

Row level transformations using map

☰ Contents

[Task 1](#)

[Print to PDF](#) ►

[Task 2](#)

Let us understand how we can perform row level transformations using `map`. Here are some of the examples.

- Derive new fields from existing fields.
- Get last 4 digits of social security number.
- Standardize phone numbers.
- Convert names to lower or upper case.
- Break down the address into street, city, state, zip code.
- Encrypt confidential information such as social security number or other unique ids such as Aadhaar.

```
%run 02_preparing_data_sets.ipynb
```

```
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)
```

```
68883
```

```
order_items[:10]
```

```
[ '1,1,957,1,299.98,299.98',  
  '2,2,1073,1,199.99,199.99',  
  '3,2,502,5,250.0,50.0',  
  '4,2,403,1,129.99,129.99',  
  '5,4,897,2,49.98,24.99',  
  '6,4,365,5,299.95,59.99',  
  '7,4,502,3,150.0,50.0',  
  '8,4,1014,4,199.92,49.98',  
  '9,5,957,1,299.98,299.98',  
  '10,5,365,5,299.95,59.99' ]
```

```
len(order_items)
```

```
172198
```

Task 1

Get day name of each date in our orders data set. Output should be tuple with 3 elements.

- `order_id` of type integer
- `order_date` of type string
- `order_day_name` of type string

```
import datetime as dt  
d = dt.datetime.strptime('2013-07-25 00:00:00.0'.split(' ')[0], '%Y-%m-%d')
```

```
d
```

```
datetime.datetime(2013, 7, 25, 0, 0)
```

```
d.weekday()
```

```
3
```

```
import calendar
```

```
list(calendar.day_name)
```

```
['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday']
```

```
calendar.day_name[d.weekday()]
```

```
'Thursday'
```

```
import datetime as dt, calendar as c
```

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

```
order.split(',')[1].split(' ')[0]
```

```
'2013-07-25'
```

```
dt.datetime.strptime(order.split(',')[1].split(' ')[0], '%Y-%m-%d')
```

```
datetime.datetime(2013, 7, 25, 0, 0)
```

```
dt.datetime.strptime(order.split(',')[1].split(' ')[0], '%Y-%m-%d').weekday()
```

```
3
```

```
c.day_name[dt.datetime.strptime(order.split(',')[1].split(' ')[0], '%Y-%m-%d').weekday()]
```

```
'Thursday'
```

```
import datetime as dt, calendar as c
```

```
order_dates = map(  
    lambda order: c.day_name[dt.datetime.strptime(order.split(',')[1].split(' ')[0], '%Y-%m-%d').weekday()],  
    orders  
)
```

```
list(order_dates)[:10]
```

```
['Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday',  
'Thursday']
```

Note

We can use lambda function as long as we do not have assignment operations. However, we might end up compromising readability.

```
import datetime as dt, calendar as c
order_dates = map(
    lambda order: (
        int(order.split(',')[0]),
        order.split(',')[1],
        c.day_name[dt.datetime.strptime(order.split(',')[1].split(' ')[0], '%Y-%m-%d').weekday()]
    ),
    orders
)
```

```
list(order_dates)[:10]
```

```
[(1, '2013-07-25 00:00:00.0', 'Thursday'),
 (2, '2013-07-25 00:00:00.0', 'Thursday'),
 (3, '2013-07-25 00:00:00.0', 'Thursday'),
 (4, '2013-07-25 00:00:00.0', 'Thursday'),
 (5, '2013-07-25 00:00:00.0', 'Thursday'),
 (6, '2013-07-25 00:00:00.0', 'Thursday'),
 (7, '2013-07-25 00:00:00.0', 'Thursday'),
 (8, '2013-07-25 00:00:00.0', 'Thursday'),
 (9, '2013-07-25 00:00:00.0', 'Thursday'),
 (10, '2013-07-25 00:00:00.0', 'Thursday')]
```

Note

Here is the example of implementation using named function.

```
def get_order_date(order):
    order_details = order.split(',')
    order_id = int(order_details[0])
    order_date = order.split(',')[1]
    order_date_as_datetime = dt.datetime.strptime(order_date.split(' ')[0], '%Y-%m-%d')
    order_day_name = c.day_name[order_date_as_datetime.weekday()]
    return (order_id, order_date, order_day_name)
```

```
order_dates = map(
    get_order_date,
    orders
)
```

```
list(order_dates)[:10]
```

```
[(1, '2013-07-25 00:00:00.0', 'Thursday'),
 (2, '2013-07-25 00:00:00.0', 'Thursday'),
 (3, '2013-07-25 00:00:00.0', 'Thursday'),
 (4, '2013-07-25 00:00:00.0', 'Thursday'),
 (5, '2013-07-25 00:00:00.0', 'Thursday'),
 (6, '2013-07-25 00:00:00.0', 'Thursday'),
 (7, '2013-07-25 00:00:00.0', 'Thursday'),
 (8, '2013-07-25 00:00:00.0', 'Thursday'),
 (9, '2013-07-25 00:00:00.0', 'Thursday'),
 (10, '2013-07-25 00:00:00.0', 'Thursday')]
```

```
order_dates = map(
    lambda order: get_order_date(order),
    orders
)
```

```
list(order_dates)[:10]
```

```
[(1, '2013-07-25 00:00:00.0', 'Thursday'),
 (2, '2013-07-25 00:00:00.0', 'Thursday'),
 (3, '2013-07-25 00:00:00.0', 'Thursday'),
 (4, '2013-07-25 00:00:00.0', 'Thursday'),
 (5, '2013-07-25 00:00:00.0', 'Thursday'),
 (6, '2013-07-25 00:00:00.0', 'Thursday'),
 (7, '2013-07-25 00:00:00.0', 'Thursday'),
 (8, '2013-07-25 00:00:00.0', 'Thursday'),
 (9, '2013-07-25 00:00:00.0', 'Thursday'),
 (10, '2013-07-25 00:00:00.0', 'Thursday')]
```

Task 2

Add weekend flag for Saturday and Sunday dates.

```
def get_order_date(order):
    order_details = order.split(',')
    order_id = int(order_details[0])
    order_date = order.split(',')[1]
    order_date_as_datetime = dt.datetime.strptime(order_date.split(' ')[0], '%Y-%m-%d')
    order_day_name = c.day_name[order_date_as_datetime.weekday()]
    weekend_flag = True if order_date_as_datetime.weekday() in (5, 6) else False
    return (order_id, order_date, order_day_name, weekend_flag)
```

```
order_dates = map(
    get_order_date,
    orders
)
```

```
list(order_dates)[:10]
```

```
[(1, '2013-07-25 00:00:00.0', 'Thursday', False),
 (2, '2013-07-25 00:00:00.0', 'Thursday', False),
 (3, '2013-07-25 00:00:00.0', 'Thursday', False),
 (4, '2013-07-25 00:00:00.0', 'Thursday', False),
 (5, '2013-07-25 00:00:00.0', 'Thursday', False),
 (6, '2013-07-25 00:00:00.0', 'Thursday', False),
 (7, '2013-07-25 00:00:00.0', 'Thursday', False),
 (8, '2013-07-25 00:00:00.0', 'Thursday', False),
 (9, '2013-07-25 00:00:00.0', 'Thursday', False),
 (10, '2013-07-25 00:00:00.0', 'Thursday', False)]
```

Note

Validate whether weekend_flag is generated properly or not.

```
def get_order_date(order):
    order_details = order.split(',')
    order_id = int(order_details[0])
    order_date = order.split(',')[1]
    order_date_as_datetime = dt.datetime.strptime(order_date.split(' ')[0], '%Y-%m-%d')
    order_day_name = c.day_name[order_date_as_datetime.weekday()]
    weekend_flag = True if order_date_as_datetime.weekday() in (5, 6) else False
    return (order_id, order_date, order_day_name, weekend_flag)
```

```
order_dates = map(
    get_order_date,
    orders
)
```

```
order_dates_without_ids = map(
    lambda order: (order[1], order[2], order[3]),
    order_dates
)
```

```
set(order_dates_without_ids)
```

[illegible]

[illegible]

[illegible]

[illegible]


```
('2014-07-16 00:00:00.0', 'Wednesday', False),  
( '2014-07-17 00:00:00.0', 'Thursday', False),  
( '2014-07-18 00:00:00.0', 'Friday', False),  
( '2014-07-19 00:00:00.0', 'Saturday', True),  
( '2014-07-20 00:00:00.0', 'Sunday', True),  
( '2014-07-21 00:00:00.0', 'Monday', False),  
( '2014-07-22 00:00:00.0', 'Tuesday', False),  
( '2014-07-23 00:00:00.0', 'Wednesday', False),  
( '2014-07-24 00:00:00.0', 'Thursday', False)}
```