

Gathering Data: Requirements and Collection

Marc S. Galli

Walden University

Professor Dr. Mike Collins

DSCI 2010: Data Science Essentials

August 1, 2020

Gathering Data: Requirements and Collection

Within the Nutri Mondo case study, a thoughtful and exploratory approach has been concluded and a solid business understanding has been reached. Our team is on, and well into, the analytical approach phase and has undergone formulation of the data requirements and determining the scope of collection. While considering all aspects of data requirements, our team, in video conference calls, set out to define the parameters of what data would be required (Laureate, 2018). This was to include the data's content, as well as what format it was to be in, and what groups the collected data would be representative of. The team outlined how to source the data. Public versus private data acquisition sources were discussed, and the team decided affirmatively on public data, particularly, governmental data sourcing. In perfunctory discussions the team agreed that publicly governmental sourced data would be transparent, and far more accurate than what regional teams at Nutri Mondo would be capable of collecting. An additional benefit, they surmised, would be realized in obtaining more variables, including controls for ethnicity and income, among others. Once the team clarified the scope of the collection net and the source of the prospective data, they set out to solidify plans to collect, understand, and work with the data (IBM, 2016). They additionally clarified how they anticipated to meet the desired outcome with data that may not be representative of all counties or countries they sought to comprehend and discerned that they may have the capability to apply assumptions learned from the collected data to a larger group. The team kept a paramount goal in mind in thinking ahead and anticipating ensuing snares.

Next, our team endeavored to outline prospective plans for data collection. As they progressed, data requirements were revised. Our team of analysts will undoubtedly make full utilization of techniques such as visualization and descriptive statistics and apply them to the dataset in a quest to assess the quality and content of the data, as well as, to make initial insights (International Business Machines, 2016). Provisions and calculated steps were discussed in detail as to how analysts sought to identify gaps and disparities in the data. (In essence, a correlation would be pursued between the counties and countries which the collected data applied to and the counties and countries which were representatively absent in the dataset.) Lastly, it was discussed what steps would be taken if the collection required more or less data and that gaps would be

identified. In short, assessment is intended to be made after collection to ascertain if the Nutri Mondo team has the datasets which they need to proceed.

In Nutri Mondo's quest to aid local communities in gaining enhanced access to healthy food and in furtherance of the collective goal for research to be tangential to the problem under discussion, all variables and factors considered and raised previously were consistent with discussed data collection and data requirement frameworks. To restate, the business understanding and analytical approach, outlined in last week's status report, align with the data requirements and the data collection process as set forth by the data science team. Necessary time was spent contemplating each aspect of the task, as well as problems, at-hand and variables which needed to be identified were delineated prior to the team members being individually dispatched on their discrete tasks. The data science team, through analytic approach, notated data requirements and clearly defined data collection procedures and roles for each team member.

References:

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

International Business Machines. (2016). *Data collection* [Video file]. Armonk, NY: Author.

Laureate. (2018). *Nutri Mondo*. Laureate International Universities. <https://mym.cdn.laureate-media.com/2dett4d/Walden/DSCI/2010/WorldOfDataScience/index.html>