

# Python\_Module13

October 18, 2020

## 0.1 Author: Joseph Vargovich

```
[2]: #Import libraries
import pandas as pd
import numpy as np
```

## 1 Exercise 1 - Table for gender and year combinations.

```
[17]: #a. Load the data
Survey = pd.read_csv("http://www.lock5stat.com/datasets/StudentSurvey.csv",
    ↪na_values=['', ' '])
Survey.head()

#b. Cleanup and select what we need.
Survey = Survey[['Year', 'Gender']]

Survey.sort_index(axis=0, level='Year')

#Grab the amount of Female and Male Students in each Year.
Survey = Survey.groupby(['Gender', 'Year'], as_index=False).size()
#Convert our Series back to a dataframe.
Survey = Survey.to_frame(name = 'n').reset_index()
print(Survey)

#Create a pivot table
pivotTable = Survey.pivot(index='Gender', columns='Year', values='n')

pivotTable = pivotTable[['FirstYear', 'Sophomore', 'Junior', 'Senior']]
pivotTable.head()
```

	Gender	Year	n
0	F	FirstYear	43
1	F	Junior	18
2	F	Senior	10
3	F	Sophomore	96
4	M	FirstYear	51

5	M	Junior	17
6	M	Senior	26
7	M	Sophomore	99

```
[17]: Year      FirstYear  Sophomore  Junior  Senior
Gender
F           43          96         18      10
M           51          99         17      26
```

### 1.1 Exercise 3 - Working with two simple data sets.

```
[71]: ADict = {'Name': ["Alice", "Bob", "Charlie"], 'Car': ["Ford F150", "Tesla Model_
↳III", "VW Bug"]}
BDict = {'First.Name': ["Bob", "Charlie", "Alice"], 'Pet': ["Cat", "Dog", "
↳Rabbit"]}

A = pd.DataFrame(data=ADict)
print(A)
B = pd.DataFrame(data=BDict)

#a. Join together with pd.concat and join options.
#Concat (like cbind on axis=1)
result = pd.concat([A, B], axis=1, sort=False)
print("\na. Concatted Tables: \n" )
print(result)

result = A.merge(B, how='left', left_on='Name', right_on='First.Name')
print("\na. Joined Tables: \n" )
print(result)

#b. Add row for Alice's guinea pig to A
guineaDict = {'First.Name': ["Alice"], 'Pet': ["Guinea Pig"]}
guineaDf = pd.DataFrame(guineaDict)
B = pd.concat([B, guineaDf], axis=0, sort=False)
print("\nb. Table B with new entry: \n" )
print(B)

#c. Combine the new dataframes together using concat and join.
#Using concat without join logic, shaping errors will occur.
placeholderDict = {'Name': [None], 'Car' : [None]}
placeholderDf = pd.DataFrame(placeholderDict)
A = pd.concat([A, placeholderDf], axis=0, sort=True)
result = pd.concat([A, B], axis=1, sort=False) #This will error unless we add_
↳an None row to A.
print("\nc. Concatted Tables: \n" )
print(result)
```

```
#Using join, we can join with different shapes.
result = A.merge(B, how='left', left_on='Name', right_on='First.Name')
print("\nc. Joined Tables: \n" )
print(result)
```

	Name	Car
0	Alice	Ford F150
1	Bob	Tesla Model III
2	Charlie	VW Bug

a. Concatted Tables:

	Name	Car	First.Name	Pet
0	Alice	Ford F150	Bob	Cat
1	Bob	Tesla Model III	Charlie	Dog
2	Charlie	VW Bug	Alice	Rabbit

a. Joined Tables:

	Name	Car	First.Name	Pet
0	Alice	Ford F150	Alice	Rabbit
1	Bob	Tesla Model III	Bob	Cat
2	Charlie	VW Bug	Charlie	Dog

b. Table B with new entry:

	First.Name	Pet
0	Bob	Cat
1	Charlie	Dog
2	Alice	Rabbit
0	Alice	Guinea Pig

c. Concatted Tables:

	Car	Name	First.Name	Pet
0	Ford F150	Alice	Bob	Cat
1	Tesla Model III	Bob	Charlie	Dog
2	VW Bug	Charlie	Alice	Rabbit
0	None	None	Alice	Guinea Pig

c. Joined Tables:

	Car	Name	First.Name	Pet
0	Ford F150	Alice	Alice	Rabbit
1	Ford F150	Alice	Alice	Guinea Pig
2	Tesla Model III	Bob	Bob	Cat
3	VW Bug	Charlie	Charlie	Dog

4	None	None	NaN	NaN
---	------	------	-----	-----