



# INFT6201 - BIG DATA

## LAB PROJECT 1

### (PREPARATION)

This lab project is based on the dataset “Pokemon with Stats” by Parradas (2016).

#### PREPARATION

In preparation for lab project 1, load the pokemon.csv dataset into your Jupyter notebook and use it to practice the topics that we discussed in the lecture up to this point. In the Week 2 lab, practice the following concepts:

- Load and explore the \*.csv datasets
- Look up further details of the Pandas import functions in:  
[https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read\\_csv.html](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html)
- Explore the loaded DataFrame using Pandas  
(<https://pandas.pydata.org/pandas-docs/stable/reference/frame.html>)
- Use the slice notation to select rows and columns
- Calculate basic statistics on the dataset (e.g. variance, standard deviation, percentiles, etc)
- Explore further calculations in numpy

The specification for the lab project 1 will be released separately at the start of Week 3.

#### REFERENCES

Barradas, A. (2016). Pokemon with Stats. <https://www.kaggle.com/abcsds/pokemon>