Jacob Daugherty

Cohen

CS 1030

5/1/2023

**Data Analysis 101**

**Data Tools**

Data tools are something that I am extremely familiar with due to the background of college and high school classes I have taken. This section reminded me much of statistics, which I took in high school. There were multiple references such as negative correlation, significance, and overall reading data to come to a solution. The “pre-test” that Khan Academy gave in my opinion was super easy, due to my previous understanding and ability to interpret data. The Post-Test that was offered at the end went just as well as the pre-test. The Post Test felt like a more complex version of the pre-test, in which I scored a 4/4 on both tests. With that in mind, Data Tools were an easier topic to understand due to said statistics class and the computer information classes I am taking right now.

**Big Data**

Big Data, on the other hand, is something that is a bit newer to me. As previously mentioned, I have studied a small portion of data in Computer Information Systems at school. However, the level of data in said class is on a much smaller business level, not a mass storage system. This section was actually pretty interesting to read about and just read into. For example, the estimates of how much data there is out there are crazy to put into perspective. Khan Academy’s resource tends to believe there are 176 zettabytes worth of data. A zettabyte is 2^70 bytes of information, an unbelievable amount of information. On the other side of Big Data is storage, which I have much more familiarity with. The types of storage systems used to store big data are familiar alongside understanding the value of scalable systems. After reading through this section, I have a lot more cohesion on the idea of mass data and how it works, how its stored, and the problems with it.

**Bias in Machine Learning**

Bias in Machine Learning, where have we read this before? Oh, that’s right, Hannah Fry’s Hello World talks about machine learning and algorithms which we have read all semester long. There are multiple points that we have already touched on in Hello World, that were mentioned in this section of learning. For example, there was a section about bias in machine learning and algorithms in crime, and how it may unfairly discriminate against people. There were also biases in terms of main ideas such as medicine and the medical field. This section while in cahoots with Hannah Fry’s book discusses the problems in machine learning, especially bias towards one race or another. This section perfectly encapsulates the problems that we have with machine learning and points in the correct direction to help fix these biases.

**Unit Test**

The Unit Test combined the three previous sections of Data Tools, Big Data, and Bias in Machine Learning, and made one 9-question quiz on the topics. I noticed that there were 3 questions for each topic, in order to gauge if you improved, worsened, or had no change at all. After taking the exam, I got 7 out of the 9 total questions correct, leading to improvements in 2 categories and no change in the other. So overall, I did learn and advance in two areas, which could be expected after studying and reading over information for the end goal being improvement. Overall, this Khan Academy course, I felt was effective and improved my skills in those three categories.