



신규 호스트를 위한 숙소 가격 예측 모델

프랑스 파리를 중심으로

황형기

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

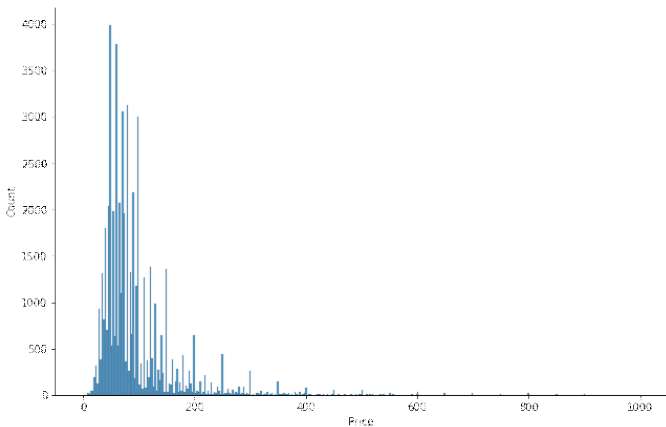
2) Feature : Target을 예측할 때 사용되는 변수

- 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 30/60/90/365:
- Number of Reviews : 위도, 경도
- Review Scores : 각종 리뷰 점수(청결, 위치 등)
- Cancellation Policy: 환불 정책
- Reviews per Month : 월별 리뷰 수

2. 데이터 분석 | 데이터 탐색

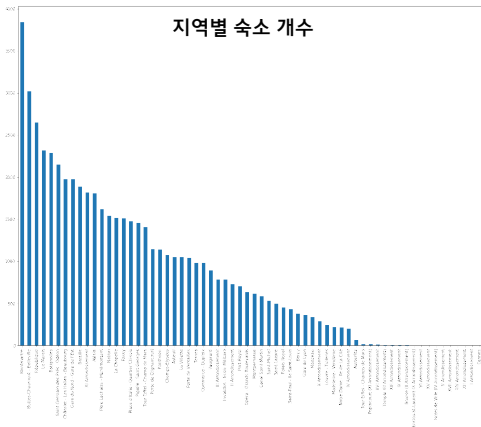
1) 가격의 분포

- 한쪽으로 지나치게 **쏠려있는** 모습을 보임



2. 데이터 분석

데이터 탐색



2) 숙소가 가장 많은 Top 3 지역

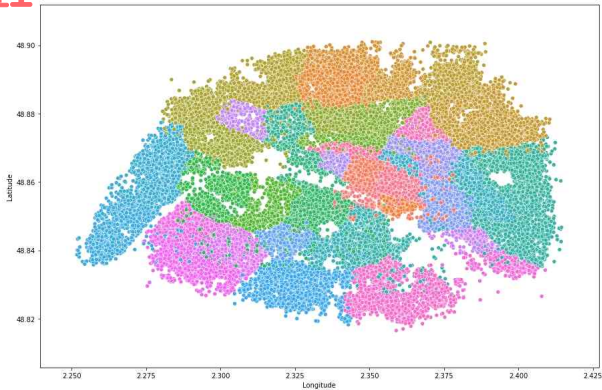
(1) Montmartre

(2) Buttes=Chaumont - Belleville

(3) Republique

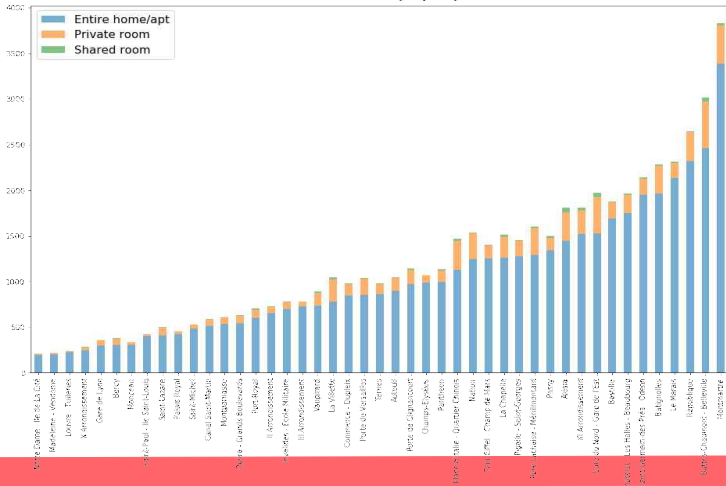
2. 데이터 분석 | 데이터 탐색

3) 지역별 숙소 분포



데이터 탐색

The number of Airbnb listings by Neighbourhood



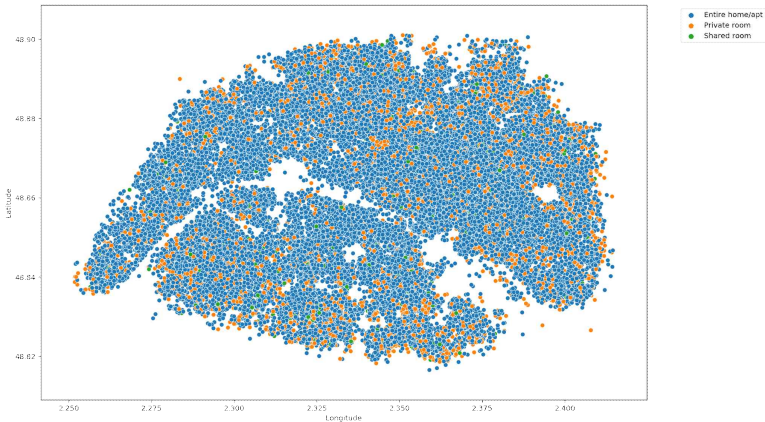
4) 지역별 숙소 유형 분포

- 대체로 비슷한 비율의 분포를 보여줌

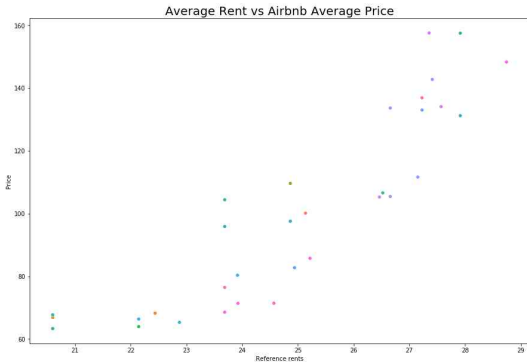
2. 데이터 분석 | 데이터 탐색

4) 지역별 숙소 유형 분포

- 주택, 혹은 아파트 전체를 빌려주는 유형이 압도적으로 많음



2. 데이터 분석 | 데이터 탐색



가설 1.
지역별 평균 월임대료가 높을수록
숙박료가 높을 것이다.

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

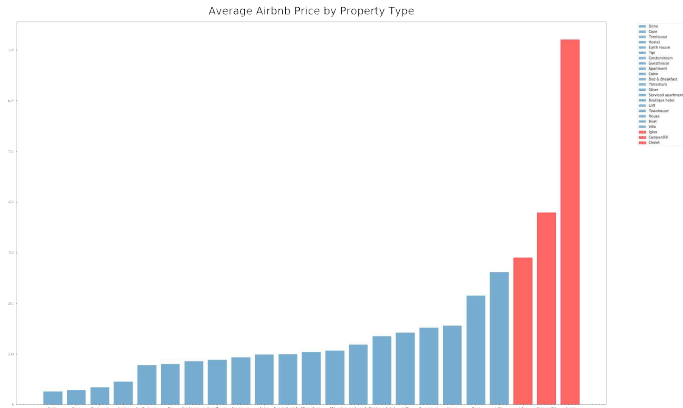
건물의 월 평균 임대료가 높을수록 숙소 가격이 높아지는
경향이 보임

2. 데이터 분석 | 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 타
입이 숙박료가 높은 경향을 보임



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	5375.1633164878085	73.31550529381768	45.17688286355283	0.0
훈련데이터	2.153899e+03	4.641011e+01	2.703064e+01	6.093965e-01
검증데이터	4.184329e+26	2.045563e+13	2.195979e+11	-7.793852e+22

3. 모델링

LGBM 회귀모델

(2) LGBM 회귀모델

	MSE	RMSE	MAE	R2
기준모델	0.308750	0.555653	0.427517	0.0
훈련데이터	0.046693	0.216087	0.161769	0.849108
검증데이터	0.077073	0.277620	0.208930	0.747935

3. 모델링

XGBoost 회귀모델

(3) XGBoost 회귀모델

	MSE	RMSE	MAE	R2
기준모델	0.308750	0.555653	0.427517	0.0
훈련데이터	0.029768	0.172535	0.127763	0.903802
검증데이터	0.078042	0.279359	0.210290	0.744767

(4) 모델 튜닝

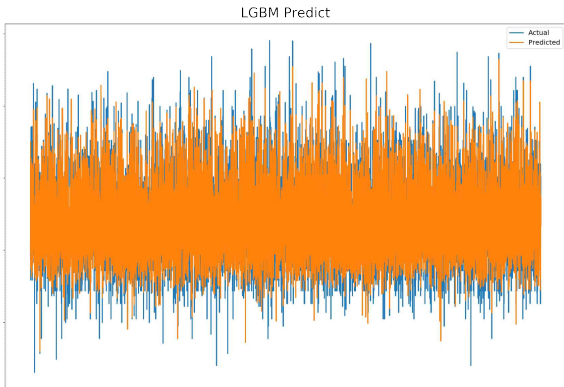
- 세가지 기법중 가장 성능이 좋은 LGBM 회귀모델을 튜닝하여 최적의 성능을 얻고자 함

	MSE	RMSE	MAE	R2
기준모델	0.308750	0.555653	0.427517	0.0
훈련데이터	0.052912	0.230026	0.171880	0.829012
검증데이터	0.077021	0.277526	0.207896	0.748106
최종성능	0.073912	0.271868	0.204589	0.760617

4. 결론

4. 결론

모델 예측 성능 확인

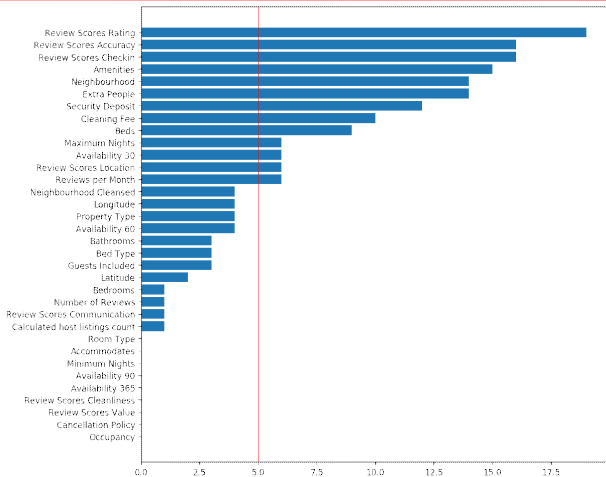


(1) 모델 예측 성능 확인

- 최적의 모델로 학습된 모델로 테스트데이터를 예측한 결과, 실제값을 비슷하게 예측하는 것으로 보임

4. 결론

특성중요도



(2) 특성중요도 확인

- 기존 예측과는 다르게 Review Score가 모델의 성능에 가장 영향을 미치는 요소로 나타났음

- 또한, 방의 타입이나 청결도 점수같은 일반적으로 가격에 영향을 미칠 것 같은 특성이 그렇게 중요하지 않게 나타났다

4. 결론

(3) 결론

- 신규호스트의 경우 리뷰가 없는 상태에서 시작하니, 이를 제외한 중요한 가격 결정 요인은 편의/부대시설, 지역, 청소요금이나 보증금 등으로 확인됨
- 또한, Checkin 리뷰의 평가기준인 친절성, 신속성등과 같이 호스트의 역량에 따라 가격이 달라질 수 있다
- 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

Thank you
