DATA SCIENCE INFINITY

13 QUESTIONS YOU MUST ASK FOR EVERY DATA SCIENCE PROJECT

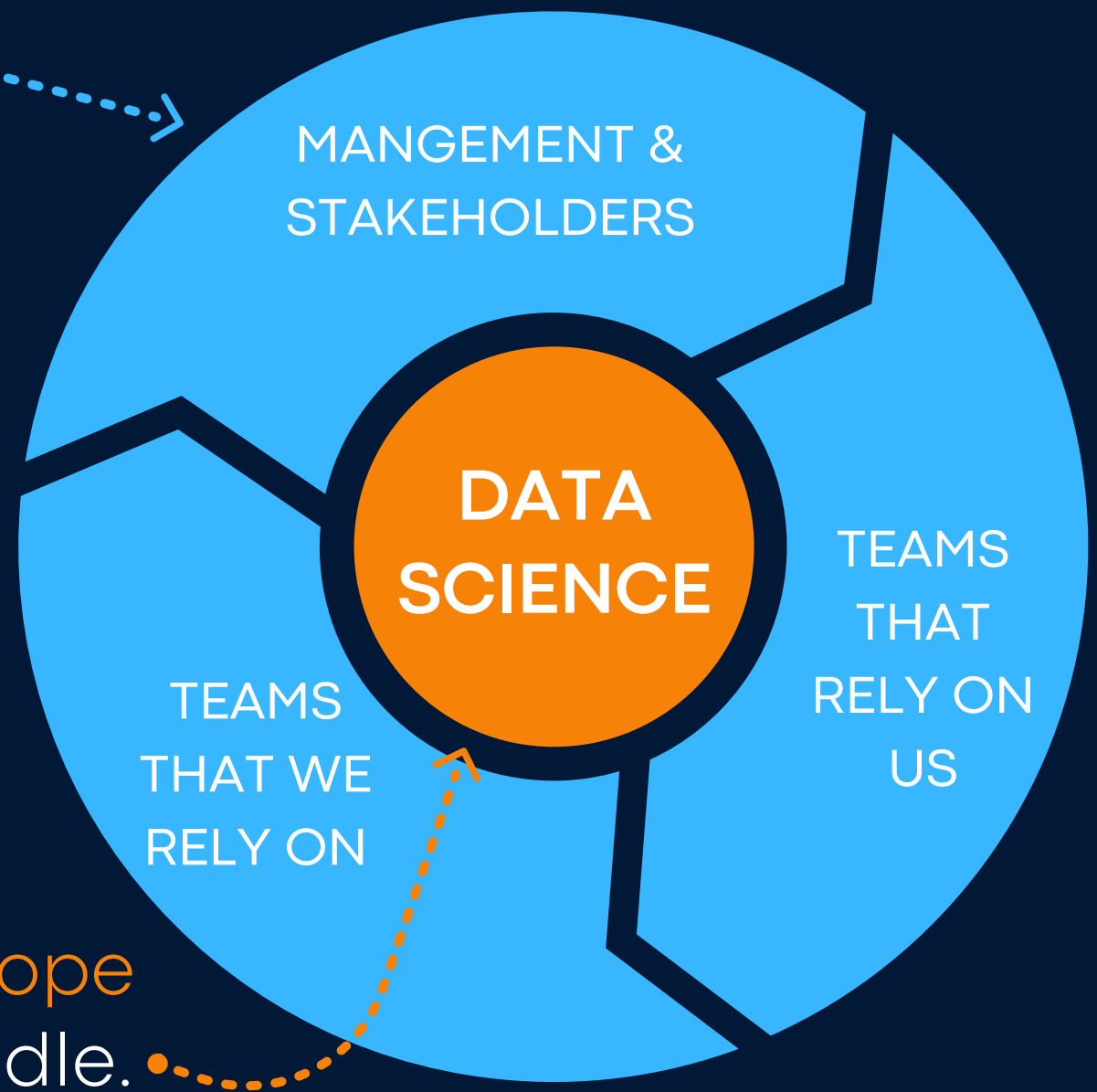
While all 13 questions won't apply to every project - having this template at hand will faciliate...

- A more efficient workflow
- A solution more likely to be adopted by the business
- A solution more likely to add tangible value

DATA SCIENCE INFINITY

Firstly, let's think of our world as two inter-conected parts...

The Business Scope is holistic. It overarches the whole project from inception through to delivery or application!



The Data Science Scope fits in the middle.

If a project is to be successful, Data Science must have clear connections into the other business areas.

Business Scope

What is the **business goal** for this project?

This might seem obvious, but it is so common to see this skimmed over too quickly!

The aim here is to get, at bare minimum, a high level understanding around what the goal of this project is, what the outputs look like, who the end user is, and what metric will be used to measure success.

For example, is the main goal to cut costs, is it to increase revenue, or perhaps make the customer experience better?

Push for answers on specific goals - is there a number the business has in mind?

Remember however, it might be that you, the Data Scientist, that needs to help scope these numbers out based on some early investigation into the data.

Why do we ask this?

This high level information helps frame and direct everything else in the entire project - if we know why we're working on a task, it helps us know what the next question to ask is and who it is we might ask. This all guides our thinking, and directs everyone towards the best solutions.

Business Scope

Can this project be solved or enhanced with Data Science?

The reality is, many aspects of business operation can run or be enhanced without ever needing magic from the Data Science team.

We always want to offer our skills - but a Data Science team has no shortage or tasks that need to be done.

For efficiency's sake, we should always make sure Data Science isn't forced onto every project without valid justification.

Based on the goal of the project, we, as Data Scientists want to work with the key stakeholders to answer this question.

We're the data experts, we know what data we have access to, and what can be done with it.

Our guidance is key in answering this question.

Business Scope

What does the Minimum Viable Product (MVP) look like?

A Minimum Viable Product (MVP) is a version of a product or solution with just enough features to satisfy the high level goal of the project.

As the name would suggest - it's accepted that it's not a final solution, and that many improvements and enhancements are likely possible...

The benefit of putting in place a Minimum Viable Product however, is that it can be built & implemented quickly, as a way to test the waters and prove to the business that there is value to be added.

Why do we ask this?

The more simplistic approach of the MVP can be easily understood by all key decision makers, meaning we have a high probability of getting it into production. The MVP gives all teams involved an entry point for the project, and scaling up from there becomes a much more manageable and efficient task.

If this MVP isn't successful, this can be equally valuable information for the business. As it's built and/or put into production relatively quickly, the business can pivot or change direction easily without any heavy investment of time and money.

Business Scope

When does the project need to be completed?

What is the desired timeframe, at a business level?

Sometimes this will be a hard date, for example an official product release or launch date (or it could even be a date that the business needs to comply to new regulations)

Often however, the date will be somewhat unknown. This will need to be revisited and refined once further exploration by all teams has been conducted.

Why do we ask this?

This information helps us start to think about what is and is not possible, and what options we may consider to achieve the goal - if the idea of an MVP is possible, when do we need this completed by?

For Data Science team leaders, this will also influence how much resource they allocate to a project.

Business Scope

Are there any considerations around ethics or compliance?

This mostly applies to external customer-facing projects, but it's worth asking or at least considering for any significant project that comes up. We always want our projects to be as efficient & effective as possible but we must not trade off fairness & bias to achieve this.

Why do we ask this?

While there are ethical and compliance considerations across the board, it is also very relevant to those in the data industry. We need to be more and more aware of the permissions we have for certain data sources, and we also need to consider any bias.

If the historical data that we're using to train our Machine Learning models, contains a hidden bias - the model simply learns those patterns and relationships and the bias continues.

This is something we of course want to avoid, so asking this question of the business as early as possible is vital.

Data Science Scope

What work has been done before in this area?

What information or knowledge does the business already hold and who holds this? What exploration do we need to do before starting any hands-on work?

Why do we ask this?

Efficiency!

Perhaps this is a completely new concept, and we're going to need to start from scratch - but if someone in the business has insights on what works and doesn't work, what data exists where, or just generally some specific knowledge pertaining to the task at hand - then this is gold-dust!

Asking this question can quite literally save weeks of exploration and mean we hit the ground running. Remember, our goal is to get this project or solution implemented in some way that will add business value. Take any help and any information you can, to ensure you quickly find the right path!

Start with key stakeholders and managers across all data teams, and follow any leads. Ask, ask, ask!

Data Science Scope

Who are we relying on, and who is relying on us?

Who else in the business will be working on this project, and in the overall flow who comes before, and after our team?

In other words - who might we be relying on in order to start our work, and who is relying on us before they can start their work?

Why do we ask this?

For a project to be successful, we must always remember that we, the Data Science team, are just one piece of the overall puzzle.

Try to avoid ever getting too inward focused as you can lose sight on the overall goal.

Perhaps we're relying on certain data being made available to us, who do we speak to about this?

Perhaps once we've built our code, or report, or model, it needs to be handed off to an engineering team to be put into production.

If so, what considerations are there in terms of format, code, size - in other words what are they expecting, and when are they expecting it?

Data Science Scope

What data is available, where is it, and what state is it in?

Through speaking with others in the business (Question 6), we should already have a feel for the type of data we need.

Of the data we need, what data is available to us? What data could be made available? Are there data sources in other parts of the business we can utilise? is there some crucial data for this particular task that could be acquired from an external source? Where do we find all this data, and how much work will it be to get it into a usable state?

Why do we ask this?

This really all boils down to timelines, if the needed data isn't ready, we need to report this back to the project manager.

Similarly, if the data is available, but in a bad state then we also need to consider what is required to make this usable. Data cleaning and preparation can be a time intensive task, so let's make sure we factor that into the equation...and keep everyone in the loop as we do.

This really sits with you, the Data Scientist. We're now in a phase where we should be driving our parts of the project!

Data Science Scope

What are the initial insights from the data?

We've got the data we need, we now need to start exploring it, and feeding back initial insights to key stakeholders.

Why do we ask this?

Think back to the very first question we asked in the whole process...

"What is the business goal for this project?"

This is an important time to stay aligned with that thinking - and to adjust timelines if necessary!

Early insights from the data will help all stakeholders stay close to the project, and keep an eye on the development & progress.

Data Science Scope

What data is valuable (and can we enhance it further)?

After exploring the data, we now want to narrow the focus down to only that which has value to the overall goal.

For Machine Learning we might look to feature selection and/or feature engineering, whereas for a report or dashboard we might look to discard anything that is not relevant to the key goal at hand.

Why do we ask this?

We always want to see if we can enhance any of the data through our own human intuition as well as what we've learned from speaking to key figures in the business about this particular project. This is another facet of Data Science that can separate good from great.

Always stop and think about what the data contains and what the overall goal is.

Put yourself in the shoes of the end user, and really push to empathise with their needs.

Data Science Scope

What approach or tool should we use (and why)?

Depending on the task at hand, there will be a variety of tools or approaches that might be applicable.

For certain projects, the options may be limited, but for others, let's say a project that requires a Machine Learning model to be built, there are all sorts of other considerations too, not only what model or algorithm to use, but how to validate the model, what hyper-parameters to use and so on.

While the technical considerations are extremely important, remember that we must also take the time to keep our head above this level of detail.

Start by again assessing the business-level goal for the project, and break this down. What type of problem is this? What techniques address that type of problem? What possible ways are there to apply each approach?

Create a shortlist of potential approaches and test.

Document the findings from your tests. This will help you justify your choice to the business.

Data Science Scope

How can we sync DS metrics with the overall business metric?

In Data Science & Machine Learning, we're often concerned with success metrics such as accuracy, precision, f1-score and so on.

These may not mean a lot to the key business decision makers in isolation, so while it's still very important to share and explain these, we must ensure that we communicate to stakeholders in a language they understand - in the terms that drive their goals.

To do this, we must again pivot from the Data Science solution back up to the overall business goal.

If the overall goal of the project is to acquire new customers then as well as showing of the accuracy of the model, put together example scenarios showing how many customers might be acquired with the model at various levels of marketing expenditure (for example).

Data Science Scope

How will this product or solution change over-time?

Don't push this question back. Ask this at the time of building the product or solution rather than down the line.

Is this solution one that will perform just as well in 6 months time or a years time as it will now, or not?

How will we measure this?

For a report or a dashboard, is there anything that will change in the near future, whether that be at a business level, a change in the data, or a change in regulations that might mean we need an update?

For a Machine Learning model, how will put in performance tracking to understand if the model degrades over time? What should be the threshold at which we would consider it to no longer meet the overall business goal, and at this point what work will need to be done?

Often, we're so busy getting the solution built or put into production that we don't factor this thinking in - make sure this is considered early!

DATA SCIENCE INFINITY

Want to land an incredible role in the exciting, future-proof, and lucrative field of **Data Science**?

DATA SCIENCE INFINITY

"I had over 40 interviews without an offer. After **DSI** I quickly got **7 offers** including one at **KPMG** and my amazing new role at **Deloitte**!"

- Ritesh

"The **best program** I've been a part of, hands down"

- Christian

"DSI is **incredible** - everything is taught in such a clear and simple way, even the more complex concepts!"

- Arianna

DATA SCIENCE INFINITY

>> https://data-science-infinity.teachable.com