HOTEL REVIEWS

NLP Topic Modeling and Hotel Recomender

“HOTEL CALIFORNIA”

The Back Story:

An International Hotel Chain approached us to study the possibility of setting up “Hotel California” chain in Europe.

They wanted us to get insights on the postive and negative comments for their Strategic Business Plan Department.

They also want us to prototype a Hotel Recommender for their application.

TABLE OF CONTENTS

01 Tools

Exploratory Data Analysis

Preprocess text
Count Vectorize

Topic Modeling:
LSA, NMF, LDA

Display Findings
Hotel Recommender

01 Tools
Topic Modeling Work Flow

Preprocess text

Remove numbers, capital letters and punctuation

Prepare data for modeling

Divide data into X and y data sets
Convert words into Vectors

Classification into topics

Using different models to classify the data into possible topics

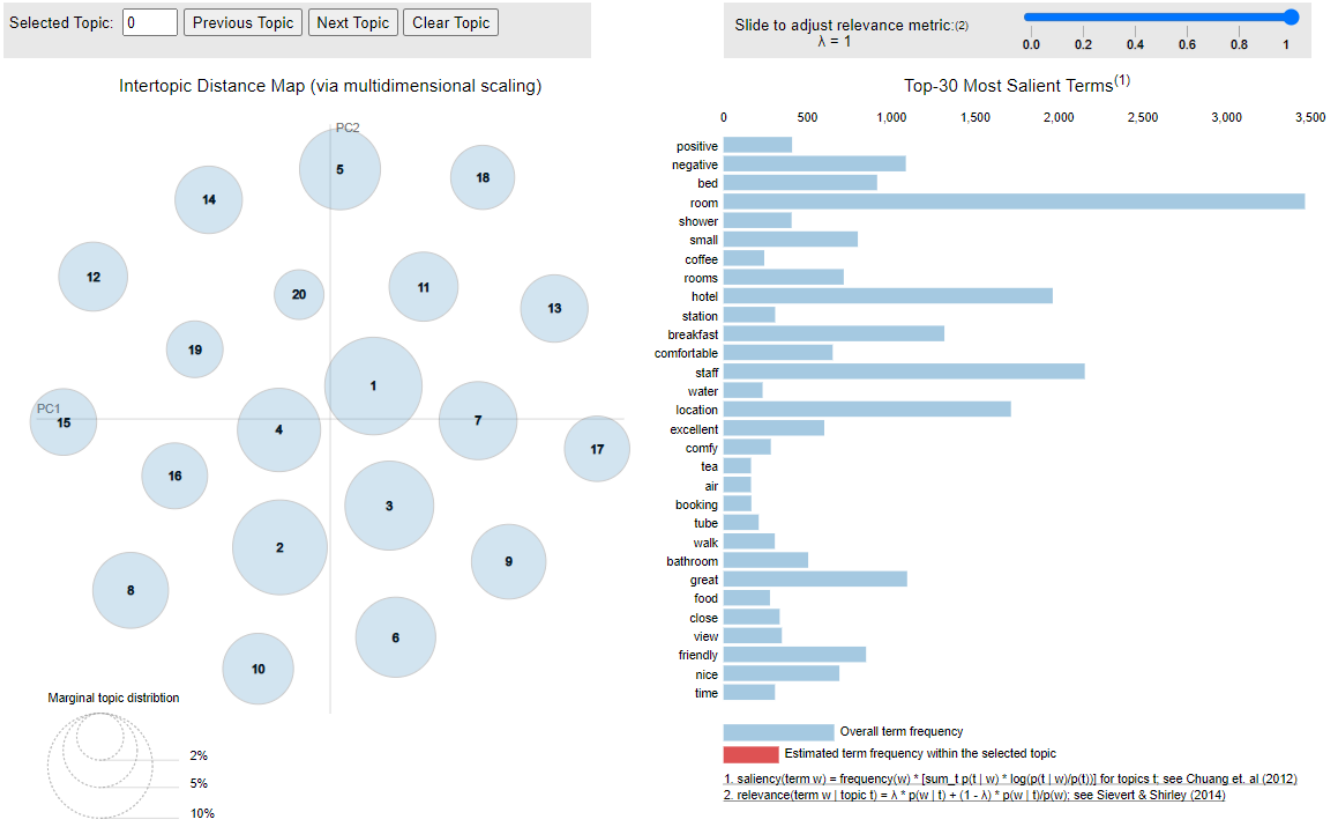
Intepret the data

Infer the possible meaning the topics are about

02 The DataSet by location

02 The DataSet by keywords

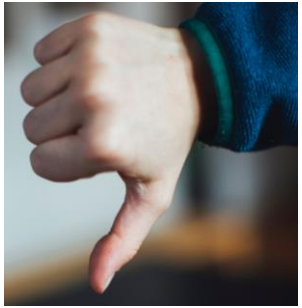
Out[149]:



03 NLP
Common Positive and Negative Comments

Positive Comments

1) Friendly and Helpful staff
2) Comfortable Rooms and Beds
3) Good Hotel Location
4) Excellent Breakfast
5) Clean Bathroom



Negative Comments

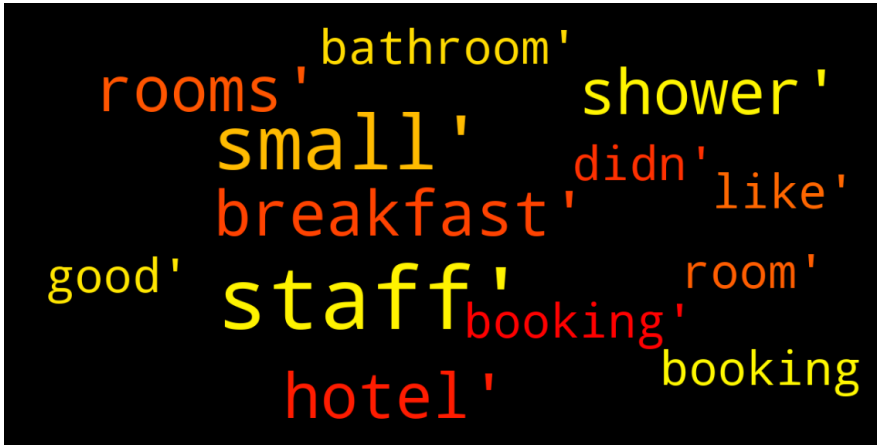
1) Small Hotel Room and Bed
2) Did not have breakfast
3) Lack of staff at night
4) Expensive breakfast,
5) Took long time to find booking

03 NLP
Latent Semantic Analysis (LSA)

Sklearn's TruncatedSVD With **Count Vectorizer**

Positive words

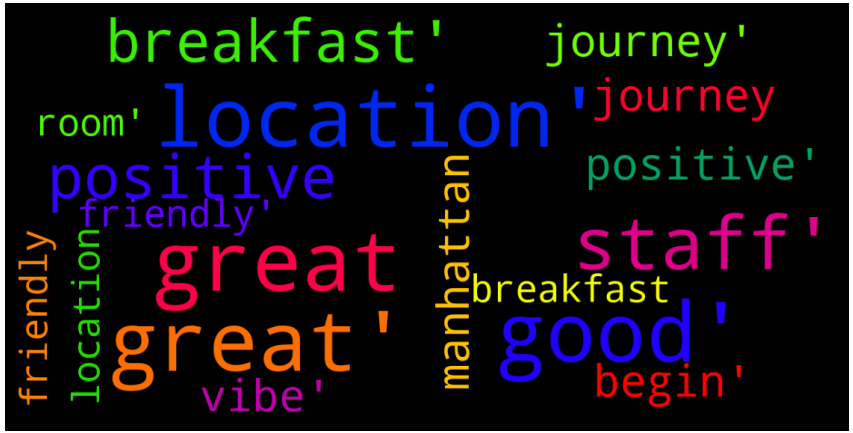
Negative words



03 NLP
Latent Semantic Analysis (LSA)

Using Sklearn's TruncatedSVD With **TFIDF Vectorizer**

Positive words



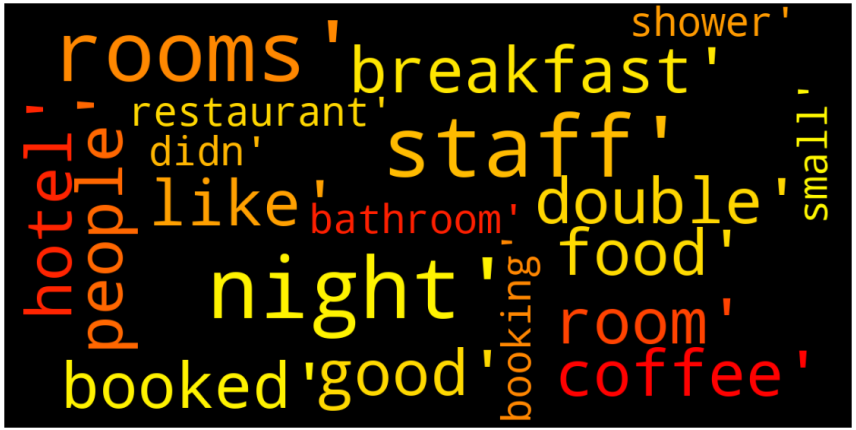
Negative words

03 NLP
Non-Negative Matrix Factorization (NMF)

NMF model With **Count Vectorizer**

Positive words

Negative words



03 NLP
Latent Dirichlet Allocation(LDA)

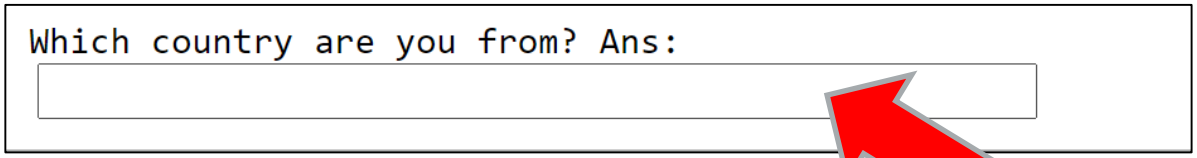
Using Gensim and LDA With **Count Vectorizer**

Positive words

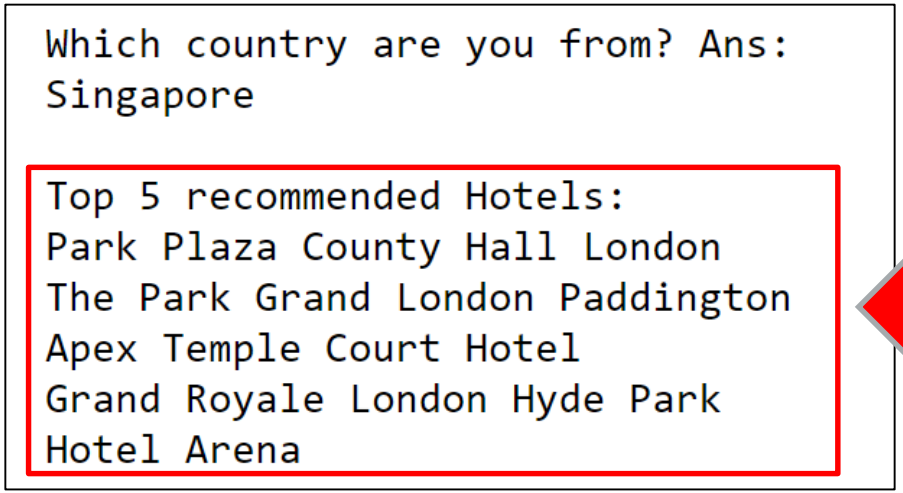
Negative words

04 Hotel Recommender

04 Hotel Recommender



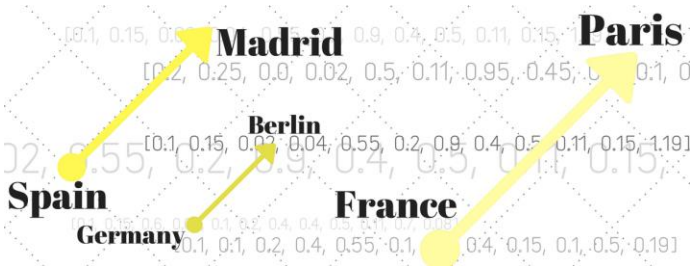
Input Country from List



Recommend Top 5 Hotel Choices

05 Future Improvements

How can we do better?



Explore Word Embedding

Use Word2vec, Glove, Bert for topic modeling

Explore Clusters in each Country

Use Kmeans on individual location to explore potiential clusters

Hotel Recommender

Build a better Hotel Recommender with Country Search option

What do you like or dislike in your last hotel stay?

THANKS!

DOES ANYONE HAVE ANY QUESTIONS?

Follow me on linkedin:
https://www.linkedin.com/in/andrew-tan-41878671
https://github.com/guitarfly78