Priya Panati

Problem Set One

1. Naur thinks data is formal and structured. He thinks that data science is just the compiling of data but the analysis would fall into a different field like statistics, computer science, or the field the data relates to. This is problematic because if someone is dealing with multiple datasets they would have to make choices about what is and isn’t relevant so that already goes against his rule that data is formal and structured. There will be inevitable human bias involved. There are also other steps a data scientist needs to do. They need to have background knowledge and do research to understand what the data means.
2. The Journal of Data Science emphasizes the applications of what the data represents and the national science foundation emphasizes all the fields that data science falls into. This is very different than what Naur and Tukey defined it as. Unlike Naur these definitions show that data science isn’t just the compiling of data it is also the analysis and the applications.
3. Personal computers- Even small businesses and people can record data, Cell phones- Track usage and location data, Machine Learning- the ability to predict outcomes based on past data. Cheaper and faster technology- Enables everyone to have access to computers, hard drives, and people and companies who can afford more have larger amounts of data.

Facebook, Twitter, social media- Has lots of data based on likes, comments, and clicks. Facebook data is being used to target advertisements to each individual. Another source of data is E commerce- E commerce data tracks your buying patterns and thus can create a profile based on your spending patterns. Credit and debit card data is also another data source. Large financial institutions can predict fraud by looking at past spending patterns and locations the card has been used.

1. 1. Root: cd/
   2. Home: cd ~
   3. Parent: cd ..
2. 1. Cd~: home directory
   2. Mkdir ./problem\_set\_1: go to problem\_set\_1 from mkdir
   3. Cd.. : change current dictionary to parent dictionary
   4. Pwd: print working dictionary
3. Print everything in working dictionary problem1
4. Cd~

Ls

Cd/ datasets/activedata