Regionality-Centric Sentiment and Topic Modeling Pattern Analysis: Luxury Hotel in Europe

Wonha Shin / wshin7@ur.rochester.edu
Raymond Yu / cyu45@ur.rochester.edu

Agenda

A. Problem motivation

B. Dataset Overview & Preprocessing

C. Topic Modelings & Algorithms

D. Findings

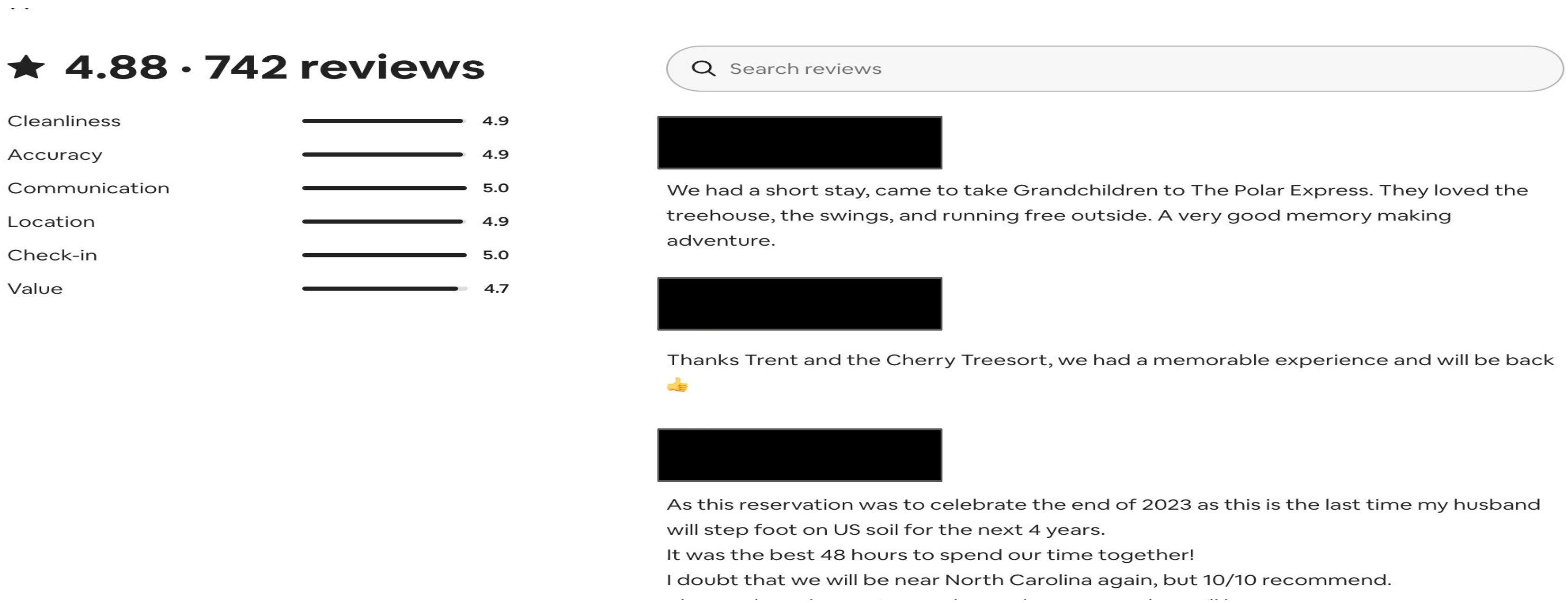
E. Conclusion and Potential Application

Problem Motivation & Statement

- As a foodie, we always count on the rating score and reviews, but sometimes got tricked
    - Google Review

Problem Motivation & Statement

- As a travel lover, we count on various booking platform’s rating and comments
  - Airbnb, Booking.com, Hotel.com, Trivago, Trip.com



Problem Motivation & Statement

We would like to find and analyze is there any interesting findings with the following:

- Contents in past reviews and comments, instead of only counting on rating
- What is the metric priority for people staying in the luxury hotels
- Whether different nationality has a different perspective on choosing a hotel to stay, and what is the sequence of properties each nationality prefer/does not prefer

Agenda

A. Problem motivation

B. Dataset Overview & Preprocessing

C. Topic Modelings & Algorithms

D. Findings

E. Conclusion and Potential Application

Overview & Preprocessing - Dataset and Cleansing

- Data Sources:
  - Kaggle: “515K Hotel Reviews Data in Europe”
  - Original sources: booking.com
- Dataset Overview:
  - Rows: 515K rows of data
  - Columns: 17 different columns
  - Unique hotels: 1,493 hotels across France, United Kingdom, Netherlands, Italy, Austria, Spain in Europe
- Data Cleansing
  - Space trimming for string data type
  - Drop the missing data which affect the outcome
  - Discretization (Concept Hierarchy)

Dataset overview & Preprocessing - Column Overview

- Column Overview
  - Discretization (Concept Hierarchy)
    - Nationality -> Sub-region -> Region

Hotel_Address	**Positive_Review**	Additional_Number_of_Scoring
Review_Date	Review_Total_Positive_Word_Counts	lat
Average_Score	Reviewer_Score	lng
Hotel_Name	Total_Number_of_Reviews_Reviewer_Has_Given	**Region**
Reviewer_Nationality	Total_Number_of_Reviews	**Sub-region**
Negative_Review	Tags	
Review_Total_Negative_Word_Counts	days_since_review	

Dataset overview & Preprocessing - Column Overview

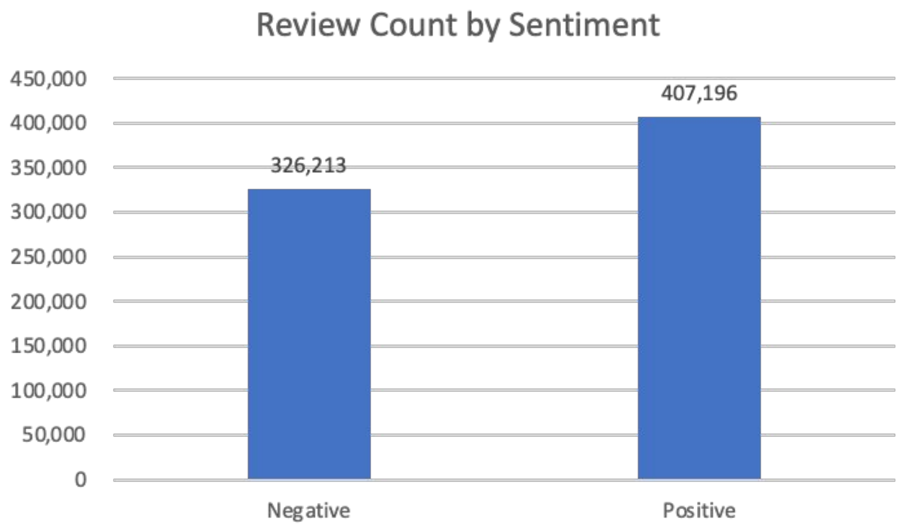
To structure our analysis, we aligned our regional and sub-regional classifications with those defined by Google Maps as follows.

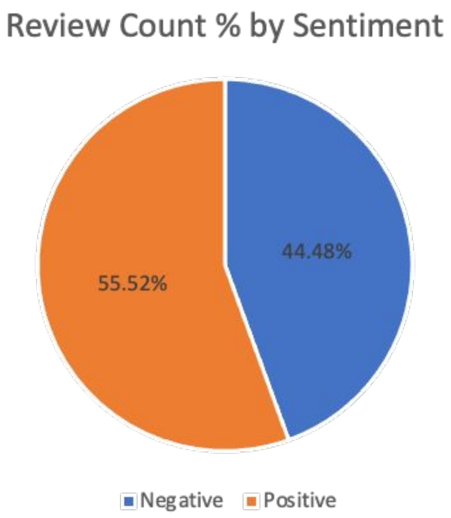
- Regions : {'Africa', 'Americas', 'Antarctica', 'Asia', 'Europe', and 'Oceania'}
- Sub-regions :{ 'Antarctica', 'Australia and New Zealand', 'Central Asia', 'Eastern Africa', 'Eastern Asia', 'Eastern Europe', 'Latin America and the Caribbean', 'Melanesia', 'Micronesia', 'Middle Africa', 'Northern Africa', 'Northern America', 'Northern Europe', 'Polynesia', 'South-eastern Asia', 'Southern Africa', 'Southern Asia', 'Southern Europe', 'Western Africa', 'Western Asia', and 'Western Europe'. }

Dataset overview & Preprocessing - Reviews by Geography

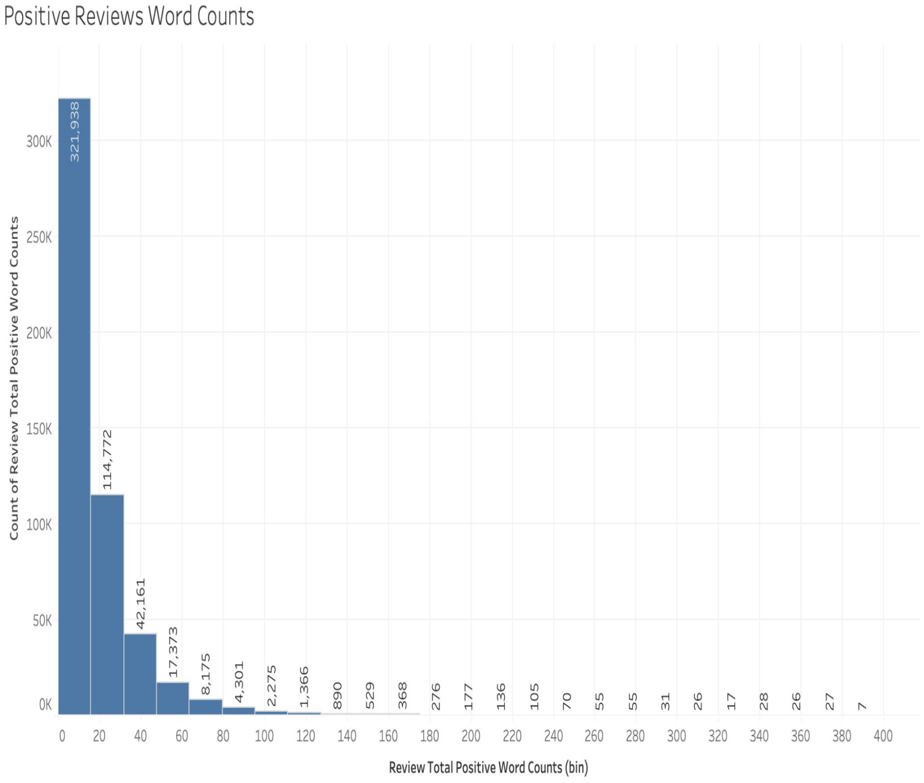
Hotel Geography

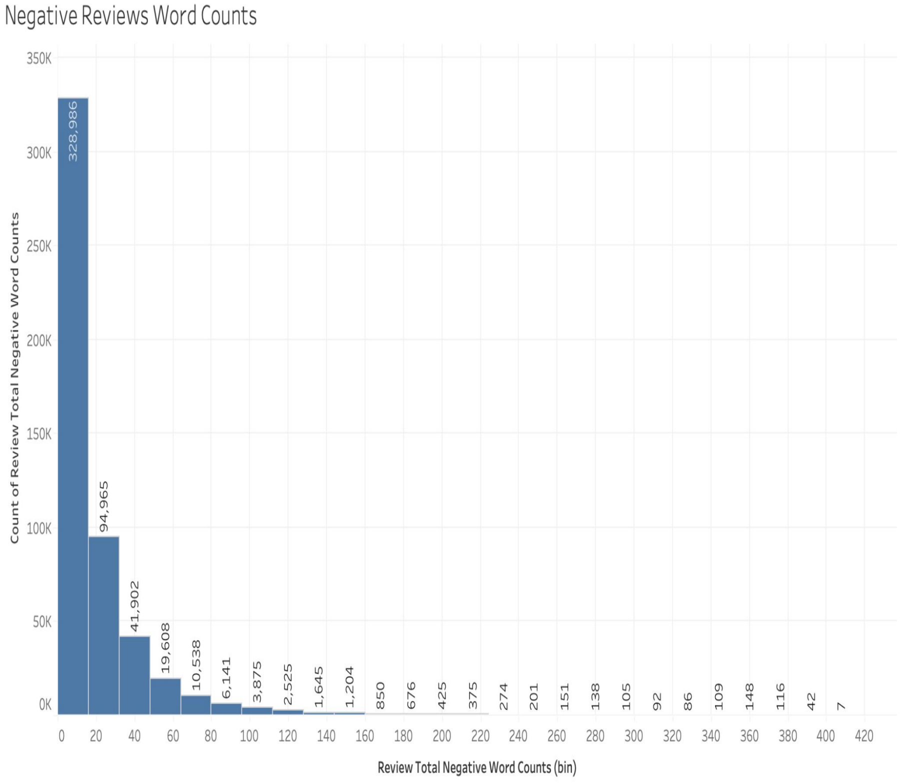
Dataset overview & Preprocessing - Sentiment Count of Reviews





Dataset overview & Preprocessing - Word Counts for Positive/Negative





UNIVERSITY of ROCHESTER

Dataset overview & Preprocessing - Preprocessing Overview

- Merged into Combined One Review Dataset
- NLP Review Cleansing
  - Lower case only
  - Tokenization
  - Lemmatization
  - Stopwords Removal

Dataset overview & Preprocessing - Merging into Combined Dataset

- Merged into Combined One Review Dataset
    - to conduct a comprehensive topic modeling

Negative_Review	Positive_Review	Merged_Review
Rooms are nice but for elderly a bit difficult...	Location was good and staff were ok It is cute...	Rooms are nice but for elderly a bit difficult...
My room was dirty and I was afraid to walk bar...	Great location in nice surroundings the bar an...	My room was dirty and I was afraid to walk bar...
′ou When I booked with your company on line yo...	Amazing location and building Romantic setting	You When I booked with your company on line yo...

Dataset overview & Preprocessing - Cleansing of Datasets

- Cleansing Reviews Datasets
  - Lower case
  - Tokenization: Breaking down text into individual tokens
    - ex) The runner runs so fast ➡️ the/ runner/ runs / so / fast
  - Lemmatization: Reducing words to their base or root form

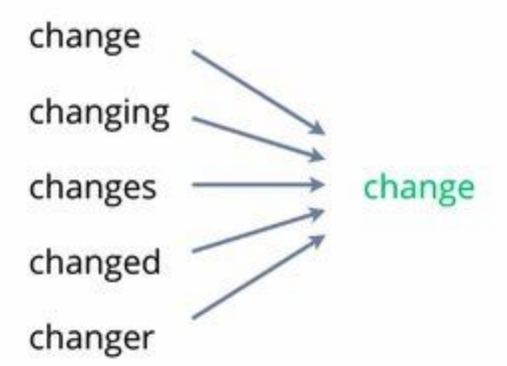


Image Source: https://www.quora.com/What-is-lemmatization-in-NLP

Dataset overview & Preprocessing - Cleansing of Datasets

- Cleansing Reviews Datasets
  - Stopwords Removal : A set of commonly used words in a language (ex) a, the, I, am, are, some, their, themselves, such
  - NLTK Stopwords Library & manually listed and added

```
stop_words = set(stopwords.words('english'))
stop_words_nouns = ["hotel", "place", "room", "5", "4", "3", "2", "1",
                    "wa", "u", "one", "two",
                    "everything", "nothing", "thing"]
stop_words_verbs = ["enjoy", "would", "like", "could",
                    "get", "stay", "visit", "go", "think", "make"]
stop_words_adverbs = ["highly", "within", "really", "well", "even",
                      "though", "bit", "little", "extremely","definitely",
                      "much", "also", "always", "never", "often", "usually"]
stop_words_adjectives =["great", "positive", "negative", "good", "nice", "lovely"
                        "fantastic", "excellent",b"amazing", "wonderful", "super"
                        "horrible", "terrible", "disappointing", "awful", "poor",
```

Dataset overview & Preprocessing - Cleansing of Datasets

- Cleansing Reviews Datasets
    - A set of commonly used words in a language (ex) a, the, I, am, are, some, their, themselves, such
    - Stop Words Removal - NLTK Stopwords Library & Manually listed library

```
data['Merged_Review'][0:8]
```

```
0    I am so angry that i made this post available ...
1                                                 NaN
2    Rooms are nice but for elderly a bit difficult...
3    My room was dirty and I was afraid to walk bar...
4    You When I booked with your company on line yo...
5    Backyard of the hotel is total mess shouldn t ...
6    Cleaner did not change our sheet and duvet eve...
7    Apart from the price for the brekfast Everythi...
Name: Merged_Review, dtype: object
```

```
data['Merged_Review'][0:8]
```

```
0    [angry, made, post, available, via, possible, ...
1                                                   []
2    [elderly, difficult, story, narrow, step, ask,...
3    [dirty, afraid, walk, barefoot, floor, looked,...
4    [booked, company, line, showed, picture, thoug...
5    [backyard, total, mess, happen, star, restaura...
6    [cleaner, change, sheet, duvet, everyday, made...
7    [apart, price, brekfast, location, set, park, ...
Name: Merged_Review, dtype: object
```

Agenda

A. Problem motivation & Statement

B. Overview & Preprocessing

C. Topic Modelings & Algorithms

D. Findings

E. Conclusion and Potential Application

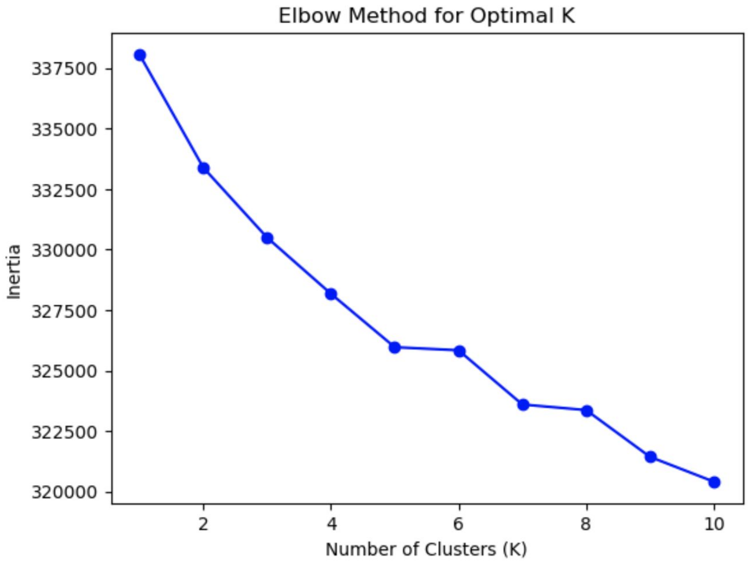
UNIVERSITY of ROCHESTER

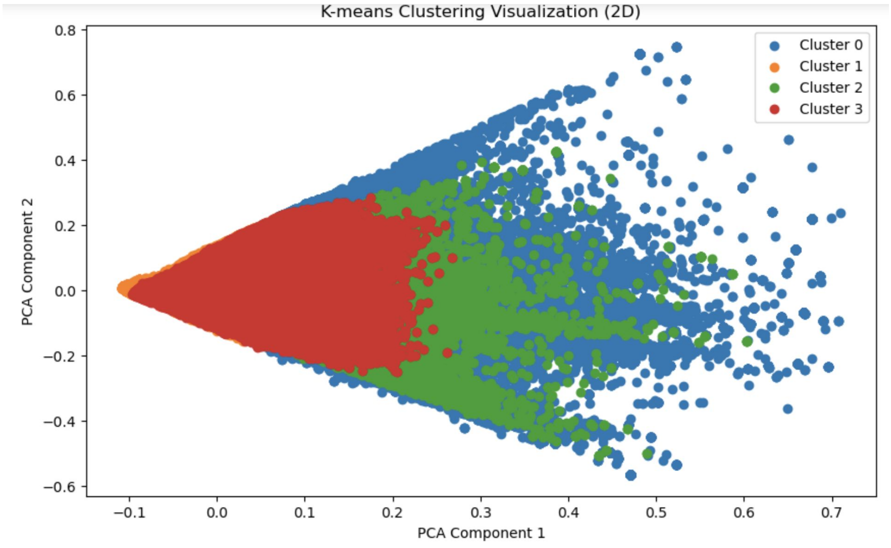
Topic Modelings & Algorithms - Three Methods to select

- For specific topic modeling, three prevalent methods are used
  - K-means Clustering
  - TF-IDF
  - LDA Topic Modeling with Gensim

Topic Modelings & Algorithms - K means Clustering

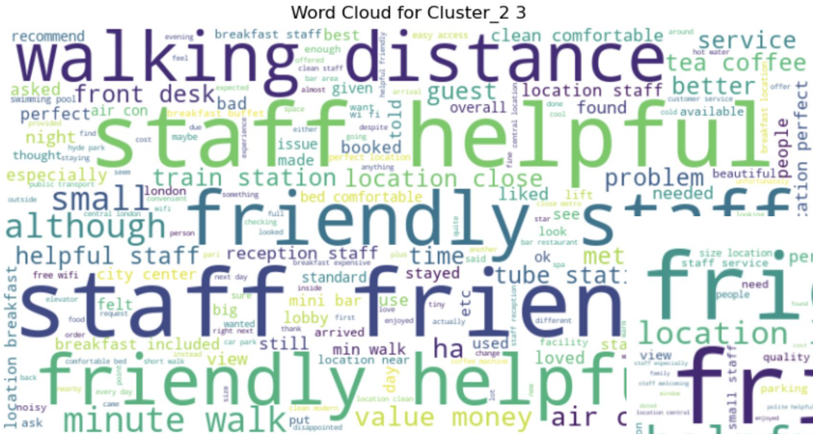
- K means is bad at modeling topics of datasets. **HoW bAd It Is?**
- After conducting Elbow test, found the appropriate number K as 4

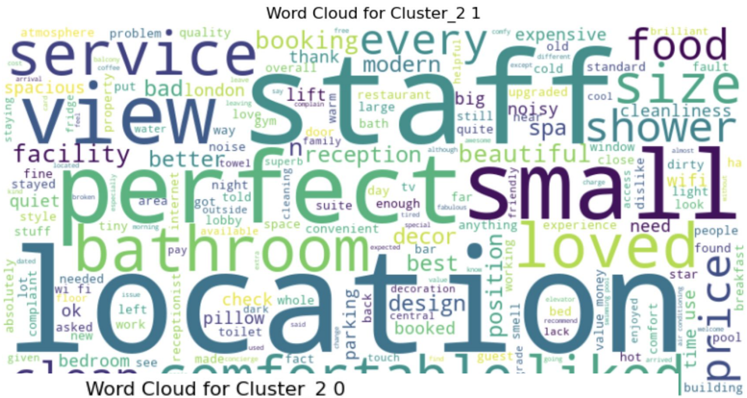


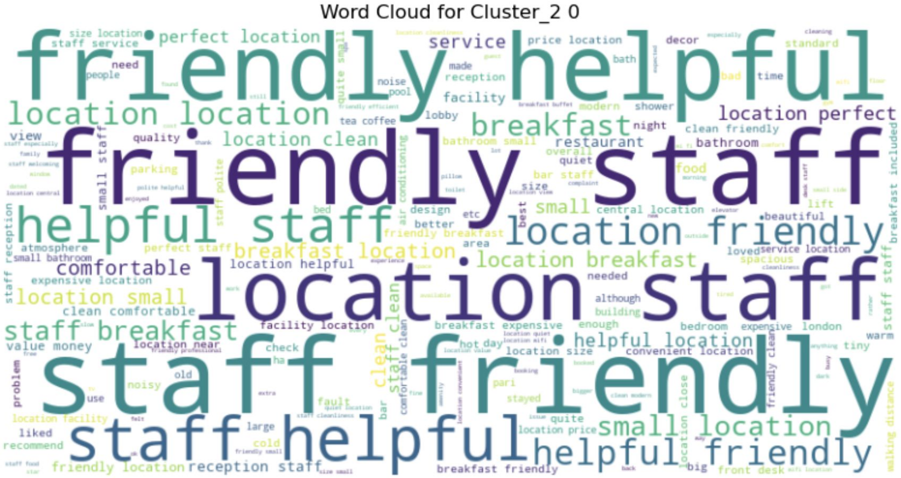


Topic Modelings & Algorithms - K means Clustering

- Due to the limitation of K-means, meaningless cluster of words







Topic Modelings & Algorithms - TF-IDF / NMF

- TF-IDF (Term Frequency-Inverse Document Frequency)
  - Term Frequency (TF): how frequently a term occurs in a document.
  - Inverse Document Frequency (IDF): the importance of a term across a set of documents.
  - TF-IDF is combined and obtained by multiplying the TF and IDF scores of a term.
    how frequently they appear across multiple documents.

- NMF(Non-Negative Matrix Factorization)
  - a dimensionality reduction and matrix factorization technique
  - for interpretation of topics

Topic Modelings & Algorithms - TF-IDF / NMF

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
staff	location	bed	breakfast	close	small	clean	service	perfect	friendly
helpful	staff	comfortable	expensive	station	bathroom	comfortable	food	beautiful	staff
reception	facility	comfy	choice	bar	size	modern	customer	every	efficient
polite	price	shower	included	restaurant	quite	spacious	slow	loved	comfortable
pleasant	central	bathroom	buffet	view	shower	quiet	bar	london	professional
welcoming	size	pillow	price	night	space	tidy	restaurant	view	atmosphere
attentive	value	double	selection	area	double	new	quality	absolutely	beautiful
kind	cleanliness	size	delicious	time	side	bathroom	reception	recommend	stuff
food	convenient	hard	variety	bathroom	bedroom	big	bad	best	welcome
facility	comfort	uncomfortable	better	walk	old	value	concierge	trip	quiet
professional	wifi	large	quality	shower	single	wifi	charge	back	warm

Topic Modelings & Algorithms - LDA Topic Modeling with Gensim

- LDA (Latent Dirichlet Allocation)
  - A powerful tool identifying central themes within textual data
  - Uses probabilistic modeling to analyze extensive text, determining the distribution of topics across documents and identifying key words associated with each topic.

- Gensim
  - Topic Modeling and Natural Language Processing
  - Useful for the semantic topics

Topic Modelings & Algorithms - LDA Topic Modeling with Gensim

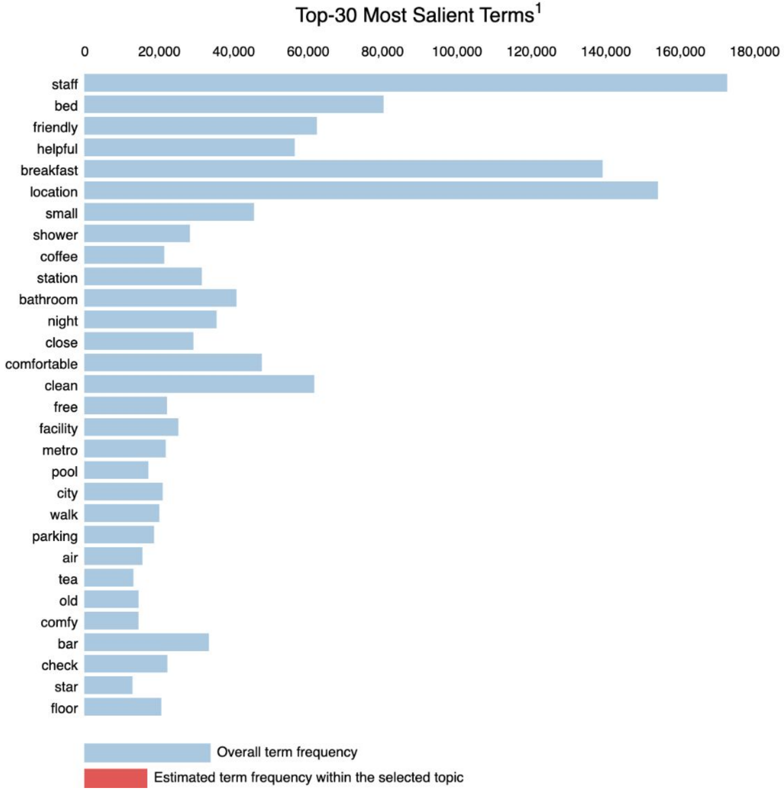
- Established a model with different parameters with Gensim
  - After running (over) 50 times of modeling, found the most appropriate parameters
    - The highest coherence level (objective)
    - The most meaningful topic separations (subjective)
  - A brief explanation of each parameter is as follows:
    - num_topics: The number of topics to be generated
    - chunksize: The number of documents to be processed in a single training session
    - passes: The number of training passes over the entire corpus
    - iterations: The number of iterations per document

Topic Modelings & Algorithms - Topic Selection

- Collected rating metrics from five popular and well-known booking platforms, including Airbnb, Trip.com, Hotels.com, Trivago, and Booking.com

#	Airbnb	Trip.com	Hotel.com	Trivago	Booking.com
1	Cleanliness	Cleanliness	Cleanliness	Cleanliness	Cleanliness
2	Location	Location	NA	Location	Location
3	Communication	Service	Staff & Service	Service	Staff
4	Value	NA	NA	Value for money	Value for money
5	NA	Amenities	Amenities	Building	Facilities
6	Check-in		Property conditions & facilities	Rooms	
7	Accuracy			Comfort	
8				Food	

Topic Modelings & Algorithms - LDA Visualization for Topic Modeling

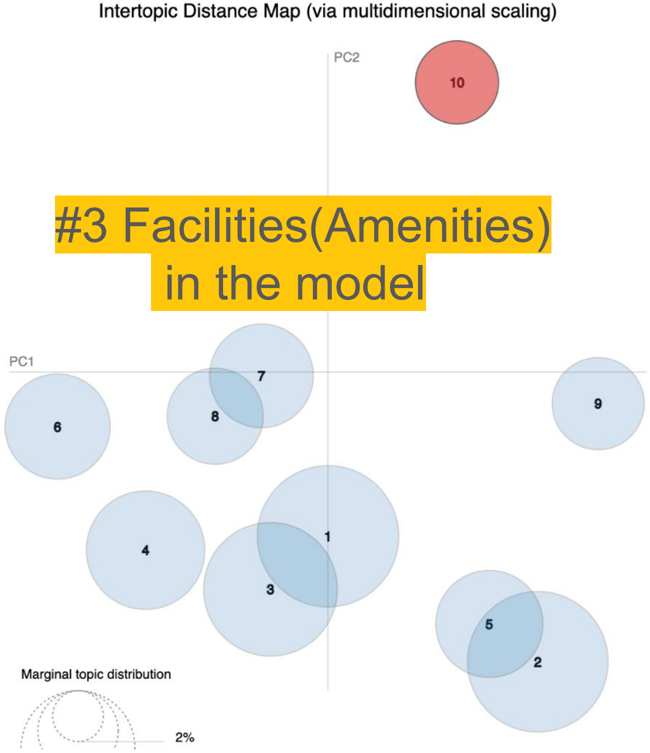


Topic Modelings & Algorithms - Topic Selection

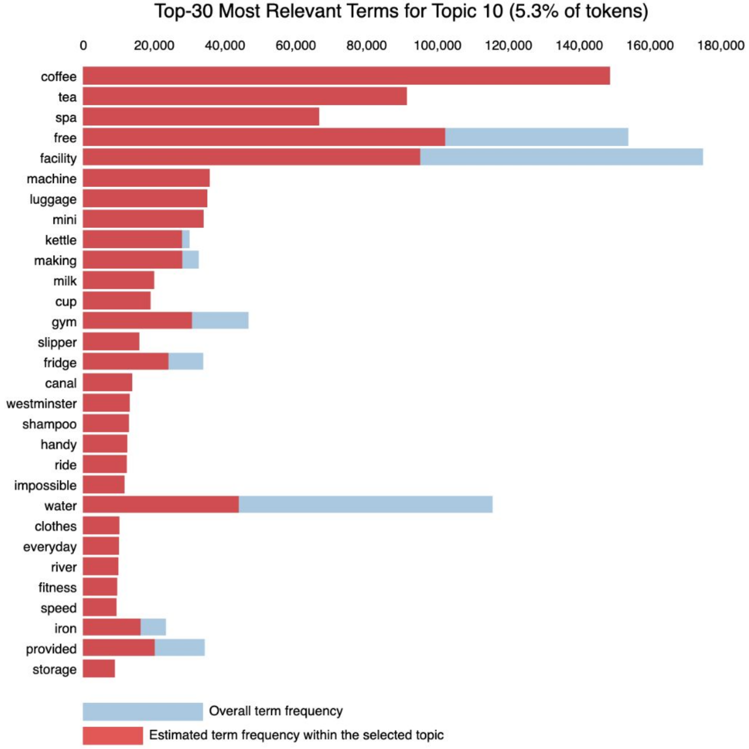
- Based on the references, and running models with several trials differentiating whole parameters including topic numbers, we selected 10 topics

#	Selection	Examples
1	Location	station, close, location, metro, city, walk, mitue, near
2	Room Conditions	bed, small, comfortable, comfy, size
3	Facilities (Amenities)	coffe, free, facility, tea, spa, wifi, bar, machine
4	Meals	breakfast, bar, food, restaurant service, area, buffet
5	Customer Services	staff, day, time, check (in / out), reception, service
6	Cleanliness	shower, bathroom, air, bed, water, clean
7	Friendly Staff	staff, friendly, helpful, service
8	Value	star, value, small, money
9	Easy Process	booking, pay, booked, paid
10	Noise and Comfort	noise, sleep, hear, sound, construction

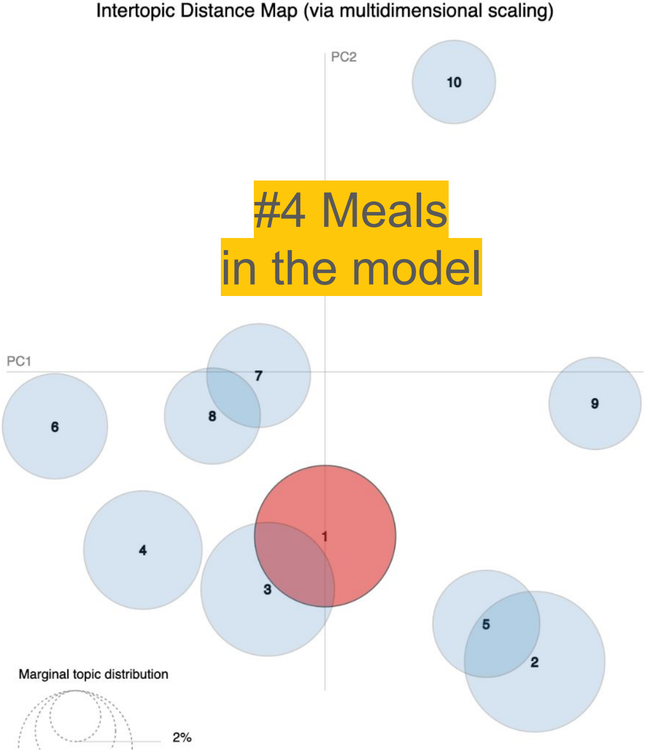
Topic Modelings & Algorithms - LDA Visualization for Topic Modeling



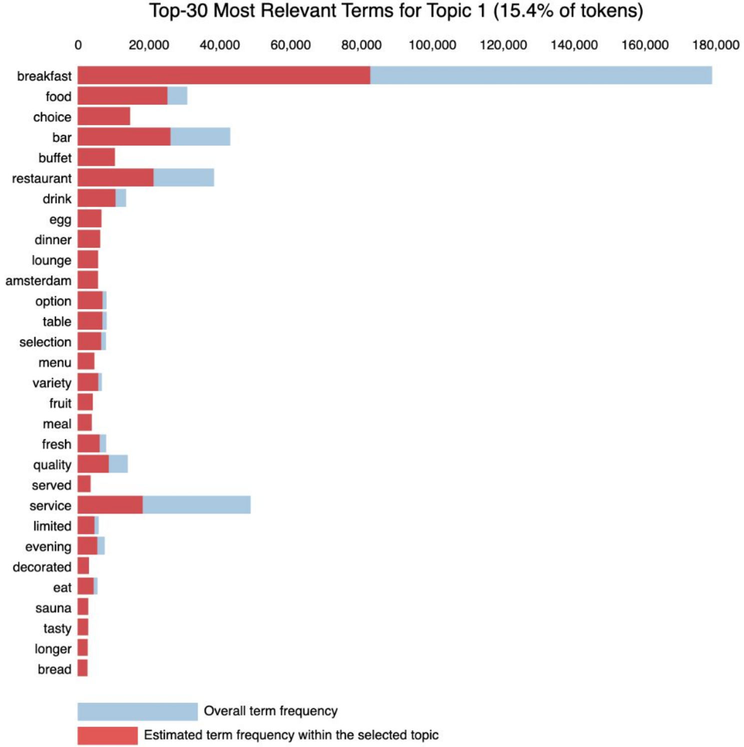
#3 Facilities(Amenities) in the model



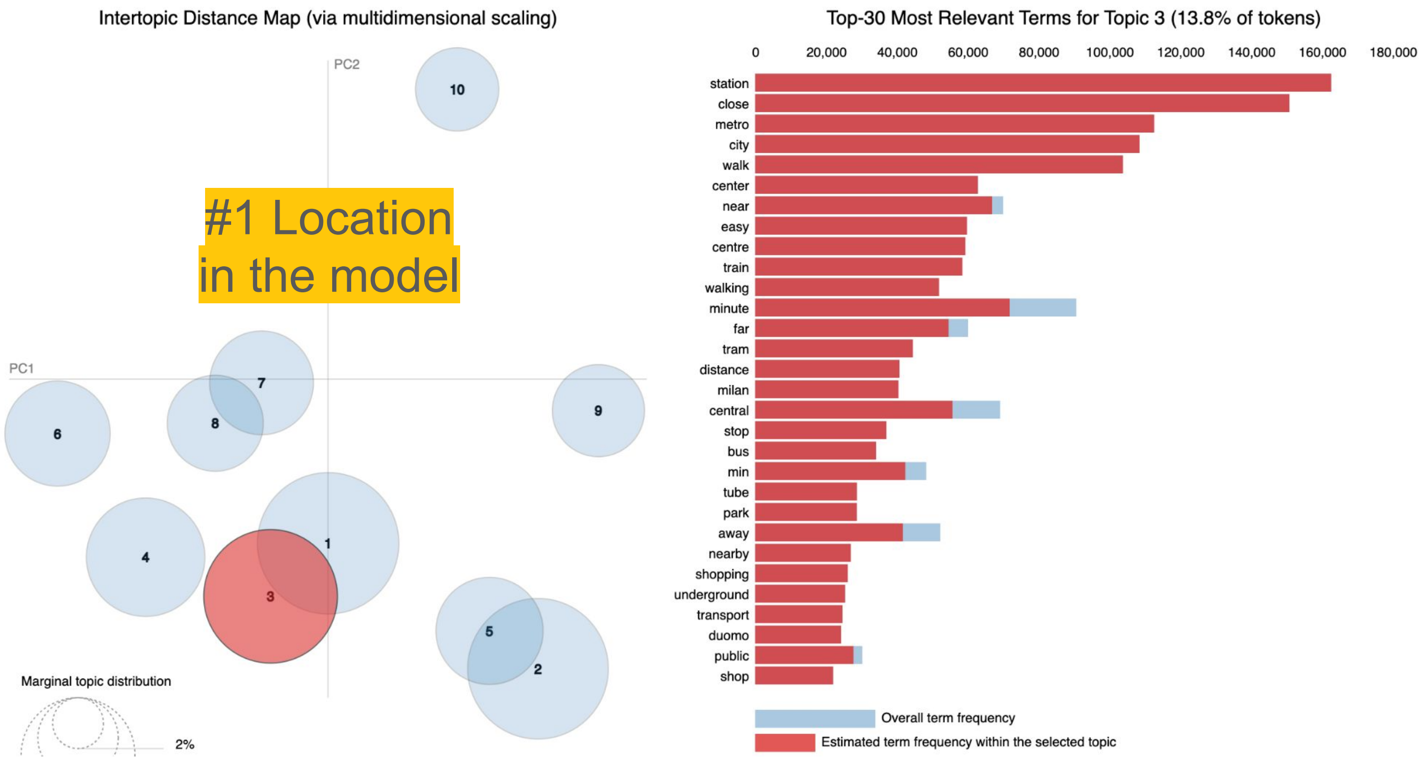
Topic Modelings & Algorithms - LDA Visualization for Topic Modeling



#4 Meals
in the model



Topic Modelings & Algorithms - LDA Visualization for Topic Modeling



Topic Modelings & Algorithms - LDA Topic Modeling with Gensim

- With the established model, applied to the reviews respectively
  - Merged Reviews, Positive Reviews, Negative Reviews

    ex) Merged Review : Top 7 Topics and give them labels accordingly

	Merged_Review	Top_Topics	Top_Topics_Labels
0	[angry, made, post, available, via, possible, ...	[(9, 0.29813027), (8, 0.24832857), (4, 0.22751...	[Noise and Comfort, Easy Process, Customer Ser...
1	[]	[(0, 0.1), (1, 0.1), (2, 0.1), (3, 0.1), (4, 0...	[Location, Room Conditions, Facilities, Meals,...
2	[elderly, difficult, story, narrow, step, ask,...	[(2, 0.415281), (7, 0.20591322), (9, 0.0935646...	[Facilities, Value, Noise and Comfort, Meals, ...
3	[dirty, afraid, walk, barefoot, floor, looked,...	[(4, 0.38528657), (9, 0.19621842), (5, 0.18289...	[Customer Services, Noise and Comfort, Cleanli...
4	[booked, company, line, showed, picture, thoug...	[(8, 0.61148167), (4, 0.24193136), (3, 0.09421...	[Easy Process, Customer Services, Meals, Noise...
5	[backyard, total, mess, happen, star, restaura...	[(0, 0.28715238), (7, 0.2293834), (6, 0.218429...	[Location, Value, Friendly Staff, Meals, Custo...
6	[cleaner, change, sheet, duvet, everyday, made...	[(1, 0.27602145), (4, 0.20604055), (2, 0.20507...	[Room Conditions, Customer Services, Facilitie...
7	[apart, price, brekfast, location, set, park, ...	[(3, 0.48896974), (6, 0.2711778), (0, 0.189831...	[Meals, Friendly Staff, Location]
8	[]	[(0, 0.1), (1, 0.1), (2, 0.1), (3, 0.1), (4, 0...	[Location, Room Conditions, Facilities, Meals,...
9	[aircondition, noise, hard, sleep, night, big,...	[(9, 0.37672615), (1, 0.2406975), (3, 0.187581...	[Noise and Comfort, Room Conditions, Meals, Lo...

Topic Modelings & Algorithms - LDA Topic Modeling with Gensim

- With the established model, applied to the reviews respectively
  - Merged Reviews, Positive Reviews, Negative Reviews

    ex) Merged Review : Top 7 Topics and give them labels accordingly

	Negative_Review_2	Top_Topics_Neg	Top_Topics_Labels_Neg
0	[angry, made, post, available, via, possible, ...	[(9, 0.30675742), (4, 0.2258605), (3, 0.097307...	[Noise and Comfort, Customer Services, Meals, ...
1	[]	[(7, 0.5499771), (8, 0.05001126), (0, 0.050001...	[Value, Easy Process, Location, Room Condition...
2	[elderly, difficult, story, narrow, step, ask,...	[(9, 0.40442112), (4, 0.25689125), (0, 0.09865...	[Noise and Comfort, Customer Services, Locatio...
3	[dirty, afraid, walk, barefoot, floor, looked,...	[(4, 0.25055608), (9, 0.24485967), (8, 0.13539...	[Customer Services, Noise and Comfort, Easy Pr...
4	[booked, company, line, showed, picture, thoug...	[(3, 0.32847464), (9, 0.19419841), (4, 0.18155...	[Meals, Noise and Comfort, Customer Services, ...
5	[backyard, total, mess, happen, star]	[(4, 0.49579525), (9, 0.22788627), (7, 0.08638...	[Customer Services, Noise and Comfort, Value, ...
6	[cleaner, change, sheet, duvet, everyday, made...	[(3, 0.37577936), (9, 0.3146134), (8, 0.151607...	[Meals, Noise and Comfort, Easy Process, Room ...
7	[apart, price, brekfast]	[(9, 0.5585725), (3, 0.18862244), (2, 0.182787...	[Noise and Comfort, Meals, Facilities, Friendl...
8	[picture, show, clean, actual, quit, dirty, ou...	[(9, 0.28636014), (8, 0.27139696), (5, 0.13922...	[Noise and Comfort, Easy Process, Cleanliness,...
9	[aircondition, noise, hard, sleep, night]	[(6, 0.23211722), (9, 0.20946416), (5, 0.19399...	[Friendly Staff, Noise and Comfort, Cleanlines...

UNIVERSITY of ROCHESTER

Topic Modelings & Algorithms - LDA Topic Modeling with Gensim

- With the established model, applied to the reviews respectively
  - Merged Reviews, Positive Reviews, Negative Reviews

    ex) Merged Review : Top 7 Topics and give them labels accordingly

	Positive_Review_2	Top_Topics_Pos	Top_Topics_Labels_Pos
0	[park, outside, beautiful]	[(8, 0.4656045), (4, 0.26420465), (9, 0.192398...	[Easy Process, Customer Services, Noise and Co...
1	[real, complaint, location, surroundings, amen...	[(9, 0.3381056), (4, 0.25318256), (8, 0.180592...	[Noise and Comfort, Customer Services, Easy Pr...
2	[location, staff, ok, cute, breakfast, range, ...	[(8, 0.3354647), (4, 0.29399502), (3, 0.207109...	[Easy Process, Customer Services, Meals, Noise...
3	[location, surroundings, bar, restaurant, outd...	[(8, 0.3362651), (4, 0.17547598), (9, 0.136365...	[Easy Process, Customer Services, Noise and Co...
4	[location, building, romantic, setting]	[(7, 0.3500131), (4, 0.32061303), (8, 0.212706...	[Value, Customer Services, Easy Process, Noise...
5	[restaurant, modern, design, chill, park, near...	[(9, 0.4401033), (8, 0.2572065), (0, 0.1494226...	[Noise and Comfort, Easy Process, Location, Cu...
6	[spacious, bright, located, quiet, beautiful, ...	[(4, 0.4258423), (8, 0.28729737), (3, 0.236808...	[Customer Services, Easy Process, Meals]
7	[location, set, park, friendly, staff, food, h...	[(4, 0.57311565), (9, 0.2239613), (3, 0.086177...	[Customer Services, Noise and Comfort, Meals, ...
8	[]	[(4, 0.54996616), (2, 0.050019223), (9, 0.0500...	[Customer Services, Facilities, Noise and Comf...
9	[big, enough, bed, breakfast, food, service, o...	[(4, 0.38471362), (8, 0.18335861), (3, 0.13279...	[Customer Services, Easy Process, Meals, Facil...

Agenda

A. Problem motivation

B. Dataset Overview & Preprocessing

C. Topic Modelings & Algorithms

D. Findings

E. Conclusion and Potential Application

UNIVERSITY of ROCHESTER

Findings - Merged Reviews

Merged Reviews: Region

Metric	Europe	Americas	Asia	Africa	Oceania
Room Conditions	1	2	2	2	2
Location	2	1	1	1	1
Meals	3	3	3	3	3
Customer Services	4	4	4	4	4
Facilities	5	5	5	5	5
Friendly Staff	6	6	6	6	6
Value	7	7	7	7	7
Noise and Comfort	8	8	9	9	8
Cleanliness	9	9	8	8	9
Easy Process	10	10	10	10	10

Slight different priority compared with different regions, but not much

Top 5 metrics in Merged Reviews

Findings - Merged Reviews

Merged Reviews: Sub-region

Metric	Europe	Eastern Europe	Northern Europe	Western Europe	Southern Europe	Asia	Western Asia	Southern Asia	South-eastern Asia	Eastern Asia
Room Conditions	1	2	1	2	2	2	1	2	2	2
Location	2	1	3	1	1	1	2	1	1	1
Meals	3	3	2	3	3	3	3	3	3	3
Customer Services	4	4	4	4	4	4	4	4	4	4
Facilities	5	5	5	5	5	5	5	5	5	5
Friendly Staff	6	7	6	6	6	6	6	6	6	6
Value	7	6	8	7	7	7	7	7	7	7
Noise and Comfort	8	9	9	8	9	9	9	9	9	9
Cleanliness	9	8	7	9	8	8	8	8	8	8
Easy Process	10	10	10	10	10	10	10	10	10	10

Metric	Americas	Northern America	Latin America and the Caribbean	Africa	Southern Africa	Eastern Africa	Northern Africa	Western Africa	Australia and New Zealand
Room Conditions	2	2	2	2	2	1	1	1	2
Location	1	3	1	1	1	2	2	2	1
Meals	3	1	3	3	3	3	3	3	3
Customer Services	4	4	4	4	4	4	4	4	4
Facilities	5	5	5	5	5	5	6	5	5
Friendly Staff	6	6	6	6	6	6	5	6	6
Value	7	7	7	7	7	7	7	7	7
Noise and Comfort	8	8	9	9	9	9	9	9	8
Cleanliness	9	9	8	8	8	8	8	10	9
Easy Process	10	10	10	10	10	10	10	8	10

Small difference within sub-region

Findings - Negative Reviews

Negative Reviews: Region

Metric	Europe	Americas	Asia	Africa	Oceania
Noise and Comfort	1	4	2	3	3
Meals	2	3	1	1	4
Easy Process	3	2	3	2	2
Value	4	1	6	6	1
Customer Services	5	5	4	4	6
Location	6	6	5	5	5
Facilities	7	7	7	7	7

Critical factors leads to negative reviews:

- issues related to noise levels and overall comfort significantly impact the negative experiences of guests in these regions
- efficiency in services and perceived value for money are critical factors affecting guest satisfaction

Findings - Negative Reviews

Negative Reviews: Sub-region

Metric	Europe	Eastern Europe	Northern Europe	Western Europe	Southern Europe	Asia	Western Asia	Southern Asia	South-eastern Asia	Eastern Asia
Noise and Comfort	1	1	1	1	1	2	2	1	2	2
Meals	2	2	2	2	2	1	1	2	3	3
Easy Process	3	4	5	4	4	3	3	3	1	4
Value	4	3	3	3	3	6	6	6	6	6
Customer Services	5	6	4	5	6	4	4	4	4	1
Location	6	5	6	6	5	5	5	5	5	5
Facilities	7	7	7	7	7	7	7	7	7	7

Metric	Americas	Northern America	Latin America and the Caribbean	Africa	Southern Africa	Eastern Africa	Northern Africa	Western Africa	Australia and New Zealand
Noise and Comfort	4	4	4	3	3	1	2	2	3
Meals	3	3	3	1	1	2	1	1	4
Easy Process	2	1	2	2	6	3	3	3	2
Value	1	2	1	6	4	4	6	6	1
Customer Services	5	5	6	4	2	6	4	4	6
Location	6	6	5	5	5	5	5	5	5
Facilities	7	7	7	7	7	7	7	7	7

Findings - Positive Reviews

Positive Reviews: Region

Metric	Europe	Americas	Asia	Africa	Oceania
Customer Services	1	1	1	1	1
Easy Process	2	2	2	2	2
Noise and Comfort	3	4	3	3	4
Meals	4	3	4	4	3
Location	5	5	6	5	5
Cleanliness	6	6	7	7	6
Facilities	7	7	5	6	7

- Cleanliness are more prominent in positive reviews compared to negative reviews, indicating that when they are well-maintained, they significantly contribute to a positive experience.

Top 3 critical factors leads to positive reviews:

- Customer Services' is the top-ranked topic in positive reviews for all regions, highlighting the crucial role of service quality in creating positive guest experiences.
- emphasizing that efficient processes and comfortable, quiet environments are key contributors to satisfaction.

Findings - Positive Reviews

Positive Reviews: Sub-region

Metric	Europe	Eastern Europe	Northern Europe	Western Europe	Southern Europe	Asia	Western Asia	Southern Asia	South-eastern Asia	Eastern Asia
Customer Services	1	1	1	1	1	1	1	1	1	1
Easy Process	2	2	2	2	2	2	2	2	2	2
Noise and Comfort	3	3	3	4	3	3	3	3	4	4
Meals	4	4	4	3	4	4	4	4	3	3
Location	5	5	5	5	5	6	7	7	6	5
Cleanliness	6	6	6	6	6	7	6	5	5	6
Facilities	7	7	7	7	7	5	5	6	7	7

Metric	Americas	Northern America	Latin America and the Caribbean	Africa	Southern Africa	Eastern Africa	Northern Africa	Western Africa	Australia and New Zealand
Customer Services	1	1	1	1	1	1	1	1	1
Easy Process	2	2	2	2	2	2	2	2	2
Noise and Comfort	4	4	3	3	3	3	3	3	4
Meals	3	3	4	4	4	4	4	4	3
Location	5	5	6	5	5	5	6	5	5
Cleanliness	6	6	5	7	7	7	7	7	6
Facilities	7	7	7	6	6	6	5	6	7

UNIVERSITY of ROCHESTER

Findings

- The difference in ranking between negative and positive reviews for the same region indicates areas where improvements can significantly enhance guest experiences.

- Cultural and regional differences can influence the emphasis on certain topics. For example, 'Customer Service' is more prominent in specific sub-region than others, possibly reflecting cultural values or hospitality standards.

- The importance of 'Noise and Comfort' in both negative and positive reviews underlines the significant impact of the physical environment on guest experiences.

Agenda

A. Problem motivation

B. Dataset Overview & Preprocessing

C. Topic Modelings & Algorithms

D. Findings

E. Conclusion and Potential Application

Conclusion and Potential Application

- Conclusion
    - There’s slightly difference between travelers from different sub-regions and regions from the dataset, and with some interesting topics
    - Within luxury hotels in main city in Europe, there are different preference leads to negative and positive reviews

- Restriction
    - Access to different booking platform sources, locations, and various hotel style

- Future work
    - Access to different booking platform sources, locations, and various hotel style
    - Recommendation System(perhaps)

THANK YOU

TEAM ISLAND