

데이터 분석 & 시각화 Team Project

● 서울특별시 아파트 실거래가 분석

2019년과 2021년 비교를 중심으로

7조_김한결 박준석 장서희

목차

CONTENTS

1

개요 INTRO

- 데이터 소개
- 데이터 전처리
 - 주제 선정

2

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

- 분석 및 시각화 과정
 - Color Chart
 - 분석 및 시각화

3

결론 및 제언 CONCLUSION

TITLE 01

개요 INTRO

- 데이터 소개
- 데이터 전처리
 - 주제 선정

TITLE 01

개요 INTRO 데이터 소개

•법정동 코드.xlsx

[법정동코드, 시도명, 시군구명, 읍면동명, 동리명, 생성일자, 말소일자]

	법정동코드	시도명	시군구명	읍면동명	동리명	생성일자	말소일자
0	1100000000	서울특별시	NaN	NaN	NaN	19880423	NaN
1	1111000000	서울특별시	종로구	NaN	NaN	19880423	NaN
2	1111010100	서울특별시	종로구	청운동	NaN	19880423	NaN
3	1111010200	서울특별시	종로구	신교동	NaN	19880423	NaN
4	1111010300	서울특별시	종로구	궁정동	NaN	19880423	NaN

•2019년 서울 아파트 실거래가.csv

[건축년도, 거래년도, 거래월, 거래일자, 법정동, 거래금액, 아파트이름, 전용면적, 지역코드, 층, 지번]

	건축년도	거래년도	거래월	거래일자	법정동	거래금액	아파트이름	전용면적	지역코드	층	지번
0	2008	2019	1	7	사직동	105,000	광화문풍림스페이스본(101동~105동)	97.61	11110	10	9
1	2008	2019	1	11	사직동	162,000	광화문풍림스페이스본(106동)	163.33	11110	11	9-1
2	2008	2019	1	19	사직동	119,000	광화문풍림스페이스본(101동~105동)	131.44	11110	13	9
3	2008	2019	1	30	사직동	119,000	광화문풍림스페이스본(101동~105동)	147.31	11110	1	9
4	2004	2019	1	25	내수동	160,000	경희궁의아침4단지	150.48	11110	14	73

•2021년 서울 아파트 실거래가.csv

[시군구, 번지, 본번, 부번, 단지명, 전용면적(㎡), 계약년월, 계약일, 거래금액(만원), 층, 건축년도, 도로명, 해제사유발생일, 거래유형, 중개사소재지]

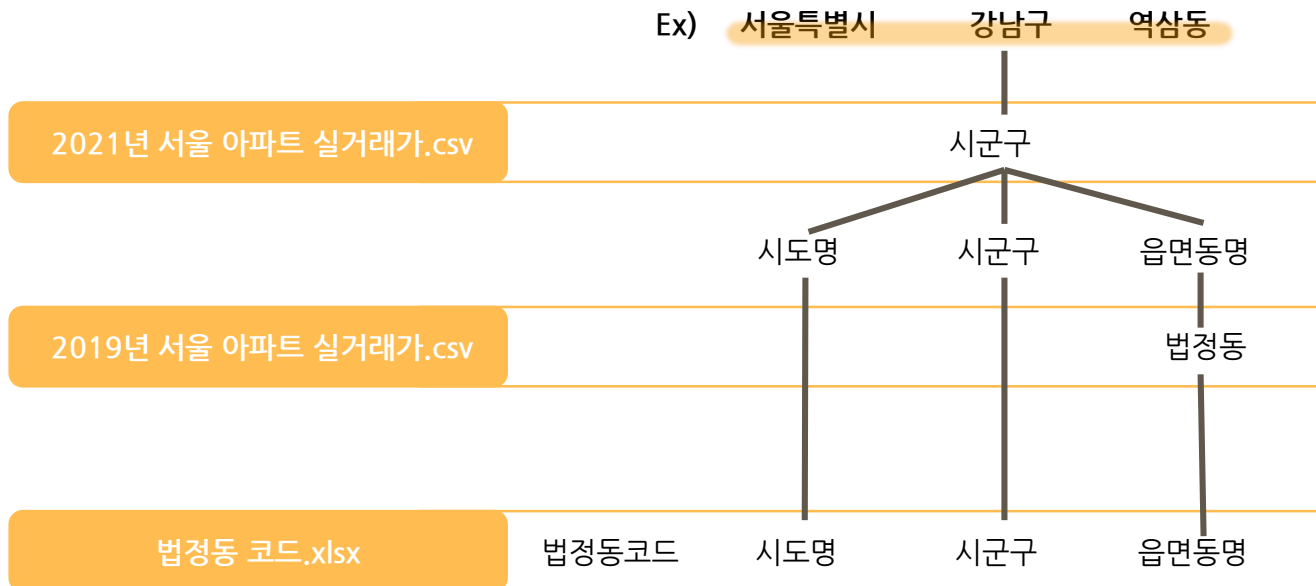
	시군구	번지	본번	부번	단지명	전용면적(㎡)	계약년월	계약일	거래금액(만원)	층	건축년도	도로명	해제사유발생일	거래유형	중개사소재지
0	서울특별시 강남구 개포동	655-2	655.0	2.0	개포2차현대아파트(220)	77.75	202104	15	174,000	9	1988	연주로 103	NaN	-	-
1	서울특별시 강남구 개포동	658-1	658.0	1.0	개포6차우성아파트1동~8동	79.97	202105	15	195,000	5	1987	연주로 3	NaN	-	-
2	서울특별시 강남구 개포동	658-1	658.0	1.0	개포6차우성아파트1동~8동	79.97	202108	11	200,000	3	1987	연주로 3	NaN	-	-
3	서울특별시 강남구 개포동	658-1	658.0	1.0	개포6차우성아파트1동~8동	79.97	202109	2	220,000	2	1987	연주로 3	NaN	-	-
4	서울특별시 강남구 개포동	658-1	658.0	1.0	개포6차우성아파트1동~8동	79.97	202112	4	215,000	3	1987	연주로 3	NaN	중개거래	서울 강남구

TITLE 01

● KB국민은행 IT's your Life

개요 INTRO 데이터 전처리

전처리 전략1_같은 값을 저장한 서로 다른 컬럼 통일 시키기



- 같은 값에 대해 서로 다른 형식으로 표현되어 있음
- 모두 같은 값인 시도명에 대한 컬럼은 제외하고 **구와 동 정보를 담은 컬럼을 통일** 시켜주기로 함
- 2021년 데이터의 경우 시군구 컬럼을 시도명(서울특별시), 시군구(강남구), 읍면동명(역삼동)으로 나누어 줌
- 2019년 데이터의 경우 법정동 코드 데이터와 읍면동명을 기준으로 merge해서 구에 대한 컬럼 추가

→ 'gu'와 'dong'컬럼으로 통일

TITLE 01

● KB국민은행 IT's your Life

개요 INTRO 데이터 전처리

전처리 전략2_결측치 확인 후 Drop&Fill

```
apt_21.isnull().sum() # 결측치 존재
```

시군구	0
번지	14
본번	5
부번	5
단지명	0
전용면적(m²)	0
계약년월	0
계약일	0
거래금액(만원)	0
층	0
건축년도	0
도로명	0
해제사유발생일	42003
거래유형	0
중개사소재지	0

dtype: int64

```
apt_21.loc[apt_21['본번'].isnull()]
```

	시군구	번지	본번	부번	단지명	전용면적(m²)	계약년월	계약일	거래금액(만원)	층	건축년도	도로명	해제사유발생일	거래유형	중개사소재지
27658	서울특별시 서초구 신원동	NaN	NaN	NaN	힐스테이트 서초 센트리스	84.95	202103	11	163,000	2	2014	현릉로8길 10-12	NaN	-	-
27659	서울특별시 서초구 신원동	NaN	NaN	NaN	힐스테이트 서초 센트리스	101.90	202104	15	187,500	4	2014	현릉로8길 10-12	NaN	-	-
27660	서울특별시 서초구 신원동	NaN	NaN	NaN	힐스테이트 서초 센트리스	101.90	202104	27	184,500	3	2014	현릉로8길 10-12	NaN	-	-
27661	서울특별시 서초구 신원동	NaN	NaN	NaN	힐스테이트 서초 센트리스	84.95	202105	26	165,000	4	2014	현릉로8길 10-12	NaN	-	-
27662	서울특별시 서초구 신원동	NaN	NaN	NaN	힐스테이트 서초 센트리스										

힐스테이트서초센트리스

아파트 · 256세대 · 중9동 · 2014.12. · 114㎡ ~ 153㎡

🔍 길찾기 | 📌 관심단지 | 🔄 공유

☎ 관리사무소 02-578-3332

🏠 면적 114A㎡, 114B㎡, 135㎡, 153㎡ [평면도 보기](#)

📍 배정학교 - [연남초](#)

📍 서울시 서초구 현릉로8길 10-12 · [지번](#) 신원동 557 [지도보기](#)

apt_21['본번'].fillna(10, inplace=True)
apt_21['부번'].fillna(12, inplace=True)

- '해제사유발생일' 컬럼의 경우 분석에 사용하지 않을 것이기 때문에 결측치 처리 하지 않고 컬럼을 drop
- '번지' 컬럼의 경우 결측치 외에도 잘못 입력된 값이 존재 --> 때문에 해당 컬럼을 사용하지 않고 drop
- '본번'과 '부번'의 내용을 합쳐 19년 실거래가 데이터의 '지번'과 같은 정보를 담은 컬럼을 만들 수 있음 따라서, 결측치의 수가 적기 때문에 내용 확인 후 채워넣는 전략 수립!

TITLE 01

● KB국민은행 IT's your Life

개요 INTRO 데이터 전처리

전처리 전략3&4_법정동코드를 읍면동 단위까지 처리&강남or강북 컬럼 추가



- 데이터 분석의 범위를 읍면동까지 살펴볼 예정이었기 때문에 해당 코드의 내용도 읍면동까지 코드를 남기기로 결정
- 추후, 지도 시각화를 하게 된다면 읍면동 단위의 shp파일과 법정동 코드를 통해 merge하기 위해 코드를 남기기로 결정



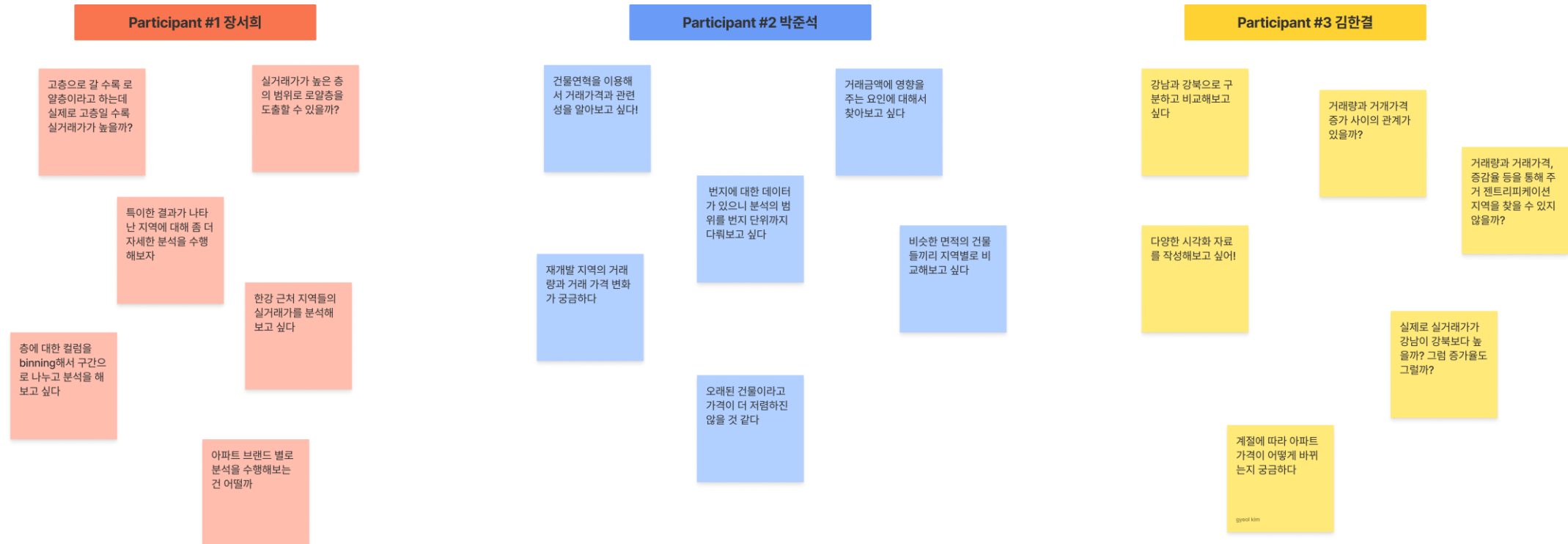
- 아파트 가격의 경우 일반적으로 강남과 강북의 차이가 클 것이라고 예상됨
- 따라서 관련 분석을 수행하기 위해 강남과 강북을 구분하는 컬럼 추가적으로 생성

TITLE 01

● KB국민은행 IT's your Life

개요 INTRO 주제선정

BrainStoming을 통해 분석에서 다뤄 보고 싶은 것들을 나열



- 팀원 별 6가지 썩의 아이디어 도출

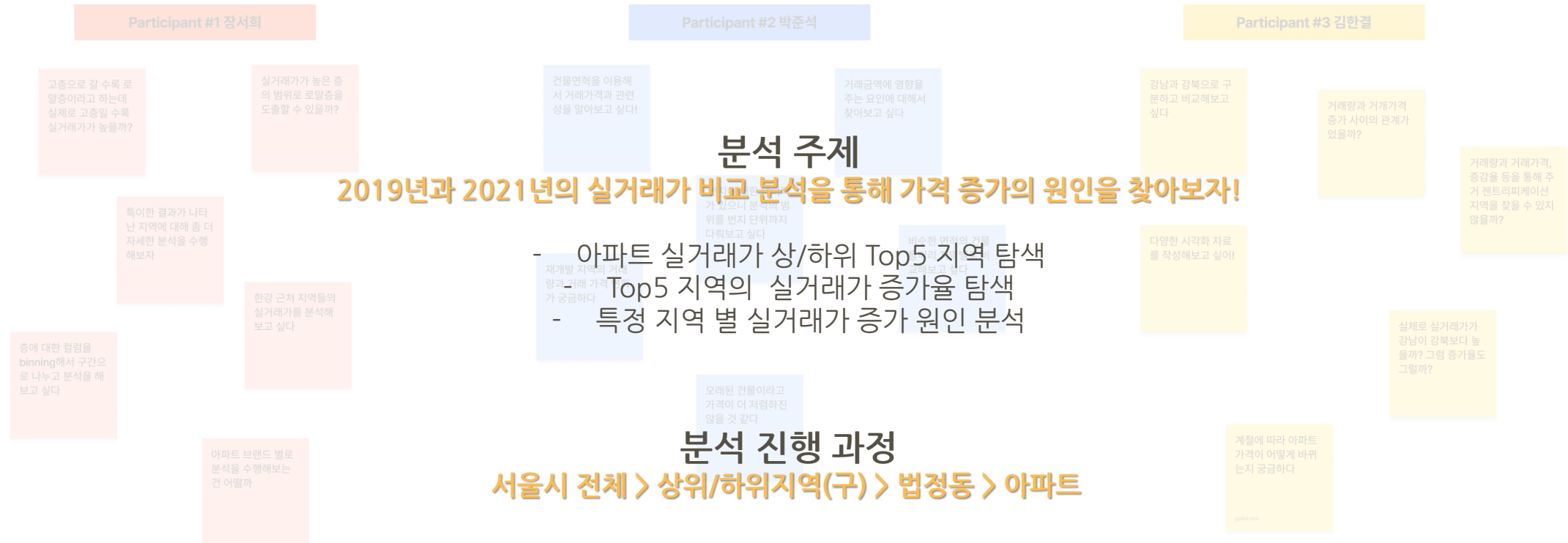
- 아이디어 중 몇 가지를 선정해 실제로 수행해보고
최종적으로 유의미하면서 다양한 분석이 나올 수 있는 주제로 최종 결정

TITLE 01

● KB국민은행 IT's your Life

개요 INTRO 주제선정

BrainStoming을 통해 분석에서 다뤄 보고 싶은 것들을 나열



- 팀원 별 6가지 썬크의 아이디어 도출
- 아이디어 중 몇 가지를 선정해 실제로 수행해보고 최종적으로 유의미하면서 다양한 분석이 나올 수 있는 주제로 최종 결정

TITLE 02

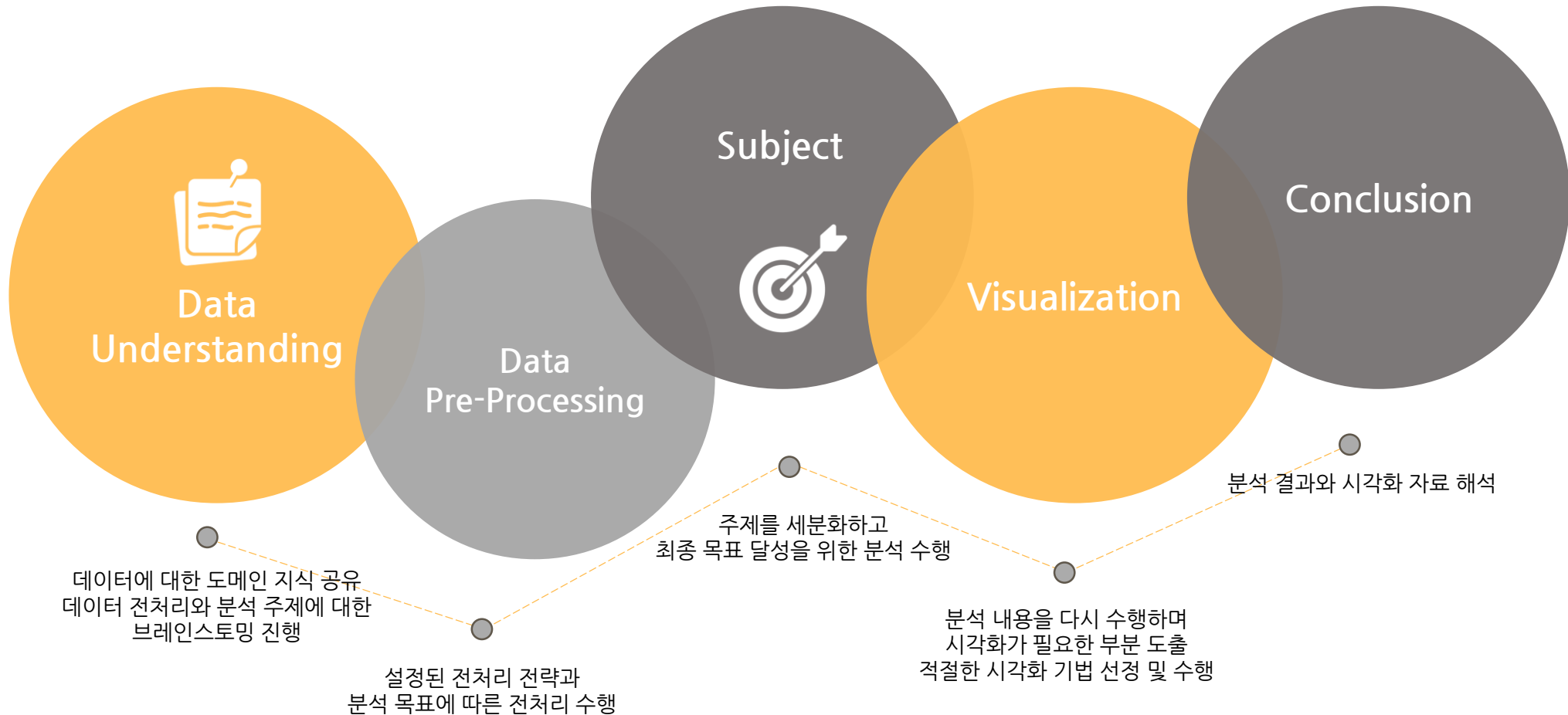
데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

- 분석 및 시각화 과정
 - 데이터 분석

TITLE 02

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

분석 및 시각화 과정

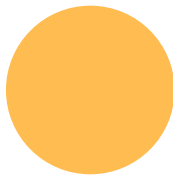


TITLE 02

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

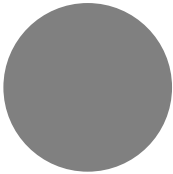
Color Chart

서울시 전체 분석



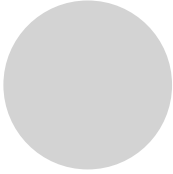
하이라이팅

성동구 분석



강조와 대비

노원구 분석



강조와 대비

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

아파트 실거래가가 높은 지역과 낮은 지역 파악하기

2019년 상위 TOP 5

구	평균 실거래가
강남구	172,963.05
서초구	165,753.55
용산구	153,938.59
송파구	120,263.64
성동구	96,172.01

2019년 하위 TOP 5

구	평균 실거래가
도봉구	40,032.84
금천구	43,253.85
노원구	43,992.32
중랑구	44,128.97
강북구	46,512.15

2021년 상위 TOP 5

구	평균 실거래가
강남구	207,051.75
서초구	204,859.02
용산구	199,079.17
송파구	154,369.24
성동구	144,537.89

2021년 하위 TOP 5

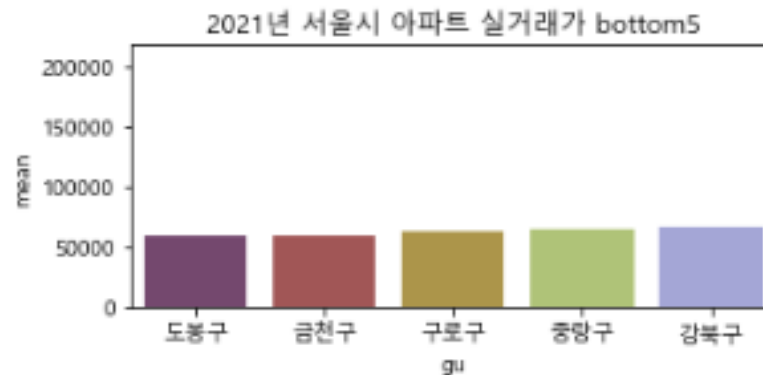
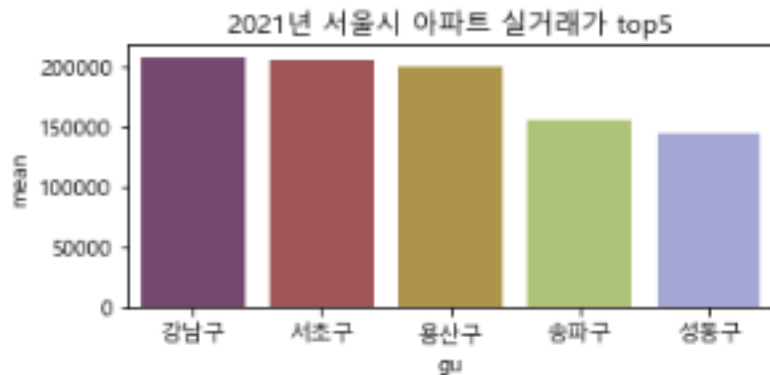
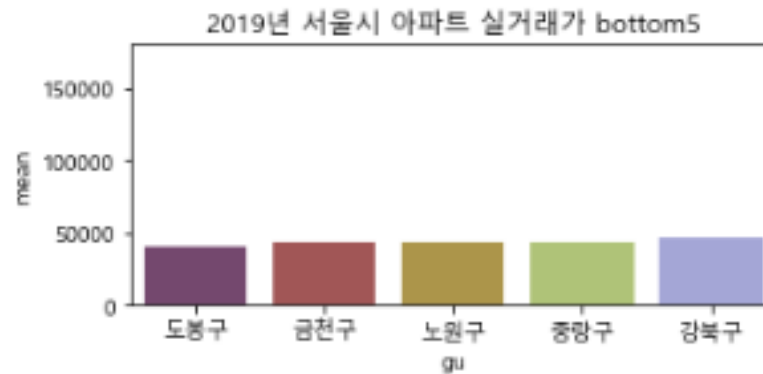
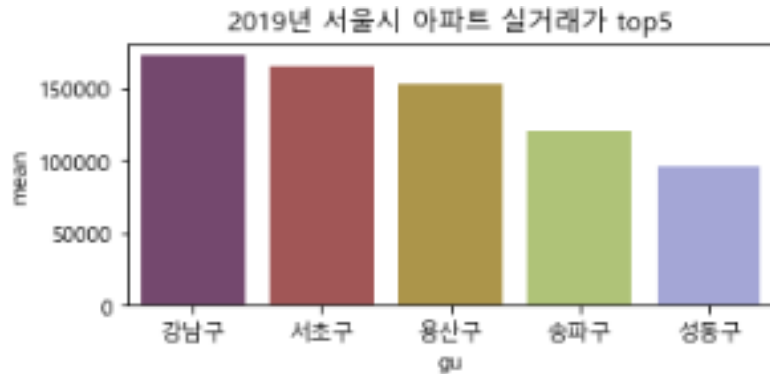
구	평균 실거래가
도봉구	58,448.15
금천구	59,569.06
구로구	61,995.49
중랑구	64,459.75
강북구	66,961.10

TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

아파트 실거래가가 높은 지역과 낮은 지역 파악하기



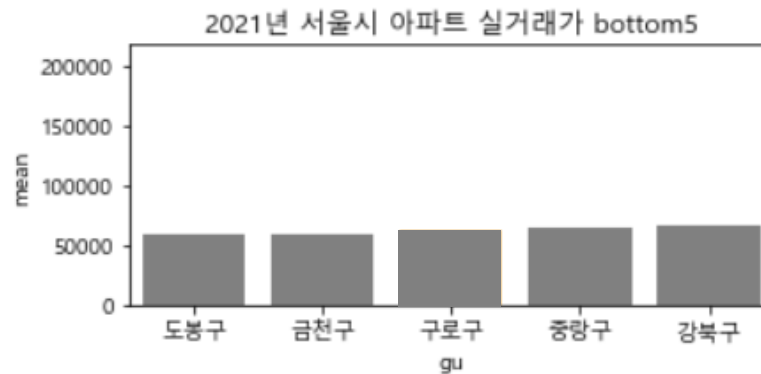
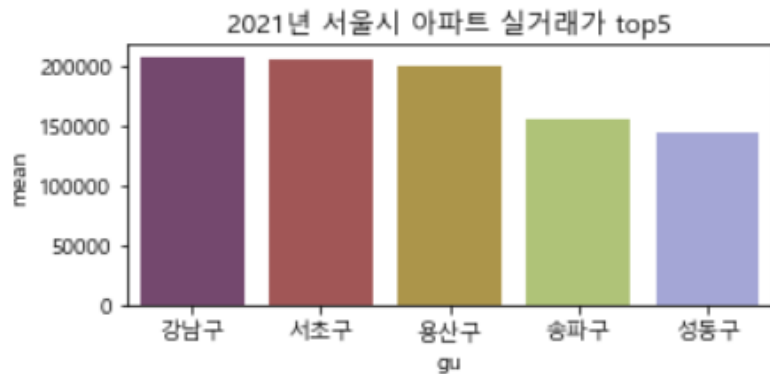
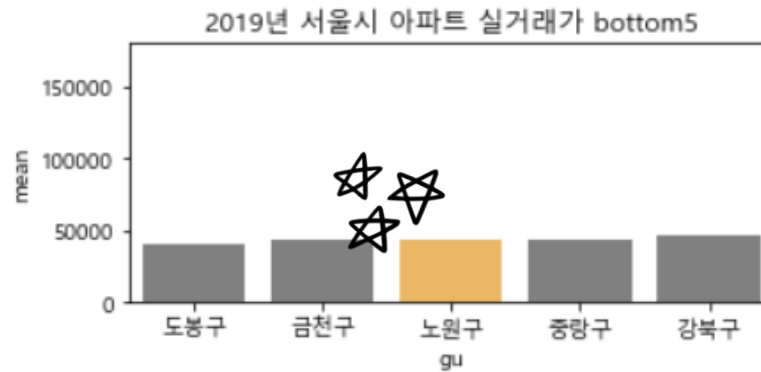
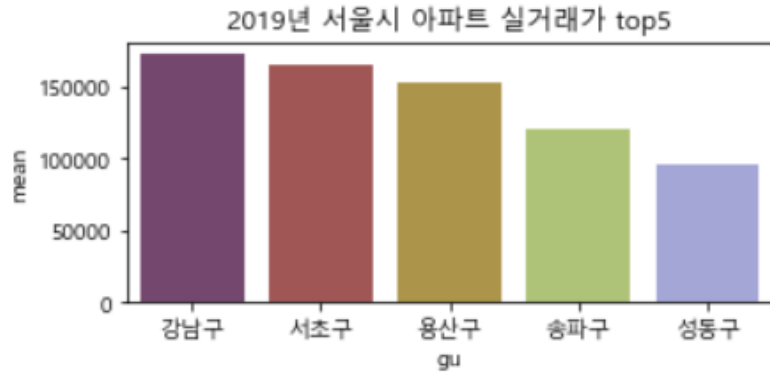
- 각 자치구의 평균 실거래가가 높은 5개 지역과 낮은 5개 지역을 살펴봄
- 실거래가가 높은 지역의 경우 순위의 변화가 없었지만, 실거래가가 낮은 지역은 3위 노원구가 순위 밖으로 밀려남

TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

아파트 실거래가가 높은 지역과 낮은 지역 파악하기



- 각 자치구의 평균 실거래가가 높은 5개 지역과 낮은 5개 지역을 살펴봄
- 실거래가가 높은 지역의 경우 순위의 변화가 없었지만, 실거래가가 낮은 지역은 3위 노원구가 순위 밖으로 밀려남

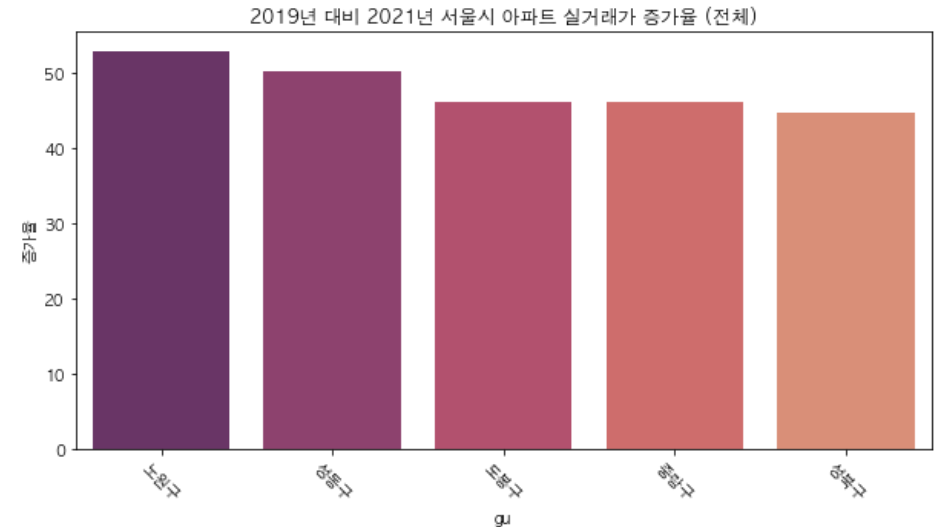
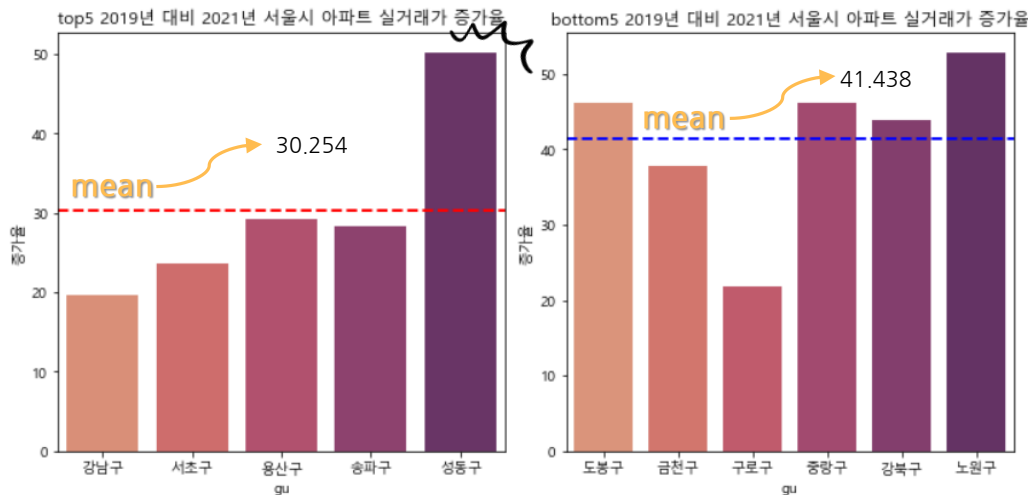
→ '노원구'를 집중적으로 분석할 필요가 있다고 판단!

TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

상위 5, 하위 5 지역의 아파트 실거래 상승률 파악하기



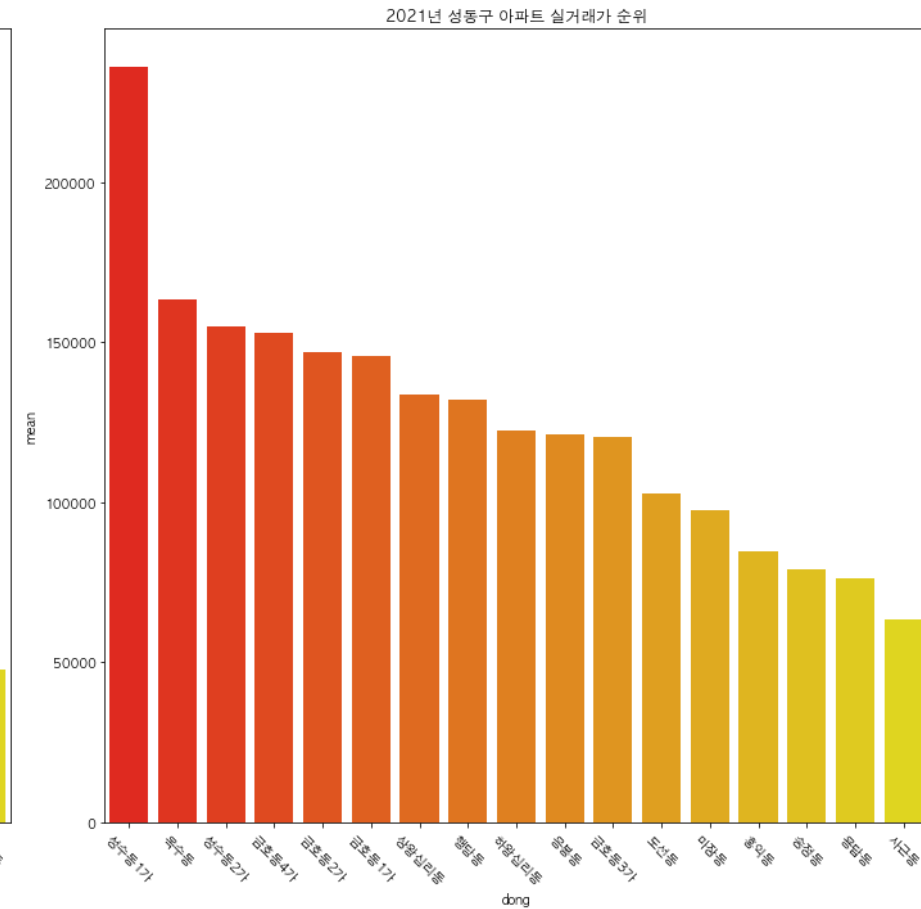
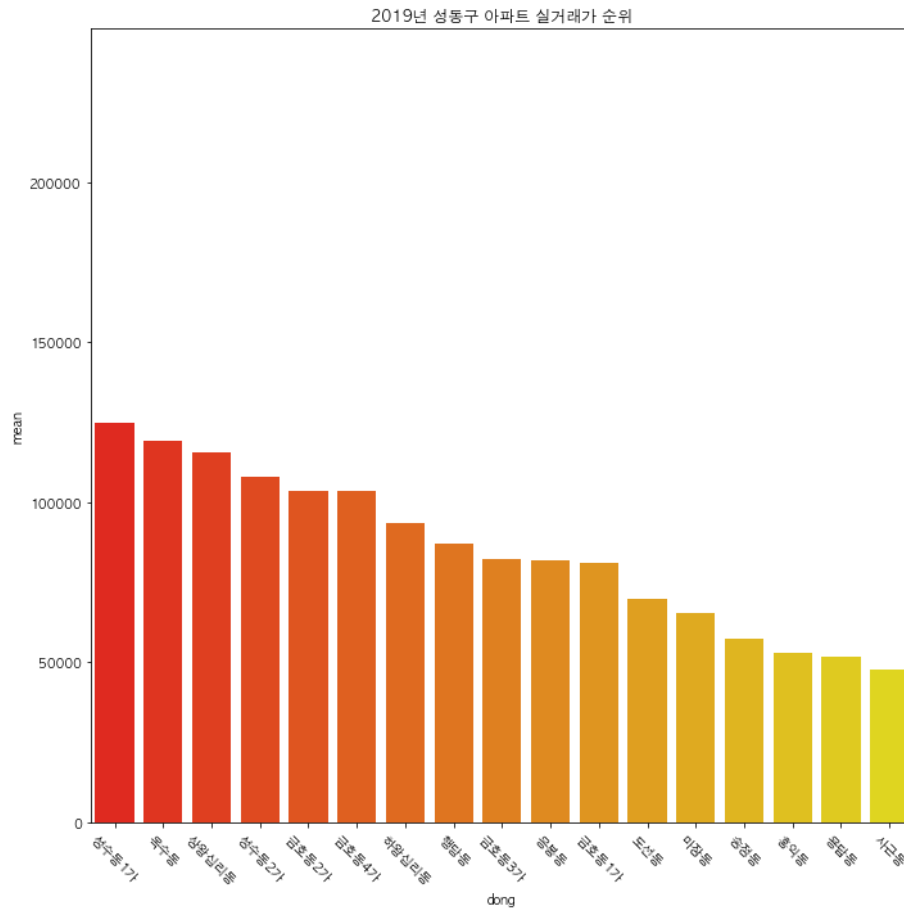
- 실거래가 상/하위 Top5지역의 증가율을 평균선과 그래프로 파악해보면
평균 증가율 자체는 상위 지역이 하위 지역에 비해 더 낮음을 알 수 있음
 - 또한, 서울시 전체의 범위에서 따졌을 때도 증가율 상위 5곳을 뽑으면
노원구, 도봉구, 중랑구처럼 실거래가가 낮음 지역들이 실거래가 증가율이 높음을 알 수 있음
 - BUT, **성동구의 경우 기존의 거래가가 매우 높은 지역임에도 불구하고
거래가 상승률 또한 서울시 전체에서 2위를 기록**
- 기존 저평가 되어있던 강북 지역의 집값 상승률의 대폭 상승으로는 설명되지 않는 부분
 - 다른 상위권 지역과 다른 양상을 띄는 **성동구**를 집중적으로 파악하기로 함

TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

성동구 소속 동별 2019년, 2021년 실거래가 순위

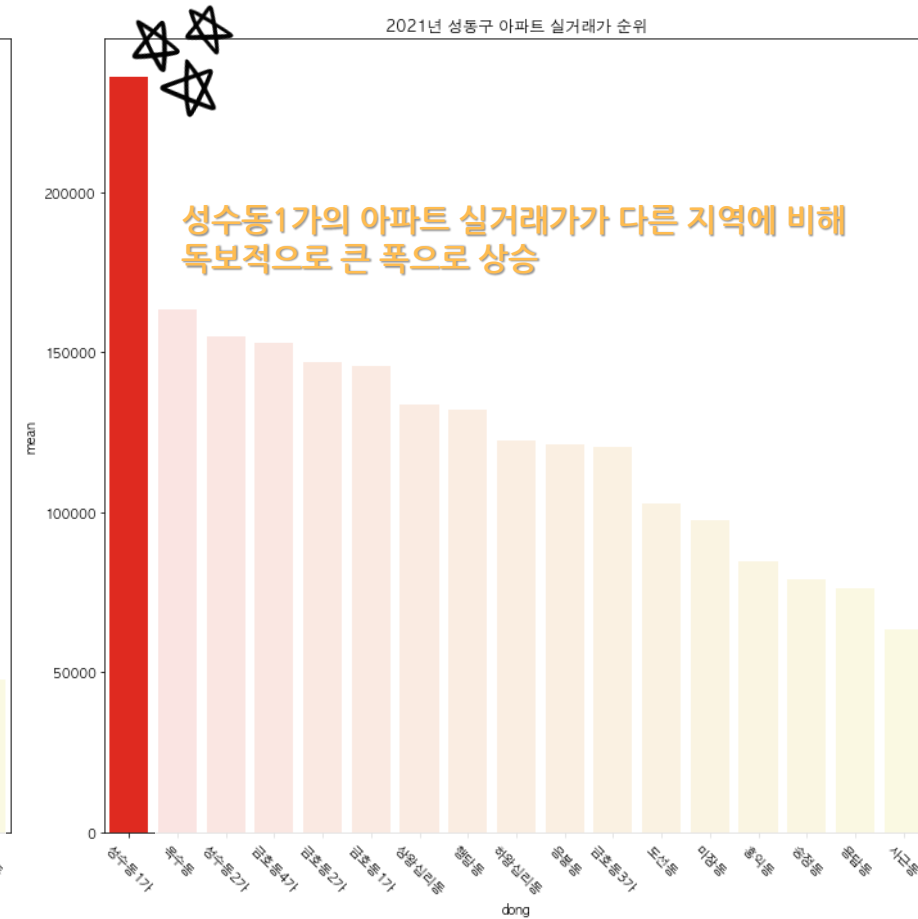
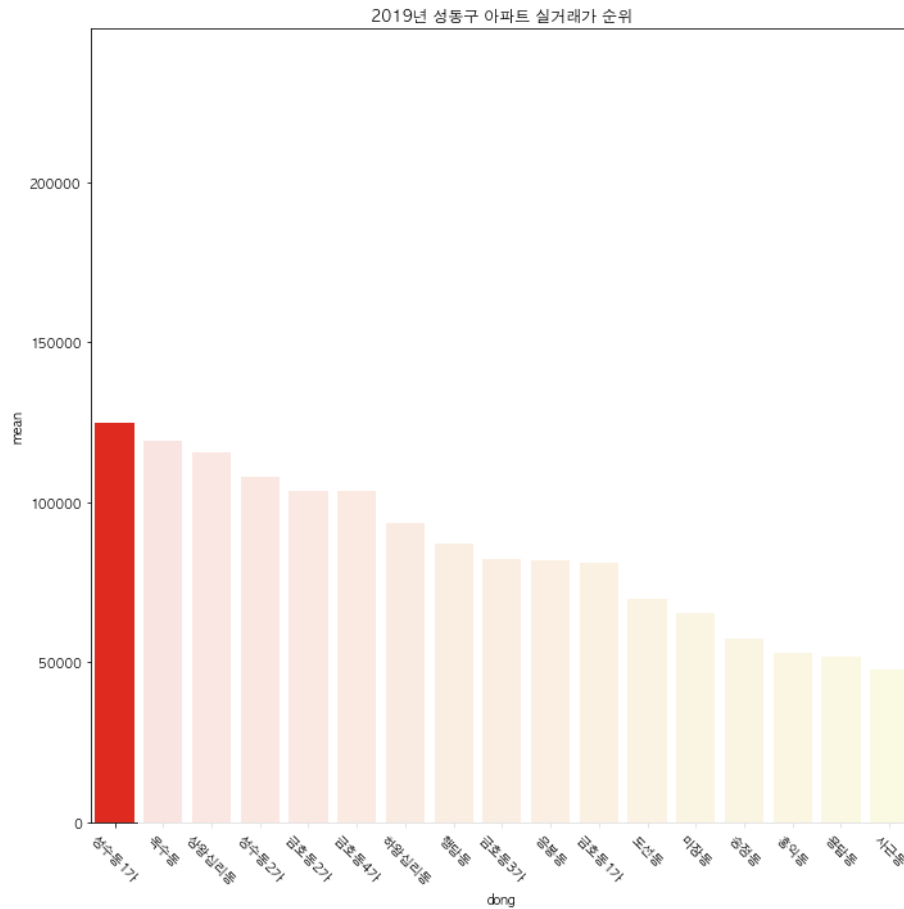


TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

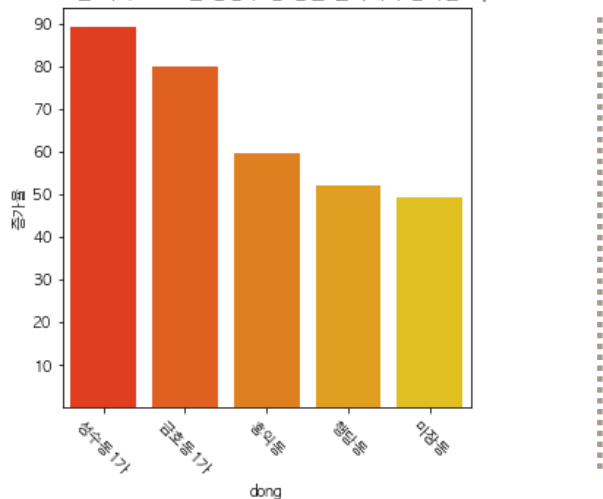
성동구 소속 동별 2019년, 2021년 실거래가 순위



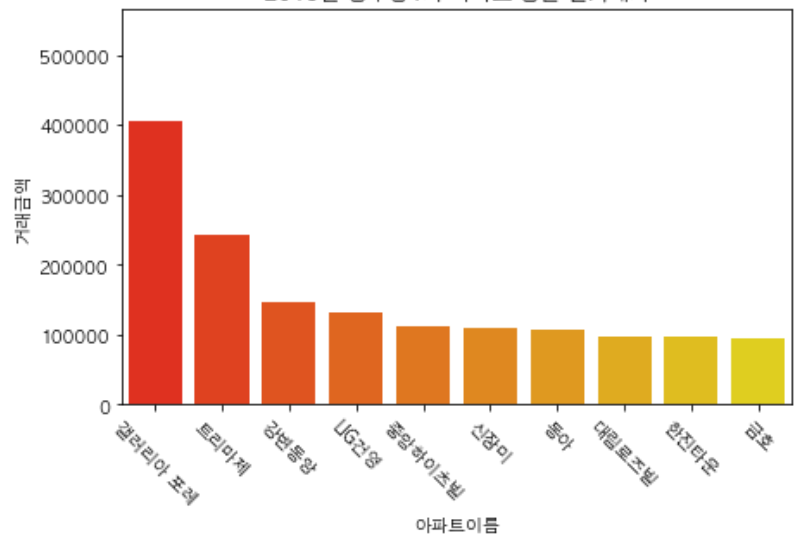
데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

시각화를 통한 성동구의 거래가 상승 원인 파악

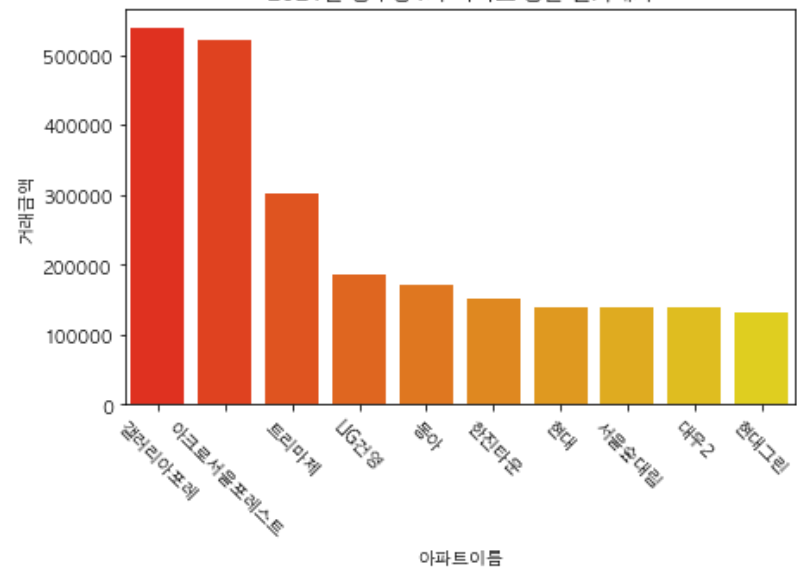
2019년 대비 2021년 성동구 동 평균 실거래가 증가율 top5



2019년 성수동1가 아파트 평균 실거래가



2021년 성수동1가 아파트 평균 실거래가



법정동	증감액	증가율(%)
성수동1가	111,291.02	89.18
금호동1가	64,644.90	79.89
홍익동	31,535.0	59.54
행당동	45,101.26	51.93
마장동	32,156.55	49.29

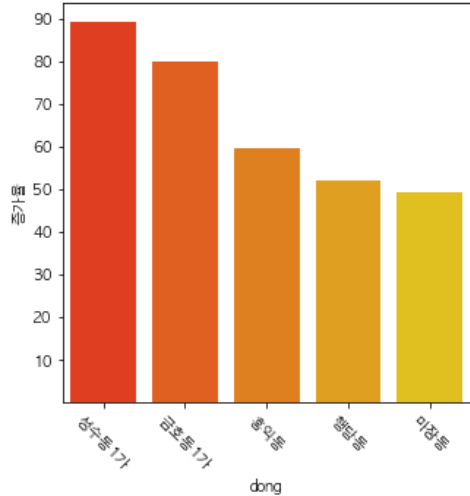
TITLE 02

● KB국민은행 IT's your Life

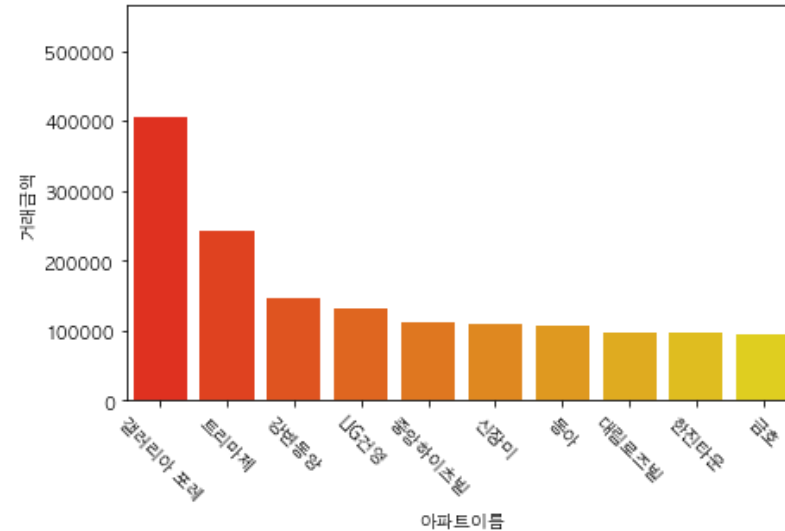
데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

성동구 소속 동별 2019년, 2021년 실거래가 순위

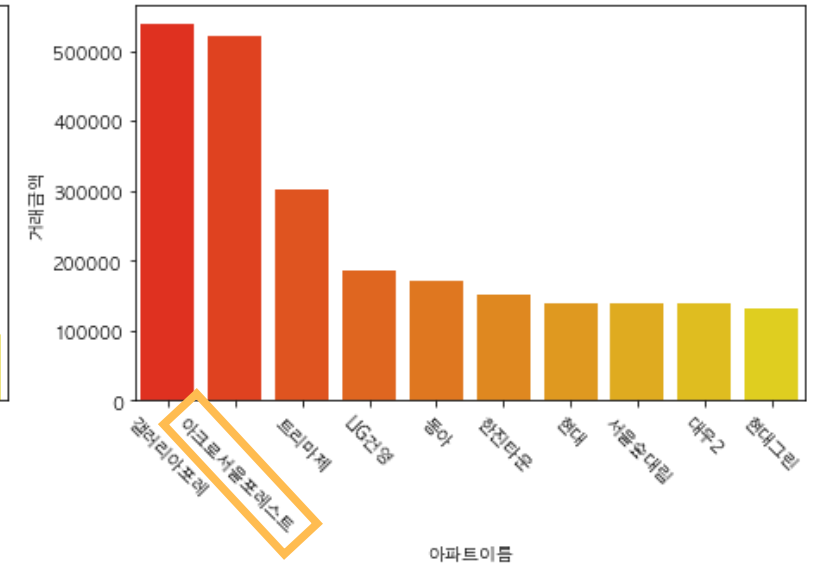
2019년 대비 2021년 성동구 동 평균 실거래가 증가율 top5



2019년 성수동1가 아파트 평균 실거래가



2021년 성수동1가 아파트 평균 실거래가



법정동	증감액	증가율(%)
성수동1가	111,291.02	89.18
금호동1가	64,644.90	79.89
홍익동	31,535.0	59.54
행당동	45,101.26	51.93
마장동	32,156.55	49.29

Why?

데이터를 통해 파악한 원인

- '아크로 서울 포레스트'의 등장
해당 건축물이 20년 말에 사용 승인이 나면서 21년 들어 매매가 활성화 됨
약 50억 이상의 매매가를 기록하며 성수동1가의 평균 실거래가 상승의 주 원인이 됨

자료 조사를 통해 파악한 원인

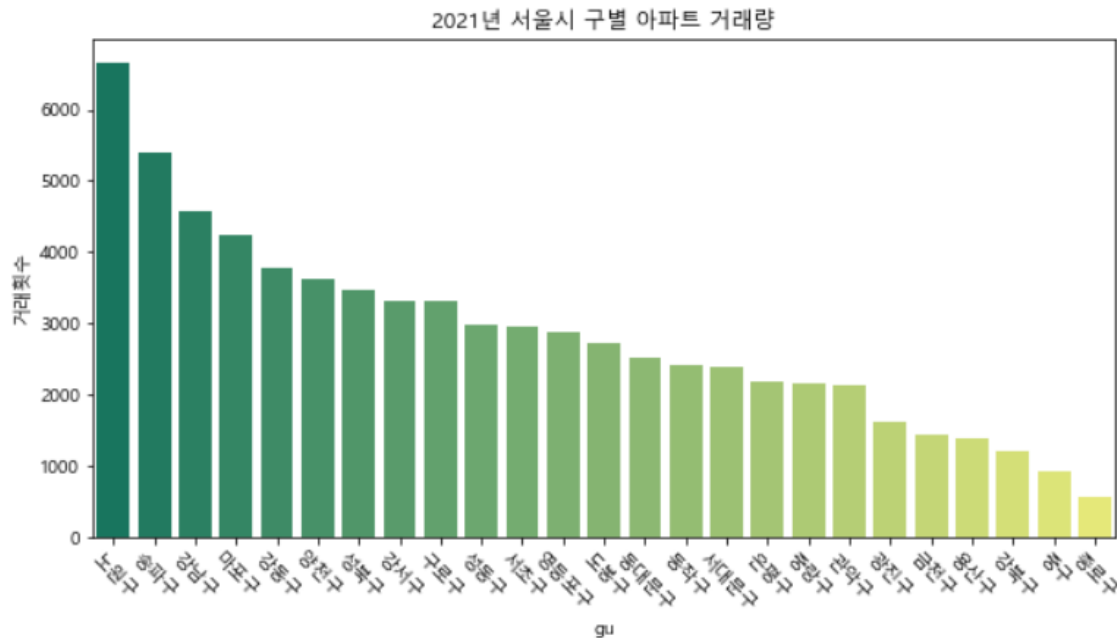
- 19년과 21년 모두 1위를 차지하고 있는 **갤러리아 포레**의 경우 근처 노선에 재개발을 앞두고 있어 이에 따른 부동산 가격 상승에 대한 기대감이 반영된 것으로 보임
- 또한, 강남권의 아파트 노후화에 따라 **강남권 접근성이 좋은 성수동 1가 지역의 고급 아파트 단지**에 대한 부유층의 선호 증가

TITLE 02

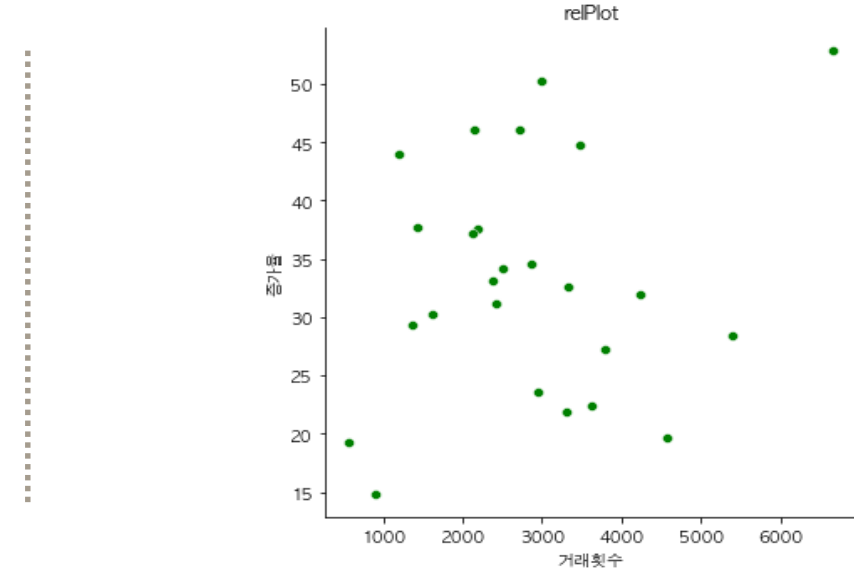
● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

노원구의 실거래가 하위권 탈출 원인 분석



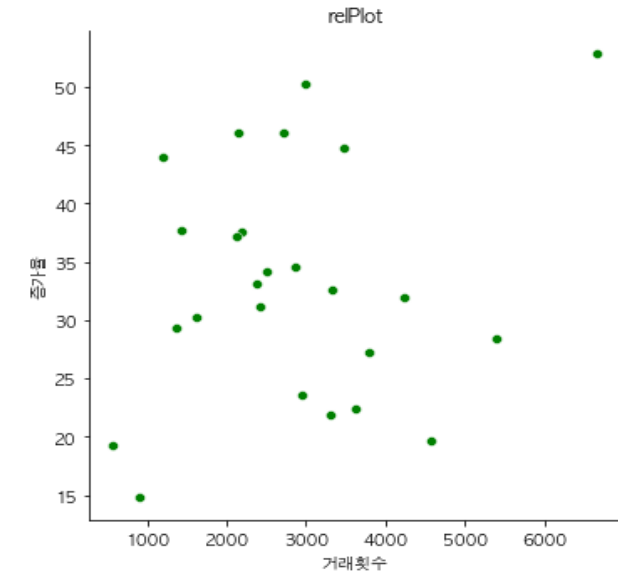
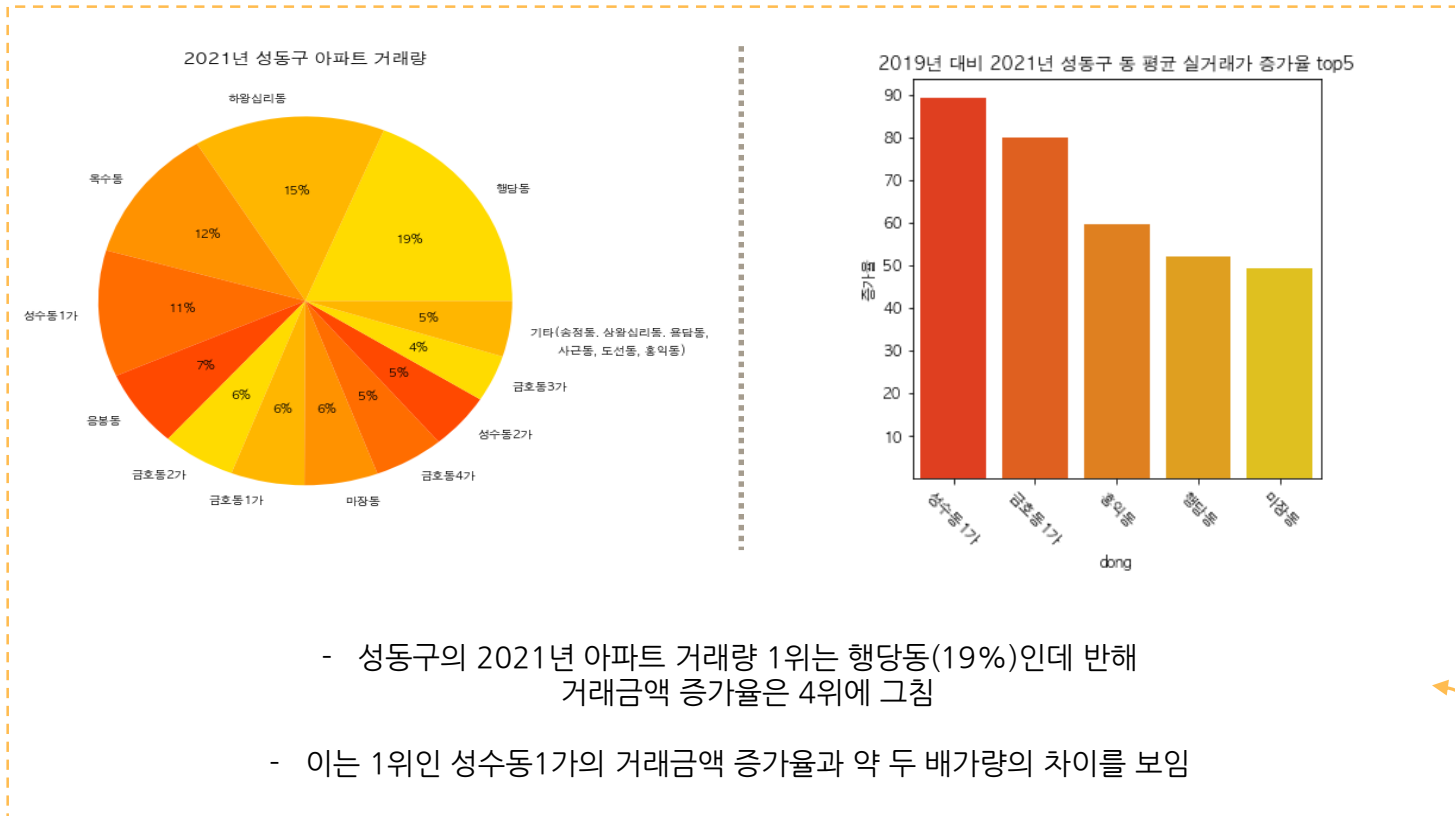
- 노원구는 2021년 서울 전체 지역들 중 아파트 거래량 1위를 차지
- 노원구의 경우 거래량과 거래가 증가율 모두 1위인 지역이기 때문에 거래량과 증가율 사이의 관계가 있는지 알아보기 위한 시각화 작업을 수행



- 거래량과 실거래 증가율 사이의 선형관계가 나타나지 않음

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

노원구의 실거래가 하위권 탈출 원인 분석



- 거래량과 실거래 증가율 사이의 선형관계가 나타나지 않음

TITLE 02

● KB국민은행 IT's your Life

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

노원구의 실거래가 하위권 탈출 원인 분석

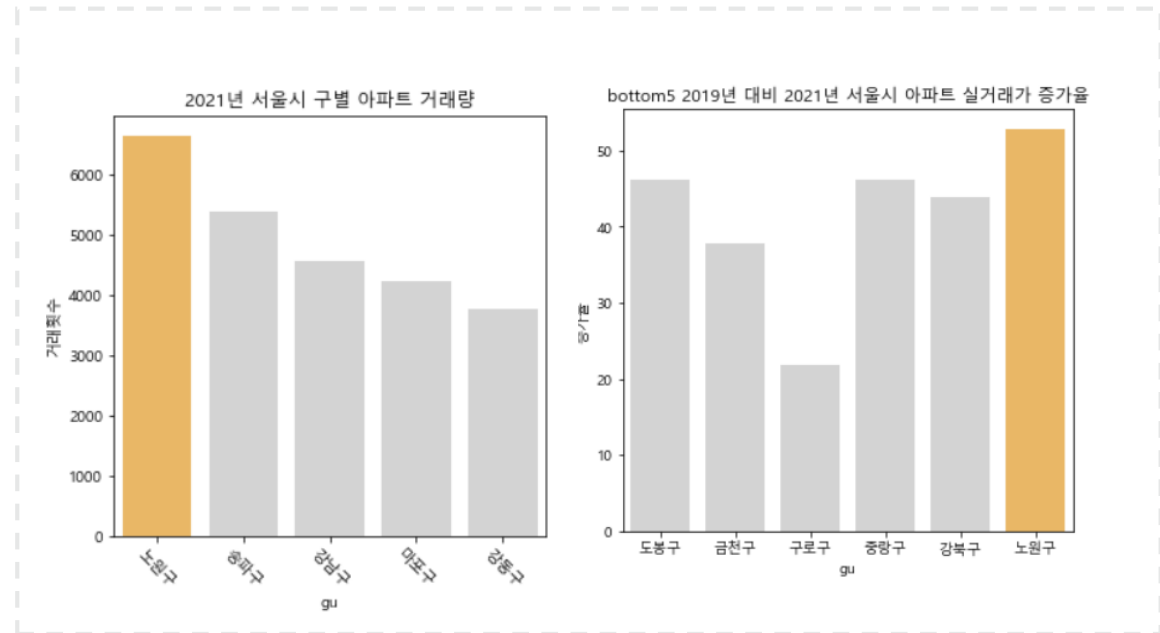
서울시 전체에 대한 상관관계수

	거래량	증가율
거래량	1.0000000	0.185698
증가율	0.185698	1.0000000



노원구에 대한 상관관계수

	거래량	증가율
거래량	1.0000000	0.738369
증가율	0.738369	1.0000000



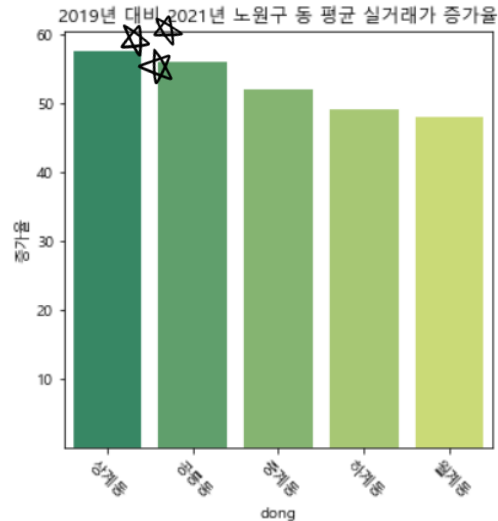
- 노원구의 경우 거래량도 최상위(6,658건)이고 실거래가 증가율(52.89%)도 이에 비례하여 가장 높음
- 따라서, 거래량 1위, 실거래가 증가율 1위의 근거를 데이터를 통해 조금 더 자세히 살펴보고자 노원구의 어느 동에서 거래량이 가장 많았는지 그래프를 통해 파악

TITLE 02

● KB국민은행 IT's your Life

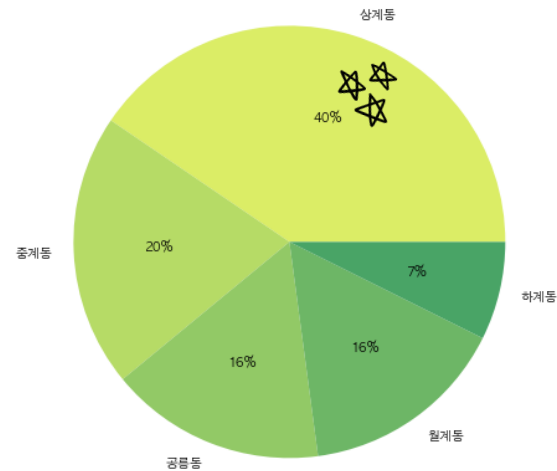
데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

노원구의 실거래가 하위권 탈출 원인 파악



법정동	증감액	증가율(%)
상계동	23,745.88	57.61
공릉동	21,749.68	56.00
중계동	25,981.55	52.03
하계동	23,429.12	49.20
월계동	21,720.50	47.99

2021년 노원구 아파트 거래량

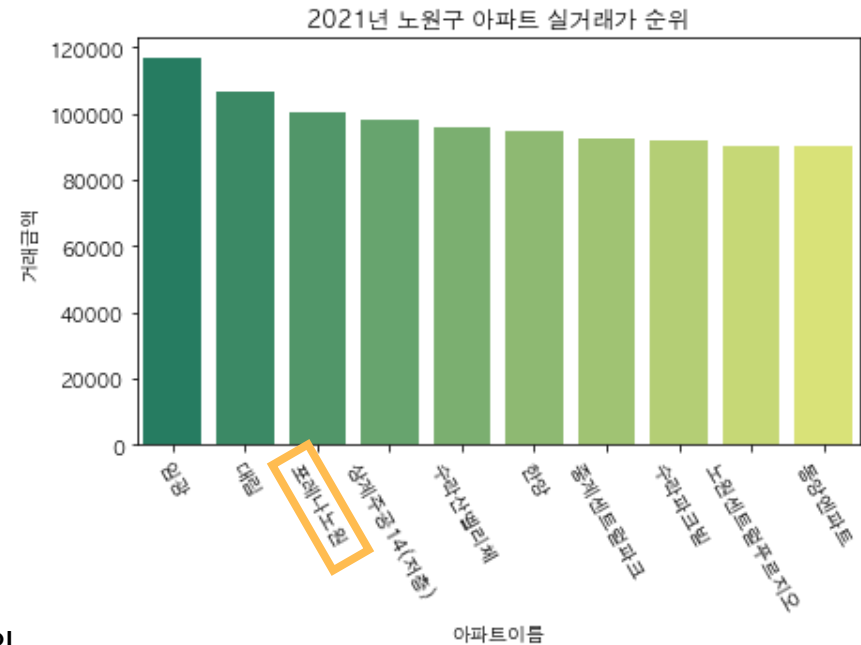
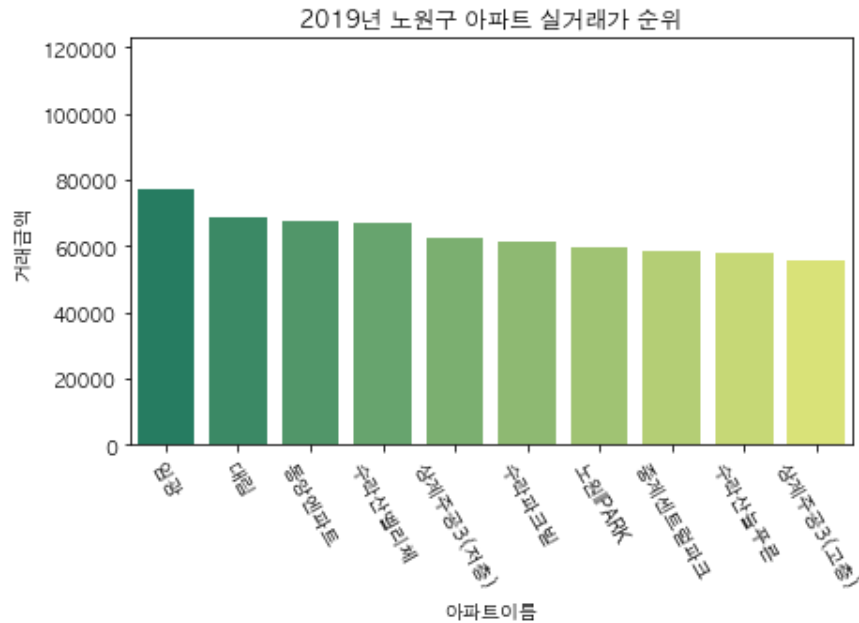


법정동	거래량
상계동	1609
중계동	813
공릉동	642
월계동	617
하계동	292

- 노원구의 실거래가 증가율과 21년의 거래량을 그래프를 통해 파악
- 두 경우 모두 상계동이 1위를 차지
- 특히, 아파트 거래의 경우 노원구 전체에서 약 40%가 상계동에서 발생함을 알 수 있음
- 이에 따라 상계동에 위치한 아파트에 대한 분석을 수행

데이터 분석 & 시각화 ANALYSIS & VISUALIZATION

노원구의 실거래가 하위권 탈출 원인 분석



데이터를 통해 파악한 원인

- 전반적으로 모든 아파트 가격이 상승
- 1,2위의 순위는 변함이 없지만 3위에 19년에 없던 '포레나 노원'이 등장
- 20년 말에 사용 승인 및 매매가 이뤄지면서 21년의 실거래가 상승에 영향을 줌

자료 조사를 통해 파악한 원인

- 포레나 노원의 경우 상계동에서 거의 유일한 신축 아파트로 용적률 문제로 인해 재건축 허가가 어려운 환경이기 때문에 입지에 비해 높은 가격이 형성되었음
- 포레나 노원 건축 이후, 재개발에 대한 기대감으로 인해 해당 지역의 거래가 상승

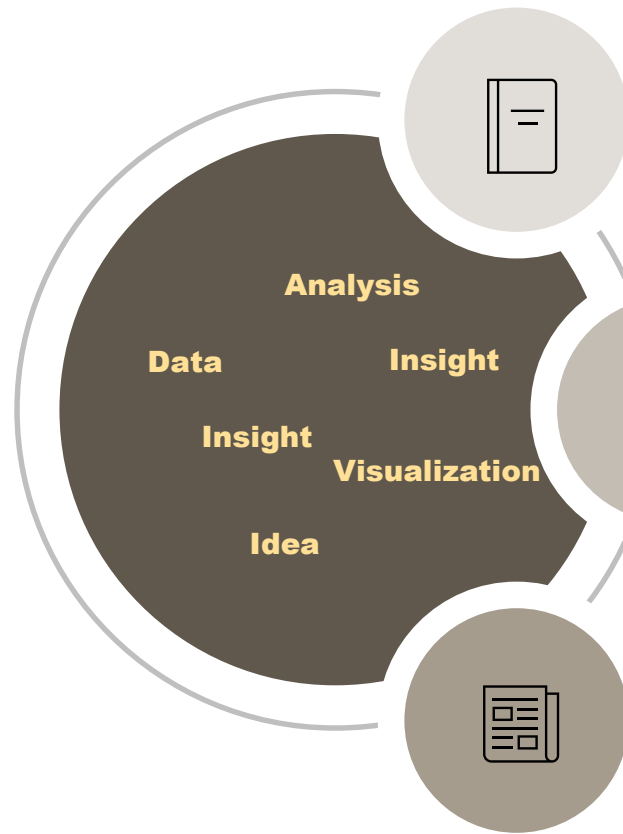
TITLE 03

결론 및 제언

CONCLUSION

TITLE 03

결론 및 제언 CONCLUSION



상/하위 지역의 실거래가 증가 이유

같은 서울시 내에서도 아파트 실거래가의 차이는 큰 편이었다
하지만 거래가격이 큰 지역과 적은 지역을 한 곳 씩 선정해 분석을
수행한 결과, 거래 가격의 증가 원인 자체는 비슷함을 알 수 있었다

아파트 가격과 같은 데이터를 다룰 때?

다른 지역에 비해 이상치라고 여길 정도로 높은 가격의 강남구, 거래량과
거래가격의 증가율이 상관이 없던 성동구, 이와 달리 상관이 있던 노원구...
이처럼 큰 범위에서 같은 지역이라 할지라도 서로 다른 특성을 보이는
데이터를 다룰 땐 다양한 관점에서 많은 시도가 필요함을 깨달았다

데이터 분석 결과의 해석?

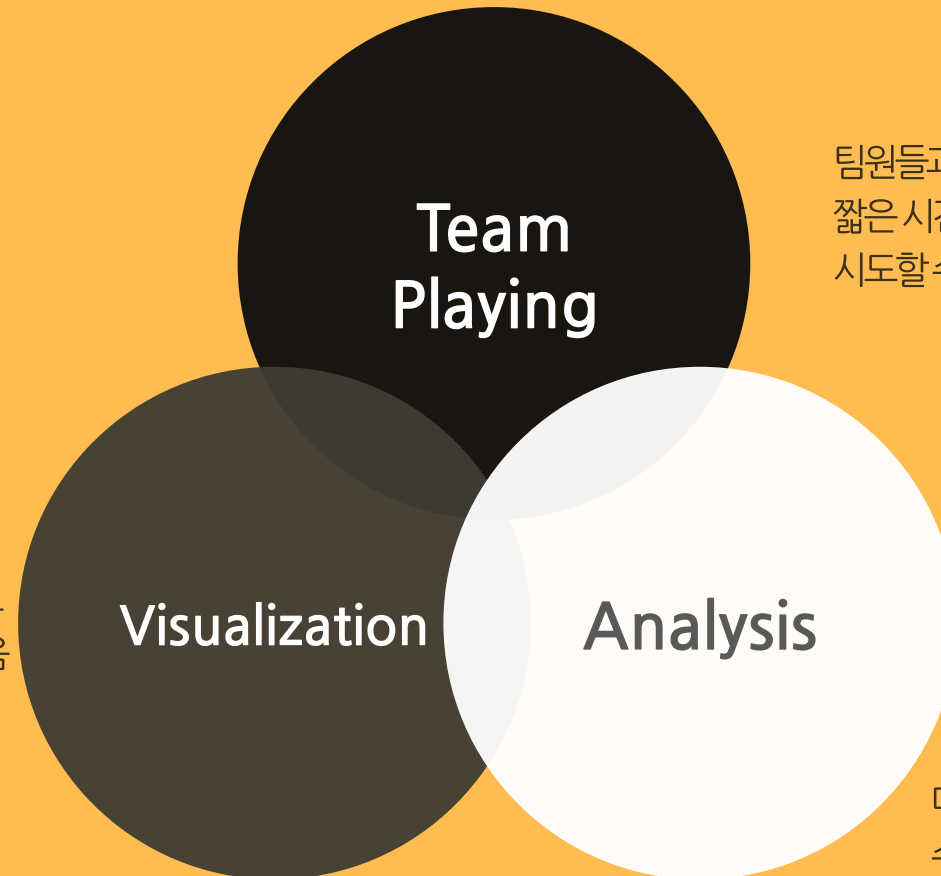
- 분석결과를 바탕으로 실제 그 지역의 현황을 함께 파악하는 과정을 통해
데이터를 좀 더 폭 넓게 이해할 수 있었다

TITLE 03

결론 및 제언 CONCLUSION

GOOD!

다양한 시각화 사례를 직접 만들어 볼 수 있었음
한 데이터에 대해 다양한 시각화 방법을 적용하고
최선의 방법 모색하는 과정을 통해 배울 수 있었음



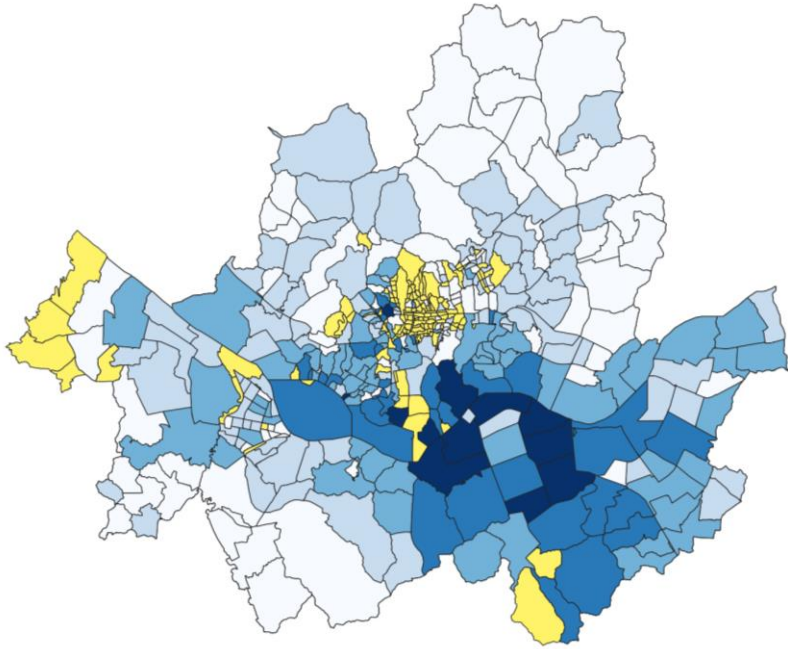
팀원들과 아이디어를 공유하면서
짧은 시간에 더 많은 분석과 시각화를
시도할 수 있었음

데이터에 대한 이해를 바탕으로 전처리 전략을
수립하고 분석을 수행해 체계적인 분석 방법론을
적용해볼 수 있었음

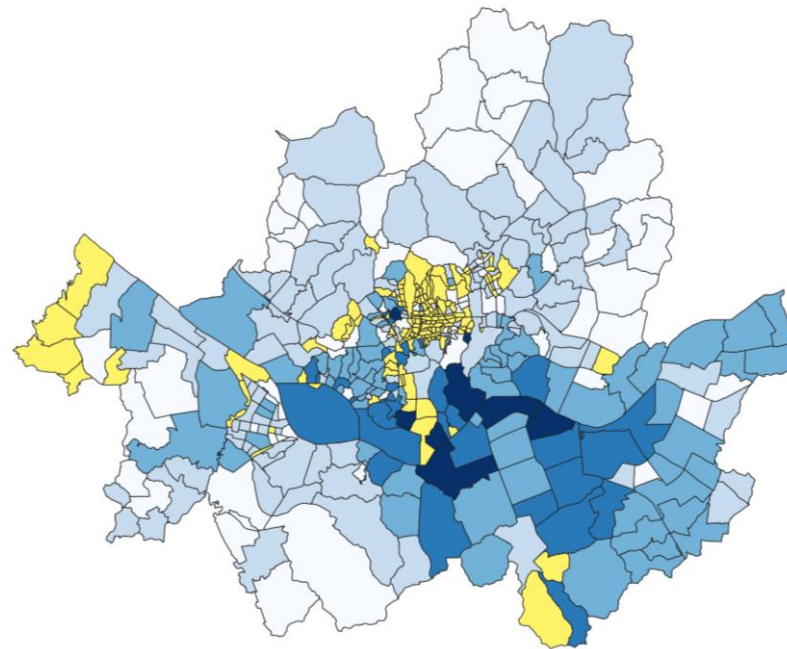
TITLE 03

결론 및 제언 CONCLUSION

2019년의 평균 실거래가



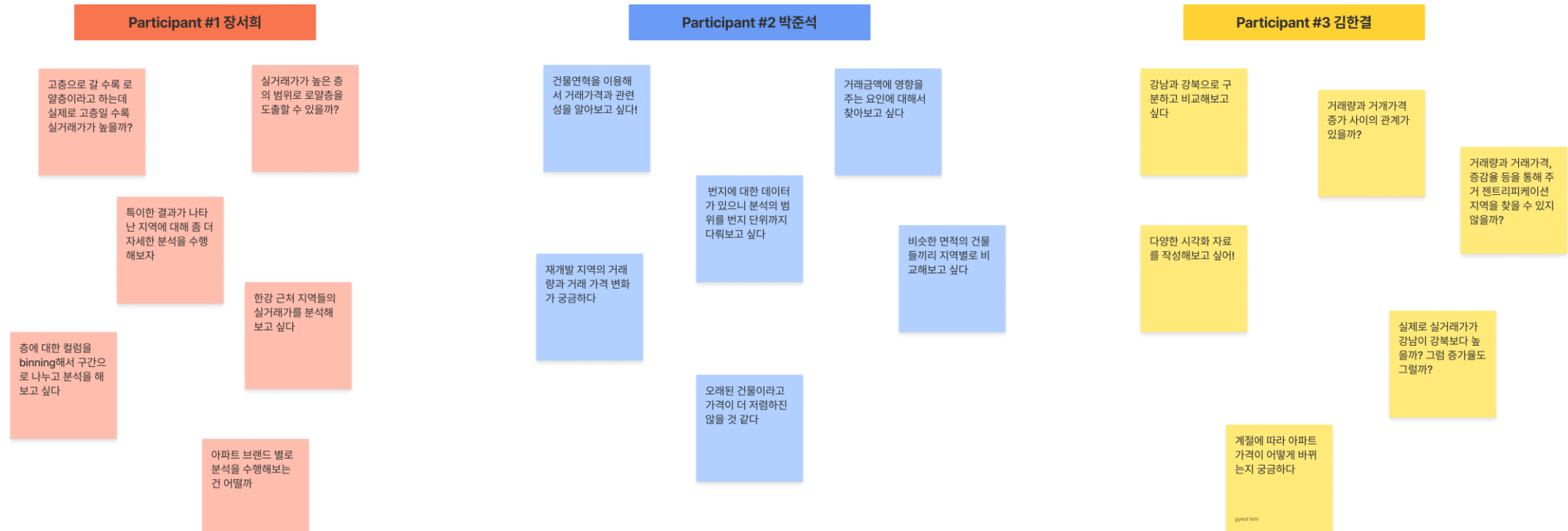
2021년의 평균 실거래가



아쉬운 점 1_ 지도를 이용한 시각화를 제대로 해보지 못한 점이 아쉽다
지도 시각화를 위해 해당 분석엔 사용하지 않았던 법정동 코드를 남겨 놓았지만
제대로 활용하지 못했고, 간단하게 시도해봤지만 일부 지역에 대한 누락이 발생한 것을 해결하지 못함

TITLE 03

결론 및 제언 CONCLUSION



아쉬운 점 2_ 범주화가 필요한 데이터 전처리
여러 아이디어들 중 일부 컬럼을 범주화해서 사용해자는 의견이 있었음
하지만, 이 부분에 대한 이해 부족으로 시도해 보지 못해서 아쉬움이 남음

Ex) 로얄층 분석, 건축년도에 따른 거래 금액 분석 등



김한결

이 멤버, 리멤버...



박준석(25)

좋은 사람과 좋은 시간...



장서희

다른 조 가도... 나 잊지마...

감사합니다.

7조_김한결_박준석_장서희