General Assembly

STUDY GUIDE

INTRO TO PANDAS

Key Terms and Definitions

» **Pandas:** A package that provides us with DataFrame operations. Pandas is mainly used for data manipulation, cleaning, and parsing. Pandas Series and DataFrames are objects that hold data in row-and-column tables.

- A DataFrame holds one table.
- A Series is a single-column DataFrame.

» **Attributes and Methods:** An object in Python has inherent qualities it knows about itself (attributes) as well as functions it inherently knows how to perform (methods).

» **Reading Files:** .read_csv()/.read_json(),/.read_table() creates a DataFrame based on an external file source.

» **Useful DataFrame Attributes:**

- .shape: Used to return the number of rows and columns.
- .columns: Returns a list of the column names.
- .index: Returns a list of the row names.
- .dtypes: Used to return the type of data stored in each column.

» **Useful DataFrame Methods:**

- .head(): Returns the top rows of the "df" DataFrame.
- .sort_values(): Returns a copy of the DataFrame sorted by one or more columns.
- .describe(): Returns the summary statistics (count, mean, and standard deviation, along with various percentiles) of all numeric columns in the DataFrame.

Guiding Questions

1. Why is it important to always check the.dtypes of a Pandas DataFrame?
2. How are Series and DataFrame objects related?

Additional Resources

1. DataCamp:
    - » Intermediate Python for Data Science. See Section 2, "Dictionaries & Pandas," specifically "Pandas, Part 1" and "Pandas, Part 2". You might also check out the "Filtering Pandas DataFrame" content in Section 3, "Logic: Control Flow and Filtering."
    - » Pandas Foundations See Section 1, "Data Ingestion & Inspection."
2. GA Pandas Demo Video
3. Pandas Repl.it Challenges:
    - » Challenge 1
    - » Challenge 2
    - » Challenge 3
4. Pandas DataFrame Documentation