Troy Siegler

Problem Set 1

INFO 3401

1. What Peter Naur is trying to say in this quotation is that while data science is a growing field and important, the only job as a data scientist is to manage data and clean it. As opposed to looking deeper into data and finding connections, real world solutions, etc, Naur is claiming that data scientists should simply be managing and cleaning the numbers and then turning that information back over to the professionals within the field. This is problematic because you cannot simply treat data basic numbers; there is always a bigger story that one has to keep in mind while evaluating this data. It’s always important to have a story when evaluating data; if one just churns and cleans data blindly, they could miss the overall pattern/story that the specific field is trying to address within said data.
2. Tukey and Naur in the 1970s described data science sort of as a subfield from other common forms of science at the time. They believed that data science was simply a practice within its own field regarding the cleaning, sorting, and manipulation of data/numbers. During this time period, these two scientists (Tukey and Naur) believed that it was simply the data scientists job to take the data, clean it up and turn it back in as opposed to the definitions that we began to see later on in the 2000s. Whereas in 2002, we began to see these definitions change. Data science became less about the simple science of exploring data sets and crunching numbers, but the science was more about the relativity to the field in which the data was coming from. Instead of just defining this practice as a spreadsheet or field of numbers, data science actually meant “almost everything that has something to do with data: Collecting, analyzing, modeling...... yet the most important part is its applications--all sorts of applications”. This began the era in which data scientists starting using data for the good of their businesses, companies, schools, etc. They began to see that there is so much to learn from this data that is applicable to our everyday lives.
3. Data continues to grow more and more everything single day. Presidential candidate Andrew Yang was recently quoted stating the “data is becoming more expensive an oil”, but unlike oil, data in this day and age is a renewable resource. A few reasons/sources as to how/why this data is growing are listed in the following sentences. First, hardware is improving each and every year. Hardware is quickly becoming faster, cheaper, and more accessible to everyone. For example, Moore’s law states that microchips get smaller and smaller every two years. Next, data is much more relevant and ubiquitous in our lives now more than ever. For example, social media and apps like Venmo generate so much data from extremely basic parts of our lives. This data is then sold or used for target advertising in the future. Lastly, I believe the rise in data can also be contributed to the rise of online consumerism. Sites like amazon have made our lives, as consumers, much easier but also leaves an unbelievably large data trail that can be used by larger companies in the future.
4. A. In order to get back to the root, which is the highest-level directory in the entire computer, you could use the cd / function.

B. To move back to the home directory, the default directory for each computer, you would use the cd ~ (tilda) function.

C. To get back to the parent directory from the place you are currently in, you would use the cd .. function and that would take you back a section to the parent directory.

Having these functions to navigate through the file structure are extremely important because they make it much easier, quicker, and more efficient to get from place to place within your computer’s files. It would take much longer and require so much more effort if we did not have these commands and were forced to navigate from section to section manually.

1. These commands are fairly basic, and I believe I am beginning to understand them quite well. First, the cd ~ function will take us to our home directory, which is the default for all computers that contains most of the files/directories. Next, the mkdir ./problem\_set\_1 will create a new directory titles “problem\_set\_1” within the home directory that we are currently in. After that, using the cd .. function will take us back to the parent directory that we were previously in. And lastly, using the pwd function will output the current directory that we are in. For this particular case, the pwd function will output /Users (at least that is what came up on my terminal), which shows us the current “parent” directory that we are in because we used the cd .. function.