

# DATA SCIENCE SOURCES CHECKLIST

<b>Data source:</b> <input type="text"/>		<b>Availability</b> Is this an open or proprietary data source? Can anyone in your organization access it? Why (not)?	<b>Origin</b> How was this data collected? Is it generated automatically by sensors or is it based on human input?
<b>Volume</b> What data does this source contain? How much data in terms of gigabytes, records, etc?	<b>Velocity</b> Is this a fairly static data source, or is new data added on a regular basis?	<b>Gatekeeper</b> If this is a proprietary data source, who is responsible for controlling access?	<b>Granularity</b> Is this data source aggregated or does it provide granular insights? What is the highest level of detail it supports (e.g. seconds, products)?
<b>Variety</b> What types of data does this source contain? Does it contain logs, product information, images, audio/video recordings?	<b>Veracity</b> What do you think about the quality of the data contained in this source? Is it accurate, timely, and relevant?	<b>Access</b> How would you access this data source? Through static exports, an API, or a direct link?	<b>Foreign keys</b> What unique identifiers does this data source have that may connect it to other data sources?
		<b>Criticality</b> How business critical is this data source? Is it a production data source? If so, is a staging or backup environment available?	<b>Privacy</b> Is this a sensitive data source from a privacy perspective?
		<b>Credentials</b> What sort of credentials would you need to access this data? Username and password, access keys, other?	<b>Consistency</b> Was all information in this data source collected in a uniform fashion?

## What data sources would contribute to solving your question?

<b>RELEVANCE</b>	Highly relevant, unavailable	Highly relevant, available
	Not relevant, unavailable	Not relevant, available
	<b>AVAILABILITY</b>	

## How can you move data sources to the top-right quadrant?

<b>Data source:</b>	<b>Actions:</b>
<b>Costs:</b>	

# DATA SCIENCE SOURCES CHECKLIST

<b>Data source:</b> <input type="text"/>		<b>Availability</b> Is this an open or proprietary data source? Can anyone in your organization access it? Why (not)?	<b>Origin</b> How was this data collected? Is it generated automatically by sensors or is it based on human input?
<b>Volume</b> What data does this source contain? How much data in terms of gigabytes, records, etc?	<b>Velocity</b> Is this a fairly static data source, or is new data added on a regular basis?	<b>Gatekeeper</b> If this is a proprietary data source, who is responsible for controlling access?	<b>Granularity</b> Is this data source aggregated or does it provide granular insights? What is the highest level of detail it supports (e.g. seconds, products)?
<b>Variety</b> What types of data does this source contain? Does it contain logs, product information, images, audio/video recordings?	<b>Veracity</b> What do you think about the quality of the data contained in this source? Is it accurate, timely, and relevant?	<b>Access</b> How would you access this data source? Through static exports, an API, or a direct link?	<b>Foreign keys</b> What unique identifiers does this data source have that may connect it to other data sources?
		<b>Criticality</b> How business critical is this data source? Is it a production data source? If so, is a staging or backup environment available?	<b>Privacy</b> Is this a sensitive data source from a privacy perspective?
		<b>Credentials</b> What sort of credentials would you need to access this data? Username and password, access keys, other?	<b>Consistency</b> Was all information in this data source collected in a uniform fashion?

<b>Data source:</b> <input type="text"/>		<b>Availability</b>	<b>Origin</b>
<b>Volume</b>	<b>Velocity</b>	<b>Gatekeeper</b>	<b>Granularity</b>
<b>Variety</b>	<b>Veracity</b>	<b>Access</b>	<b>Foreign keys</b>
		<b>Criticality</b>	<b>Privacy</b>
		<b>Credentials</b>	<b>Consistency</b>