List of tuples

Let us see an example of how we can read data from a file into **list of tuples** using Python as programming language.

- When we read data from a file into a `list`, typically each element in the list will be of type binary or string.
- We can convert the element into `tuple` to simplify the processing.
- Once each element is converted to `tuple`, we can access elements in the `tuple` using positional notation.
- Let us see an example to read the data from a file into **list of tuples** and access dates.

```
%%sh

ls -ltr /data/retail_db/orders/part-00000
```

```
-rw-r--r-- 1 root root 2999944 Nov 22 16:08 /data/retail_db/orders/part-00000
```

```
%%sh

tail /data/retail_db/orders/part-00000
```

```
68874,2014-07-03 00:00:00.0,1601,COMPLETE
68875,2014-07-04 00:00:00.0,10637,ON_HOLD
68876,2014-07-06 00:00:00.0,4124,COMPLETE
68877,2014-07-07 00:00:00.0,9692,ON_HOLD
68878,2014-07-08 00:00:00.0,6753,COMPLETE
68879,2014-07-09 00:00:00.0,778,COMPLETE
68880,2014-07-13 00:00:00.0,1117,COMPLETE
68881,2014-07-19 00:00:00.0,2518,PENDING_PAYMENT
68882,2014-07-22 00:00:00.0,10000,ON_HOLD
68883,2014-07-23 00:00:00.0,5533,COMPLETE
```

```
# Reading data from file into a list
path = '/data/retail_db/orders/part-00000'
# C:\\users\\itversity\\Research\\data\\retail_db\\orders\\part-00000
orders_file = open(path)
```

```
type(orders_file)
```

```
_io.TextIOWrapper
```

```
orders_raw = orders_file.read()
```

```
type(orders_raw)
```

```
str
```

```
str.splitlines?
```

```
Docstring:
S.splitlines([keepends]) -> list of strings

Return a list of the lines in S, breaking at line boundaries.
Line breaks are not included in the resulting list unless keepends
is given and true.
Type:      method_descriptor
```

```
orders_raw[:10]
```

```
'1,2013-07-'
```

```
orders = orders_raw.splitlines()
```

```
type(orders)
```

```
list
```

```
orders[:10]
```

```
['1,2013-07-25 00:00:00.0,11599,CLOSED',
 '2,2013-07-25 00:00:00.0,256,PENDING_PAYMENT',
 '3,2013-07-25 00:00:00.0,12111,COMPLETE',
 '4,2013-07-25 00:00:00.0,8827,CLOSED',
 '5,2013-07-25 00:00:00.0,11318,COMPLETE',
 '6,2013-07-25 00:00:00.0,7130,COMPLETE',
 '7,2013-07-25 00:00:00.0,4530,COMPLETE',
 '8,2013-07-25 00:00:00.0,2911,PROCESSING',
 '9,2013-07-25 00:00:00.0,5657,PENDING_PAYMENT',
 '10,2013-07-25 00:00:00.0,5648,PENDING_PAYMENT']
```

```
len(orders) # same as number of records in the file
```

```
68883
```

```
order = '1,2013-07-25 00:00:00.0,11599,CLOSED'
```

```
order
```

```
'1,2013-07-25 00:00:00.0,11599,CLOSED'
```

```
order.split(',')
```

```
['1', '2013-07-25 00:00:00.0', '11599', 'CLOSED']
```

```
tuple(order.split(','))
```

```
('1', '2013-07-25 00:00:00.0', '11599', 'CLOSED')
```

```
(*order.split(','), )# special operator to convert list to tuple
```

```
('1', '2013-07-25 00:00:00.0', '11599', 'CLOSED')
```

```
order_tuples = [(*order.split(','),) for order in orders]
```

```
order_tuples = [tuple(order.split(',')) for order in orders]
```

```
type(order_tuples)
```

```
list
```

```
order_tuples[0]
```

```
('1', '2013-07-25 00:00:00.0', '11599', 'CLOSED')
```

```
order_tuples[:3]
```

```
[('1', '2013-07-25 00:00:00.0', '11599', 'CLOSED'),
 ('2', '2013-07-25 00:00:00.0', '256', 'PENDING_PAYMENT'),
 ('3', '2013-07-25 00:00:00.0', '12111', 'COMPLETE')]
```

```
len(order_tuples)
```

```
68883
```

```
order_dates = [order[1] for order in order_tuples]
```

```
order_dates[:3]
```

```
['2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0', '2013-07-25 00:00:00.0']
```

```
len(order_dates)
```

```
68883
```

```
# We can also change the data types of elements in the tuples
def get_order_details(order):
    order_details = order.split(',')
    return (int(order_details[0]), order_details[1], int(order_details[2]), order_details[3])
```

```
order_tuples = [get_order_details(order) for order in orders]
```

```
order_tuples[:3]
```

```
[(1, '2013-07-25 00:00:00.0', 11599, 'CLOSED'),
 (2, '2013-07-25 00:00:00.0', 256, 'PENDING_PAYMENT'),
 (3, '2013-07-25 00:00:00.0', 12111, 'COMPLETE')]
```

```
order_customer_ids = [order[2] for order in order_tuples]
```

```
order_customer_ids[:3]
```

```
[11599, 256, 12111]
```

```
type(order_customer_ids[0])
```

```
int
```

```
path = '/data/retail_db/orders/part-00000'
# C:\\users\\itversity\\Research\\data\\retail_db\\orders\\part-00000
orders_file = open(path)
orders_raw = orders_file.read()
orders = orders_raw.splitlines()
order_tuples = [(*order.split(','),) for order in orders]
order_dates = [order[1] for order in order_tuples]
```

```
unique_dates = set(order_dates)
```

```
len(unique_dates)
```

```
364
```