

CS982: Big Data Technologies  
CS989: Big Data Fundamentals

## Lab 2 – Data Handling and Manipulation

1. Import the necessary libraries (numpy and pandas)
2. Import the Pokemon dataset available at <https://www.kaggle.com/alopez247/pokemon>
3. Print the first 10 entries
4. How many observations and columns are there?
5. Print the names of all of the columns
6. Sort by attack power from high to low
7. Describe the defence power for each Pokemon type
8. What are the mean, median, max and minimum of the total column for each Pokemon type?
9. What is the most common Pokemon type?
10. Find an interesting dataset online and download it for analysis. Good places to look include:
  - i. Kaggle - <https://www.kaggle.com/datasets>
  - ii. UCI Machine Learning Repository - <https://archive.ics.uci.edu/ml/datasets.html>Explore the dataset starting with some techniques that we explored in the lecture and this lab