Kevin Dyer

10/27/21

HWRS 501

Pandas Cheat Sheet

* Pandas is a software library for python used for data interpretation and analysis. It provides fast, flexible, and intuitive data analysis tools that give python greater utility for running tasks. Pandas are different from other object types we’ve worked with because they are so broad and have so many functions and tools associated with them.
* There are multiple options to make a pandas dataframe from scratch. One option is to make a dataframe and single out certain variables from a list: “df = pd.DataFrame(data, columns = ['Name', 'Age'])”. Another option is to simply just use the whole data: “df = pd.DataFrame(data)”. And one can also use arrays as well: “df = pd.DataFrame(data, index =['rank1','rank2','rank3','rank4'])”.
* To slice certain parts of data in a pandas dataframe using loc one would enter something like: “df1.loc[1:5, 'Film':'EA1']”. Using iloc, one would type: “dfl.iloc[1:3, 0:6]”. These functions allow for the selection of rows and columns by their associated number in python. Once selected they can be sliced out into their own array or list to be analyzed without data/information that isn’t pertinent and may make certain tasks more difficult. Loc is label based whereas iloc is integer based.
* Indexing is a way to refer to individual objects within a dataframe by its positions. It differs from other columns in that it does not need to be entire rows and columns selected, it can just parts of them, allowing for greater control over data analysis and manipulation.
* Key methods associated with pandas dataframes include: loc[:], iloc[:], head(), describe(), and drop\_duplicates().
* Key attributes associated with pandas dataframes include: index, Dtypes, shape, count, size, and columns.