

Data Structures

Python basics

Kunal Khurana

2024-02-23

Table of contents

Learning outcomes-	3
Reading and writing data in text format	3
Binary Data Formats	3
Interacting with web APIs	3
Interacting with Databases	3
Reading and writing data in text format	3
Interacting with Databases	4

Learning outcomes-

Reading and writing data in text format

- Reading text files in pieces
- Writing data to text format
- Working with other delimited formats
- JSON data
- XML and HTML: Web Scraping
- Parsing XML and lxml.objectify

Binary Data Formats

- Reading Microsoft Excel Files
- Using HDF5 Format

Interacting with web APIs

Interacting with Databases

```
import pandas as pd
import numpy as np
```

Reading and writing data in text format

```
df = pd.read_csv('username.csv')

df.head()
```

	Username; Identifier;First name;Last name
0	booker12;9012;Rachel;Booker
1	grey07;2070;Laura;Grey
2	johnson81;4081;Craig;Johnson
3	jenkins46;9346;Mary;Jenkins
4	smith79;5079;Jamie;Smith

```
# without header
df = pd.read_csv('username.csv', header = None)
df.head()
```

0	
0	Username; Identifier;First name;Last name
1	booker12;9012;Rachel;Booker
2	grey07;2070;Laura;Grey
3	johnson81;4081;Craig;Johnson
4	jenkins46;9346;Mary;Jenkins

```
df = pd.read_csv('E:\pythonfordatanalysis\username.csv', names= ['onboarding details'])
```

```
df
```

	onboarding details
0	Username; Identifier;First name;Last name
1	booker12;9012;Rachel;Booker
2	grey07;2070;Laura;Grey
3	johnson81;4081;Craig;Johnson
4	jenkins46;9346;Mary;Jenkins
5	smith79;5079;Jamie;Smith

Interacting with Databases

```
import sqlite3

query = """
Create table test
```

```

(a tatata(29), n Blabla(20),
c Real, d integer);
"""

con = sqlite3.connect('mydata.sqlite')

con.execute(query)

```

<sqlite3.Cursor at 0x27ccb6f6420>

```

con.commit()

# insert a few rows of data

data= [('Atlanta', 'Georgia', 1.25, 6),
        ('Tallahassee', 'Florida', 2.26, 3),
        ('Sacramento', 'California', 1.5,3)]

stmt = 'Insert into test values (?, ?, ?, ?)'

con.executemany(stmt, data)

```

<sqlite3.Cursor at 0x27ccb6f6340>

```

# most SQL drivers return a list of tuples when selecting data from table

cursor = con.execute('SELECT * FROM test')

rows = cursor.fetchall()

rows

```

```

[('Atlanta', 'Georgia', 1.25, 6),
 ('Tallahassee', 'Florida', 2.26, 3),
 ('Sacramento', 'California', 1.5, 3)]

```

```

cursor.description

```

```
((('a', None, None, None, None, None, None),
  ('n', None, None, None, None, None, None),
  ('c', None, None, None, None, None, None),
  ('d', None, None, None, None, None, None))
```

```
pd.DataFrame(rows, columns=[x[0] for x in cursor.description])
```

	a	n	c	d
0	Atlanta	Georgia	1.25	6
1	Tallahassee	Florida	2.26	3
2	Sacramento	California	1.50	3

```
!pip install sqlalchemy
```

```
! pip install gradio typing_extensions
```

```
! pip install jiwer
```

Requirement already satisfied: jiwer in c:\users\khurana_kunal\appdata\local\packages\python...

Requirement already satisfied: click<9.0.0,>=8.1.3 in c:\users\khurana_kunal\appdata\local\p...

Requirement already satisfied: rapidfuzz<4,>=3 in c:\users\khurana_kunal\appdata\local\packag...

Requirement already satisfied: colorama in c:\users\khurana_kunal\appdata\local\packages\pytl...

```
! pip install gradio typing-extensions
```

```
import sqlalchemy as sqla
import pandas as pd
```

```
db = sqla.create_engine('sqlite:///mydata.sqlite')
```

```
pd.read_sql('SELECT * FROM test', db)
```

a	n	c	d
---	---	---	---