

Enhancing Airbnb Ratings Through Topic Model Scores



Jon Yu - Metis
August 19, 2021

Airbnb and the U.S. as of 2021



- ∅ Low mortgage rates causing surge in new landlords*
 - ∅ Multifamily investment up 56% quarter to quarter
- +
- ∅ 2.9 million hosts with 7 million listings worldwide**
 - ∅ 14,000 new hosts each month
- +
- ∅ U.S. unemployment up 50% since 2019 (3.6% to 5.4%)***

* <https://money.com/historic-low-mortgage-rates-new-investment-property/>

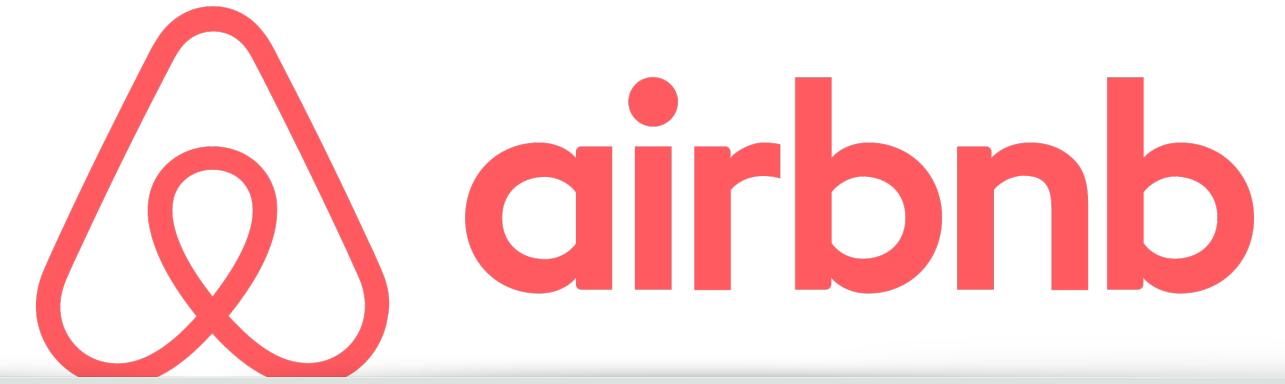
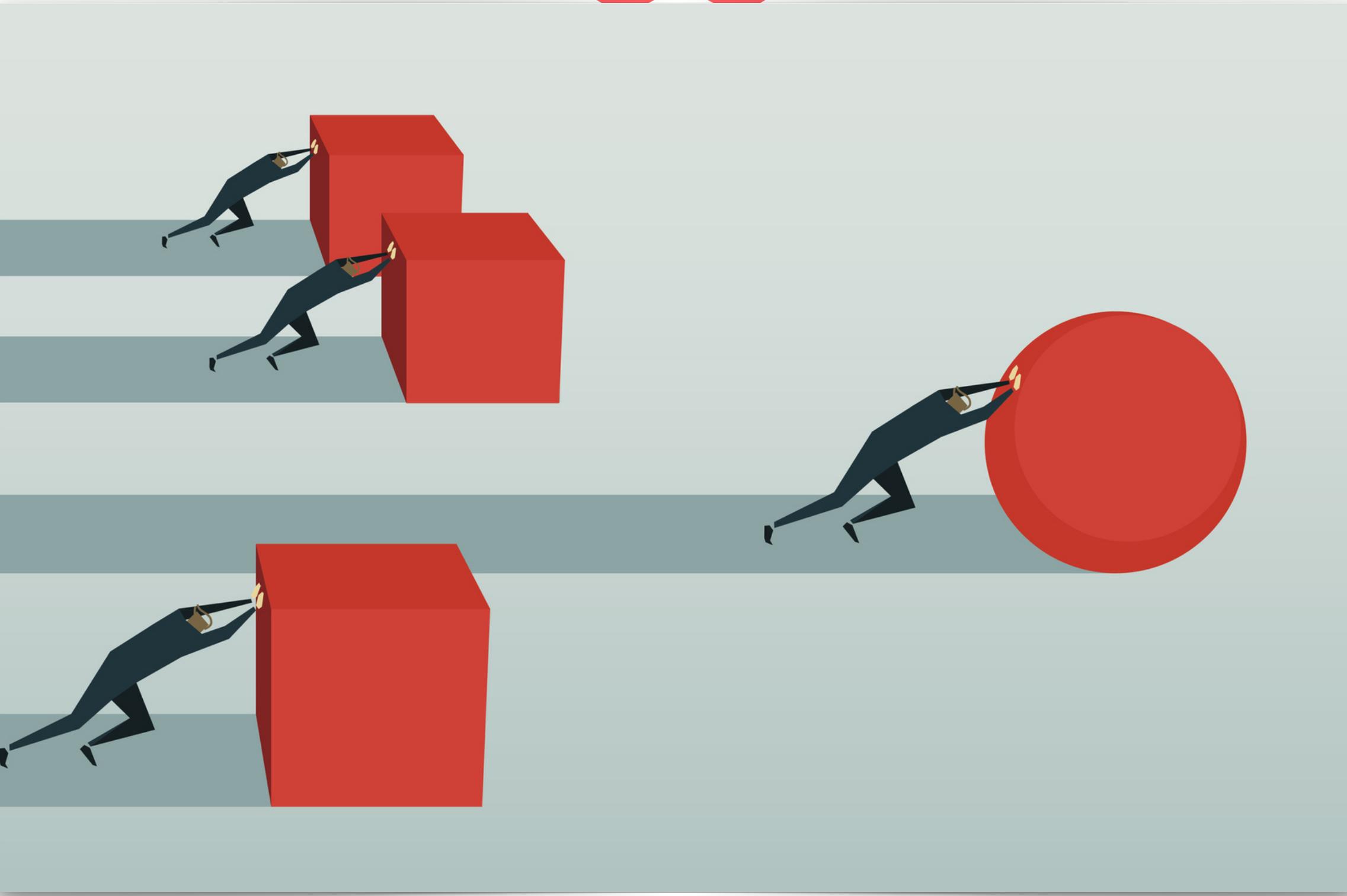
** <https://capitalcounselor.com/airbnb-statistics/>

*** <https://data.bls.gov/timeseries/LNS14000000>

Airbnb as of 2021



- ⌚ Ever competitive market for hosts
- ⌚ Struggle to maintain occupancy rate

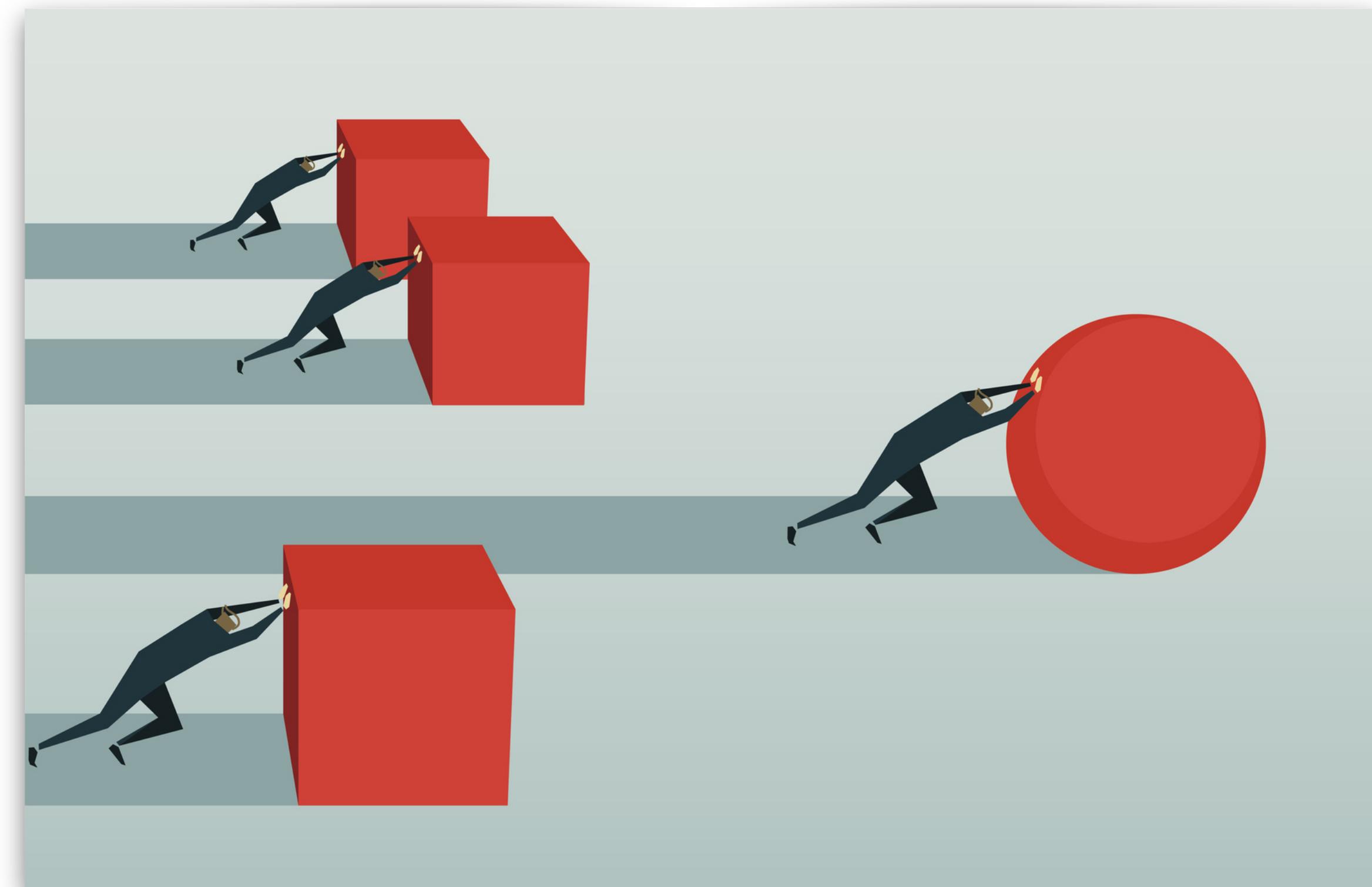
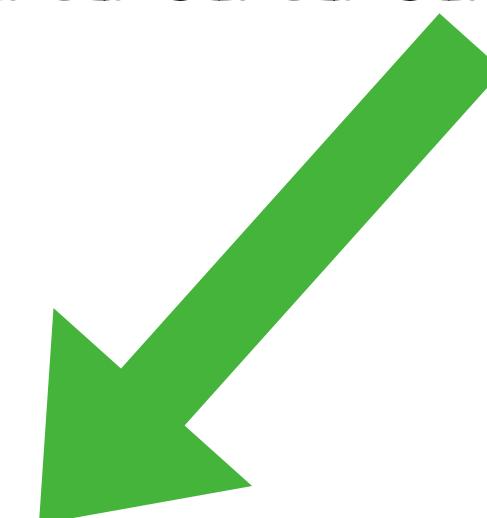


Motivation



Discover **HIDDEN METRICS** driving guest ratings

HOSTS: set your space
apart from the rest



Data



Inside Airbnb

Adding data to the debate

HAWAII*



~603k reviews (docs)

~22k listings

*Honolulu, Hawaii, has the highest Airbnb occupancy rate in the US — 68%

(<https://capitalcounselor.com/airbnb-statistics/>)

Preprocessing & EDA



```
# Set custom stopwords

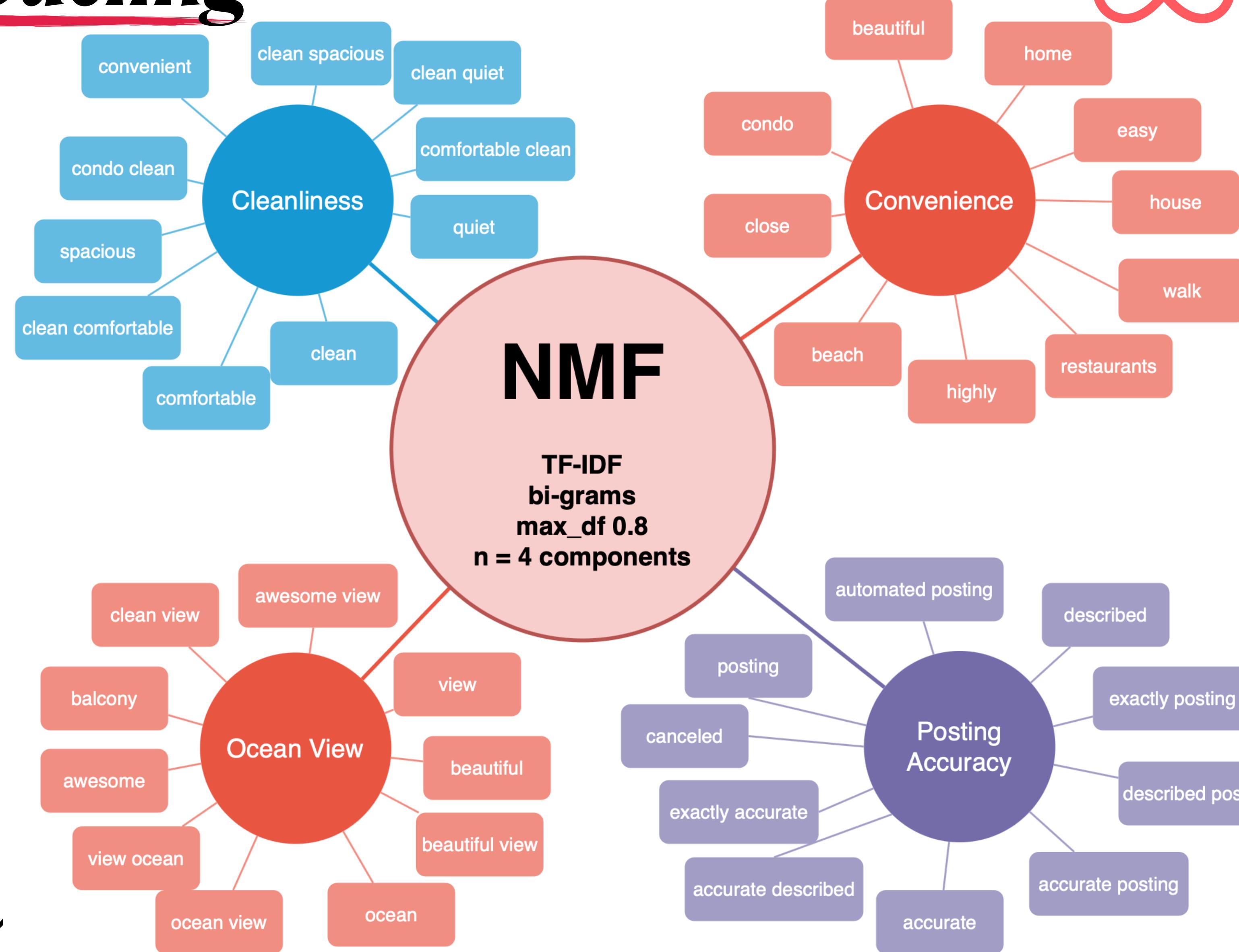
custom_stopwords = ["hawaii", "airbnb", "bnb", "oahu", "honolulu", "island", "big", "a", "aa", "aaa",
                     "moloka", "ahu", "kau", "ni", "ihau", "i", "kaho", "olawe", "lana", "maui", "hawai",
                     "good", "great", "perfect", "br", "amazing", "recommend", "place", "like", "nice",
                     "room", "stay", "overall", "really", "check", "host", "unit", "did", "didn", "day",
                     "night", "just", "super", "famly", "location", "love", "loved", "days",
                     "definitely", "wonderful", "reservation", "arrival", "s"]
final_stopwords = text.ENGLISH_STOP_WORDS.union(custom_stopwords)
```



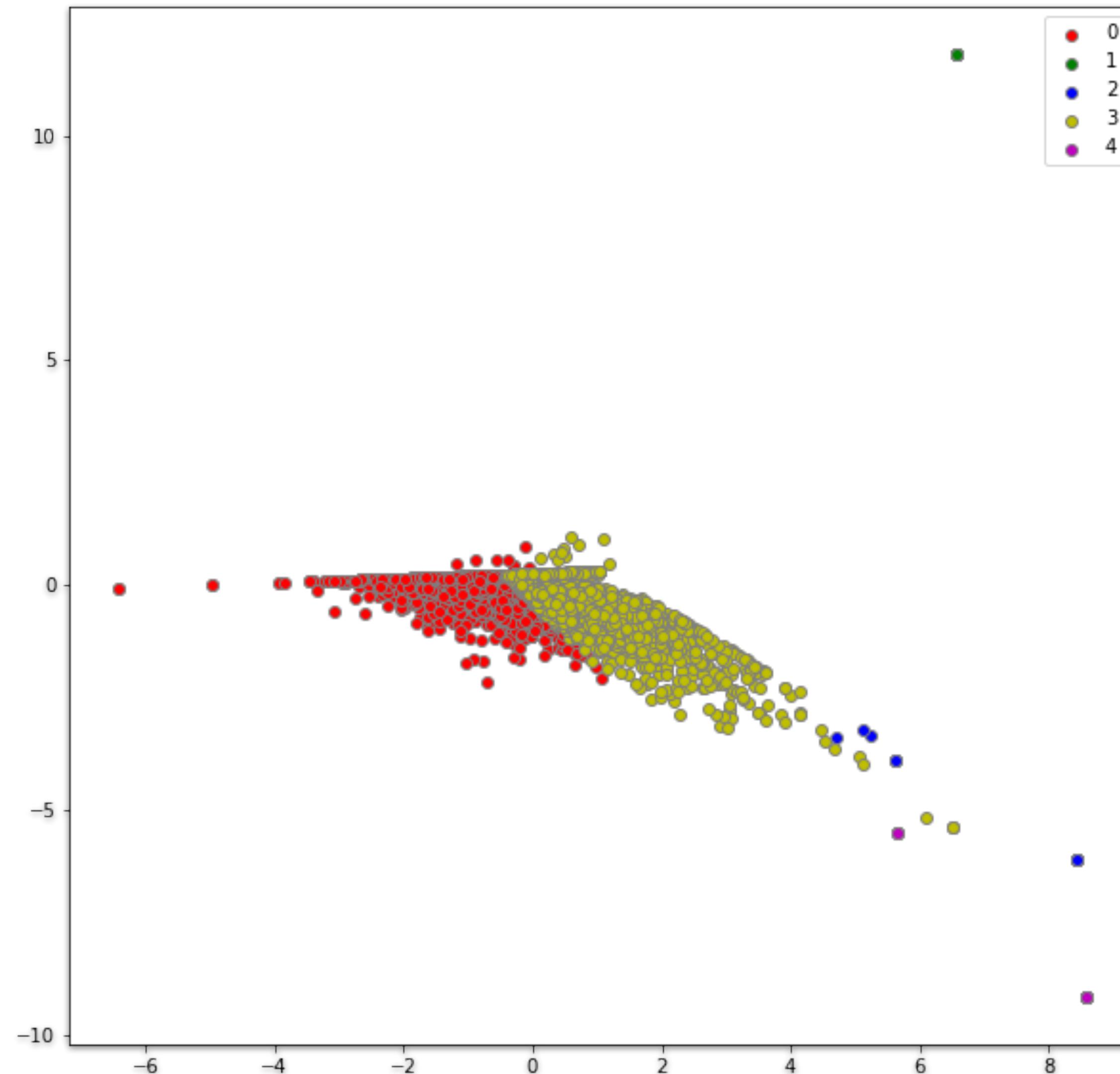
Preprocessing & EDA



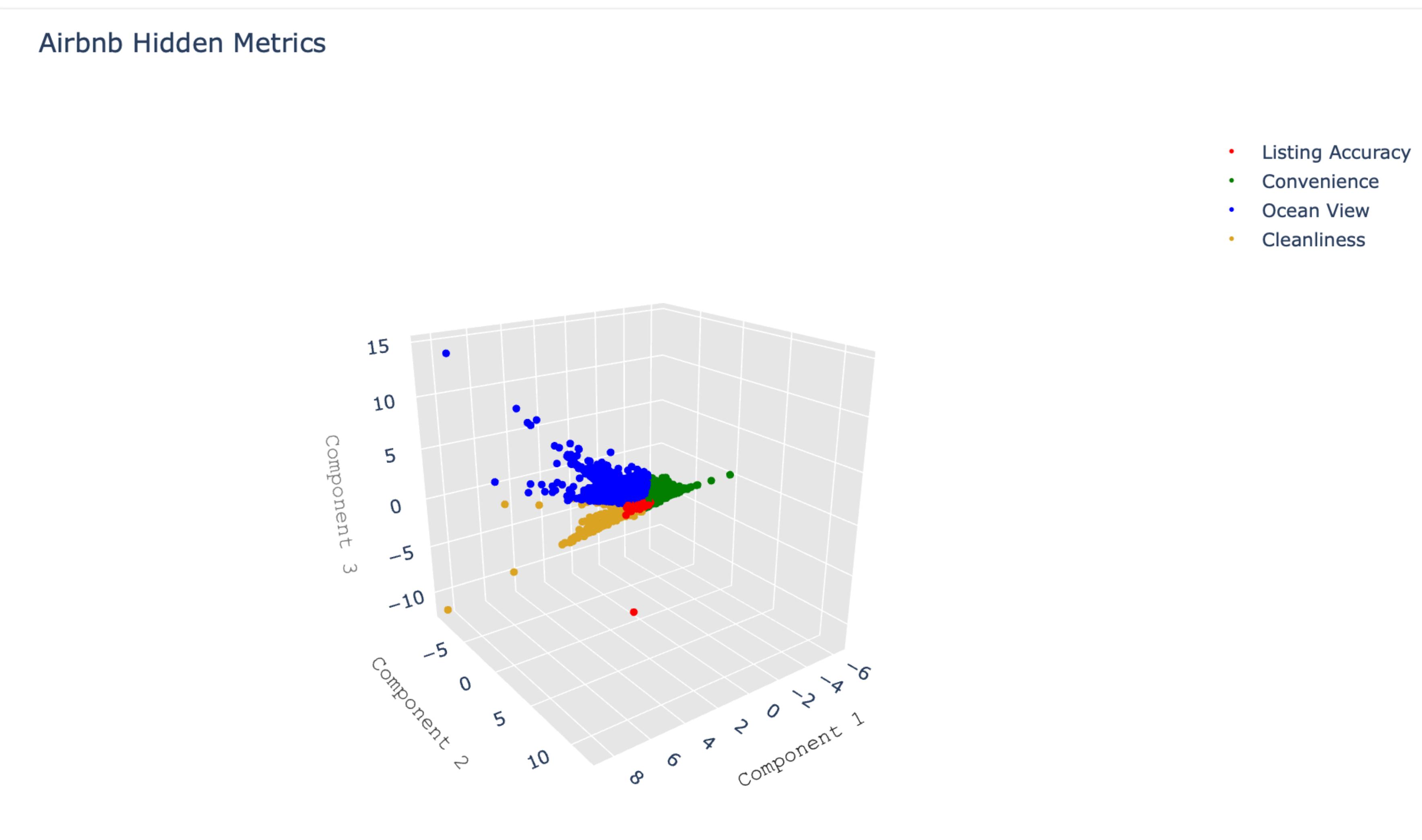
Topic Modeling



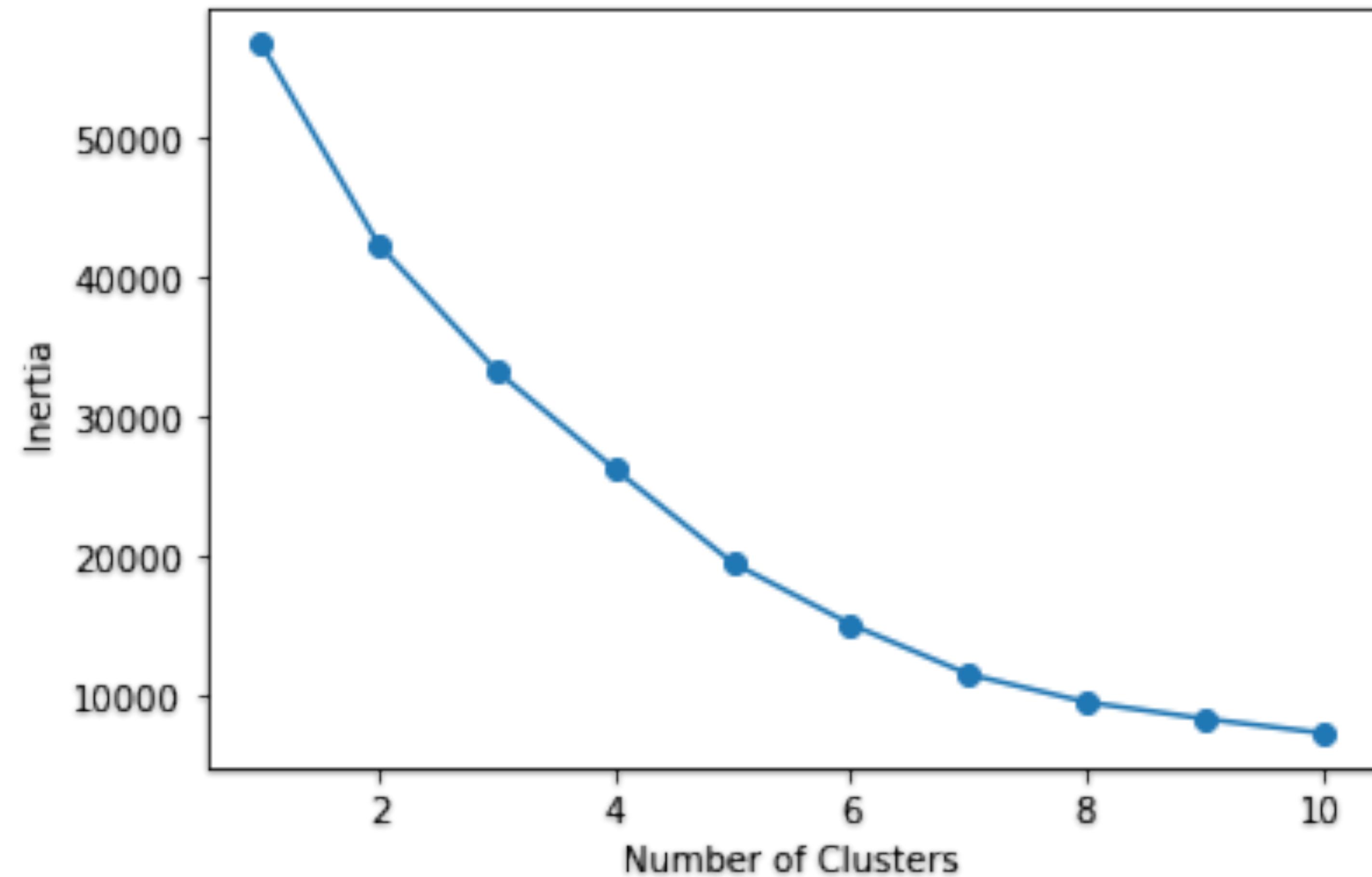
Dimensionality Reduction: 2-D PCA



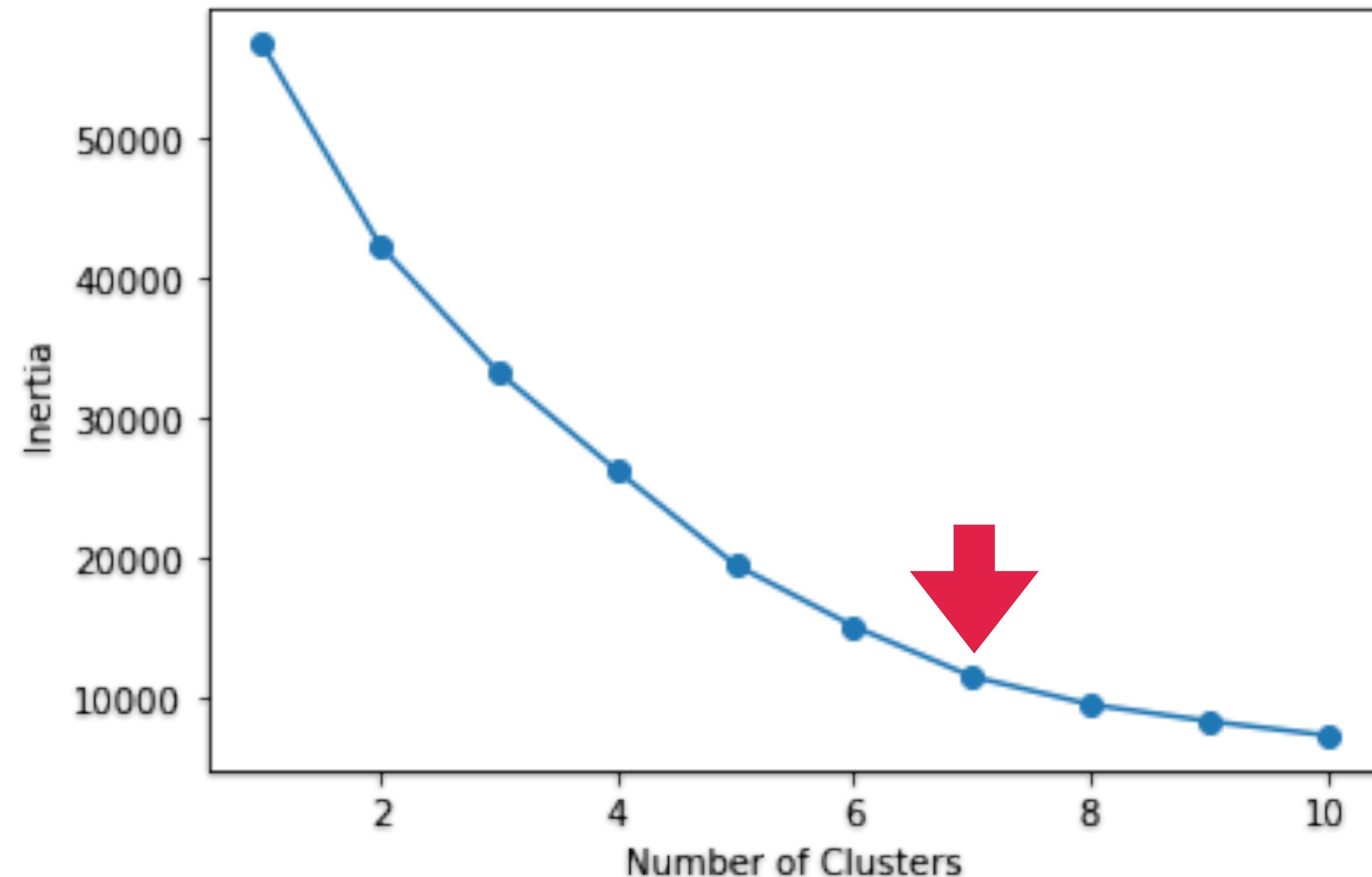
Dimensionality Reduction: 3-D PCA



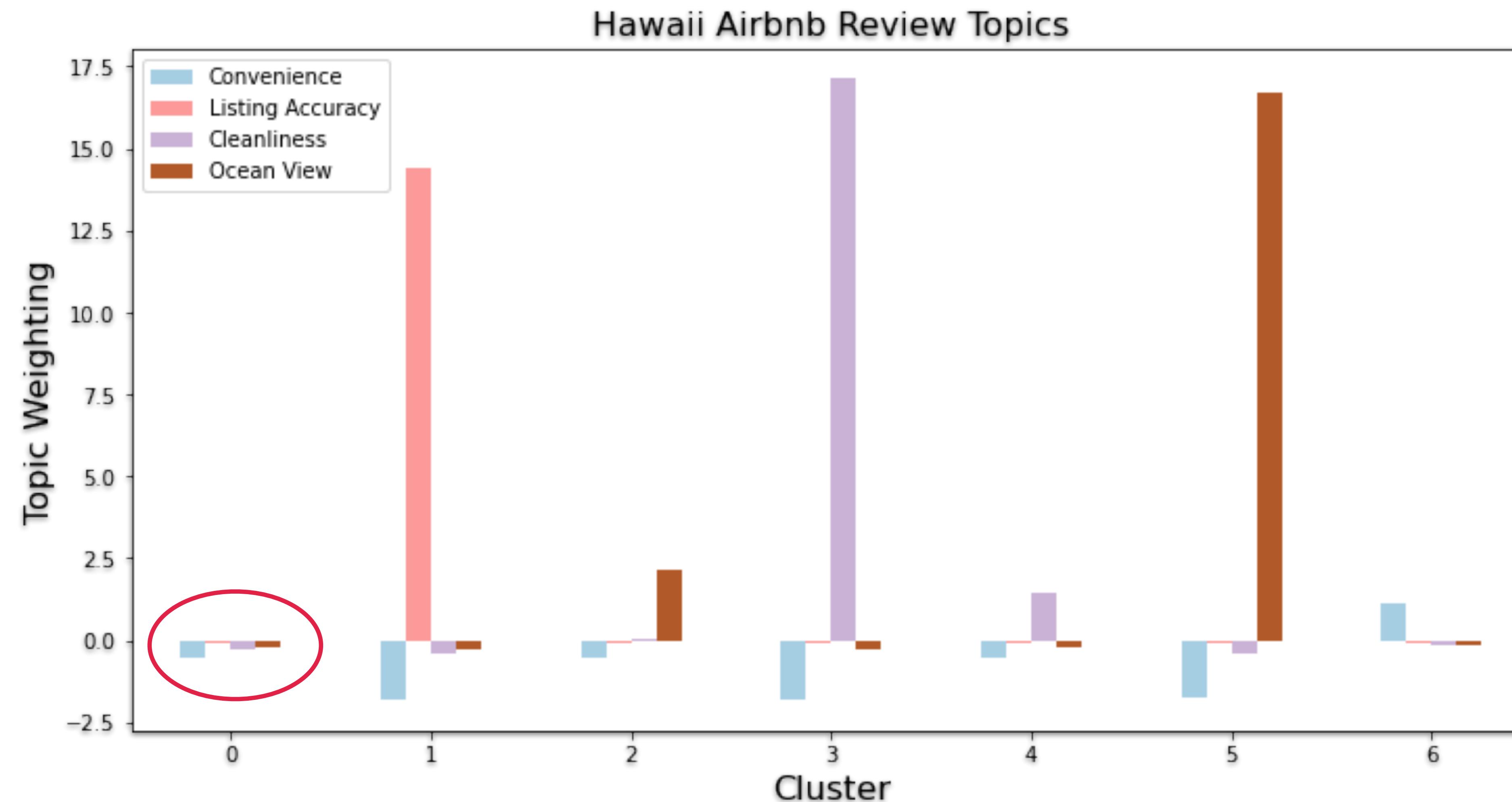
K-Means Clustering: Inertia



K-Means Clustering: Inertia



K-Means: 7 Clusters



K-Means: 5 Clusters

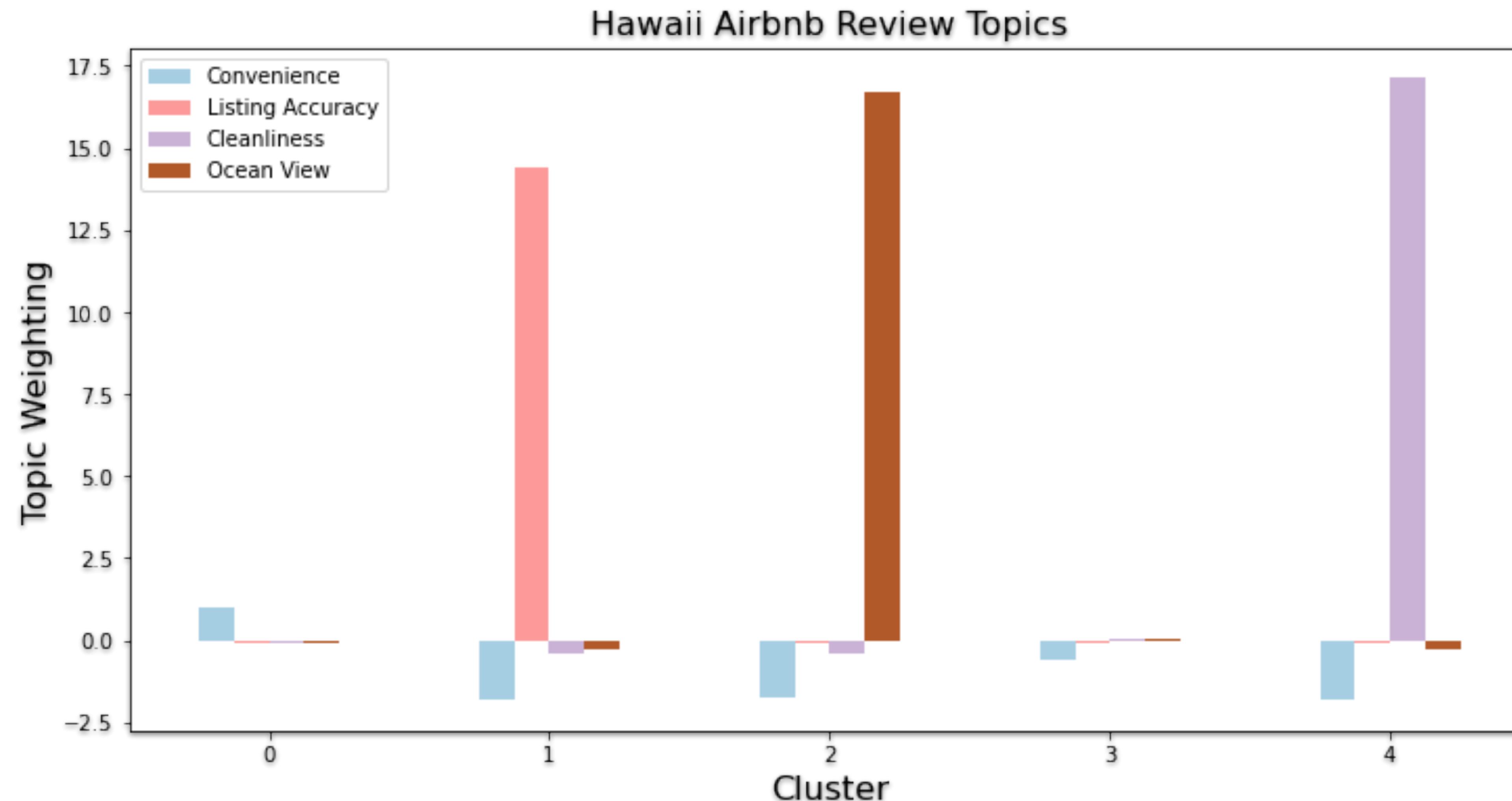


Tableau Demo

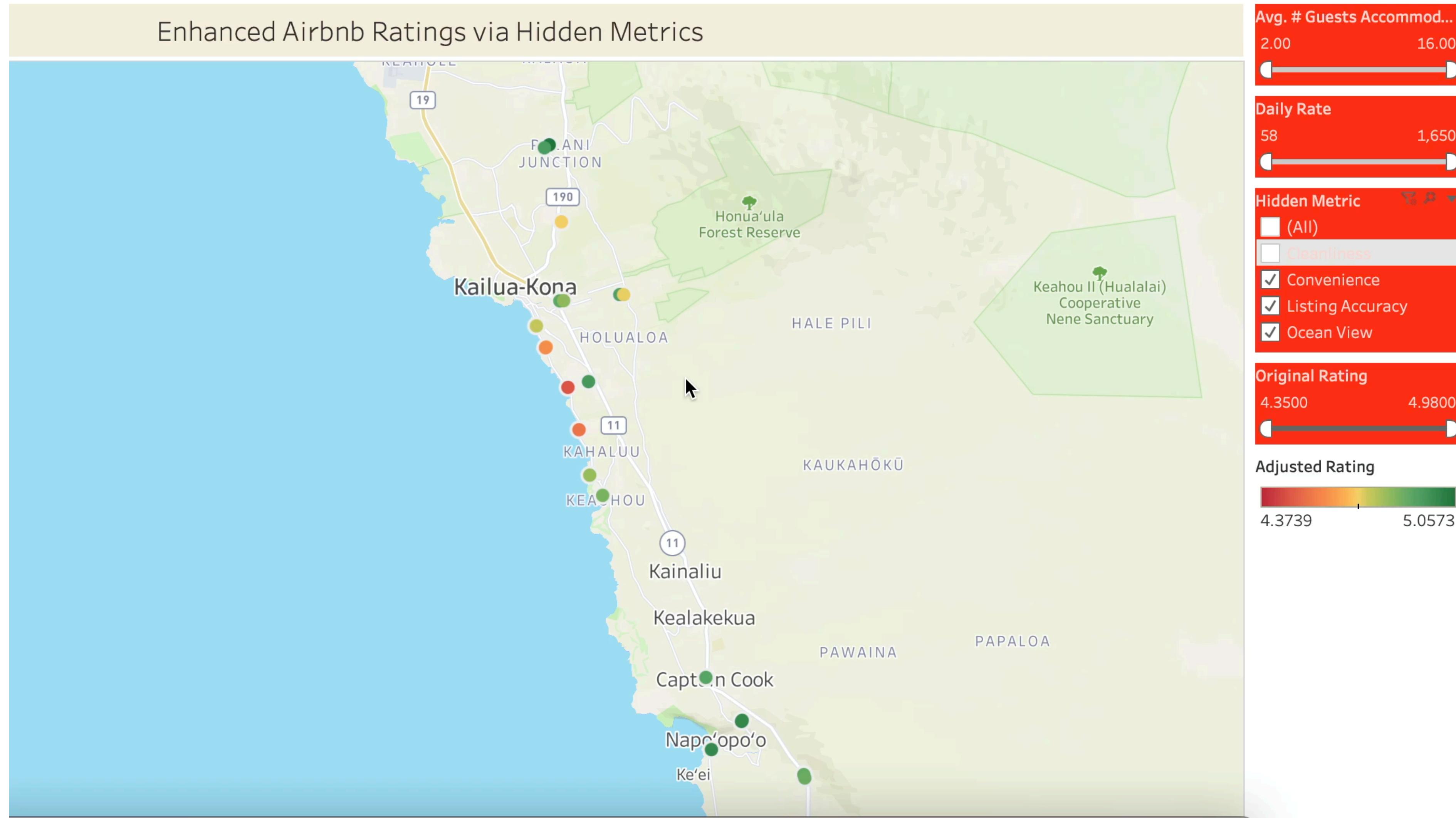
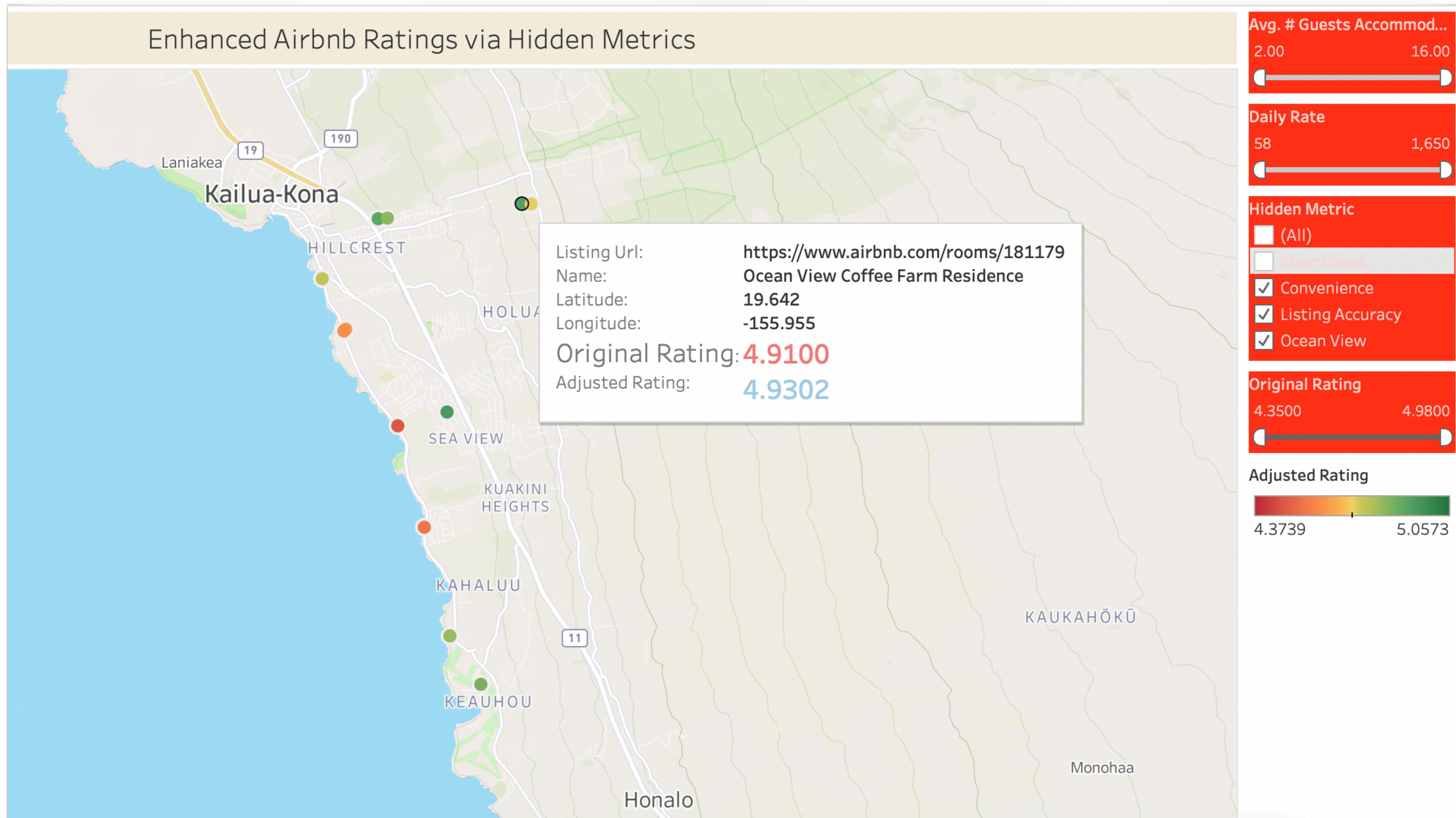


Tableau Demo



Conclusions



⊗ **TF-IDF NMF**
enables optimal
Interpretability

Conclusions



⊗ **TF-IDF NMF**

enables optimal
Interpretability

As a Host

FOCUS ON

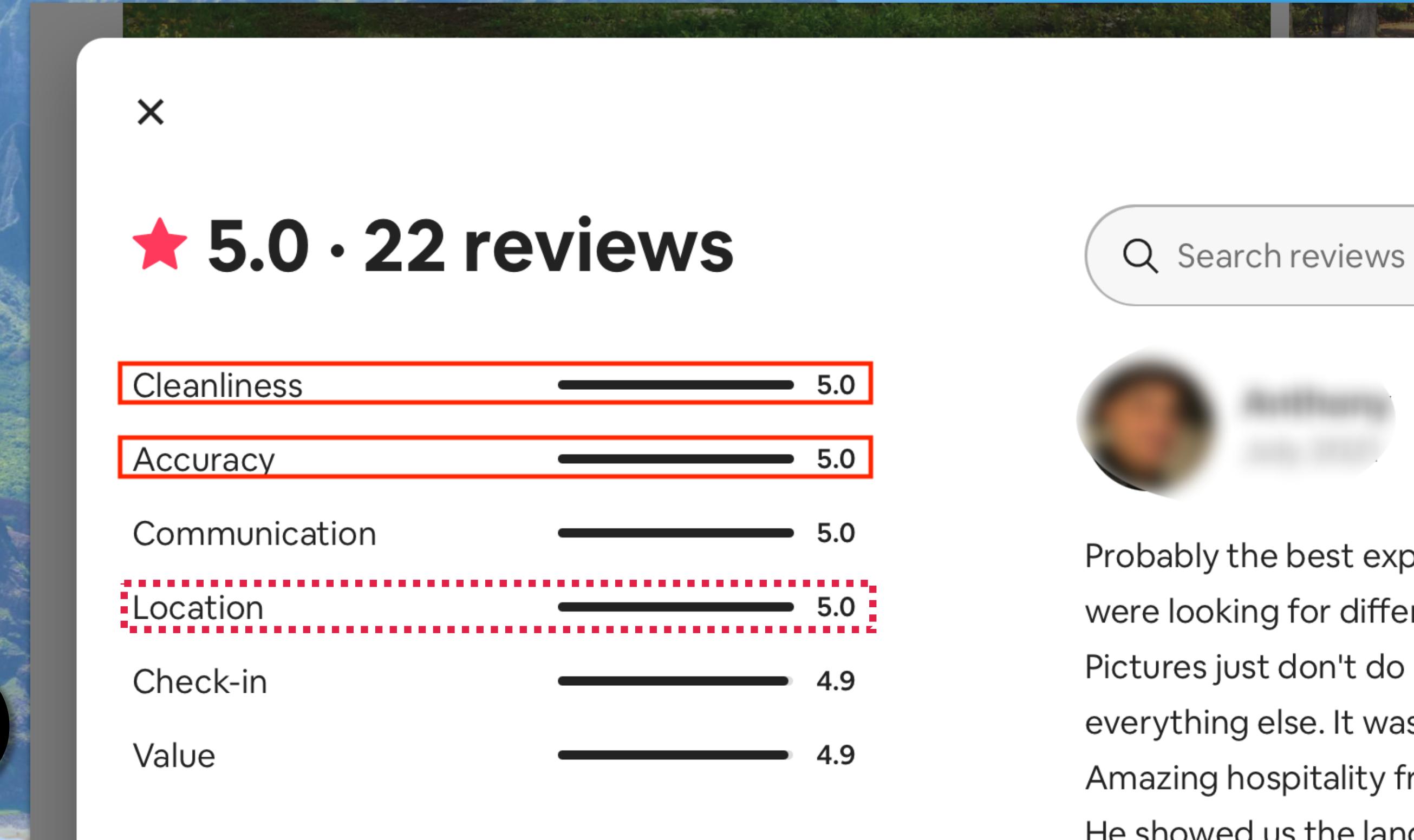
Convenience
Listing Accuracy
Cleanliness
Ocean View (location)

Conclusions



FOCUS ON

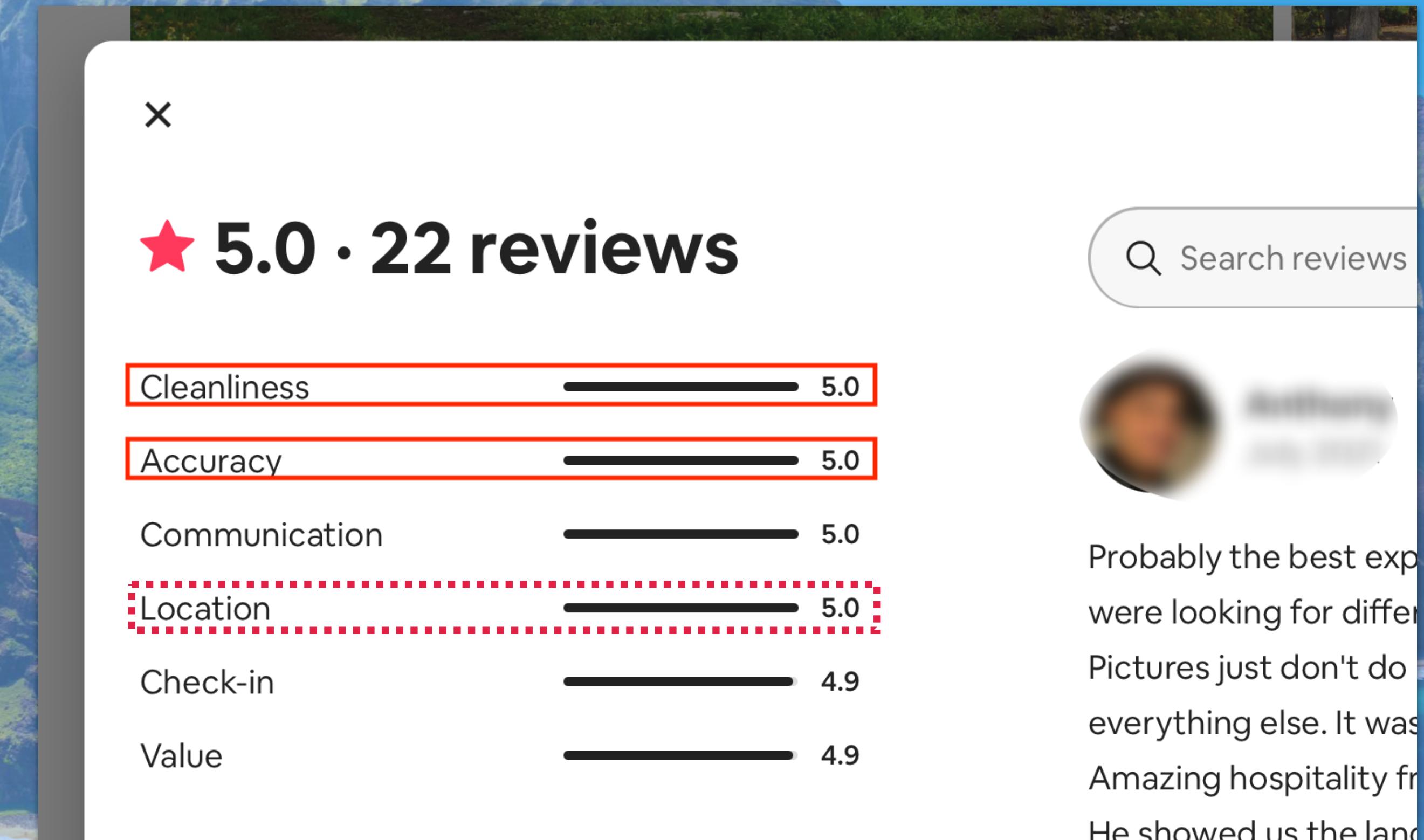
Convenience
Listing Accuracy
Cleanliness
Ocean View (location)



Conclusions

Airbnb has already conducted extensive research & implemented thoughtful sub-ratings

Consider region specific metrics





Future Work

⌚ Selectively filter out existing
sub-ratings using stop words

Thank you!



PEOPLE



PLACES



LOVE



AIRBNB

