



신규 호스트를 위한 숙소 가격 예측 모델

프랑스 파리를 중심으로

활형기

목차

1. 배경

2. 데이터 분석

3. 모델링

2. 결론

1. 배경

1. 배경

Airbnb란?

1. 배경

Airbnb Recovers From Pandemic Dip

Number of nights and experiences booked on Airbnb*



* net of cancellations and alterations

Source: Airbnb



statista

Airbnb Recovers From Pandemic Dip, <https://www.statista.com/>

1. 배경

관광트렌드의 변화



한국관광공사(2022)

1. 배경



The Aribnb Blog(2018)

2. 데이터 분석

2. 데이터 분석 | 개요

● 파리에 숙소를 등록하고자 하는 신규 호스트는 **어떻게** 가격을 책정해야 할까?

가설 1. 지역별 평균 **월임대료**가 높을수록 숙박료가 높을 것이다.

가설 2. Apartment, Hotel room 등 일반적인 숙소보다는 **특별한 종류의 숙소 타입**이 숙박료가 높을 것이다.

2. 데이터 분석 | 개요

1) Target : 예측하고자 하는 값

- Price : 숙박료

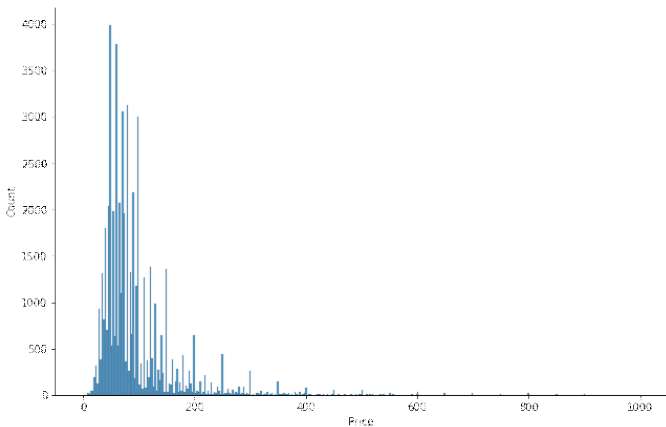
2) Feature : Target을 예측할 때 사용되는 변수

- Neighbourhood : 지역
- Neighbourhood Cleansed : 관광지를 기준으로 나누어진 지역
- Latitude, Longitude : 위도, 경도
- Property Type : 건물 종류
- Room Type : 숙소 종류
- Accomodates : 최대 수용 가능 인원
- Bathrooms/Bedrooms/Beds : 욕실, 침실, 침대의 개수
- Bed Type : 침대 종류
- Amenities : 편의/부대시설
- Security Deposit / Cleaning Fee: 보증금 / 청소비
- Availabilty 30/365: 검색일로부터 30/365일 간 예약 가능한 날짜
- Number of Reviews : 총 리뷰 수
- Review Scores : 각종 리뷰 점수(청결, 위치 등)
- Cancellation Policy: 환불 정책
- Reviews per Month : 월별 리뷰 수

2. 데이터 분석 | 데이터 탐색

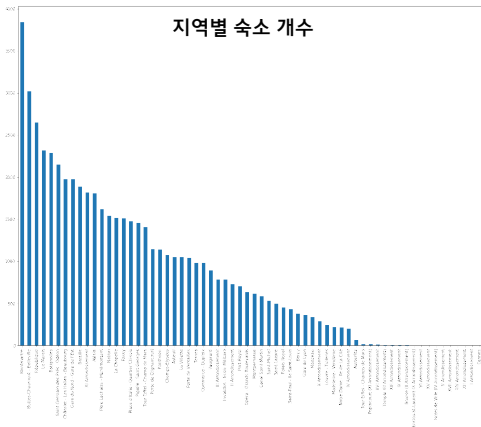
1) 가격의 분포

- 한쪽으로 지나치게 **쏠려있는** 모습을 보임



2. 데이터 분석

데이터 탐색



2) 숙소가 가장 많은 Top 3 지역

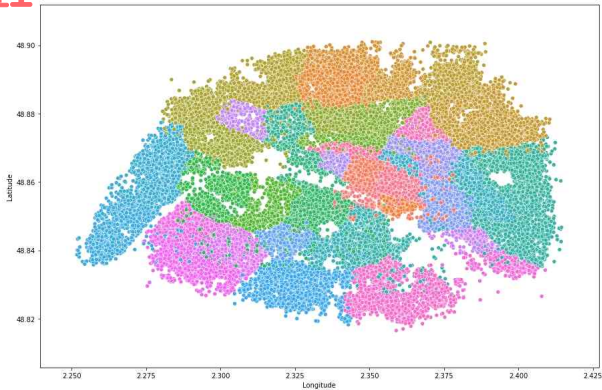
(1) Montmartre

(2) Buttes=Chaumont - Belleville

(3) Republique

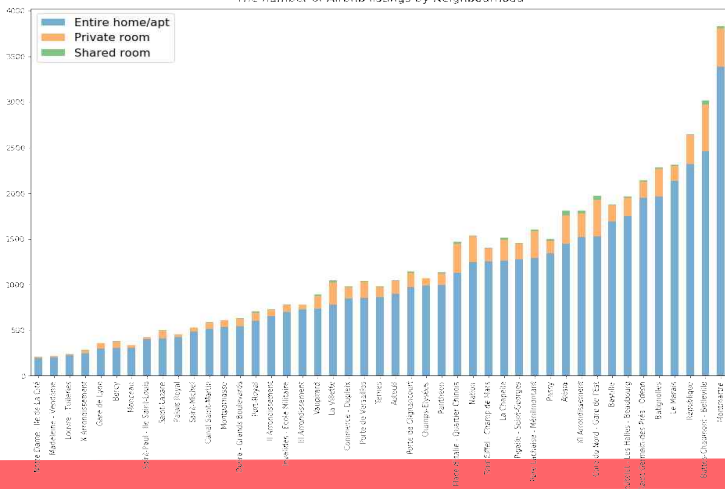
2. 데이터 분석 | 데이터 탐색

3) 지역별 숙소 분포



데이터 탐색

The number of Airbnb listings by Neighbourhood



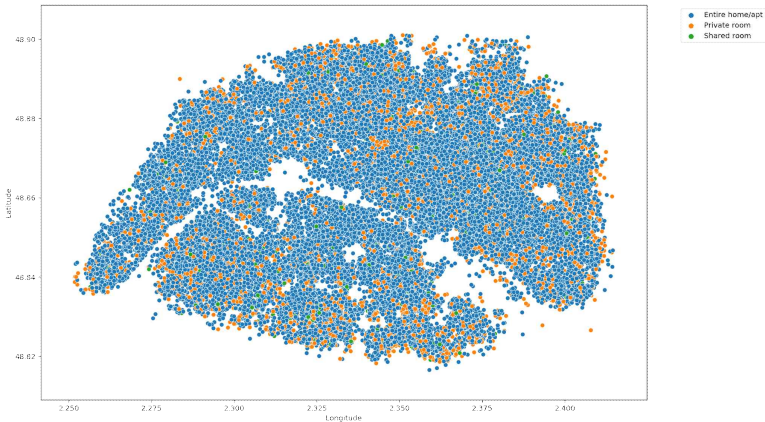
4) 지역별 숙소 유형 분포

- 대체로 비슷한 비율의 분포를 보여줌

2. 데이터 분석 | 데이터 탐색

4) 지역별 숙소 유형 분포

- 주택, 혹은 아파트 전체를 빌려주는 유형이 압도적으로 많음



데이터 탐색



가설 1.

- 2017년 파리지역의 월 평균 임대료 데이터와 Airbnb 숙소 가격을 비교해 본 결과,

건물의 월 평균 임대료가 높을수록 숙소 가격이 높아지는
경향이 보임

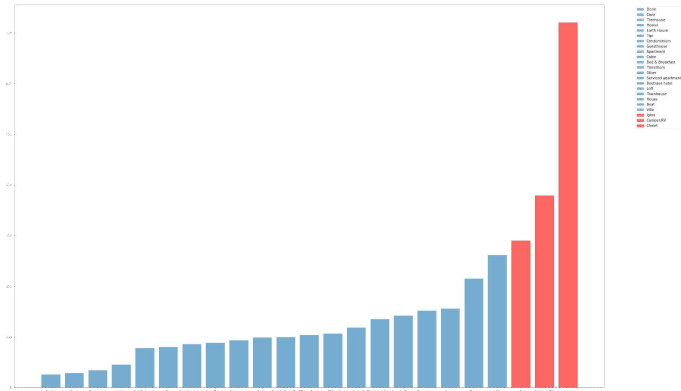
데이터 탐색

가설 2.
일반적인 숙소보다는 특별한 종류의
숙소 타입이 숙박료가 높을 것이다.

- 숙박료와 숙소의 건물 종류를 비교해 본 결과,

호텔, 일반 아파트 등의 숙소 보다는 이글루, 캠퍼 등 일상생활에서 즐기지 못하는 숙소 타입이 숙박료가 높은 경향을 보임

Average Airbnb Price by Property Type



3. 모델링

3. 모델링

1) 모델링 기법

(1) 다중선형회귀모델

(2) LGBM 회귀모델

(3) XGBoost 회귀모델

- 모델을 훈련시킬 훈련데이터, 훈련된 모델을 검증할 검증데이터, 검증된 모델의 최종 성능을 확인하기 위한 테스트데이터로 구분

- 위 세가지 기법을 사용하여 예측력을 비교한 뒤, 성능이 가장 좋은 모델을 선택

3. 모델링

다중선형회귀모델

(1) 다중선형회귀모델

	MSE	RMSE	MAE	R2
기준모델	0.309225	0.556080	0.427786	0.0
훈련데이터	0.095182	0.308516	0.230745	0.690395
검증데이터	0.095354	0.308795	0.231877	0.693646

3. 모델링

LGBM 회귀모델

(2) LGBM 회귀모델

	MSE	RMSE	MAE	R2
기준모델	0.309225	0.556080	0.427786	0.0
훈련데이터	0.049174	0.221753	0.165214	0.840047
검증데이터	0.077612	0.278590	0.209286	0.750647

(3) XGBoost 회귀모델

	MSE	RMSE	MAE	R2
기준모델	0.309225	0.556080	0.427786	0.0
훈련데이터	0.046771	0.216265	0.162551	0.847866
검증데이터	0.078920	0.280927	0.210622	0.746445

3. 모델링

모델 튜닝

(4) 모델 튜닝

- 세가지 기법중 가장 성능이 좋은 LGBM 회귀모델을 튜닝하여 최적의 성능을 얻고자 함

	MSE	RMSE	MAE	R2
기준모델	0.309225	0.556080	0.427786	0.0
훈련데이터	0.075687	0.275112	0.205921	0.753809
검증데이터	0.080316	0.283401	0.212375	0.741960
최종성능	0.079530	0.282010	0.210742	0.746059

과적합



일반화 성능

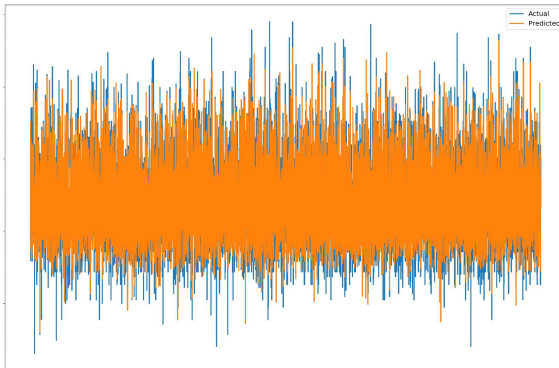


4. 결론

4. 결론

모델 예측 성능 확인

LGBM Predict



	Price	Predict
38338	5.141664	4.916660
53354	3.688879	3.753853
32059	4.532599	4.373907
18809	3.931826	4.003782
748	5.916202	5.440200
...
51727	5.017280	4.800951
44645	4.189655	3.888173
34691	6.282267	5.933025
49108	4.189655	4.139804
1697	5.017280	4.801121

(1) 모델 예측 성능 확인

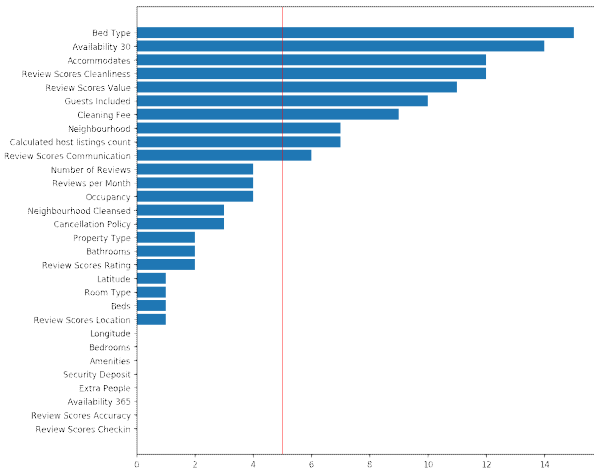
- 최적의 모델로 학습된 모델로

테스트데이터를 예측한 결과,

실제값을 비슷하게 예측하는 것으로 보임

4. 결론

특성중요도



(2) 특성중요도 확인

- 침대의 타입, 검색일로부터 30일간 예약 가능한 날짜, 최대수용인원, 그리고 청결도 점수가 가격에 가장 큰 영향을 미치는 요소로 나타났다

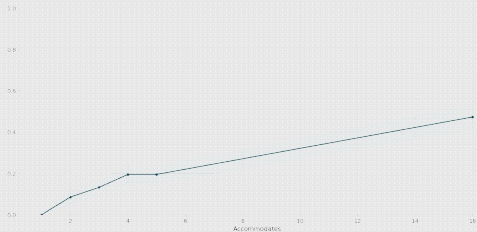
- 또한, 방, 또는 건물 타입이나 리뷰 점수 등 일반적으로 가격에 영향을 미칠 것 같은 특성이 그렇게 중요하지 않게 나타났다

4. 결론

부분의존도

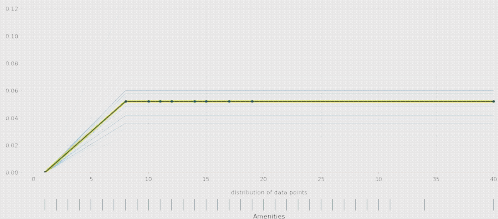
PDP for feature "Accommodates"

Number of unique grid points: 6



PDP for feature "Amenities"

Number of unique grid points: 10



(3) 부분의존도 확인

- 일반적인 상식과 같이 최대 수용인원이 늘어날수록 가격이 상승하는 경향을 보여줌
- 어메니티는 6~7종류까지 가격에 영향을 미치며, 그 이후로는 별다른 영향을 주지않는 것으로 분석됨

4. 결론

(4) 결론

- 신규호스트의 경우 리뷰가 없는 상태에서 시작하니, 이를 제외한 중요한 가격 결정 요인은 침대의 타입, 최대수용인원, 검색일로부터 30일간 예약가능한 날짜, 그리고 청결도로 확인됨
- 또한, 침대의 타입, 최대수용인원 등의 요소로 미루어보아 얼마나 수용가능한지, 침대의 크기는 얼마나 되는지 등 많은 사람을 수용할 수 있는 숙소일수록 가격이 높은 것으로 판단됨
- 숙소의 지리적 위치가 가격선정에 큰 영향을 미치기 때문에, 숙소의 가격을 책정할 때 인근의 Airbnb 숙소 뿐 아니라 일반 숙소들의 렌트가격도 고려하여 합리적으로 선정해야할 것으로 판단됨

4. 결론

(5) 한계점

- 당 모델은 시간에 따른 수요 변화를 반영하지 못하는 한계점이 있음. 숙박 시설은 계절성에 큰 영향을 받으며, COVID-19 등과 같은 세계 경제에 큰 타격을 주는 요인이 있는 경우 가격의 변동성이 커지나 이 모델은 이를 적절히 반영하지 못하고 있음
- 또한, 건물의 월 임대료가 숙박비에 영향을 미치는 것으로 분석되었지만 모델에 적용하지 못하였음. 추후 이를 반영한다면 더욱 정확한 모델이 될 수 있을 것이라 생각됨

Thank you
