The Learning Hub

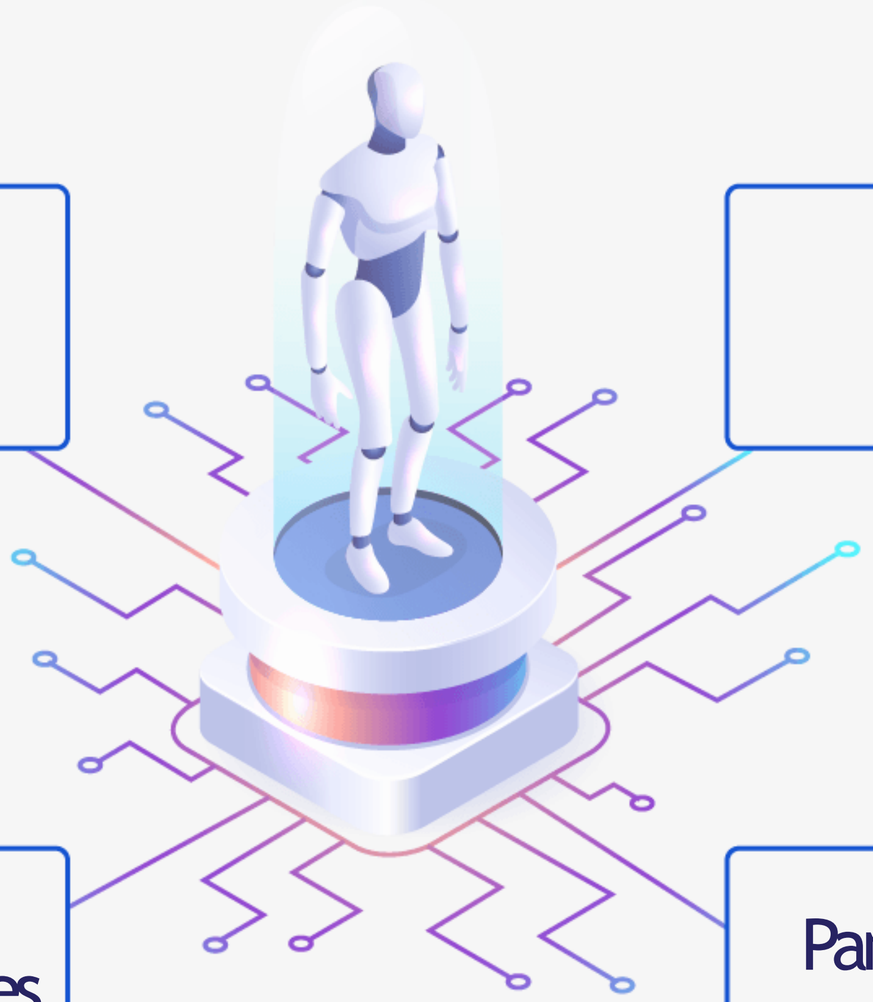
Pandas

Education and Training Solutions 2023

The Learning Hub

Introduction to Pandas

Pandas Series



Pandas Data Frames

Pandas Analyzing Data

The Learning Hub

Introduction to Pandas

What is Pandas?

- Pandas come from Panel Data and Python Data Analysis.
- Pandas is mainly popular as an open-source library that offers various data structures and operations for managing numerical data and time series.
- Pandas contain different functions used for analyzing, exploring, manipulating, and cleaning data.

Why Use Pandas?

- Pandas Library allows us to analyze big data, clean messy data sets, and make them readable and relevant to drive conclusions based on statistical theories.
- you can with pandas discover the correlation between two or more columns, Max and Min value, and delete rows that are not relevant, or contains wrong values, like empty or NULL values.

Installation and Import of Pandas

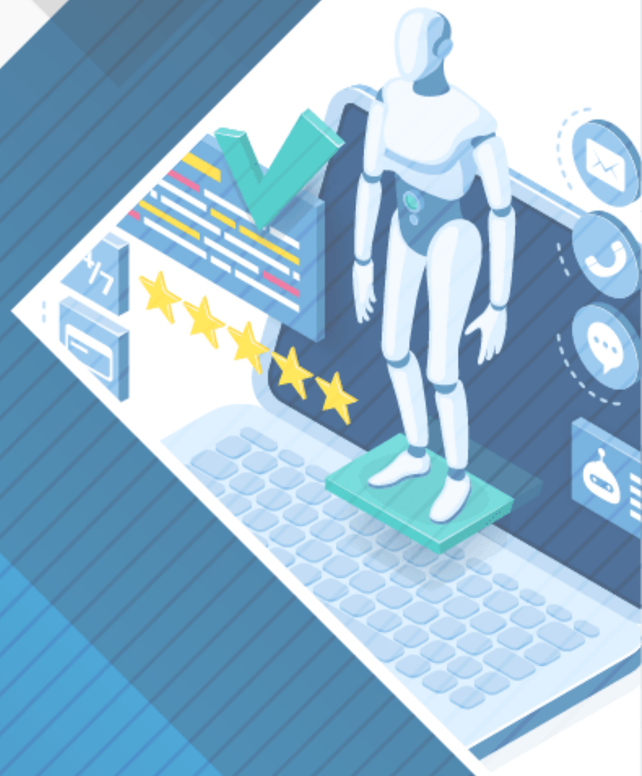
- Install Pandas using this command:

```
pip install pandas
```

- Import Pandas into your applications by adding the import keyword under the pd alias:

```
import pandas as pd
```

The Learning Hub



Pandas Series

Pandas Series

- What is a Pandas Series? A Pandas Series is like a column in a table which is a one-dimensional array holding data of any type.

```
:  1  lis = [5,3,1]
   2  ps = pd.Series(lis)
   3  print(ps)
   4
   5  print("--------------------------")
   6
   7  print(ps[0])
   8
```

```
0    5
1    3
2    1
dtype: int64
--------------------------
5
```

Pandas Series

- Create Labels- you can name your own labels with the index argument, Where you can access an item by referring to the label.

```
1
2  lis = [5,3,1]
3
4  ps = pd.Series(lis, index = ["a", "b", "c"])
5
6  print(ps)
7
8  print(ps["c"])
9
```

```
a    5
b    3
c    1
dtype: int64
1
```

Pandas Series

- Dictionary Series- to create a series from the dictionary, we have to first create a dictionary after that we can make a series

```
1
2  diction = {"id1":123,"id2":456,"id3":789}
3
4  ds = pd.Series(diction)
5
6  print(ds)
7  print("-----------------------------------------------------")
8  ds2 = pd.Series(diction,index = ["id1","id2"])
9
10 print(ds2)
11
12
```

```
id1    123
id2    456
id3    789
dtype: int64
-----------------------------------------------------
id1    123
id2    456
dtype: int64
```

The Learning Hub

Pandas Data Frames

Pandas Data Frames

- Pandas Data Frame will be created by loading the datasets from existing storage, storage can be CSV or Excel files. Pandas Data Frame can be created from the lists, dictionaries, and from lists of dictionaries, etc.
- Pandas Data Frame consists of three principal components, the data, rows, and columns. Data frames can be created in multiple ways.

Pandas Data Frames

- Datasets in Pandas are usually multi-dimensional tables, called Data Frames.
- Series is like a column, Data Frame is the entire table.

```
:  1 data = {"colours":["Red","Green","Blue"],"numbers":[1,2,3] }
   2 df = pd.DataFrame(data)
   3 print(df)
   4
```

```
  colours  numbers
0     Red        1
1   Green        2
2    Blue        3
```

Pandas Data Frames

- Locate a Row (df.loc[ ])- use the loc attribute to return one or more specified row(s)

```
1  data = {"id": [1554,475,13],"length":[4,3,2] }
2  df = pd.DataFrame(data)
3  print(df.loc[0])
4  print("-----------------------------------------------------")
5  print(df.loc[[0,1]])
6
7
```

```
id        1554
length       4
Name: 0, dtype: int64
-----------------------------------------------------
     id  length
0  1554       4
1   475       3
```

Pandas Data Frames

- Data Frames Indexes

```
1  data = {"id": [1554,475,13],"length":[4,3,2] }
2
3  df = pd.DataFrame(data, index =["first","second" ,"third"])
4  print(df)
5
```

```
          id  length
first   1554       4
second   475       3
third     13       2
```

Pandas Data Frames

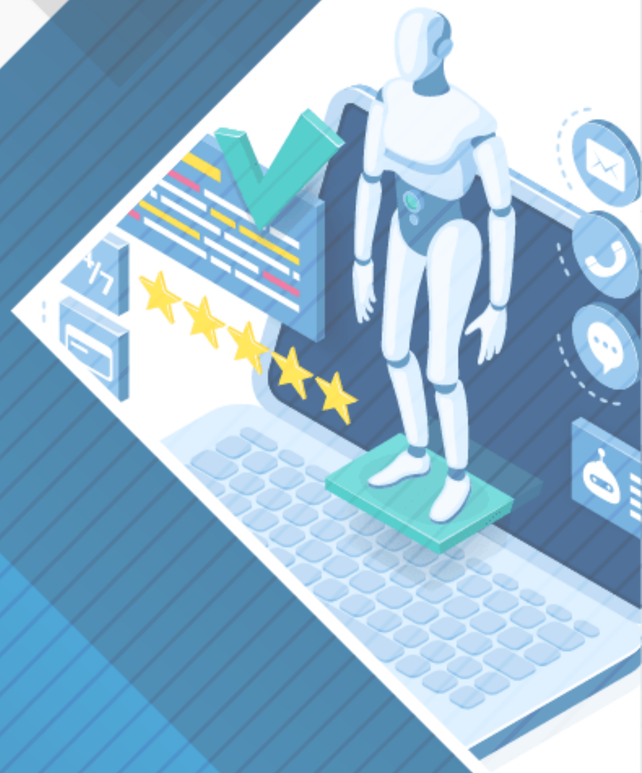
- Locate Named Indexes

```
1  data = {"id": [1554,475,13],"length":[4,3,2] }
2  df = pd.DataFrame(data, index = ["first", "second", "third"])
3  print(df.loc["third"])
4
5
```

```
id        13
length     2
Name: third, dtype: int64
```

The Learning Hub

Pandas Analyzing Data



Load & Save Data Frames

- Load & Save Data Frames

```
1  df = pd.read_csv("data.csv")
2
```

- Save Data Frame Functions

```
1  df.to_csv("df.csv")
2  df.to_html("df.html")
3  df.to_json("df.json")
4
```

Analysing Data Frames

Viewing the Data

- df.head() – returns the first 4 rows in the data frame.
- df.tail() – returns the last 4 rows in the data frame.

Information about Data Frame.

- df.info() – returns information about the data frame Columns.

```
#   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   phone  10 non-null      int64
 1   Names  10 non-null      object
 2   salary 10 non-null      float64
```

Data Frame Functions

fillna() – replace empty (NaN) cells with a value.

mean(), median(), mode() – calculate the respective values for specified column.

describe() – generate a descriptive statistics analysis for the entire data frame.

value_counts() – return a Pandas Series containing the counts of unique values.

- You can check **Pandas Documentation** for more and more of these functions and methods.

References

1. GeeksforGeeks. (2020). Pandas Tutorial. [online] Available at: https://www.geeksforgeeks.org/pandas-tutorial/?ref=lbp [Accessed 30 Sep. 2022].

2. www.w3schools.com. (n.d.). Pandas Tutorial. [online] Available at: https://www.w3schools.com/python/pandas/default.asp.

3. pandas.pydata.org. (n.d.). DataFrame — pandas 1.1.5 documentation. [online] Available at: https://pandas.pydata.org/pandas-docs/stable/reference/frame.html.

THANK YOU