9. Pandas Definition, Indexing, Read from file

PYTHON COURSE

SIN YONG TENG

1

Pandas : Role and Usage

1. Pandas is a fast and flexible package for data analysis.
2. NumPy is more suitable for numeric matrix, while Pandas is suitable for Tabular. You can call Pandas the “Excel of Python”.
3. Pandas have good support for importing/exporting data, pre-processing data, sorting, etc. However Pandas is less efficient for computation.
4. You can convert data from Pandas to NumPy and vice-versa.

2

Install Pandas

1. Right click on a tab in Notepad++

2. Open containing folder in CMD

3. Install pandas from pip

>pip install pandas

4. Import pandas

```
import pandas as pd
```

3

How to define a DataFrame?

Pandas uses DataFrame (df), this is how to define one:

Data: List/Dict/ndarray

```
df = pd.DataFrame (data, columns = ['A','B', 'C'], index=['First', 'Second', 'Third'])
```

Horizontal header

Vertical title

	A	B	C
First			
Second			
Third			

4

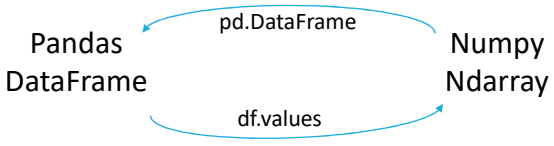
Challenge 1: Try to define.

List vs dict: What is the difference?

```
import pandas as pd

data = {'A':  [1,1,1],
        'B': [2,2,2],
        'C': [3,3,3]}

data=[[1,1,1],[2,2,2],[3,3,3]]

df = pd.DataFrame (data, columns = ['A','B', 'C'], index=['First', 'Second', 'Third'])

print(df)
```

5

DataFrame to Ndarray and back




```
import pandas as pd

data=[[1,2,3],[4,5,6],[7,8,9]]

df = pd.DataFrame (data, columns = ['A','B', 'C'], index=['First', 'Second', 'Third'])

print(df)

#Pandas to Numpy
x=df.values
print(x)


#Numpy back to Pandas
df2=pd.DataFrame(x, columns=df.columns, index=df.index)
print(df2)
```

6

Pandas Indexing

```
import pandas as pd

data=[[1,2,3],[4,5,6],[7,8,9]]

df = pd.DataFrame (data, columns = ['A','B', 'C'], index=['First', 'Second', 'Third'])

print(df)

#Directly index column name
print(df[['A']])

#Use iloc for integer index
print(df.iloc[0,1])

#Use loc for index name
print(df.loc['Second'])
```

7

Challenge 2: Indexing

Prepare a Pandas **DataFrame** as the following:

	A	B	C
First	“Hello”	2	3
Second	4	“World”	6
Third	7	8	“Pandas”

Use indexing to find “Hello” “World” and “Pandas” **from the DataFrame**.

Then print “Hello World Pandas” using them.

8

Read from Excel

Directly read/write from excel file (xlsx):

pd.read_excel(filename) **df.to_excel(filename)**

Read/write from csv file (csv):

pd.read_csv(filename) **df.to_csv(filename)**

Prerequisites:

>pip install xlrd

>pip install openpyxl

9

Challenge 3: Technology Evaluation

1. Prepare the following matrix in an excel file. Read it using pandas

Technology	Energy	Environmental Impact	Cost
A	30	70	70
B	40	50	60
B	40	50	60
C	30	80	50
D	70	50	50

2. Drop duplicated technology “B” using df.drop_duplicates()

3. Sort the technology DataFrame by “Energy” using df.sort_values(column_name).

4. Find the cost of the technology that creates the highest energy using indexing.

10

Homework: Technology selection II

1. Prepare the following table in excel, and read using Pandas. Drop the repeating data.

Technology	Power	Steam Consumption	Cost
A	80	70	70
B	20	50	60
C	40	30	20
D	70	50	50
D	70	50	50

2. Remove the technology with the highest cost.
3. Normalize the data using min-max normalization (criteria-wise). You can do this in NumPy.
4. Using weight of Power=1.3, Steam Consumption=1 , Cost=-0.9 : Find the weighted average of each technology.
5. Find the technology with the highest weighted average.

11

Conclusion

1. Pandas Roles and Usage
2. Install Pandas
3. DataFrame Define
4. DataFrame definition: List vs Dict
5. DataFrame to Ndarray
6. DataFrame indexing
7. Example for indexing
8. Read/Write in Pandas
9. Technology Evaluation Example
10. Technology Selection Example

12