TOPIC MODELING &
SENTIMENT ANALYSIS

Airbnb Reviews

airbnb

TABLE OF CONTENTS

OBJECTIVE

- Topic modeling with reviews will help Airbnb to improve customer experience

- Applying sentiment analysis to differentiate between negative, neutral, and positive reviews.

WORKFLOW

FROM START TO FINISH

Raw Data

from inside Airbnb

Data Preprocessing

preprocessing the raw data

Topic Modeling

NMF, LSA, LDA, and CoreX

Sentiment Analysis

Negative, positive, and neutral reviews

ABOUT OUR DATASET

From inside Airbnb

80k+ review

6 columns

Tools

Pandas

Numpy

Matplotlib

NLTK

pickle

Seaborn

Wordcloud

Data Preprocessing

1. REMOVE NON-ENGLISH WORDS
2. REMOVE NUMBERS, CAPTIAL LETTERS, PUNCTUATION, AND SPELL CHECKING
3. REMOVING STOP WORDS
4. TOKENIZING AND LEMMATIZING

Topic Modeling

Using Count Victorizer & TF-IDF Victorizer

NMF	LSA	LDA	CoreX
3 Topics	Not good	3 Topics	Our model 4 Topics

CoreX

BEST MODEL

ROOMS

Kitchen, bed, bathroom, window

LOCATION

distance, market, downtown, shop

HOST COMMUNICATIONS

friendly, generous, considrate, host

CLEANLINESS

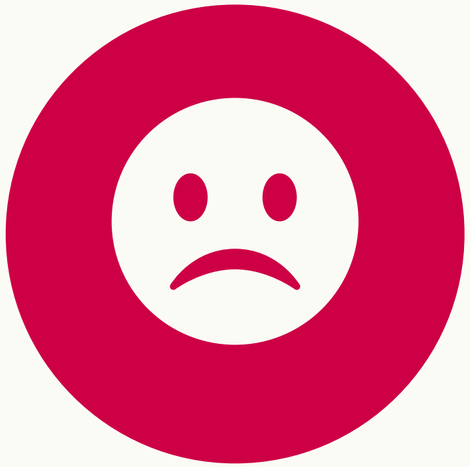
clean, bright, organize, tidy

SENTIMENT ANALYSIS

USING VADER

97.1%

2.6%



0.3%

FUTURE WORK

Scrapping the reviews from Airbnb

to get good result with the sentiment analysis

topic modeling on positive and negative reviews

to detect what are the topics that occur the most on positive and negative reviews

Build a recommendation system

to help Airbnb find listings that have similar reviews

THANK YOU!

APPENDIX

LDA WITH 3 TOPICS