data.world

The Modern Data Project
Checklist

data.world

Data projects can get messy fast.

That's why we created this checklist to break them down into a clear, repeatable process. Try it on your next project. See how it clarifies questions, gets everyone on the same page, helps them find and use more valuable knowledge, and delivers a better analysis.

This sequence is modeled on the practices of the world's most accomplished data teams. But every company is different, and no two data projects are exactly alike. As you work through the checklist, adapt it to your needs.

If you'd like extra detail on any step or suggestion, we're happy to help. We'd also love your feedback! Did this help you as much as it has helped so many other businesses on data.world? How did you tweak it to match your unique team? Let us know: research@data.world

Use the checklist as an editable pdf, typing in text fields and checking boxes as you go. You also have plenty of writing room if you prefer to work with pen and paper. Either way, this checklist is ready for your team to immediately put to use.

Hint: *copy this checklist into your favorite task management tool to manage the workflow.*

WRITE DOWN THE QUESTION OR HYPOTHESIS:

REVIEW THE CONTEXT:

I know HOW this relates back to a business goal.

I know WHY the question was asked.

I know WHO will use the analysis.

I know WHAT they will do with the analysis.

I know WHEN they need the analysis.

CLARIFY THE QUESTION:

Shared definitions and clear questions reduce friction and back-and-forth between stakeholders, increase the number of people who can benefit and contribute to the analysis, and align results to expectations.

It is only ONE question (if not, separate them).

There is little chance that another stakeholder will misinterpret the question.

All ambiguous terms are defined.

There are no acronyms.

It is concise.

It can be answered with data you have or data you can get.

PLAN

IDENTIFY THE STAKEHOLDERS:

ROLE	CONTRIBUTION	NAME(S)
Requester(s)	*Asks the question*	
Subject Matter Expert(s)	*Understands the market, product, or topic involved; not necessarily technical*	
Data consumer(s)	*Makes decisions from the analysis produced; often the same as the requester*	
Data steward(s)	*Manages the data needed, or knows where to find it*	
Data practitioner(s)	*Answers the question using data and related context*	

Step 2: PREP DATA

BUILD YOUR BANK OF RESOURCES:

Determine what data you need, who owns it (data stewards), and where it currently lives.

Organize your data by grouping related files into datasets.

Place datasets into a shared environment for access, analysis, and discussion.

Schedule data updates if it will require refreshing over course of project.

Review usage, security, and licensing restrictions.

Clean and normalize the data. Use internal fact tables, ontologies, or directories to streamline this work where you can.

MAKE YOUR DATA EASY TO UNDERSTAND:

Briefly describe each file to help others understand the inputs.

Create a data dictionary that defines columns in your datasets.

Write a project summary. Use the chart on the following page to outline initial ideas.

PREP DATA

What question are you trying to answer?

Who are the stakeholders? (Refer to the project stakeholders table from earlier in Step 1)

Requester(s):

Subject Matter Expert(s):

Data Consumer(s):

Data Steward(s):

Data Practitioner(s):

When will the project be completed?

How can people use the analysis?

What restrictions apply to this analysis and/or data?

What related materials or projects should people reference to understand this project?

Will this project and/or analysis be updated, and on what schedule?

Whom should they contact if they'd like to contribute?

Step 3: ANALYZE

DIG IN:

Empower your team to work with the data and share what they learn as they go.

Connect the right tools to your common data workspace. (Use the Data Tech Stack Audit template if you're not sure what tools your team needs to integrate. Get your free copy at data.world/audit)

Create, save, and share queries so people can build on the work of their peers.

Share scripts created through R, Python, Jupyter, or similar tools to document your work and ensure reproducibility.

Encourage people to share insights and visualizations, ask questions, and post ideas throughout the life of the project. (Showing your work increases trust in your analysis. It also lets others learn from your process, springboarding future projects.)

CHECK IN:

Build in time to gut-check your project with your original requester and Subject Matter Experts. They can ensure your work is still aligned with their initial objective, resolve any confusion, and spot early insights your practitioners may be unaware of. **Try these 10 prompts to guide your check-in session:**

1. Here's what we've learned so far, and why we think it's relevant.*

2. Does this help answer your question?

3. What's surprising (or not) about this project so far?

4. Are any of these insights useful to you today? If so, how?

5. What would make these insights easier for you to act on?

6. How would you summarize what we've learned to your team (or manager)?

7. Do you agree or disagree with this interpretation of the data? Why?

8. Who else could use what we've shown you today?

9. What would they need to easily understand and act on the information?

10. What other questions do you have now that you've seen our progress?

Useful if you've been saving insights as you go!

Hint: *If you're on track, keep moving through the process. If your check-in uncovers gaps or opportunities, go through the "Prep Data" and "Analyze" exercises again so fresh resources meet the same standards for clean, contextualized data.*

Step 4: DEPLOY

PACKAGE AND SHIP:

Once analysis is complete, make sure the right people know about it. Keep in mind that different audiences will benefit from the data project's results being packaged in different ways. Try prepping your final results for these three data consumer types.

AUDIENCE:

TYPE	NEEDS	EMPHASIS
High-level consumer	Values anything that helps them keep a pulse on the business; values transparency and content that can quickly improve business decisions	• Specific, concise conclusions and recommendations • Visualizations and text explanations • Supporting third-party research • How-to guidance on using the analysis • Link to the data project
Data interpreter and/or strategist	Helps discover and translate the story in the data; needs detail and summary information to translate and convey to other stakeholders, especially clients and executives	• Project summary and process updates • Activity records • Access to, and understanding of, data used • Visualization-ready data • Saved queries to quickly explore raw data • Ability to securely share project, insights, and data with teams or individuals • Open channel to ask questions about the project • Ability to subscribe to project activity alerts
Data analyst	Checks conclusions; reuses and adapts relevant parts of the project for other projects	• Access to detailed context, data, and source information • Option to view, clone, or create new queries to add new insights to the project • Open channel to make suggestions on additional data, hypotheses, or other improvements

Ready for more Collective Data Empowerment?

This tool is most effective when used in conjunction with other tools, practices, and strategies to make everyone you work with more productive with data. Visit *Distinct Values* to to get more tools and a comprehensive field guide to Collective Data Empowerment.

About us

data.world is the modern catalog for data and analysis. It wakes up the hidden data workforce by replacing outdated barriers and disparate silos with deep connections between data, people, and analysis. When every employee can find, understand, and use data, your advanced practitioners can focus on high-impact work, and you can build a data-driven culture, faster. data.world is a Public Benefit Corporation headquartered in Austin, Texas. Visit data.world for more information and expert guidance.

data.world