Python Programming

Akshita Chanchlani

Pandas

Pandas

- Pandas is Python package for data analysis.
- Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures.
- The name Pandas is derived from the **word Panel Data** – an Econometrics from Multidimensional data.
- It Provides built-in data structures which simplify the manipulation and analysis of data sets. provides tools for data manipulation: reshaping, merging, sorting, slicing, aggregation etc.
- Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Pandas Essential Concepts / Data Structures of Pandas

- **Series**
  - A Series is a named Python list (dict with list as value), one-dimensional array like structure with homogeneous data.
  - **Example:**
    - { 'grades' : [50,90,100,45] }
    - a one-dimensional, labeled array capable of holding any data type – axis labels are collectively referred to as the "index"
- **Data Frame**
  - A DataFrame is a dictionary of Series (dict of series).
  - DataFrame is a two-dimensional array with heterogeneous data.
  - Example:
    - a 2-dimensional data structure with columns of potentially different types (essentially a high performance table object)
    - **Example:**

      { { 'names' : ['bob','ken','art','joe']} { 'grades' : [50,90,100,45] } }

Creating Data Frame

- A pandas Data Frame can be created using various inputs like − Lists, dictionary , Series , Numpy ndarrays
- **Creating Empty Data frame**

```
import pandas as pd
df = pd.DataFrame()
 print (df)
```

- **Creating Dataframe using list**

```
import pandas as pd
data = [1,2,3,4,5]
df = pd.DataFrame(data)
 print (df)
```

- **Creating Dataframe by giving column names**

```
import pandas as pd
data = [[stud1',10],[‘stud2',12],[‘stud3',13]]
df = pd.DataFrame(data,columns=['Name','Age'],dtype=float)
print(df)
```

- **Creating Dataframe using Dictionary**

```
import pandas as pd
 data = {'Name':[‘n1, ‘n2', ‘n3', ‘n4'],'Age':[68,44,26,45]}
df = pd.DataFrame(data, index=['rank1','rank2','rank3','rank4'])
```

Reading data using pandas

- `#Read csv file`
- `df = pd.read_csv("filename.csv")`

  - **There is a number of pandas commands to read other data formats:**
  - pd.read_excel('myfile.xlsx',sheet_name='Sheet1', index_col=None, na_values=['NA'])
  - pd.read_stata('myfile.dta')
  - pd.read_sas('myfile.sas7bdat')
  - pd.read_hdf('myfile.h5','df')