UNIT 1: COLLABORATIVE DISCUSSION 1

Unit 1 was cantered around learning about Big Data Characteristics and management techniques. The main artefact of this topic was a collaborative discussion of the rationale behind the Internet of Things (IOT), in the context of the article by Huxley et al (2020) .

Learning outcomes:

1. Identify and manage challenges, security issues and risks, limitations and opportunities in data wrangling.
2. Critically analyse data wrangling problems and determine appropriate methodologies, tools and techniques (involving preparing, cleaning, exploring, creating, optimising and evaluating big data) to solve them. (University of Essex Online , 2024)

REFERENCES:

- University of Essex Online. (2024) Collaborative Discussion 1: The Data Collection Process: https://www.my-course.co.uk/mod/forum/view.php?id=910307 [Accessed 16th November 2023].

Feedback from Peer:

by Sahr Solar Sumana - Monday, 12 February 2024, 5:10 PM

It is understood that the concept and integration of the IOT has a multitude of benefits as listed by Simbiso including being able to automate the collection of big data, and have that data collected and processed in real time.

Although the limitation of cloud services due to large capacity storage requirements has been listed, there is an outstanding limitation of data cleaning and data integrity requirements. Hellerstein (2008) has stated that "sensor deployment often affects data quality, and many sensors are subject to errors including miscalibration and interference from unintended signals furthermore, any procedure that integrates data from multiple sources can lead to errors". To extend upon this statement a major downside is that data from the IOT can be susceptible to erroneous entries and with the scale at which data is collected a huge number

of erroneous entries can be recorded and not be picked up upon until later stages of the data lifecycle such as during the cleaning or analytical phases.

I agree with Simbiso's point of devices and systems being able to communicate with one another to ensure a productive WSN. Not making this a priority can lead to missing data or erroneous entries as mentioned before, a WSN where devices are not in sync with one another can lead to extreme values where the distribution of data is not normal. With the IOT it may be harder to screen data due to the size of the data and how time consuming it can be Huxley (2020) has stated that "data screening process is often a neglected stage of data preparation".

Reference List:

Hellerstein, J. (2008) *Quantitative Data Cleaning for Large Databases* [Preprint].

Huxley (2020) 'Data Cleaning', *SAGE Research Methods Foundations* [Preprint]. doi:10.4135/9781526421036842861.