Data Science Starts with Asking Questions

Marc S. Galli

Walden University

Professor Dr. Mike Collins

DSCI 2010: Data Science Essentials

July 25, 2020

Data Science Starts with Asking Questions

Data science is founded on specific techniques and procedures. It starts with an exploratory approach, asking questions. Although data science ultimately seeks an answer to the question at large, fundamentally, it does so much more. Data scientists follow a system of methods, collectively referred to as a methodology, to identify, select, process, and analyze information (International Business Machines, 2016). This systematic study of structured and unstructured data uncovers trends and permits data scientists to formulate insights about the data (IBM, 2016). Data science enables staff to make strategic corporate decisions and informed choices.

The data science methodology this treatise references has been outlined by John Rollins, a senior data scientist and analyst for IBM. Rollins' comprehensive methodology can be relied on due to his extensive background as an engineering consultant, professor, and researcher. Additionally, he holds doctoral degrees in economics and engineering (International Business Machines, 2016). Data science begins with a business understanding, then takes on an analytic approach. From there, the methodology enters a feedback loop considering data requirements, data collection, data understanding, and data preparation. What follows next is another feedback loop consisting of modeling, evaluation, deployment, and finally: Feedback. Given feedback, data scientists can return to modeling and proceed again all while continually refining their study until trends and insight are acquired. As evidenced, each step in the methodology depends on the preceding step and if a proper business understanding is not ascertained then the data science team will be answering the wrong questions.

Ergo, it all starts with the business understanding. A business understanding is critical to any work assignment which may rely on data science or require analytics to achieve success. In order for research to be tangential to the problem which needs a solution, all variables and factors must be considered and raised before data collection and even before data requirements can be framed. The same holds true for other tasks in a business setting. It is necessary for time to be spent contemplating all aspects of the problem and to delineate any variables which need to be identified or constants that need to be defined before your team members are individually dispatched on their disparate tasks. Getting clarity around the problem to be solved allows one to determine what data will be used to answer the principal question (Big Data University, 2016).

In business understanding, and beginning with an exploratory approach to problem-solving, data scientists start framing questions to be answered. In our case study, Nutri Mondo, which aids local communities in gaining enhanced access to healthy food, determines what data they intend to access and what that data tells them about the relationships. They subsequently determine the source of the data, develop a timeline, and establish a clear listing of what variables exist (Laureate, 2016). Other questions surface and get answered to plainly establish the data requirements and the parameters of where data will be collected from. Upon exhausting this exploratory process and determining every possible question which may arise, then, and only then, should the keen data scientist progress to the analytic approach stage where the team can organize and plan how to tackle the core question and problem. In our case study, the data science team, in their analytic approach, then proceeded to notate data requirements and clearly define data collection procedures and roles for each team member.

In conclusion, the methodology of data science starts with an exploratory approach, asking questions. In the end, properly employing the data science methodology results in uncovering trends and the formulation of insights that are key to strategic, corporate decision-making. Defining all tangential variables is foundational and essential to success. Data science starts with asking questions and each step in the methodology depends on the preceding step and if a proper business understanding is not ascertained then the data science team will be answering the wrong questions and irrelevant data will be mined and analyzed leading to precious loss of time, manpower, and resources.

References:

Big Data University. (2016). *Data Understanding* [Video file]. Armonk, NY: Author.

IBM. (2016). *What is Data Science?* [Video file]. Armonk, NY: Author.

International Business Machines. (2016). *Welcome to Data Science Methodology* [Video file].

    Armonk, NY: Author.

Laureate. (2016). *Nutri Mondo*. Laureate International Universities. https://mym.cdn.laureate-

    media.com/2dett4d/Walden/DSCI/2010/WorldOfDataScience/index.html