

★★★★★

# AirBnB

## Keyword Analysis

- Project 3 - Group 3 -

- Stephanie
- Junny
- Wes
- Zi Ming





**Problem  
Statement**



**Dataset &  
Methodology**



**EDA &  
Visualization**



**Modelling &  
Evaluation**



**Conclusion,  
Recommendations &  
Continuous  
Improvement**

# Contents



# Problem Statement



As part of the AirBNB Data Insights department, we are tasked to leverage on insightful reddit curated text data for designing and implementing strategies for sales, marketing and campaign outreach.



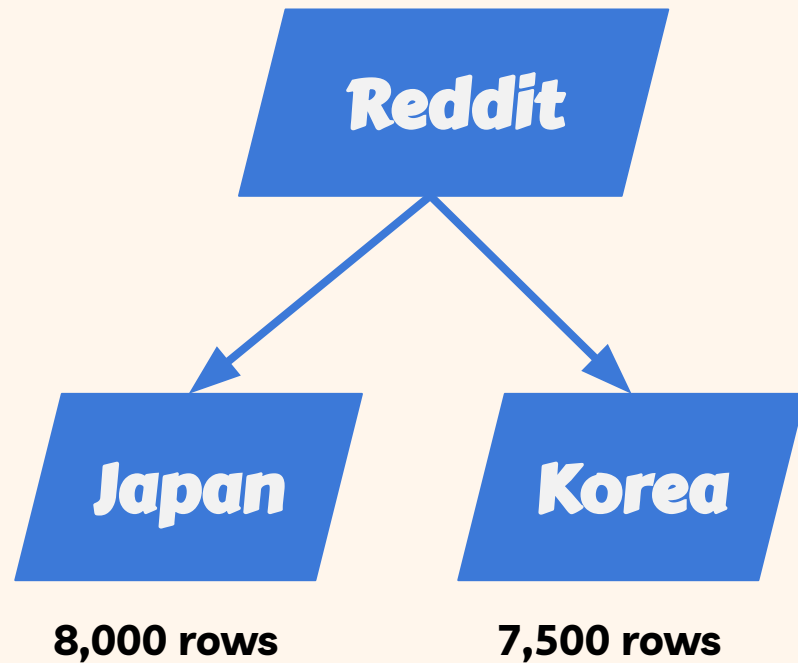
Using NLP machine learning algorithms developed, we would be able to classify and predict popular trends, preferences and travel updates inbound Japan and Korea.

# Data Collection & Methodology



- 1 Using Reddit API to extract comments, no. of upvotes, time of comments and etc
- 2 Normalize text, removing noise data, stop words and etc
- 3 Extract significant, most frequently used keywords or phrases via TF-IDF analysis and Count-Vectorizer transformations
- 4 Classify up/downvotes based on keywords analysis of the frequency and context of those words
- 5 Compile insights report that highlights most important trends, popular destinations, seasonal travellers preferences

# Exploratory Data Analysis



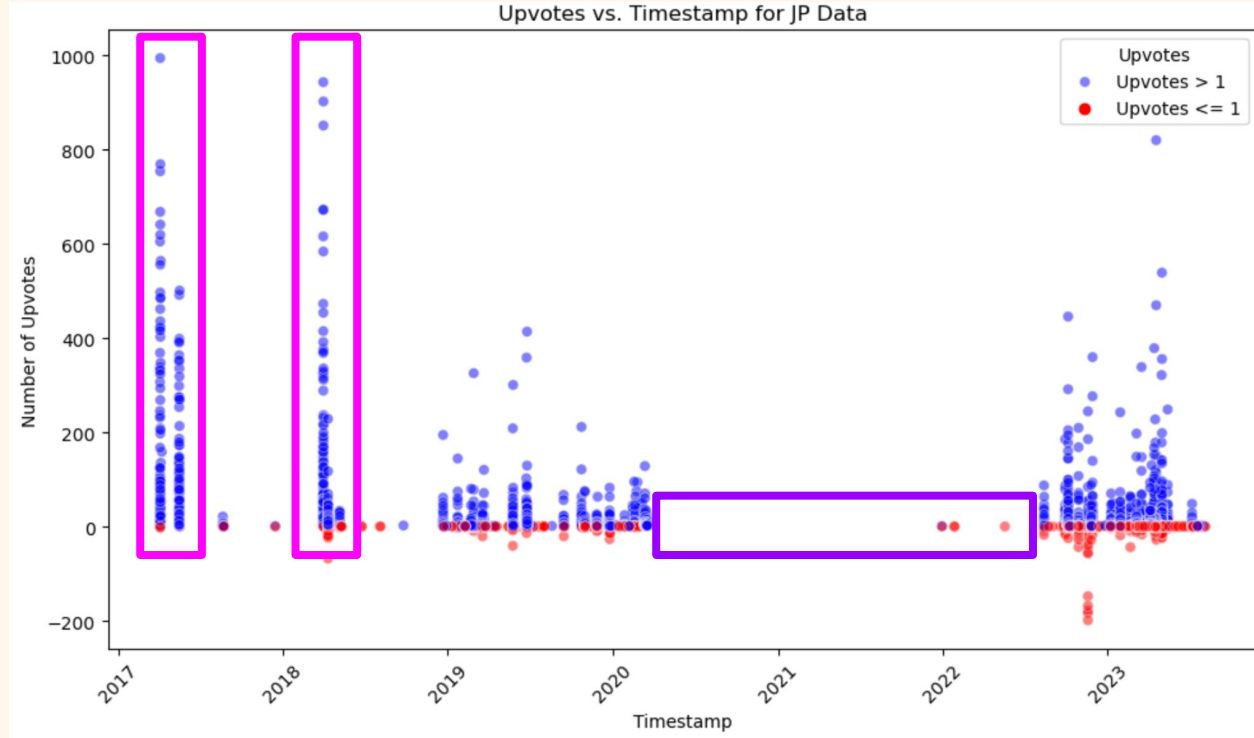
# Exploratory Data Analysis - JAPAN



# Exploratory Data Analysis - KOREA

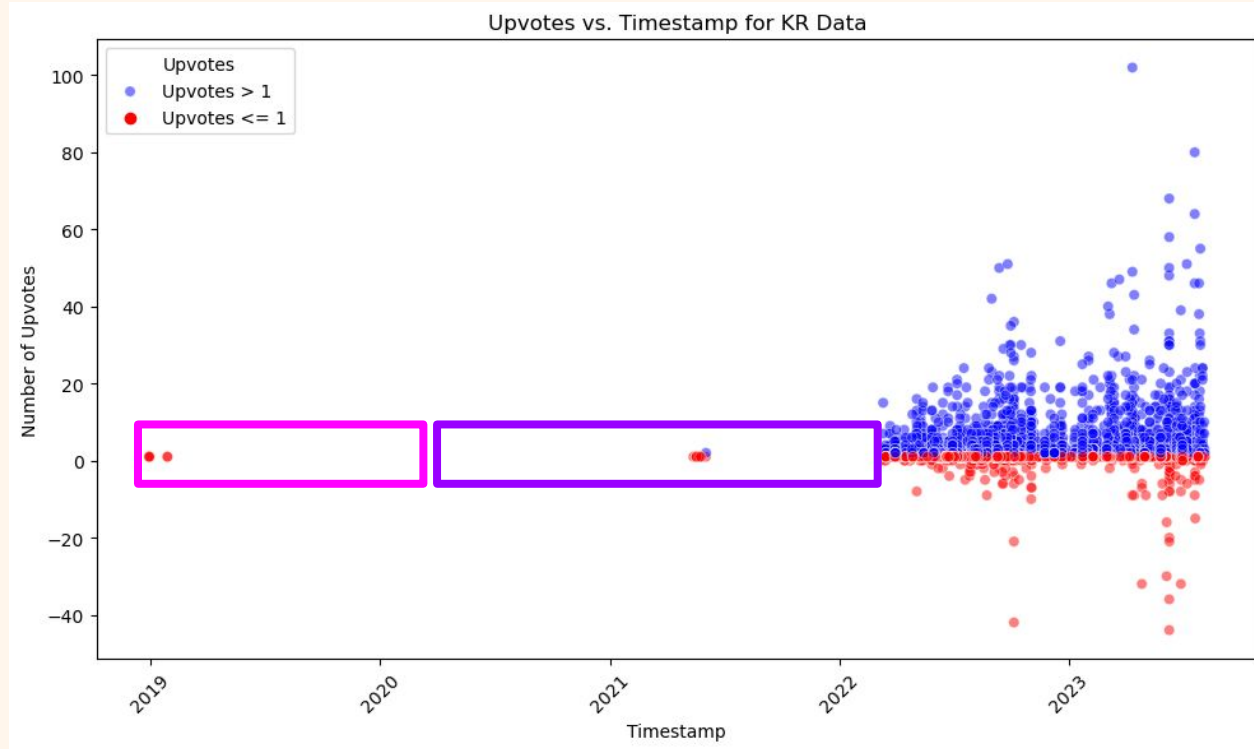


# Time x Upvotes – JAPAN





# Time x Upvotes – KOREA



# ***Modelling and Evaluation***



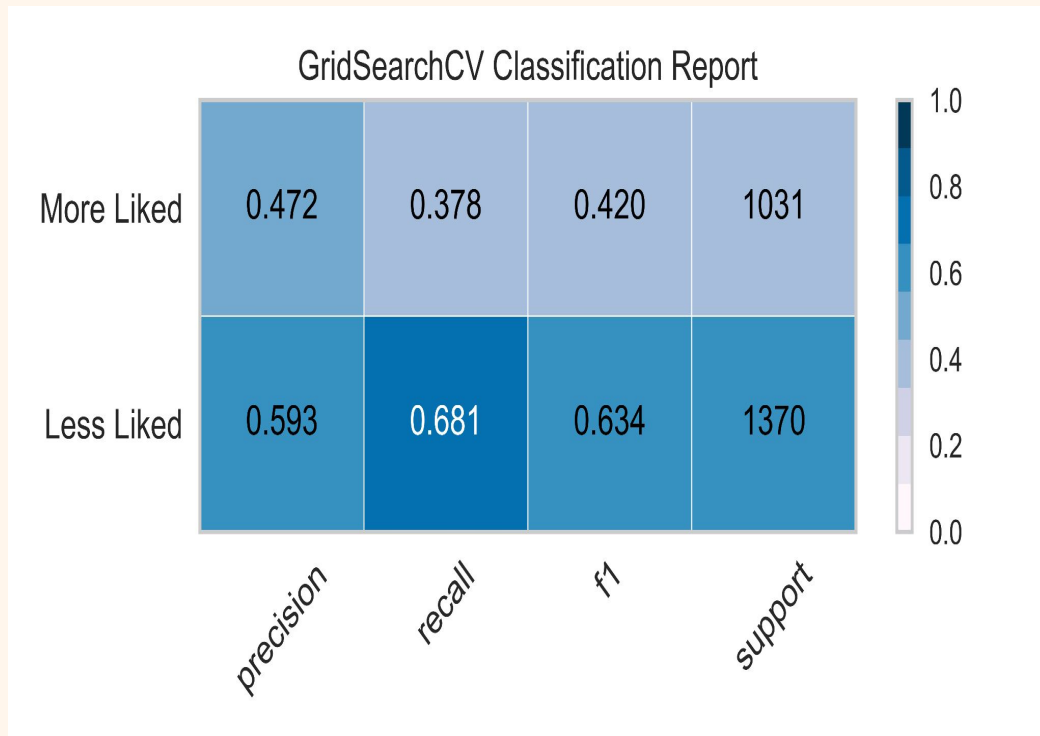
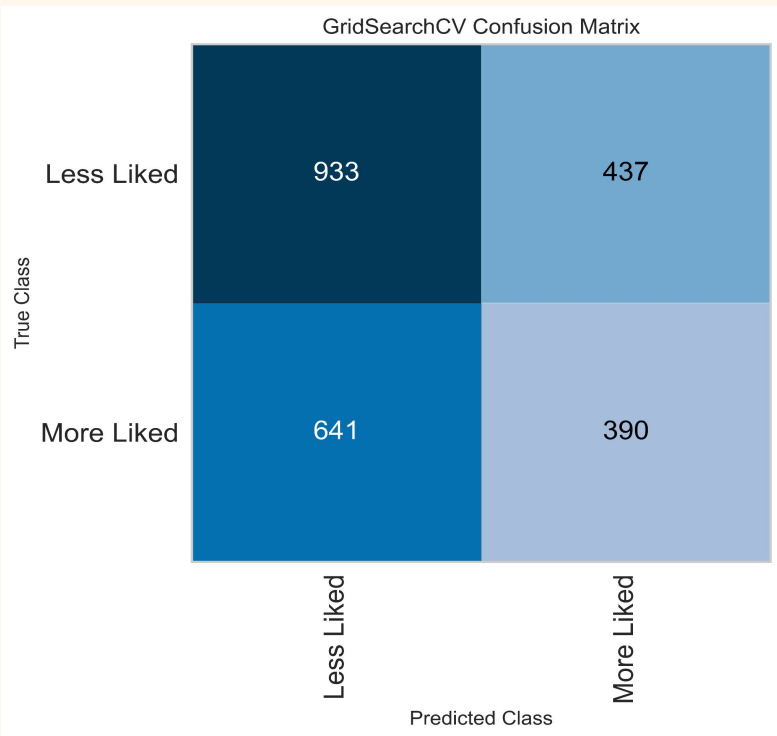
**Count  
Vectorizer**

**TF-IDF  
Vectorizer**

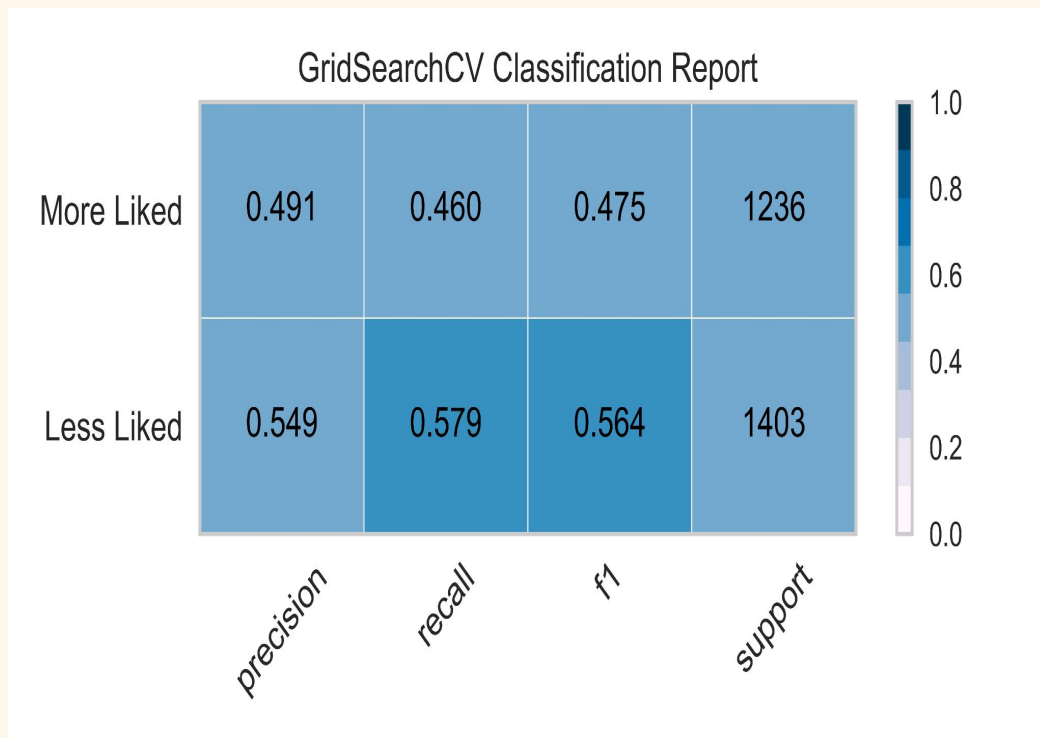
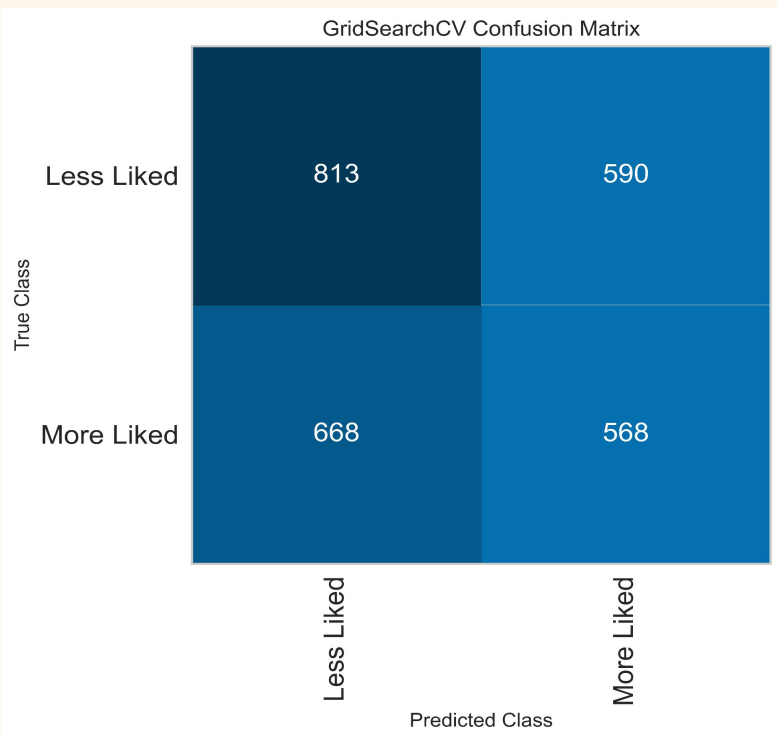
**Multinomial  
Naïve Bayes**

**Random  
Forest**

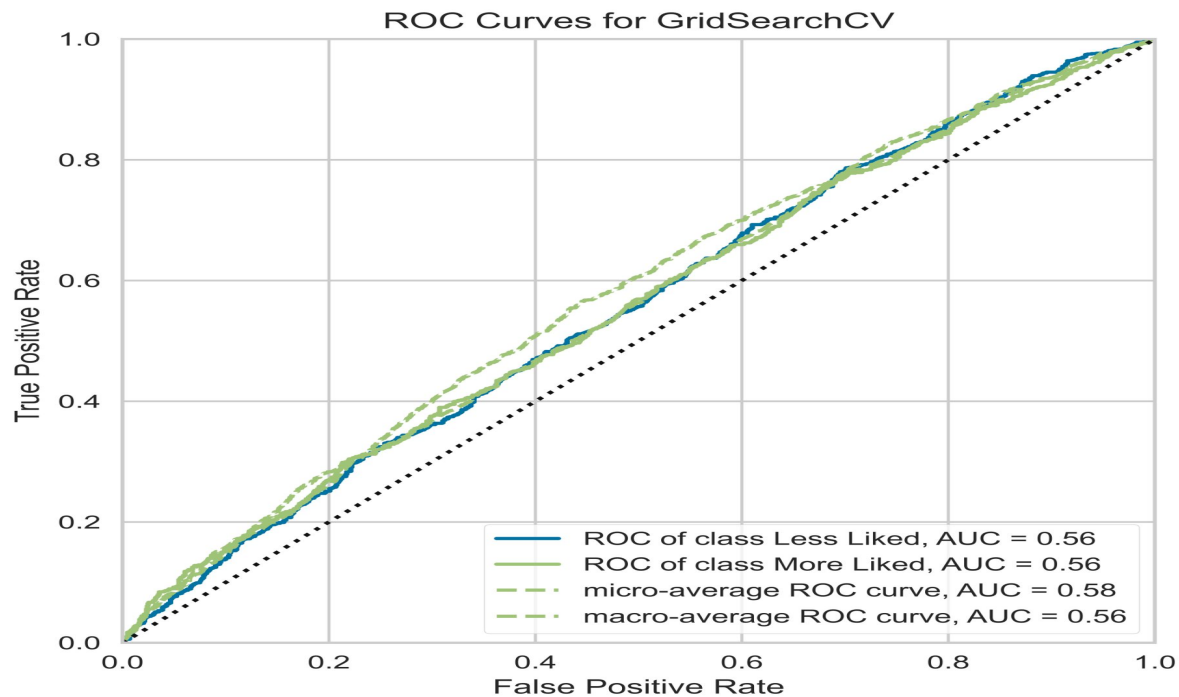
# Japan



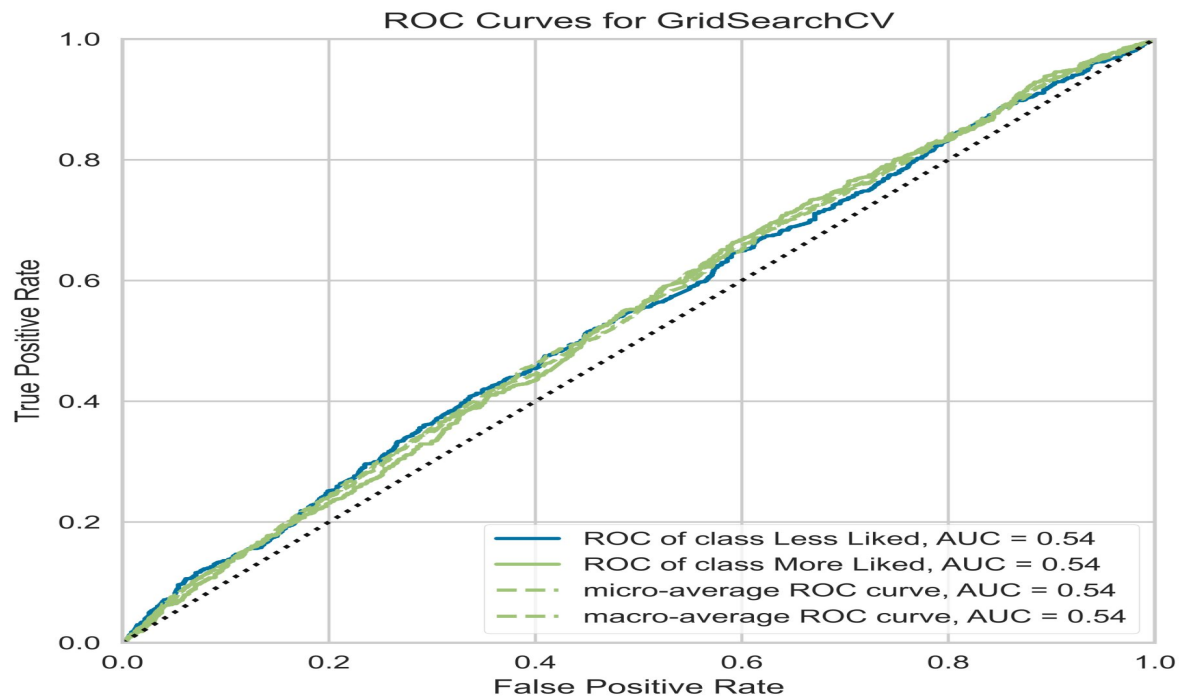
# Korea



# Japan



# Korea



# Conclusion



## Model Performance:



- Achieved around 57% accuracy.
- Provides insights into travelers' prevailing sentiments and interests.

## Benefits for Airbnb:



- Insights allow for strategic curation of offerings.
- Ability to target current travel trends.
- Promote offerings from specific destinations to themed experiences.

# Recommendations



## Decision-making Aid:

1

- Leverage model's accurate classification and prediction.
- Make data-driven decisions.
- Example: Incentivize hosts in emerging travel locales in Korea.

## Validation Against Observable Trends:

2

- Match model insights with real-world observations.
- Correlation with actual trends confirms model effectiveness.



# Continuous Improvements



## Domain-Specific Knowledge:

1

- Focus on keywords specific to travel in Japan and Korea.
- Curate list of cities, attractions, and cultural references.

## Exploring Topic Modelling:

2

- Investigate algorithms such as Latent Dirichlet Allocation (LDA).
- Identify clusters of frequently co-occurring words.

## Incorporating a Feedback Mechanism:

3

- Value the importance of feedback.
- Share initial findings with experts or team members.

***Thank you***