Python for Economics

Zhenhua He
Morning session, 9/24/2021
Table of Contents

This course is divided into numbered lessons

14. Pandas
15. Pandas Additional Materials

- Lunch
- (More Lessons after Lunch)
Lesson 14
Pandas

Use Python Pandas library to manipulate data
Pandas Review - Series

- One-dimensional labeled array
- Capable of holding any data type (integers, strings, floating point numbers, etc.)
- Example: time-series stock price data

Index | Value
--- | ---
A | 0
B | 1
C | 2
D | 3
E | 4
Pandas Review - DataFrame

- Primary Pandas data structure
- A dict-like container for Series objects
- Two-dimensional size-mutable
- Heterogeneous tabular data structure

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>x</td>
<td>0.1</td>
<td>True</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>y</td>
<td>2.4</td>
<td>False</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>z</td>
<td>1.9</td>
<td>True</td>
</tr>
<tr>
<td>D</td>
<td>NA</td>
<td>w</td>
<td>8.3</td>
<td>False</td>
</tr>
<tr>
<td>E</td>
<td>9</td>
<td>a</td>
<td>6.8</td>
<td>False</td>
</tr>
</tbody>
</table>
Pandas Review - Creating a Data Frame

Ways to do so:

• from Dictionary
• from Numpy array
• from file (csv)
### Example Data for Exercises

National Economics Data; a selection of countries; data sourced from [https://worldpopulationreview.com/countries/countries-by-gdp](https://worldpopulationreview.com/countries/countries-by-gdp)

<table>
<thead>
<tr>
<th>Country</th>
<th>Continent</th>
<th>GDP (Billion dollars)</th>
<th>Population (Millions)</th>
<th>GDPPerCapita (Thousand dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>America</td>
<td>18624.5</td>
<td>332.9</td>
<td>66.7</td>
</tr>
<tr>
<td>China</td>
<td>Asia</td>
<td>11218.3</td>
<td>1444.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Japan</td>
<td>Asia</td>
<td>4936.2</td>
<td>126.1</td>
<td>43.6</td>
</tr>
<tr>
<td>Germany</td>
<td>Europe</td>
<td>3477.8</td>
<td>83.9</td>
<td>49.5</td>
</tr>
<tr>
<td>India</td>
<td>Asia</td>
<td>2259.6</td>
<td>1393.4</td>
<td>2.3</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Europe</td>
<td>2647.9</td>
<td>68.2</td>
<td>42.9</td>
</tr>
<tr>
<td>France</td>
<td>Europe</td>
<td>2465.5</td>
<td>65.4</td>
<td>44.0</td>
</tr>
<tr>
<td>Italy</td>
<td>Europe</td>
<td>1858.9</td>
<td>60.4</td>
<td>34.6</td>
</tr>
<tr>
<td>Brazil</td>
<td>America</td>
<td>1795.9</td>
<td>214.0</td>
<td>9.6</td>
</tr>
<tr>
<td>Canada</td>
<td>America</td>
<td>1529.8</td>
<td>38.1</td>
<td>48.1</td>
</tr>
</tbody>
</table>
Exercises - Common Notebook

Go to Google Classroom assignment Pandas “National Economics”

Tasks
1. Upload the example data into your Colab session
2. Create the DataFrame
3. Keep this notebook open for later exercises
DataFrame: data retrieval

- Retrieve a column
- Retrieve multiple columns
- Retrieve a row
- Retrieve multiple rows
- Drop entries
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-2”

Tasks

● Follow the examples for Data Retrieval

Return to National Economics Data notebook

● Complete Data Retrieval Exercises
Break Time Reminder Slide

10 minutes break
DataFrame: Input and Output

- Read and Write
- Different file formats
- Quick Stats Facts
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-4”

Tasks
- Follow instructions for the examples

Return to National Economics Data notebook
- Complete Input and Output Exercises
DataFrame Grouping

• `groupby()`

Group DataFrame by value in one or more columns

The grouped dataframe is like a list of sub-dataframes.
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-5”

Tasks

● Follow instructions for the examples

Return to National Economics Data notebook

● Complete Grouping Exercises
DataFrame Plotting

- `plot()`: Make plots of Series or DataFrame. By default, `matplotlib` is used.
- `hist()`: Make a histogram of the DataFrame columns.
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-6”

Tasks
● Follow instructions for the examples

Return to National Economics Data notebook
● Complete Plotting Exercises
DataFrame - Handle Missing Data

Handle missing data (NaN in DataFrame)

- `dropna()`: drop rows or columns with missing values
- `fillna()`: forward or backward fill from existing values
- `interpolate()`: linear fill between existing values
DataFrame - Handle Duplicate Data

Handle duplicate data

- `duplicated()`: check duplicate values
- `drop_duplicates()`: drop duplicate rows, 
or drop duplicates from a particular column
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-7”

Tasks

- Follow instructions for the examples

Return to National Economics Data notebook

- Complete Missing and Duplicate Data Exercises
Break Time Reminder Slide

10 minutes break
Lesson 15
Pandas Additional Materials

Use Python Pandas library to manipulate data
DataFrame: Operations/Manipulation

- Selecting with slicing
- Filtering
- Sorting
  - sort by index
  - sort by values
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-3”

Tasks

● Follow instructions for the examples

Return to National Economics Data notebook

● Complete Operations/Manipulation Exercises
Merge DataFrame with a database-style join.

- left join
- right join
DataFrame - Merging Data

Merge DataFrame with a database-style join.

- inner join
- outer join
Examples and Exercises

Go to Google Classroom assignment Pandas “DataFrame-8”

Tasks

● Follow instructions for the examples

Return to National Economics Data notebook

● Complete Merging Exercises
Lunch Break Reminder Slide

1 hour break