Discussion
Group 6 — Yiqin Zhou, Jiaqi Wei, Jiachen Gao, Zihao Huang

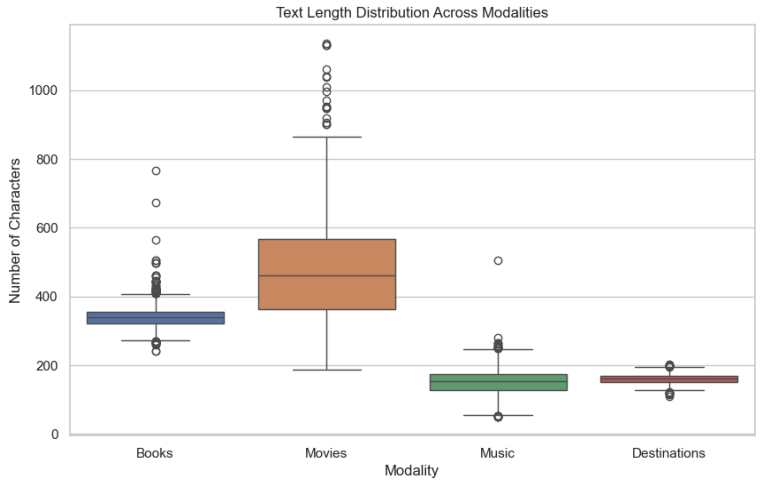
1. Introduction

This project develops a multi-modal travel recommendation system that maps user preferences in movies, books, and music to semantically aligned travel destinations. The system is motivated by the observation that cultural media encode rich atmospheric and emotional signals—such as tone, genre, themes, and mood—that reflect users' experiential preferences. By extracting semantic embeddings from diverse cultural artifacts and aligning them with city descriptions, the system generates personalized and conceptually grounded travel recommendations.

To support this goal, we curated a high-quality dataset consisting of movies from TMDB, books from Goodreads, tracks from Spotify, and global destinations from GeoNames and Wikipedia. Rather than collecting large but noisy datasets, we deliberately prioritized data quality, semantic richness, and diversity to ensure stable performance in lightweight deployment environments such as Hugging Face Spaces. The project includes end-to-end components for data preprocessing, embedding generation, semantic search, fallback strategies for deployment constraints, and a fully functional front-end interface.

2. Exploratory Data Analysis

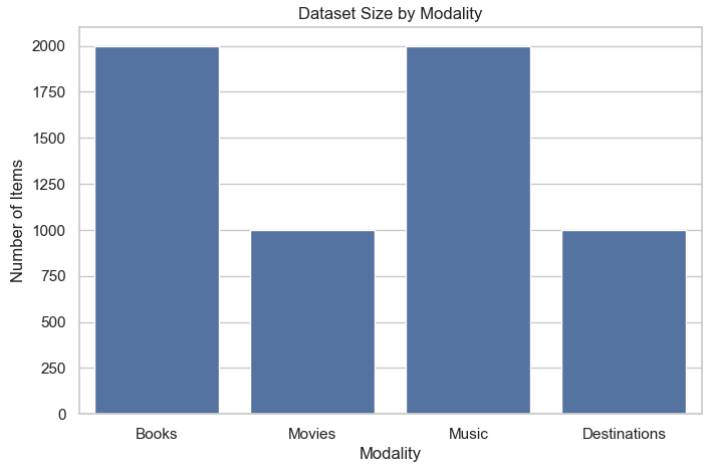
Our EDA focused on assessing dataset quality, representation balance, and semantic suitability for embedding-based retrieval.



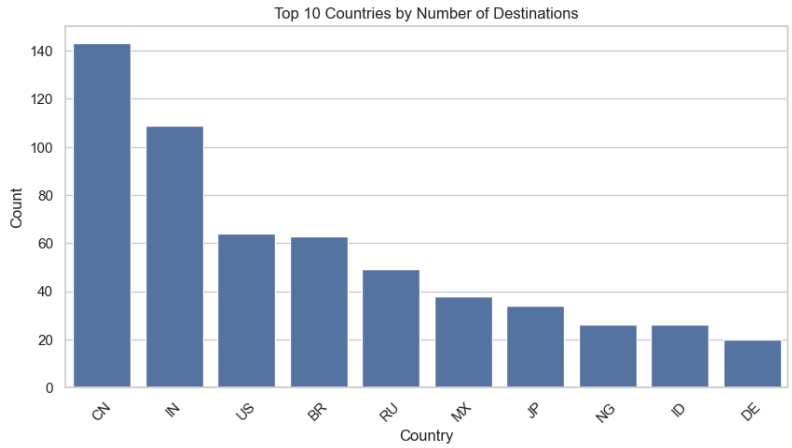
Across all media types, we observed substantial variability in text description length. Movie overviews averaged 200–300 characters, book descriptions varied greatly (from one sentence to

multi-paragraph summaries), and Spotify track metadata often contained minimal descriptive information. This unevenness influenced how much semantic detail each domain contributed to the shared embedding space.

Destination data exhibited similar heterogeneity. Some locations had detailed multi-paragraph Wikipedia summaries, while others only contained short factual statements. To ensure suitability for semantic matching, we retained only destinations with structured names and meaningful textual descriptions. We also filtered out non-touristic entities such as administrative regions, transport hubs, or ambiguous geographic terms.



In terms of dataset size, the curated collection consists of several thousand media items and around 1,000 global destinations. Distribution analysis confirmed an imbalance across genres and regions (e.g., more English-language media, concentration in certain tourist-heavy countries). These characteristics informed later design choices in stratified sampling, debiasing strategies, and fallback retrieval logic.



3. Preprocessing and Engineering Decisions

- Data Cleaning and Field Selection

We retained only text-rich fields essential for semantic modeling: movie *title* and *overview*, book *description*, music *track name* and *artist metadata*, and destination *name* and *summary*. Items lacking meaningful descriptions were removed to avoid producing weak or noisy embeddings. We also eliminated duplicates, empty titles, or records referring to non-touristic entities.

- Stratified Sampling with TF-IDF Reweighting

To avoid popularity bias and promote diversity, we implemented a stratified sampling strategy combined with TF-IDF filtering. Media items were categorized into four semantic-quality tiers—high-quality classics (40%), unique strong items (30%), genre-coverage items (20%), and niche long-tail items (10%). Within each tier, TF-IDF was used to prioritize samples with richer textual content and clearer semantic signals.
This process ensured balanced coverage across genres, avoided overrepresentation of mainstream works, and increased serendipity in downstream recommendations.

- Transition from .npy to .npz Embedding Files

During deployment, Hugging Face Spaces flagged .npy files as potential unsafe binaries. To ensure smooth deployment and reduce file size, we migrated all embeddings to .npz format, which is compressed, safer, and equally easy to load. This change improved stability without altering algorithmic behavior.

- Image Source Migration (Unsplash → Pexels)

High-quality destination images were essential for the front-end experience. Unsplash API rate limits and licensing restrictions led us to switch to Pexels, which offered more permissive usage terms and better integration. However, Pexels occasionally returned unrelated images for small or ambiguous locations—highlighting a data limitation described later.

- Embedding Strategy and Multi-Modal Fusion

We used Sentence-BERT (all-MiniLM-L6-v2) to generate 384-dimensional embeddings for all media and destination texts. For multi-modal queries (movie + book + music), we adopted a simple weighted fusion of embedding vectors. Equal weights performed robustly, while preserving the option for future user-adjustable preferences.

- FAISS to NumPy Fallback in Cloud Deployment

FAISS operates efficiently on local machines but is unsupported in Hugging Face Spaces due to binary installation restrictions. To maintain cross-platform consistency, we implemented a fallback: when FAISS is unavailable, the system computes inner-product similarity directly using NumPy. Though slower, this remains effective given our relatively small dataset (~1k destinations) and allows full deployment without GPU or compiled dependencies.

- Logging

We incorporated lightweight logging throughout the retrieval and deployment pipeline to monitor embedding loading, fallback behavior, and runtime errors, which helped ensure system stability in constrained cloud environments.

4. Baseline Method and Improvements

- Baseline: TF-IDF + Cosine Similarity

Our early prototype used TF-IDF features for media and destination descriptions. While the baseline could match literal keywords, it failed to capture deeper thematic or atmospheric relationships. Recommendations exhibited lexical bias, often linking destinations and media purely on shared nouns.

- Improved Method: Transformer-Based Embeddings + Weighted Fusion

Replacing TF-IDF with Sentence-BERT substantially improved semantic alignment. Movies, books, and music with similar moods clustered in vector space, even without overlapping vocabulary. Weighted fusion of the three modalities allowed more nuanced preference modeling. Semantic search using FAISS/NumPy rankings produced coherent, diverse recommendations, and the system demonstrated meaningful cross-media alignment—for example:

- contemplative or melancholic music → coastal or quieter historical towns
- fantasy novels → destinations with rich cultural or architectural heritage
- energetic pop music + adventure movies → nightlife-oriented global cities

This improvement addressed nearly all limitations observed in the baseline.

5. Evaluation

Traditional accuracy-based evaluation is unsuitable for this task because no ground-truth destination labels exist. Instead, we evaluated the system using qualitative, structural, and user-centric methods:

(1) Content Coherence

Manual inspection confirmed strong alignment between input themes and recommended destinations. Multi-modal queries especially produced rich, interpretable results.

(2) Diversity and Geographic Coverage

Our stratified sampling and debiasing techniques reduced overconcentration in popular Western countries. Top-k destination sets generally showed meaningful geographic diversity.

(3) Robustness Tests

Input perturbation tests—substituting one media item with a similar one—produced stable, predictable changes in recommendations, indicating consistent semantic structure.

(4) User Feedback Logs

The system logs feedback locally, enabling analysis of user-rated relevance. Early testers reported positive experiences, especially praising unexpected but reasonable creative matches. Cloud deployment does not persist logs, but the mechanism supports further iterative refinement.

(5) Qualitative Case Studies

Case studies demonstrated that cross-media fusion often generated more specific and interesting travel suggestions compared to any single-media input alone.

6. Adaptation to Instructor and Peer Feedback

Throughout development, we incorporated feedback that shaped the final architecture:

- **Clarifying the connection between cultural preferences and travel behavior.**
  We strengthened our explanation through embedding visualization and alignment analysis.
- **Addressing deployment constraints.**
  Following instructor advice, we redesigned our retrieval pipeline to support non-FAISS cloud environments via NumPy fallback.
- **Improving data diversity and reducing popularity bias.**
  Peers suggested avoiding domination by mainstream media; this motivated our stratified sampling and TF-IDF refinement approach.
- **Enhancing user experience.**
  We shifted from Unsplash to Pexels, added Wikipedia summaries, and improved front-end responsiveness.

These adaptations improved system reliability, interpretability, and alignment with project goals.

7. Challenges and Limitations

Several challenges emerged during development:

- Data Quality and Coverage

Media datasets were relatively small, limiting semantic diversity. Some destinations were culturally inappropriate or politically unstable—highlighting the need for future safety filters (e.g., travel advisories, blacklist regions).

- Image Search Limitations

Pexels occasionally retrieved unrelated images for small cities, underscoring a dependence on third-party search relevance.

- Deployment Constraints

FAISS incompatibility required fallback solutions. Hugging Face storage quotas necessitated the use of compressed .npzfiles and smaller datasets.

- Lack of Real-Time Travel Itinerary Generation

We initially planned to incorporate LLM-generated travel guides but deferred due to API costs and time constraints. This remains a major opportunity for future enhancement.

8. Conclusion

This project demonstrates that cultural preferences expressed through movies, books, and music can be effectively transformed into personalized travel recommendations using modern NLP techniques. Our system integrates high-quality multi-modal datasets, transformer-based embeddings, debiasing strategies, efficient similarity retrieval, and a user-friendly interface. Despite deployment constraints and data limitations, the final system is stable, interpretable, and extensible. Future iterations will expand dataset scale, integrate safety and relevance filters, enhance image sourcing, and incorporate LLM-based itinerary generation to create a richer, more practical travel planning tool.

References

- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.
- Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data.
- The Movie Database (TMDB) API: https://www.themoviedb.org/documentation/api
- Spotify Tracks Dataset. Kaggle: https://www.kaggle.com/datasets/maharshipandya/spotify-tracks-dataset
- Goodbooks-10k Dataset. Kaggle: https://www.kaggle.com/datasets/zygmunt/goodbooks-10k
- GeoNames Geographical Database: https://www.geonames.org
- Wikipedia API: https://www.mediawiki.org/wiki/API:Main_page