DATA SUMMIT
KEY TAKEAWAYS
MAY 2018

BACKGROUND

On May 14, 2018, the NYC Tech Talent Pipeline (TTP) held a Data Summit at Etsy, bringing together tech education and training providers from 11 organizations and 20 industry advisors. In a series of focused conversations, accelerated training providers, as well as faculty and adjuncts from CUNY Computer Science departments, met with data-focused professionals from local companies to discuss feedback on data-related curricula and workforce trends. This document summarizes several high-level takeaways that emerged from those conversations.

OVERVIEW OF DATA ROLES

Demand for data professionals in New York City is growing across a number of different roles. While there are a wide variety of roles that require data proficiency, they can be broadly grouped into three main types of roles:

- **Data Analysts** glean insights from data to help businesses make good decisions. To do this, data analysts must be able to set up a framework for analysis, ask the right questions to get at an answer, clean up a messy dataset to make it useful, use a variety of tools and methods to identify an answer, and then present these findings clearly to technical and non-technical stakeholders in a business environment.

- **Data Scientists** take the work of data analysts a step further. Data Scientists also glean insights from data in order to help businesses make good decisions, but they are generally expected to be able to set up and conduct complex, sophisticated analyses using advanced statistical, programming, and other skills to produce more powerful insights. That said, the work of Data Analysts and Data Scientists is closely intertwined. Data Analysts are generally focused on trends and helping the business remain on track toward goals. If this work reveals that a business is off track somewhere, a Data Scientist may step in to undertake an experiment or research study to identify the source of the problem.

- **Data Engineers** make all of this work possible. In general, Data Engineers are responsible for building the frameworks and infrastructure to capture and store the data that Analysts and Scientists use. Data Engineers help to ensure that data is stored efficiently, securely, and in a way that will make it easy for others to retrieve and use. In some companies (typically larger companies), Data Engineers are also responsible for acquiring and cleaning data; in other companies, Data Analysts and Data Scientists must do this work themselves.

It's important to note that these roles can vary significantly from company to company; what a Data Scientist does in one company may be very different from a role with the same title in a different setting. For example, in some companies, responsibilities for Machine Learning fall with Data Scientists; in others, dedicated engineers are responsible for this area. Applicants for these jobs should ask questions during the job application process to ensure they understand what they'll have the opportunity to work on at various companies.

COMMON AREAS FOR IMPROVEMENT

Employers provided training and education providers with specific feedback about areas for improvement in their curricula. Common areas for improvement include:

- **USING DATA SKILLS TO SOLVE BUSINESS PROBLEMS:** Many industry representatives across a variety of companies emphasized that it's not enough to master technical skills; data professionals must be able to apply these skills to solve business problems. Employers see many candidates who are technically excellent but flounder when asked to answer interview case questions that require using data skills to answer a business problem. Training and education providers should incorporate applied projects and case studies to help students practice using their technical skills to solve business problems.

  Specifically, data professionals must be able to understand a complex business problem or question being asked; break that problem down into parts; set up an appropriate approach and collect the data needed to support that approach; perform an analysis; and produce a recommendation or solution to a business problem.

- **TELLING A STORY WITH DATA:** Data professionals must be able to clearly communicate their findings in writing, through presentations, and in conversations with a variety of stakeholders. To have an impact, data professionals must clearly communicate their findings, why they're important, and recommend a course of action. Specific areas that educators and training providers should help students hone include:

  o **Data visualization:** Employers repeatedly emphasized the need to incorporate data visualization across all programs. Students should learn how to communicate their findings through clear, concise visualizations.

  o **Communicating technical ideas simply:** Data professionals must be able to communicate the findings of complex analyses to a variety of audiences, including non-technical colleagues and decision-makers.

  o **Presentation skills:** Related to both of the above, students should practice designing and delivering clear, compelling presentations that convey complex information as simply as possible.

- **FOUNDATIONAL STATISTICS:** A solid foundation in statistics is critical for all data professionals. The level of statistics needed varies greatly by occupation and by company, but at minimum candidates need a solid understanding of descriptive statistics and probability. Most data professionals need a more advanced statistical toolkit, which can include but is not limited to more sophisticated experimentation, inference, and attribution methods.

- **CREATING USEFUL DATASETS:** Many students are good at manipulating clean datasets to come up with insights, but much of the data that professionals encounter in a live business setting is messy when first encountered. While some large companies have engineers or others dedicated to cleaning data for analysis, in general, data professionals need to know how to clean and transform data into useful datasets.

- **PRIVACY AND SECURITY:** Several employers noted that educators should touch on privacy, security, and data ethics in their curricula. As the collection, storage, and analysis of large amounts of data becomes the norm across many businesses, data professionals should understand best practices for safeguarding data.

ADDITIONAL SKILLS FOR DATA SCIENTISTS

- **COMPUTER SCIENCE FUNDAMENTALS:** Data Scientists should have a solid foundation in computer science basics, especially data structures and algorithms.

- **WORKING WITH DATA AT SCALE:** While all data professionals need to understand how to work with large amount of data, Data Scientists need to have an awareness of how their work can be deployed on a large scale. While some employers continue to use Hadoop to work with big data, others have migrated to Spark. Training and education providers should stay abreast of needs as they continue to evolve.