

# Menu

Introduction

Lesson 1: Identify types of data

Lesson 2: Locate common data errors

Lesson 3: Resolve common data errors

Lesson 4: Identify data sources

Lesson 5: Identify data gaps

Lesson 6: Build a basic data flow

Lesson 7: Pre-process data

Assessments

Reference 1 – Data types

Reference 2 – Data error types

Reference 3 – Locate data quality

Reference 4 – Common data type errors

Reference 5 – Quality data best practices

Reference 6 – Quality data and machine learning

Reference 7 – Pre-process data



**Skillbuilder Intro Video**



Skillbuilder 3: Create a high-quality data set



## Jessica

- Chief Financial Officer  
Palo Verde Valley Realty (PVV).

# The rationale

Reduce office rent

Locate new office space

Estimate lease liabilities

Clients need lower costs



# High quality data sets are foundational to AI initiatives

## Data availability

Not all data types are available and contribute to AI initiatives

## Quality data types

Quality data types are foundational to AI initiatives.

## Data acquisition

Artificial Intelligence has provided a basic understanding of the importance of data and how all AI components come together to form a business case.





# Objectives

**By the end of this Skillbuilder, you should be able to:**

- Identify types of data
- Locate common types of data errors
- Resolve common data errors
- Identify data sources
- Identify data gaps
- Build a basic data flow
- Pre-process data





# Glossary

Term	Definition
Useless	Useless data is unique, discrete data with no potential relationship with the outcome variable.
Nominal	Nominal data is made of discrete values with no numerical relationship between the different categories—mean and median are meaningless.
Ordinal	Ordinal data are discrete integers that can be ranked or sorted.
Binary	Binary data is discrete data that can be in only one of two categories—either yes or no, 1 or 0, off or on, etc.
Count	Count data is discrete whole number data—no negative numbers here. Count data often has many small values, such as zero and one.







# Glossary

Term	Definition
Useless	Useless data is unique, discrete data with no potential relationship with the outcome variable.
Nominal	Nominal data is made of discrete values with no numerical relationship between the different categories—mean and median are meaningless.
Ordinal	Ordinal data are discrete integers that can be ranked or sorted.
Binary	Binary data is discrete data that can be in only one of two categories—either yes or no, 1 or 0, off or on, etc.
Count	Count data is discrete whole number data—no negative numbers here. Count data often has many small values, such as zero and one.





# Glossary

Term	Definition
Time	Time data is a cyclical, repeating continuous form of data.
Interval	Interval data has equal spaces between the numbers and does not represent a temporal pattern.
Image	Image data (jpeg or png) can be converted into pixels and stored into a matrix table.
Video	Video data that can be converted into images, then converted into a matrix table.
Audio	Audio data can be converted into text, then converted into words and then numeric values for processing.





# Glossary

Term	Definition
Completeness and Comprehensiveness	Having a full representation of the subject as intended. Gaps in data collection lead to a partial view of the overall picture to be displayed.
Availability and Accessibility	Open to access by a system. This characteristic can be tricky at times due to legal and regulatory constraints.
Granularity and Uniqueness	The level of detail at which data is collected is important, because confusion and inaccurate decisions can otherwise occur.
System ID	A unique ID is a great practice for creating any inventory of application names.
System / application name	Self-explanatory field, but good to note that a lot of organizations use internal application nicknames as well as official names.

# Lesson 1: Identify types of data



# Learning objectives

By the end of this lesson, you should be able to:

- Identify common types of data for use in AI 





ock  
Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

isto  
by Getty

iStock  
by Getty Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

ock  
Images™

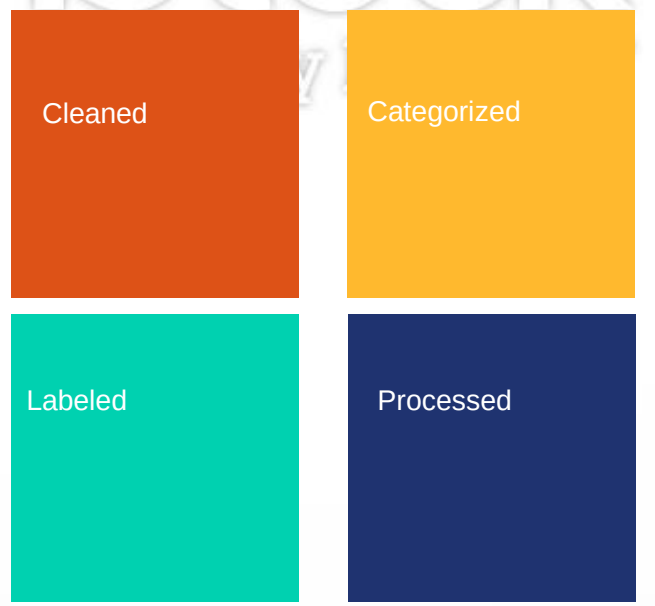
iStock  
by Getty Images™

iStock  
by Getty Images™

isto  
by Getty

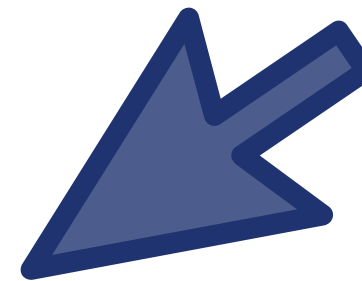
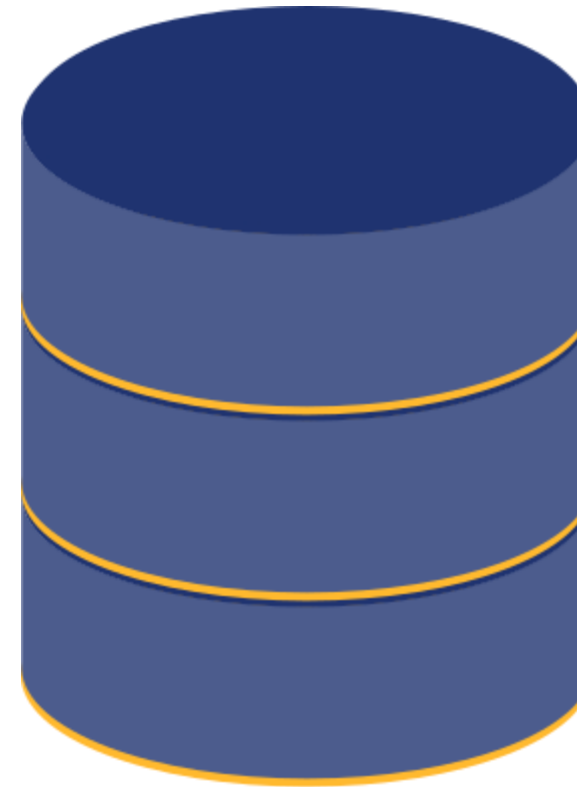
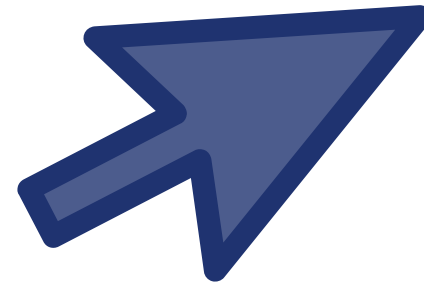
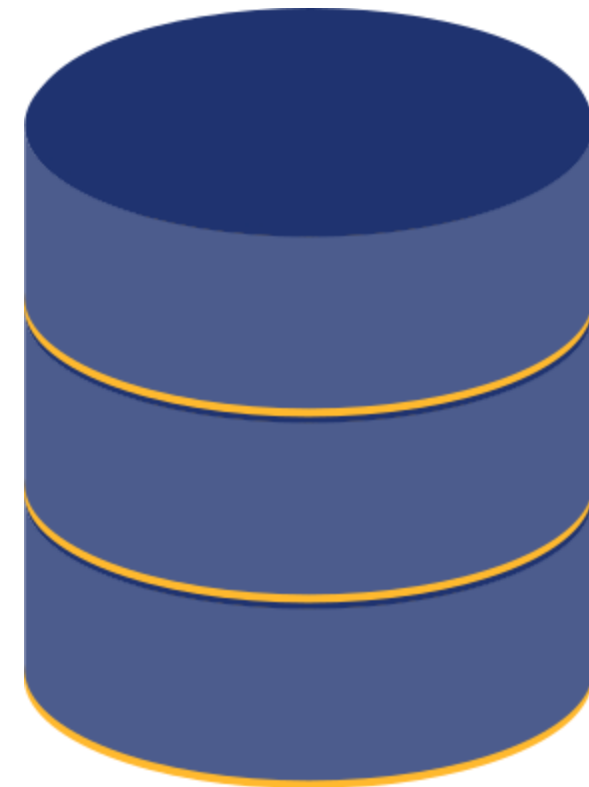


### Quality data sets are critical and must be:



You can only create a quality data set after you understand the data types.

**Understand  
data types**

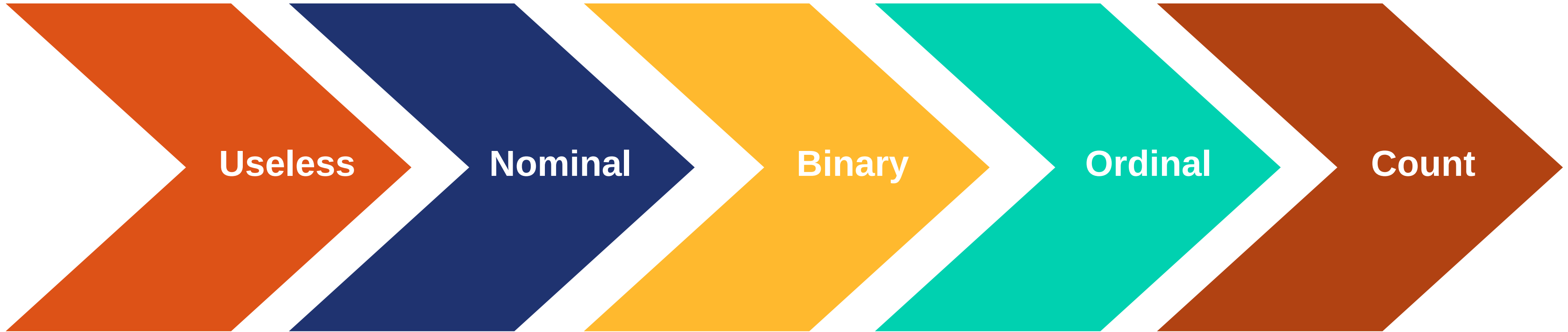


**Create a  
quality data set**





# Types of data



# Types of data



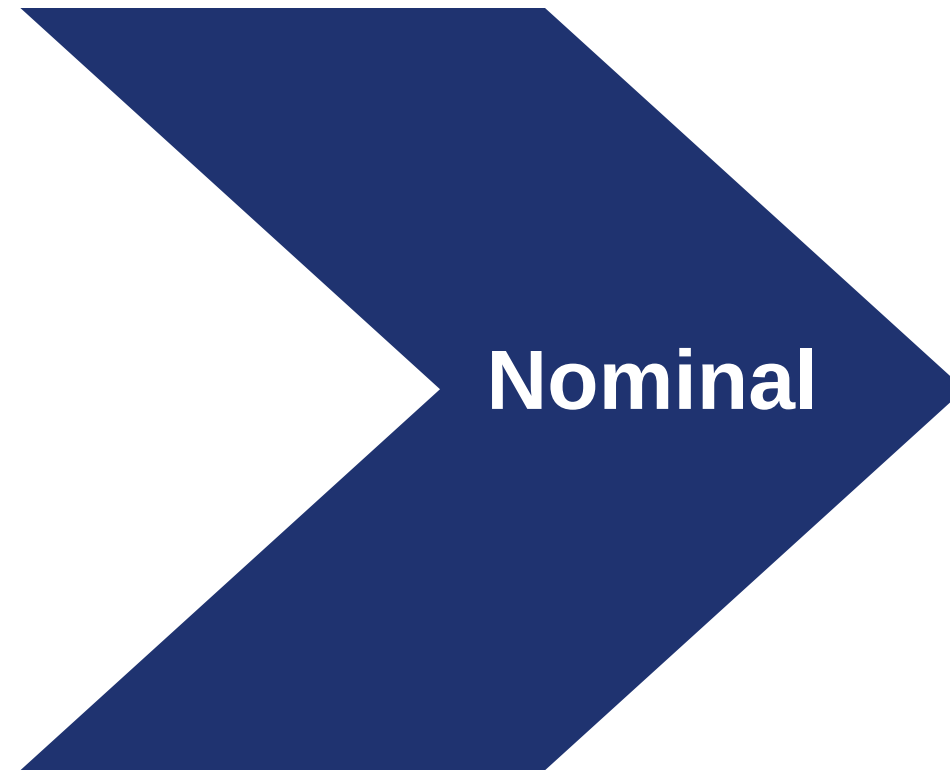
## How are they used?

Random password or bank account numbers

## Examples

Password: Wx9R31\$JG4891L2

# Types of data



## How are they used?

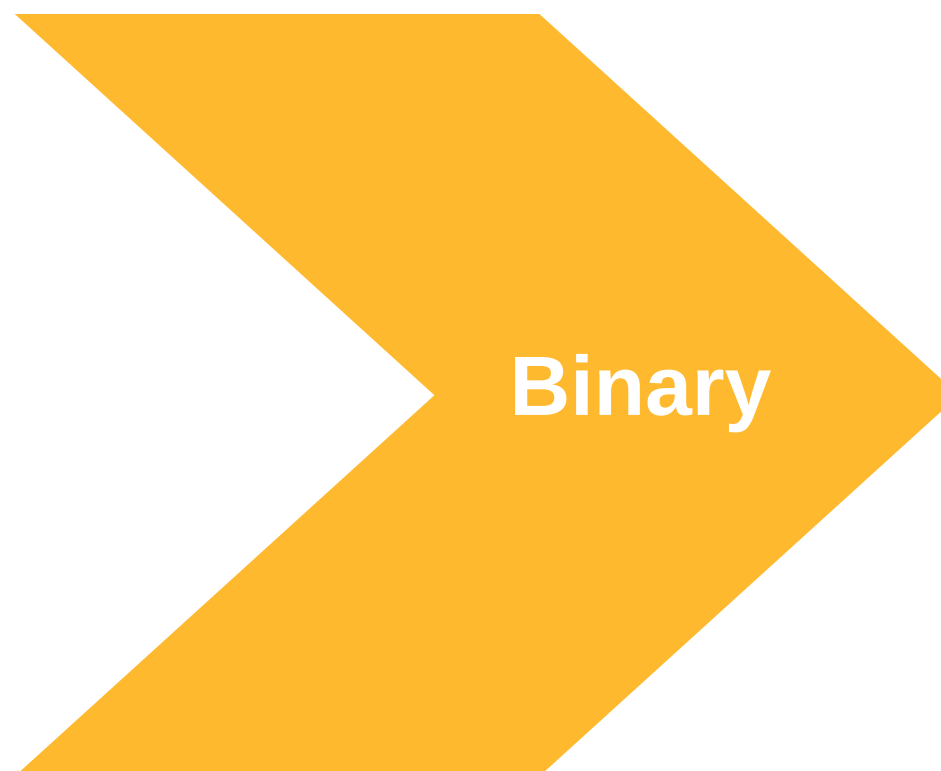
Nonrelated, discrete data for real world items

## Examples

Animal groups: Dog, cat, horse



# Types of data



## How are they used?



0 or 1, yes or no, in or out

## Examples

Pixels labeled with binary values form images.



# Types of data

## How are they used?

Infers an order order without exact relations between them

## Examples

1 to 10 = Bad to good





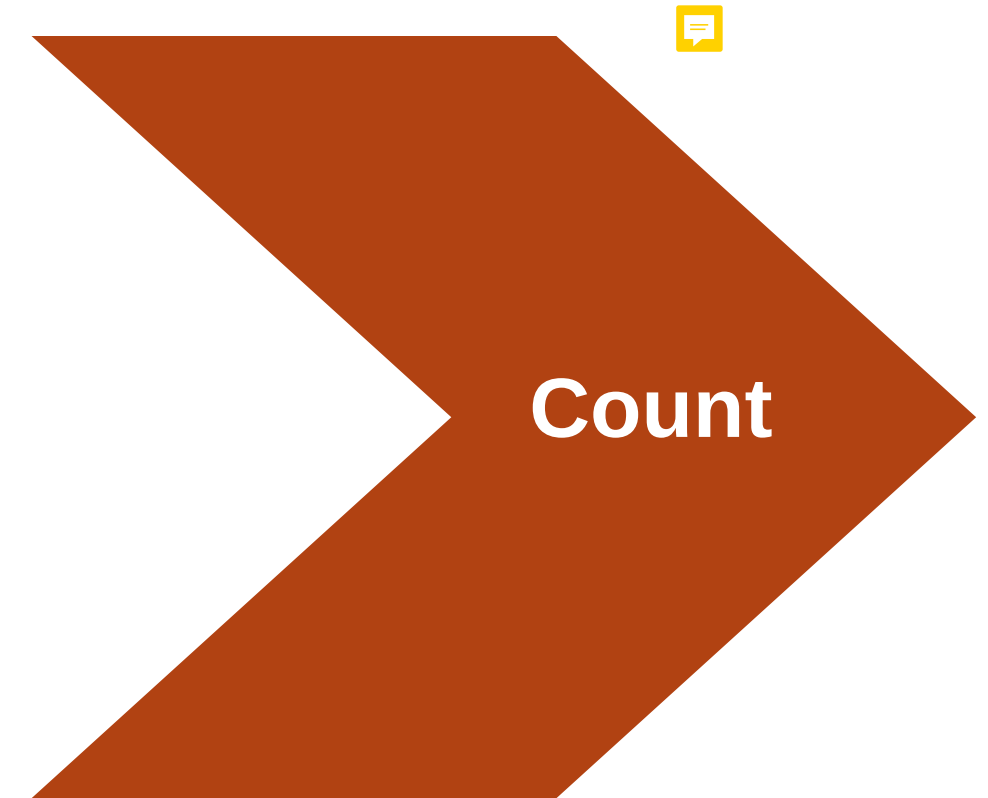
# Types of data

## How are they used?

Response numerical data without exact relations between them


## Examples

How many times have you used this product? 0



## Activity: Identify data types


Jessica has requested that you use RPA Bots to search the databases for a specific office building entry designs. Jessica tells you that AI computer vision can be used to search this by comparing the client's photo to search for this general design in other images. You need to select the best data choice for this task. If you need help, check out the helpful Reference tab.

<input checked="" type="radio"/>	Binary		<input type="button" value="Reference tab"/>
<input type="radio"/>	Nominal		
<input type="radio"/>	Ordinal		
<input type="radio"/>	Count		

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Activity: Identify data types

Recalling an earlier conversation, you are considering the idea of using social media to poll a targeted audience to optimize the marketing strategy and gain client insights using AI. You start asking website visitors if they are thinking about relocating within the next 2 years, using a Likert scale from 1 to 5, with 1 being very unlikely and 5 being very likely. AI seems like a worthwhile solution to help. What other data types should you consider? If you need help, check out the helpful Reference tab.

<input type="checkbox"/>	Binary		<input type="checkbox"/>
<input type="checkbox"/>	Nominal		
<input checked="" type="checkbox"/>	Ordinal		
<input type="checkbox"/>	Count		

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.



## Activity: Identify data types

You need to also consider using ordinal data to optimize the marketing strategy and gain more client insight by adding a like/dislike button to one of our recently searched online advertisements. If you need help, check out the helpful Reference tab.



Yes, ordinal data can be used.



Reference tab


No, ordinal data can be used.

Confirm

Select two answers then select Confirm. If you need help, select the Reference tab button.

## Activity: Identify data types

You have been asked to determine how many office moves your client has experienced in the past 10 years. What data type should you consider?  
If you need help, check out the helpful Reference tab.

<input type="checkbox"/>	Ordinal		<input type="checkbox"/>
<input type="checkbox"/>	Useless		
<input checked="" type="checkbox"/>	Count		
<input type="checkbox"/>	Nominal		

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

# Classifying types of data



# Classifying types of data



How are they used? 

Any metric for time: second, minute, day

Examples

Due on July 1, 2020

# Classifying types of data



How are they used? 

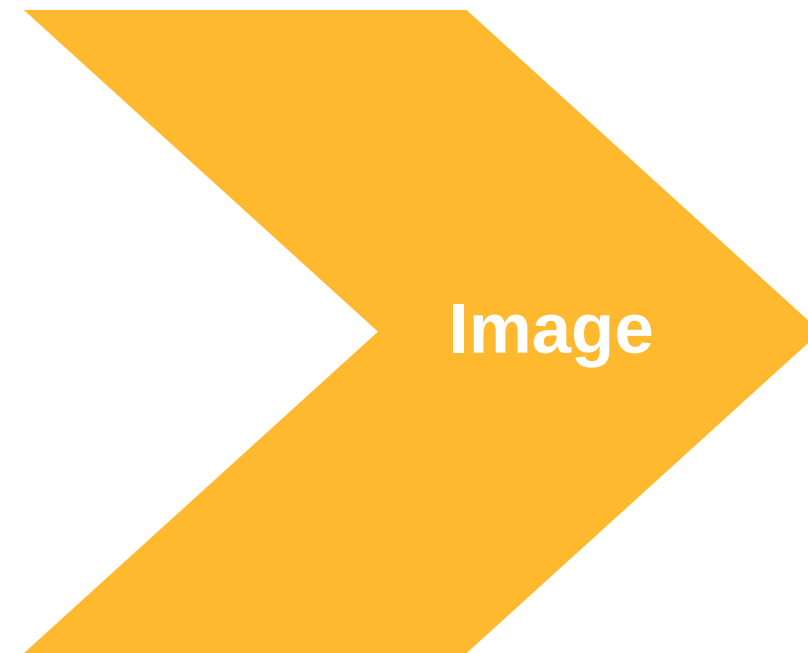
Percentages, income, temperature

Examples

95% probability



# Classifying types of data



How are they used? 

Visualizations, pictures, photographs

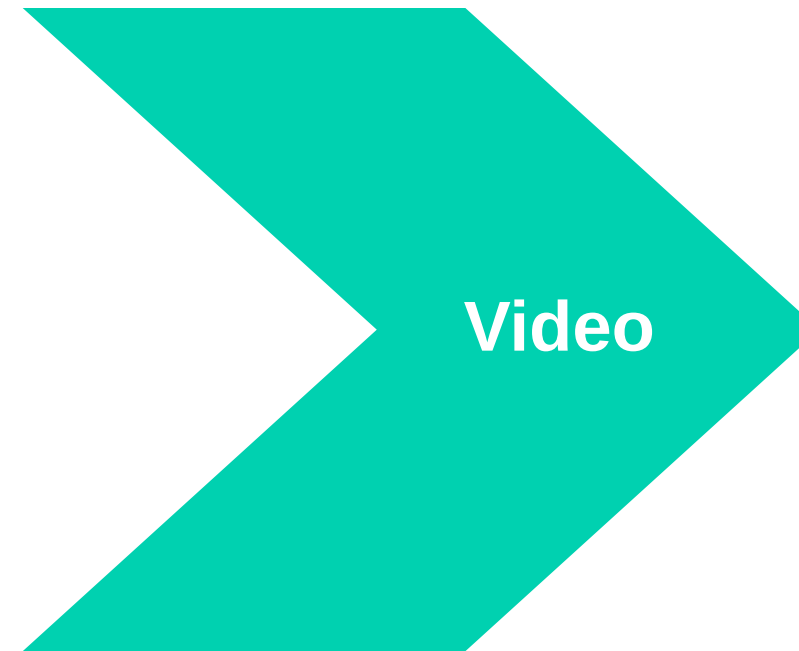
Examples

Any image - visualization





# Classifying types of data



How are they used? 

Active, moving visualization

Examples

Any video





# Classifying types of data

How are they used?

Audible sounds or speech

Examples

Any audio







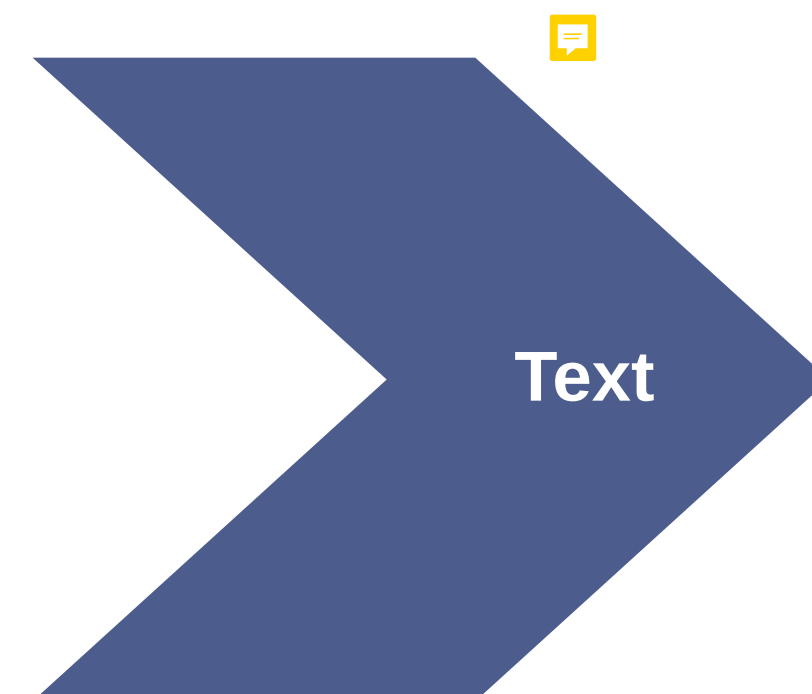
# Classifying types of data

## How are they used?

Words, sentences, writing, characters

## Examples

Analyze an office lease contract



Activity: Identify data types

You discover that AI uses binary data to form images, which is then image data-to analyze the office building images. You learn there are benefits. Why should you use this data for this AI solution to search for office building design entry? If you need help, check out the helpful Reference tab.



The image data provides the client's desired features.



Using binary data, pixels are analyzed to closely match the clients image features—and saves valuable time.

Text data from contract discussions can be easily compared to building layouts.

Stored voicemails from the client contain their list of office space must-haves.

Reference tab



Confirm

Select two answers then select Confirm. If you need help, select the Reference tab button.

## Activity: Identify data types

PVV is also considering producing and using videos to market its services for relocation on its website and social media posts. PVV is excited that the AI Initiative should support the growth from marketing. You have been asked to determine the type of data and its attributes and identify potential issues to consider in using video data.

Video data is commonly used and there are no limits to be concerned about. Consider making longer videos for more content.



Video data requires specialized storage and data should be managed by separately from other data.

Video data is relatively large and requires greater bandwidth for Wi-Fi users. Consider making shorter videos for data size and other reasons.


Video data also uses nominal data, so storage and limits are not issues for concern.

Confirm

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Knowledge check - 1

You need to be clear on certain types of data being used for our AI platform that will generate client reports for its relocation analysis. You see the table below with missing data types. Can you provide assistance in filling in gaps? You will need to enter the correct terms.

Data types	Example	Industry use	
Useless	Bank account numbers Password	Finance, security banking	
Nominal	Hair color Animal groups	Any response/descriptive use in business, or science	
Binary	Pixels are labeled with binary values and form images	Face recognition, any image or feature search	
Count	How many kids do you have? 0	CRM calendars. Deadlines in finance, contracts, or projects.	

[Confirm](#)

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Knowledge check - 2

You need to be clear on certain types of data being used for our AI platform that will generate client reports for its relocation analysis. You see the table below with missing data types. Can you provide assistance in filling in gaps? You will need to enter the correct terms.

Data types	Example	Industry use
Interval	95% probability Test score of 97	Any business performance, medical or science related business
Image	Any image - visualization Specific image attributes	Marketing for all businesses. Medical diagnosis. AI analysis.
Audio	Any audio: educational podcasts, or seminars	Podcasts, virtual assistants, music, entertainment, or education
Text	Analyze an office lease contract for early termination	Chat bots, contract analysis, search engines, other content

[Confirm](#)

Select two answers then select Confirm. If you need help, select the Reference tab button.

## Lesson 2: Locate common data errors



# Learning objectives

By the end of this lesson, you should be able to:

- Locate common data errors









## Common data errors

Root causes of poor data must be resolved

Found in inputs, software, system-level issues, and formatting

Quality data may become subjective

We use definitive characteristics to define data quality



# Locating common data errors



This is the list of the most common data error types. If you do not resolve them, you will get poor performance or failure out of your AI system.



- Invalid or missing values
- Corrupted data or variable format
- Duplicates
- Variable units of measures
- Incomplete data
- Broken URLs
- Data gaps
- Mis-mapped elements

## Activity: Locate common data error types

You located some of the daily logs for some of your clients. You need you to complete the log, determine if the data is quality data or not, and if there are any errors. For each piece of data, enter yes or no for quality data, then enter the number of errors, if any. If you need help, refer to the Reference tab.

Date	Client name	Office sq. footage	Relocation interest	Quality data? Yes or No	No. Errors
12/25/2020	Quick Delivery Enterprises	5,600	No	No	1
2/25/2020	Major Tom Financial	7,200		No	1
2/25/1918	Foundation Data Inc.	large	No	No	2
2/25/2020	Crown Dentistry, L.P.	3700	No	No	1

Confirm



<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

Reference tab



Activity: Locate data quality

All database and non-networked office lease PDFs are either marked: draft, final executed, in progress, unexecuted, or are marked with just initials and a date. You need to determine what quality characteristics we should focus on and why? Choose four of the characteristics and match them with your reason why these are important for this data quality control. If you need help, check out the Reference tab.

Accessibility	Passwords are locked on various PCs making them decentralized and variably secured
Completeness	Data reflects documents may be in progress which implies missing data
Validity	Missing signatures would legal contracts unusable for the purpose intended
Consistency	File names are inconsistent and ambiguous

Reference tab




Confirm

Select two answers then select Confirm. If you need help, select the Reference tab button.

## Knowledge check - 1

You reviewed office lease availability for clients to relocate next year. The database of available office lease metrics will contain square footage, number of offices, available start dates, rental rates, grade of amenities, parking, and utility costs. The date available, square footage and rental rate data are the most critical to clients. So that any data errors can be flagged in this process, enter the best data characteristic and poor data effect for each data error event and use.

Error event	Used for	Characteristic	Poor data effect	Reference tab
August is the correct date and June is the date shown in the data	Date available	Corrupted data or variable format	Validity	
The square footage is shown to be 6,500 and is 66,500	Square footage	Invalid or missing values	Accuracy	
The rental rate is shown as \$45/sq ft and is actually \$35.75	Rental rates	Invalid or missing values	Consistency	

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Lesson 3: Resolve common data errors

# Learning objectives

By the end of this lesson, you should be able to:

- Identify common data type errors
- Resolve common data type errors
- Identify data quality best practices





ock  
Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

iSto  
by Getty

iStock  
by Getty Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

ock  
Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

iSto  
by Getty





## Resolving common data errors

Use a Chatbot to prompt and respond

Chatbot script - ask visitors a strategic question

Script will direct them to a relocation expert

Checking a Chatbot script for errors



Activity: Resolve common data errors

Here is the chat Bot script. Errors may be surprisingly difficult to spot in some cases, regardless of effort. You have been tasked with finding 15 errors. What should you do next? Refer to the Reference tab for help.

Chat Bot Script for directing prospective client users on the website. The Bot's name is Simeon.

Reference tab

Simeon: I see 1 thatyou 2 may be interested in Palo Verde 3 exceptional real estate services. May I answer a question about office space or our award-winning office relocation services 4

Visitor: Yes, I am possibly, going to relocate to new office space in the suburbs. 5

Simeon: Great. Palo Verde has Realty 6 won the RelocationProfessionals of America Excellence Award for each of the last 10 years. May I direct you 7 one of our office space location experts? 8

Visitor: Sure, that would be fine. 9

Simeon: Excellence. Julian Delgado would like 10 contact you by phone 11 . His contact information is 800- 12 13 14  15  
973-0001 and j.m.delgado@palovr.com. Please provide your contact information and time to call, and Julian will contact you at the requested time.

Visitor: Okay. My name is H.L. Hunter and my cell is 900-867-5309 15

Confirm

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

Activity: Quality data best practices

Resolving data errors are a priority for quality data. Best practices with data quality are a must for achieving optimal performance from an AI system. When reviewing some of Palo Verde’s data management practices, you realized there are gaps in the use of best practices for addressing quality data issues. Consider consulting the Reference tab and then assign best practices to the common data issue. You look for best practices to address data concern. Text the correct response in the designated text box.

Data issue found

- File and folders contained highly variable non-descriptive names
- The CFO could not locate a record of her 2019 emails
- The IT assistant created the file structure and does not know business aspects of real estate

Best practices

- Policy governance and dedicated data managers
- Leadership engagement, data issue awareness
- Naming protocols, data proximity awareness



Reference tab

Confirm

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Knowledge check - 1

Jessica indicates that there are gaps in best practice throughout data quality. She asks you to report other issues and share other observations via text message.

Data issue found

Many Lease Files were located only in locked computer hard drives

The CEO did not realize lack of data quality would negatively affect an AI solution performance

The IT Manager installed the CRM with direct access to sensitive Lease File data

Best practices

Policy governance and dedicated data managers

Leadership engagement, data issue awareness

Naming protocols, data proximity awareness



Reference tab

Confirm

<i-text> Select two answers then select Confirm. If you need help, select the Reference tab button.

## Lesson 4: Identify data sources



# Learning objectives

By the end of this lesson, you should be able to:

- Identify data sources







# Introduction to identifying data sources

Contacts for financial accounting system, CRM, and office database sources 

Ref #	System/App Name	Vendor	URL/Path	Description	Technical Contact	Business Contact	Status
1	Sales and accounting system	Sales and accounting system	Sales and accounting system.com	ERP used to capture all financial transactions	M. Rivera 800-727-0099 m.rivera@riverb1.com	R. Murphy	Active
2	Customer relationship management system	Customer relationship management system	Customer relationship management system.com	System used by the sales department to capture contacts for sales	Julie Marsh 623-983-7475 Jm.marsh@pvcrm.net	Jamie Hudson jj.hudson23@pvcrm.net	Active
3	Net1office	Net1office	Officenet1.com	Database for office space	Natasha Gupta 746-23-7364 n.gupta@net1office.com	John Holter JHolter@net1office.com	Active

# Introduction to identifying data sources

Contacts for office improvements, office database, and client database sources

Ref #	System/App Name	Vendor	URL/Path	Description	Technical Contact	Business Contact	Status
4	Ecodev2020	Eco Partners	Eco Partners.com	Office improvements allowance costs	Bill Humphrey 212-987-6453 bill@ecodevp.com	Mark Smith M.Smith@ecod evp.com	Inactive
5	Subfindnow	GenSpaces	GenSpaces.com	Database for office space	Janet Harlow 823-845-1029 j.harlow@genspace.com	Brenda Hollis Brendash@genspace.com	Active
6	Datadrop	MyDataSpot	MyDataSpot.com	Client office lease and other data repository	Palo Verde IT Manager	Kalinda Carlson K.carlso@mydataspot.com	Active

# Explore data sources

1

2

3

4

5

6

7

8

Ref #	System/App Name	Vendor	URL/Path	Description	Technical Contact	Business Contact	Status
1	Sales and accounting system	Sales and accounting system	Sales and accounting system.com	ERP used to capture all financial transactions	M. Rivera 800-727-0099 m.rivera@riverb1.com	R. Murphy	Active
2	Customer relationship management system	Customer relationship management system	Customer relationship management system.com	System used by the sales department to capture contacts for sales	Julie Marsh 623-983-7475 Jm.marsh@pvcrm.net	Jamie Hudson jj.hudson23@pvcrm.net	Active
3	Net1office	Net1office	Officenet1.com	Database for office space	Natasha Gupta 746-23-7364 n.gupta@net1office.com	John Holter JHolter@net1office.com	Active
4	Ecodev2020	Eco Partners	Eco Partners.com	Office improvements allowance costs	Bill Humphrey 212-987-6453 bill@ecodevp.com	Mark Smith M.Smith@ecodevp.com	Inactive
5	Subfindnow	GenSpaces	GenSpaces.com	Database for office space	Janet Harlow 823-845-1029 j.harlow@genspace.com	Brenda Hollis Brendash@genspace.com	Active
6	Datadrop	MyDataSpot	MyDataSpot.com	Client office lease and other data repository	Palo Verde IT Manager	Kalinda Carlson K.carlso@mydataspot.com	Active



## Activity: Identify data sources

The data source table provides much more than just a source for data. You discover that there has been an emergency about processing a final quarterly financial report from the financial accounting software. IT personnel were not available to help with this problem. What should you know in this situation?

The data source table is comprehensive and may provide a user with the ability to repair most system issues on their own.



The data source table includes basic information which is sufficient for any personnel to reference system contacts for an emergency.

The data source table does not include information about financial systems due to confidentiality.

Confirm

<i-text> Select your answer then select Confirm.

Activity: Identify data sources



The office data for available office relocation metrics was provided from two vendors. The office metric data from both vendors included input into the CRM for the Sale Department to access. You have been asked to expand the data source description to include these important factors. What should you know in this situation?



Office database providing office metrics and sales data respectively to our sales team, DQI required



Office database for our CRM DQI and Sales Dept.

Office database metrics with access to our CRM DQI

Confirm



<i-text> Select your answer then select Confirm.

## Lesson 5: Identify data gaps

# Learning objectives

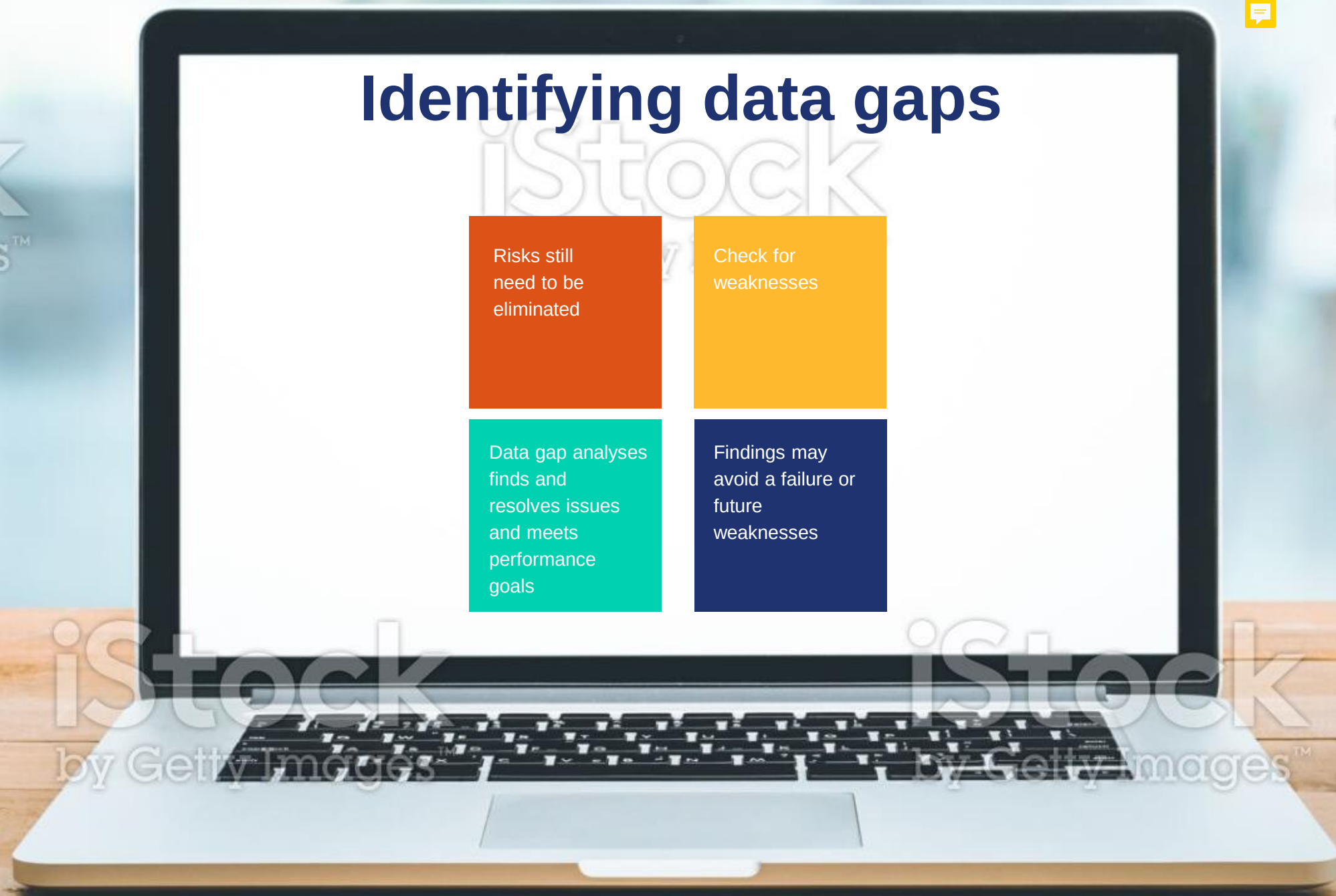
By the end of this lesson, you should be able to:

- Identify data gaps









## Identifying data gaps

Risks still need to be eliminated

Check for weaknesses

Data gap analyses finds and resolves issues and meets performance goals

Findings may avoid a failure or future weaknesses

## Activity: Identify data gaps

Jessica asked you to start looking for data gaps in data use and business purposes for the data. After you review the table below, Jessica asked you to text her your responses from the Data use column. What should you do next?

Ref #	Data Source	Data Inputs	Data use	Business Purpose
1	Sales and accounting system	Sales, revenues, costs	Financial	Provide financial systems and reports
2	Customer relationship management system	Clients, contracts, actions	Sales, productivity	Provide sales team and client relationship management
3	Net1office	Office metrics and info	Available space	Calculate new office costs for Client Report
4	Ecodev2020	Office improvements allowance cos	Improvement calculator	Calculates repayment of unamortized tenant improvement allowance
5	Subfindnow	Office metrics and info	Available space	Calculate new office costs for Client Report
6	Datadrop	Client office leases	Termination provisions	Calculate liability for Client Report

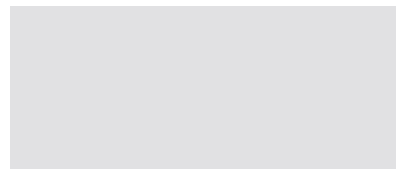


Confirm

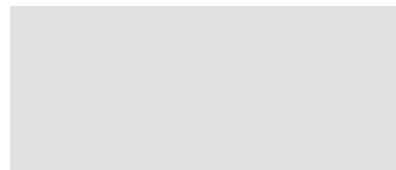


Activity: How company KPIs influence data gaps

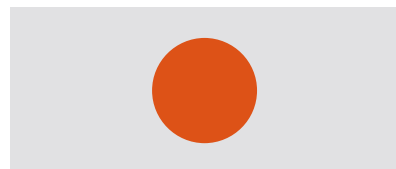
You can see that codev2020 is the inactive source for this data and that this data is missing from the Report's liability calculation. What should you select to resolve this issue?



Form an action plan to estimate improvements internally and incorporate estimates in calculations



Form an action plan to research office lease termination provisions



Form an action plan to activate ecodev2020 and incorporate their missing data in calculations

Confirm



<i-text> Select your answer then select Confirm.



## Knowledge check - 1

You can see that codev2020 is the inactive source for this data and that this data is missing from the Report's liability calculation. What should you select to resolve this issue?

Sales, revenues, costs



Office metrics and information

Clients, contacts, actions

Confirm



<i-text> Select your answer then select Confirm.

## Lesson 6: Build a basic data flow



# Learning objectives

By the end of this lesson, you should be able to:

- Build a basic data flow







# Building a basic data flow

Now that you have identified the data sources and have resolved the data gaps, the next step is to build a Data Flow Diagram.

## Build a Data Flow Diagram (DFD)

Provides a view of data flow

DFDs reveal strengths and weaknesses

Logical DFD: What occurs in the data flow to perform key functions of a solution

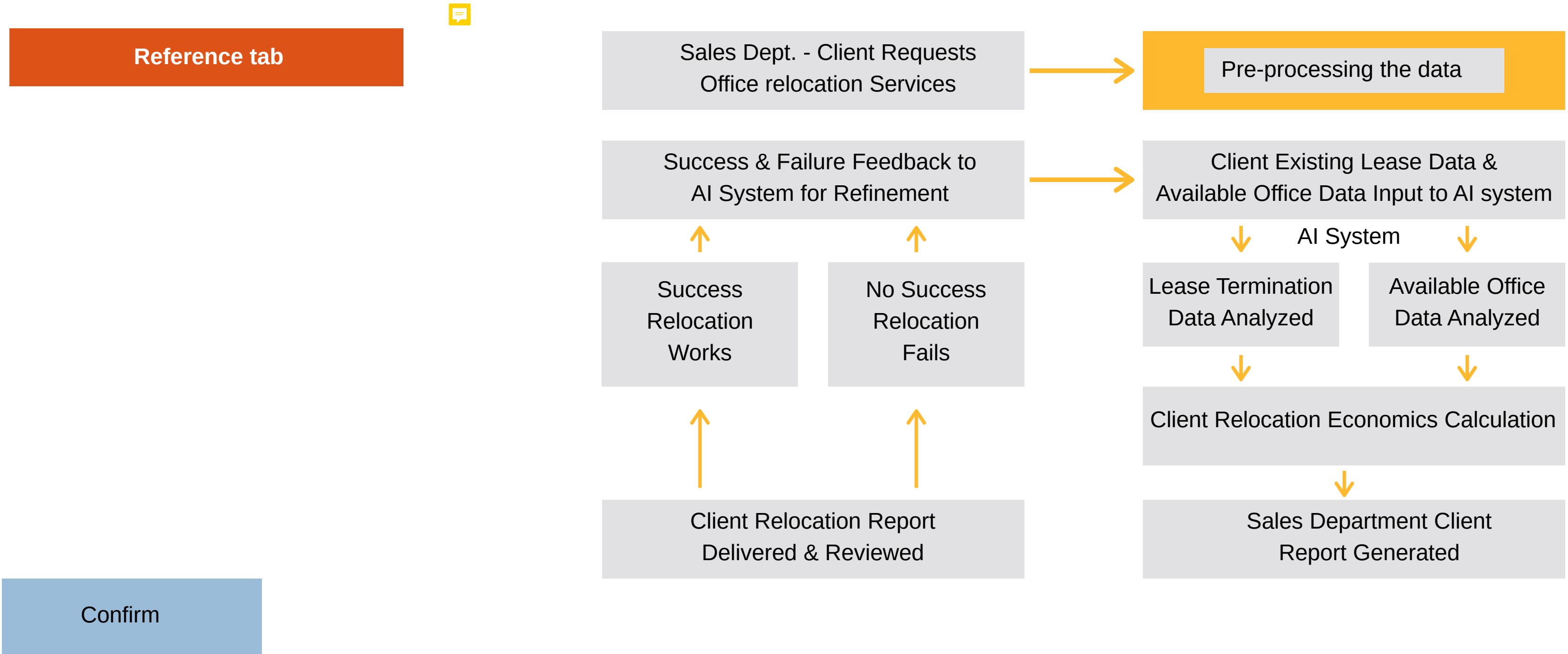
Physical DFD: Implementation of data to a system






Activity: Build a basic data flow

You started creating a simple Data Flow Diagram for an AI solution. You drew a step between the Client request for a Report and the Client data input into the system. You need to complete the diagram. Jessica asked you to text the correct response. What do you perceive as the logical data flow step between the Client request and the Data Input into the AI system?





Knowledge check - 1 

Consider taking a closer look at data flows. It is important that you know the difference between logical and physical DFDs. Jessica has requested that you provide her with information through text messaging about the data flow and decide which DFD type each data flow is: logical or physical.

Data flow	Type
Sales Dept. – Client Requests – Office Relocation Services	Physical
Pre-processing the data	Logical
Client Existing Lease Data & Available Office Data Input to AI System	Physical
Available Office Data Analyzed	Physical
Client Re-location Economics Calculation	Logical
Sales Department Client Report Generated	Physical
Client Relocation Report – Delivered & Reviewed	Physical
No Success – Relocation Fails	Logical
Success – Relocation Works	Logical
Success & Failure Feedback to AI System for Refinement	Logical

Confirm

Select your answer then select Confirm.

# Lesson 7: Pre-process data

# Learning objectives

By the end of this lesson, you should be able to:

- Pre-process data





ock  
images™

iStock  
by Getty Images™

iStock  
by Getty Images™

iSto  
by Getty

iStock  
by Getty Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

ock  
Images™

iStock  
by Getty Images™

iStock  
by Getty Images™

iSto  
by Getty



## Data and human supervision

Supervision:  
Human  
processing  
step

Data pre-  
processing starts  
with pre-  
processing text

Tokenization,  
lower casing,  
stop word  
removal ... are  
required for  
unstructured  
documents and  
meta-data.

Structured text  
needs less data  
pre-processing

Less structured  
text requires  
more robust  
pre-processing  
with human  
supervision



## Knowledge check - 1

Data types that you should examine include: office lease analysis, improvement analysis, and office space metrics. Jessica asks you to consider raw data types. Jessica indicated that she is making business decisions about the types of data that will require human supervision or augmentation to provide optimal results. What are your recommendations? Go ahead and check out the reference tab for additional help.

Reference tab

- Office lease analysis
- Improvement analysis
- Office space metrics

Confirm



<i-text> Select your answer then select Confirm.



# Skillbuilder

## Summary

Let's take a look at what you have learned.

Data quality is critical for AI success and performance

Robotics Process Automation (RPA) seeks and delivers targeted data to an AI system.

Machine Learning (ML) is a type of AI in which a computer uses iterative data processing and intelligent algorithms to learn from data.

Natural Language Processing (NLP) uses ML with text or voice data to synthesize amazing solutions with human language data.

If data types are misapplied, or data errors or poor data quality exists, even an excellent AI system may fail.







You have completed the Skillbuilder on creating a high-quality data set.

You will be presented with seven multiple choice questions. Once you have selected your answer, then select Confirm and move to the next question.

You have 10 minutes to complete the assessment. Select Begin to start your assessment.

Begin



### Assessment question 1

Imagine that you work for the Sales Department at Palo Verde Valley Realty. Your manager wants to provide clients seeking a Client Relocation Report with an information portal to share preliminary office metrics during the phase where clients decide on new office preferences for relocation. The IT Manager requires all clients to have a random generated 12-character password. The first 5 client passwords are generated and are listed below. After analyzing the passwords, you need to indicate the data type.

<input checked="" type="radio"/>	Useless
<input type="radio"/>	Nominal
<input type="radio"/>	Encrypted
<input type="radio"/>	Interval

- 49105x5\$hYsZ
- tcUNd2bSt58U
- RmNx!++vvE!b
- pPpJ#HG-Zr-\$
- VAqvSd+GkM5V



Confirm



<i-text> Select your answer then select Confirm.



## Assessment question 2

As an IT Manager, you were recently advised of several blatant errors within 6 Office Lease Contracts, being truncated sentences with many missing words. The same error appears in every lease contract's PDF file for this one specific Office Building Management Company, for all 6 leases. Instead of flagging 6 documents for errors to repair, you need to determine the root cause of this poor data quality, so the errors are not continually produced. To find and resolve the root cause, you need to search more places at the office and call the landlord who creates the lease forms. Where should you look for these errors?

- Palo Verde's email server to find where the leases were transmitted electronically
- The landlord's lease contract inputs, software, system, and formatting
- The landlord's data backup system where the final lease data is stored
- The Palo Verde transaction assistant who facilitated the final lease execution



Confirm



<i-text> Select your answer then select Confirm.



### Assessment question 3

Jessica just noticed that the IT Manager has cc'd her on an email to an office building landlord regarding some lease contract errors. The Sales Department is also seeking and resolving errors. The CFO is concerned that each person may be resolving errors but not sharing the information with each other. You are tasked with determining the data quality characteristics of an organization that best describes a solution for this issue. Where should you research?

- Palo Verde's email server to find where the leases were transmitted electronically
- The landlord's lease contract inputs, software, system, and formatting
- The landlord's data backup system where the final lease data is stored
- The Palo Verde transaction assistant who facilitated the final lease execution



Confirm



Select your answer then select Confirm.



#### Assessment question 4

The IT Manager has discussed the repeating data errors in lease contracts with the office landlord. They agree that the root cause of the error is in the original contract template input data from one of the landlord's many law firm counsels. The landlord wishes to contact the legal counsel and retrieve the correct language missing from the lease template but is unsure of which law firm prepared the source document. A data source table would have saved time and resources in tracking down this source. The IT Manager enlists your help to determine what data source information would be most useful for the landlord to have.

- A system/application name so they may search for law firms using the system
- The URL/Path simply provides the link to the system's end user for quick access
- Business contact for business application or license questions
- A reference number to provide an accurate method for locating the data source



Confirm



<i-text> Select your answer then select Confirm.



## Assessment question 5

The business is quickly adopting resources and methods to improve its data quality so the AI solution performance will be optimized. When seeking to improve the performance of the AI solution, data gaps may be identified and resolved to improve the quality of AI output and performance, such as when you located and resolved a data gap for calculating lease termination liability. What do you know about the primary purpose for data gap identification and analysis?

Establishes a data gap reference document for existing data, data sources and purpose

A methodology for understanding data gaps and their value to system performance

A systematic approach to finding and resolving missing data to meet performance goals

Confirm

<i-text> Select your answer then select Confirm.



### Assessment question 6

The Data Flow Diagram provides a view of data flow through organizations and systems, giving a greater understanding for process steps, functions and transfer of data to meet important objectives in providing client relocation reports in this scenario. You need a combination of a logical DFD and physical DFD, and as a result, enhanced AI system performance by adding a data processing step. You are asked to describe the differences between a logical DFD and physical DFD?

Logical DFD describes a logical progression for data inputs. Physical DFD primarily involves data outputs.



Logical DFD describes what happens in data flow to perform key functions. Physical DFD involves data implementation into a system.

Logical DFD describes preprocessing steps. Physical DFD describes whole system performance.

Logical DFD describes data error processing and quality elements. Physical DFD describes the physical attributes of an AI system.

Confirm

Select your answer then select Confirm.



### Assessment question 7

Natural Language Processing (NLP) is an AI technology used for text analysis use cases, such as contract analysis. Machine learning (within NLP) may be used to automatically capture and process meta-data and less structured data, which requires greater data pre-processing resources. After building a Data Flow Diagram for the AI system, Palo Verde Valley Realty realized there were certain data processing and other tasks which may be best performed, or supervised, by humans, collaboratively with the AI system. Jessica asks you to describe the characteristics of data or data processing for human supervision or collaboration with the AI system.



Office space metrics with highly unstructured text, schedules, calculations, and images

Capture, scan, analyze, and process highly structured data all in a single system

Both highly unstructured and highly structured data

Office lease contract data containing termination liability provisions

Confirm



<i-text> Select your answer then select Confirm.



Congratulations!

You have finished the assessment.

You scored <Learner's %> of the required <PASSING%>.

You answered X out of 7 questions correctly.

<l-text>: Select X to exit.

Sorry!

You did not pass the assessment.

You scored <Learner's %> of the required <PASSING%>.

You answered X out of 7 questions correctly.

Click Retake Assessment to try again.

Retake  
Assessment

<l-text>: Select X to exit.