



Follow

580K Followers

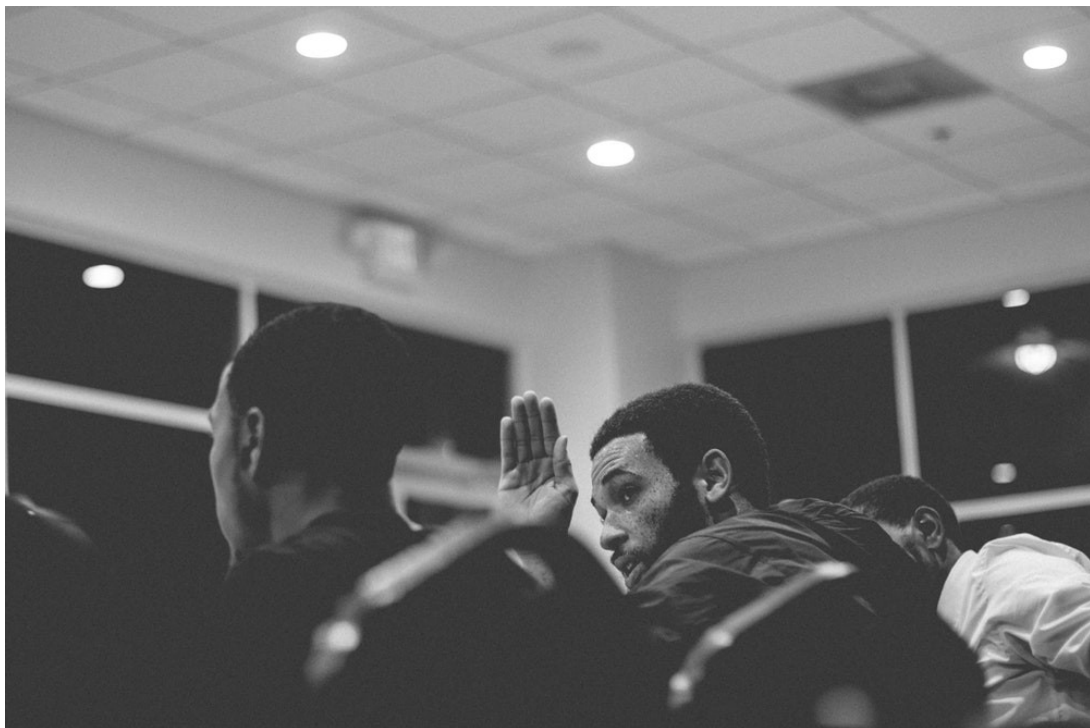


# How To Ask The Right Questions As A Data Scientist

To Define a Problem Statement



Admond Lee · Dec 8, 2018 · 5 min read



Before we talk about how to define a problem statement by asking the right questions as a data scientist, let's try to understand **why asking the right questions is so important.**

Long story short, when I first started [my first data scientist internship](#), I was very excited with the project and just wanted to get my hands dirty as soon as possible, without a clear understanding of the big picture.

I understood the problems that I was trying to tackle. But I did not drill down to the details to define goals and objectives. Even worse, I did not question the dataset given to



---

understanding of the problems and data.

This is my little story.

And I believe asking right questions and defining problem statements are some of the challenges faced by many beginners in data science (including me).

You see. Asking questions is easy. Everyone can do that. But asking the right questions is somewhat subtle because we don't know what right questions are considered **RIGHT**.

In this article, I'll share with you some of my takeaways and guides on how to ask the right questions and subsequently define a problem statement. I hope that this would re-define or help in your method in some ways to approach these challenges.

Let's get started!

## How to define a problem statement by asking the right questions?



(Source)

Admit it or not, defining a problem statement (or data science problem) is one of the most important steps in data science pipeline.

---

*A problem well defined is a problem half-solved*

— Charles Kettering

---

In the following part, we'll go through the four stages to define a problem statement.

All questions should be geared towards this direction to gain a better understanding of a project before formulating a problem statement.

### **1. Understand the problem that needs to be addressed and solved**



Very often, problem statements on [Kaggle competitions](#) are well-defined and we are given datasets to work with without having to worry about how the problem statement will be beneficial to others or how to get the data etc.

Sweet.

Now the thing is, problems are not defined in real work environment. They look ambiguous. They are vague.

And most of the time (if not all), stakeholders will just give us a question: **I have this “problem”, can you help me solve this?** Period.

Short but not sweet.

It is our task to help them frame the problem into a data science problem statement by really putting ourselves in their shoes and see things and problems from their perspective.

In other words, we need to have empathy.

Ask questions that can help you gain a better and deeper understanding of the problem as stakeholders have domain knowledge in the problem. Our task is to learn the domain knowledge from them and combine our technical knowledge with data to come up with a solution to drive business values.

## 2. Assess the situation with respect to the problem

Once we’ve framed a data science problem, the next thing to do is to assess the situation with respect to the problem.

This means we need to exercise caution analyzing risks, costs, benefits, contingencies, regulations, resources and requirements of the situation.

To illustrate further, this could be broken down into these following questions in general:

- What are the requirements of the problem?
- What are the assumptions and constraints?
- What resources are available? This is in terms of both personnel and capital, such as computer systems (GPU, CPU available), instruments etc.

## 3. Understand the potential risks and benefits of the project

This step is optional, depending on the size and scale of your project.



production.

- What are the main costs associated with this project?
- What are the potential benefits?
- What risks are there in pursuing the project?
- What are the contingencies to potential risks?

Answering to these questions help you get a better overview of the situation as well as a better understanding of what the project involves. And having an in-depth understanding of the project helps us assess the validity of the problem statement defined earlier.

#### **4. Define a success criteria (or metric) to assess the project**

This is important.

You don't want to have an ambitious project with a problem statement to be solved, only to realize that you don't have any metrics to gauge and evaluate the success of the project at the end.

This boils down to a simple question: **What do you hope to achieve by the end of a project?**

The achievement should be measurable and not be something abstract that could not be quantified. Some metrics might not be immediately available and therefore require data collection and preprocessing.

It is imperative that you discuss with stakeholders what metrics to be used and this discussion should always come in the early phase when asking the **RIGHT** questions.

Defining success criteria is so important as this will help you to assess a project throughout its life cycle.

## **Final Thoughts**



(Source)

Ultimately, our final goals are to formulate better questions and well-defined problem statements to solve using data science approach and generate business insights and drive actionable plans.

Thank you for reading. I hope this article gave you a glimpse of the importance of asking the right questions and how to frame problem statements.

As always, if you have any questions or comments feel free to leave your feedback below or you can always reach me on [LinkedIn](#). Till then, see you in the next post! 😊

### About the Author

**Admond Lee** is now in the mission of making data science accessible to everyone. He is helping companies and digital marketing agencies achieve marketing ROI with actionable insights through innovative data-driven approach.

With his expertise in advanced social analytics and machine learning, Admond aims to bridge the gaps between digital marketing and data science.

Check out his [website](#) if you want to understand more about Admond's story, data science services, and how he can help you in marketing space.

You can connect with him on [LinkedIn](#), [Medium](#), [Twitter](#), and [Facebook](#).

#### **Admond Lee**

In the mission of making data science accessible to everyone.  
Admond is helping companies and digital marketing agencies achie...

[www.admondlee.com](http://www.admondlee.com)

---

[Open in app](#)



---

[Data Science](#)

[Towards Data Science](#)

[Education](#)

[Motivation](#)

[Learning](#)

[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

