UECS3213 / UECS3453 Data Mining

SESSION: January 2019

Lab 4: Introduction to Pandas DataFrame

Introduction

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Pandas is well suited for many different kinds of data:

- Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
- Ordered and unordered (not necessarily fixed-frequency) time series data.
- Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
- Any other form of observational / statistical data sets. The data actually need not be labeled at all to be placed into a pandas data structure

The two primary data structures of pandas, Series (1-dimensional) and DataFrame (2-dimensional). This tutorial covers Pandas DataFrames, from basic manipulations to advanced operations. Pandas DataFrames make manipulating your data easy, from selecting or replacing columns and indices to reshaping your data.

Objectives

At the end of this lab, you are expected to acquire the following:

a) How To Create a Pandas DataFrame
b) How To Select an Index or Column From a DataFrame
c) How To Add an Index, Row or Column to a DataFrame
d) How To Delete Indices, Rows or Columns From a DataFrame
e) How To Rename the Columns or Indices of a DataFrame
f) How To Format the Data in Your DataFrame
g) How To Create an Empty DataFrame
h) Does Pandas Recognize Dates When Importing Data?
i) When, Why and How You Should Reshape Your DataFrame
j) How To Iterate Over a DataFrame
k) How To Write a DataFrame to a File

Instruction

1. Visit the "10 Minutes to Pandas Tutorial" at the following link: https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html

2. Follow the step-by-step instructions in the tutorial.

Other Related References

- https://www.tutorialspoint.com/python_pandas
- https://pandas.pydata.org/pandas-docs/stable/getting_started/dsintro.html
- https://www.datacamp.com/community/tutorials/pandas-tutorial-dataframe-python
- https://www.dataquest.io/blog/pandas-python-tutorial/
- https://hackernoon.com/intro-to-pandas-1-an-absolute-beginners-guide-to-machine-learning-and-data-science-a1fed3a6f0f3
- https://pandas.pydata.org/
- https://pandas.pydata.org/pandas-docs/stable/

The End