

# ROLE OF THE DATA SCIENTIST IN EARLY PHASE STARTUPS



**Kristina Young**  
Lead Data Scientist/  
Engineer, Author

## 1 USER FRICTIONS

This refers to the **user need, and the obstacles** that the user currently goes through to achieve a desired goal. There is no point in building a product that no one actually needs.

The team identifies these frictions by conducting user interviews. These are skilfully designed to emerge the problems that the users want solved, whilst not biasing them towards solutions that drove the founders to create the business in the first place.



### USER FRICTIONS | DATA SCIENTIST ROLE

- Understand the problem that the user needs solved
- Understand the high level context that will be represented by the data
- Suggest interview questions to emerge data-related needs and current state

## 2 IDEATION

Once the frictions have been identified, a team is gathered to **come up with ideas that may solve those user needs**. This is usually done in a workshop format over a period of multiple days, using a variety of frameworks to get the ideas flowing.

The aim is to do divergent exploration at first, where the team gathers as many ideas as possible. This is then followed by a convergent approach, where the top concepts are refined to make them more meaningful.



### IDEATION | DATA SCIENTIST ROLE

- Come up with ideas that could solve user frictions
- Expand other's ideas and product visions
- Refine concepts by proposing data science approaches
- Bring up ethical AI considerations

## 3 VALIDATION

A set of top concepts is selected at the end of the ideation. **An idea can be great, but it won't stand a chance if it can't be realised as a self-sustaining business.**

In a validation phase, the team determines whether the idea can succeed by looking at it through these three lenses:

- Desirability: does anyone want it?
- Feasibility: is it possible to build it?
- Viability: can the business make money from it?

This is usually done through further interviews, desk research and proofs of concept.



### VALIDATION | DATA SCIENTIST ROLE

- Understand what kinds of data may be useful to solve the use case, and which of these exist/do not exist
- Understand where the data comes from (data inputs, iot devices, etc) and what it means
- Explore whether there is any data that can be purchased for said use-case
- Explore whether there are any vendors that can already solve the data science problem at hand
- Determine whether a data gathering campaign will be needed and how to run it
- Determine what high level software infrastructure needs the solution would require
- Build a proof of concept
- Assess ethical AI risks and guardrails

## 4 INCUBATION

Once the business has been validated, it is time to build it. **This involves not only building the first version of the product, but setting up an independent venture** along with its operations.

It is important to note that a product continually develops. In this phase the team aims to build the minimum viable product (MVP), that will allow them to test their vision with real users. For example, if the desired product is a dashboard showcasing advanced insights from 10 systems, the minimum viable product might be a basic dashboard that integrates 2 of those systems and shows 10 key graphs.

This is usually done through further interviews, desk research and proofs of concept.



### INCUBATION | DATA SCIENTIST ROLE

- Provide input for the design of the data gathering application (if relevant)
- Analyse and understand the data and the relationship to the user need
- Implement lean data science solution – sometimes a basic business rule is just enough to test with users
- If a model is needed for MVP, implement a first version
- Prepare your minimum viable software infrastructure setup
- Bring your model to production to be used by the product, it does nobody any good sitting on your local machine
- Iterate, pivot based on what the team learns from user testing
- Implement ethical risk mitigations

## 5 COMMERCIALISATION

In this phase the startup aims to grow, **getting more users and evolving beyond the minimum viable product that was first implemented**. This means more features, more users, more data, and more scale.

Just like the product continues to evolve, so do the strategies required to scale it further.



### COMMERCIALISATION | DATA SCIENTIST ROLE

- Analysis to understand how users are using the product
- Start moving towards data engineering and pipelining for scale (where needed)
- Start working on more advanced modeling / business rules (where needed)
- Perform and analyse results of A/B testing
- To Implement new/alternative features