**A Comparison of Data Science in Academia and Industry**

A Ph.D. student’s exploration of the surprising parallels between academic and industrial data science

[[Nazlı Alagöz](https://medium.com/@nalagoz13?source=post_page-----f9c4cb3fda92--------------------------------)](https://medium.com/@nalagoz13?source=post_page-----f9c4cb3fda92--------------------------------)

[[Towards Data Science](https://towardsdatascience.com/?source=post_page-----f9c4cb3fda92--------------------------------)](https://towardsdatascience.com/?source=post_page-----f9c4cb3fda92--------------------------------)

[Nazlı Alagöz](https://medium.com/@nalagoz13?source=post_page-----f9c4cb3fda92--------------------------------)

·

[Follow](https://medium.com/m/signin?actionUrl=https%3A%2F%2Fmedium.com%2F_%2Fsubscribe%2Fuser%2F4ba02da50bf&operation=register&redirect=https%3A%2F%2Ftowardsdatascience.com%2Fcrossing-the-bridge-a-comparison-of-data-science-in-academia-and-industry-f9c4cb3fda92&user=Nazl%C4%B1+Alag%C3%B6z&userId=4ba02da50bf&source=post_page-4ba02da50bf----f9c4cb3fda92---------------------post_header-----------)

Published in

[Towards Data Science](https://towardsdatascience.com/?source=post_page-----f9c4cb3fda92--------------------------------)

·

8 min read

·

May 29

232

2

A person working on a computer

Description automatically generated with medium confidence

Photo by [Campaign Creators](https://unsplash.com/@campaign_creators?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on [Unsplash](https://unsplash.com/photos/pypeCEaJeZY?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText" \t "_blank)

As a current Ph.D. student who decided to leave academia for the industry, I’ve devoted substantial time to exploring how my academic experience could translate into an industry setting. Initially, the two domains seemed to stand at opposite ends of the spectrum. However, after extensive research and discussions with many who’ve traversed the path from academia to industry, I found more parallels than I anticipated. In this blog post, I’ll share my insights about the differences, and commonalities between academic and industrial data science.

I should note that lacking firsthand industry experience, my comparisons between industry and academia rely on insights from others who have traversed this path. Additionally, I am a Ph.D. candidate in quantitative marketing and thus more familiar with academic research in economics and business. So, when I make comparisons I compare industry practices to the academic research process in these areas.

For those unfamiliar with the process of academic research and the application of data science within this context, I define data science as the process of driving insights from data using scientific methods and algorithms. For example, I mainly use causal inference and machine learning methods to answer the research questions in my dissertation projects. I will next provide an overview of academic research in quantitative marketing.

In academia, we need to find relevant questions that warrant answers (e.g., how do paywalls affect revenue, how do certain campaigns affect sales). We then gather the data required to answer these questions (e.g., by collaborating with an online newspaper, web scraping, using APIs, or procuring data). Once we have the data, we can start the process of preparing the data for analysis and use the data to test our hypotheses. Once we have some initial results, we communicate these through presentations and draft papers to get feedback. We update the analyses, presentations, and papers to address the feedback received. This process (i.e., get feedback, address the feedback) repeats until we reach a high-quality project/paper that clears the scientific bar.

Now that we understand how data science underpins the academic research process, let’s explore some key aspects worth discussing when comparing data science in academia versus industry.

**Objectives and Impact Definition**

The objectives of data science in academia and industry, though both valuable, have distinct focus areas and consequently, varied definitions of impact. In academia, the primary objective is often the advancement of knowledge. Scholars aim to fill gaps in current understanding, challenge existing theories, or develop new frameworks, methodologies, or tools. The impact here is largely measured by the influence of their work on the academic community, as seen through citations, advancements in the field, and contributions to teaching and learning.

In industry, data science objectives are typically more directly tied to business goals such as improving operations, optimizing products or services, and aiding strategic decision-making. The impact is often assessed in terms of tangible outcomes like enhanced user experience, increased revenue, cost savings, or other key performance indicators relevant to the business. This does not mean that the work in the industry is purely transactional. Many industry projects also contribute to broader societal goals, such as improving healthcare, advancing sustainable practices, or enhancing data privacy and security.

In both sectors, the ultimate goal is to create value, but the way that value is defined and measured can vary significantly. Recognizing these differences can help us appreciate the unique contributions of data science in both academia and industry.

A pair of glasses on a notebook

Description automatically generated with low confidence

Photo by [Dan Dimmock](https://unsplash.com/@dandimmock?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on [Unsplash](https://unsplash.com/photos/sNwnjxm8eTY?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText" \t "_blank)

**Formulation of Research Questions and Data Acquisition**

In academia, the creation of research questions and the task of gathering data are usually an intrinsic part of a scholar’s role. The impetus to develop and answer questions is driven by intellectual curiosity, gaps in existing knowledge, and the potential to contribute new insights to our collective understanding. This process can be demanding and time-consuming as it may involve an extensive literature review, experimentation, surveys, or fieldwork.

In contrast, within the industry setting, the origin of research questions and data often comes in a different form. Questions are generally derived from specific business needs, to yield insights that enhance operations, products, or services. Data scientists in the industry typically work with data that is readily available — a byproduct of business activities or procured from third-party providers.

However, these are not strict rules and there can be quite a bit of variation in both sectors. In both academia and industry, the ability to formulate meaningful questions and understand data quality and suitability are crucial skills for a data scientist.

**Timelines and Scope of Projects**

Both academic and industry projects have their unique timelines and scopes that align with their specific objectives. Academic research often follows a longer timeline due to the meticulous nature of data collection, rigorous analysis (e.g., robustness checks), peer reviews, and publication processes inherent in academia. As scholars, our purpose is to add to the reservoir of scientific knowledge through robust, thoroughly verified results, a process that naturally requires time.

However, in an industry setting, timelines tend to be shorter in response to the dynamic nature of the market and business needs. This doesn’t mean that quality is compromised. On the contrary, the industry also follows stringent quality control processes to ensure the reliability of results. The pace is driven by the goal to deliver actionable insights swiftly to maintain or gain a competitive edge. This agility allows companies to adapt and innovate rapidly, which is a strength in the fast-paced business environment.

**Critical and Scientific Thinking**

Critical and scientific thinking is highly valued both in academia and data science. Hypothesis testing, for example, is at the core of the scientific process in academia and the causal inference track of industry data science (e.g., experimentations). The ability to formulate good questions, validate answers, and critically evaluate arguments are skills esteemed in both realms.

**Communication and Collaboration**

A crucial skill in both sectors, the essence of communication, varies depending on the audience. In academia, we communicate with fellow researchers, professors, students, or stakeholders from the industry, while in industry, data scientists communicate with a wider spectrum of stakeholders, including other data scientists, managers, executives, or clients. The knack for simplifying complex ideas for non-specialist groups is a common requisite.

There is a misconception that doctoral students lack communication skills because of their deep specialization, potentially struggling to explain complex ideas in simple, accessible language to non-specialists or business stakeholders. However, this is fundamentally untrue. Success in academia hinges upon the ability to communicate effectively with diverse audiences. For example, I regularly need to communicate with people from different backgrounds such as researchers from my field, researchers from other fields, students, and people from the industry. For each of these groups of people or some combination of these groups, I communicate differently.

A computer screen with text on it

Description automatically generated with low confidence

Photo by [Irvan Smith](https://unsplash.com/@mr_vero?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText" \t "_blank) on [Unsplash](https://unsplash.com/photos/5eBW5GomfhY?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText" \t "_blank)

**Handling data and good coding practices**

With the advent of big data, both academic and industry researchers grapple with processing large datasets and complex data pipelines. Although the specific tools utilized may vary, the principles of good coding practices, rigorous documentation, and efficient data management are universal for both industry and academia. In academia, these principles form the foundation for reproducible and open science. Thus, many researchers implement these principles to make science more transparent, accessible, collaborative, and reproducible.

**Project Management**

Effective project management and prioritization skills are critical in both academia and industry, despite differing perceptions. The notion that longer timelines in academia lessen the importance of project management is indeed a misconception. In academia, scholars often balance a diverse set of responsibilities, including teaching, mentoring, and administration, alongside complex research projects. Each stage of research requires meticulous planning and efficient task management, emphasizing the need for strong project management skills.

Similarly, in industry, careful planning, coordination, and execution are crucial across all stages of data science projects. The rapid pace of the industry, multiple simultaneous projects, and the demand for quick, actionable insights further underscore the importance of effective project management. Despite differences in timelines and context, the need for successful project management in both sectors is paramount, ensuring not just timely completion, but also work quality and resource efficiency.

**Soft Skills**

Independence, self-discipline, effective collaboration, and strong organizational skills are vital whether you’re aiming to publish a paper or deliver a successful business project. These skills might be one of the most important things I learned during my Ph.D. It is important to be able to work independently but also collaboratively.

For my projects, I need to be able to come up with ideas, solve problems, and do other tasks independently but at the same time, we need to be on the same page with my collaborators (co-authors). I have learned how important effective communication is especially when your collaborators come from different backgrounds and cultures. Another skill that one develops for effective collaboration is to be the ability to take, appreciate and implement feedback. These skills are also crucial for industry roles.

**Learning and Adaptability**

Both academia and industry necessitate a consistent commitment to learning and adaptability, keeping pace with evolving methods and technologies. In academia, as it is in the industry you cannot ignore any new useful tools to not lose your competitive edge.

For example, while I was doing my first Ph.D. project and using a rigorous statistical method used in both academia and industry. However, there was a lot of new research coming out showing that in certain situations this method does not perform that well. I ended up educating myself on issues that arise and potential new solutions to these issues.

**Technical Skills**

Data scientists across both academia and industry must master an assortment of technical tools. Languages like Python and R are widespread in both spheres for data science tasks. However, tools such as SQL, Spark, and Tableau may see more frequent usage within industry circles compared to academia.

**Conclusions**

My exploration into the intersection of data science in academia and industry has been enlightening, revealing more parallels than I initially anticipated. Many academic researchers find that their backgrounds provide a solid foundation for roles in the industry. Their educational and technical training enables them to easily grasp new methods, while their experience in communication, problem-solving, independent research, and collaboration provides them with a strong skill set that can be applied in the industry context.