

Barok Tsegaye
9.15.23
Pandas

Pandas

- A python library that can make analyzing data easier
 - Primarily used for data science

Dataframes

- A Pandas object that is used to store a dataset
- Information is organized in rows and columns
- Dataframes simplify common operations like storing data

Series

- A Pandas object used to create dataframes
- Seen as a one dimensional list of data.
 - Think of it as a single column in a dataframe

Indexing

- Main techniques:
 - `df.loc[]`
 - `name.loc[row_label, col_label]`
 - `df.iloc[]`
 - `name.iloc[row_index, col_index]`

Selection

- The process of accessing a subset of a dataframe.
 - You can select these subsets by using loc and iloc

Filtering

- Selecting values of a dataset where certain conditions are true.
 - Popular pattern:

- Df[condition]

Combining Dataframes

- Three Techniques:
 - Concatenate: Natively combines along an axis
 - Merge: Combine through shared column
 - Join: Combine using shared indices.
 - Inner
 - Left
 - Right
 - Outer