

합리적인 일본 여행을 위한 도쿄 숙소 데이터 분석

2022. 12. 08



목차

1

데이터 분석 목적

2

데이터 수집

3

데이터 가공

4

데이터 분석

5

분석 결과

1. 데이터 분석 목적

❖ 분석 목적

- 일본 여행 시 가격, 평점 등 다양한 사항을 고려하여 가장 합리적인 호텔을 선정하고자 함.
- 다음의 가설을 세워 데이터 분석 후 결과를 확인함.
- ✓ 12월, 1월이 연말연시이기 때문에 성수기라서 숙소가격이 2월보다 높을 것이다.
 - 성수기에 가장 합리적인 숙소는 어디인가?
- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다.
 - 도시 중심부에서 가까울수록 좋은 호텔인가?
- 가장 합리적인 숙소는 어디인가?

2. 데이터 수집

❖ 데이터 수집 사이트: HostelWorld

- 가장 큰 온라인 호텔 예약 사이트인 HostelWorld에서 데이터를 수집함.

city check in ~ check out

← **HOSTELWORLD**

Tokyo, Japan 7 12월 - 10 12월 2

Tokyo, Japan Showing 29 properties

Sort Filters Map

hostel_name

UNPLAN Shinjuku

8.6 Fabulous (655) num_reviews

Hostel - 1.8km from city centre

distance

room_charges

Privates From ₩517,780 Dorms From ₩52,737

Quick Filters

- ☐ Best Flexible Rates. View hostels offering great rates now, with full flexibility if your plans change.
- ☐ View only properties with promotions
- ☐ View only properties that follow Covid-19 Safety guidelines

rating

CITAN Hostel

9.7 Superb (939)

Hostel - 8.1km from city centre

Privates From ₩207,112 No Dorms Available

2. 데이터 수집

❖ 데이터 수집 사이트: HostelWorld

- hostel_name: 숙소 이름
- checkin_date: 입실 날짜
- checkout_date: 퇴실 날짜
- city: 숙소가 위치한 도시(Tokyo)
- distance: 도시중심부로부터 숙소까지의 거리
- room_type: Private(개인실) or Dorms(공용실)
- room_charge: 1박 요금
- num_reviews: 리뷰 수
- review_total_score: 숙소 평점

	hostel_name	checkin_date	checkout_date	city	distance	room_type	room_charge	num_reviews	review_total_score
0	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	Hostel - 9.8km from city centre	PRIVATE	₩ 117363	2 Total Reviews	10
1	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	Hostel - 9.8km from city centre	DORMS	₩ 45306	2 Total Reviews	10
2	Guest House Shinagawa-shuku Tokyo	2022-12-01	2022-12-02	Tokyo	Hostel - 8.7km from city centre	PRIVATE	₩ 123692	324 Total Reviews	10
3	Guest House Shinagawa-shuku Tokyo	2022-12-01	2022-12-02	Tokyo	Hostel - 8.7km from city centre	DORMS	₩ 41231	324 Total Reviews	10
4	CITAN Hostel	2022-12-01	2022-12-02	Tokyo	Hostel - 8.1km from city centre	PRIVATE	₩ 198482	935 Total Reviews	9.7

3. 데이터 가공

1. room_charge = e 인 데이터 삭제

2. num_reviews 숫자만 남기기

3. num_reveiws = 0인 데이터 삭제

4. review_total_score = No인 데이터 삭제

5. 주말, 평일정보 추가

6. distance 숫자만 남기기

7. room_charge 숫자만 남기기 (\ 삭제)

8. room_type 숫자로 표시 (Private=0, dorms=1)

	hostel_name	checkin_date	checkout_date	city	distance	room_type	room_charge	num_reviews	review_total_score
0	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	Hostel - 9.8km from city centre	PRIVATE	¥ 117363	2 Total Reviews	10
1	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	Hostel - 9.8km from city centre	DORMS	¥ 45306	2 Total Reviews	10
2	Guest House Shinagawa-shuku Tokyo	2022-12-01	2022-12-02	Tokyo	Hostel - 8.7km from city centre	PRIVATE	¥ 123692	324 Total Reviews	10
1496	Top Edge Hostel Koenji	2023-02-27	2023-02-28	Tokyo	Hostel - 4.2km from city centre	PRIVATE	e	12 Total Reviews	NO

3. 데이터 가공

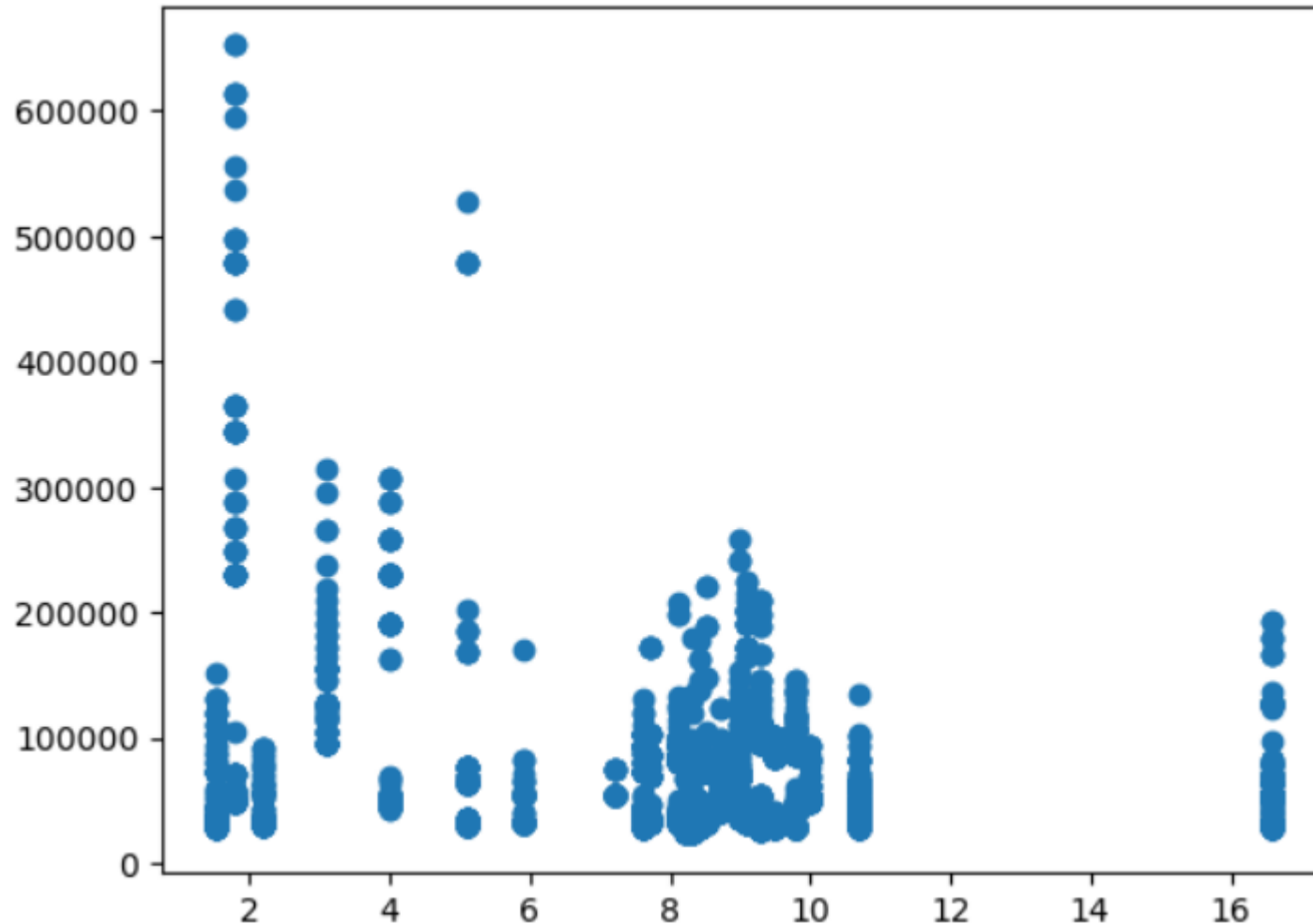
가공 전 4038개의 데이터 -> 가공 후 2888개의 데이터
총 1150개의 데이터 손실

	hostel_name	checkin_date	checkout_date	city	distance	room_type	room_charge	num_reviews	review_total_score	weekends
0	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	9.8	0	117363	2	10	0
1	Time Sharing Stay Asakusa	2022-12-01	2022-12-02	Tokyo	9.8	1	45306	2	10	0
2	Guest House Shinagawa-shuku Tokyo	2022-12-01	2022-12-02	Tokyo	8.7	0	123692	324	10	0
3	Guest House Shinagawa-shuku Tokyo	2022-12-01	2022-12-02	Tokyo	8.7	1	41231	324	10	0
4	CITAN Hostel	2022-12-01	2022-12-02	Tokyo	8.1	0	198482	935	9.7	0
...
2883	&AND HOSTEL MINAMISENJU	2023-02-27	2023-02-28	Tokyo	10.7	1	36910	5	8.0	0
2884	Little Japan	2023-02-27	2023-02-28	Tokyo	8.5	0	189823	66	7.6	0
2885	Little Japan	2023-02-27	2023-02-28	Tokyo	8.5	1	31637	66	7.6	0
2886	Juyoh Hotel	2023-02-27	2023-02-28	Tokyo	10.7	0	30966	124	7.4	0
2887	Almond hostel & cafe Shibuya	2023-02-27	2023-02-28	Tokyo	2.2	1	29528	254	7.3	0

2888 rows × 10 columns

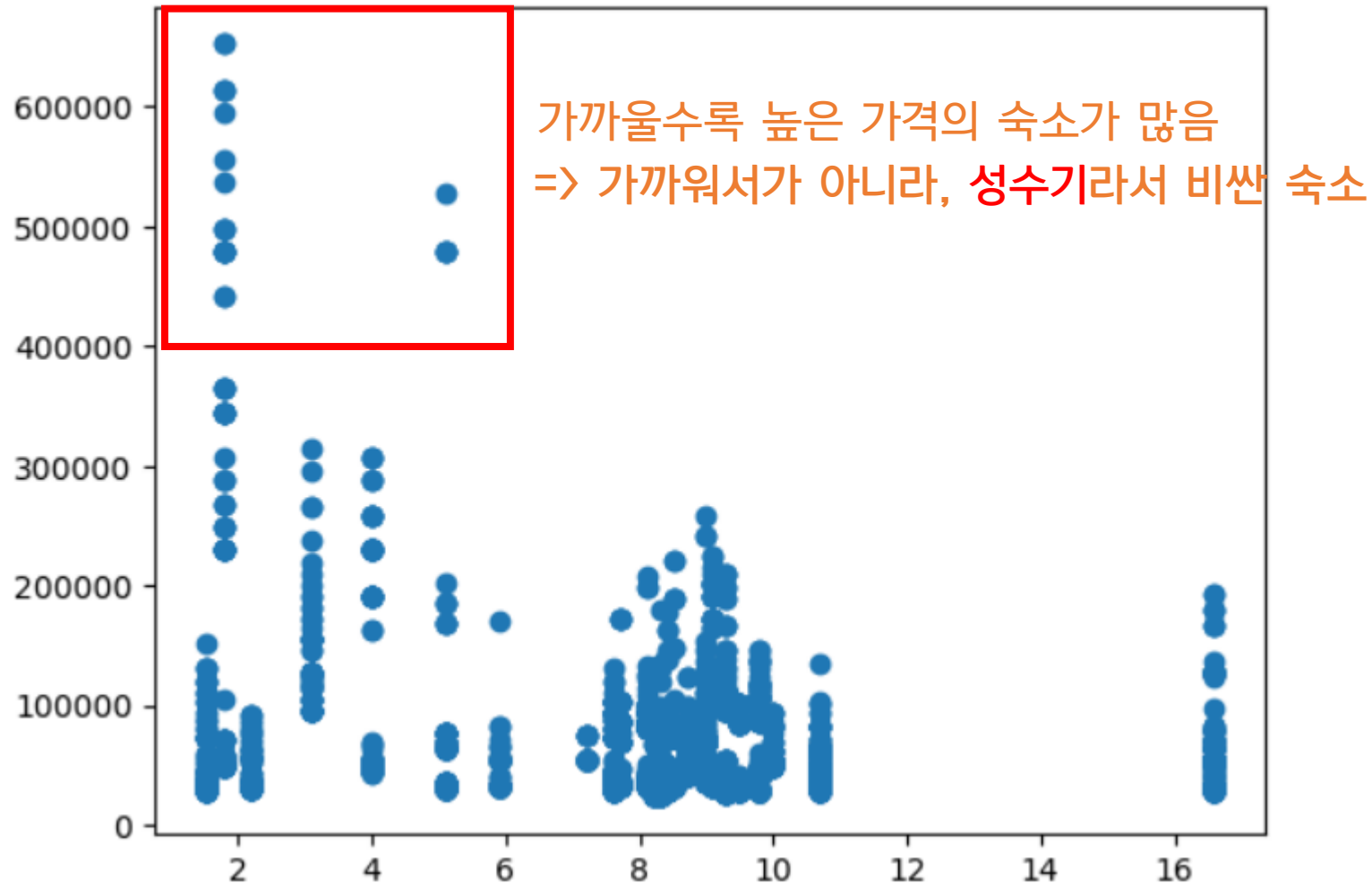
4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다.
- ✓ 12월, 1월이 연말연시이기 때문에 성수기라서 숙소가격이 2월보다 높을 것이다.



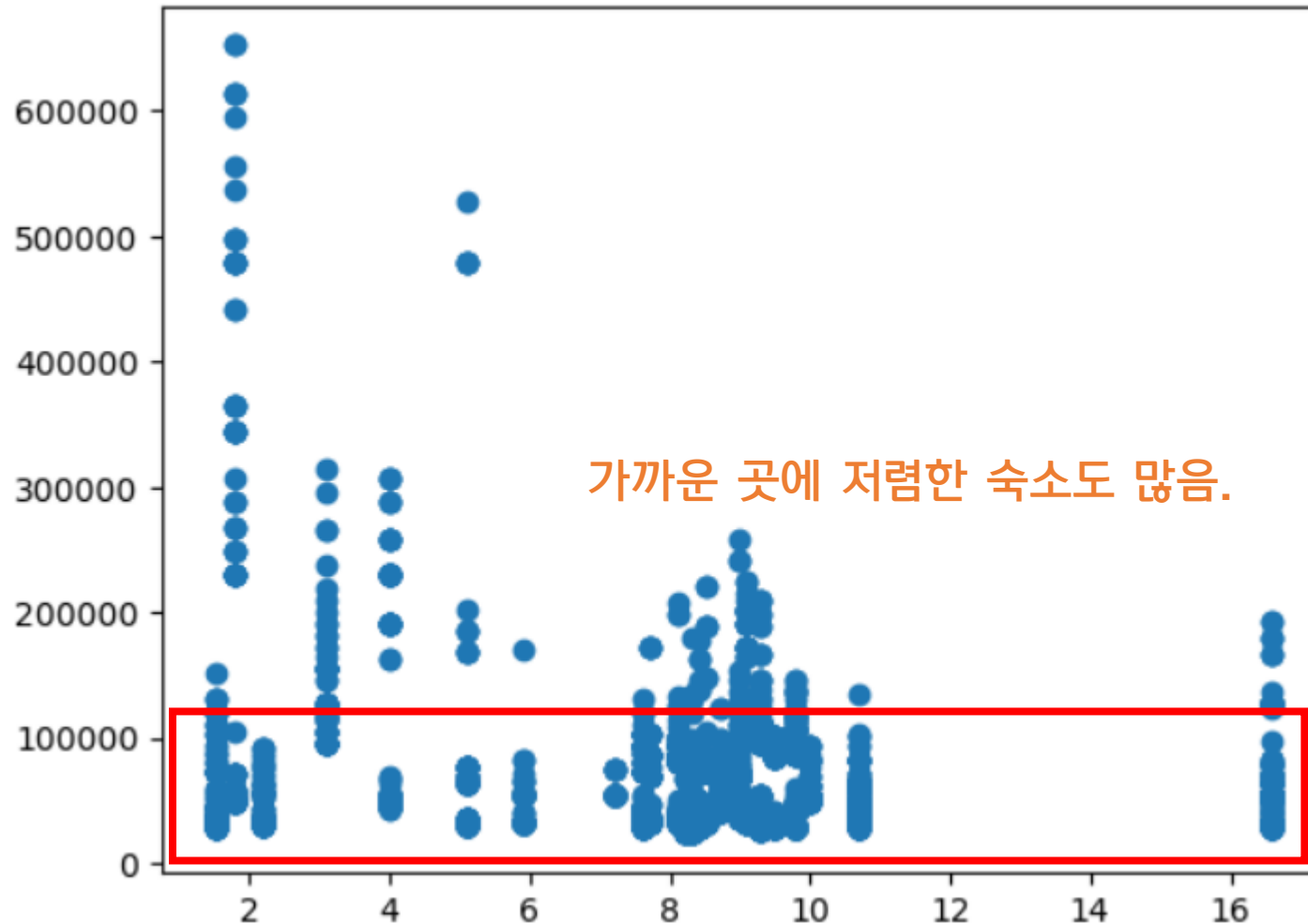
4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 12월, 1월이 연말연시이기 때문에 성수기라서 숙소가격이 2월보다 높을 것이다.



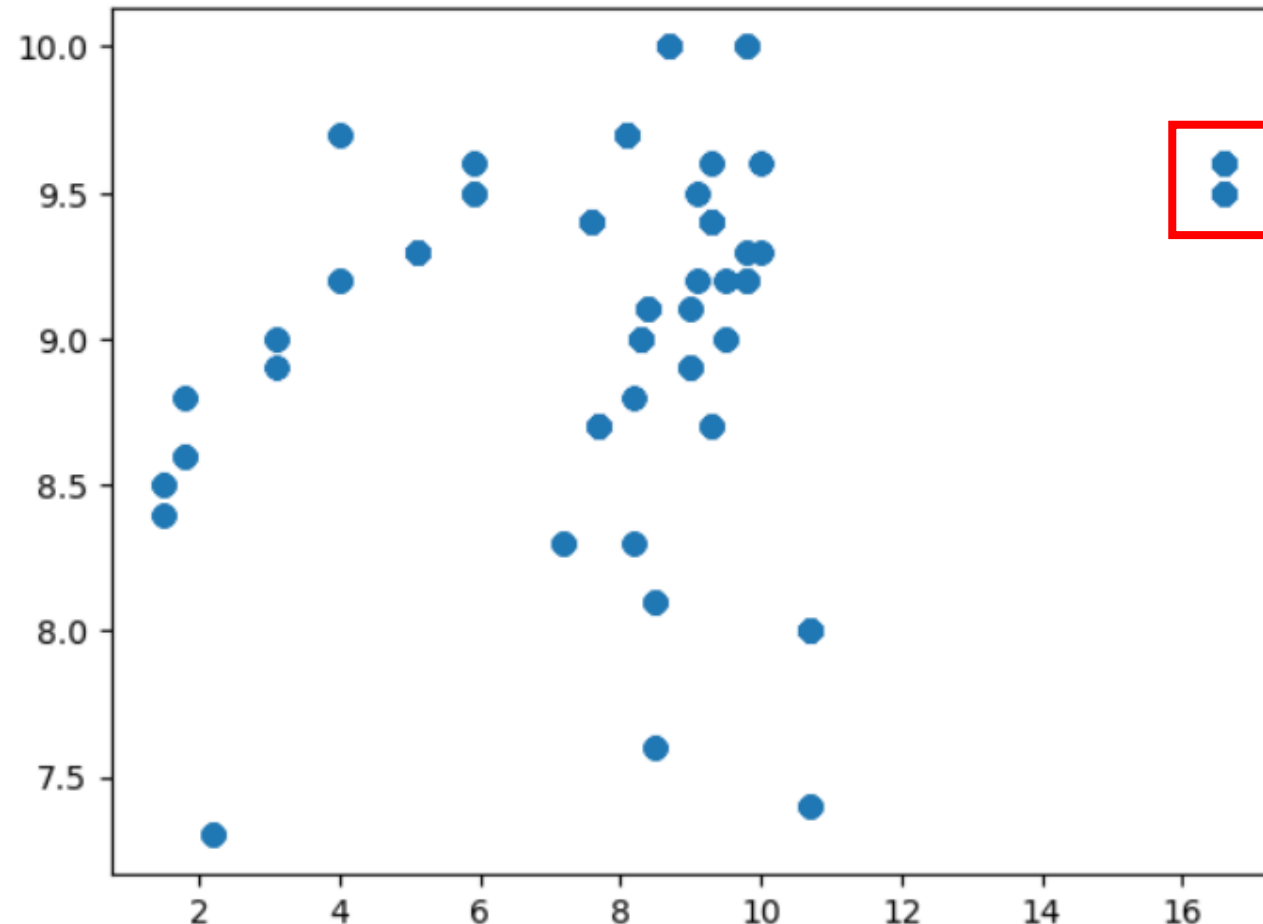
4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 12월, 1월이 연말연시이기 때문에 성수기라서 숙소가격이 2월보다 높을 것이다.



4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

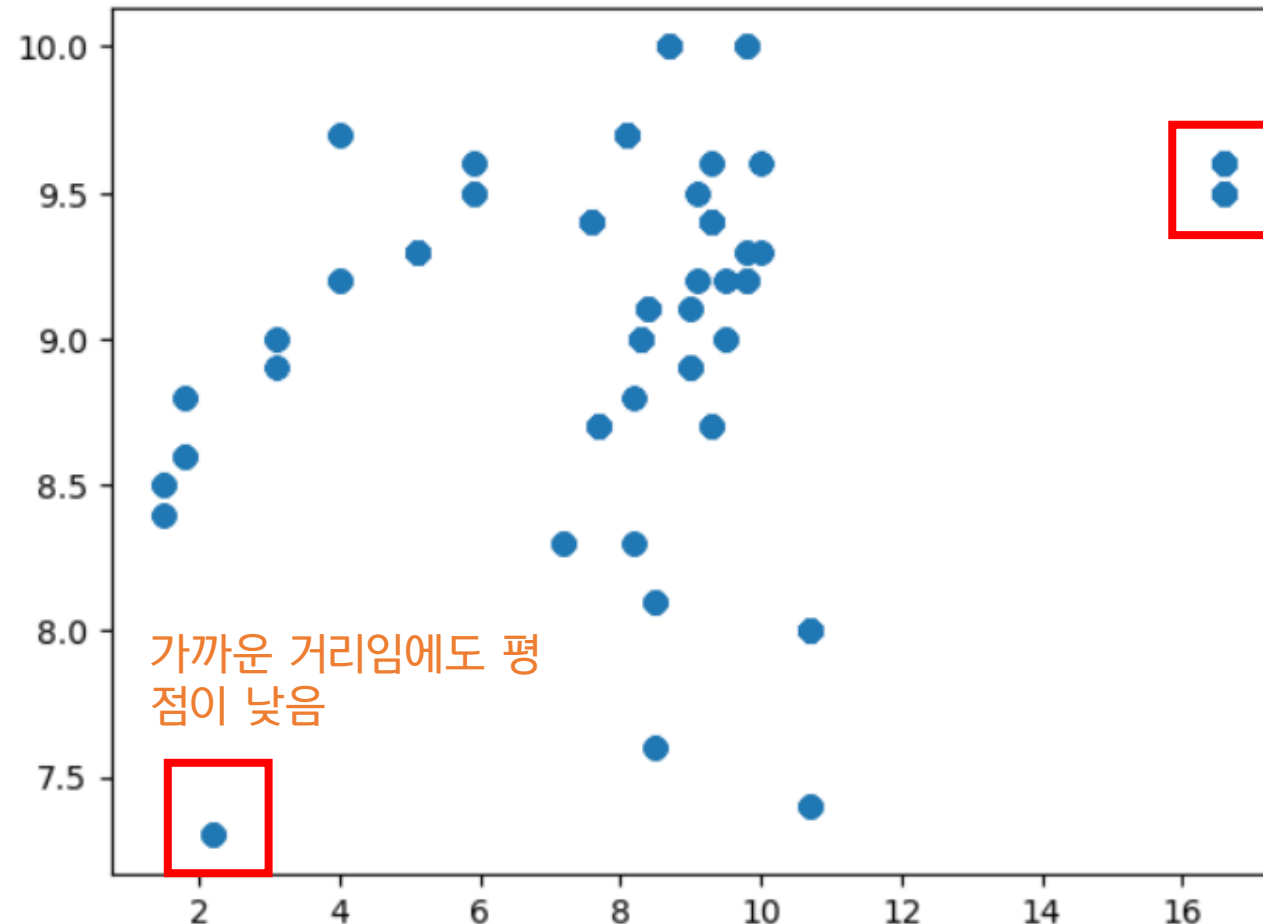
- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 도시 중심부에서 가까울수록 좋은 호텔인가? => 평점이 좋은가? (X)



매우 먼 거리임에도 평점이 높음.

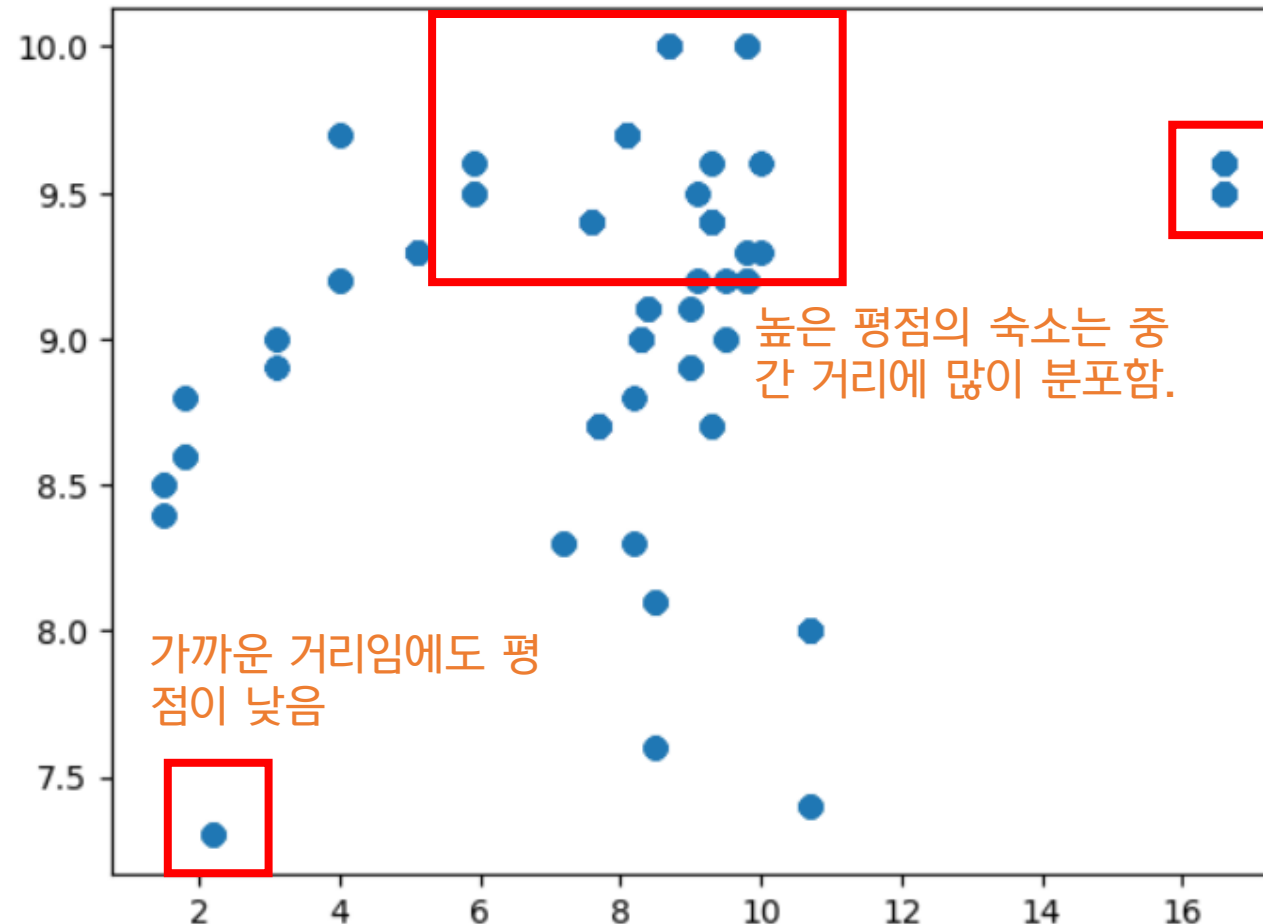
4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 도시 중심부에서 가까울수록 좋은 호텔인가? => 평점이 좋은가? (X)



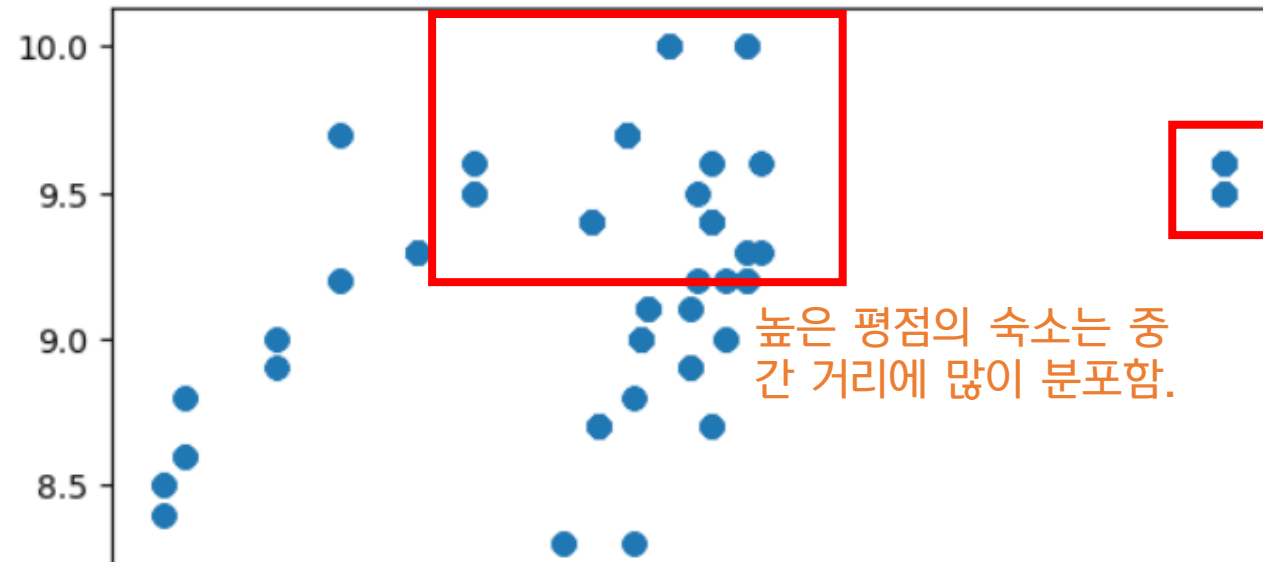
4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 도시 중심부에서 가까울수록 좋은 호텔인가? => 평점이 좋은가? (X)



4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다. (X)
- ✓ 도시 중심부에서 가까울수록 좋은 호텔인가? => 평점이 좋은가? (X)



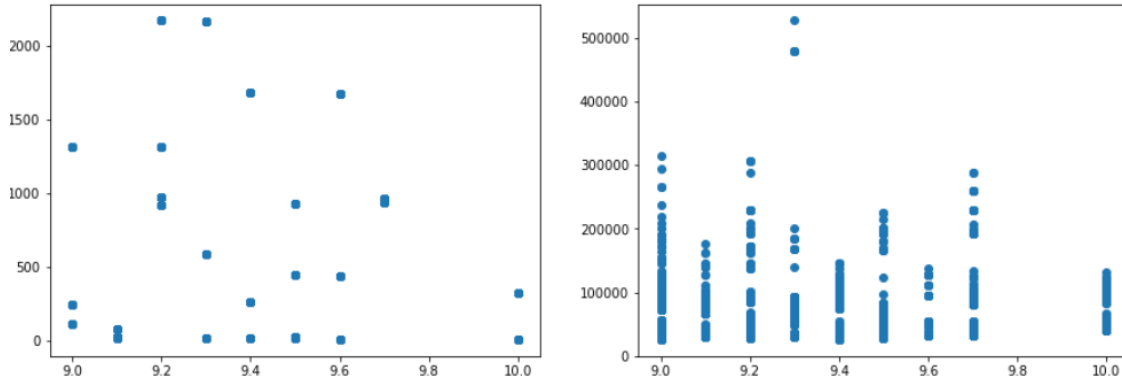
매우 먼 거리임에도 평점이 높음.

높은 평점의 숙소는 중간 거리에 많이 분포함.

가까운 거리에 낮은 평점의 숙소가 많고,
높은 평점의 숙소는 중간 거리에 많음.
=> 평점이 높은 숙소의 특성은 무엇일까?

4. 데이터 분석 - 1) 거리에 따른 숙소 데이터 분석

❖ 평점별 리뷰 수 및 1박 가격 비교

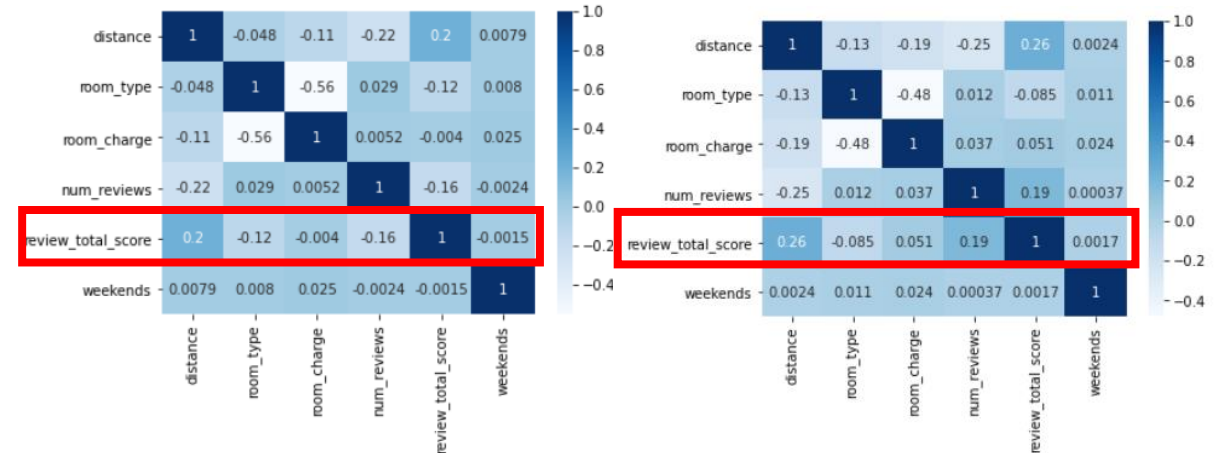


평점별 리뷰 수

평점별 1박 가격

- ✓ 평점별 리뷰 수와 1박 가격을 확인하였으나, 관계성을 찾기 어려움.
- ✓ 10점 만점 평점의 경우 리뷰 수가 적음.
- ✓ 10점 만점 평점의 경우 20만원을 넘지 않음.

❖ 9점 이상 평점 데이터의 correlation map



9점 이상 평점 데이터

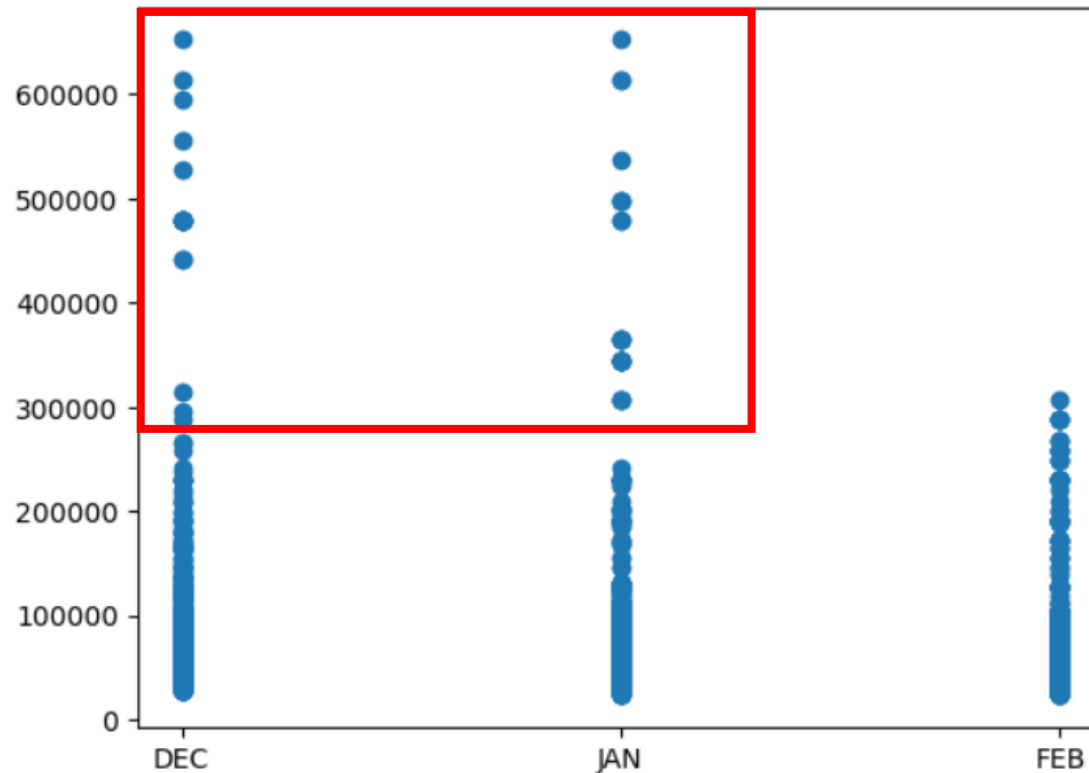
전체 데이터

- ✓ 9점 이상 평점 데이터에서는 distance 외에는 모두 리뷰 평점과 음의 상관관계를 가짐.
 - ✓ 전체 데이터에서는 리뷰 평점과 room_charge, num_reviews, weekends와 양의 상관관계를 가짐.
- => 좋은 평점의 숙소일수록 가격이 낮고, 리뷰 수가 적고, 평일에 예약 가능함(=주말 예약이 완료되었을 것임.)

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

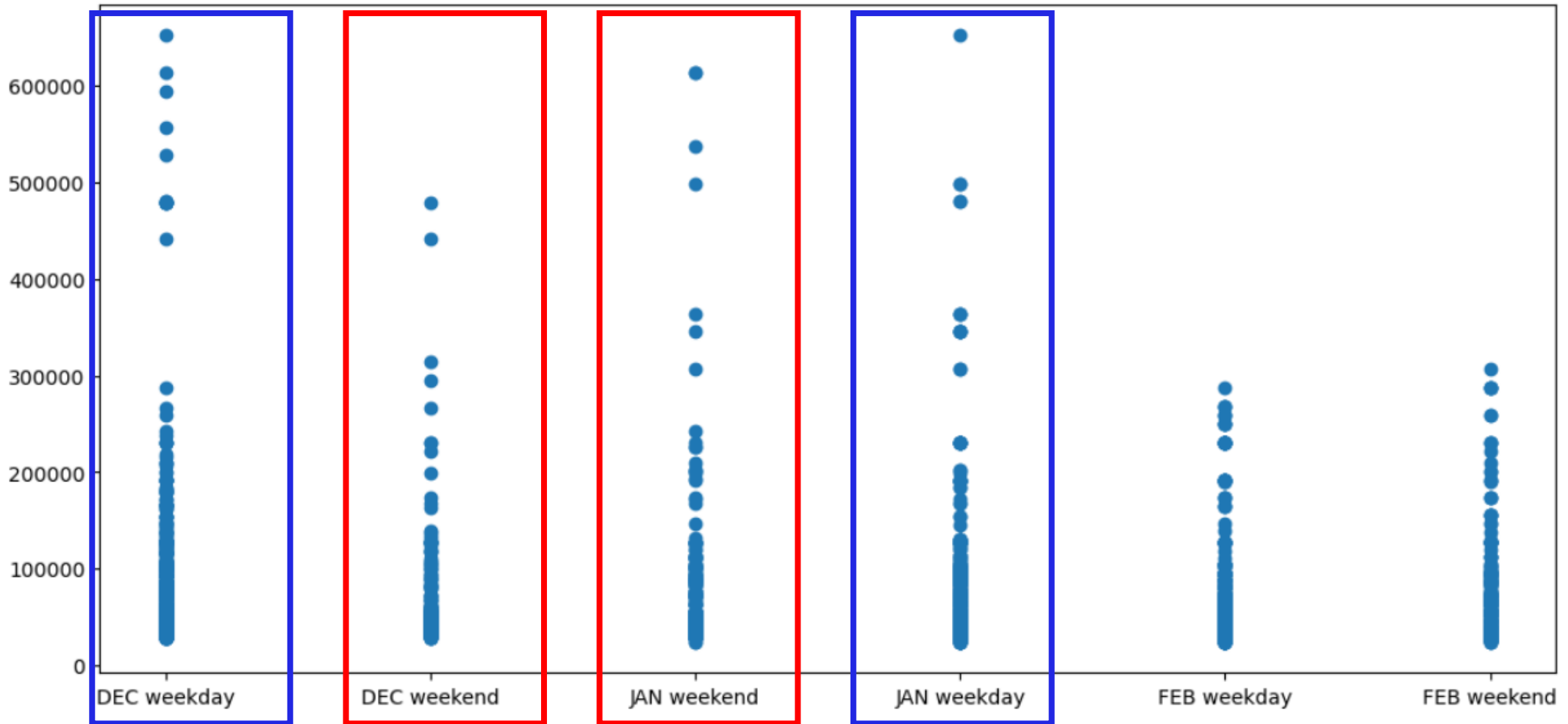
- ✓ 도시 중심부에서 가까울수록 가격이 높을 것이다.
- ✓ 12월, 1월이 연말연시이기 때문에 성수기라서 숙소가격이 2월보다 높을 것이다.

12월과 1월에서 2월보다 높은 가격의 숙소가 있음.



4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

- ✓ 주말, 주중 가격 비교
- ✓ 주말 가격이 더 낮은 것을 확인



4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

2022-12월						
SUN	MON	TUE	WED	THU	FRI	SAT
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

2023-1월						
SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

- ✓ 크리스마스가 있는 12월 4주부터 1월 1주가 가장 비쌘.
- ✓ 크롤링 날부터 가까운 날인 12월 달의 가격이 비쌘.

모두 Private room(개인실), checkin_date 기준

	$x < 40$ 만		$50\text{만} \leq x < 60\text{만}$
	$40\text{만} \leq x < 50\text{만}$		$60\text{만} \leq x$

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

2022-12월

SUN MON TUE WED THU FRI SAT

성수기의 합리적인 숙소를 찾기 위해
12월 4주 ~ 1월 1주까지의 데이터 분석

checkin date 기준
12/18 ~ 1/8 데이터 분석

18	19	20	21	22	23	24
25	26	27	28	29	30	31

2023-1월

SUN MON TUE WED THU FRI SAT

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

- ✓ 크리스마스가 있는 12월 4주부터 1월 1주가 가장 비쌘.
- ✓ 크롤링 날부터 가까운 날인 12월 달의 가격이 비쌘.

모두 Private room(개인실), checkin_date 기준

	$x < 40$ 만		$50\text{만} \leq x < 60\text{만}$
	$40\text{만} \leq x < 50\text{만}$		$60\text{만} \leq x$

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

- ✓ checkin_date 12/18 ~ 1/8 까지의 숙소
- ✓ '합리적인 의미'를 기준으로 하여 **평균 미만의 데이터 삭제**

	distance	room_type	room_charge	num_reviews	review_total_score	weekends
count	646.000000	646.000000	646.000000	646.000000	646.000000	646.000000
mean	8.588235	0.575851	79194.222910	465.970588	9.030031	0.205882
std	3.581631	0.494596	75486.527331	586.955639	0.655351	0.404658
min	1.500000	0.000000	27960.000000	2.000000	7.300000	0.000000
25%	7.600000	0.000000	39696.000000	20.000000	8.700000	0.000000
50%	8.700000	1.000000	53600.000000	254.000000	9.200000	0.000000
75%	9.800000	1.000000	94188.000000	655.000000	9.500000	0.000000
max	16.600000	1.000000	652019.000000	2170.000000	10.000000	1.000000

기준	'합리적인' 특징의 의미
distance	거리가 가까울수록 편리한 숙소
room_charge	가격이 저렴할수록 경제적인 숙소
num_reviews	리뷰가 적을수록 편향되기 쉽고, 리뷰가 많을수록 이용객이 많은 좋은 숙소
review_total_score	평점이 높을수록 합리적인 숙소

- distance 가 8.5km 초과인 숙소 삭제
- room_charge 79,194원 초과인 숙소 삭제
- num_reviews 가 465개 미만인 숙소 삭제
- review_total_score 가 9점 미만인 숙소 삭제

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

✓ 거리, 가격, 리뷰 수, 리뷰 평점 기준의 합리적인 3개의 성수기 숙소 44개 데이터 추출

	hostel_name	checkin_date	checkout_date	city	distance	room_type	room_charge	num_reviews	review_total_score	weekends	month
0	CITAN Hostel	2022-12-18	2022-12-19	Tokyo	8.1	1	37971	935	9.7	1	DEC weekend
1	Sakura Hotel Ikebukuro	2022-12-18	2022-12-19	Tokyo	5.1	0	63514	583	9.3	1	DEC weekend
2	UNPLAN Kagurazaka	2022-12-18	2022-12-19	Tokyo	4.0	1	53696	967	9.2	1	DEC weekend
3	CITAN Hostel	2022-12-19	2022-12-20	Tokyo	8.1	1	36245	935	9.7	0	DEC weekday
4	Sakura Hotel Ikebukuro	2022-12-19	2022-12-20	Tokyo	5.1	0	63514	583	9.3	0	DEC weekday
5	Sakura Hotel Ikebukuro	2022-12-19	2022-12-20	Tokyo	5.1	1	29993	583	9.3	0	DEC weekday
6	UNPLAN Kagurazaka	2022-12-19	2022-12-20	Tokyo	4.0	1	53696	967	9.2	0	DEC weekday
7	CITAN Hostel	2022-12-20	2022-12-21	Tokyo	8.1	1	36245	935	9.7	0	DEC weekday
⋮											
38	CITAN Hostel	2023-01-06	2023-01-07	Tokyo	8.1	1	39696	935	9.7	0	JAN weekday
39	Sakura Hotel Ikebukuro	2023-01-06	2023-01-07	Tokyo	5.1	0	69881	583	9.3	0	JAN weekday
40	Sakura Hotel Ikebukuro	2023-01-06	2023-01-07	Tokyo	5.1	1	32985	583	9.3	0	JAN weekday
41	CITAN Hostel	2023-01-08	2023-01-09	Tokyo	8.1	1	44874	935	9.7	1	JAN weekend
42	Sakura Hotel Ikebukuro	2023-01-08	2023-01-09	Tokyo	5.1	0	76248	583	9.3	1	JAN weekend
43	Sakura Hotel Ikebukuro	2023-01-08	2023-01-09	Tokyo	5.1	1	35976	583	9.3	1	JAN weekend

- 'CITAN Hostel'
- 'UNPLAN Kagurazaka'
- 'Sakura Hotel Ikebukuro'
- 12/25, 12/30, 12/31, 01/02, 01/05, 01/07
숙소 X

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

- ✓ 거리, 가격, 리뷰 수, 리뷰 평점 기준의 합리적인 3개의 성수기 숙소
- ✓ 'CITAN Hostel' 'UNPLAN Kagurazaka' 'Sakura Hotel Ikebukuro'

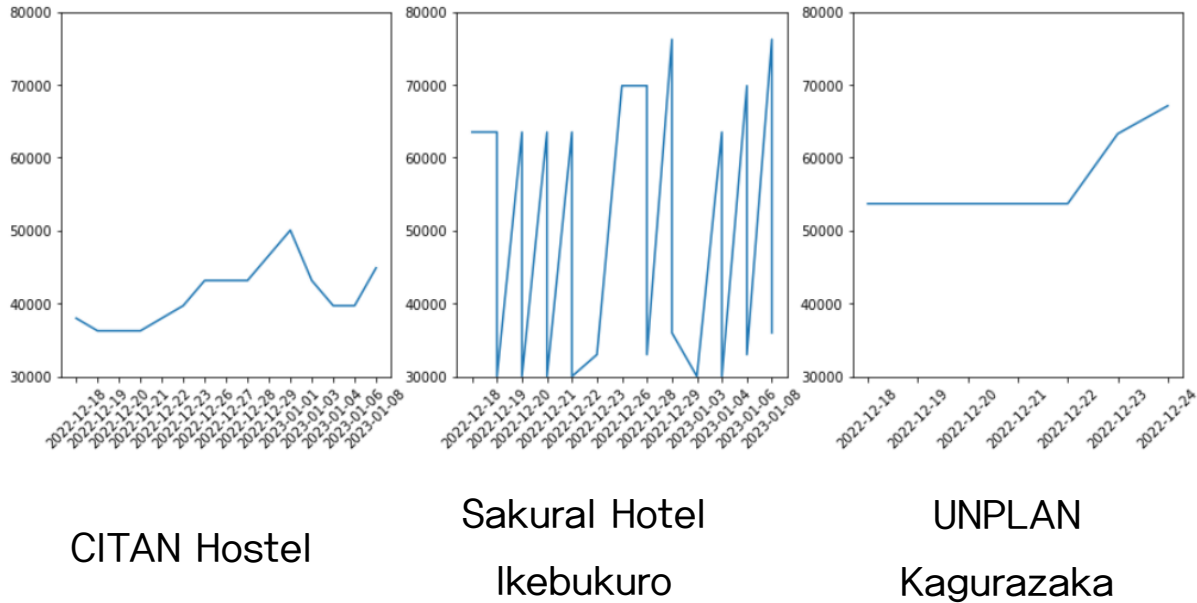
2022-12월						
SUN	MON	TUE	WED	THU	FRI	SAT
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18 Sakura CITAN	19 Sakura CITAN	20 Sakura CITAN	21 Sakura CITAN	22 Sakura CITAN	23 - CITAN	24 - UNPLAN
25 - -	26 Sakura CITAN	27 - CITAN	28 Sakura CITAN	29 Sakura CITAN	30 -	31 -

2023-1월						
SUN	MON	TUE	WED	THU	FRI	SAT
1 - CITAN	2 - -	3 - CITAN	4 Sakura CITAN	5 - -	6 Sakura CITAN	7 - -
8 Sakura CITAN	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Best Private room : Best review_total_score private room
Best Dorms room : Best review_total_score dorms room

4. 데이터 분석 - 2) 성수기 숙소 데이터 분석

❖ Room_charge 비교



- CITAN Hostel 이 가장 저렴함.
- Sakura Hotel Ikebukuro 은 개인실, 공용실이 있어 가격 변동이 심하지만, 공용실이 매우 저렴함.
- UNPLAN Kagurazaka 은 방이 별로 없고, 공용실임에도 가격대가 높음.

❖ 호텔별 조건 비교

	CITAN	Sakura	UNPLAN
distance	8.1	5.1	4.0
num_reviews	935	583	967
review total_score	9.7	9.3	9.2
room_type	only dorms	private, dorms	only dorms

* 위 표는 12/18 ~ 2023-01/08 까지의 숙소 조건임.

* room type 의 경우 이후 기간에는 달라질 수 있음.

- UNPLAN Kagurazaka 가 제일 가까움
- UNPLAN Kagurazaka 의 리뷰가 가장 많음
- CITAN Hostel 의 평점이 가장 높음
- Sakura Hotel Ikebukuro 에만 개인실이 있음.

4. 데이터 분석 - 3) 합리적인 숙소 탐색

❖ 합리적인 숙소 탐색

	distance	room_type	room_charge	num_reviews	review_total_score	weekends
count	2888.000000	2888.000000	2888.000000	2888.000000	2888.000000	2888.000000
mean	8.354294	0.560249	71039.753116	496.829640	9.032618	0.255886
std	3.638733	0.496443	63958.098732	591.370449	0.657141	0.436434
min	1.500000	0.000000	23561.000000	2.000000	7.300000	0.000000
25%	7.200000	0.000000	35970.000000	20.000000	8.700000	0.000000
50%	8.500000	1.000000	52737.000000	262.000000	9.200000	0.000000
75%	9.800000	1.000000	84379.000000	917.000000	9.500000	1.000000
max	16.600000	1.000000	652019.000000	2170.000000	10.000000	1.000000

- distance 가 8.3km 초과인 숙소 삭제
- num_reviews 가 496개 미만인 숙소 삭제
- review_total_score 가 9점 미만인 숙소 삭제



- 'CITAN Hostel'
- 'UNPLAN Kagurazaka'
- 'Sakura Hotel Ikebukuro'

5. 분석 결과

❖ 합리적인 숙소

	CITAN Hostel	Sakura Hotel Ikebukuro	UNPLAN Kagurazaka
distance	8.1	5.1	4.0
num_reviews	935	583	967
review total_score	9.7	9.3	9.2
room_type	private, dorms	private, dorms	private, dorms
room_charge	private: 81,490 ~ 101,878 dorms: 32,788 ~ 50,052	private: 63,504 ~ 479,886 dorms: 29,993 ~ 35,976	private: 191,740 ~ 306,833 dorms: 44,100 ~ 67,120



- ✓ 전체 기간에서 합리적인 숙소가 성수기에도 좋은 숙소로 분석됨.
- ✓ 도시중심부에서 멀어도 여행지에 따라 좋은 숙소가 될 수 있음.
- ✓ 향후 더 많은 도시별 숙소 데이터를 수집하여 분석하면, 계절별 합리적인 숙소를 탐색할 수 있음.
- ✓ 데이터 시각화를 통해 선호 특징별 숙소 파악을 용이하게 할 수 있음.

(가격별 숙소 분포 지도, 리뷰 평점별 숙소 분포 지도 등)

- ✓ 세부 평점(청결 평점, 보안성 평점 등) 데이터를 추가로 수집하여 선호도에 따른 가장 좋은 숙소 탐색 가능



Thank You