



Group by and Visualization

1. Now you first have to create a new cell in your notebook.
2. Then if we write this code to group by the operation name then we can have the data accordingly.

The screenshot shows a Databricks notebook cell with the following code:

```
display(df.groupBy(df("Operationname")).count())
```

Below the code, the output is displayed as a table with 2 columns: **Operationname** and **count**. The table contains 15 rows of data. At the bottom, it indicates 217 rows and 2.05 seconds runtime.

	Operationname	count
1	Delete record set of type A	2
2	Create or Update Route	3
3	Delete Private Endpoint Connection Proxies	2
4	Create or Update Virtual Network Peering	18
5	Create an Azure Automation job	2
6	Revoke Instant Item Recovery for Protected Item	6
7	VaultsOnPremPostRegistrationAction Microsoft.RecoveryServices	3
8	Create or update data collection rule association	24
9	Delete insights component	3
10	Regenerate Storage Account Keys	2
11	Create or update an private endpoint.	4
12	Delete website	21
13	Create or Update Restore Point Collection	6
14	Create or Update Application Gateway	4
15	Register Subscription for Compute	2

3. Now if we just want to see the data above 100 then we can add a filter. Below you can see the results.
4. Now if you click on the plus symbol which is highlighted. You will have two options visualization and data profile. You have to choose visualization.

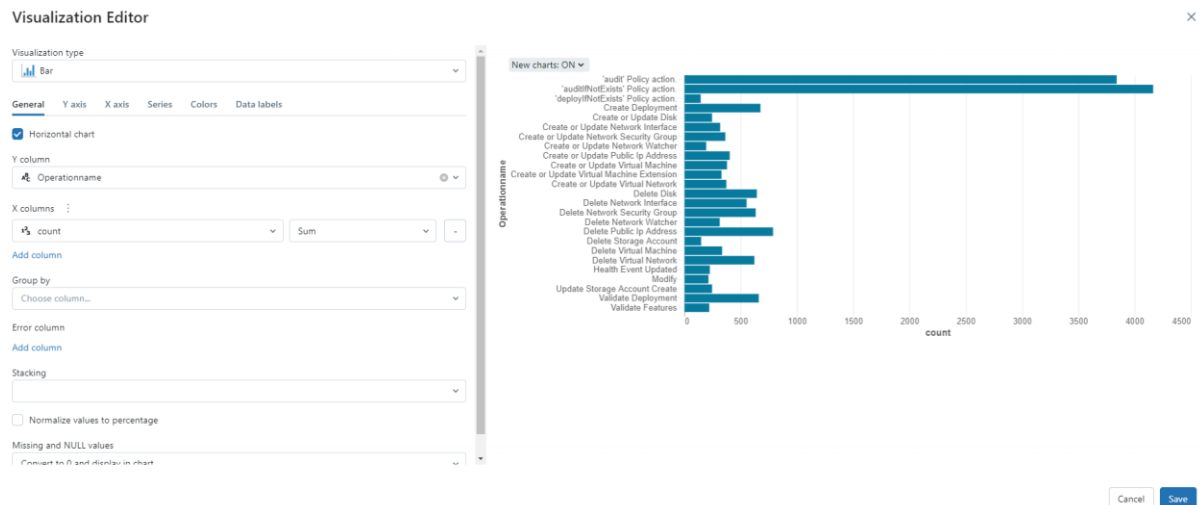
The screenshot shows a Databricks notebook cell with the following code:

```
display(df.groupBy(df("Operationname")).count().alias("Count").filter(col("Count") > 100))
```

Below the code, the output is displayed as a table with 2 columns: **Operationname** and **Count**. The table contains 15 rows of data. At the bottom, it indicates 25 rows and 0.75 seconds runtime. A red box highlights the plus symbol in the table header area.

	Operationname	Count
1	'audit' Policy action.	3841
2	Delete Network Interface	552
3	Create or Update Public Ip Address	402
4	Delete Virtual Machine	334
5	Health Event Updated	225
6	Modify	213
7	Create or Update Disk	245
8	'deployIfNotExists' Policy action.	144
9	Create Deployment	675
10	Validate Features	220
11	Delete Public Ip Address	787
12	Delete Disk	643
13	Validate Deployment	660
14	Create or Update Virtual Network	372
15	Delete Storage Account	148

5. You can explore this visualization editor on your own. You have so many options for visualization types. Now just click on save.



6. Then in your notebook you can see your data in visual form.

