

Pandas - 시각화 하기

전처리한 데이터를 기반으로 그래프, 차트를 그려보는 시각화에 대해서 알아보겠습니다. Pandas에 내장된 Matplotlib 기능을 이용한 시각화 기능에 대해서 알아보겠습니다.

1. 시각화를 위한 폰트 설정하기

```
# 폰트 설정을 하지 않으면 오류가 발생합니다.

import matplotlib.pyplot as plt
plt.rc("font", family="Malgun Gothic")

# 맥(mac) 사용자는 apple gothic
# plt.rc("font", family="Malgun Gothic")
```

```
C:\Users\seon\anaconda3\lib\site-packages\matplotlib\backends\backend_agg.py:201: RuntimeWarning: Glyph 51648 missing from current font.
font.set_text(s, 0, flags=flags)
C:\Users\seon\anaconda3\lib\site-packages\matplotlib\backends\backend_agg.py:201: RuntimeWarning: Glyph 50669 missing from current font.
font.set_text(s, 0, flags=flags)
C:\Users\seon\anaconda3\lib\site-packages\matplotlib\backends\backend_agg.py:201: RuntimeWarning: Glyph 47749 missing from current font.
font.set_text(s, 0, flags=flags)
```

2. 시각화하기

```
# pandas의 plot을 활용하면 다양한 그래프를 그릴 수 있습니다
# 지역명으로 분양가격의 평균을 구하고 선그래프로 시각화 합니다.
g = df_last.groupby(['지역명'])['평당분양가격'].mean()
# 정렬 > .sort_values(ascending=False)

# 선 그래프 그리기(기본값)
g.plot()
# g.plot(kind='bar')

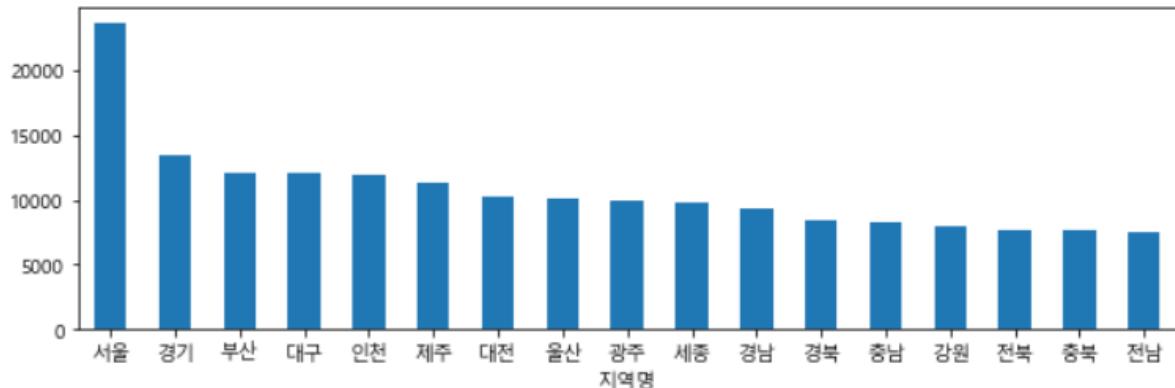
'''
'line' : line plot (default)
'bar' : vertical bar plot
'barh' : horizontal bar plot
'hist' : histogram
'box' : boxplot
'kde' : Kernel Density Estimation plot
'density' : same as 'kde'
'area' : area plot
'pie' : pie plot
'scatter' : scatter plot
'hexbin' : hexbin plot.
'''

# 히스토그램 그리기 - 평당분양가격 분포
# g.hist()

# 글씨를 바르게 회전, 간격 맞추기
# g.plot.bar(rot=0, figsize=(10, 3))
```

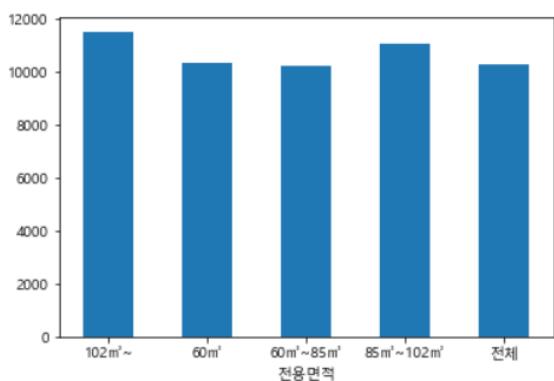
```
# 지역명으로 분양가격의 평균을 구하고 막대그래프(bar)로 시각화 하기  
g.plot.bar(rot=0, figsize=(10, 3))
```

<AxesSubplot:xlabel='지역명'>



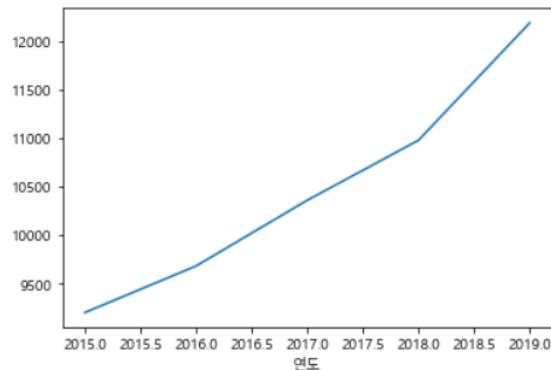
```
# 전용면적으로 분양가격의 평균을 구하고 막대그래프(bar)로 시각화 합니다.  
df_last.groupby(['전용면적'])['평당분양가격'].mean().plot.bar(rot=0)  
  
# df_last.groupby(['연도'])['평당분양가격'].mean().plot()  
# df_last.groupby(['연도'])['평당분양가격'].mean().hist()
```

<AxesSubplot:xlabel='전용면적'>



```
# 연도별 분양가격의 평균을 구하고 막대그래프로 시각화 합니다  
df_last.groupby(['연도'])['평당분양가격'].mean().plot
```

```
<AxesSubplot:xlabel='연도'>
```



```
# box plot (https://ko.wikipedia.org/wiki/%EC%83%81%EC%9E%90\_%EC%88%98%EC%97%BC\_%EA%B7%B8%EB%A6%BC)
# 상자수염그림
# 5가지 요약수를 이용해 통계를 확인할 수 있음

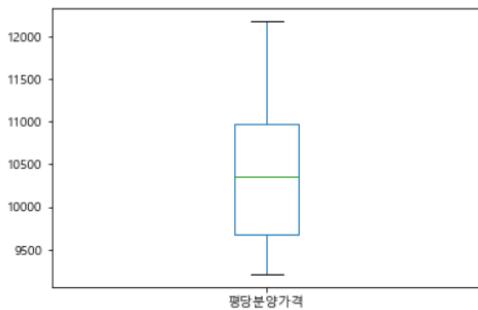
...
최솟값
제 1사분위수
제 2사분위수( ), 즉 중앙값
제 3 사분위 수( )
최댓값
...

# 연도별 평당분양가격
df_last.pivot_table(index='연도', values='평당분양가격').plot.box()

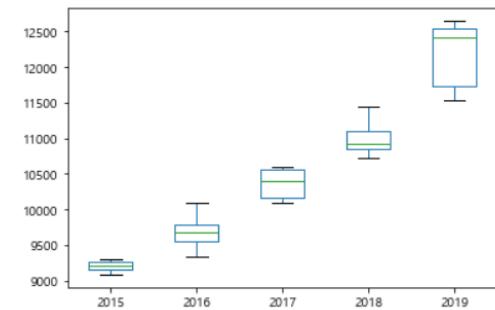
# 월별 평당분양가격
df_last.pivot_table(index='월', columns='연도', values='평당분양가격').plot.box()

# 전용면적별 평당분양가격
# df_last.pivot_table(index='전용면적', columns='연도', values='평당분양가격').plot.box()
```

```
<AxesSubplot:>
```

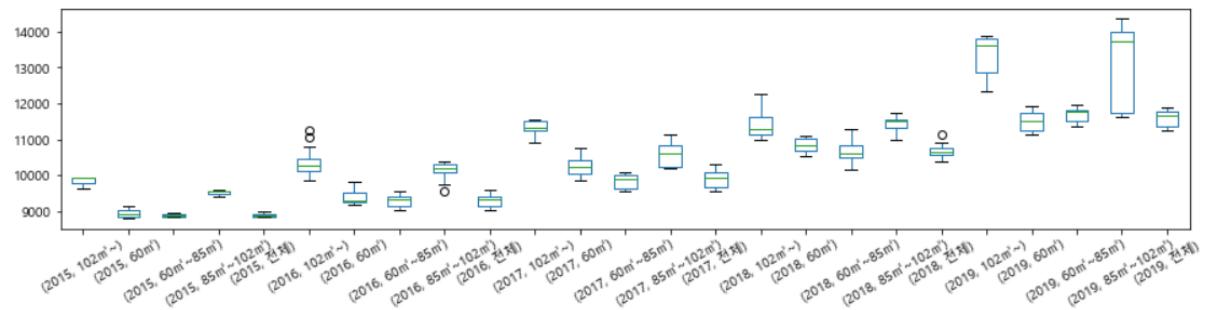


```
<AxesSubplot:>
```



```
# columns에 "연도", "전용면적"을 추가해서 pivot_table을 만들고 시각화 합니다.
p = df_last.pivot_table(index='월', columns=['연도', '전용면적'], values='평당분양가격')
p.plot.box(figsize=(15, 3), rot=30)
```

<AxesSubplot :>



```
# index를 월, columns 를 지역명으로 구하고 평당분양가격 으로 pivot_table 을 구하고 선그래프를 그립니다.
```

```
p = df_last.pivot_table(index='월', columns=['지역명'], values='평당분양가격')
p.plot.bar(figsize=(15, 3), rot=30)
```

<AxesSubplot:xlabel='월'>

