

DATA SCIENCE SOURCES CHECKLIST

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

What data sources would contribute to solving your question?

RELEVANCE	Highly relevant, unavailable	Highly relevant, available
	Not relevant, unavailable	Not relevant, available
	AVAILABILITY	

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

Data source:	Actions:
Costs:	

DATA SCIENCE SOURCES CHECKLIST

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

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