	98	
	Male	1033	320	
October Total		1472	418	
Grand Total		**9892**	**3126**	

The first row field added in the Rows element **creates a list of the unique items** in that column. When we add a second row field, it creates a sub-category, allowing our pivot table to display **all the possible combinations** of the two fields between the columns in the pivot table.

Back to the questions we wanted to investigate

Distribution of **Covid-19 patients** on life support in **Kenya**

How many patients required supplementary oxygen **per month per gender group**?

How many patients required ventilatory support **per month per gender group**?

How many new Covid-19 cases were recorded **per month per gender group**?

With just a few clicks and no formula, we can answer the above questions and more!

Month	*Gender*	SUM of Supplementary Oxyg	SUM of Ventilatory Support	SUM of New Co
− December	Female	1966	713	11687
	Male	1896	661	13589
December Total		3862	1374	25276
− November	Female	2839	706	23916
	Male	1719	628	9961
November Total		4558	1334	33877
− October	Female	439	98	6973
	Male	1033	320	6687