
AirBnB Superhost 예측하기

Content

1. Introduction and hypothesis
2. Data explanation
3. Visualization
4. Model application
5. Conclusion

Why AirBnB?

- **AirBnB**
 - Data availability

Why Superhost? - Credibility & Profit



Superhost: Recognising the best in hospitality

The Superhost programme celebrates and rewards Airbnb's top-rated and most experienced hosts.

슈퍼호스트 자격 요건

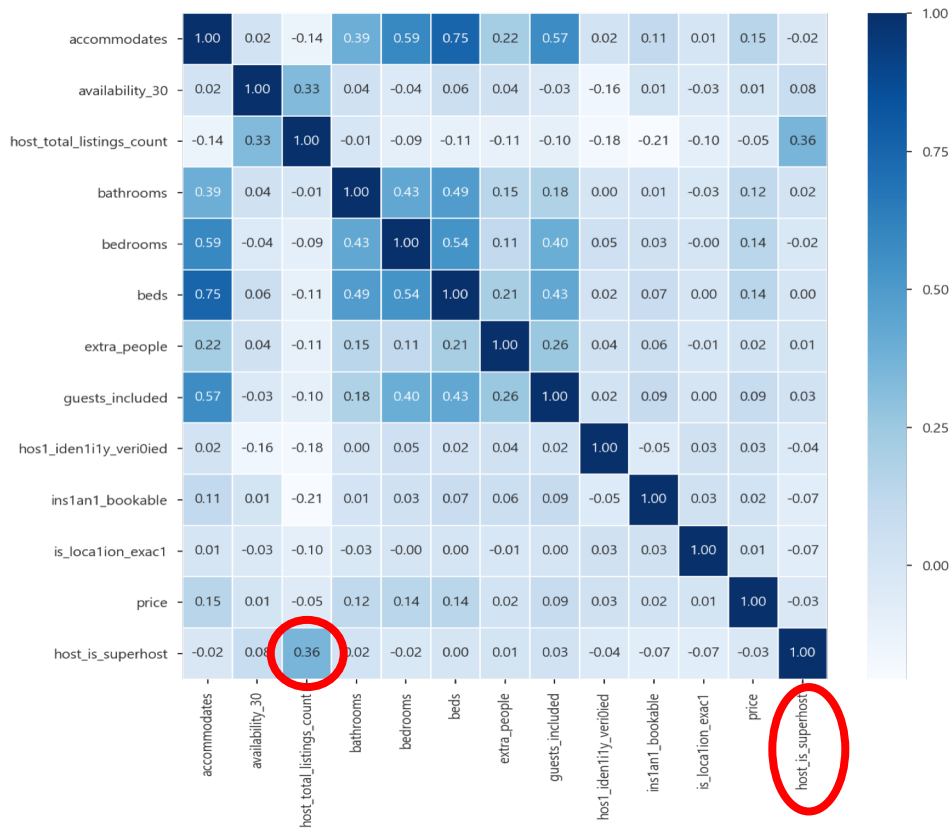
- 숙박 10건 이상 호스팅, 또는 3건의 예약에 걸쳐 100박 이상을 호스팅
- 응답률 90% 이상 유지
- 정상참작이 가능한 상황을 제외하고 예약 취소율 1%(예약 100건당 1건 취소) 유지
- 평점 4.8점 이상 유지

Research aim & hypothesis

- Superhost badge를 얻는 것은 호스트의 **profit**과 직결
- 현재 Airbnb의 superhost 선정기준은 host 중심이 아닌 guest 중심 선정기준이다 (review와 숙박 일수는 guest에 의해 결정되는 요인이다) → 새로운 host들은 superhost가 아니기 때문에 guest에게 선택될 확률이 낮다 (**high barrier to entry**) → host가 superhost가 되기 위해서 무엇을 해야하는지 host의 중심 기준이 필요하다
- 신생 Host가 Superhost가 되기 위해서는 어떻게 해야할까?
- Superhost가 되기 위한 최적의 조건은 무엇인가?
- Review 정보를 고려하지 않고도 Superhost 달성 유무 가능성을 파악할 수 있다!

Data explanation

accommodates	beds
availability_30	extra_people
host_total_listings	guests_included
bathrooms	host_identity_verified
bedrooms	instant_bookable
is_location_exact	price



What we want to know: Is this host a superhost?

Data explanation

accommodation_id	availability_30	host_total_listings_count	bathrooms	bedrooms	beds	extra_people	guests_included	host_id	host_is_verified	instant_bookable	is_location_exact	number_of_reviews	price	host_is_superhost
3	9	2	1	1	2	156	2	1	1	0	1	130	1408	0
1	16	1	1	1	1	0	1	1	0	0	0	11	430	1
6	22	12	1	2	3	196	2	0	0	0	1	260	853	0
6	16	12	1	2	3	227	3	0	0	0	0	259	1056	0
2	1	1	1	1	1	0	1	1	0	0	1	27	939	0
2	23	0	1	1	1	0	1	1	0	0	1	0	1048	0
6	14	12	1	2	3	196	2	0	0	0	1	208	696	0
6	11	12	1	2	3	196	2	0	0	0	1	216	775	0
8	12	11	1	2	3	150	4	0	0	0	1	119	1580	0
2	30	1	1	1	1	0	1	0	0	0	1	17	696	0
2	0	1	1	1	1	100	1	0	0	0	1	4	352	0
2	14	1	1	1	1	0	1	1	0	0	1	153	994	1
4	4	18	1	0	2	102	1	1	1	1	1	73	1314	0
8	0	1	3	1	7	0	1	0	0	0	1	0	4303	0
4	26	1	1	2	3	0	4	1	0	0	0	231	1080	0
4	11	7	1	1	1	156	2	0	1	1	1	41	900	0
11	8	8	2	3	6	156	5	1	1	1	1	238	1549	0
6	0	7	2	3	3	156	2	0	1	1	1	100	939	0
3	0	1	1	2	2	156	2	1	0	0	1	22	626	0
4	0	2	1	1	2	313	4	1	1	1	1	34	782	0
2	0	2	1	1	1	0	1	0	0	0	0	20	1056	0
3	26	18	1	1	1	117	1	1	0	0	1	15	1001	0
10	30	18	3	3	6	141	3	1	1	1	1	102	2253	0
5	16	1	2	3	4	156	2	1	0	0	1	226	1158	0
10	0	1	3	5	5	0	1	0	1	0	0	2	2504	0
1	0	6	1	1	1	80	1	1	0	0	1	5	297	0
1	29	2	1	1	1	0	1	0	0	0	0	40	485	0
2	0	1	1	1	1	0	1	0	0	0	1	0	1017	0
3	26	18	1	1	2	117	1	1	0	0	1	18	782	0
5	21	3	1	1	3	130	2	1	1	1	1	217	548	0
6	7	2	1	2	2	99	4	1	0	0	1	434	1001	1
4	0	1	1	1	2	274	2	1	0	0	1	10	1244	0
2	30	16	1	1	1	0	1	0	0	0	1	74	383	0
3	15	2	2	2	2	0	1	1	0	0	1	20	1322	0
7	18	1	1	3	4	150	5	0	0	0	1	197	1103	1
4	2	4	1	1	2	156	2	1	1	1	1	170	868	0

Model application

학습하기

```
In [10]: 1 from sklearn.tree import DecisionTreeClassifier
          2 from sklearn.model_selection import train_test_split
```

```
In [11]: 1 X = data.iloc[:, :-1]
          2 y = data.iloc[:, -1]
```

```
In [12]: 1 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

```
In [13]: 1 dt = DecisionTreeClassifier()
          2 dt.fit(X_train, y_train)
```

```
Out[13]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
                                max_features=None, max_leaf_nodes=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, presort=False, random_state=None,
                                splitter='best')
```

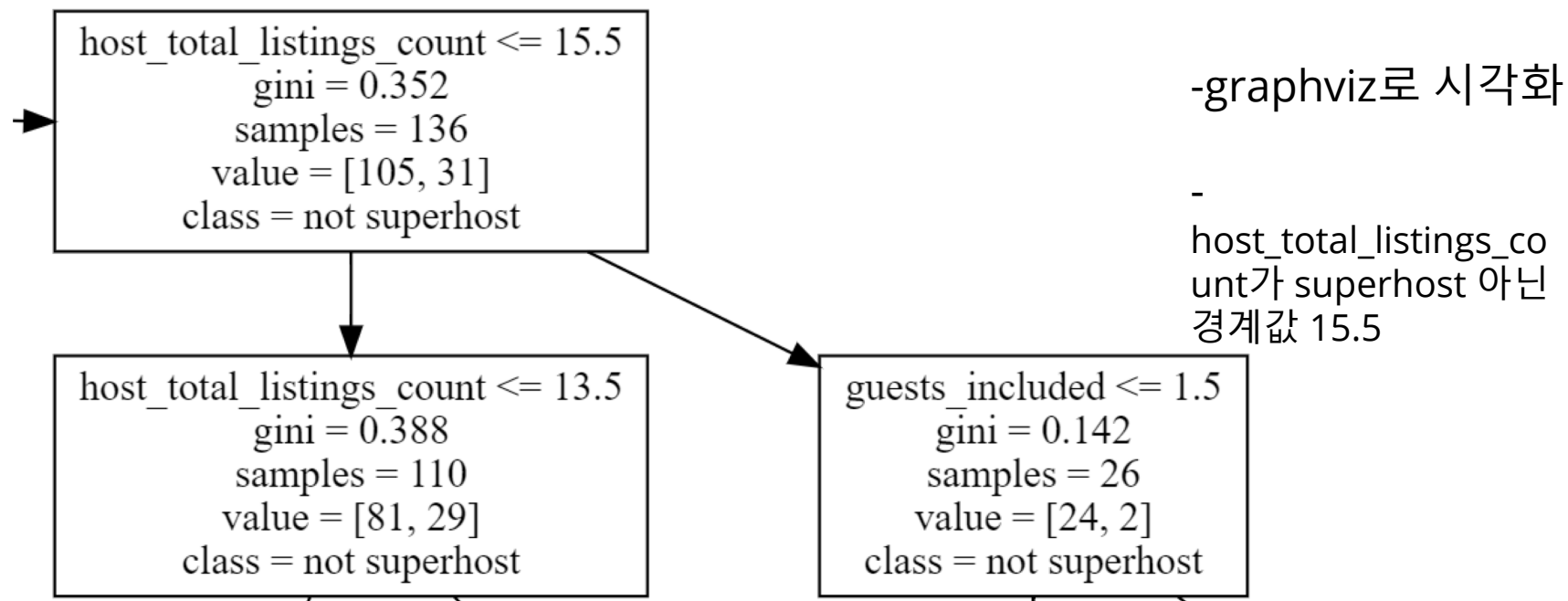
```
In [14]: 1 dt.score(X_test, y_test)
```

```
Out[14]: 0.8573322726067356
```

선택한 모델: Decision
Tree

학습 결과: 0.85(!!!!!)

Visualization



Conclusion

- 결론

- host가 몇개의 AirBnB listing을 가지고 있는지가 중요한 변수(corr 0.36)
- 중요한 것은 host의 hosting 경험이다!
- 신규 host들이 superhost가 되려면, 경험을 쌓는 수 밖에....

- 한계점과 보완점

- 과적합의 문제
- 실제 superhost 선정 기준에서 review가 차지하는 비중이 높음->근본적인 AirBnB 평가 시스템을 변화시킬 수는 없음
- 다른 도시들에도 적용가능한지 확인