

Data Preparation Checklist

The following checklist can help you think through all the steps needed for data preparation. Then consult this table again as you validate your data preparation and consider if your data is ready for output.

Input Data

Topic	Item	Yes	No	N/A
Gather Data	Are all tables that contain relevant fields included to support the data analysis?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Input Filters	Can input step filters be applied to data that is unnecessary or restricted for the end user?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Clear Field Selection	Consider the users and the purpose of the data analysis: Are the non-relevant fields cleared from the input selection?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sampling	For large data sets, has the sampling size and method been set to best prepare the majority of the data? Is the full data set used in the flow only when needed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gather Data	Are all tables that contain relevant fields included to support the data analysis?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Data Structure and Cleaning Operations

Topic	Item	Yes	No	N/A
Field Names	Are naming conventions followed? Are names short, accurate, understandable, and descriptive of the underlying data?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Types	Are field data types set correctly based on the underlying data?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Roles	Can available data roles or published custom data roles be used to validate data?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Remove Fields	Consider the users and the purpose of the data analysis: Are the non-relevant fields removed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Split Fields	Are fields that contain multiple pieces of information split into separate fields for easier analysis?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Calculated Fields	Do calculated fields need to be added to perform analysis? Do calculated fields add additional value in support of the data analysis? Have common string cleaning operations, e.g., trim spaces, remove punctuation, been performed? Are date/time fields in a usable format?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Topic	Item	Yes	No	N/A
Filters	Can filters be applied to exclude data that is unnecessary or restricted for the end user?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nulls	When needed, are nulls handled by calculated fields, filters, or default values?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Incorrect or Duplicate Data	Are incorrect or near-duplicate data values edited, grouped and replaced, filtered, or otherwise excluded?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Outliers	When needed, are outliers edited, filtered, or excluded?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Combine Data

Topic	Item	Yes	No	N/A
Joins	Are joins set up when fields from other data sources are needed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Are tables at the same granularity (level of detail) or is an aggregate step needed?			
	Do join keys need cleaning to join correctly?			
	Are join keys the same data type?			
	Are join types (e.g., inner, left, etc.) set up correctly?			

Topic	Item	Yes	No	N/A
Unions	Are tables that have the same structure unioned to provide all the information needed?			
	Are unions in the flow used when data sources have different data types or cleaning is needed to create a common structure?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Are wildcard unions used when files with the same source type and structure are in the same directory?			
	Have mismatched fields been merged, when possible?			

Shape Data

Topic	Item	Yes	No	N/A
Pivots	Is normalized data pivoted to the Tableau Desktop preferred denormalized data structure?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Is a pivot needed to turn columns into rows?			
	Is a pivot needed to turn rows into columns?			
Aggregations	Are aggregates set up when data is needed at a different level of granularity?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

