

Data Analysis 101

Data Tools:

This module explains how we can store data in different ways and what we can do once it is stored. We can use a spreadsheet for example to store, organize, and analyze data more accessible. I learned that CSV (comma-separated values) is an easy format to read the data from. Once the data is stored it can be used for many things. I learned that once we get the data it is good to start computing stats from it like we can find an average of something from it. Also, it is useful to find correlations between numbers from the data, which can be important to know.

Big Data:

I learned that in 2025 the digital data will be approximately 175 zettabytes. There are challenges to storing big data sets, but can be very useful. The huge amount of data comes from several smaller sources, since there are about 8 billion people the data is saved and stored to become one big data sets. This also includes government research, like satellite imagery, digital books, medical records, and so much more. There are locations called data centers, which are where they entirely house computers and data storage devices. It takes a while to store and needs to be handled responsibly.

Bias in Machine Learning:

It talks about the different types of bias when it comes to algorithms, like predictive algorithms, facial recognition, and language translation. In the criminal justice system, the algorithm was used to say whether or not a person would re-offend and it was unjust because it would underestimate white people and overestimate black people. This is because the bias in the world caused the program to be biased also. When it comes to facial recognition, it is biased because it tends to read lighter people better than darker people. This has caused darker people to be arrested for crimes they did not even commit because facial recognition failed, just like the predictive algorithms. Another bias in machine learning is language translation. The translation can be sexist. For example, something that is not gendered-specific will translate to he or she. Like she is a nurse is a biased way of translation because nursing is not only a female job but the translation picked up on the sexist ways of the world.

Unit Test:

The unit test activity was a mix of all three of the units. It had good questions it ensured the understanding of all the units. Also, I did have to take it more than once so it was not completely easy but it helped me understand the concepts better. Overall, it was a good experience and interesting. I definitely learned some things.