
# The Data Science Method (DSM) — A framework on how to take your data science projects to the next level.

Aiden V Johnson Follow
Nov 22, 2018 · 3 min read

Have you already landed your first data science job or are you in the midst of a data science boot camp? Or perhaps, you're a seasoned machine learning professional? Regardless, applying the Data Science Method (DSM) in your projects will elevate your work, improve your impacts, and strengthen your skills as a data scientist.

> *The biggest difference between people that are successful as data scientists and those that are not, is their ability to effectively frame data science projects and communicate project outcomes.*

To be successful, you must have all the prerequisite core knowledge of machine learning algorithms, programming abilities, and also be passionate about becoming a professional data scientist. The biggest difference between people that are successful in data science roles and those that are not is their ability to effectively frame data science projects and communicate project outcomes. Sometimes, this is referred to as data storytelling, however, that only describes part of the process involved in sharing your findings and results. You need to work to develop the entire framework for the narrative of your data science project to communicate the project as a story.

The Data Science Method (DSM) serves to identify the context of your data science story. Starting with the end in mind is one way to glean some guidance — you must know where you are headed in order to take the appropriate steps along the way. This can be difficult depending on the complexity of your data and the business needs being requested for the project. Let's consider the scientific method as a framework, as it provides clear steps that are taken along an experimental path.

Based on the scientific method, I have developed the Data Science Method (DSM) as a way to improve data science project outcomes and take your work to the next level. The DSM is detailed below.

. . .

**The Data Science Method**

1. Problem Identification
2. Data Wrangling
3. Exploratory Data Analysis
4. Pre-processing and Training Data Development
5. Modeling
6. Documentation

. . .

The DSM steps can be applied to nearly every data science or machine learning project, as it serves as a guide map for what actions you need to take. The benefits of following this approach include, but are not limited to: reducing the likelihood of needing to backtrack in your work to solve a data problem, identifying key data issues early, gauging expectations of the project outcomes appropriately, forcing yourself as the data science professional to be very clear about the goal of the project and how the data does or does not support that goal. The DSM steps and other frameworks have been developed by seasoned professionals because they have extensive experience managing projects that forced them to acknowledge when they overlooked a crucial consideration earlier in the analysis. The DSM is most helpful for aspiring data scientists because it allows you to leverage the experience of professionals that came before you.

Structured thinking is a common and well-articulated tool used in all types of industries. Thus, the DSM supports an organized approach to solving a business problem with data, and it also supports the essential communication component of articulating project outcomes to your client. As you work through the five steps of the DSM, you will build a deeper understanding of your data, the problem, and how the data can be leveraged to meet your client's expectations and solve the problem.

Review each of the linked articles in the above short outline to understand each step in more depth. For now, here is the full overview of the DSM steps.

To receive updates about DSM or Data Science Professional Development Sign up here.

Data Science Machine Learning Professional Development Data Python