Rich Wagner

Professor Zhe Deng

Data Analytics

25 April 2019

Big Data

Big Data refers to the large amount of data that is collected in this technology era. This data can be structured, semi-structured, and unstructured. Big data has a high volume, high velocity, and/or high variety. Big data is mostly generated in real time and can be collected from websites, social medias, transactional applications, devices, networks, and more. Analyzing big data can lead to better decisions and more efficient processes for the businesses. As more data is able to be collected it is becoming so much more important to store it and analyze it property (IBM).

Big data builds on the concept in the course because this course was mainly about extracting, loading, and transforming data. We used to SQL to store our data, Tableau prep to extract, text files to analyze, and finally RStudio to analyze. These all involve big data. Data is constantly being collected and these processes and programs make it easier to understand and analyze it.

One assignment that specifically interpreted big data was Assignment #8: Association Rules Mining in R. We used a Rstudio to generate a list of rules from 43,366 items purchased across 9,835 transactions at a grocery store. The rules consisted of baskets and showed what customers would likely buy if they bought some products. The rules output also contained the support, confidence, and lift which all contribute to seeing what the customer might buy next. The store could use this data to better understand their customers and see what they will potentially buy next. They can also better map out their store, placing products that customers might buy next to each other. Businesses must do this in order to compete in the tech based business world today. They will fall behind to their competition if they do not.

Works Cited

“Big Data Analytics.” IBM, www.ibm.com/analytics/hadoop/big-data-analytics.