GENERAL ASSEMBLY

Pivot Tables in Pandas

Patrick D. Smith

Lead Instructor, DSI DC

LEARNING OBJECTIVES

- Demonstrate how to use pandas.pivot_table
- Demonstrate how to use the parameters: values, columns, index, aggfunc, and margins

OPENING

Pivot Tables in Pandas

We've all heard about them; What are Pivot Tables?

- A pivot table can automatically sort, count, total or give the average of the data stored in one table or spreadsheet, displaying the results in a second table showing the summarized data
- Pivot tables are also useful for quickly creating unweighted cross tabulations. The user sets up and changes the summary's structure by dragging and dropping fields graphically.

Pivot Tables in Pandas

We've all heard about them; What are Pivot Tables?

The function pandas.pivot_table can be used to create spreadsheet-style pivot tables. It takes a number of arguments:

- **data:** A DataFrame object
- **values:** a column or a list of columns to aggregate
- **index:** a column, Grouper, array which has the same length as data, or list of them. Keys to group by on the pivot table index. If an array is passed, it is being used as the same manner as column values.
- **columns:** a column, Grouper, array which has the same length as data, or list of them. Keys to group by on the pivot table column. If an array is passed, it is being used as the same manner as column values.
- **aggfunc:** function to use for aggregation, defaulting to numpy.mean
- **margins:** boolean, default False, Add row/column margins (subtotals)

Pivot Tables in Pandas

We've all heard about them; What are Pivot Tables?

ix	Item	CType	USD	EU
0	Item0	Gold	1	1
1	Item0	Bronze	2	2
2	Item0	Gold	3	3
3	Item1	Silver	4	4

ix=Item	Bronze	Gold	Silver
Item0	2	2 = mean(1,3)	NaN
Item1	NaN	NaN	4

```
d.pivot_table(index='Item', columns='CType', values='USD',  aggfunc=np.mean)
```

Demo: Pivot Tables in Pandas

Independent Practice

Pivot Tables in Pandas - Independent Practice

Using the rock song data, complete the following requirements:

- Import pandas and numpy
- Read in rock.csv as a data frame
- The simplest pivot table must have a dataframe and an index . Let’s use the ARTIST CLEAN as our index.
- How about indexing on multiple values. Let's look at the data by ARTIST CLEAN and Release Year next.
- Explicitly define the 'PlayCount' column by using the values field.

EXIT TICKET

DON’T FORGET TO FILL OUT YOUR EXIT TICKET