



# PROJECT OVERVIEW

- Objectives
- Motivations
- Research Question



# DATA

## SOURCE

GitHub – Web scraping



## INSIDE THE FILES

Title, reviews, hotel, starts



## REVIEW SENTIMENT

Positive, Negative, Neutral



## DATASET

JSON files for 8 different vacation spots



## TOTAL REVIEWS

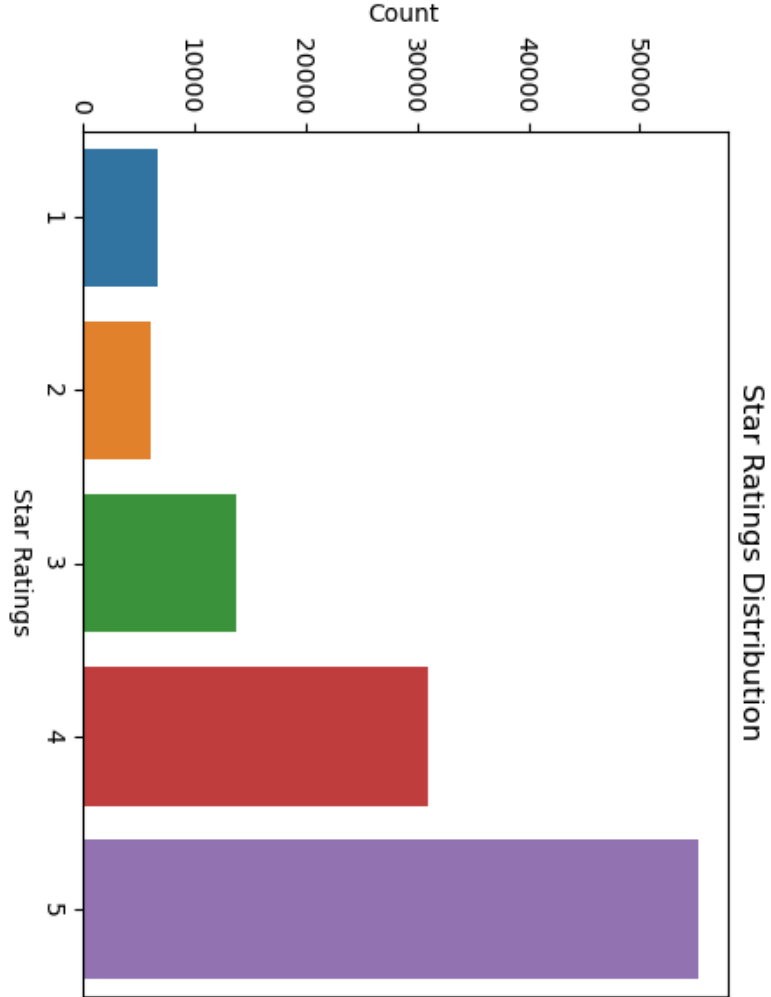
8 6001



## DATA HEAD

stars	hotel	title	content	
0	5	[The Reserve at Paradisus Punta Cana]	[Family concierge stay]	[Hands down great family vacation. Estefanía o...
1	5	[Paradisus Punta Cana Resort]	[Sunrise bartender]	[Juan Batistae has got to be the best host/bar...
2	5	[Dreams Palm Beach Punta Cana]	[Overall Excellent Experience!]	[This was our first trip! We went with our fri...
3	5	[Now Onyx Punta Cana]	[Lovely 10 days]	[Me and my partner travelled from Glasgow Scot...
4	4	[Paradisus Punta Cana Resort]	[Amazing spring break vacation 🌞🍹🥰👍]	[We had amazing time with family and friends. ...
5	5	[Dreams Palm Beach Punta Cana]	[Great vacation hotel]	[Just got back from a week long girls trip and...
6	5	[Now Onyx Punta Cana]	[Paradise personified]	[This trip was a retirement present my Mum fro...
7	5	[Grand Palladium Punta Cana Resort & Spa]	[Excellent stay]	[My husband and I had an amazing time at grand...
8	5	[Grand Palladium Punta Cana Resort & Spa]	[Family Vacation 2019]	[Very satisfied with Grand Palladium Punta Can...
9	4	[Four Points by Sheraton Puntacana Village]	[Pit stop ]	[Good location, After a few days on Bavaro be...

# RATING DISTRIBUTION



[illegible]

## Positive Word Cloud

[illegible]

## Negative Word Cloud



# METHODOLOGY

## LOADING, COMBINING, EXPLORING

- Load Libraries
- Loading JSON Data
- Combine the files into one file
- Explore the data using bar charts and word clouds

## SENTIMENT ANALYSIS

- Plotting the sentiments score
- Looking at the score ranging from -1 to 1.
- -1 = negative reviews
- 1 = positive reviews

## PREPROCESSING

- Lowercase
- Stop word removal
- Lemmatization
- Tokenization

## MODELS AND COSINE SIMILARITY

- LDA for topic modeling
- NMF for topic modeling
- Cosine Similarity Scores based on user input

## CHATBOT

- Interactive to enhance the user's experience

# FINDINGS AND RESULTS



SENTIMENT  
ANALYSIS

Plot ranging from -1  
to 1



LATENT  
DIRICHLET  
ALLOCATION

Top 10 words for 5  
topics



NON-NEGATIVE  
MATRIX  
FACTORIZATION

Top 10 words for  
5 topics



COSINE  
SIMILARITY

Similarity score  
based on user input

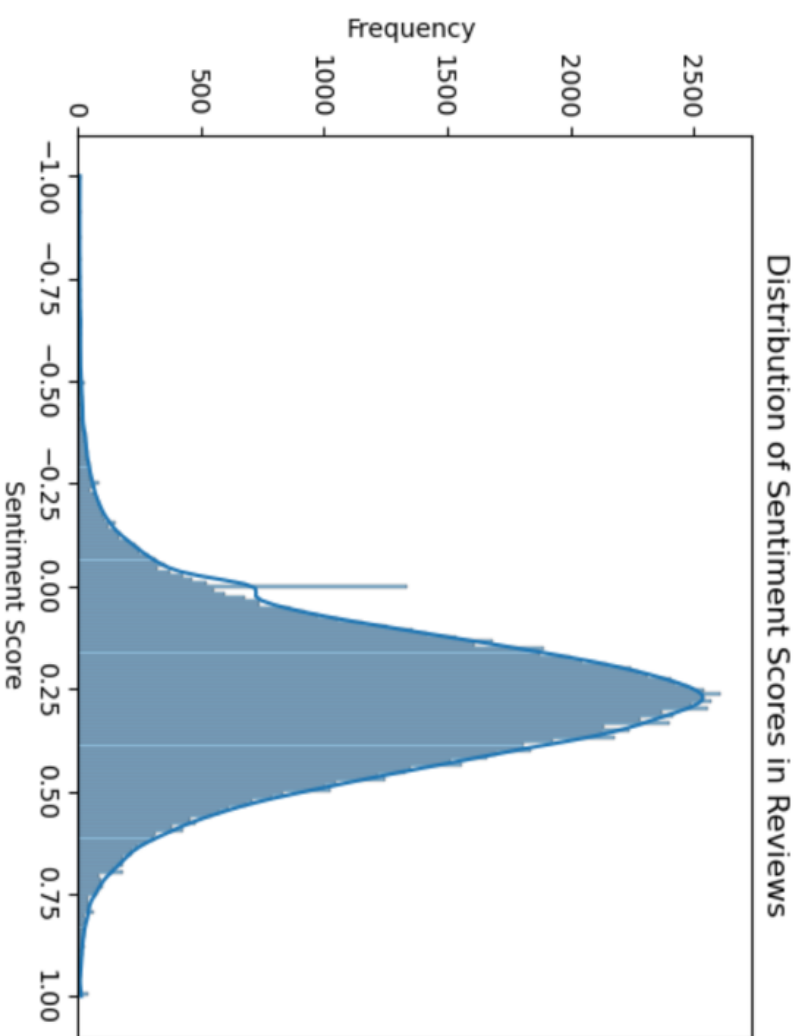


CHATBOT

Interactive to align  
with user preferences

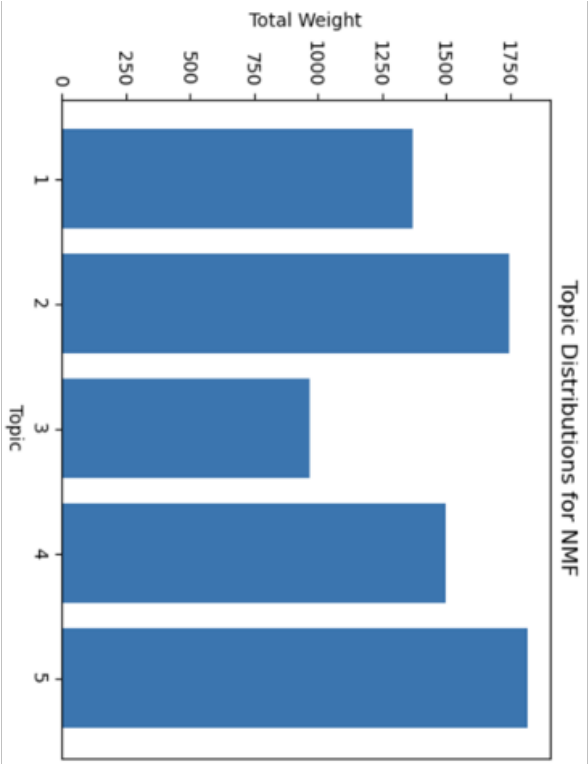
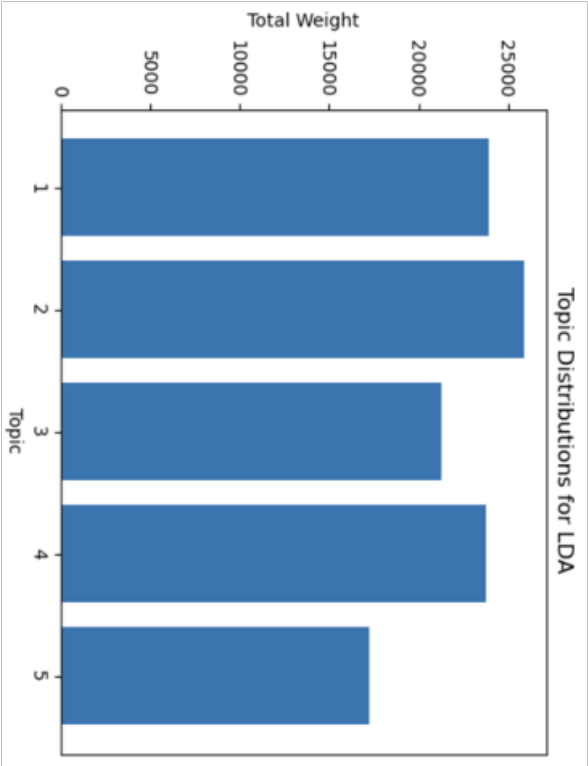


# SENTIMENT ANALYSIS





# LDA & NMF



LDA Topic 1: room, hotel, night, bed, day, good, place, didnt, stay  
LDA Topic 2: hotel, room, breakfast, good, city, location, vienna, clean, staff  
LDA Topic 3: place, great, hostel, stay, hotel, room, staff, location, beach  
LDA Topic 4: resort, food, beach, great, service, staff, pool, amazing, time  
LDA Topic 5: pool, apartment, hotel, beach, good, lovely, room, area, bar  
NMF Topic 1: hotel, good, breakfast, room, staff, vienna, location, city, clean  
NMF Topic 2: resort, beach, pool, food, service, great, restaurant, time, amazing  
NMF Topic 3: apartment, walk, minute, kitchen, area, bedroom, lovely, clean, centre  
NMF Topic 4: hostel, great, place, stay, staff, really, location, nice, friendly  
NMF Topic 5: room, night, bed, day, shower, bathroom, didnt, floor, like



# COSINE SIMILARITY

Top 5 recommended destinations:  
hotel \

28711	[Boracay Uptown]
28357	[The Tides Hotel Boracay]
23872	[La Carmela de Boracay]
68675	[Ocean Riviera Paradise]
12581	[Ocean El Faro]

	content	similarity_score \
28711	front beach resort geat view enjoy food good s...	0.554219
28357	great island beach beautiful everything much c...	0.537607
23872	one unforgettable experience im family enjoy f...	0.529129
68675	resort beautiful variety activity enjoy beach ...	0.474074
12581	enjoyed bowling food pool entertainment activi...	0.463626

stars

28711	3
28357	5
23872	5
68675	4
12581	5

# CHATBOT

## *Tutorial*

- **Interaction:** Chatbot collects user's vacation preferences.
- **Processing:** User preferences are analyzed.
- **Recommendations:** Suggestions are made based on user input.
- **Outcome:** Personalized user experience with tailored recommendations.



N

## LIMITATIONS

- **Data Collection:** Explore dataset's representativeness and comprehensiveness.
- **EDA:** Utilize advanced visualizations and analytics tools. ———
- **Text Preprocessing:** Refine model to adapt to evolving language trends.
- **Topic Analysis:** Evaluate LDA and NMF against other topic modeling techniques.
- **Similarity Algorithms:** Investigate alternative algorithms for scalability and efficiency.
- **Chatbot Intelligence:** Evaluate and enhance chatbot's responsiveness and learning capability.

## FUTURE RESEARCH

- **Personalization:** Improve travel recommendations with diverse user preferences and social influences. \_\_\_\_\_
- **Data Sources:** Explore varied data sources for a comprehensive recommendation system.
- **EDA Tools:** Refine and expand Exploratory Data Analysis tools.
- **Text Preprocessing:** Improve text preprocessing for better data quality.
- **Topic Modeling:** Assess topic modeling techniques for optimization.
- **Similarity Algorithms:** Optimize similarity algorithms for scalability and efficiency.
- **Chatbot Intelligence:** Enhance chatbot intelligence for increased user engagement.

# CONCLUSION

- **Achievements & Data Collection:** Developed NLP system; diversify data sources.
- **EDA & Preprocessing:** Uncovered patterns; refine text preprocessing.
- **Topic Analysis & Cosine Similarity:** Used LDA, NMF; consider scalability.
- **Chatbot & Future Focus:** Enhances engagement; diversify sources, enhance tools, refine protocols.
- **Final Note:** Iterative refinement makes NLP dynamic and exciting.



THANK YOU!