

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)

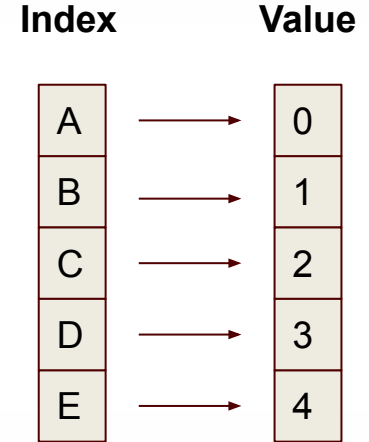
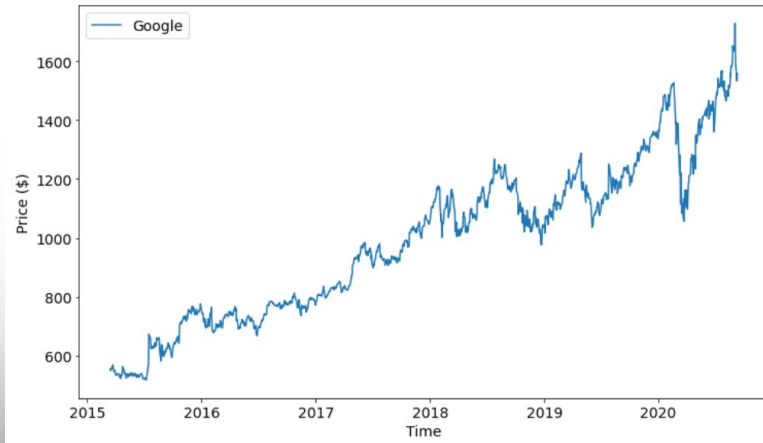
Lesson 14

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



Pandas Review - DataFrame

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

	C1	C2	C3	C4
A	0	x	0.1	True
B	1	y	2.4	False
C	2	z	1.9	True
D	NA	w	8.3	False
E	9	a	6.8	False

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>

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

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

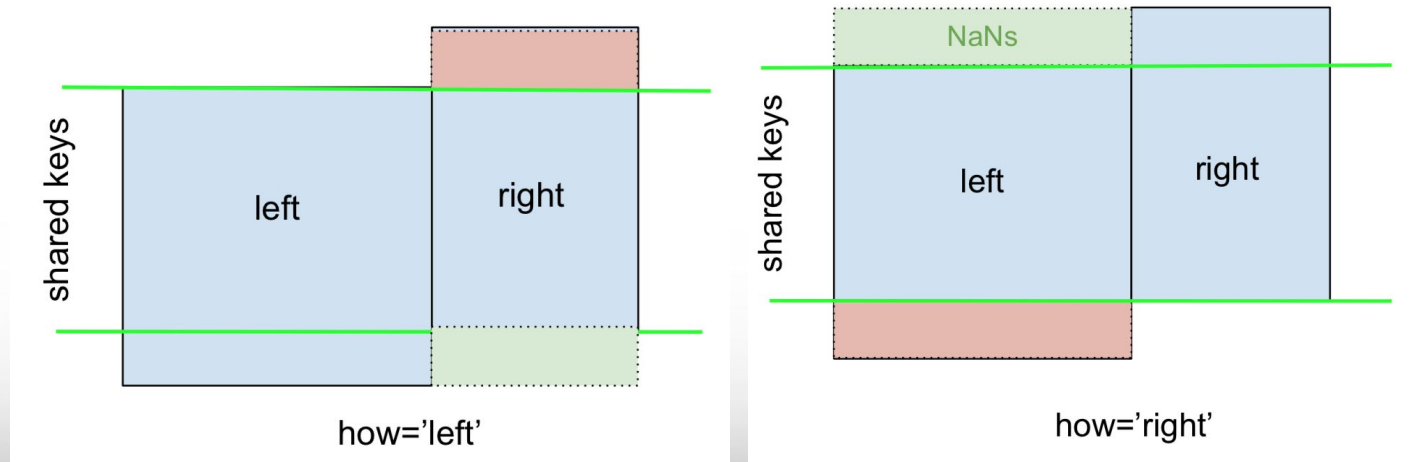
Return to National Economics Data notebook

- Complete Operations/Manipulation Exercises

DataFrame - Merging Data

Merge DataFrame with a database-style join.

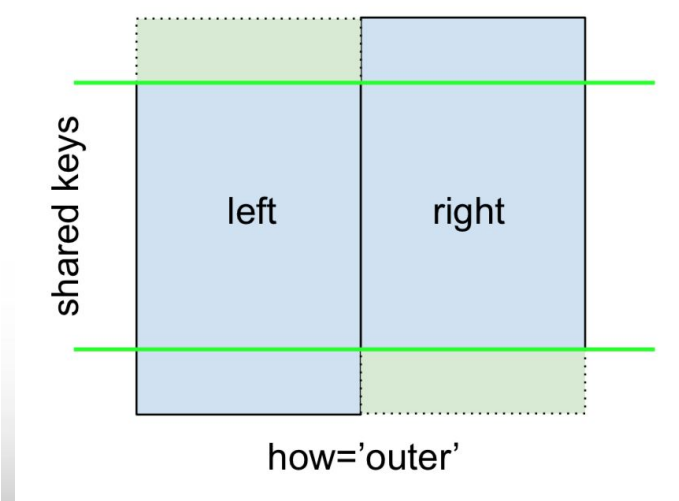
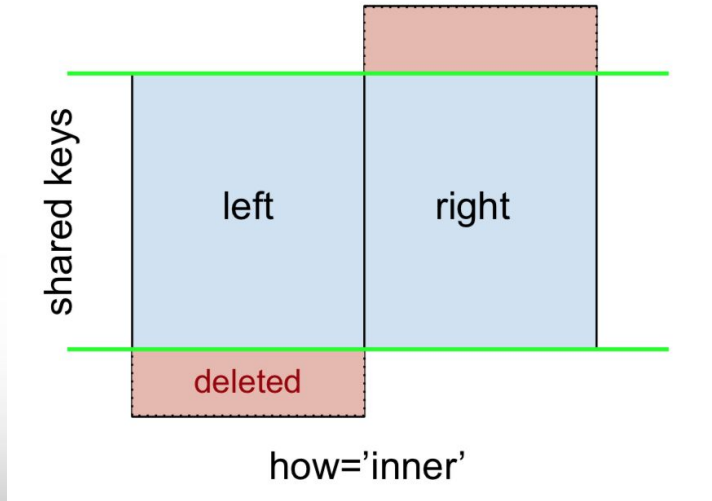
- left join
- right join



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

