

# Airbnb Review Analysis





## **The impetus- so you want the perfect vacation...**

- Many items have hundreds of reviews
- In some cases, genuine reviews are padded with fraudulent reviews
- Making an informed decision is time intensive

**Who has time for that!?!**

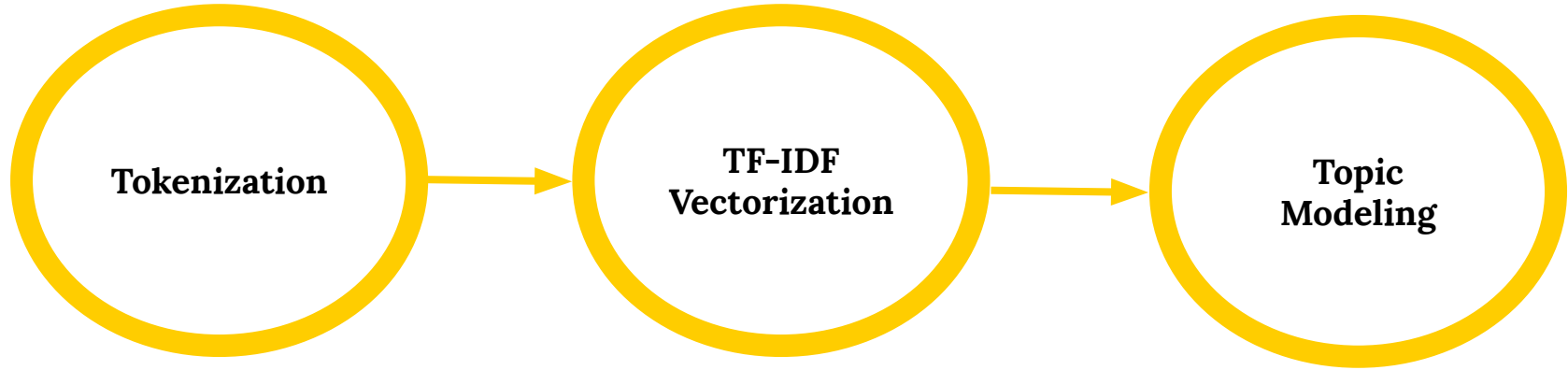


## Data Collection

- Airbnb listings details and reviews in collected in late 2017
- Data includes listing features and reviews
- Basic web crawler built for potential future endeavors



**Corpus and Document Definition:** For the purposes of our analysis a corpus was defined as all reviews for an individual listing and a document was defined as an individual review





# Natural Language Processing

## Overarching Goal

- ◉ Identify reviews that provide the most differentiating information about the listing

## Methods

- ◉ Input: TF-IDF vectors for all reviews listings with more than 100 reviews
- ◉ Approach: LDA followed by Rao's diversity coefficient
- ◉ Output: Reviews with the top-5 diversity coefficient

## Evaluation

- ◉ True/False: Information returned by algorithm contained information not referenced in first 2 pages of reviews.



## Topic Modeling Details

Rao's Diversity:

Gives a diversity coefficient based on the topic similarity and the proportion of words in topics from the document.

High diversity Score : More words from the document are in topics that are very dissimilar.

$$div^{(d)} = \sum_{i=1}^T \sum_{j=1}^T P(i|d)P(j|d)\delta(i, j)$$

## Common Themes: Location, Ambiance, Space Size

“ This place is in an **ideal location** and **centrally located** to all of Boston. The house was awesome and spacious. John was quick to respond and provided everything we needed. It was a very comfortable and cozy house. thanks ”

“ ....Small footprint but 4-floors of living/sleeping space. Great for **fitting in a bunch of people or family** for any occasion... ”

“ ....The house was **clean and beautiful**. **Plenty of space for a large group**. Multiple common areas to relax and chat.... ”

“ **Superb location** and a very **lovely and sleek** space. Everything amenity is thoughtfully provided. Lots of great details ”

# Similarities

Great place, John was an excellent host.

quaint while also being  
exceptionally  
convenient. It is clean,  
well-kept, elegantly  
decorated. The host  
offered thoughtful  
touches of hospitality

Great house and location!

....Just note that it is indeed a New England home, **so the spiraling staircase between floors can be challenging for older guests.** But other than that, John thought of just about everything...

# Differences





## Evaluation & Examples

Numerous examples were identified cases where information returned by the algorithm would not have been ascertained by review the first 30 reviews for a listing

“ The only problem is that the AC is in the living room, and it takes a bit for the apartment to cool down during the hot summer days.... ”

“ ....The only problem is that the AC is in the living room, and it takes a bit for the apartment to cool down during the hot summer days.... ”