



SPIRIT OF GOA

A Data-Driven Approach to
Capturing People's Perception

Group 11:

Advik Sharma - B2024005

Ashvath Diniz - I204002

Ashwin Rajesh - I2024006

Sakshi Chaudhari - I2024005

Rohan Sharma - I2024008





INTRODUCTION

- This GiveGoa project aims to capture perceptions in their purest form and conduct an analysis on what people think about Goa as a state
- We, as BDA students of GIM have worked in collaboration with ActForGoa, a community working towards sustainable development in Goa. ActForGoa is dedicated to preserving and promoting Goa's unique essence.
- A survey was conducted to understand people's perceptions of what defines the true spirit of Goa, and a deep analysis was performed on the collected data.
- We also performed sentiment analysis and topic modeling on posts from the r/Goa subreddit using various NLP techniques.
- Insights from this study will help drive initiatives aligned with Goa's cultural and environmental values.



ACT FOR GOA

ACTFORGOA

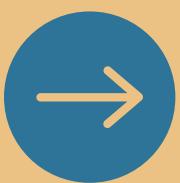
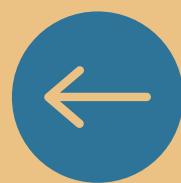
ActForGoa is a platform committed to fostering sustainable and positive action in Goa. It empowers individuals by providing fact-based resources, a Sustainability Green Pages directory, and an outreach channel to amplify impactful activities. The organization works to prevent duplication of efforts by encouraging collaboration among NGOs, government bodies, businesses, academia, and engaged citizens. ActForGoa enables public participation in governance, matches requests for support with available resources, and highlights inspiring citizen-led initiatives. Additionally, it promotes sustainability through education, volunteer opportunities, and policy advocacy, ensuring that Goa retains its unique identity while addressing critical environmental and social challenges. Our collaboration with ActForGoa was a valuable step towards collecting data as its communications team designed an outreach approach to ensure numbers of respondents.



SURVEY OVERVIEW

- Conducted in four languages: English, Marathi, Hindi, and Konkani
- Targeted respondents across different demographics (locals, visitors, diaspora)
- Designed to gather sentiment and opinions on Goa's identity, challenges, and future
- Distributed through multiple channels: WhatsApp groups, Instagram page of ActForGoa, QR code





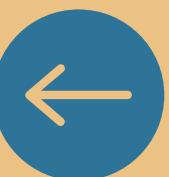
SURVEY OBJECTIVE

- 1** Main Goal: Conduct a sentiment analysis to understand people's perceptions of Goa. Understand how positively or negatively people think about Goa.
- 2** Analyze emotional responses to key themes like culture, environment, and tourism. Understand whether or not there is a unique spirit of Goa which resonates among respondents.
- 3** Identify actionable insights through visualizations and data results to help guide community and policy initiatives.



SURVEY QUESTIONS

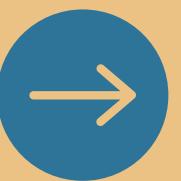
1. *How would you describe your connection to Goa?*
2. *Do you consider Goa your home?*
3. *Which area of Goa are you from or currently living in?*
4. *Do you think Goa has a unique spirit?*
5. *What aspects contribute most to Goa's identity?*
6. *What can be done to preserve Goa's true essence?*
7. *Biggest challenges in preserving Goa's spirit?*
8. *Fears & hopes for the future of Goa?*
9. *Demographics: Age, Gender*



TARGET AUDIENCE

- Goan Residents (Born & Living in Goa)
- Goan Diaspora (Returnees & families with roots in Goa)
- Non-Goan Long-Term Residents
- Students & Young Professionals
- Short-Term Residents
- Tourists & Visitors
- Senior Citizens & Generational Locals
- Environment enthusiasts
- NGO workers





SURVEY ANALYSIS METHODOLOGY

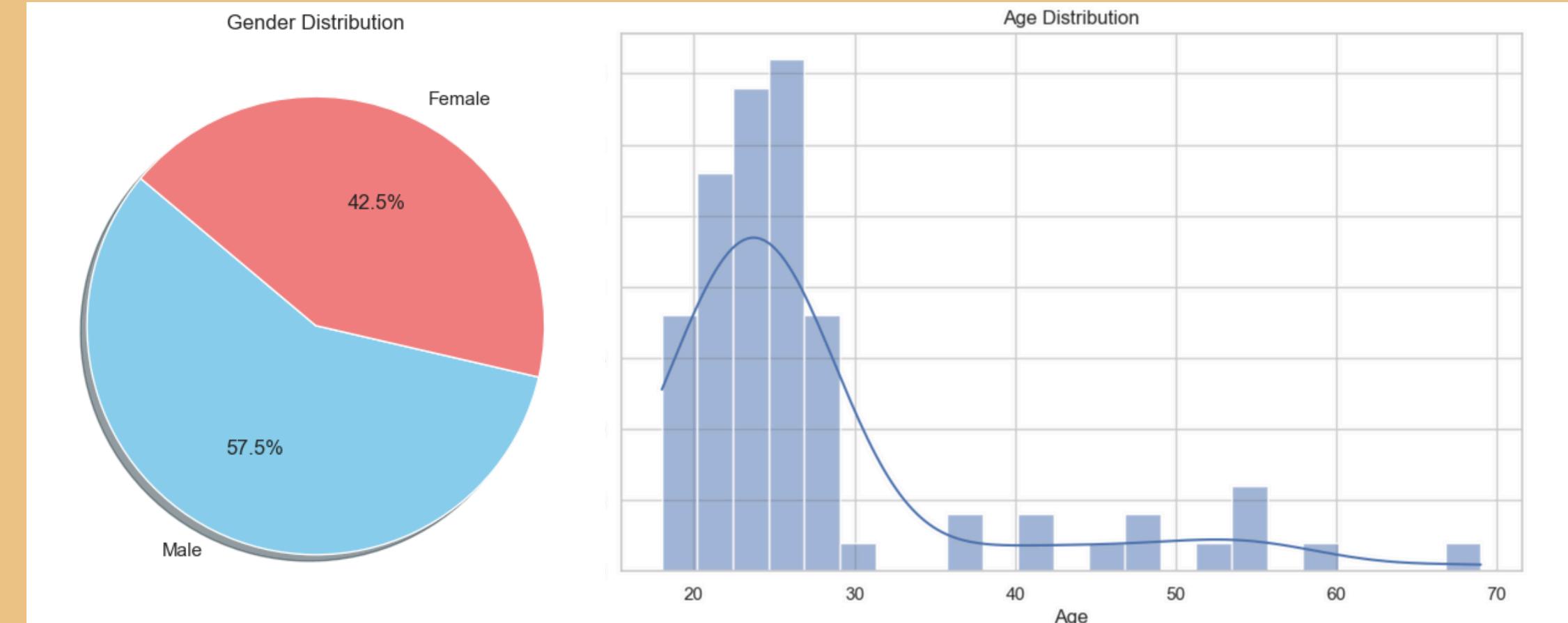
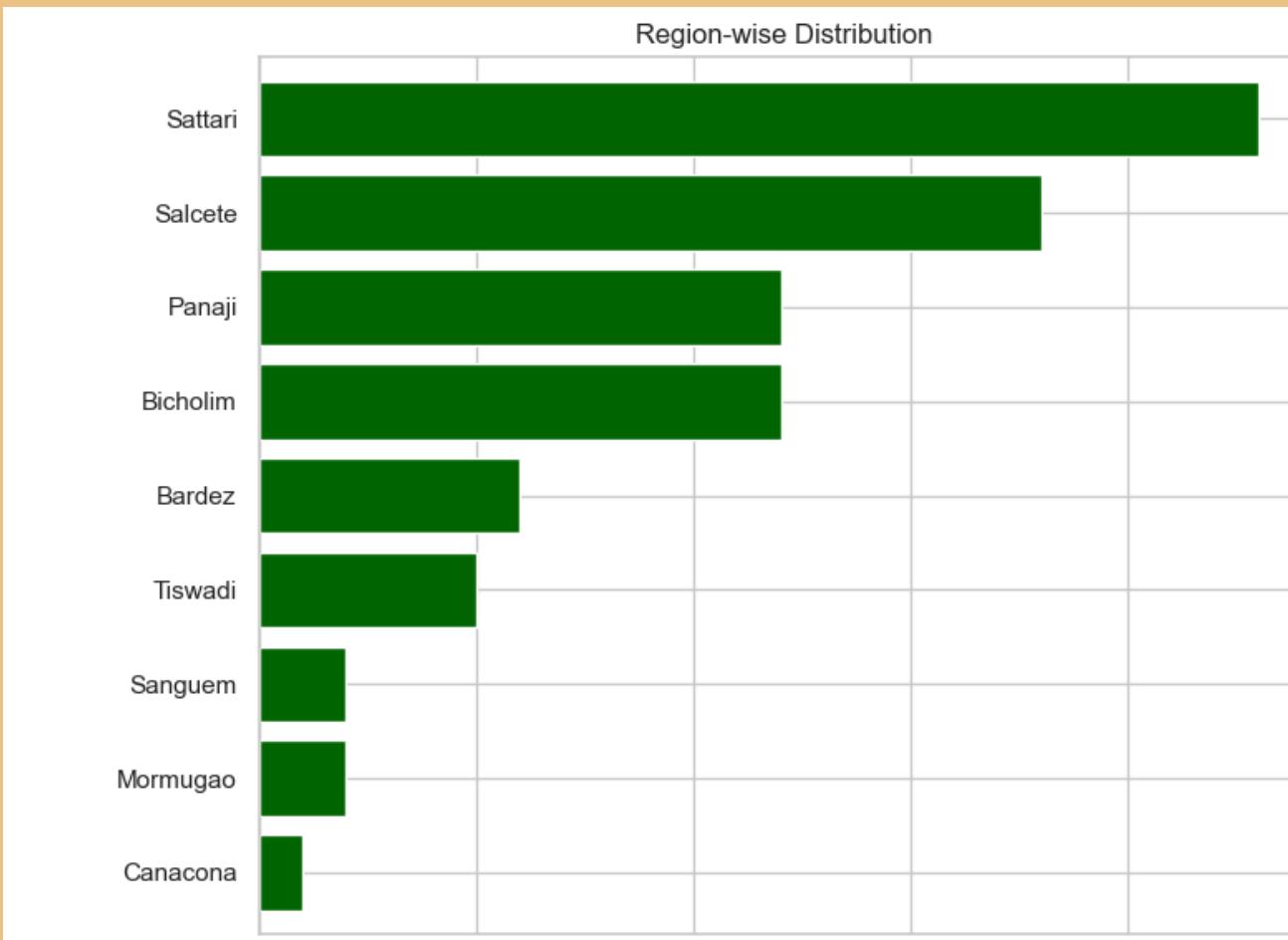
- **Data Preprocessing:** Dropped unnecessary columns, renamed for clarity, handled missing values, tokenized, lemmatized, and removed stopwords.
- **EDA:** Analyzed age distribution, regional responses, unique aspects, preservation acts, and challenges via histograms and bar plots.
- **Sentiment Analysis:** Used TextBlob & VADER to classify sentiments (Positive, Neutral, Negative) and visualized distributions for Fears & Hopes.
- **Topic Modeling:** Applied TF-IDF, LDA, and NMF to extract key topics from Fears & Hopes.
- **Visualizations:** Sentiment histograms, demographic sentiment variations, and word clouds for key themes.



KEY FINDINGS

Age & Gender Distribution

- Majority of respondents are young adults (20-35 years)
- Slightly skewed distribution, indicating varied participation across age groups
- Male respondents constituted around 58% while female respondents were 42%



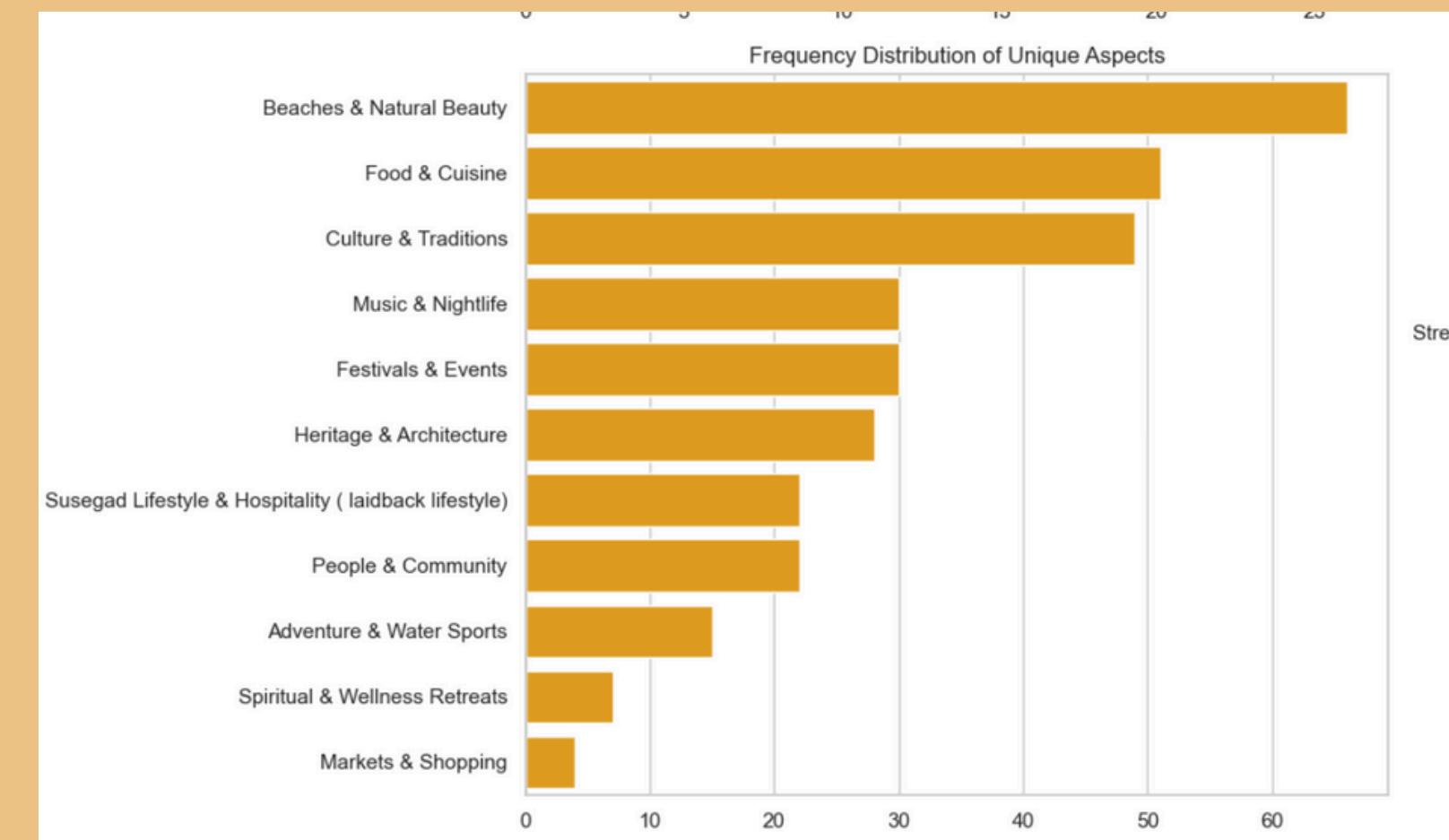
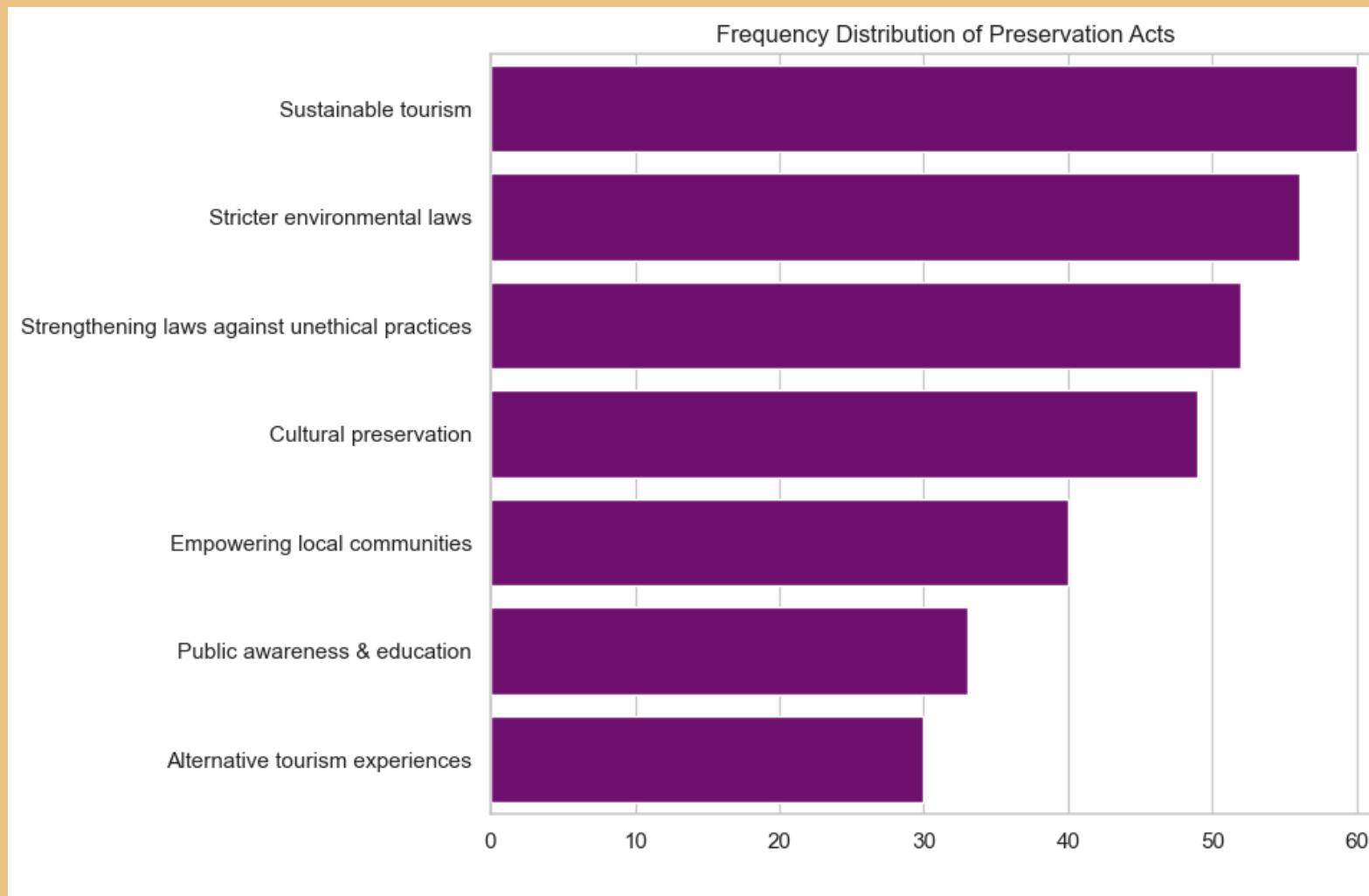
Region-wise Distribution

- Respondents were observed residing in or belong from a wide range of regions within Goa
- Sattari, Salcete and Panaji were the top 3 regions identified in terms of numbers



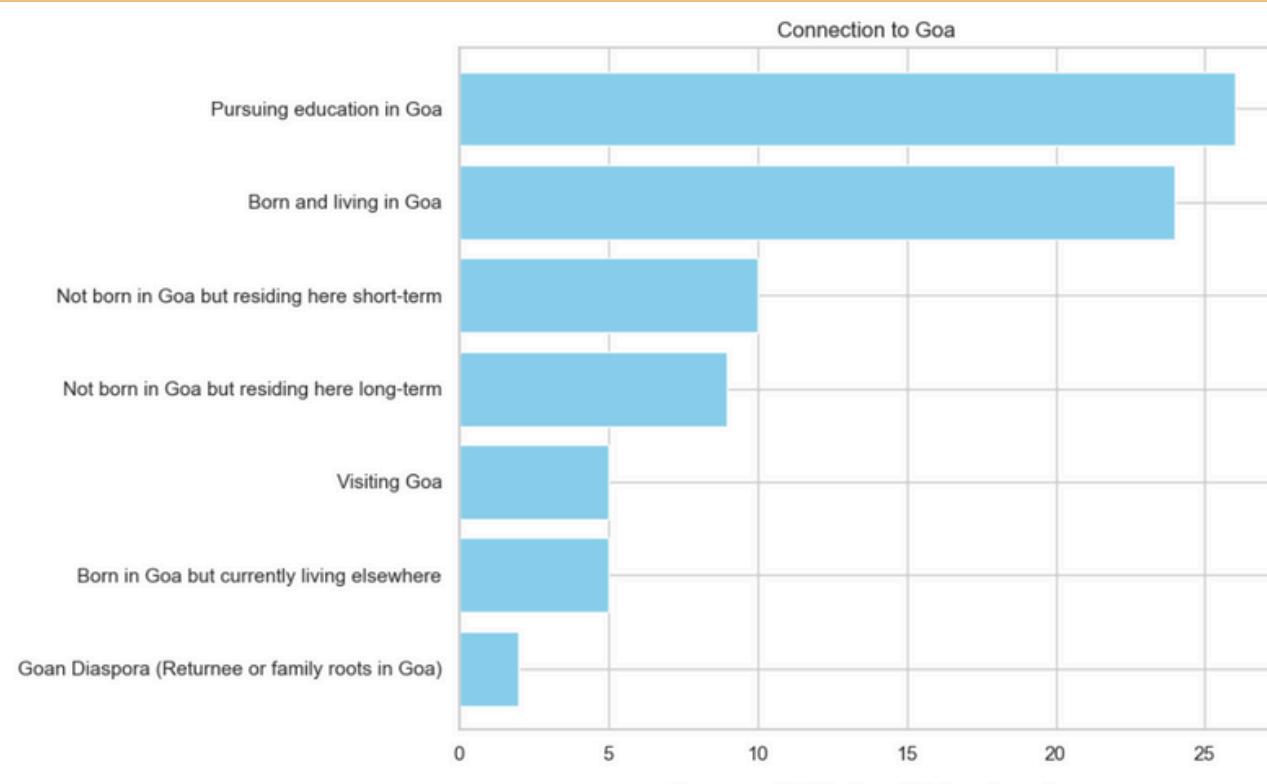
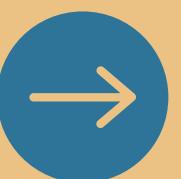
Unique Aspects

- Beaches & nature, food, and culture are the most valued aspects of Goa.
- Markets & shopping and spiritual retreats saw the least preference



Preservation Acts

- Most people think that sustainable tourism & stricter environmental laws are key to preserving the essence of Goa
- Other opinions such as cultural preservation, public awareness & empowerment of communities also had fair share of distribution

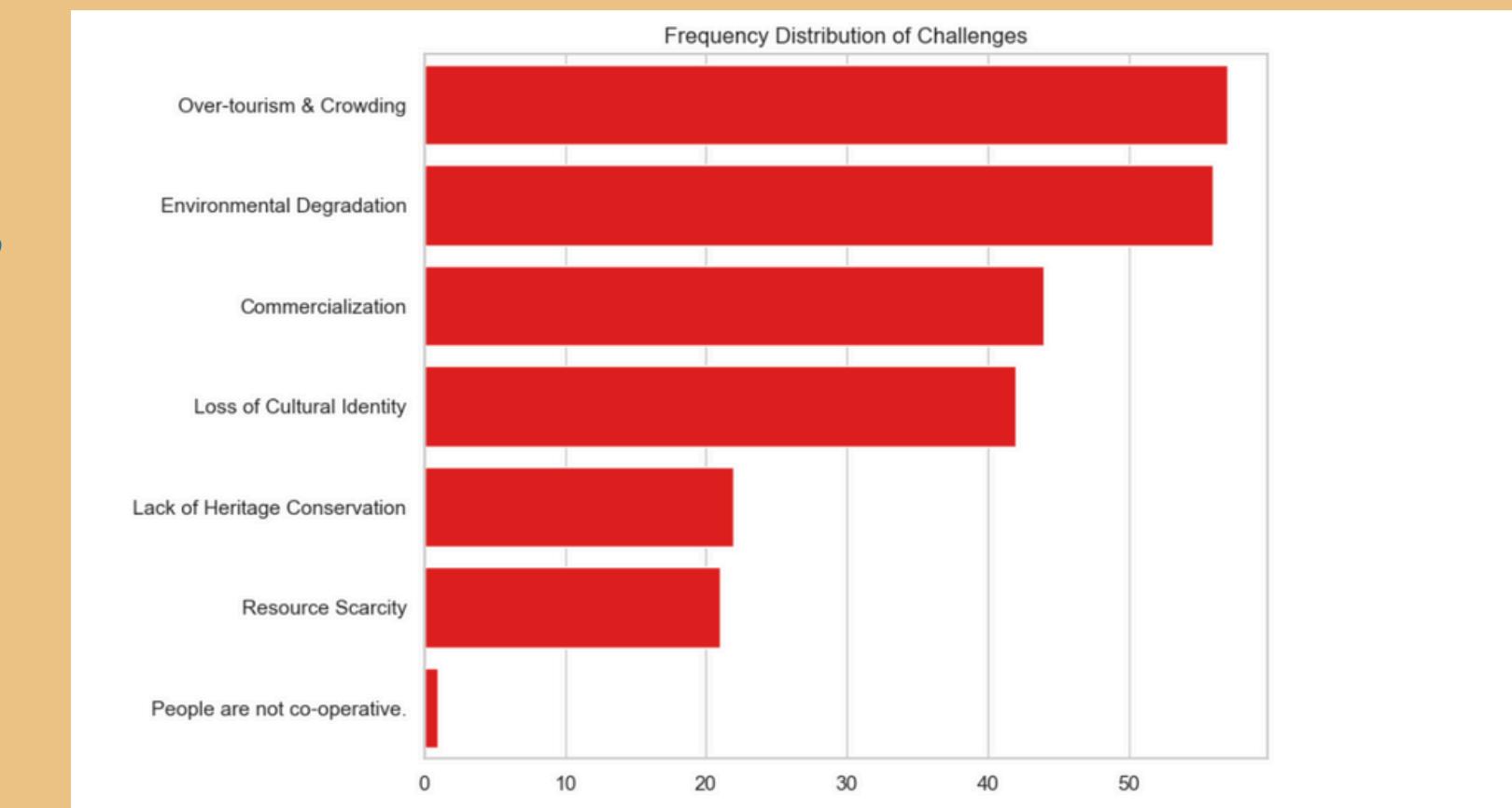


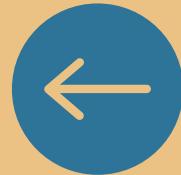
Connection to Goa

- Respondents have a strong personal or ancestral connection to Goa
- Significant variation in responses from different regions, reflecting diverse perspectives
- Most numbers of the respondents are students pursuing education in Goa

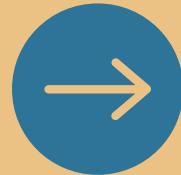
Challenges Identified

- Over-tourism, environmental degradation on beaches, and commercialization are the top concerns among respondents
- People also worry about Goa losing its natural charm and local culture





LDA & NMF TOPIC MODELLING



- LDA (Latent Dirichlet Allocation) is a technique used to automatically discover hidden topics in a collection of text documents. It works by identifying patterns of word usage and grouping words that frequently appear together into topics, then estimating how much of each topic is present in each document.
- NMF (Non-negative Matrix Factorization) is a technique used to uncover hidden patterns or topics in text data by breaking down a large word-document matrix into smaller, easier-to-understand parts. Like LDA, NMF helps in organizing and summarizing large text collections and is useful in applications such as topic modeling, document clustering, and recommendation systems.
- LDA & NMF was applied to the text columns of hopes and fears to identify key theme and top words used by respondents to answer the question
- Key Themes for Fears:
 - > Environmental degradation (beach, preserve, beauty)
 - > Cultural erosion / community impact (local, people, place)
 - > Over-commercialization through tourism (Goa, tourism, better)
- Key Themes for Hopes:
 - > Sustainable development & eco-tourism (preserve, beauty, beach, tourism)
 - > Community empowerment (local, people, place)
 - > Optimism for a better future for Goa (hope, better, place)

Top topics for fears (LDA):

Topic 0: local beach place preserve hope goa tourism better people beauty
Topic 1: beauty better goa local hope tourism beach preserve people place
Topic 2: goa tourism people preserve hope better local beach place beauty

Top topics for fears (NMF):

Topic 0: goa local beach beauty place better hope tourism preserve people
Topic 1: tourism preserve hope better beauty place people local goa beach
Topic 2: people place beach better tourism preserve local hope goa beauty

Top topics for hopes (LDA):

Topic 0: hope goa place local preserve beauty people tourism beach better
Topic 1: better beauty goa preserve place hope beach tourism local people
Topic 2: beach tourism people preserve place goa hope beauty better local

Top topics for hopes (NMF):

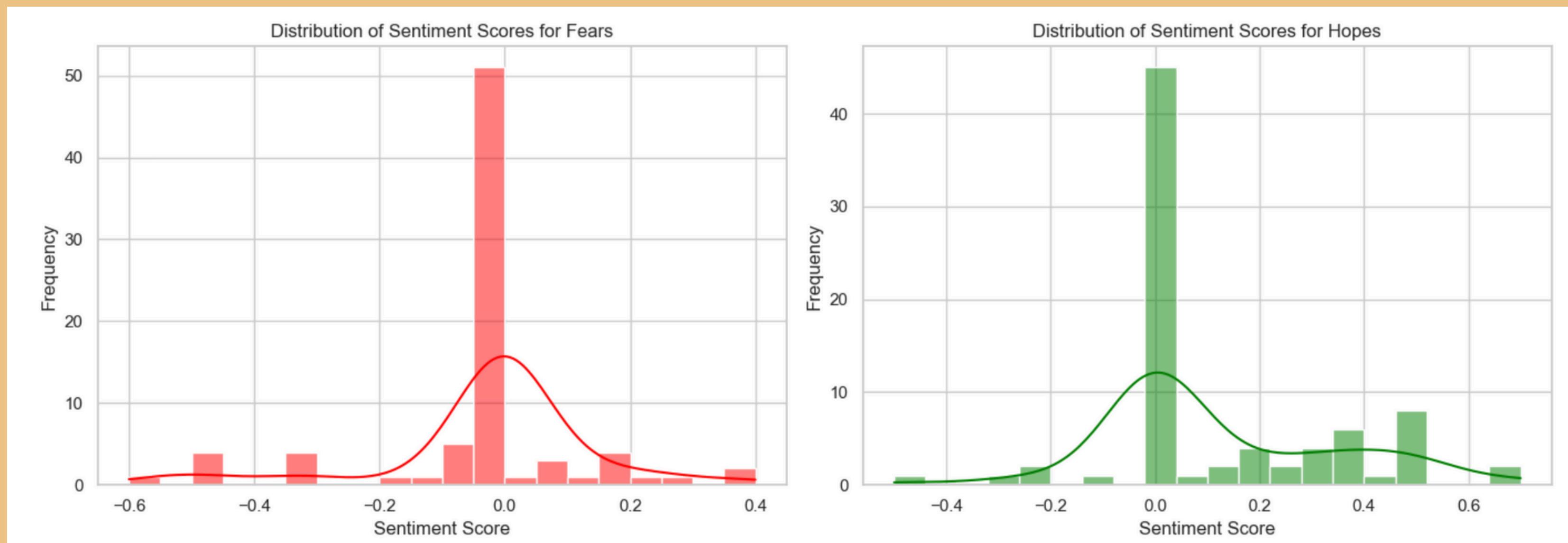
Topic 0: hope goa place beauty people local preserve tourism better beach
Topic 1: beach tourism preserve place beauty goa local people hope better
Topic 2: better tourism place goa preserve people local hope beauty beach



SENTIMENT ANALYSIS (HOPES VS FEARS)

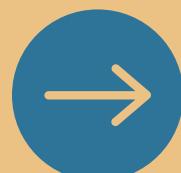


- Fears: Majority of responses are negative due to concerns about rapid urbanization, environmental damage, and loss of cultural identity
- Hopes: More positive sentiments, indicating optimism for sustainable tourism, better governance, and cultural revival
- Overall, the frequency distribution was centered around the mean, indicating that most respondent had a generally neutral feeling about hopes and fears regarding Goa

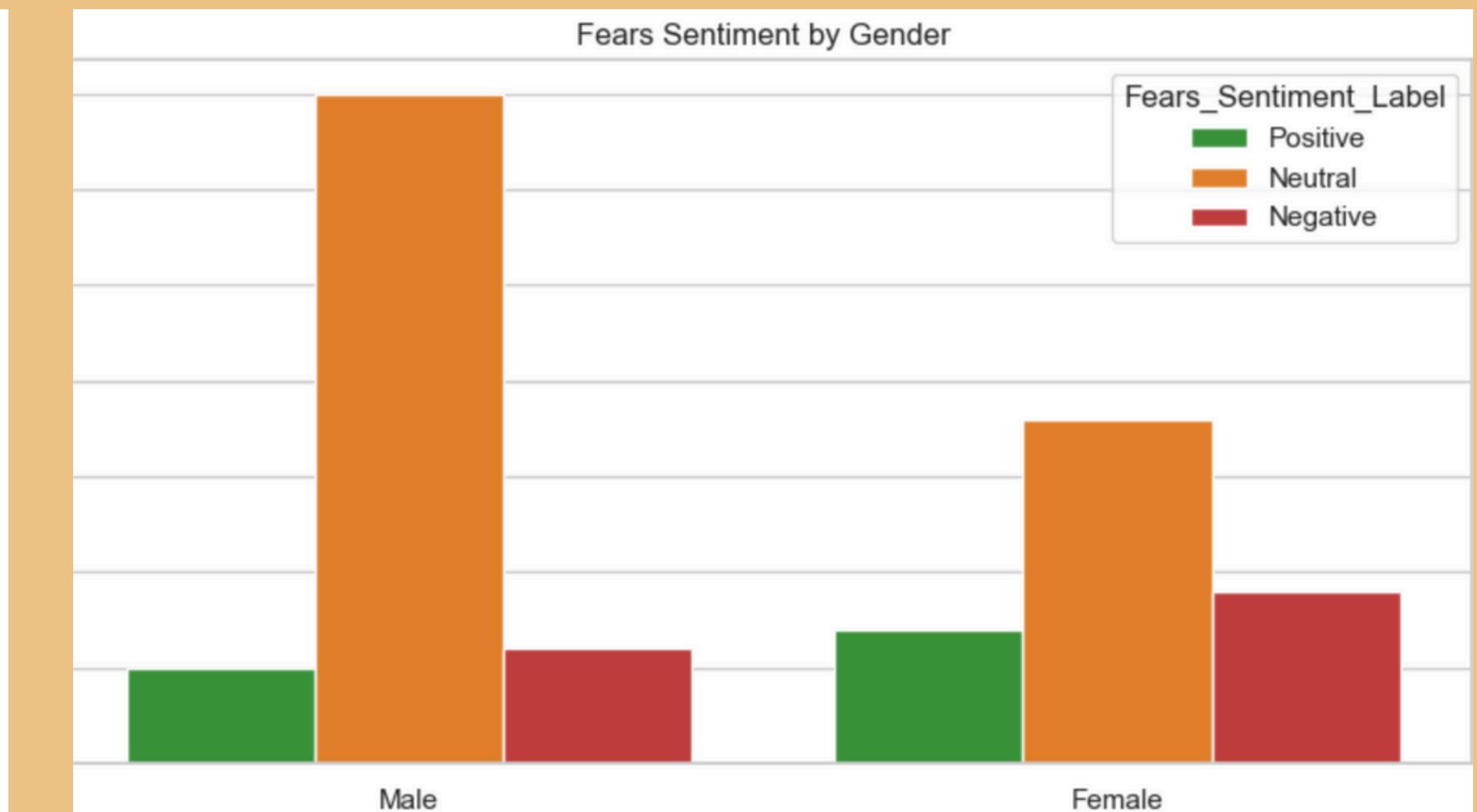
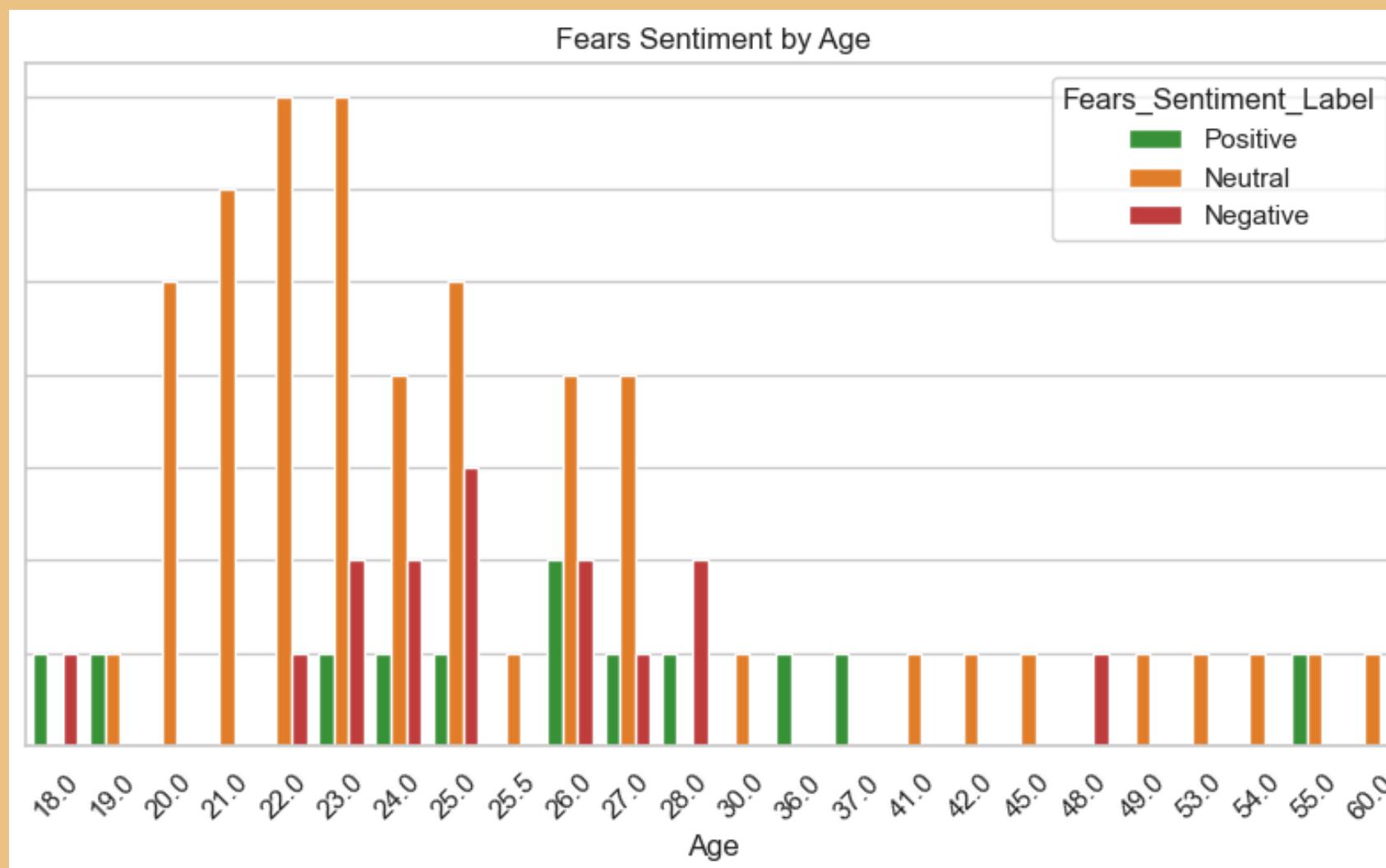




FEARS SENTIMENT BY AGE & GENDER



- Younger individuals express more concerns (negative sentiment) about Goa's future.
- Older age groups show a more balanced sentiment, with both concerns and hopes.
- No major difference in sentiment distribution between genders. However, neutral sentiment was dominant across categories

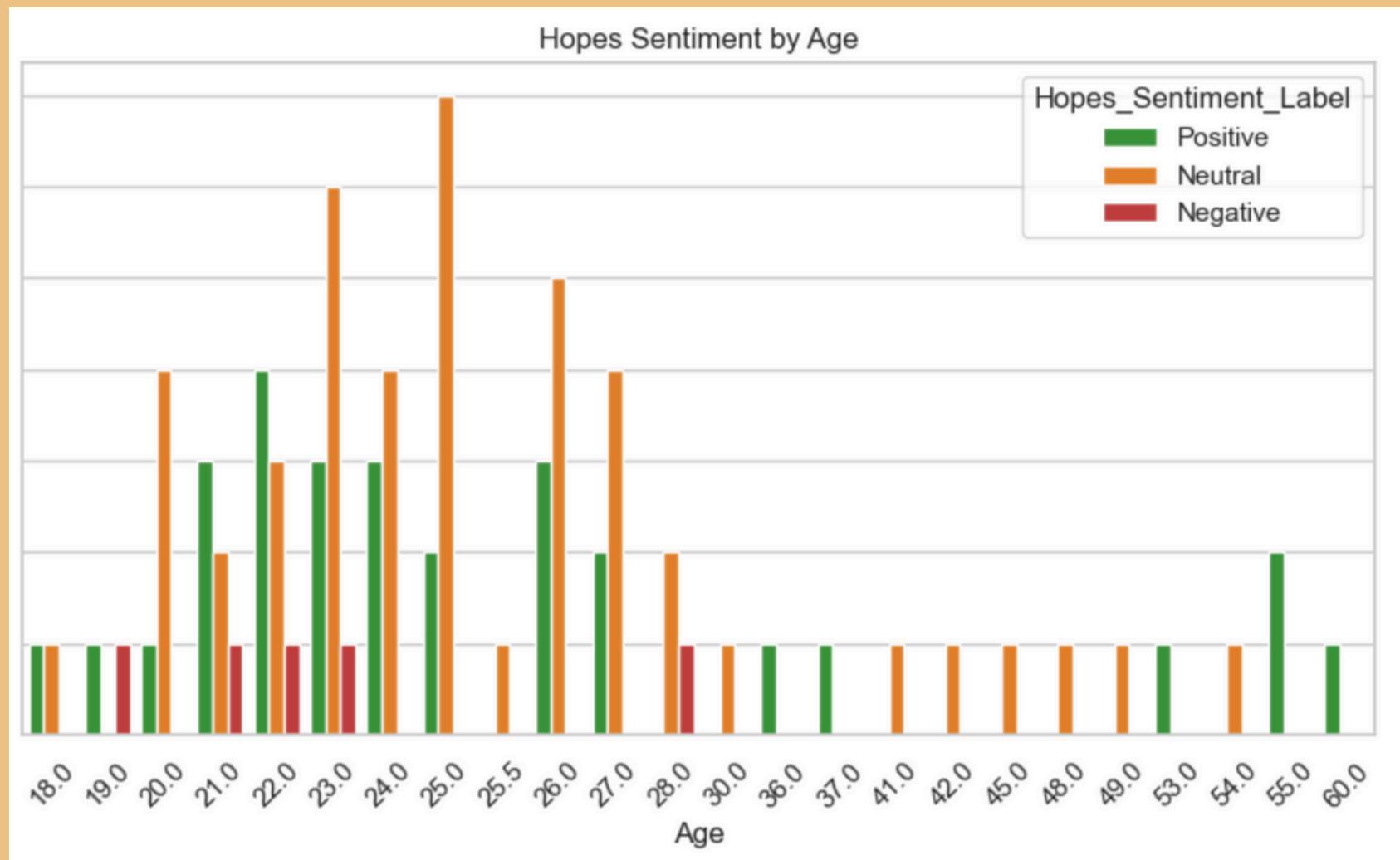




HOPES SENTIMENT BY AGE & GENDER



- Younger groups are more optimistic, while older individuals show balanced sentiments.
- Males lean neutral, while females are more polarized.
- Overall, positive sentiment dominates, reflecting optimism about Goa's future.



WORD CLOUD INSIGHTS



- Fears: Words like "pollution," "overcrowding," and "culture loss" dominate.
 - Hopes: Words like "sustainability," "protection," and "heritage" appear frequently, showing a desire for positive change.



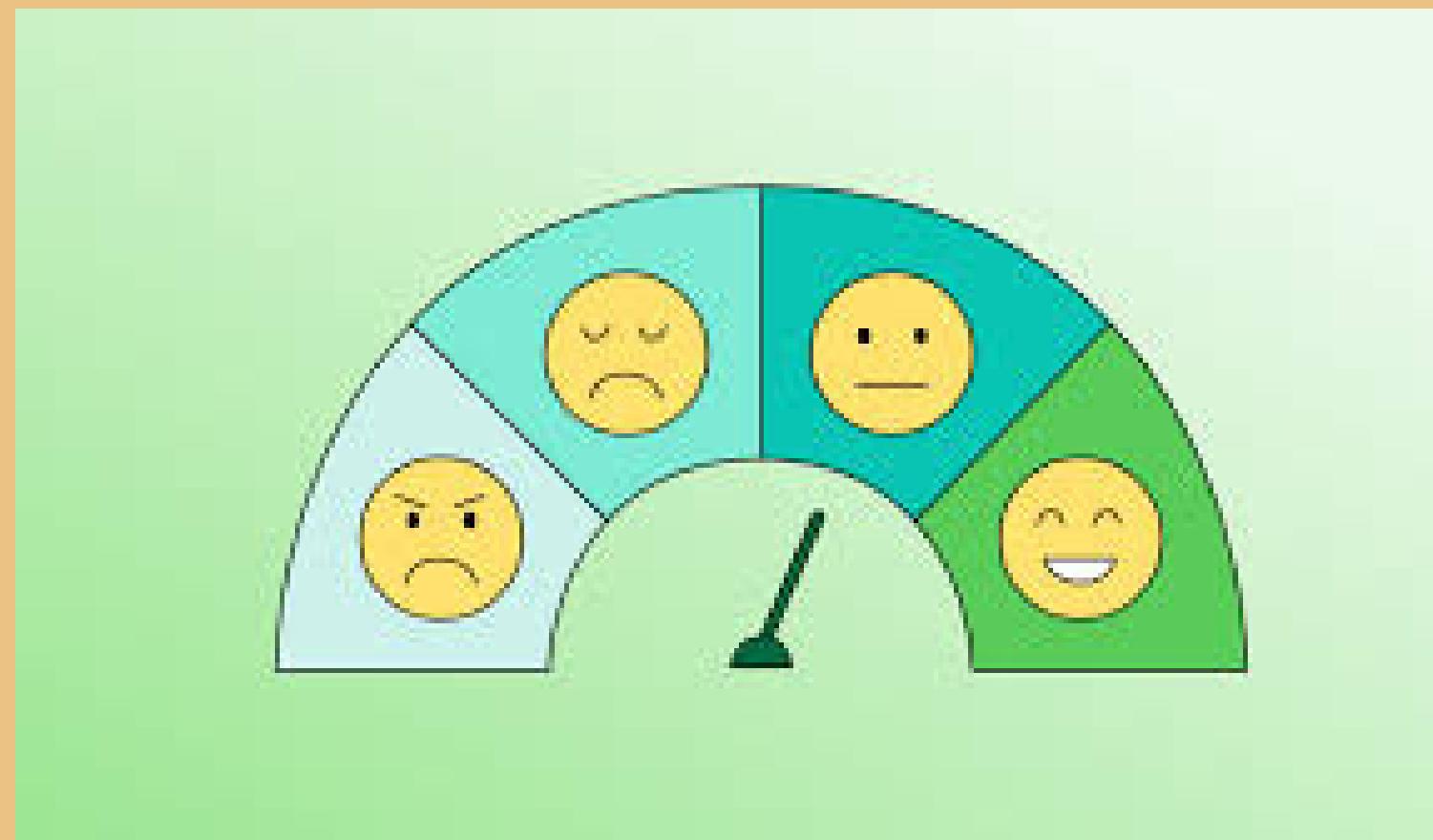
REDDIT POSTS ANALYSIS

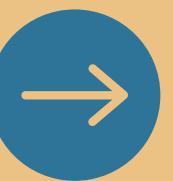
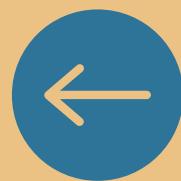


- This analysis focuses on Reddit posts from the r/Goa subreddit to understand discussions, opinions, and sentiments.
- We used Natural Language Processing (NLP) techniques to analyze text data and derive meaningful insights.
- Multiple sentiment analysis models were applied to compare results.

OBJECTIVE

- Identify key topics discussed in the subreddit.
- Analyze sentiment trends (positive, neutral, negative).
- Compare different sentiment models (TextBlob, VADER, RoBERTa, BERT).
- Provide data-driven insights about public perception and concerns.





METHODS USED

- Web scraping of “r/Goa” subreddit using praw library & Reddit API
- Text preprocessing
- Topic Modeling using TF-IDF vectorization and k-means clustering
- Word Cloud to visualize frequently used words.
- Sentiment Analysis: Measure emotions using:
 1. TextBlob (Simple rule-based Naive Bayes polarity score)
 2. VADER (Lexicon-based + Heuristics sentiment scoring)
 3. RoBERTa & BERT (Transformer-based deep learning models)

OUTCOME OF ANALYSIS

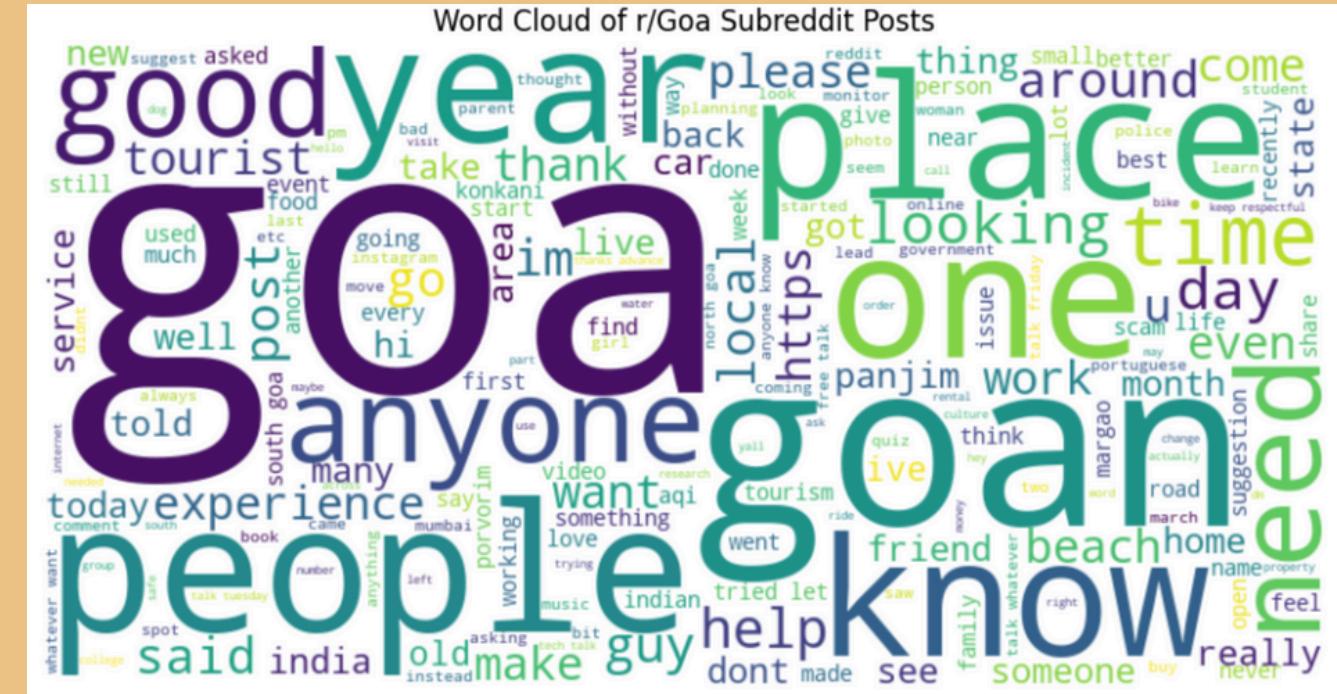


Topic Modelling results:

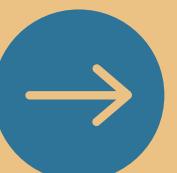
```
[nltk_data] Downloading package stopwords to  
[nltk_data]      C:\Users\ADVIK\AppData\Roaming\nltk_data...  
[nltk_data] Package stopwords is already up-to-date!  
Topic 1: last, place, internet, fish, nuste, thali, today, christmas, guess, sunset  
Topic 2: want, keep, respectful, whatever, welcome, tech, tuesday, free, friday, talk  
Topic 3: know, anyone, meetup, litter, interfaith, wedding, kakolem, near, goa, beach  
Topic 4: one, place, kind, boi, anyone, share, really, service, goa, good  
Topic 5: one, north, like, people, state, see, im, goans, goa, south  
Topic 6: dogs, time, https, watch, konkani, car, local, goa, get, goan  
Topic 7: looking, im, anyone, would, week, know, goa, help, please, need  
Topic 8: like, people, want, would, work, know, year, looking, new, goa  
Topic 9: hoping, owl, belong, cuckoo, know, nice, pic, species, feather, bird  
Topic 10: around, groups, play, pm, com, sunday, football, seqc, quiz, panjim
```

- Topic 1: Social experiences and local food culture, with festive and scenic elements.
 - Topic 2: Casual conversations blended with tech and weekly themed discussions.
 - Topic 3: Community engagement, meetups, and local event coordination.
 - Topic 4: Service reviews and place recommendations shared by users.
 - Topic 5: Regional identity, people's opinions on North vs South Goa, and local sentiments.
 - Topic 6: Local culture, pets, Konkani language, and media sharing.
 - Topic 7: General queries and community help-seeking behavior.
 - Topic 8: Conversations about work-life balance, aspirations, and settling in Goa.
 - Topic 9: Birdwatching and nature-related discussions by wildlife enthusiasts.
 - Topic 10: Group events, sports, and quiz-related social activities.

Word Cloud results:



- Frequent use of words like “people,” “anyone,” “help,” “know” reflects a highly interactive and supportive community.
 - Keywords such as “Goa,” “place,” “beach,” “panjim,” “south goa” suggest strong interest in travel, local attractions, and geographical queries.
 - Terms like “experience,” “service,” “food,” “stay,” “tourist” indicate users actively post reviews and suggestions about services and experiences in Goa.
 - Mentions of “tourist,” “local,” “goan,” “food,” “beach,” “konkani” highlight discussions on Goan identity, language, traditions, and tourist activities.



OUTCOME OF ANALYSIS

TextBlob results:

```
Number of positive sentiments = 511  
Number of neutral sentiments = 277  
Number of negative sentiments = 140
```

VADER results:

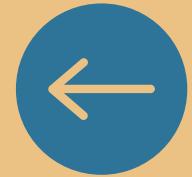
```
[nltk_data]  Package vader_lexicon  
Number of positive sentiment: 530  
Number of neutral sentiment: 234  
Number of negative sentiment: 164
```

RoBERTa results:

```
Device set to use cpu  
RoBERTa Sentiment  
LABEL_1    549  
LABEL_2    241  
LABEL_0    138  
Name: count, dtype: int64
```

BERT results:

BERT Sentiment	
1 star	376
5 stars	262
4 stars	138
3 stars	115
2 stars	37



FINAL INSIGHTS

- People love Goa's culture and natural beauty, but fear it is at risk due to commercialization and over-tourism
- Optimism exists, but strong governance, sustainable tourism, and cultural preservation are key to Goa's future
- Policymakers should focus on balancing economic growth with environmental and cultural conservation
- There was a fairly even mix of positive and negative sentiments around hopes and fears
- Majority of reddit posts are positive (~50-60%) across all models.
- RoBERTa & VADER found more positive sentiment, while BERT detected more negative sentiment.



THANK YOU

