Quantifying Data Science Training Value
| 2020 |

There is no question that AI and its application will change society, but unfortunately, easy ways to measure this change have been elusive. The economic impact of Artificial Intelligence (AI) technologies on both macro and micro levels are challenging to quantify. In this article we summarize the most important factors to consider when measuring the impact training investments have on an organization. This article also provides a contextual backdrop that helps business leaders calibrate their expectations when they run the numbers.

The key takeaways:

1. An understanding of the factors outside of talent investment that play a role in estimating a ROI
2. Key metrics to track that contribute to maximizing ROI

Talent as an innovation and improvement engine

To start out we can think of data teams (analysts, translators, scientists, engineers) collectively as **talent**. The process we are interested in assessing is that talent produces data projects and data projects are used to realize a return on investment (ROI). Examples of data products are the deployed models, dashboards, Jupyter notebooks, reports and more. ROI is often measured by carefully tracking a specific business metric like change in customer retention over time.


Off-the-shelf software platforms that produce data products have a much lower ceiling when it comes to return on investment. Having an in-house data team is akin to having an at-home cook who knows some number of recipes and is continually learning more. Outsourcing talent under this metaphor, either through contracting or via software is like ordering takeout. The usefulness of a data product is directly related to how well it is infused with knowledge from the

business. We have seen time and time again that after just a few weeks of customized training, in-house talent is more capable than off-the-shelf data science products.

It is not hard to see that taking folks who know the business well and transforming them into capable data team members is the ideal scenario, but still we should be able to measure the impact. The following three factors have effects on realized ROI that are so important that any of the three can diminish the ROI of a talent transformation.

1. **Data Culture** - The degree to which a company decides to embrace AI.
2. **Data Maturity** - The level of organization and centralization a company has when it comes to saving, storing and accessing data.
3. **Confounding Factors** - Additional factors outside of the data products themselves that influence the realized ROI.
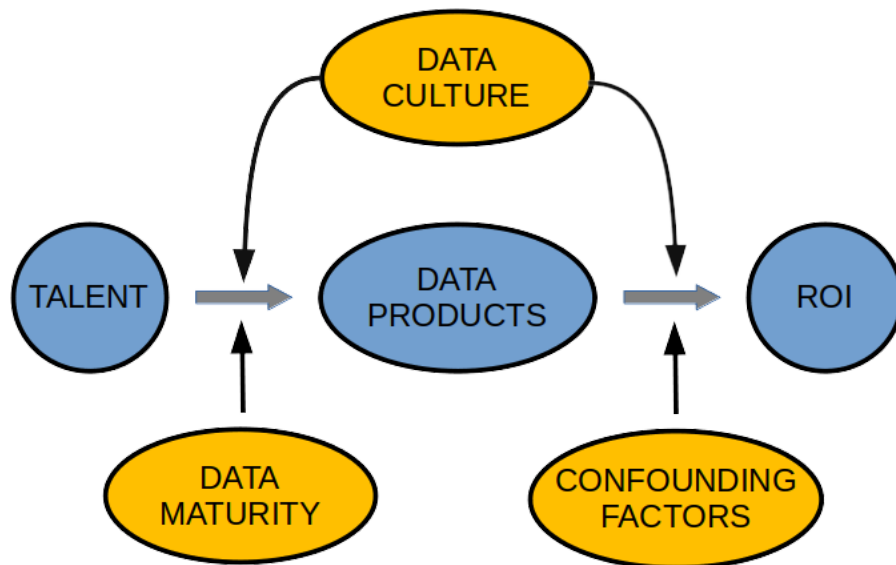
The talent pool is the engine that drives improvement and innovation. The process of leveraging that talent pool through the creation of data projects to maximize an ROI is influenced at a high level according to this diagram.

Influences on realized ROI

Most organizations are aware that data maturity is a major blocker for data teams, which is why many of them are actively engaged in centralizing and improving how data is saved and accessed. Depending on the industry and level of data maturity, it is not uncommon for data scientists to spend 50-80% of their time cleaning and preparing data [CITE].

Data culture is the degree to which a company embraces AI, but there are several things to unpack when considering what this means for your organization. Data culture affects both the ability of data professionals to do their work and the level of adoption data products attract after they have been made available. Here are some questions that you can ask to help determine the health of your company's data culture.

- Are data team members provided with clear roles and responsibilities?
- Do data team members have access to a high-performance computing platform?
- Do data team members have access to cloud resources?
- Do data scientists have full administrative privileges to install and work with the best tools for the job?
- Do data teams use version control, pair programming and some form of Agile?
- Is there an effective inter- and intra-team organization that promotes collaboration, code reuse, and business alignment?
- Do data team members have access to continued learning?
- Is there a high level of employee satisfaction?
- Do all data teams adhere to a process model that has been agreed upon for development?

Confounding factors are the factors other than the data products themselves that contribute to ROI. Perhaps you have a recommendation engine in production, but it takes a couple of seconds to display the recommendations. An investment in infrastructure may speed up the model, which then results in a better user experience and accordingly more purchases. This is an example of a confounding factor that may drive up ROI even though no changes were made to the underlying model. Seasonal fluctuations in consumer spending, demographics of a new market and a myriad of other factors influence realized profit. And because ROI is simply the amount of profit expressed as a percentage of the amount invested, the numbers can be misleading if you cannot account for the majority of influencing factors.

Deriving value from data products

Data products can be loosely categorized into three buckets. Each data product exists to address a **business opportunity**. Perhaps, a more effective revenue projection model would help leadership, or maybe an outlier detection model might help us identify potential fraudulent transactions. These examples and the others we discuss here all exist to:

To learn more about Data products at Galvanize, please contact **enterprise@galvanize.com**

1. Improve or optimize existing business opportunities
2. Create new business opportunities
3. Automate an existing process

Value is derived from each of these in a slightly different way. Improving on an existing product or service could improve revenue. A chatbot that is designed to make a website experience more efficient, may increase customer satisfaction. A well-engineered model for supply chain optimization can have a significant impact on quarterly earnings. This is the most difficult category in which to make ROI estimates, and unfortunately, it is also the most common type of data product for many organizations.

The second category, creating new business opportunities, is the one often described as the application space in data science where there is the most to gain. There is always risk that comes with creating new products, but new products are an important way to clearly demonstrate value. Value in the case of new opportunities can be directly measured in terms of revenue or customers.

The final category is automation. Like the creation of new products, automation is straightforward and has a clearly defined way to measure impact. The common unit of measurement is reduced person-hours.

It is important to note that most data products offer small but consistent improvements with respect to business metrics, e.g. a staffing optimization tool that results in a reduction of person-hours by say 0.5 per day. Over a year with 250 work days, this amounts to 125 hours or a savings or 3 weeks of time. This does not even account for improved efficiency of the model. Staffing optimization models can be developed and deployed into production in a few days with an efficient team.

Some data teams have hundreds of data products, varying in size and complexity. Although the more sophisticated data products that accomplish something new can result in an impressive ROI, it is important to keep in mind that an army of small automations and efficiency improvements can collectively produce a sizable ROI as well.

Key Metrics for Successful ROI

In many circumstances, it is difficult to measure person-hours saved. It can also be difficult to accurately measure usage of data tools and the impact that they have. A tool that improves customer satisfaction will indirectly affect profit. In this case, it might be better to use customer satisfaction instead of profit as your business metric. Here are some metrics that all businesses can use to measure the effectiveness of data teams when it is difficult to get a realistic estimate of ROI.

- Number of data projects by category (improved, new, automated)
- API hits on deployed models
- Average time to deployment for data projects by category
- Employee estimated time savings

Galvanize Leads in Talent Enablement

Data Science and Data Engineering Professional Services at Galvanize focuses on training pathways on data strategy, organizational growth, and data implementation to generate maximum return on investment. Galvanize offers industry-leading Talent Transformation training opportunities, including for data analysts, data scientists, and data engineers.

Leadership at your organization can learn the best practices for data strategy through our Executive and Data Management workshops. Offered in half-day and full-day workshops, Executive courses highlight key functional areas including Data Culture, Data Governance, the Data Science Workflow, and Relevant Data Trends.

The Galvanize AI Workflow for Business™ program offers embedded Data Scientists and Data Engineers real-world product simulation for AI workflows in production. AI Workflow for Business™ highlights design thinking, data strategy, and implementation best practices for building and scaling data products.

Life after Galvanize training programs continue with Galvanize consulting and coaching. Our experienced Lead Instructors can consult, advise, and mentor your business units, product managers, and data teams for sprints and longer term projects on strategy, product management, and product implementation.

To learn more about Data products at Galvanize, please contact **enterprise@galvanize.com**