Data Analytics Foundations

Module 4:
The data analytics lifecycle

DeepLearning.AI

The data analytics lifecycle

Module 4 introduction

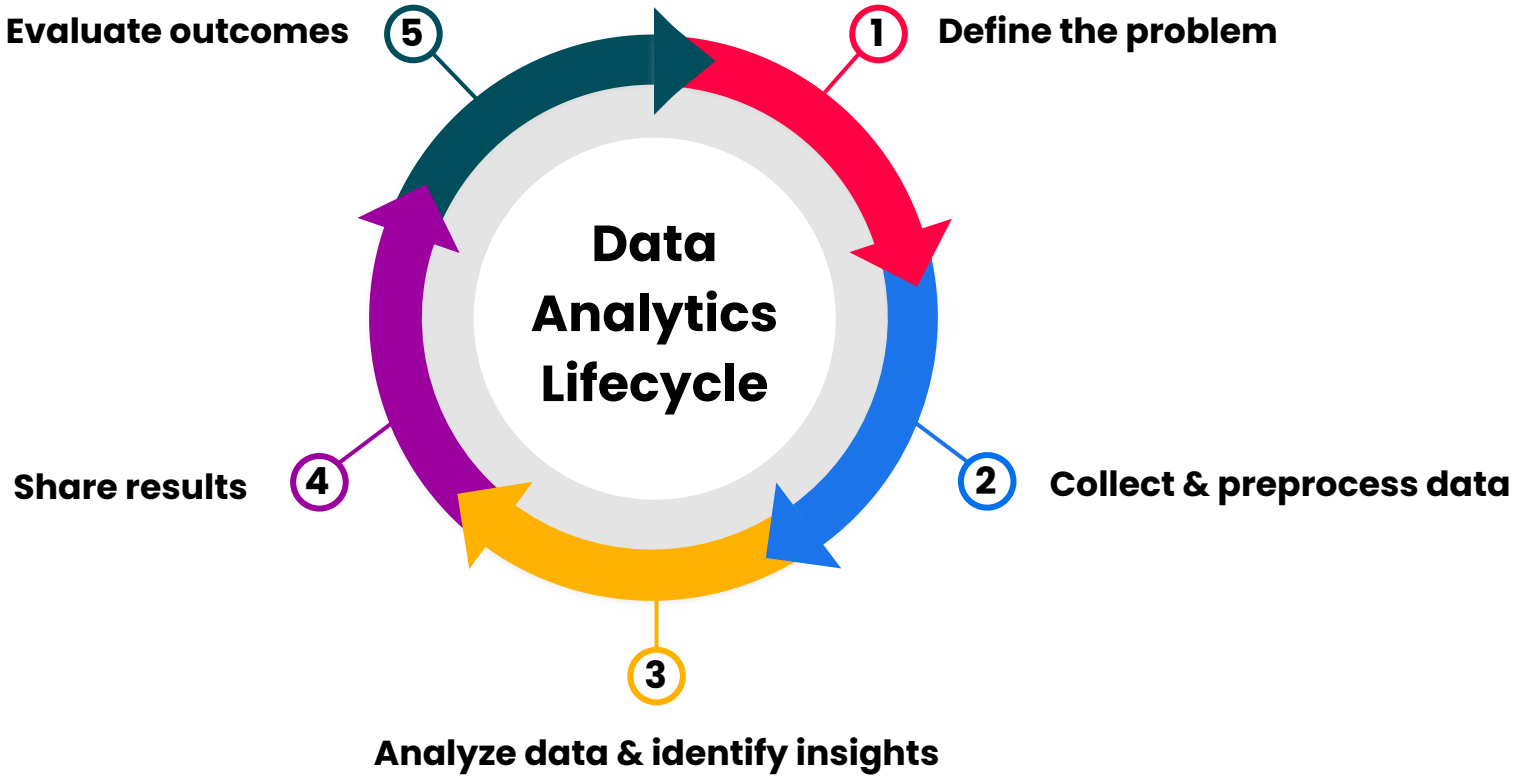
DeepLearning.AI

- **The data analytics lifecycle**
  - Each stage: defining problem to evaluating impact
- **Working with stakeholders**
  - Domain knowledge
  - Identify stakeholders
  - Determine goals
  - Check in
- **LLMs for stakeholder analysis**
  - Reconciling conflicts
  - Prioritizing
  - Developing check-in questions
- **Capstone exercise**
  - Analyzing data to reduce cancellations

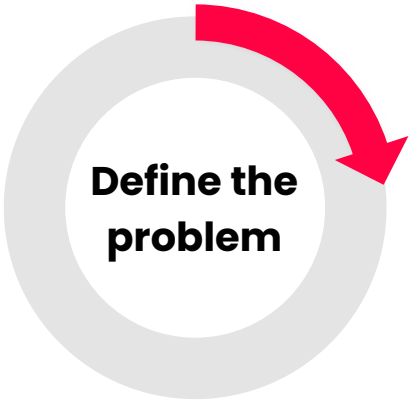
The data analytics lifecycle

The data analytics lifecycle

DeepLearning.AI


Evaluate outcomes
5
1
Define the problem
Data
Analytics
Lifecycle
Share results
4
2
Collect & preprocess data
3
Analyze data & identify insights

The data analytics lifecycle



The point:

1. Narrow decision space
2. Set expectations about success

Rule out incorrect approaches

↓

Focus on most fruitful ideas

Key questions:

- What are the **business goals**?
- What **decisions need to be made** to achieve those goals?
- Who are the **stakeholders** making those decisions, and what are their **needs**?
- What does **success** look like for this project?

"If I had an hour to solve a problem, I'd spend **55 minutes** thinking about **the problem** and **5 minutes** thinking about **solutions**."

- Albert Einstein

The data analytics lifecycle

- **Transform** raw data into usable information
  - Most data requires some transformation

Challenges

Missing information

Poor formatting

Unavailable data

Quality of **data** → Quality of **insights**

The data analytics lifecycle

Statistics

Data visualization

Machine learning

Patterns in data

Data

Insights

Story

What do they mean in context?

How do they help inform decisions?

The data analytics lifecycle

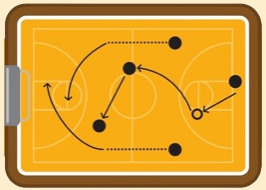
Communicating findings effectively empowers stakeholders to:

- ✅ Make informed decisions
- ✅ Achieve the project goals

The data analytics lifecycle

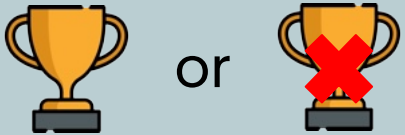
Analysis → Decision → Impact?

Insights



New game plan

Positive change



Outcome

Ask questions like:

- Did the decision lead to the **successful outcome**?
- Are the **stakeholders happy** with the results?
- What is the **long-term impact**?

The data analytics lifecycle

Defining the problem

DeepLearning.AI

1 What are the business goals?

Most businesses: Increase sales | Reduce costs | Improve satisfaction

More specific objective: Better product recommendations

2 Who are the stakeholders and what are their needs?

Who will be using your results?

- Decisions
- Information

3 What are the key unknowns?

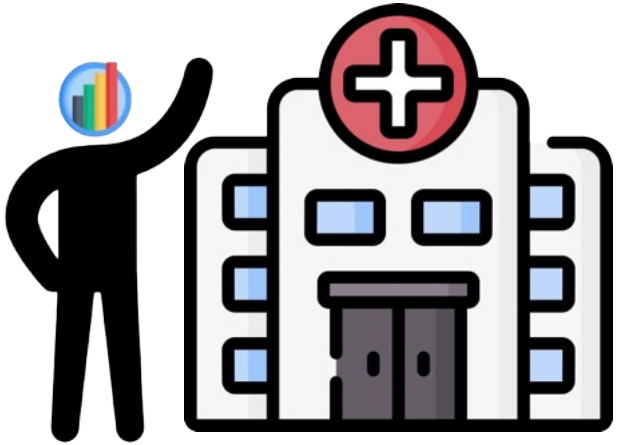
Examples:

Most effective marketing channels

Cancellation factors

Scenario

① **Business goal**: Reduce patient readmission rates



You
Data Analyst

- **Problem**: Increase in patient readmission rates for pneumonia
  - Low readmission rates are better
  - Patients want problem fixed first time
  - Repeated visits put a strain on staff
- Hospital wants to understand factors contributing to this issue

Scenario

1. **Business goal**: Reduce patient readmission rates
2. **Stakeholders:** people with vested interest in outcome of analysis

Stakeholder	Needs
Patients and families	Receive best possible care
Doctors & nurses	Identify improvements for care
Administrators	Allocate resources effectively

Scenario

1. **Business goal**: Reduce patient readmission rates
2. **Stakeholders:** people with vested interest in outcome of analysis

Stakeholder	Needs	Actionable insights
Patients and families	Receive best possible care	Learn about key risk factors
Doctors & nurses	Identify improvements for care	Explain at-home care
Administrators	Allocate resources effectively	Adjust scheduling

Finding: 10 additional minutes of at-home care education leads to improved patient outcomes

Scenario

1. **Business goal**: Reduce patient readmission rates
2. **Stakeholders:** people with vested interest in outcome of analysis

Stakeholder	Needs	Actionable insights
Patients and families	Receive best possible care	Learn about key risk factors
Doctors & nurses	Identify improvements for care	Explain at-home care
Administrators	Allocate resources effectively	Adjust scheduling

3. 💭 Most common **reasons** for readmission?
   💭 **Patterns** in the demographics, history, or treatment?
   💭 **Gaps** in at-home care?
   💭 What **interventions** could be implemented?

Collaborating with stakeholders

The data analytics lifecycle

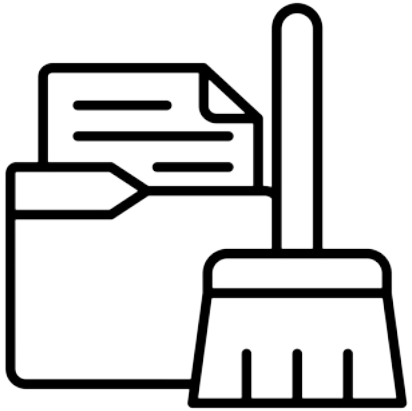
Collecting &
preprocessing data

DeepLearning.AI

Stage 2: Collect & preprocess data

Step 1: Data collection

Identify sources of information



Step 2: Data preprocessing

Transforming your data for analysis

Strategies for data collection

Revisit problem statement

Problem/question	Outcome of interest	Data that puts outcome in context
 **Customer retention**	Rate at which customers cancel	• Demographics • Purchase history • User engagement

Strategies for data collection

	Yes	No
Accessibility:	☐	☐
Quality:	☐	☐

Brainstorm potential data sources

Internal databases

Publicly available datasets

Surveys

Experiments

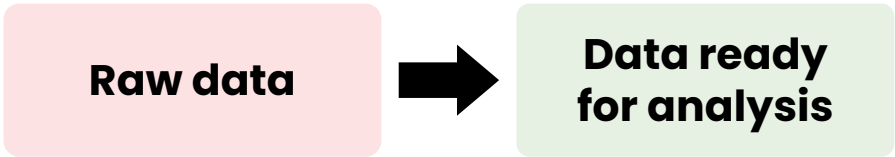
Prioritize sources that:

- Yield valuable insights
- Keep project within budget and on time

Ask domain experts:

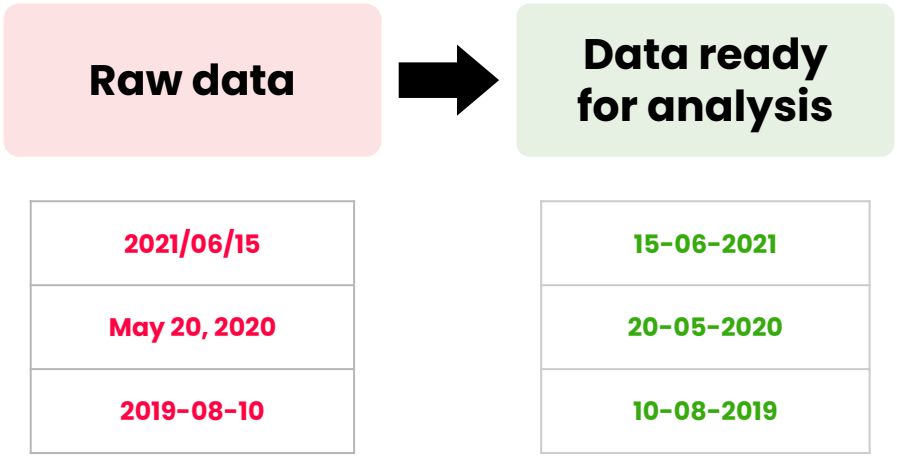
Identify data typically used to answer similar questions

Data preprocessing



Data preprocessing

Formatting

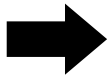


Data preprocessing

Formatting

Cleaning

Raw data



Data ready for analysis

John Doe	25	2021/06/15
John Doe	**25**	**2021/06/15**

John Doe	25	2021/06/15

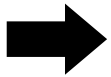
Data preprocessing

Formatting

Cleaning

Handling missing values

Raw data



Data ready for analysis

Emily	28	**UNKNOWN**

~~Emily~~	~~28~~	~~**UNKNOWN**~~

Data preprocessing

Formatting | Cleaning | Handling missing values | Transformation

Raw data → **Data ready for analysis**

Membership Status
Active
Inactive
Active

Membership Status
1
0
1

The data analytics lifecycle

Analyzing data

DeepLearning.AI

Common techniques for data analysis

Descriptive statistics

Summarize different features in data

- Frequency
- Mean
- Median
- Correlation

Data visualization

Visual summary of data

Identify trends not apparent from descriptive statistics

Statistical analysis and modeling

Evaluate hypotheses about data

Understand relationships between features

Machine learning

Sophisticated algorithms learn from data & make predictions

Less complex

More complex

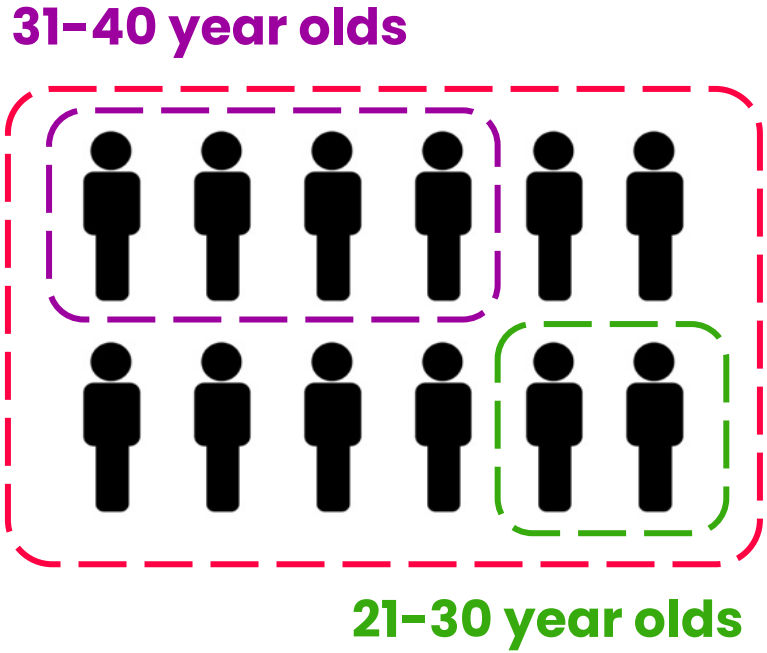
Descriptive statistics 📈

Demographic data

- Age
- Income
- Education level

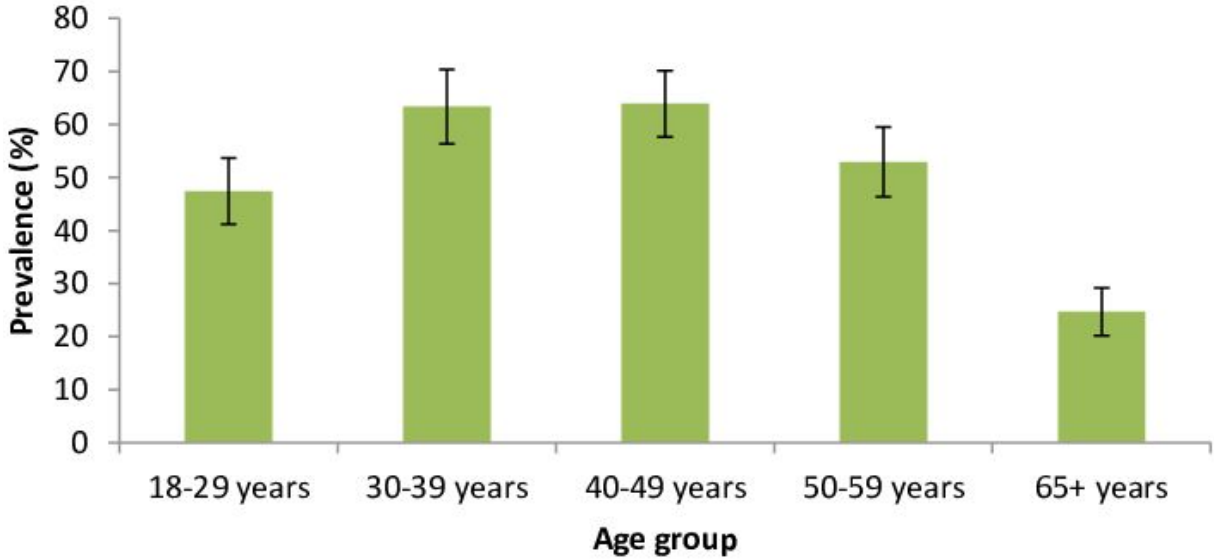
Health-related data

- Body mass index
- Smoking status
- Exercise habits



Data visualization

- **Summarize** the descriptive statistics you calculated

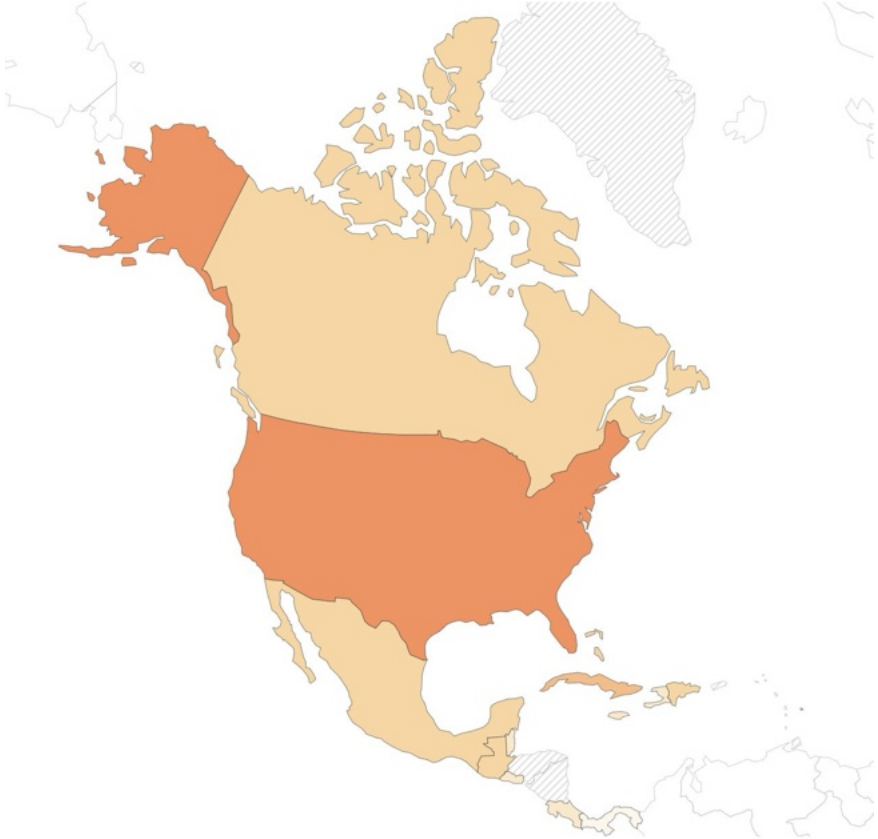


[Roberts, et al., Tobacco Use and Nicotine Dependence among Conflict-Affected Men in the Republic of Georgia., Int J Environ Res Public Health, (2013)]

Data visualization

Data visualizations really shine for:

- Complex or large data sets



[Our World in Data. (n.d.). Share of adults who smoke, 2020. Our World in Data]

Statistical analysis & modeling 🔍

- Allows you to be more confident that results are:
  - ✅ True effects
  - ❌ Random chance
- Tailor interventions to address specific risk factors

Hypothesis test

Determines if the differences between groups suggest a real effect

❓Statistically significant decrease in smoking prevalence in groups who received the intervention

Regression model

- Identify factors most strongly associated with smoking

Machine learning 🧠

A **class of techniques** that:

- Learn patterns from data
- Predict specific outcomes of interest

Previous analyses

Predictive model

✅ Identify the individuals most likely to benefit

✅ Target resources to those who would benefit the most

Clustering

✅ Segment the target population into groups

✅ Improve effectiveness for specific subgroups

Conducting a comprehensive evaluation

- ✅ Descriptive statistics
- ✅ Data visualization
- ✅ Statistical modeling
- ✅ Machine learning

Identifying segments most receptive to the initiative

Visualizing geographic trends over time

Identifying most significant factors contributing to smoking

Personalizing interventions to maximize impact

The data analytics lifecycle

Identifying insights

DeepLearning.AI

Framework for insight finding

Intervention Group:

- Personalized counseling and nicotine replacement therapy
- Quit Rate: **25%** after 6 months

Standard Group:

- Quit Rate: **10%** after 6 months

Statistically significant!

- **Key result**: Statistically significant difference between 25% and 10% quit rate
- **Business goal**: Reduce smoking rates among 21-30 year olds
- **Explanation**: Provided interventions are better at reducing smoking
- **Recommended decisions**:
  - Provide the intervention to all 21-30 year olds who smoke
  - Identify subset who benefited most
  - Promote awareness about benefits

The data analytics lifecycle

Sharing results

DeepLearning.AI

Sharing results

- What is the **best approach for the insights** you need to summarize?

Complex data → **visualized**

Numbers → **contextualized**

Several insights → **summarized**

New data → **contrasted**

- What are your **stakeholder needs**?

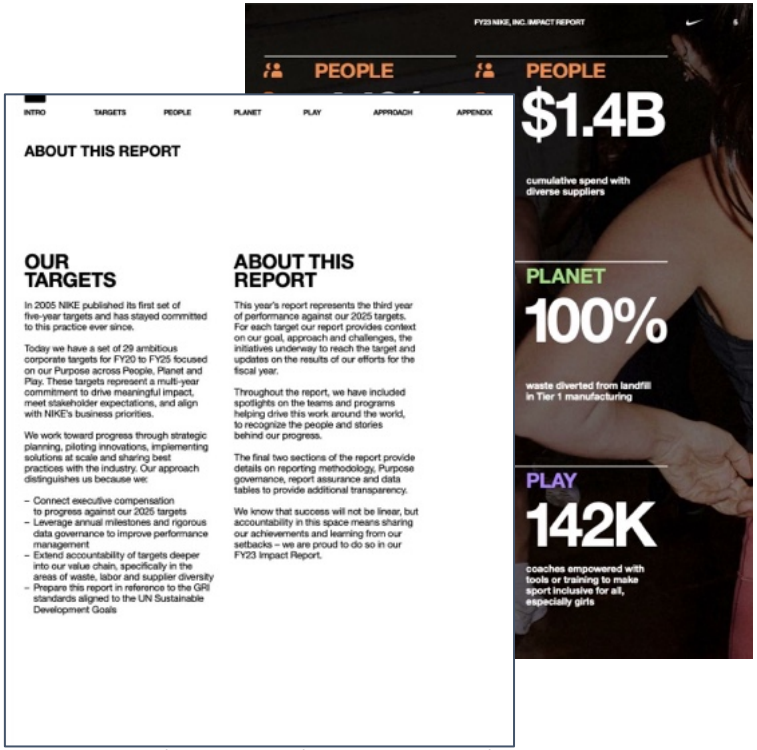
Technical knowledge

Time

Control

Reports

- Written summaries of findings
- Often include visualizations



[Nike. (2023). 2023 Nike, Inc. Impact Report. Nike]

Useful when you need to...

✅ Provide a thorough explanation

✅ Document methodology

✅ Present complex findings to a technical audience

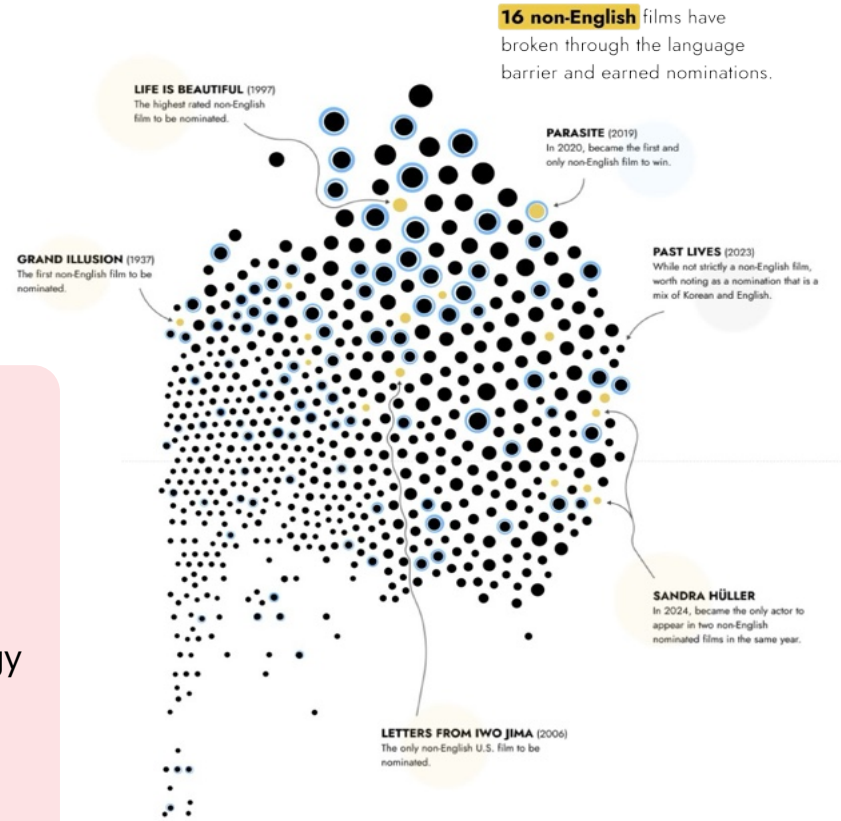
Not as useful when you need to...

❌ Communicate quickly

❌ Communicate with a non-technical audience

Dashboards

Interactive visualizations that allow users to explore data on their own



[Beccab. (n.d.). Non-English Films Breaking Best Picture Barriers. Tableau]

Useful when you need to...

- ✅ Empower stakeholders to uncover own insights
- ✅ Provide easy access to current information
- ✅ Track performance over time

Not as useful when you need to...

- ❌ Provide a detailed explanation
- ❌ Document methodology
- ❌ When interactivity isn't useful

Presentations

Live share-outs of your findings, often accompanied by visuals

[160over90. (2023). 2022-23 NFL Season Recap. 160over90]

Useful when you need to...

- ✅ Persuade your audience
- ✅ Engage stakeholders in a discussion
- ✅ Present to a large audience

Not as useful when...

- ❌ Audience needs to re-engage with information
- ❌ Audience needs different depths of explanation

Models

Deploying a machine learning model to automate decisions

Useful when you need to...

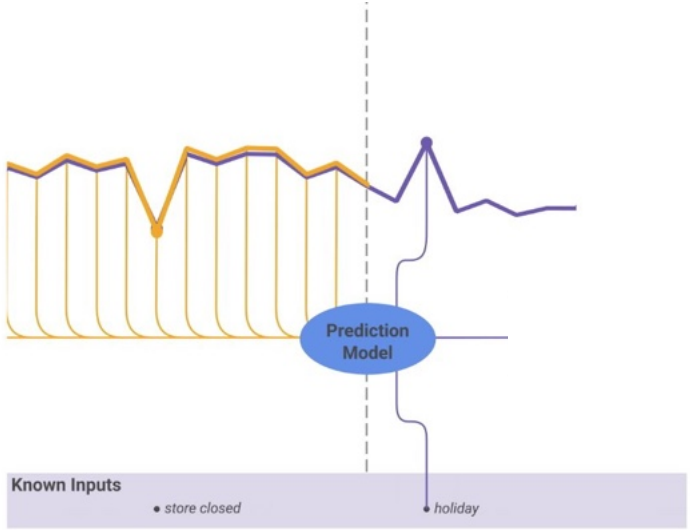
✅ Make frequent decisions based on incoming data

Not as useful when you need to...

❌ Explain reasoning behind findings

❌ Explore data interactively

❌ Communicate with non-technical audience



[Arik, S.O., & Pfister, T. (2021). Interpretable Deep Learning for Time Series Forecasting [GIF]. Google Research]

Key principles for sharing results

Clarity

- Make sure message is easy to understand
- Avoid technical jargon
- Use visuals to illustrate points
- Be aware what your audience knows

Relevance

- Focus on insights most most relevant to goals
- Resist urge to share everything you know
- Help audience understand what they need to know

Actionability

- Focus on decisions to be made
- Provide evidence-based recommendations

The data analytics lifecycle

Evaluating outcomes

DeepLearning.AI

Evaluating outcomes

Questions to ask:

- Did the decision lead to the **desired outcome**?
- Were the **stakeholders satisfied** with the results?
- What were the **key learnings** from the project?
- What could be **improved** in the future?

Did the decision lead to the desired outcome?

- Have defined success criteria

  Metrics → measurable

- **Compute** your key metrics **before** and **after** decisions were implemented

- **Monitor** relevant metrics over an extended period to assess sustainability

Your smoking prevention intervention is effective for **60%** of 21-30 year olds.

Measure % of 21-30 year olds who abstain from smoking for 6 months:

Before | Update intervention | **After**

Possible: boost in quit rates is temporary, levels out after 1 year

Assessing stakeholder satisfaction

Gather feedback through:

- Surveys
- Interviews
- Focus groups

Ask:

- What worked well?
- What could be improved?

Learning and improving from decisions

Evaluating outcomes is also about continuous improvement

[The Coca-Cola Company. (n.d.). *New Coke: The Most Memorable Marketing Blunder Ever?*]

✅ Right analysis ❌ Wrong decision

Return of Coca-Cola Classic:

📈 Led to a surge in sales

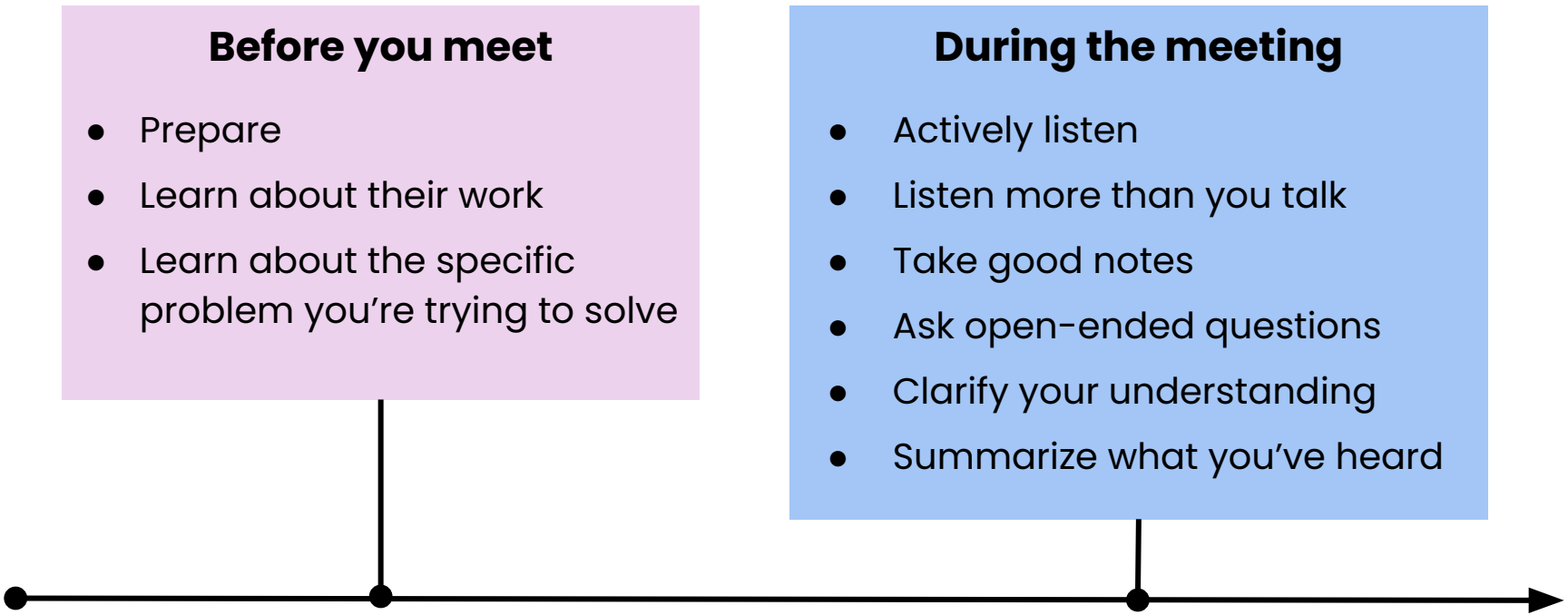
🏆 Solidified the brand's cultural significance

The data analytics lifecycle

Gathering stakeholder requirements

DeepLearning.AI

A structure for getting to know stakeholders



Questions for getting to know stakeholders

What are your business **goals**?

- Challenge with respect
- Make them articulate how their work impacts the business

How do you think about **success**?

- Define their vision for being successful
- Help develop a metric

What are the biggest **challenges** you're facing?

- Connect challenge to an important business goal

Questions for getting to know stakeholders

What **decisions** do you need to make based on this analysis?

- Determine type of analysis to perform
- Design output of analysis

What **risks** do you anticipate with this work? What would lead you to **not use** this analysis?

- Tough questions
- Ensure you use your time well

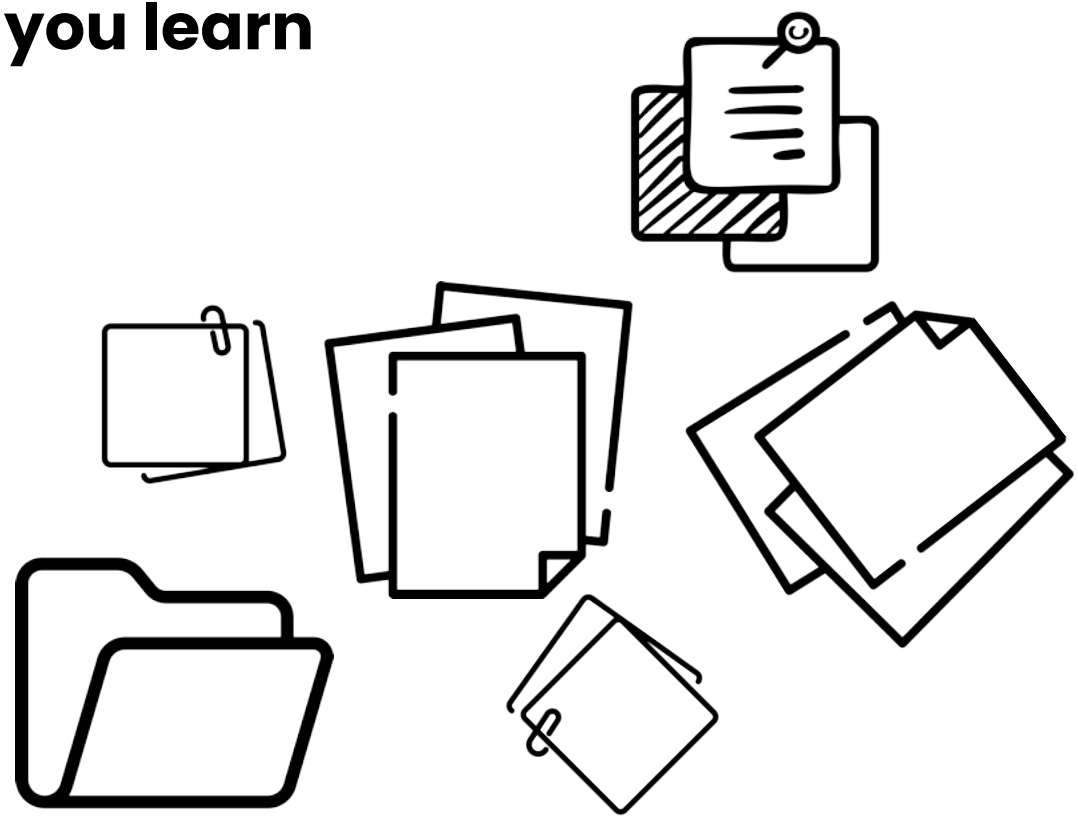
The data analytics lifecycle

Synthesizing stakeholder input

DeepLearning.AI

Synthesizing what you learn

1. **Organize**



Synthesizing what you learn

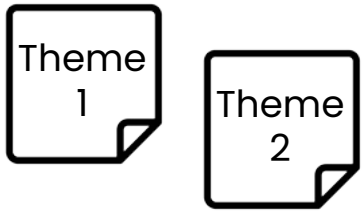
1. **Organize**

Synthesizing what you learn

1. **Organize**

2. **Identify themes**

- Look for recurring ideas

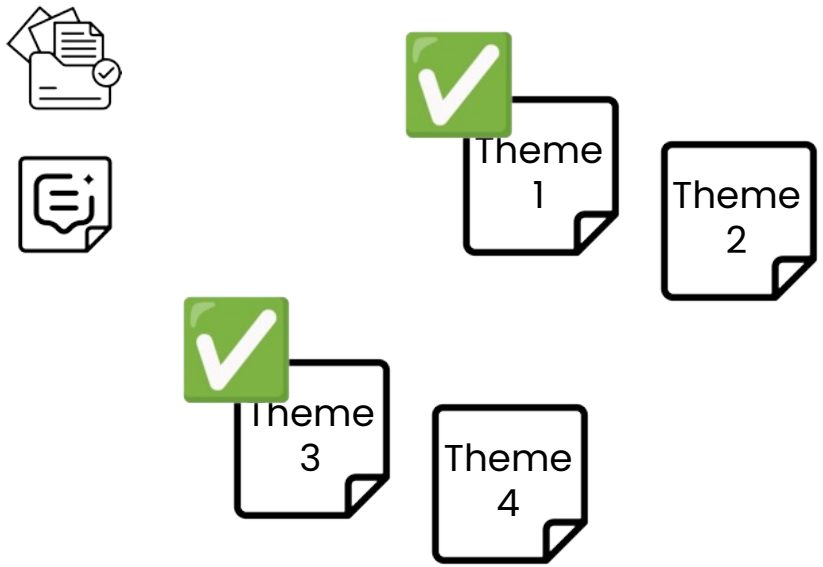


Sticky notes

LLMs

Synthesizing what you learn

1. **Organize**
2. **Identify themes**
3. **Prioritize**



Synthesizing what you learn

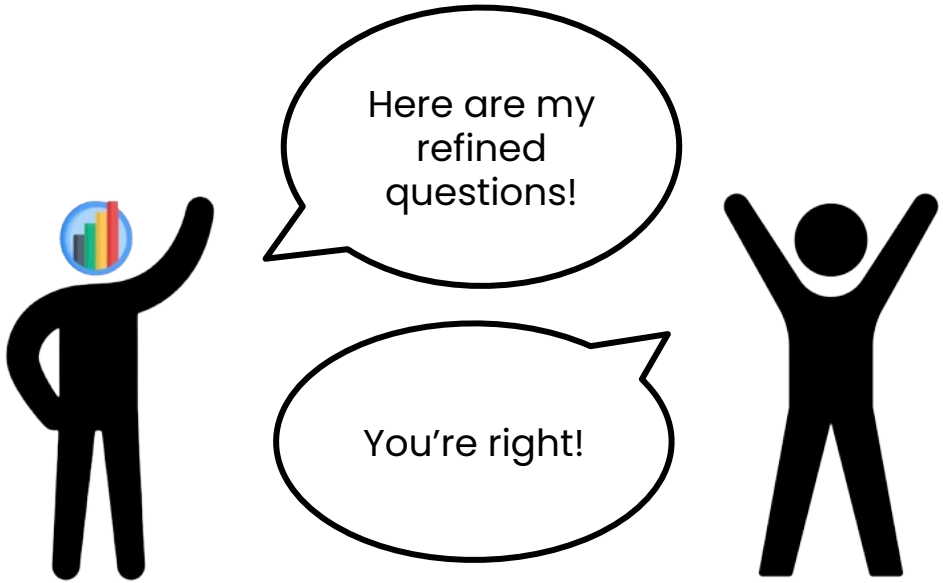
1. **Organize**
2. **Identify themes**
3. **Prioritize**
4. **Refine questions**

Specific
Measurable
Achievable
Relevant
Time-bound

S M A R T

Synthesizing what you learn

1. **Organize**
2. **Identify themes**
3. **Prioritize**
4. **Refine questions**
5. **Validate themes**



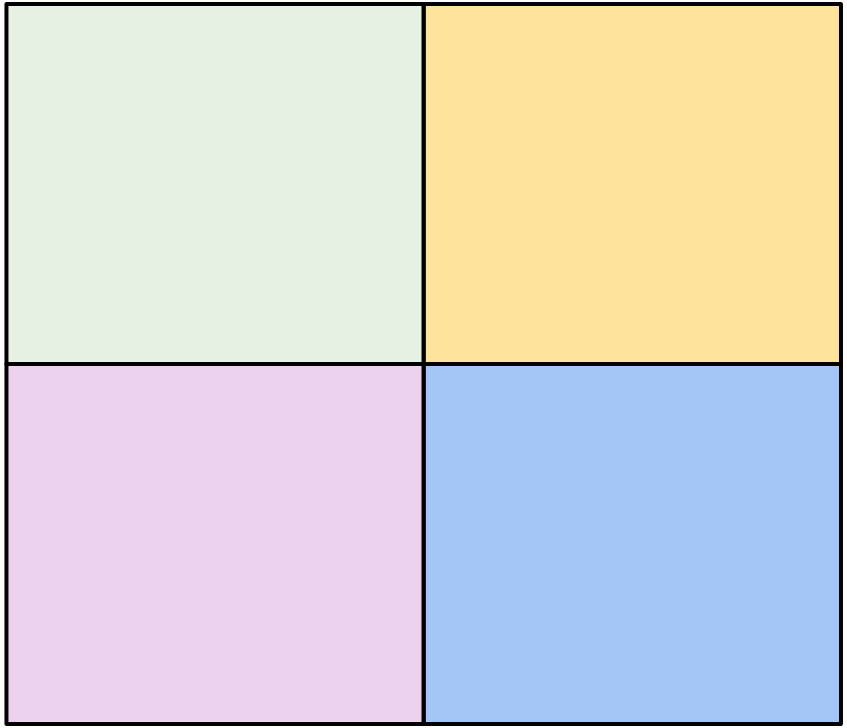
The Rumsfeld matrix

Known knowns

Known unknowns

Unknown knowns

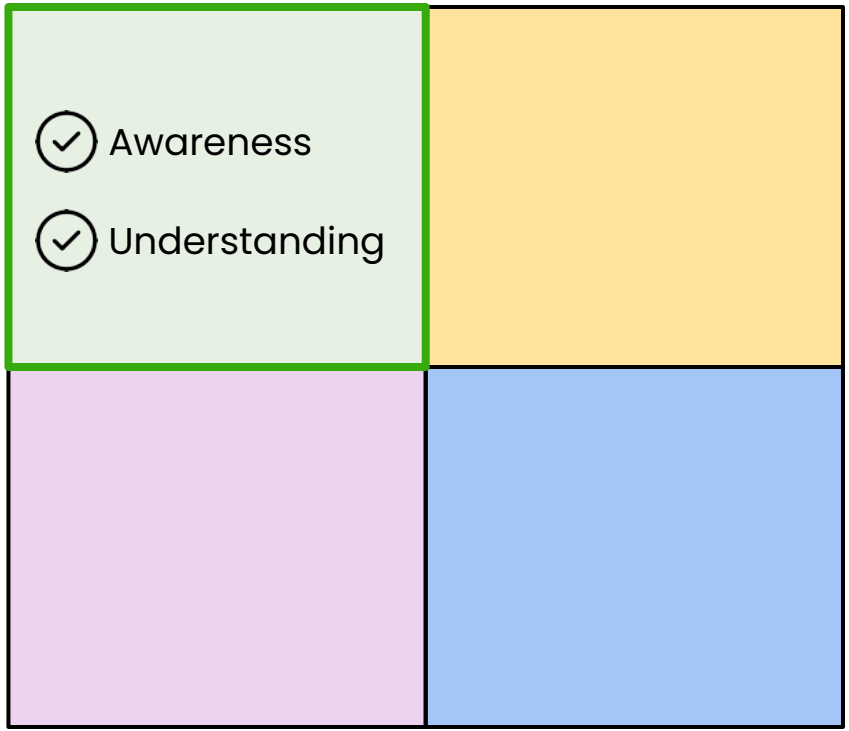
Unknown unknowns



The Rumsfeld matrix

Known knowns

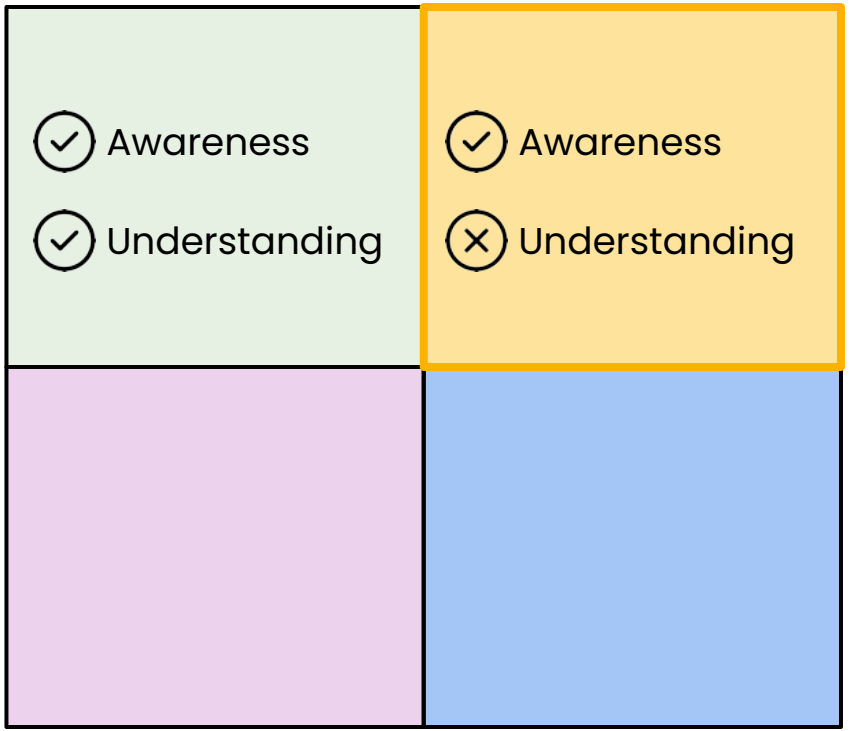
- Facts, information, and insights
- Foundation of your decision-making



The Rumsfeld matrix

Known unknowns

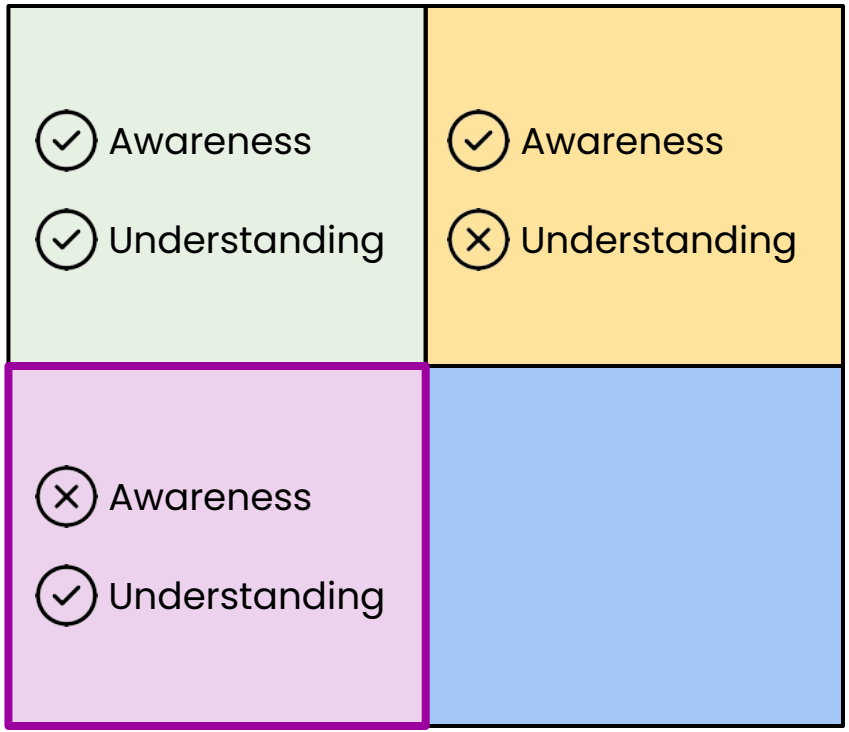
- Certain you don't know something
- Gaps in your knowledge that you need to address



The Rumsfeld matrix

Unknown knowns

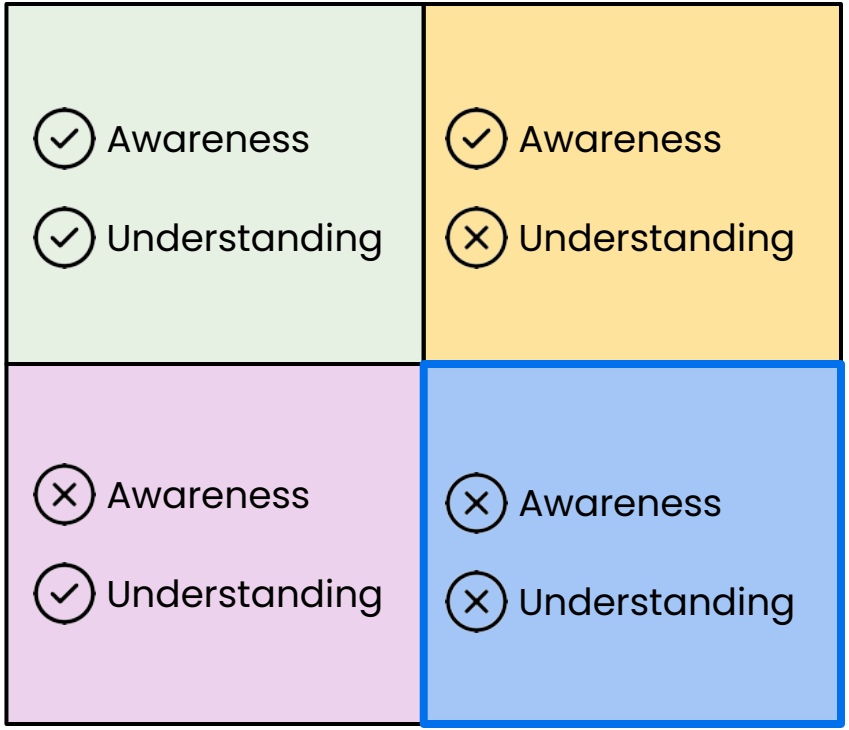
- Information you posses but haven't articulated
- Intuition, experience, anecdotal evidence



The Rumsfeld matrix

Unknown unknowns

- Unpredictable or unforeseen factors
- Risks that you are not yet aware of



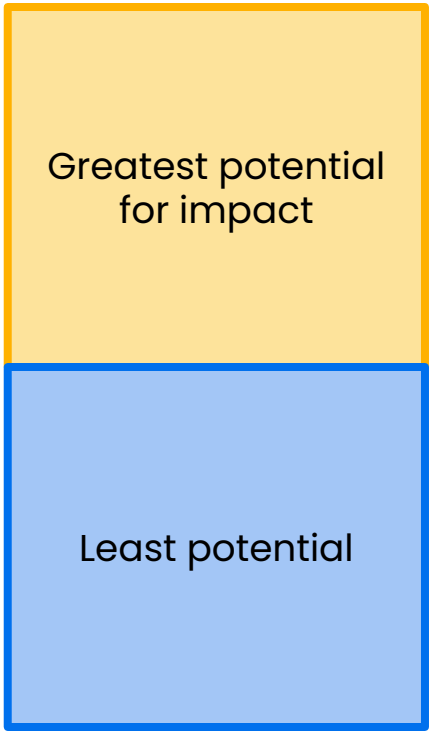
Greatest and least impact

Known unknowns

- Represent actionable questions
- Generate insights that address problem
- Mitigate potential risks
- Allocate resources more efficiently

Unknown unknowns

- Difficult to factor into your analysis



The data analytics lifecycle

Checking in
with stakeholders

DeepLearning.AI

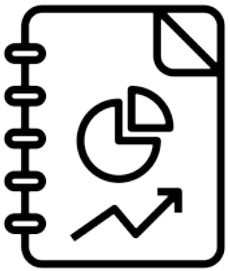
Why checking in is so important

- Ensures alignment
- Prevent rework
- Build trust
- Navigate roadblocks
- Provide continuous value

Why checking in is so important

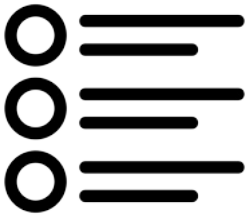
Ensures alignment	You and stakeholders are on the same page
Prevent rework	
Build trust	Demonstrates that you value stakeholders' input
Navigate roadblocks	2 brains > 1
Provide value	Make progress continuously visible

Be prepared



Concise summary

Accept feedback



Identify next steps

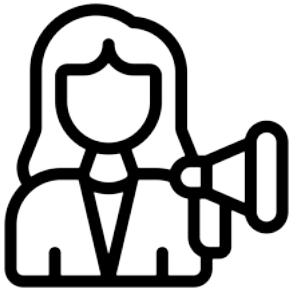
Create a collaborative atmosphere

When and how to check in

Project complexity	More complex	More frequent check-ins
Stakeholder availability	Consider schedules	Try not to overwhelm
Communication preferences	Meetings	Emails
	Quick calls	

Exotic pet store check-ins

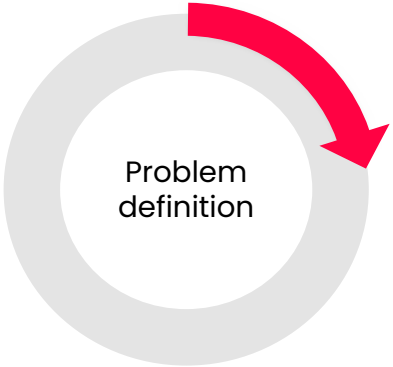
Store owner



Marketing manager

Customer service

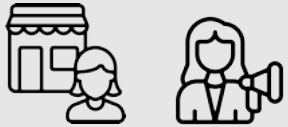
Exotic pet store check-ins



Check-ins

Initial meeting

Follow-ups



Outcomes

- Brainstorm of potential solutions
- Store's goals, audience, budget

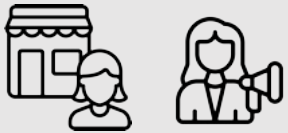
Increase customer retention and lifetime value through **loyalty program**

Exotic pet store check-ins

Check-ins

Check-in

Check-in



Outcomes

- Qualitative feedback:
  - Common pain points and preferences
  - Suggestions
- Identifying available customer data

Exotic pet store check-ins

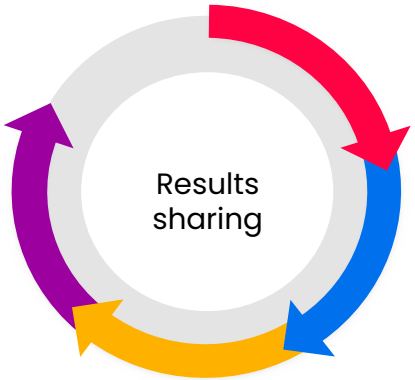
Check-ins

Check-in

Outcomes

- Share preliminary findings
- Problem solve around any unexpected challenges

Exotic pet store check-ins



Check-ins

Comprehensive presentation

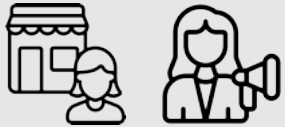
Outcomes

- Loyalty program structure
- Potential rewards
- Communication strategies
- Gather feedback on the proposed program

Exotic pet store check-ins

Check-ins

Check-in



Check-in after launch

Outcomes

- Discuss implementation plan
- Gather input on challenges and risks

- Monitor performance
- Gather feedback
- Continuously improve the program

The data analytics lifecycle

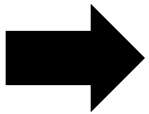
Domain knowledge

DeepLearning.AI

Domain knowledge

Understanding of specific field or industry

- Language
- Processes
- Pain points
- Goals



Ask the most relevant business questions

Identify insights in context

Choose the right analysis

Validate your analysis

The data analytics lifecycle

Demo: LLMs for stakeholder analysis

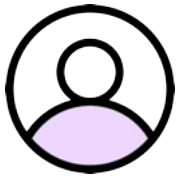
DeepLearning.AI

Scenario

You
Data Analyst

✅ **Current task:** analysis of of all hiking and biking trips in the past 5 years

Stakeholders:



Engineering Lead

VP of Sales

🏆 **Goal:** synthesize input and develop a proposed approach

Data Analytics Foundations

Your next steps
in data analytics

DeepLearning.AI