

DATA SUITABILITY CHECKLIST FOR PROCESS MINING

What exactly do you need to look for to determine whether your data can be used for process mining? Here is a checklist with the questions that you can go through to assess the suitability of your data.

Yes	Question	Description
<input type="checkbox"/>	Structured data?	Do you have data with columns and rows?
<input type="checkbox"/>	Case ID, Activity, and Timestamp columns available?	Do you have at least one column that can be your case ID, your activity name, and your timestamp? See when a timestamp is not needed here .
<input type="checkbox"/>	Same case ID in multiple rows?	Does the same case ID show up in more than one row at least sometimes? If each row has a unique case ID, your data is either not usable or you may need to reformat it .
<input type="checkbox"/>	Different activities in the same case?	Does the activity name change at least sometimes within the same case? If the activity field does not change over time, it does not contain the history and you need to look for another activity column.
<input type="checkbox"/>	Different timestamps in the same case?	Does the timestamp change at least sometimes within the same case? If the timestamp field does not change over time, it does not contain the history and cannot be used as your timestamp column. You can import your data without timestamps if it is already sorted.
<input type="checkbox"/>	Date and time in one column?	Are the date and the time portion of your timestamp placed in the same column? Because you can have multiple timestamps , each timestamp needs to be in one column.
<input type="checkbox"/>	Data in one file?	If your data was distributed across multiple files (for example, because it comes from different IT systems), have you combined it into one file?
<input type="checkbox"/>	Different timestamp patterns in separate columns?	If you have timestamps with different timestamp patterns, are they placed in different columns?
<input type="checkbox"/>	Activity names human-readable?	Are your activity names understandable (not just a numeric value like an action code, or a transaction number)?
<input type="checkbox"/>	Activity names generalized enough?	Does the same activity in another case have the same activity label (not just a free-text field that is filled differently every time)?

Can you answer 'Yes' to all of the points above? Then you can [import your data](#) into the process mining tool [Disco](#) and continue by [checking the quality of your data](#) before starting the actual process mining analysis.