제출 파일 : \*.ipynb, ….

0. 전국 신규 민간 아파트 분양가격 동향 : 공공 데이터 포털에 있는 데이터를 Pandas의 melt, concat, pivot, transpose와 같은 reshape가능을 활용해 분석해 봅니다. 그리고 groupby, pivot\_table, info, describe, value\_counts 등을 통한 데이터 요약과 분석을 해 봅니다. 이를 통해 전혀 다른 형태의 두 데이터를 가져와 정제하고 병합하는 과정을 다루는 방법도 사용합니다. 전처리 한 결과에 대해 수치형, 범주형 데이터의 차이를 이해하고 다양한 그래프로 시각화를 합니다.

1. 시각화 그래프에 한글지정을 하거나 경고메세지를 안보이게 하는 등의 설정한다.
2. 다운로드 <https://www.data.go.kr>에서 “**주택도시보증공사\_전국 신규 민간아파트 분양가격 동향**”검색한 자료는 다운 받는다.
   1. 2013년9월부터 2015년8월까지 데이터는 **df\_first** 변수에 담아 탐색
   2. 이후 데이터는 **df\_last** 변수에 담아 탐색
3. df\_last의 데이터 결측치를 확인하고 대체할지 고민한다.
4. df\_last 데이터셋에 평당분양가격 컬럼을 추가한다.(① 6,222 => 6222 수정, space가 저장된 경우 nan으로 수정하는 함수를 apply나 map적용 ②astype(‘float’)또는 pd. to\_numeric 적용)
5. df\_last 데이터셋에서 전용면적 컬럼을 추가한다(아래와 같이 규모구분 컬럼을 이용한다)

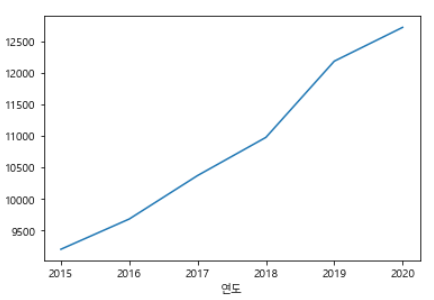
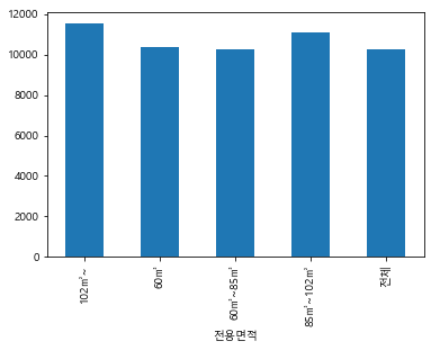
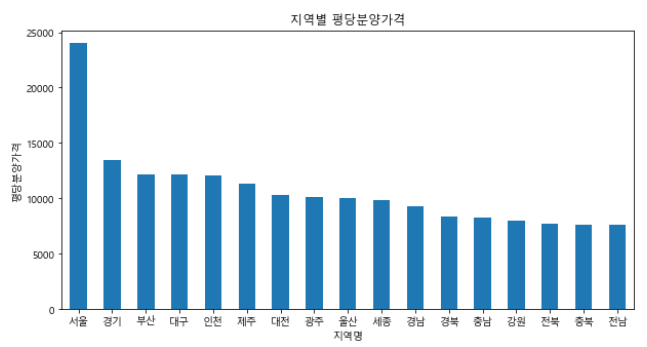
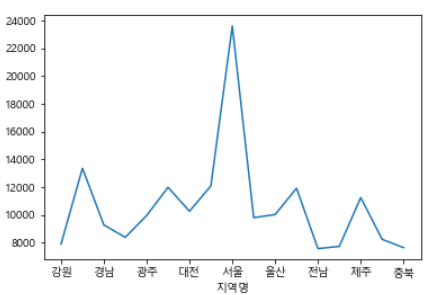
전용면적 열을 '60㎡', '60㎡~85㎡', '85㎡~102㎡', '102㎡~','모든면적'로 생성



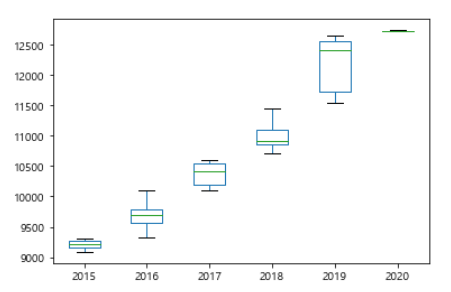
1. 메모리 사용량을 줄이기 위해 사용하지 않을 컬럼(규모구분과 분양가격((㎡)컬럼)은 제거한 후, 메모리 사용량을 전후로 확인한다.
2. GroupBy(unstack()함수 이용) vs. pivot\_table
3. 지역별 데이터 수 (cf. 평당분양가격이 있는 데이터 수)
4. 지역별 평당분양가격 (평균)
5. 전용면적별 평당분양가격(평균)
6. 지역별, 전용면적별 평당분양가격 (평균)
7. 연도, 지역별 평당분양가격(평균) – 힌트:groupby후 unstack() vs. pivot\_table



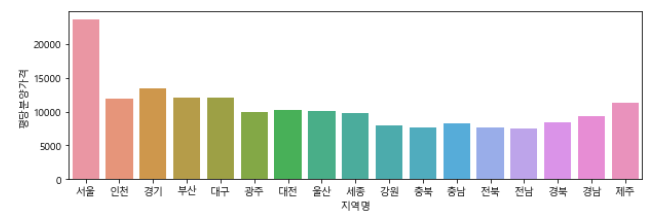
1. 위의 그룹화한 내용을 선그래프와 막대그래프로 시각화



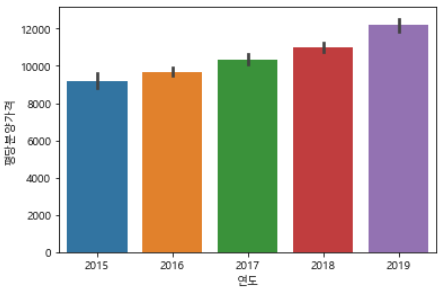
연도별(월별데이터로), 전용면적별 box플롯



1. seaborn으로 시각화(위에서 그린 시각화내용을 일부또는 그 이상을 seaborn으로 시각화
2. 지역별 평당분양가격(평균)



1. 연도별 평당분양가격(평균)



1. 그 외도 추가
2. 구조가 다른 df\_first와 df\_last의 전용면적 전체의 데이터를 하나의 같은 데이터 셋으로 합쳐서 분석을 한다. 분석은 위의 내용을 참조하여 각자 분석한다(전용면적별 분석 불가)

① 년도(int)별 평당분양가격

* 1. 년도별 지역별 평당분양가격

violinplot: <https://www.research.autodesk.com/publications/same-stats-different-graphs/>

컬러팔레트 :

<https://seaborn.pydata.org/generated/seaborn.color_palette.html#seaborn.color_palette>  
<https://matplotlib.org/tutorials/colors/colormaps.html>  
<https://seaborn.pydata.org/tutorial/color_palettes.html>

범례:<https://stackoverflow.com/questions/30490740/move-legend-outside-figure-in-seaborn-tsplot>