Unit 4-2 Intro to Pandas

* Pandas Defined:
  + An open source library that serves up high-performance, easy-to-use data structures and data analysis tools for Python; your new best friend
* Pandas
  + The Pandas package provides us with DataFrame operations. Typically, Pandas will be our package of choice for creating and manipulating tabular data that are represented as DataFrames, with each column stored as a Series object.
  + Think of DataFrames as the Python equivalent of an Excel worksheet.
* Pandas Features a Full Set of DataFrame and Series Operations
  + With Pandas you can:
    - Read data into a DataFrame or Series.
    - Explore Pandas DataFrames and Series using their attributes and methods.
    - Manipulate DataFrames and Series.
    - Create graphs and other visualizations in order to analyze data and communicate results.
* Introducing Series and DataFrames
  + Pandas Series and DataFrames are objects that hold data in row-and-column tables. These objects have a large number of methods that help us manipulate and analyze them. Note that:
  + A DataFrame holds one table.
  + A Series is a single-column DataFrame.
  + Series and DataFrames have many identical attributes and methods.
* Reading Data Into a DataFrame
  + Pandas can read in most common data formats — including data from databases — using read() functions.
  + Interacting with Pandas' read() functions is similar across various data types:
    - pd.read\_csv('data\_file.csv'): Retrieves data from a local CSV file or from the web.
    - pd.read\_table('data\_file.xml'): Retrieves data from other delimited data text files.
    - pd.read\_json('data\_file.json'): Reads in a JSON object and converts it to a DataFrame.
* Attributes and Methods
  + Attributes and methods are similar in that they are capabilities a DataFrame can provide.
  + Attributes are qualities a DataFrame knows about itself.
    - For example, a DataFrame knows its column names and the number of rows and columns.
  + Methods are functions that can be run on a data set to more deeply analyze or manipulate the DataFrame.
    - Methods can accept inputs and parameter values within their parentheses. For example:
  + df.sort\_values("column\_name", ascending=False).
  + By using attributes and methods, we can tap a rich set of functionality within Pandas without importing other packages.
* Exploring DataFrames With Attributes and Methods
  + Example attribute for a DataFrame named df:
    - df.shape: Returns an array of the number of the rows and columns in the df DataFrame.
    - df.columns: Returns a list of the column names.
    - df.index: Returns a list of the row names.
  + Example methods for a DataFrame named df:
    - df.head(): Returns the top rows of the df DataFrame.
    - df.sort\_values(): Returns a copy of the DataFrame sorted by one or more columns.
    - df.describe(): Returns the summary statistics of all numeric columns in the DataFrame.