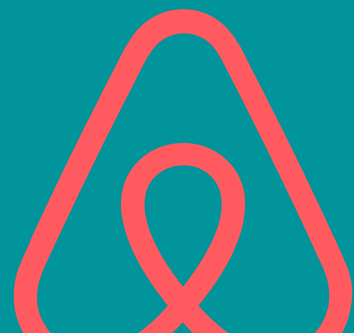


TOPIC MODELING &
SENTIMENT ANALYSIS

Airbnb Reviews





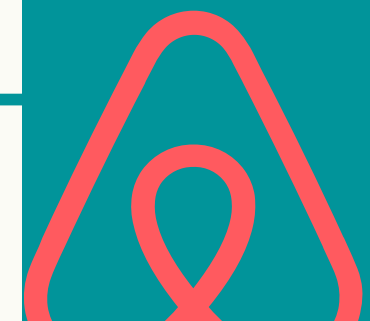
Objectives
Workflow
dataset
Tools
Data preprocessing
Topic Modeling
Sentiment analysis
Future work

TABLE OF CONTENTS



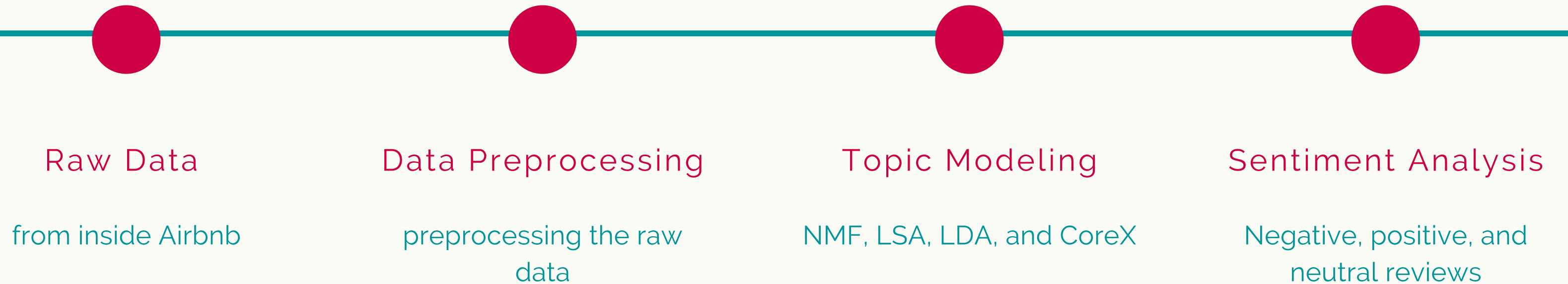
OBJECTIVE

- 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



ABOUT OUR DATASET



From inside Airbnb



80k+ review



6 columns



Tools

Pandas

Numpy

Matplotlib

NLTK

gensim

Seaborn

Wordcloud

Data Preprocessing



1

REMOVE NON-ENGLISH WORDS

2

REMOVE NUMBERS, CAPITAL LETTERS, PUNCTUATION, AND SPELL CHECKING

3

REMOVING STOP WORDS

4

TOKENIZING AND LEMMATIZING

Topic Modeling

Using Count Vectorizer & TF-IDF Vectorizer

NMF

3 Topics

LSA

Not good

LDA

3 Topics

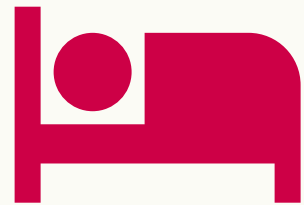
CoreX

Our model
4 Topics



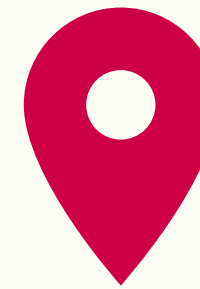
CoreX

BEST MODEL



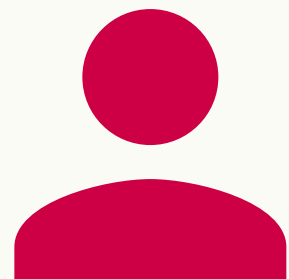
ROOMS

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

