Step by Step Guide to Proceed with the Exercises - Export and Further Analysis

Exercise A: Pivot Analysis in Excel - EBV Status in Burkitt Lymphoma

In this first exercise, we will perform a Pivot analysis on a simple structured field.

The goal is to understand:

- How PivotTables aggregate data
- How to generate frequency distributions
- How to identify missing values

Step 0 - Preparation

1. Extract the ZIP file (Right-click → **Extract All**)
2. Open the file in Excel (files are provided in **CSV** format, but they can be easily opened and converted in Excel).
3. Save working copy (File → **Save As**) to avoid modifying the original file.
4. Ensure the first row contains proper column headers.

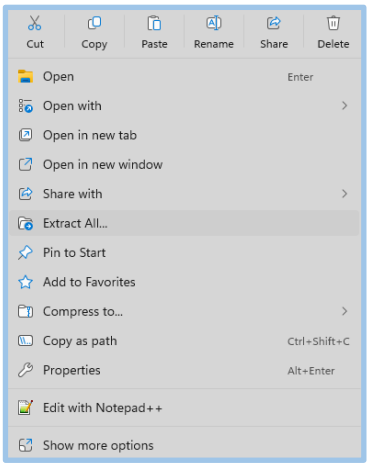


Figure 1: Extract All

Note: If the data appears in a single column, use **Data → Text to Columns** or **re-import** the file with the correct delimiter.

Step 1 - Filter for the Target Population

We will isolate patients with Burkitt lymphoma.

1. Select the header row of your dataset.
2. Add filters to the columns:
    a. In Excel: go to **Data > Filter**
    b. Or convert the range to a table with **Insert > Table**
3. Locate the relevant diagnosis column.
4. Click the filter arrow in the column header.

5. Select only **Burkitt lymphoma**.
6. Click **OK**.

Step 2 - Create a Pivot Table

Next, open the Insert tab and click on the PivotChart icon. Select "**PivotChart & PivotTable**".

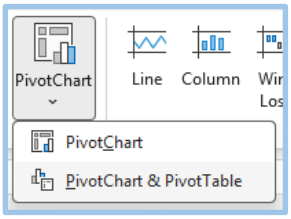


Figure 3: PivotChart

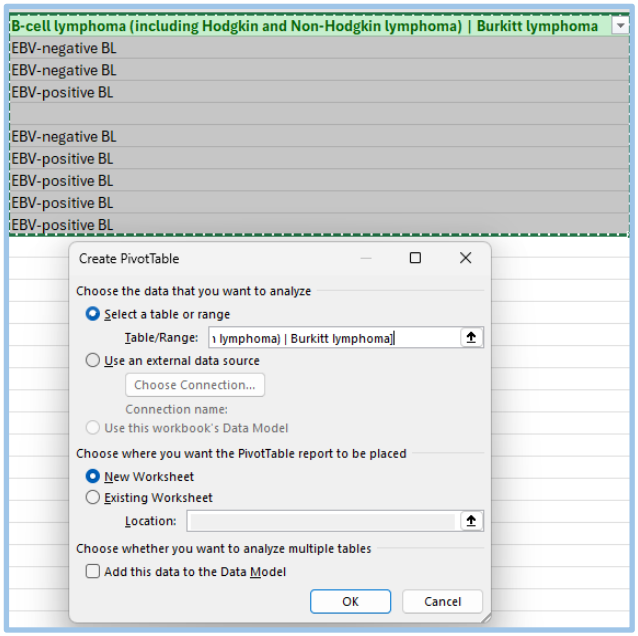


Figure 2: PivotTable & PivotChart creation

Then:

- Select the column "**B-cell lymphoma (including Hodgkin and Non-Hodgkin lymphoma) | Burkitt lymphoma**" as the Table/Range.
- Ensure that the PivotTable report is placed in a **New Worksheet.**
- Click OK.

A new tab named "**Sheet1**" will appear with an empty PivotTable.

Step 3 - Configure the Pivot Fields

On the right-hand side, in the **PivotChart Fields** pane, select:

"**B-cell lymphoma (including Hodgkin and Non-Hodgkin lymphoma) | Burkitt lymphoma**"

This will enable the axis.

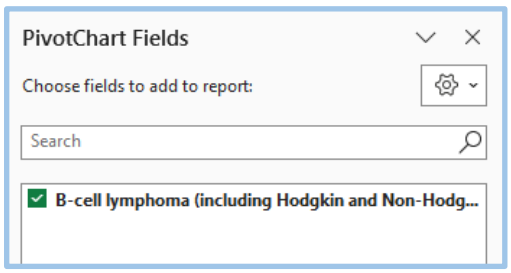


Figure 4: PivotChart Fields

As a next step, drag the selected field to the Values area in order to generate the bar plot for EBV status.

Note: Depending on your Excel settings, the PivotTable layout may differ. For example, the area used for categories may appear as **Rows** instead of "Axis (Categories)".

Empty rows will also be included and will appear in the bar plot.

See the bar plot below:

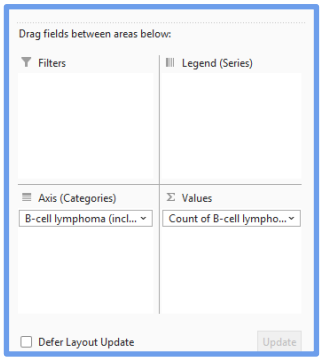


Figure 5: Field Panel

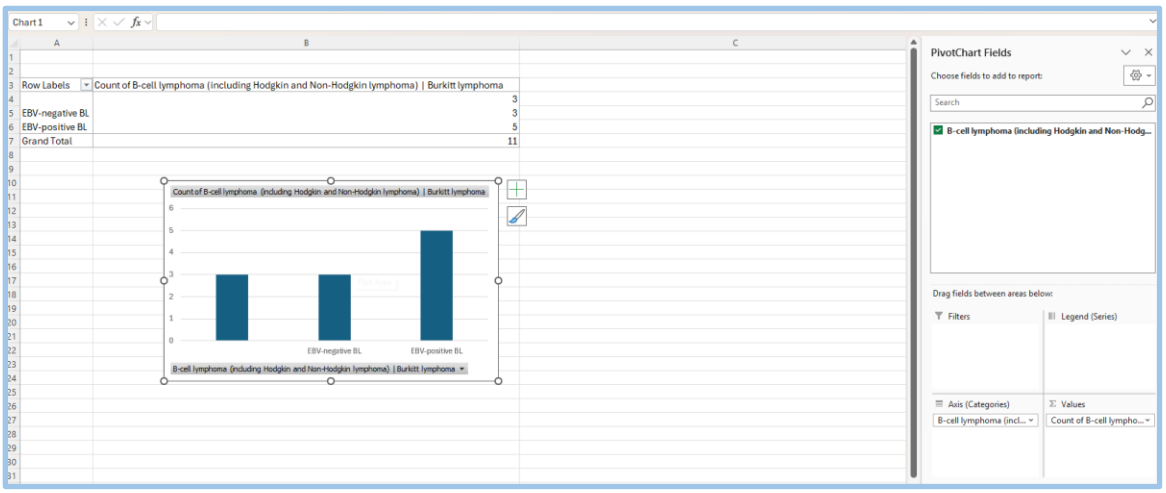


Figure 6: Overview of PivotChart

Exercise B: Handling JSON-Structured Fields in Export Files

In this exercise, we will work with a field that contains structured data (in JSON format).

This means that instead of a single value, the field stores multiple pieces of information grouped together (e.g., key–value pairs).

To analyse **repeatable question answers** (e.g. Chromosomal abnormalities), you first need to transform them into a simpler format.

This means expanding the data so that each answer appears in its own row (one value per row), instead of being grouped together in a single cell.

Let's look at the example of chromosomal abnormalities.

Step 1 - Prepare a Clean Working Area

First, isolate the column of interest in a new sheet to start with a clean layout. You may also rename the column header to something shorter (e.g., chrom_abn).

chrom_abn
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}
{"1":"del(1p) / 1p-"}
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 10"}
{"1":"-11 / monosomy 11"}
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}

Figure 7: JSON field column

Step 2 - Convert to Table and Open Power Query

1. Select the data and press **Ctrl + T (or Command + T on Mac)** to create a table.
2. Make sure the option "**My table has headers**" is checked.
3. Press **OK**.

Next:

- Go to the **Data** tab.
- Select "**From Table/Range**" to open the **Power Query Editor**.

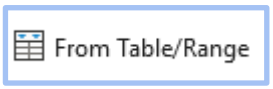


Figure 8: Power Query

Note:

Excel functions may appear differently depending on your system language settings.

For example, =SUM() in English Excel is =SOM() in Dutch Excel.

In Power Query, most functions are written in English, but if you encounter errors, please check your regional settings.

Step 3 - Expand JSON Data to a new table/column

In the **Add Column** tab of the Power Query Editor:

- Click "**Custom Column**"
- Name the new column something meaningful, such as "Data"
- In the formula field, enter the following expression

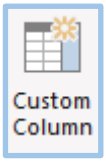


Figure 9: Custom Column

This step converts the structured data into a format that Excel can analyse:

= Record.ToTable(Json.Document([your column]))

This will convert the JSON content into a structured table that can then be expanded into separated rows.

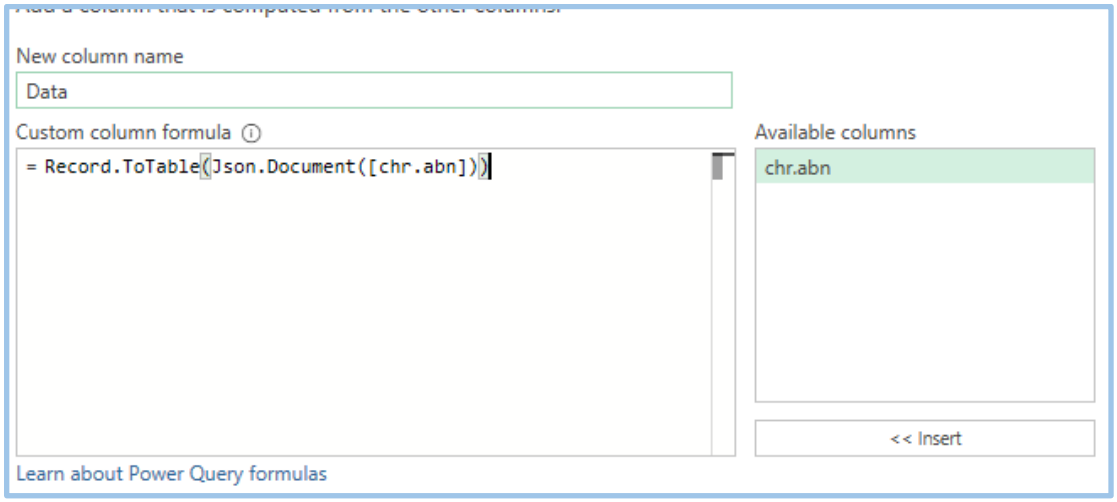


Figure 10: Formula for column formation

Step 4 - Handling Empty Rows (Error Fix)

To avoid errors for empty values, you need to modify the formula by adding a **try … otherwise** condition.

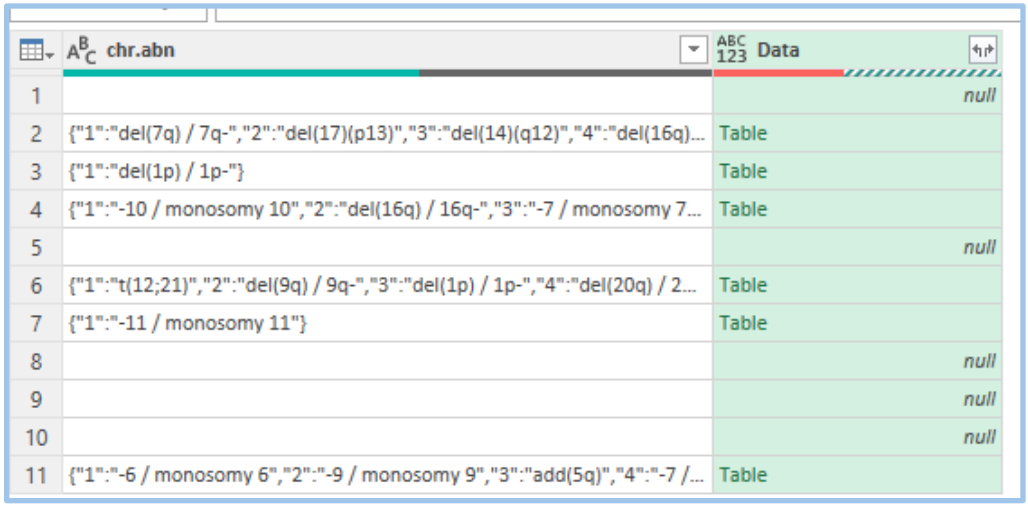
Use the following logic:

- **Try** to convert the JSON value into a table
- **Otherwise**, return **null**

```
= Table.AddColumn(#"Changed Type", "Data", each try Record.ToTable(Json.Document([chr.abn]))otherwise null)
```

Figure 11: each try.. otherwise null

After applying this formula, the errors will be replaced with null, and you should see:



	chr.abn	Data
1		null
2	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	Table
3	{"1":"del(1p) / 1p-"}	Table
4	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	Table
5		null
6	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	Table
7	{"1":"-11 / monosomy 11"}	Table
8		null
9		null
10		null
11	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	Table

Figure 12: Column Data expansion

This ensures the column can be safely expanded in the next step.

Step 5 - Expanding the JSON Data

In the next step, you need to expand the Data column to reveal the underlying tables.

- Click the expand icon (the two arrows) in the top-right corner of the newly created **Data** column.
- In the pop-up window, select the relevant fields (e.g., **Name** and **Value**)
- Click **OK**.

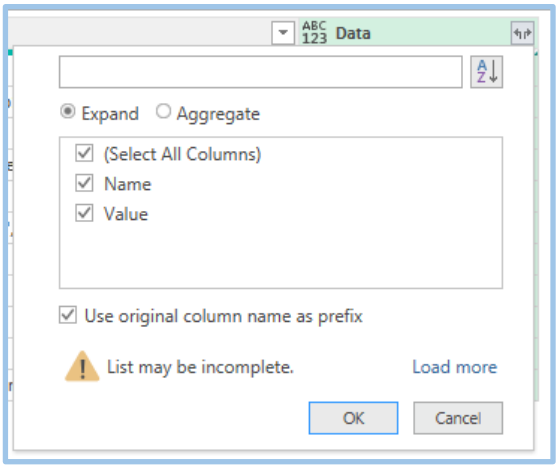


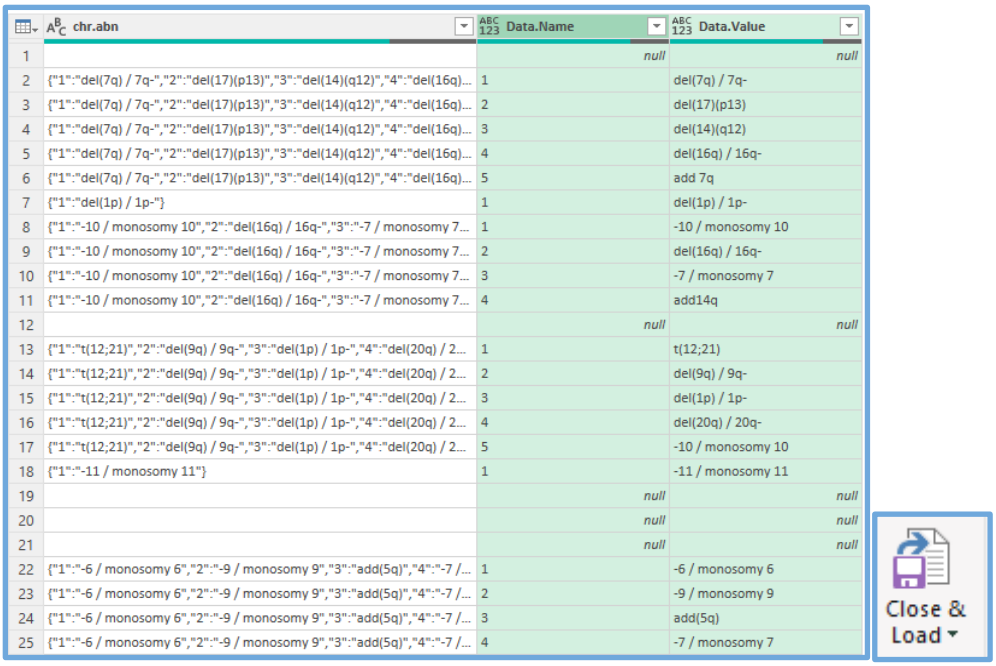
Figure 13: Column data expansion

You have now expanded the JSON-structured field into simple values, with one value per row.

To load the result back into Excel:

- Go to the **Home** tab in Power Query.
- Click "**Close & Load**".

The expanded dataset will now appear in a new Excel sheet, ready for further analysis.



	chr.abn	Data.Name	Data.Value
1		null	null
2	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	1	del(7q) / 7q-
3	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	2	del(17)(p13)
4	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	3	del(14)(q12)
5	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	4	del(16q) / 16q-
6	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	5	add 7q
7	{"1":"del(1p) / 1p-"}	1	del(1p) / 1p-
8	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	1	-10 / monosomy 10
9	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	2	del(16q) / 16q-
10	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	3	-7 / monosomy 7
11	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	4	add14q
12		null	null
13	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	1	t(12;21)
14	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	2	del(9q) / 9q-
15	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	3	del(1p) / 1p-
16	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	4	del(20q) / 20q-
17	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	5	-10 / monosomy 10
18	{"1":"-11 / monosomy 11"}	1	-11 / monosomy 11
19		null	null
20		null	null
21		null	null
22	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	1	-6 / monosomy 6
23	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	2	-9 / monosomy 9
24	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	3	add(5q)
25	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	4	-7 / monosomy 7

Figure 14: Result of Data column expansion: Data.Name, Data.Value

Exercise C: Pivot Analysis after JSON expansion

Following the previous exercise, once the JSON field has been expanded, we can proceed with a short analysis using a PivotTable.

By selecting the newly generated structured table, we can create a PivotChart.

chr.abn	Data.Name	Data.Value
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	1	del(7q) / 7q-
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	2	del(17)(p13)
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	3	del(14)(q12)
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	4	del(16q) / 16q-
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	5	add 7q
{"1":"del(1p) / 1p-"}	1	del(1p) / 1p-
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}	1	-10 / monosomy 10
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}	2	del(16q) / 16q-
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}	3	-7 / monosomy 7
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}	4	add14q
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 1	1	t(12;21)
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 1	2	del(9q) / 9q-
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 1	3	del(1p) / 1p-
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 1	4	del(20q) / 20q-
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 1	5	-10 / monosomy 10
{"1":"-11 / monosomy 11"}	1	-11 / monosomy 11
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}	1	-6 / monosomy 6
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}	2	-9 / monosomy 9
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}	3	add(5q)
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}	4	-7 / monosomy 7

Figure 15: Structured table

To do this:

- Select the expanded table (with columns such as **chr.abn, Data.Name, and Data.Value**).
- Go to Insert → **PivotChart** (or PivotTable)
- Place the report in a **New Worksheet**.

Then:

- Enable the Data.Value of Rows (Axis) area.
- Drag Data.Value again to the Values area.

This will generate a bar plot showing the frequency of each chromosomal abnormality.

The result should look like the example below with each abnormality displayed on the axis and its corresponding count shown in the chart.

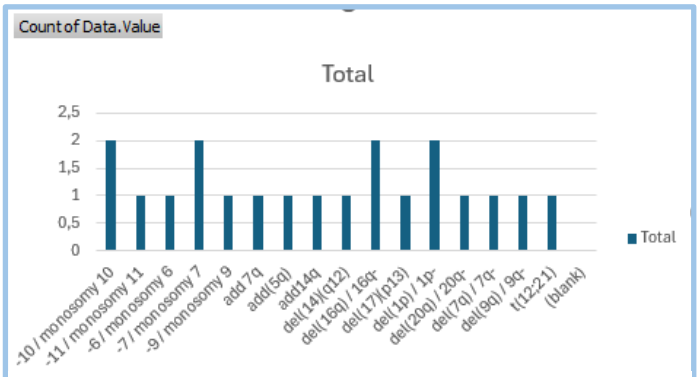


Figure 16: Bar plot for Chromosomal abnormalities and its corresponding count

Exercise D: Merging two Expanded JSON Fields

This exercise is slightly more advanced than the previous ones.

The first step is to expand both JSON fields, which are aligned because they share the same index number within the repeatable group.

Our goal is to **combine**:

- The type of chromosomal abnormality (e.g., monosomy 13)

  with

- It's corresponding indication (e.g., present or absent)

Step 1 - Prepare the Data

First, isolate the two relevant columns in a new working sheet.

Example fields:

1. "**Chromosome Analysis (All above mentioned B-cell lymphomas) | Chromosome Analysis (All above mentioned B-cell lymphomas) | Chromosomal abnormalities**"
2. **"Chromosome Analysis (All above mentioned B-cell lymphomas) | Chromosome Analysis (All above mentioned B-cell lymphomas) | Indicate whether the abnormalities were absent or present"**

For clarity, rename the columns to:

- Chr.abn
- indication

Chr.abn	indication
{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q) / 16q-","5":"add 7q"}	{"1":"Present","2":"Present","3":"Present","4":"Present","5":"Present"}
{"1":"del(1p) / 1p-"}	{"1":"Present"}
{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7","4":"add14q"}	{"1":"Present","2":"Present","3":"Present","4":"Absent"}
{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 20q-","5":"-10 / monosomy 10"}	{"1":"Absent","2":"Present","3":"Present","4":"Present","5":"Present"}
{"1":"-11 / monosomy 11"}	{"1":"Present"}
{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 / monosomy 7"}	{"1":"Present","2":"Present","3":"Present","4":"Present"}

Figure 17: Two columns aligned 1:1 based on repeatable group index.

Step 2 - Convert to Table

- **Select** both columns.
- Press **Ctrl + T** to create a table.
- Ensure "**My table has headers**" is checked.

Then:

- Go to **Data** → **From Table/Range** to open the **Power Query Editor**

Step 3 - Expand Chr.abn

Add a Custom Column to expand the Chr.abn JSON field (as shown in the previous exercise). Make sure to properly handle empty values.

Name the new column:

Data_C

Use the following formula:

= Record.ToTable(Json.Document([Chr.abn]))

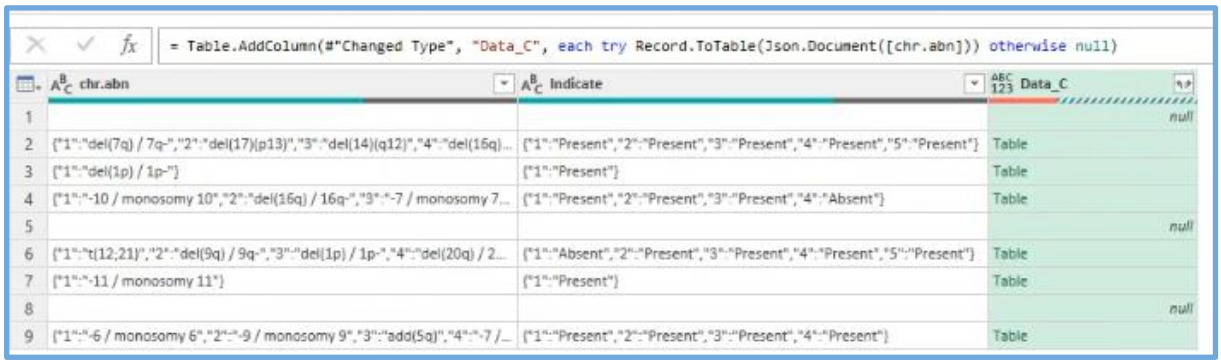


Figure 18: Chr.abn column/table formation

Step 4 - Expand Indication

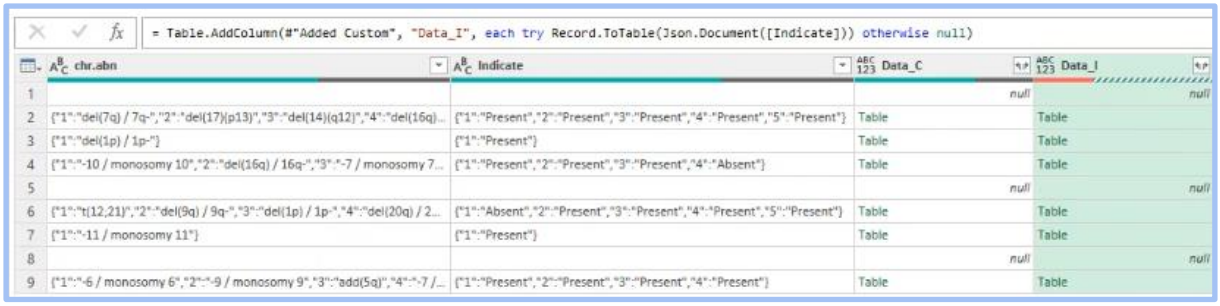
Next, create another Custom Column to expand the Indication column in the same way.

For example, name it:

Data_I

Use:

= Record.ToTable(Json.Document([Indication]))



	chr.abn	Indicate	Data_C	Data_I
1			null	null
2	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	{"1":"Present","2":"Present","3":"Present","4":"Present","5":"Present"}	Table	Table
3	{"1":"del(1p) / 1p-"}	{"1":"Present"}	Table	Table
4	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	{"1":"Present","2":"Present","3":"Present","4":"Absent"}	Table	Table
5			null	null
6	{"1":"t(12,21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	{"1":"Absent","2":"Present","3":"Present","4":"Present","5":"Present"}	Table	Table
7	{"1":"-11 / monosomy 11"}	{"1":"Present"}	Table	Table
8			null	null
9	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	{"1":"Present","2":"Present","3":"Present","4":"Present"}	Table	Table

Figure 19: Indicate column/table formation

Important Note

Be careful to correctly apply the try … otherwise null logic to avoid errors from empty rows.

After this step, you will have two structured table-columns (Data_C and Data_I) that can be expanded and aligned based on their shared index.

Step 5 - Create the Merged Column

Next, we will create a merged column.

To do this, add another Custom Column. This time, we will combine the two previously created structured columns (Data_C and Data_I) using a nested join.

Use the following formula:

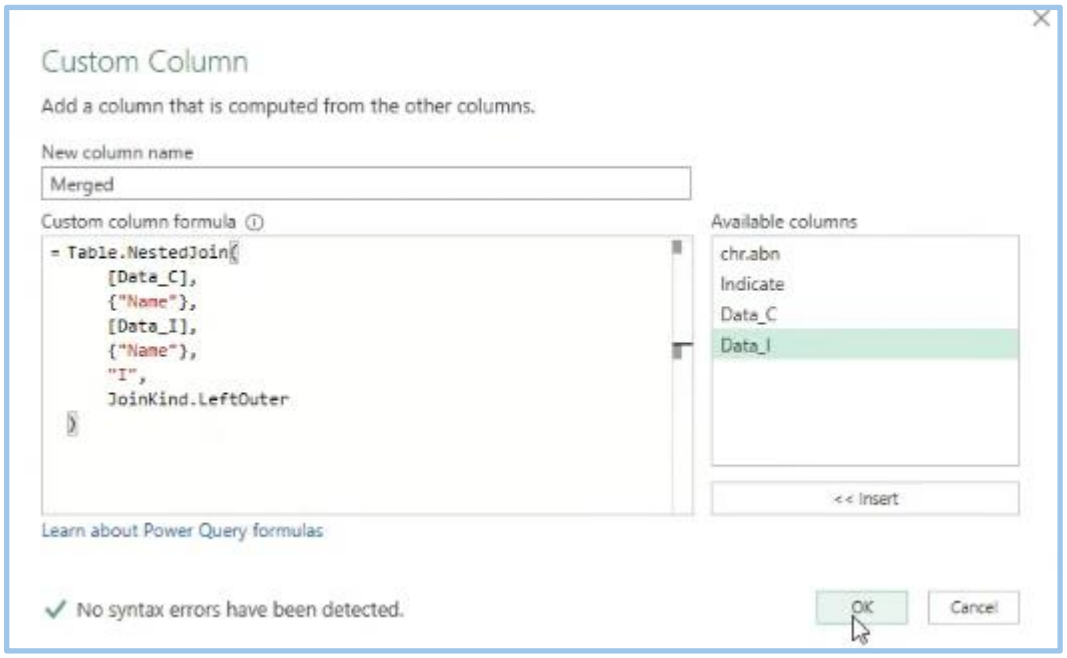


Figure 20: Formula for making the Merged Column

Name the new column:

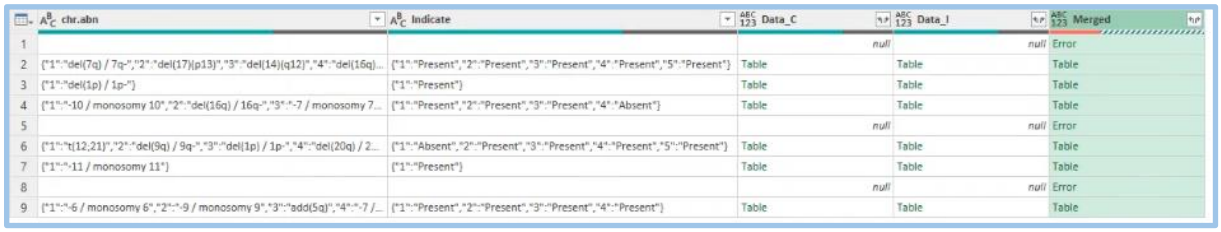
Merged

What does this formula do?

- **Name →** This is the key column used to align both tables. It represents the index of the repeatable group (1,2,3 etc.), ensuring that each chromosomal abnormality is matched with its corresponding indication.
- **JoinKind.LeftOuter** → This keeps all rows from Data_C (chromosomal abnormalities), even if no matching value exists in Data_I (indication)

In simple terms:

We are matching each abnormality with its corresponding present/absent value based on their share index.



	chr.abn	Indicate	Data_C	Data_I	Merged
1			null	null	Error
2	{"1":"del(7q) / 7q-","2":"del(17)(p13)","3":"del(14)(q12)","4":"del(16q)...	{"1":"Present","2":"Present","3":"Present","4":"Present","5":"Present"}	Table	Table	Table
3	{"1":"del(1p) / 1p-"}	{"1":"Present"}	Table	Table	Table
4	{"1":"-10 / monosomy 10","2":"del(16q) / 16q-","3":"-7 / monosomy 7...	{"1":"Present","2":"Present","3":"Present","4":"Absent"}	Table	Table	Table
5			null	null	Error
6	{"1":"t(12;21)","2":"del(9q) / 9q-","3":"del(1p) / 1p-","4":"del(20q) / 2...	{"1":"Absent","2":"Present","3":"Present","4":"Present","5":"Present"}	Table	Table	Table
7	{"1":"-11 / monosomy 11"}	{"1":"Present"}	Table	Table	Table
8			null	null	Error
9	{"1":"-6 / monosomy 6","2":"-9 / monosomy 9","3":"add(5q)","4":"-7 /...	{"1":"Present","2":"Present","3":"Present","4":"Present"}	Table	Table	Table

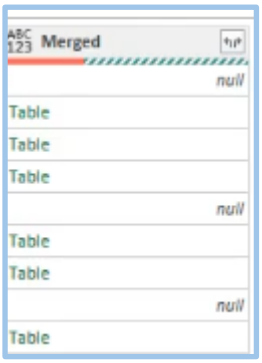
Figure 21: Merged column

Important:

To avoid errors from empty rows, make sure this formula is wrapped with:

try … otherwise null

This prevents failures when one of the nested tables is missing.



Merged
null
Table
Table
Table
null
Table
Table
null
Table

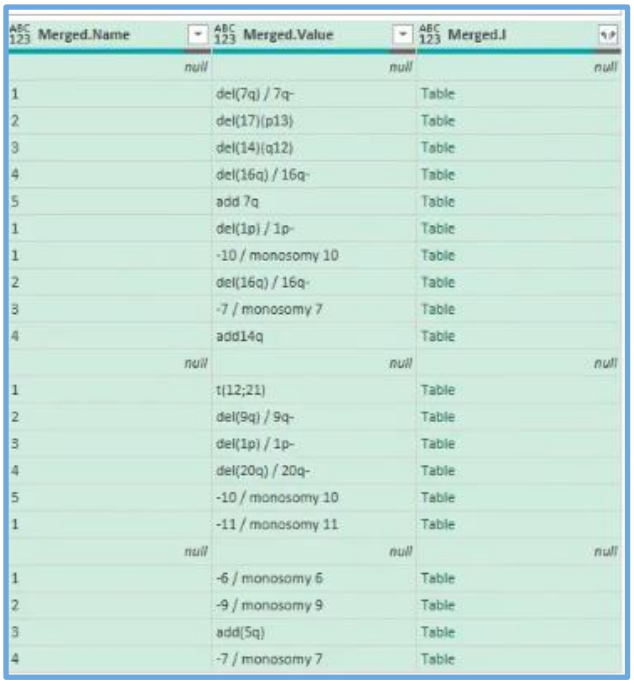
Figure 22: Merged column – nullify errors

Step 6 - Expanding the Merged Column

Next, expand the **Merged** column.

This will generate three columns:

- Merged.Name
- Merged.Value (this corresponds to the chromosomal abnormality value)
- Merged.I (a nested table containing the indication)



	Merged.Name	Merged.Value	Merged.I
	null	null	null
1		del(7q) / 7q-	Table
2		del(17)(p13)	Table
3		del(14)(q12)	Table
4		del(16q) / 16q-	Table
5		add 7q	Table
1		del(1p) / 1p-	Table
1		-10 / monosomy 10	Table
2		del(16q) / 16q-	Table
3		-7 / monosomy 7	Table
4		add14q	Table
	null	null	null
1		t(12;21)	Table
2		del(9q) / 9q-	Table
3		del(1p) / 1p-	Table
4		del(20q) / 20q-	Table
5		-10 / monosomy 10	Table
1		-11 / monosomy 11	Table
	null	null	null
1		-6 / monosomy 6	Table
2		-9 / monosomy 9	Table
3		add(5q)	Table
4		-7 / monosomy 7	Table

Figure 23: Expand Merged Column to its values: Merged.Name, Merged.Value, Merged_I

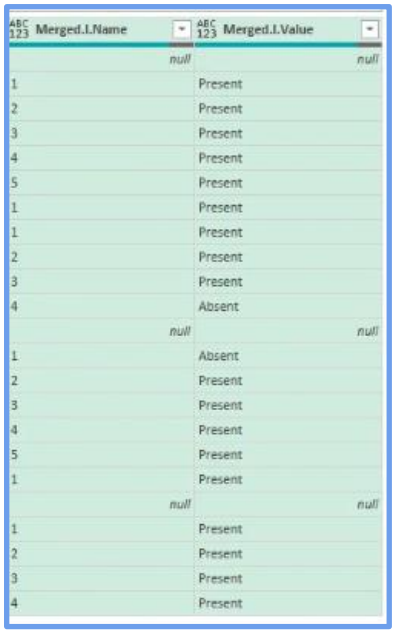
Step 7 - Final Expansion Step

Now expand Merged.I into:

- Merged_I.Name
- Merged_I.Value (this contains the indication: Present / Absent)

After this step, each row will contain:

- The abnormality index
- The abnormality description
- The corresponding indication



Merged.I.Name	Merged.I.Value
null	null
1	Present
2	Present
3	Present
4	Present
5	Present
1	Present
1	Present
2	Present
3	Present
4	Absent
null	null
1	Absent
2	Present
3	Present
4	Present
5	Present
1	Present
null	null
1	Present
2	Present
3	Present
4	Present

Figure 24: Expand Merged_I column to its components: Merged_I.Name, Merged_I.Value

Finally, go to:

Home → Close & Load

This will load the fully expanded and merged table into a new Excel sheet, ready for analysis.

Exercise E: Analysis of Merging JSON files using Pivot Table

As in the previous exercises, we can use a **Pivot Table** to analyse the merged JSON output and understand the distribution of values.

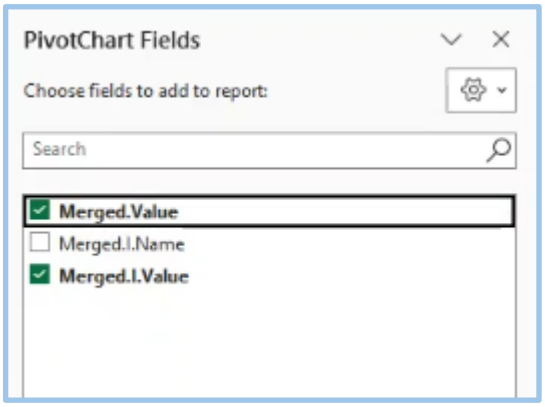


Figure 25: Pivot Chart Fields

Step 1 - Create the PivotTable

1. Select the relevant part of the table containing:
    a. Merged.Value
    b. Merged_I.Value
2. Go to **Insert → Pivot Table**
3. In the PivotChart Fields panel:
    a. Select **Merged.Value**
    b. Select **Merged_I.Value**

Step 2 - Assign Fields to the Correct Areas

Drag the fields as follows:

- Axis (Categories) → **Merged.Value**
- Legend (Series) → **Merged_I.Value**
- Values → **Merged.Value** (set as Count)

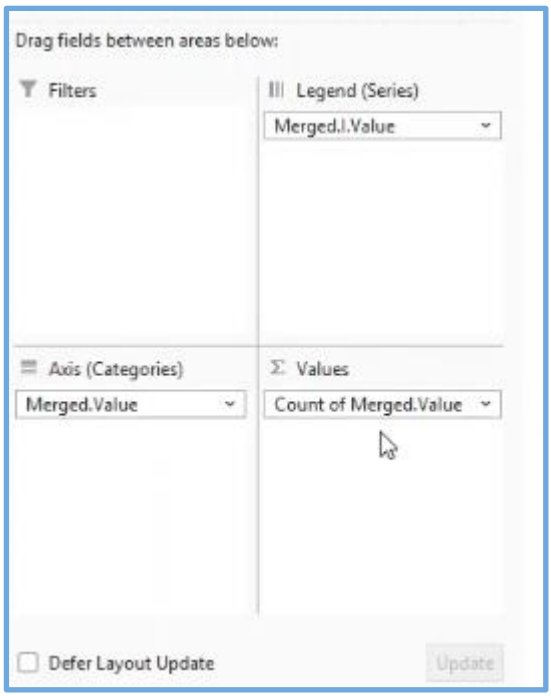


Figure 26: Pivot Chart Panel

What does this mean?

- Axis (Categories)

  Represents the chromosomal abnormalities (e.g, monosomy 7, del(5q), etc.).

- Legend (Series)

  Represents the status of each abnormality:

  - Present
  - Absent
  - (blank)

- Values (Count of Merged.Value)

  Counts how many times each abnormality appears with each status

So effectively, the chart answers:

For each chromosomal abnormality, how many cases are Present vs Absent?

Step 3 - Understanding the Generated Plot

The resulting bar chart shows:

- **X-axis** → Chromosomal abnormalities
- **Y-axis** → Number of occurrences
- Colours (**Legend**) → Status (Present/Absent)

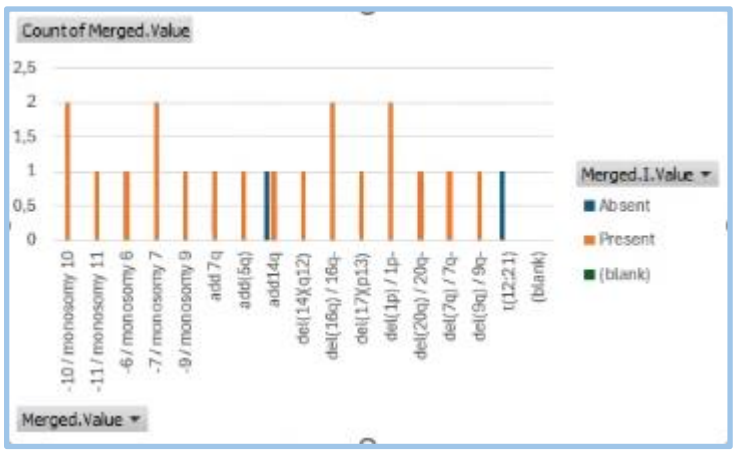


Figure 27: Clustered Bar plot

Step 4 -Improving the Visualisation

By default, Excel may generate a clustered column chart. However, for this type of comparison, a **Stacked Column Chart** is often clearer.

To change the chart type:

1. Right-click on the chart.
2. Select **Change Chart Type**.
3. Go to **Column**.
4. Select **Stacked Column**.
5. Click **OK**.

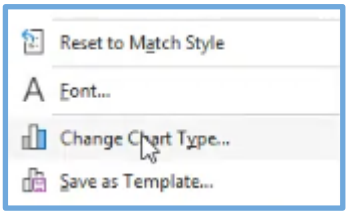


Figure 28: Change Chart Type

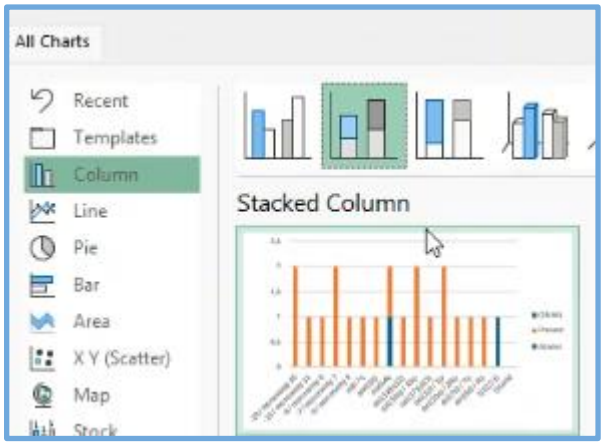


Figure 29: Stacked Bar Blot