

In [1]:



```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.rc('figure',figsize=(10,6))
```

In [2]:



!dir

C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호: E262-B56F

C:\Users\Wdonghyunkim 디렉터리

```

2020-05-07 오후 02:11 <DIR> .
2020-05-07 오후 02:11 <DIR> ..
2020-05-05 오후 10:18 <DIR> .android
2020-03-20 오후 07:08 <DIR> .AndroidStudio3.6
2020-04-20 오후 05:07 2,184 .bash_history
2019-06-27 오후 12:49 <DIR> .eclipse
2019-06-24 오후 01:43 16 .emulator_console_auth_token
2020-04-20 오후 04:58 177 .gitconfig
2019-10-01 오후 09:24 <DIR> .gradle
2019-10-04 오후 10:24 <DIR> .IdeaIC2019.2
2019-12-10 오전 11:01 <DIR> .idlerc
2020-05-07 오후 02:11 <DIR> .ipynb_checkpoints
2019-11-19 오후 04:15 <DIR> .ipython
2019-11-19 오후 04:21 <DIR> .jupyter
2019-09-05 오후 08:11 <DIR> .m2
2020-04-23 오후 02:58 <DIR> .matplotlib
2019-12-17 오후 09:50 <DIR> .p2
2019-08-02 오후 05:05 <DIR> .PyCharmCE2019.2
2019-09-27 오후 09:31 <DIR> .pylint.d
2019-06-27 오후 12:49 <DIR> .tooling
2020-02-05 오후 11:27 <DIR> .VirtualBox
2019-11-09 오후 04:42 <DIR> .vscode
2020-04-15 오후 10:00 <DIR> 3D Objects
2019-11-24 오후 04:13 <DIR> AlcoholTemp
2020-04-16 오후 03:19 <DIR> anaconda3
2019-10-01 오후 10:17 <DIR> AndroidStudioProjects
2020-02-21 오후 09:28 <DIR> BackUpAndroidBoostCourse
2019-12-04 오후 02:38 <DIR> connector_mysql
2020-04-15 오후 10:00 <DIR> Contacts
2020-04-10 오후 04:23 <DIR> Creative Cloud Files
2019-08-30 오후 07:39 <DIR> data
2020-05-05 오후 10:46 <DIR> Desktop
2020-04-20 오후 04:59 <DIR> Documents
2019-07-02 오후 02:46 <DIR> DoltAndroidRev5-master
2020-05-07 오전 11:34 <DIR> Downloads
2019-10-15 오후 11:54 <DIR> eclipse-workspace
2020-04-15 오후 10:00 <DIR> Favorites
2020-04-27 오후 11:06 <DIR> GoUnity
2019-04-25 오전 09:49 632 image.svg
2019-11-08 오후 10:53 502 image.zip
2019-12-03 오후 04:24 <DIR> jobphoto
2019-11-29 오후 02:00 <DIR> kakaosdk
2020-04-15 오후 10:00 <DIR> Links
2020-01-12 오후 07:43 <DIR> MaliProject
2020-03-05 오후 08:19 <DIR> MovieProject2
2020-04-15 오후 10:00 <DIR> Music
2019-11-24 오후 04:05 1,461,054 MyAlcohol.zip
2020-02-10 오후 01:09 <DIR> MyDrawer
2020-02-10 오후 09:43 <DIR> MyDrawerExample
2020-04-29 오후 04:22 <DIR> New Unity Project
2020-04-15 오후 09:01 <DIR> OneDrive

```

```

2019-11-29 오후 12:47 <DIR> OpenSSL
2020-04-15 오후 10:00 <DIR> Pictures
2019-10-06 오후 02:00 <DIR> PycharmProjects
2019-11-19 오후 04:20 23,879 recommendation_system.ipynb
2019-10-09 오후 12:50 9,501 robot.jpg
2020-04-15 오후 10:00 <DIR> Saved Games
2020-04-13 오전 10:14 <DIR> SDDH
2020-04-10 오후 07:19 <DIR> sddhs
2020-04-15 오후 10:00 <DIR> Searches
2019-06-22 오전 09:27 <DIR> source
2019-12-03 오후 01:31 6,012,107 sqlite-jdbc-3.27.2.1.jar
2019-11-29 오후 12:47 <DIR> ssl
2020-02-08 오후 02:21 <DIR> TempBoostCourse
2020-04-21 오후 11:02 <DIR> ToDoList
2019-11-19 오후 04:17 555 Untitled.ipynb
2020-04-16 오후 03:34 7,954 Untitled1.ipynb
2020-04-23 오후 03:23 1,042,052 Untitled2.ipynb
2020-04-30 오전 10:46 22,273 Untitled3.ipynb
2020-04-30 오전 11:03 807,592 Untitled4.ipynb
2020-04-30 오전 11:02 108,342 Untitled5.ipynb
2020-04-30 오전 10:43 95,402 Untitled6.ipynb
2020-05-07 오후 02:11 72 Untitled7.ipynb
2020-04-15 오후 10:00 <DIR> Videos
2020-04-10 오후 05:09 <DIR> VirtualBox VMs
2020-04-12 오후 06:21 <DIR> VisualStudio2017
2019-08-21 오후 01:44 192 _netrc
2019-11-24 오후 04:14 <DIR> 안드로이드(Alcohol)프로젝트 보관 파일
18개 파일 9,594,486 바이트
60개 디렉터리 53,440,634,880 바이트 남음

```

In [3]:



```
!type ex1.csv
```

```

a,b,c,d,message
1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo

```

In [4]:



```
df=pd.read_csv('ex1.csv')
```

In [5]:



```
df
```

Out[5]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [6]:

```
df=df.set_index('message')
df
```

Out[6]:

	a	b	c	d
message				
hello	1	2	3	4
world	5	6	7	8
foo	9	10	11	12

In [7]:

```
#read table(구분자를 가지고 그대로 표현)
dft=pd.read_table('ex1.csv')
dft
```

Out[7]:

	a,b,c,d,message
0	1,2,3,4,hello
1	5,6,7,8,world
2	9,10,11,12,foo

In [8]:

```
dft=pd.read_table('ex1.csv',sep=',')#sep:구분자
dft
```

Out[8]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [9]:

```
!type ex2.csv
```

1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo

In [10]:

```
dfc=pd.read_csv('ex2.csv')
dfc
```

Out[10]:

	1	2	3	4	hello
0	5	6	7	8	world
1	9	10	11	12	foo

In [12]:

```
#헤더 없애기
dfc=pd.read_csv('ex2.csv',header=None)
dfc
```

Out[12]:

	0	1	2	3	4
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [13]:

```
#헤더 만들기
dfc=pd.read_csv('ex2.csv',names=['a','b','c','d','message'])
dfc
```

Out[13]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [14]:



```
names=['a','b','c','d','message']
pd.read_csv('ex2.csv',names=names,index_col='message')
```

Out[14]:

	a	b	c	d
message				
hello	1	2	3	4
world	5	6	7	8
foo	9	10	11	12

In [15]:



```
!type ex4.csv
```

```
# hey!
a,b,c,d,message
# just wanted to make things more difficult for you
# who reads CSV files with computers, anyway?
1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo
```

In [16]:



```
pd.read_csv('ex4.csv',skiprows=[0,2,3])#0,2,3번째 skip
```

Out[16]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [17]:



```
!type ex5.csv
```

```
something,a,b,c,d,message
one,1,2,3,4,NA
two,5,6,,8,world
three,9,10,11,12,foo
```

In [18]:



```
result=pd.read_csv('ex5.csv')
result
```

Out[18]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

In [19]:



```
pd.isnull(result)
```

Out[19]:

	something	a	b	c	d	message
0	False	False	False	False	False	True
1	False	False	False	True	False	False
2	False	False	False	False	False	False

In [21]:



```
result=pd.read_csv('ex5.csv',na_values=['NULL','world'])
#na_values는 NaN으로 만들고 싶은거
result
```

Out[21]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	NaN
2	three	9	10	11.0	12	foo

In [22]:

```
sentinels={'message':['world','NA'],'something':['two']}  
pd.read_csv('ex5.csv',na_values=sentinels)
```

Out[22]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	NaN	5	6	NaN	8	NaN
2	three	9	10	11.0	12	foo

In [23]:

```
## writing Data to Text Format
```

In [24]:

```
result
```

Out[24]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	NaN
2	three	9	10	11.0	12	foo

In [25]:

```
result.to_csv('out.csv')
```

In [26]:

```
!type out.csv
```

```
,something,a,b,c,d,message  
0,one,1,2,3.0,4,  
1,two,5,6,,8,  
2,three,9,10,11.0,12,foo
```

In [27]:

```
result.to_csv('out.csv',na_rep='NULL')  
!type out.csv
```

```
,something,a,b,c,d,message  
0,one,1,2,3.0,4,NULL  
1,two,5,6,NULL,8,NULL  
2,three,9,10,11.0,12,foo
```


In [28]:



```
result.to_csv('out.csv', index=False, header=False)
!type out.csv
```

```
one,1,2,3.0,4,
two,5,6,,8,
three,9,10,11.0,12,foo
```

In [29]:



```
result.to_csv('out.csv', sep='&')
!type out.csv
```

```
&something&a&b&c&d&message
0&one&1&2&3.0&4&
1&two&5&6&&8&
2&three&9&10&11.0&12&foo
```

In [31]:



```
#JSON Data
```

In [36]:



```
obj="""
{"name": "Wes",
 "places_lived": ["United States", "Spain", "Germany"],
 "pet": null,
 "siblings": [{"name": "Scott", "age": 30, "pets": ["Zeus", "Zuko"]},
 {"name": "Katie", "age": 38,
 "pets": ["Sixes", "Stache", "Cisco"]}]
}
"""
obj
```

Out[36]:

```
'Wn{"name": "Wes", Wn"places_lived": ["United States", "Spain", "Germany"], Wn"pet": nul
l, Wn"siblings": [{"name": "Scott", "age": 30, "pets": ["Zeus", "Zuko"]}, Wn{"name": "Katie",
"age": 38, Wn"pets": ["Sixes", "Stache", "Cisco"]}] Wn} Wn'
```

In [39]:



```
import json
result=json.loads(obj)
result
```

Out[39]:

```
{'name': 'Wes',
 'places_lived': ['United States', 'Spain', 'Germany'],
 'pet': None,
 'siblings': [{'name': 'Scott', 'age': 30, 'pets': ['Zeus', 'Zuko']},
 {'name': 'Katie', 'age': 38, 'pets': ['Sixes', 'Stache', 'Cisco']}]}
```

In [40]:



```
dd=pd.DataFrame(result['places_lived'],columns=['place'])  
dd
```

Out[40]:

	place
0	United States
1	Spain
2	Germany

In [41]:



```
siblings=pd.DataFrame(result['siblings'],columns=['name','age','pets'])  
siblings
```

Out[41]:

	name	age	pets
0	Scott	30	[Zeus, Zuko]
1	Katie	38	[Sixes, Stache, Cisco]

In [42]:



```
#Getting Data Using API
```

In [43]:



```
pip install pandas_datareader
```

Collecting pandas_datareader

Downloading pandas_datareader-0.8.1-py2.py3-none-any.whl (107 kB)

Requirement already satisfied: lxml in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas_datareader) (4.5.0)

Requirement already satisfied: requests>=2.3.0 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas_datareader) (2.22.0)

Requirement already satisfied: pandas>=0.21 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas_datareader) (1.0.1)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from requests>=2.3.0->pandas_datareader) (1.25.8)

Requirement already satisfied: idna<2.9,>=2.5 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from requests>=2.3.0->pandas_datareader) (2.8)

Requirement already satisfied: certifi>=2017.4.17 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from requests>=2.3.0->pandas_datareader) (2019.11.28)

Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from requests>=2.3.0->pandas_datareader) (3.0.4)

Requirement already satisfied: python-dateutil>=2.6.1 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas>=0.21->pandas_datareader) (2.8.1)

Requirement already satisfied: numpy>=1.13.3 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas>=0.21->pandas_datareader) (1.18.1)

Requirement already satisfied: pytz>=2017.2 in c:\Users\Wdonghyunkim\Anaconda3\lib\site-packages (from pandas>=0.21->pandas_datareader) (2019.3)

Requirement already satisfied: six>=1.5 in c:\Users\Wdonghyunkim\AppData\Local\Programs\Python\Python37\site-packages (from python-dateutil>=2.6.1->pandas>=0.21->pandas_datareader) (1.12.0)

Installing collected packages: pandas-datareader

Successfully installed pandas-datareader-0.8.1

Note: you may need to restart the kernel to use updated packages.

In [44]:



```
import pandas_datareader as pdr
```

C:\Users\Wdonghyunkim\Anaconda3\lib\site-packages\pandas_datareader\compat__init__.py:7: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead.

```
from pandas.util.testing import assert_frame_equal
```

In [45]:



```
df=pdr.get_data_yahoo('005930.KS')#삼성전자 주식 from 야후
df
```

Out[45]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2015-05-11	27200.0	26720.0	27200.0	26720.0	9640450.0	259.122040
2015-05-12	27060.0	26480.0	26900.0	26620.0	10633200.0	258.152252
2015-05-13	26980.0	26520.0	26980.0	26660.0	9877250.0	258.540161
2015-05-14	26860.0	26560.0	26820.0	26800.0	8488500.0	259.897827
2015-05-15	27120.0	26420.0	27100.0	26520.0	8922350.0	257.182495
...
2020-03-09	56500.0	56500.0	56500.0	56500.0	0.0	56500.000000
2020-03-10	54900.0	53700.0	53800.0	54600.0	32106554.0	54600.000000
2020-03-11	54400.0	52000.0	54300.0	52100.0	45707281.0	52100.000000
2020-03-12	52100.0	52100.0	52100.0	52100.0	0.0	52100.000000
2020-05-07	49300.0	48700.0	49200.0	48950.0	10862339.0	48950.000000

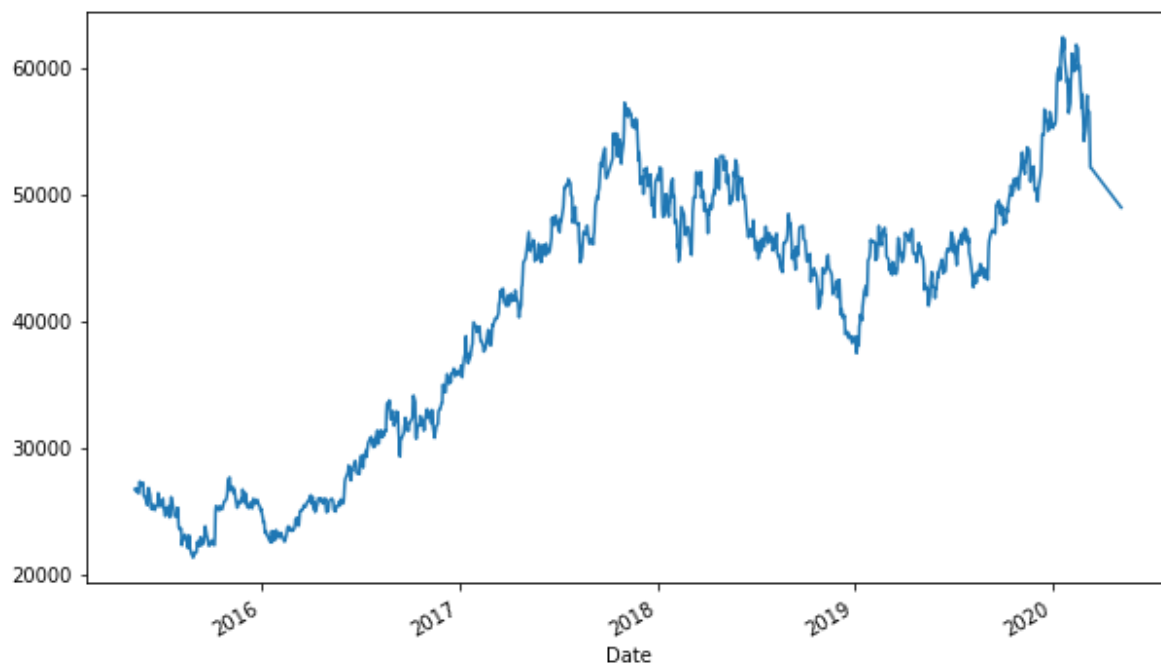
1185 rows × 6 columns

In [46]:

```
df['Close'].plot()
```

Out[46]:

<matplotlib.axes._subplots.AxesSubplot at 0x142d654d248>



In [49]:

```
df=pdr.get_data_yahoo('035760.KQ')#CJ_ENM  
df
```

Out[49]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2020-05-07	124200	121800	123200	122600	57795	122600

In [50]:



```
df=pdr.get_data_yahoo('035420.KS')#Naver
df
```

Out [50]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2015-05-11	122800.0	120600.0	122000.0	121200.0	424635.0	117981.023438
2015-05-12	122200.0	119600.0	120000.0	121000.0	447655.0	117786.343750
2015-05-13	121200.0	120000.0	120800.0	120800.0	404655.0	117591.648438
2015-05-14	122000.0	120200.0	120200.0	121400.0	307855.0	118175.718750
2015-05-15	122400.0	121200.0	122000.0	122000.0	311575.0	118759.773438
...
2020-03-09	179500.0	179500.0	179500.0	179500.0	0.0	179500.000000
2020-03-10	176000.0	165000.0	168000.0	172000.0	666532.0	172000.000000
2020-03-11	174000.0	167500.0	172000.0	170000.0	483822.0	170000.000000
2020-03-12	170000.0	170000.0	170000.0	170000.0	0.0	170000.000000
2020-05-07	220000.0	210500.0	211500.0	215500.0	765242.0	215500.000000

1185 rows × 6 columns

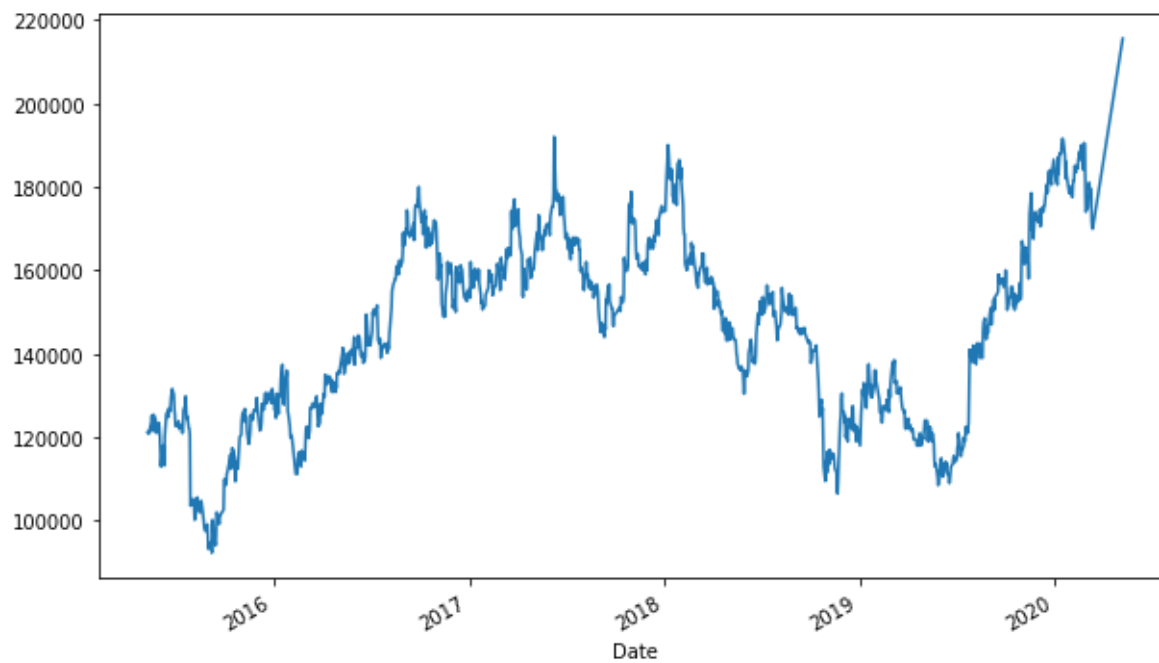
In [51]:

```
df['Close'].plot()
```



Out[51]:

<matplotlib.axes._subplots.AxesSubplot at 0x142dc41e688>



In [52]:



!dir

C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호: E262-B56F

C:\Users\Wdonghyunkim 디렉터리

```

2020-05-07 오후 03:07 <DIR> .
2020-05-07 오후 03:07 <DIR> ..
2020-05-05 오후 10:18 <DIR> .android
2020-03-20 오후 07:08 <DIR> .AndroidStudio3.6
2020-04-20 오후 05:07 2,184 .bash_history
2019-06-27 오후 12:49 <DIR> .eclipse
2019-06-24 오후 01:43 16 .emulator_console_auth_token
2020-04-20 오후 04:58 177 .gitconfig
2019-10-01 오후 09:24 <DIR> .gradle
2019-10-04 오후 10:24 <DIR> .IdeaIC2019.2
2019-12-10 오전 11:01 <DIR> .idlerc
2020-05-07 오후 02:11 <DIR> .ipynb_checkpoints
2019-11-19 오후 04:15 <DIR> .ipython
2019-11-19 오후 04:21 <DIR> .jupyter
2019-09-05 오후 08:11 <DIR> .m2
2020-04-23 오후 02:58 <DIR> .matplotlib
2019-12-17 오후 09:50 <DIR> .p2
2019-08-02 오후 05:05 <DIR> .PyCharmCE2019.2
2019-09-27 오후 09:31 <DIR> .pylint.d
2019-06-27 오후 12:49 <DIR> .tooling
2020-02-05 오후 11:27 <DIR> .VirtualBox
2019-11-09 오후 04:42 <DIR> .vscode
2020-04-15 오후 10:00 <DIR> 3D Objects
2020-05-06 오전 10:12 3,730 accident.csv
2019-11-24 오후 04:13 <DIR> AlcoholTemp
2020-04-16 오후 03:19 <DIR> anaconda3
2019-10-01 오후 10:17 <DIR> AndroidStudioProjects
2020-02-21 오후 09:28 <DIR> BackUpAndroidBoostCourse
2019-12-04 오후 02:38 <DIR> connector_mysql
2020-04-15 오후 10:00 <DIR> Contacts
2020-04-10 오후 04:23 <DIR> Creative Cloud Files
2019-08-30 오후 07:39 <DIR> data
2020-05-05 오후 10:46 <DIR> Desktop
2020-04-20 오후 04:59 <DIR> Documents
2019-07-02 오후 02:46 <DIR> DoltAndroidRev5-master
2020-05-07 오전 11:34 <DIR> Downloads
2019-10-15 오후 11:54 <DIR> eclipse-workspace
2019-02-18 오전 10:37 58 ex1.csv
2019-02-18 오전 10:37 42 ex2.csv
2019-02-18 오전 10:37 163 ex4.csv
2019-02-18 오전 10:37 78 ex5.csv
2020-04-15 오후 10:00 <DIR> Favorites
2020-04-27 오후 11:06 <DIR> GoUnity
2019-04-25 오전 09:49 632 image.svg
2019-11-08 오후 10:53 502 image.zip
2019-12-03 오후 04:24 <DIR> jobphoto
2019-11-29 오후 02:00 <DIR> kakaosdk
2020-05-07 오후 02:04 2,462 Lecture8.zip
2020-04-15 오후 10:00 <DIR> Links
2020-01-12 오후 07:43 <DIR> MallProject
2020-03-05 오후 08:19 <DIR> MovieProject2

```


2020-04-15	오후 10:00	<DIR>	Music
2019-11-24	오후 04:05		1,461,054 MyAlcohol.zip
2020-02-10	오후 01:09	<DIR>	MyDrawer
2020-02-10	오후 09:43	<DIR>	MyDrawerExample
2020-04-29	오후 04:22	<DIR>	New Unity Project
2020-04-15	오후 09:01	<DIR>	OneDrive
2019-11-29	오후 12:47	<DIR>	OpenSSL
2020-05-07	오후 02:44		87 out.csv
2020-04-15	오후 10:00	<DIR>	Pictures
2019-10-06	오후 02:00	<DIR>	PycharmProjects
2019-11-19	오후 04:20		23,879 recommendation_system.ipynb
2019-10-09	오후 12:50		9,501 robot.jpg
2020-04-15	오후 10:00	<DIR>	Saved Games
2020-04-13	오전 10:14	<DIR>	SDDH
2020-04-10	오후 07:19	<DIR>	sddhs
2020-04-15	오후 10:00	<DIR>	Searches
2019-06-22	오전 09:27	<DIR>	source
2019-12-03	오후 01:31		6,012,107 sqlite-jdbc-3.27.2.1.jar
2019-11-29	오후 12:47	<DIR>	ssl
2020-02-08	오후 02:21	<DIR>	TempBoostCourse
2020-04-21	오후 11:02	<DIR>	ToDoList
2019-11-19	오후 04:17		555 Untitled.ipynb
2020-04-16	오후 03:34		7,954 Untitled1.ipynb
2020-04-23	오후 03:23		1,042,052 Untitled2.ipynb
2020-04-30	오전 10:46		22,273 Untitled3.ipynb
2020-04-30	오전 11:03		807,592 Untitled4.ipynb
2020-04-30	오전 11:02		108,342 Untitled5.ipynb
2020-04-30	오전 10:43		95,402 Untitled6.ipynb
2020-05-07	오후 03:07		138,704 Untitled7.ipynb
2020-04-15	오후 10:00	<DIR>	Videos
2020-04-10	오후 05:09	<DIR>	VirtualBox VMs
2020-04-12	오후 06:21	<DIR>	VisualStudio2017
2019-08-21	오후 01:44		192 _netrc
2019-11-24	오후 04:14	<DIR>	안드로이드(Alcohol)프로젝트 보관 파일
	25개 파일		9,739,738 바이트
	60개 디렉터리		53,398,040,576 바이트 남음

In [54]:



```
df=pd.read_csv('accident.csv',engine='python')
df
```

Out [54]:

	도로종류	시간대	사고건수	사망자수	부상자수	중상	경상	부상신고
0	일반국도	00시-02시	878	87	1428	481	894	53
1	일반국도	02시-04시	535	51	820	273	502	45
2	일반국도	04시-06시	554	67	876	355	479	42
3	일반국도	06시-08시	1262	85	2180	727	1374	79
4	일반국도	08시-10시	1757	71	3175	993	2062	120
...
79	기타	14시-16시	1943	42	2426	735	1520	171
80	기타	16시-18시	2242	56	2820	876	1766	178
81	기타	18시-20시	2228	50	2746	768	1779	199
82	기타	20시-22시	1610	22	1964	529	1287	148
83	기타	22시-24시	1256	10	1577	437	1031	109

84 rows × 8 columns

In [57]:



```
df[df['도로종류']=='일반국도'].plot()
```

Out [57]:

<matplotlib.axes._subplots.AxesSubplot at 0x142dac91408>

In []:

