**Khan Academy CS Data Analysis**

**Data Tools**

The lesson covered three main topics: storing data sets using text files, spreadsheets, and databases, basic computer statistics including statistical functions in spreadsheets and SQL for data analysis, and identifying patterns in data sets using trend lines and correlation analysis. Text files are simple but lack tools, spreadsheets organize and analyze data but do not act as a storage mechanism, and databases help store and access data using query language. Basic statistical functions such as sum, average, and max/min are easy to understand, while more complex functions like countif and pivot tables require more knowledge. Identifying patterns in data is essential for predicting trends and finding correlations, but correlation does not imply causation.

**Big Data**

The world is consuming vast amounts of data, estimated to reach 175 zettabytes by 2025, which presents both storage and processing challenges but also new opportunities for analysis. Big data is a collection of various sources, including scientific research and user-facing applications, with groups such as NASA EOSDIS and Internet Archive contributing to the collection. Large storage should be a consideration for running computing systems. There are systems, such as disk arrays and data centers that can store large amounts of data. Engineers use parallel computing to process huge data sets for computer systems in a functioning society.

**Bias in Machine Learning**

One of the most popular types of machine learning is supervised machine learning, where the popular approach is using neural networks, but they can carry biases. I learned that neural networks start with a training phase, and the neural network starts with sets of random values to classify, where it will make the most mistakes and correct them to improve an accurate result. When it comes to applying to the real world, such as predicting the risk of an offense, critiques argue that it can have bias. In an investigation conducted by Propublica, they discovered that algorithms underestimated whites and overestimated blacks. Using algorithms can pose significant consequences for individuals of a certain group.

**Unit Test**

The unit test was relatively easy, with most of the questions coming from lesson 1, specifically computing basic statistics. The focus was on finding commonalities in the presented data. I correctly answered a question about a scatter plot and predicting results, which involved identifying patterns. However, I found three questions from lesson 3 difficult and answered them incorrectly. My overall result for the unit test was six out of nine, which may be a failure, but at least I got all the computer statistics right.