

Classifying Subreddits Using NLP

DSB-122
Elaine Chen

Table of contents

01 Objective

02 Subreddits Scrapping

03 Data Cleaning & EDA

04 Modeling & Evaluation

05 Limitation and Recommendation

Objective

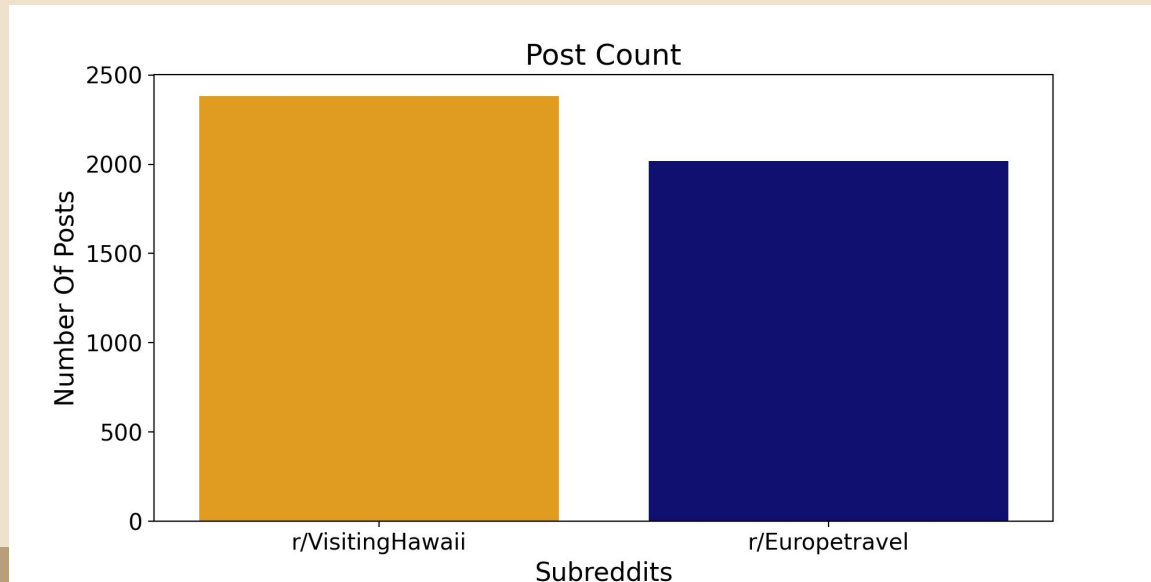
Perform a comparative analysis of user engagement patterns and content characteristics within the subreddits `r/VisitingHawaii` and `r/Europetravel`.

By examining various text-based metrics and features, the goal is to differentiate between the two communities based on the contents of the posts.

Subreddits Scrapping

r/VisitingHawaii

r/Europetravel



Data Cleaning

URLs

```
We are eloping on Oahu in May and have to decide on a Hotel for 7 nights.  
We looked at a lot of hotels in the last few weeks and these 3 all fit in our budget and look very nice so far. We  
want to explore the island and won't be at the hotel for most of the time, so the Halekulani or Turtle Bay are not  
worth it for us.  
Any opinions on these hotels and what offer would you book for the occasion?
```

```
[View Poll](https://www.reddit.com/poll/189dqh7)
```

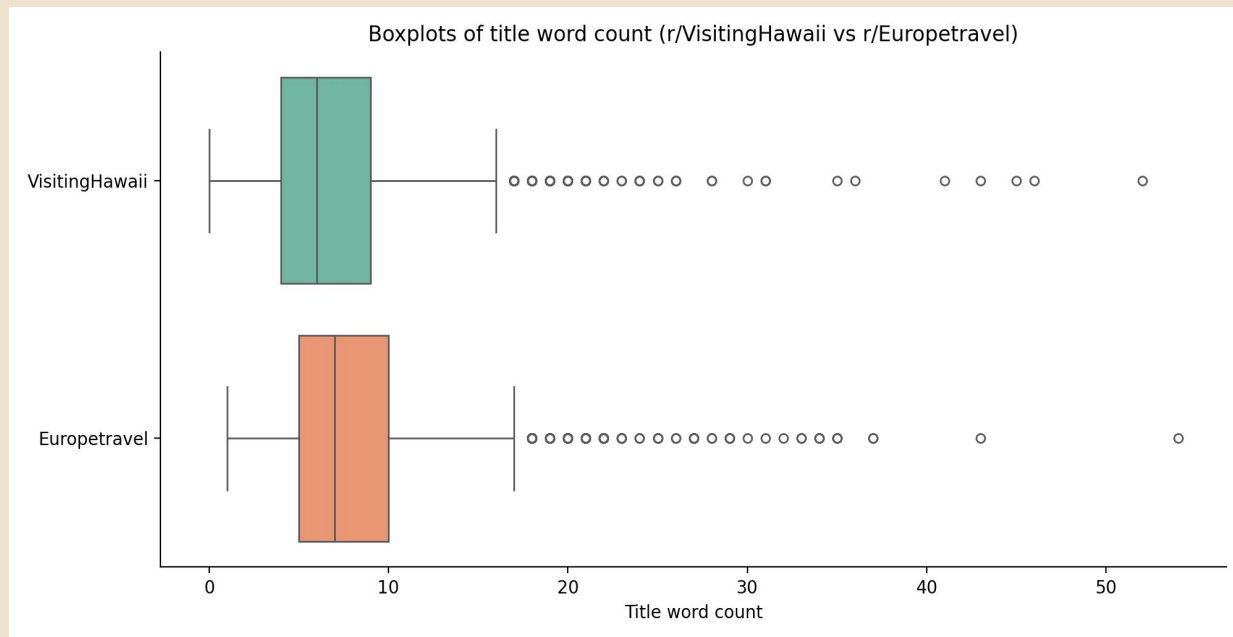
HTML encoded characters

```
"&#x200B;\n\nhttps://preview.redd.it/zrm5yww2b9lc1.png?width=781&format=png&auto=webp&s=e404da96f6b56196f22056ea5cd  
0d0a43897db38\n\nFirst time travelling solo in Europe for a week the end of August. I'm flying into Amsterdam as fl
```

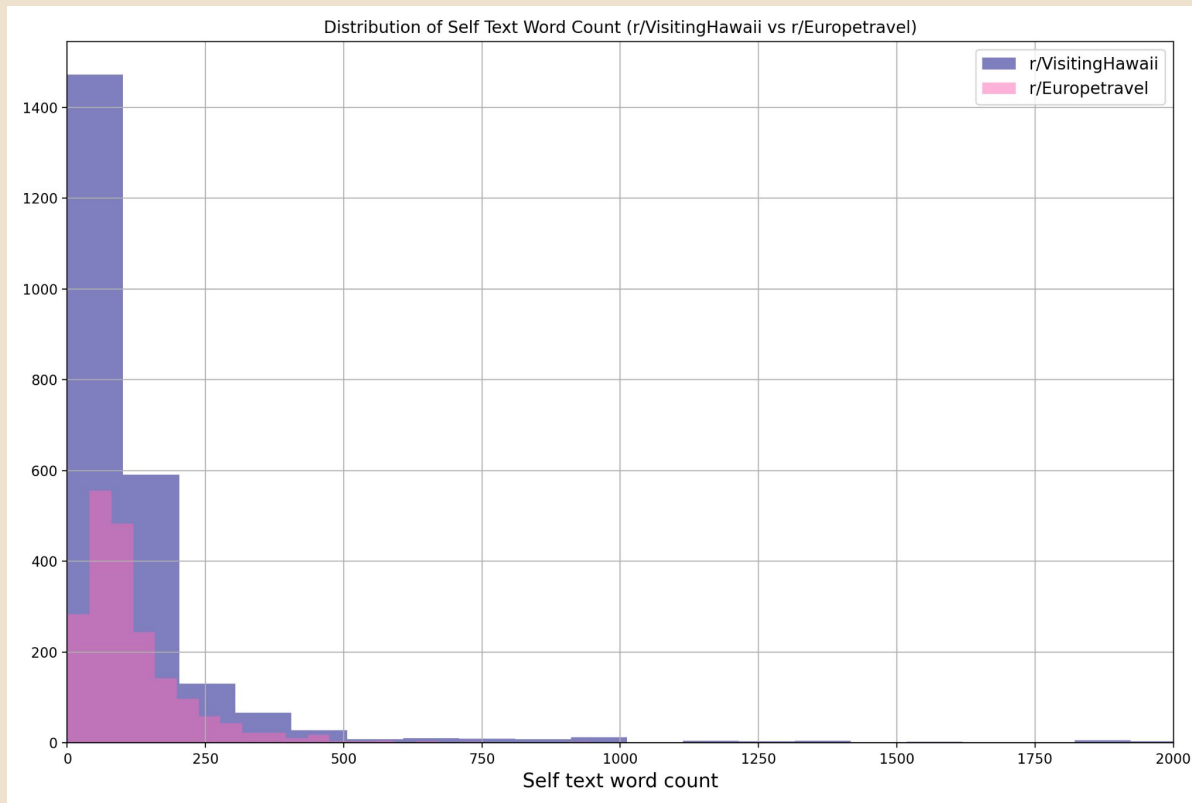
Emojis

```
"First time to Europe ! It's my honeymoon 💜💜"
```

EDA



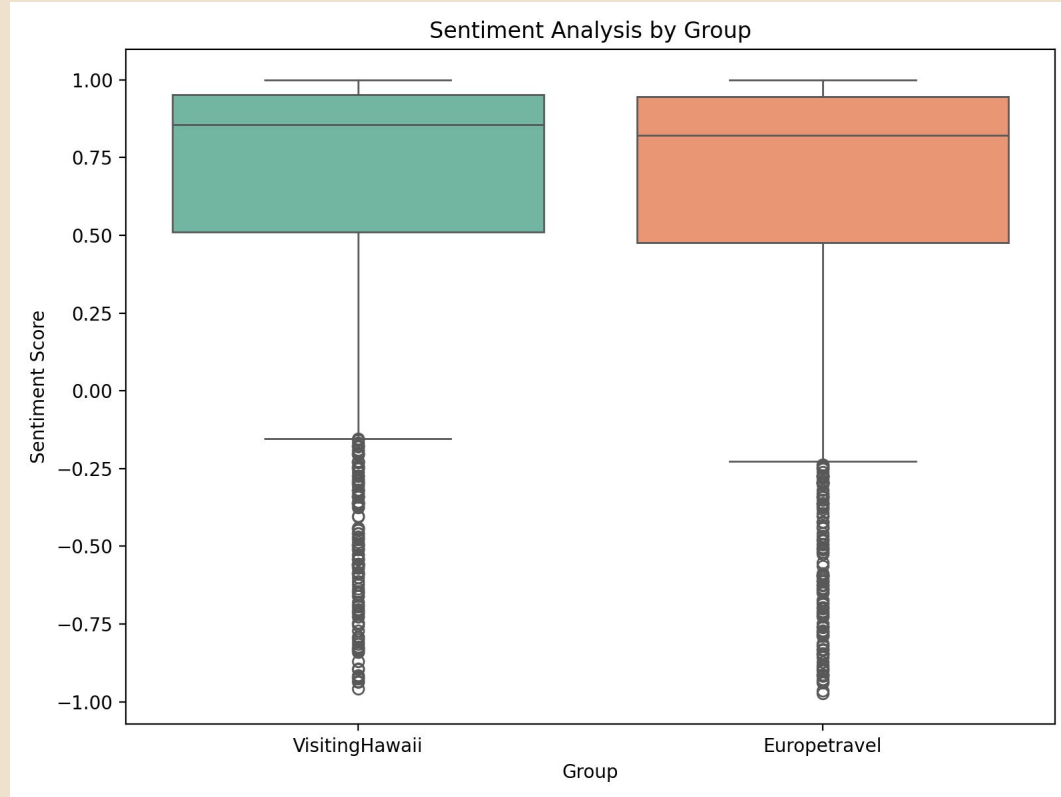
EDA



Sentiment Analysis



Sentiment Analysis



r/VisitingHawaii Frequent Words



A word cloud featuring various terms related to visiting Hawaii. The most prominent words are 'day', 'island', and 'beach', all in large, bold, orange letters. Other visible words include 'time', 'trip', 'visit', 'plan', 'stay', 'hawaii', 'oahu', 'look', 'want', 'big', 'like', 'just', and 'rental'.

stay plan visit
time day like
hawaii oahu just
island trip
look want big
beach



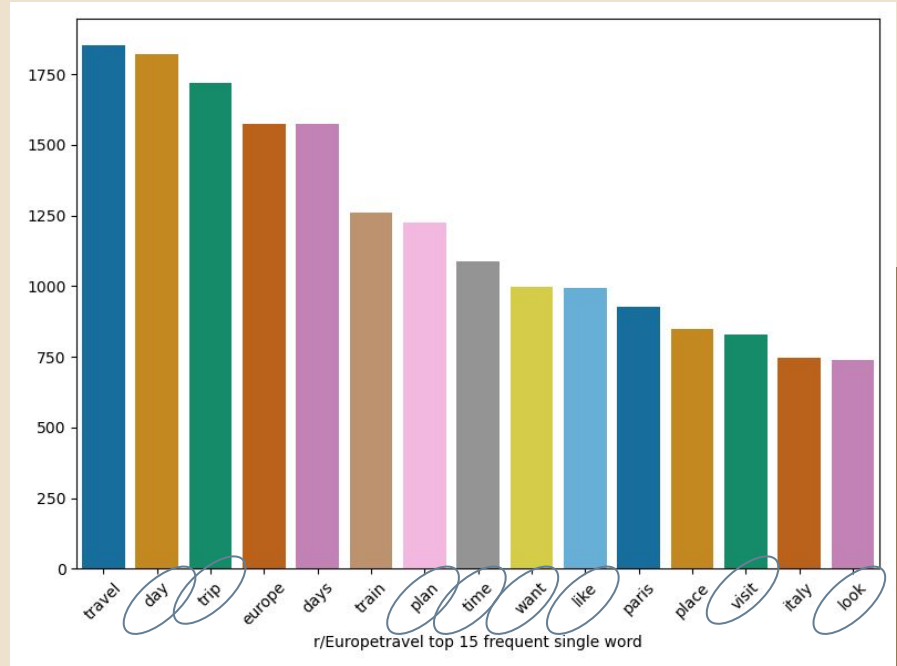
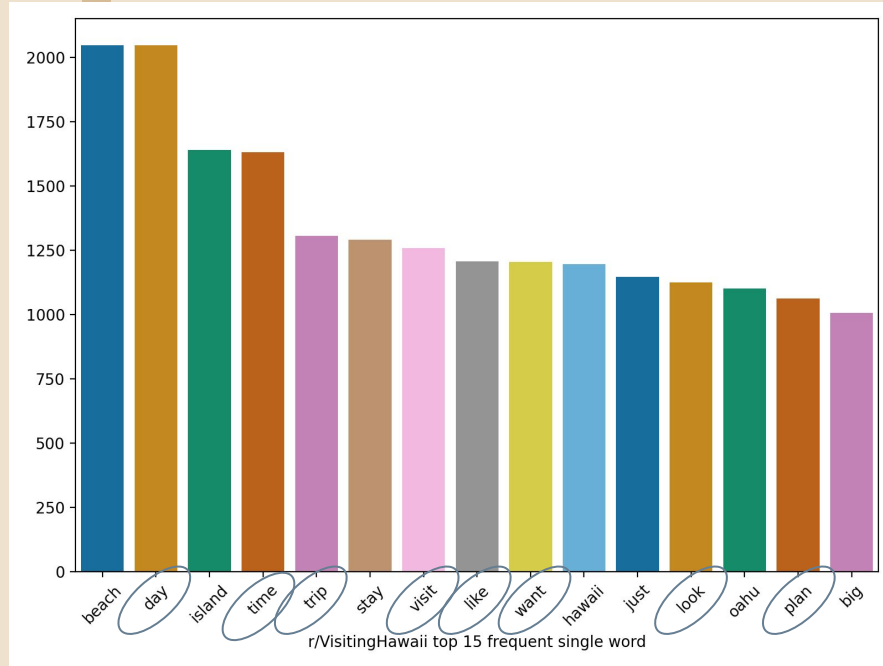
A word cloud featuring various terms related to visiting Hawaii. The most prominent words are 'big island', 'rental car', and 'visit oahu', all in large, bold, blue letters. Other visible words include 'waikiki', 'beach', 'year', 'old', 'thank', 'advance', 'manta ray', 'north shore', 'visit hawaii', 'rent car', 'plan trip', 'diamond head', 'pearl harbor', 'mauna kea', and 'national park'.

waikiki beach
year old rental car
thank advance manta ray
visit oahu
north shore
visit hawaii
rent car plan trip
big island
diamond head
pearl harbor
mauna kea
national park

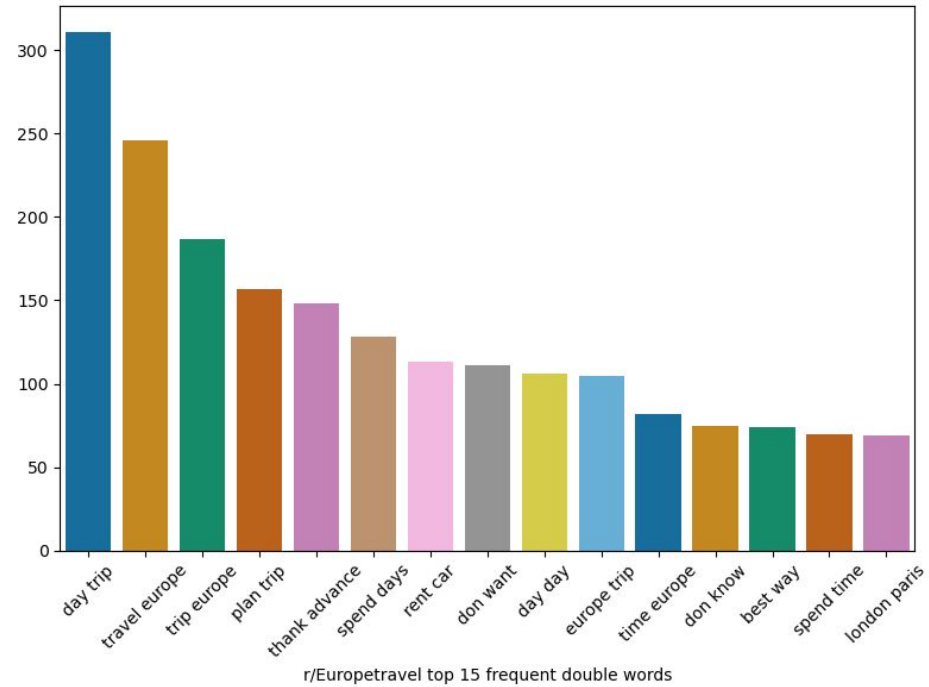
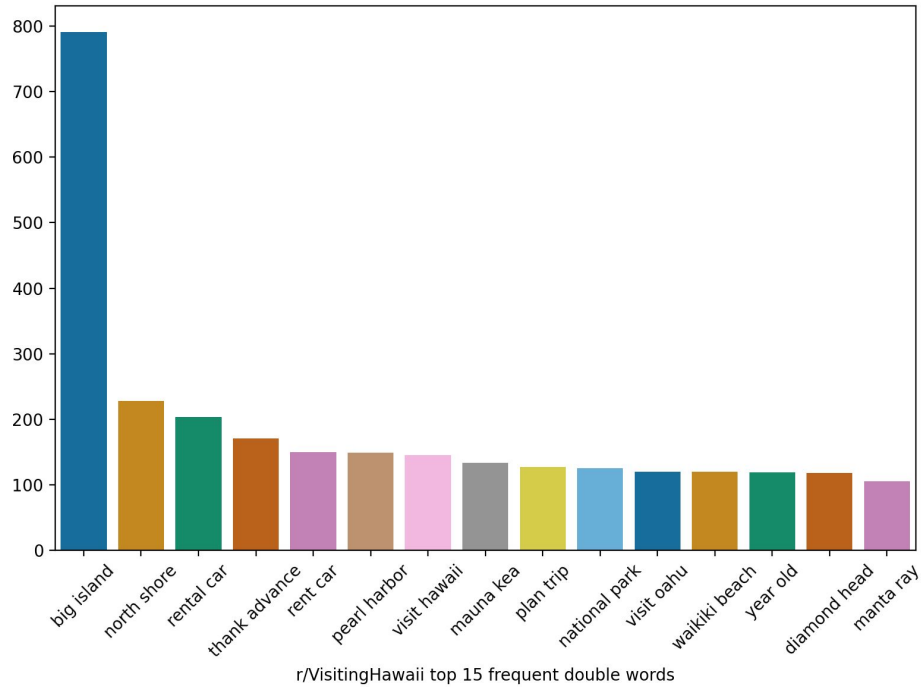
r/Europetravel Frequent Words



Top 15 Frequent Single Word



Top 15 Frequent Double Words



Modeling and Evaluation

Model	Logistic Regression	Multinomial Naive Bayes	Random Forest Classifier	SVC (Count Vectorizer)	SVC (Tfidf Vectorizer)
Training Score	100%	99.18%	98.76%	99.06%	100%
Testing Score	97.8%	97.45%	96.27%	95.46%	98.18%

Limitations and Recommendations

Data Bias - During the data cleaning process, non-English posts were excluded, potentially due to the composition of the user base. This filtering may result in biased model predictions and inaccuracies when classifying posts in other languages for subreddit classification.

Recommendation

- Multilingual data collection to improve model robustness

- Language identification to automatically detect and classify the language

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution