CSE231 - Lab 15

A Soft Intro to Pandas

Pandas? Like, 🐼 ← these pandas?

No, not those pandas. Pandas is a widely used data science module for Python that works around the concept of creating "data frames".

Pandas, matplotlib, and numpy are like the three modules of the data science apocalypse. They are super essential to learn if you're interested in data science subjects, like machine learning.

You will come to hate this class even more than you already do after learning about this module. This module cuts down the amount of CSV-reading code by… a substantial amount.

The Basics

There are two fundamental classes included with pandas. The `Series()` class, and the `DataFrame()` class.

It might be easier to think of a `Series()` object as simply a dictionary with `list` values. *Exclusively*, `list` values. Example:

{ "grades" : [50, 90, 100, 45] }

A `DataFrame()` object, then, is a dictionary of `Series()` objects. Example:

{ {"names": ["Billy", "Wungus", "Dirty Dan", "Enbody"]},

  {"grades": [50, 90, 100, 45] } }

It's time for a loooong demo

We'll be jumping into my code editor here since we're going to be working with big data sets