

Lecture-04-CMSC320

- Homework #1 released on Monday
- Slides will have suggested readings attached, submitting notes or a summary of the reading will earn extra credit points

Python:

- Weakly typed language
- Interpreted - not compiled
- Slow to run
- Used frequently in web dev and data analysis
- All for loops are for-each loops

Pandas:

- Python package used for data analysis
- Used for analyzing data locally that can fit in memory
- Almost entirely functional
- Most basic object is a **dataframe** that is a collection of **series**
- **Series:** A one dimensional array of data with an index, index is most commonly integers (a column)
- We almost never access a series by an index
- Never loop over a series
- Pandas can do anything with a series you can do with an array
- **apply():** Takes a function as an argument and returns a series with that function applied

- Most Pandas functions generate a new series, not in place
- **Dataframe:** A collection of series organized by columns
- We do not access rows via index
- We do access columns via their names
- Dataframe to series: 'df["Column Name"]'
- 'df[df["Column Name"] > 10]' returns a filtered dataframe of column names with value greater than 10, first call returns series of booleans
- Dataframe indices are static, removal of a entry removes that index

SQL:

- One SQL homework
- Databases are used when data won't fit in memory
- **SQL:** Structured Query Language
- SQL is how we ask our database questions, not a programming language