



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 positive 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

List of tools used for this project

04

Hotel Recommender

Where are you from?
Top 5 recommended Hotels:

02

Exploring DataSet

Quick look at the DataSet

05

Future Improvements

How can we do better?

03

Using NLP

Topic Modeling with
LSA, NMF, LDA



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]:

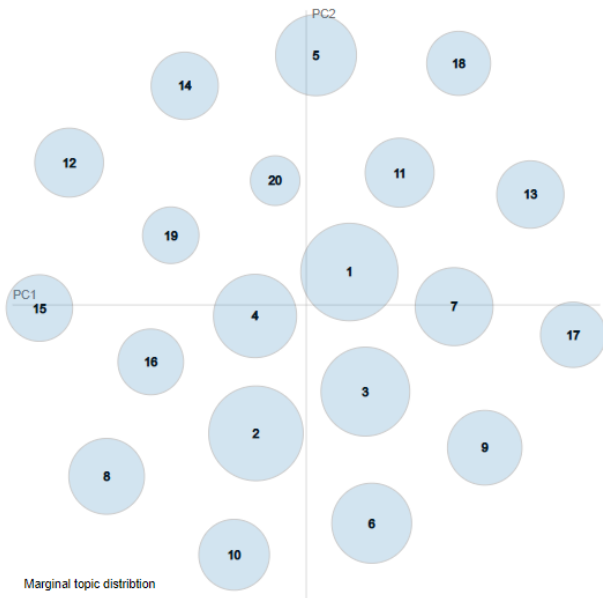
Selected Topic:

Slide to adjust relevance metric:(2)

$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1

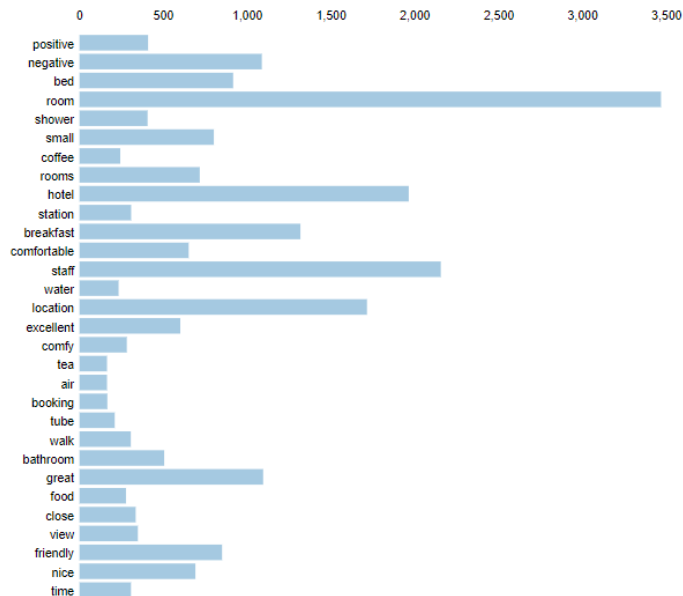
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Salient Terms⁽¹⁾



Overall term frequency

Estimated term frequency within the selected topic

1. $saliency(term\ w) = frequency(w) * [\sum_t p(t|w) * \log(p(t|w)/p(t))]$ for topics t : see Chuang et. al (2012)

2. $relevance(term\ w | topic\ t) = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$: see Sievert & Shirley (2014)

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

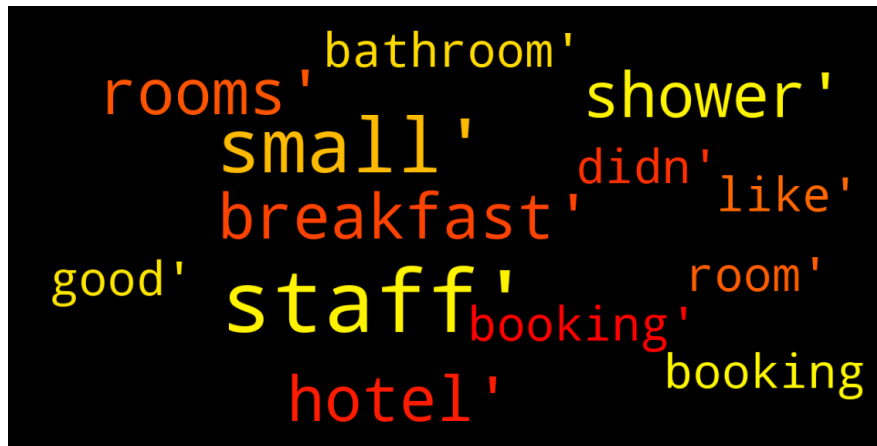
Latent Semantic Analysis (LSA)

Sklearn's TruncatedSVD With **Count Vectorizer**

Positive words

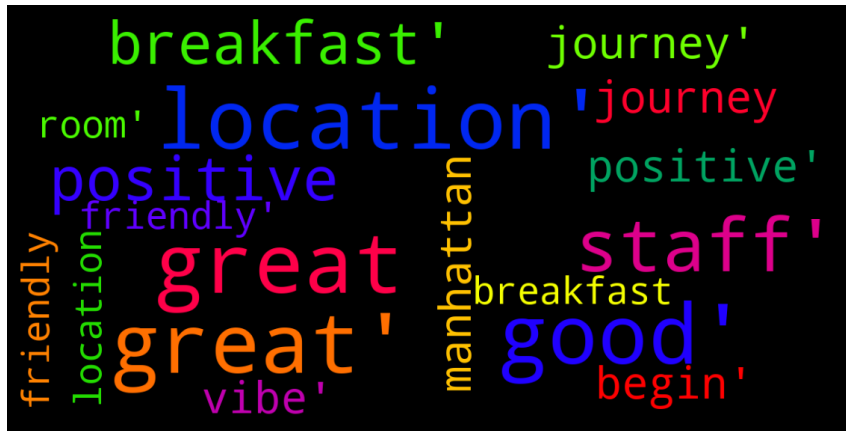


Negative words



Latent Semantic Analysis (LSA)

Positive words



night' uncomfortable' bathroom' hotel' didn't price' room' little' booking' small' sofa' rooms' breakfast' walk'

Non-Negative Matrix Factorization (NMF)

Positive words



rooms' breakfast' shower' restaurant' didn't staff' hotel' people' like' bathroom' double' small' night' food' room' coffee' booked' good' booking'

03 NLP

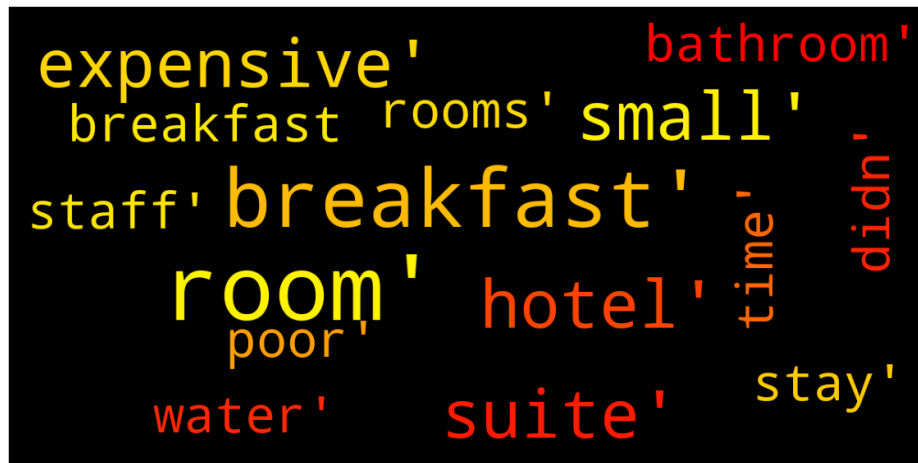
Latent Dirichlet Allocation(LDA)

Using Gensim and LDA With Count Vectorizer

Positive words



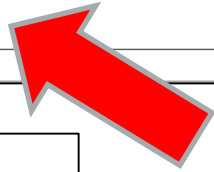
Negative words



[illegible]

04 Hotel Recommender

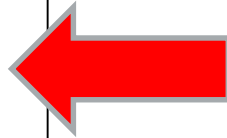
Which country are you from? Ans:



Input Country from List

Which country are you from? Ans:
Singapore

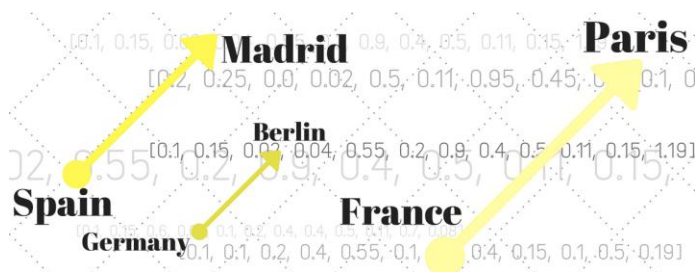
Top 5 recommended Hotels:
Park Plaza County Hall London
The Park Grand London Paddington
Apex Temple Court Hotel
Grand Royale London Hyde Park
Hotel Arena



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 potential 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>



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution.