

08_시계열 데이터

- 시계열 데이터

- > 시계열 데이터란?

- 일정 시간 간격으로 저장된 데이터

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 122 entries, 0 to 121
```

```
Data columns (total 18 columns):
```

#	Column	Non-Null Count	Dtype
0	Date	122 non-null	datetime64[ns]

Date	
0	2015-01-05
1	2015-01-04
2	2015-01-03
3	2015-01-02
4	2014-12-31

- 시계열 데이터

- > 시계열 데이터 불러오기

- 데이터를 불러오면서 변환

```
import pandas as pd
```

```
ebola = pd.read_csv('country_timeseries.csv', parse_dates=['Date'])
```

- to_datetime을 이용한 변환

```
pd.to_datetime(ebola['Date'])
```

- astype('datetime64[ns]')을 이용한 변환

```
ebola['Date'].astype('datetime64[ns]')
```

• 시계열 데이터

- > 시간 정보 추출하기
 - dt 접근자

```
print(ebola['Date'].dt.year)
```

```
Out : 0      2015
      1      2015
      2      2015
      3      2015
      4      2014
      ...
     117     2014
     118     2014
     119     2014
     120     2014
     121     2014
      Name: Date, Length: 122, dtype: int32
```

```
pandas.Series.dt.date
pandas.Series.dt.time
pandas.Series.dt.timetz
pandas.Series.dt.year
pandas.Series.dt.month
pandas.Series.dt.day
pandas.Series.dt.hour
pandas.Series.dt.minute
pandas.Series.dt.second
pandas.Series.dt.microsecond
pandas.Series.dt.nanosecond
pandas.Series.dt.dayofweek
pandas.Series.dt.day_of_week
pandas.Series.dt.weekday
pandas.Series.dt.dayofyear
pandas.Series.dt.day_of_year
pandas.Series.dt.days_in_month
pandas.Series.dt.quarter
pandas.Series.dt.is_month_start
pandas.Series.dt.is_month_end
pandas.Series.dt.is_quarter_start
pandas.Series.dt.is_quarter_end
pandas.Series.dt.is_year_start
pandas.Series.dt.is_year_end
```

```
pandas.Series.dt.is_leap_year
pandas.Series.dt.daysinmonth
pandas.Series.dt.days_in_month
pandas.Series.dt.tz
pandas.Series.dt.freq
pandas.Series.dt.unit
pandas.Series.dt.normalize
pandas.Series.dt.isocalendar
pandas.Series.dt.to_period
pandas.Series.dt.to_pydatetime
pandas.Series.dt.tz_localize
pandas.Series.dt.tz_convert
pandas.Series.dt.normalize
pandas.Series.dt.strftime
pandas.Series.dt.round
pandas.Series.dt.floor
pandas.Series.dt.ceil
pandas.Series.dt.month_name
pandas.Series.dt.day_name
pandas.Series.dt.as_unit
pandas.Series.dt.qyear
pandas.Series.dt.start_time
pandas.Series.dt.end_time
pandas.Series.dt.days
pandas.Series.dt.seconds
pandas.Series.dt.microseconds
pandas.Series.dt.nanoseconds
pandas.Series.dt.components
pandas.Series.dt.unit
```

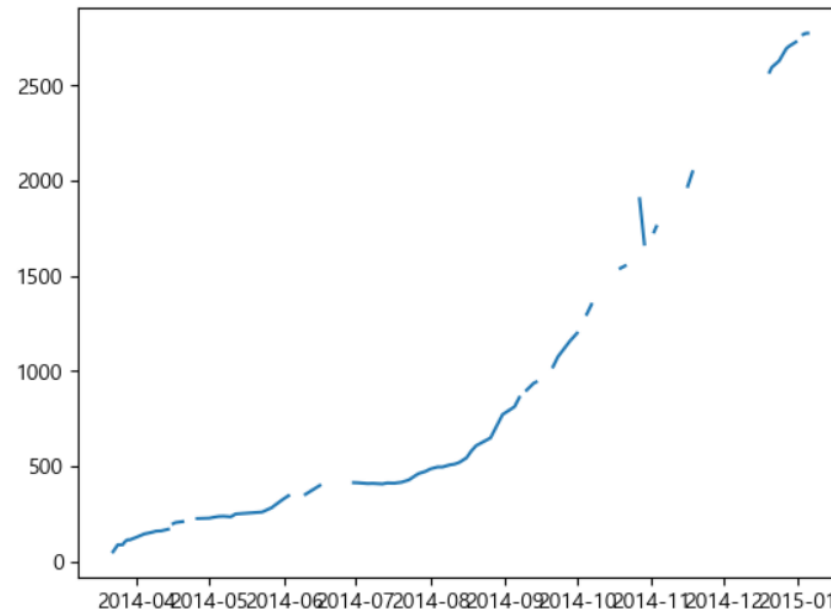
- 시계열 데이터

> 시계열 그래프 그리기

```
import matplotlib.pyplot as plt

plt.plot(ebola['Date'], ebola['Cases_Guinea'])
```

Out :



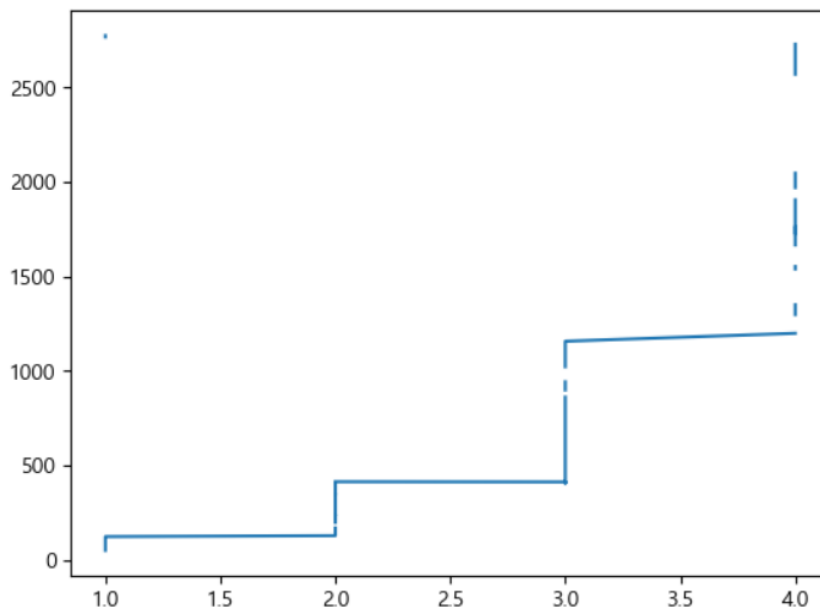
- 시계열 데이터

- > 시계열 그래프 그리기
 - 분기별 그래프 그리기

```
ebola['Date_quarter'] = ebola['Date'].dt.quarter
```

```
plt.plot(ebola['Date_quarter'], ebola['Cases_Guinea'])
```

Out :



- 시계열 데이터

- > 시계열 그래프 그리기
 - 분기별 그래프 그리기

```
ebola['Date_year'] = ebola['Date'].dt.year  
ebola_quarter = ebola.groupby(['Date_year', 'Date_quarter'])['Cases_Guinea'].mean()  
ebola_quarter
```

Out :

Date_year	Date_quarter	
2014	1	94.500000
	2	252.185185
	3	636.633333
	4	1989.800000
2015	1	2773.333333

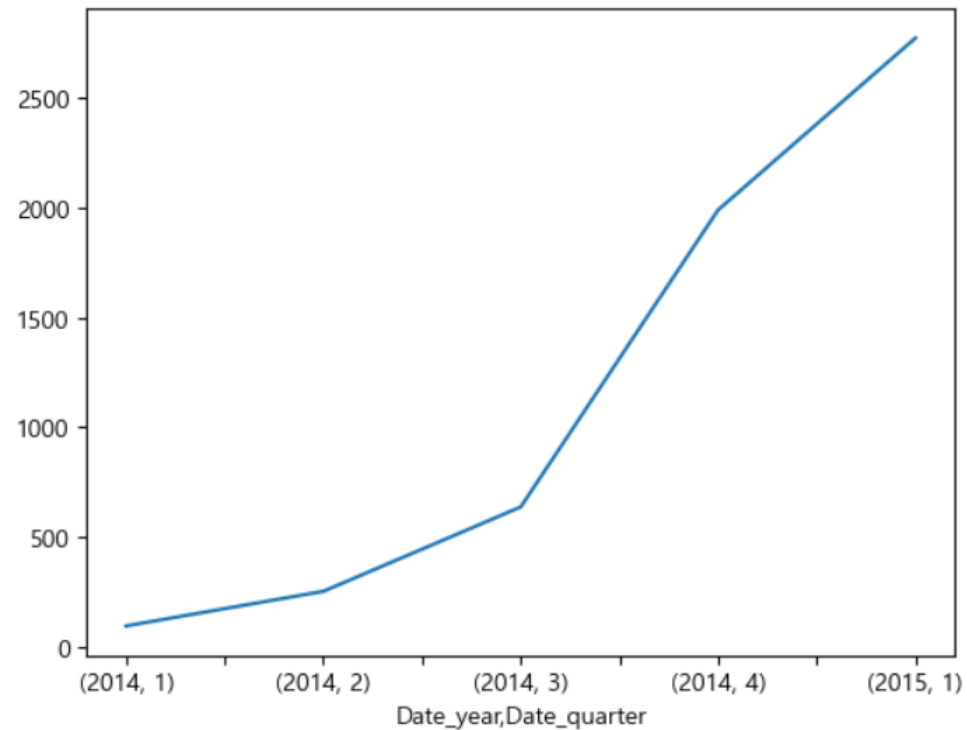
Name: Cases_Guinea, dtype: float64

- 시계열 데이터

- > 시계열 그래프 그리기
 - 분기별 그래프 그리기

```
ebola_quarter.plot()
```

Out :



- 시계열 데이터

- > 시간 범위 다루기
 - pd.date_range

```
pd.date_range('2014-12-31', '2015-01-05', freq = 'D')
```

Out : DatetimeIndex(['2014-12-31', '2015-01-01', '2015-01-02', '2015-01-03',
 '2015-01-04', '2015-01-05'],
 dtype='datetime64[ns]', freq='D')

- 시계열 데이터

- > 시간 범위 다루기

- https://pandas.pydata.org/docs/user_guide/timeseries.html#offset-aliases

지정자	설명	지정자	설명
B	평일	QS	분기의 시작일
C	사용자가 정의한 평일	BQS	휴일을 제외한 QS
D	일자 단위	A	연 마지막 날
W	주 단위	BA	휴일을 제외한 A
M	월 마지막 날	AS	연 시작일
SM	15일과 월 마지막 날	BAS	유일을 제외한 AS
BM	휴일을 제외한 M	BH	업무 시간 단위(9~16시)
CBM	BM에 사용자 정의	H	시간 단위
MS	월 시작일	T	분 단위
SMS	월 시작일과 15일	S	초 단위
BMS	휴일을 제외한 MS	L	밀리초 단위
CBMS	BMS에 사용자 정의	U	마이크로초 단위
Q	분기의 마지막 날	N	나노초 단위
BQ	휴일을 제외한 Q		

- 시계열 데이터

- > 시간 주기 변경하기
 - resample

```
ebola_month = ebola.set_index('Date').resample('M').mean()  
ebola_month
```

Out :

	Day	Cases_Guinea	Cases_Liberia	Cases_SierraLeone	Cases_Nigeria	Cases_Senegal	Cases_UnitedStates
Date							
2014-03-31	4.500000	94.500000	6.500000	3.333333	NaN	NaN	NaN
2014-04-30	24.333333	177.818182	24.555556	2.200000	NaN	NaN	NaN
2014-05-31	51.888889	248.777778	12.555556	7.333333	NaN	NaN	NaN
2014-06-30	84.636364	373.428571	35.500000	125.571429	NaN	NaN	NaN
2014-07-31	115.700000	423.000000	212.300000	420.500000	1.333333	NaN	NaN
2014-08-31	145.090909	559.818182	868.818182	844.000000	13.363636	1.000000	NaN
2014-09-30	177.500000	967.888889	2815.625000	1726.000000	20.714286	1.285714	NaN
2014-10-31	207.470588	1500.444444	4758.750000	3668.111111	20.000000	1.000000	2.555556
2014-11-30	237.214286	1950.500000	7039.000000	5843.625000	20.000000	1.000000	4.000000
2014-12-31	271.181818	2579.625000	7902.571429	8985.875000	20.000000	1.000000	4.000000
2015-01-31	287.500000	2773.333333	8161.500000	9844.000000	NaN	NaN	NaN

- 시계열 데이터

- > 시간 주기 변경하기
 - resample

```
plt.plot(ebola_month.index, ebola_month['Cases_Guinea'])
```

Out :

