# CLEAN & REORDER TABLES

1. Transform csv to .xlx
2. Import. xlx to PowerBI
3. Transform with query: The steps used to clean and reorder are described in the query
4. As good practice I always re order the tables the most efficient way. To save process and storage. In this case you´ll see I had to make a parent & child table from the original.
5. If you have numbers abbreviated in your original table = Ex: In this case we have them in millions (1mill instead of 1,000,000). Then you can choose 2 alternatives. You can change them now to whole numbers or make a measure afterward multiplying the number by 1million in this case. Otherwise, we have a problem, because when we do a graph and filter it by: region & genre. The numbers will be too small. Example 0,005 mill. Better for PowerBI to auto-abbreviate. In this case I will make the measure afterwards. You can see the graphs and filter them to understand my point.
6. DATES= In most reports this is a big problem. If you have at least 1 table with a dates column or related you need to do this:

### Dates

If your date column comes in the right format, ex= 1/1/1980. Great for you, you´ll save a lot of time.

If you have a dates column without proper format, you need to correct it now.

For example, if you have just the years = 1980 as in this case. Or say you have the month & year. You need to transform it into dates. You choose any month and day you prefer, better assign all values the same day and month. In this case we´ll use 1/1/1980. 1/1/1981. For that we do:

Add column 🡪 custom column 🡪 put this code= "01-01-" & Number.ToText([Year])

For more info about this, go to video: <https://www.youtube.com/watch?v=84yEtx7MT04>

Right click on new column 🡪 change type 🡪 date

There you have, your new column date.

You might have other date related issues; I suggest you search on a forum or YouTube video. Clean the date column now in the first steps.

1. Close and apply changes= PowerBI auto connects tables on key columns. But check if its fine. In this case on “info\_id”. Because we put same column title and values match.
2. **Dates again=** You need to create now a date table. Don’t use the date column in your table. It will bring lots of problems afterwards when you try to do more complex analysis. The best way I know to create a date table is with “Bravo”. Before you do so, be sure to uncheck the auto-date/time. This video explains everything about date tables on how to make them: <https://youtu.be/WybnTHDl-AM?t=261>
3. Now, if you are going to use Bravo, I suggest to make some measures before but you can do them afterwards. You don’t always know which measures you´ll be using. For example count/sum of rows. Bravo then automatically makes all date linked measures for those measures ex= YTD sum (it´s explained on the video).
4. Now you should have a beautiful dates table. And if you did it with bravo, you have tons of date related measures to work with. Just connect the Calendar table with all your tables that have a date column. (also explained in video)
5. Clean up and table structuring is complete. Time to analyze:

# Visuals

* Great TIP= Before your 1st graph, set a nice background. Look at some dashboards on the internet and try to copy their style.
* Make your 1st graph according to you report style, make it perfect.
* For the next graphs= You just copy paste the original graph. Then you change the graph type and measures. Doing this, keeps most of the format, so you don’t have to apply the format for EVERY graph in your dashboard.
* I am not going to put much emphasis on esthetics on this project but good to keep in mind for other ones.
* Simplistic graphs (principles of data visualization) data labels with values, no axis titles, no lines. No background for graphs or rounded borders looks good. Titles in the form of questions, Rename each measure for that specific graph. It helps a lot to understand when you hover the mouse over any point in the graph.

# EXPLORATION & QUESTIONS

First, I write the questions I want solved, and then I chose and do the appropriate measures and graphs for each one.

1. Total Games Sold?
2. Most Popular Genres?
3. Which Region sells more?
4. Top 5 games?
5. Now I want to know the answers above but filtered by: genre, region, platform
6. Games Released each year?
7. Total games sold up to date for each year (cumulative)
8. How do the 2 questions above change for each genre (slicers)
9. How many games were sold according to the game´s release date. Ex= games released in 2001 sold 300million copies.
10. Slicer for region and genre for the above question
11. We saw that certain years had a lot of sales. But lest compare each year according to how many titles were released that year= AVG sold by game´s release year.
12. Slicer: Genre, for the graph above
13. Most Popular Platforms? With slicer for genre and region.

# Useful functions

These are some useful functions that you´ll need for other projects

## Cumulative totals =

If you did the date table with bravo you should have them there. But I leave you here the code:

Meassure\_name =

CALCULATE (

COUNT ( game\_info[Genre] ),

# In this case I am using COUNT, for most you´ll use SUM()

FILTER (

ALL ( calendar ),

calendar[Date] <= MAX ( calendar[Date] )

)

)

It´s important that you use the date column in the calendar table. And not the date column in your measure table. Otherwise, the function works wrong.

Be careful, when doing the linegraph with cumulatives be sure to select years/months and not ALL dates. Otherwise, PowerBI needs a lot of memory and process to calculate and plot every single date with its cumulative total.

## Declare vars

VAR \_Varname =

    SELECTEDVALUE ( game\_info[Genre] )  
 // You can declare it like this, other function or just put a value

RETURN

// Here you put any calculation you need for that measures. calculate, sums, etc.

Generally I recommend to try your var before doing the whole calculation:

VAR \_Varname =

ANY CALCULATION

RETURN

\_Varname

Now you make a simple card and put the measure on it. See the result to check your var. If it´s good, keep on.

## Hasonevalue

IF(

HASONEVALUE(table\_Name[column\_Name]),result if true calculation , result if false calc

)

Very useful to apply certain filters. For example if a slicer has only 1 value selected, we could order something to happen. If it has 1<x another calcution.

## Double Filter

If you are going to use two filters on the SAME table, then:

CALCULATE (

FILTER (

ALL ( table\_name ),

// The ALL is commonly used with FILTER, you can take it out if needed

Filter\_1 &&

Filter\_2

)

For example:

FILTER (

ALL ( game\_info ),

game\_info[Date2] <= MAX ( game\_info[Date2] ) &&

game\_info[Genre] = VALUES(game\_info[Genre])

If filters apply for DIFFERENT tables, then:

CALCULATE (

FILTER (

ALL ( table\_name\_1 ),

Filter\_1

),

// You repeat the same function after the comma

FILTER (

ALL ( table\_name\_2 ),

Filter\_2

)

)