

**Team: Hitachi #2**

Weekly Report #7

04/08-04/14

Team Members: Members: Jay Naidu, Daniel Day, Andrew Urquhart, and Kebba Leigh

# Summary

This week, our team finalized major backend data tasks and began merging front-end components with our analytical pipeline. Substantial progress was made in sentiment analysis, embedding expansion, and exploratory visualizations. The UI prototype continues to evolve toward a polished, responsive website structure, and we’re preparing to integrate graphs and clustering interactions powered by Pinecone and Power BI.

# Project Status

* The project is organized into four primary phases: Data Extraction, Cleaning, Vectorization, and User Integration & Visualization.
* Data Extraction - 100% |Daniel finalized the cleaned dataset and added the final year’s data.
* Cleaning -100 % | Dataset is finalized and cleared of all duplicates and foreign/incomplete entries.
* Vectorization - 85% | Jay and Andrew are adding new embeddings and performing cluster and sentiment analysis.
* UI & Visualizations - 90% | Kebba developed a finalized Frontend and Backend. He is currently integrating Power BI dashboards.

## Key Achievements This Week

* **Daniel**
  + Cleaned up code and added additional documentation.
  + Finished the User Manual for the 100% complete portion of the project.
* **Andrew**
  + Embedded new metadata fields for filtering (genre, decade, etc.).
  + Fixed uploading error
  + Used T-SNE Cluster to visualize data
  + Begin to export data for visualization
* **Kebba**
  + Finalized Figma UI for all five pages of the site (Home, Search, Cluster Visualizations, About, Feedback).
  + Set up front-end structure to match Figma using mobile-responsive layout guidelines.
  + Began Power BI dashboard development for real-time lyric and metadata cluster visualizations.
  + Created custom “How It Works” vector visualization using GPT + Pinecone diagram.
* **Jay**
  + Performed sentiment analysis using VADER across genres and time.
  + Initiated similarity analysis using KMeans clustering.
  + Began working on visualizations: heatmaps, sentiment histograms, and clustered bar charts.
  + Explored t-SNE/PCA visualizations on embedding space.

Meetings/Submissions:

* Meeting with Dr. Sadovnik & Additional Team Meeting afterwards on 04/22/2025

## Next Steps

* **Kebba**:
  + Begin front-end development using React based on Figma mockups.
  + Embed live Power BI dashboards using iframe or Power BI REST API.
  + Link filtering inputs (genre, sentiment, popularity) to Pinecone queries and dashboard visuals.
* **Daniel:**
  + Complete User Manual
  + Prepare Final Presentation
  + Offer additional support where needed
* **Jay:** 
  + Finish remaining visualizations and sentiment models.
  + Assist in embedding similarity scores into Pinecone and connecting results to front-end song cards.
* **Andrew:** 
  + Explore normalization of embedding distances and alternative vector search strategies.
  + Refine Visuals
  + Export T-SNE and other cluster visualizations

**Risks/Roadblocks**

* Website interactivity requires stable embedding + metadata indexing, which is still evolving.
* Some features (e.g., dynamic filtering of vectors in Power BI) may require workaround strategies or code embedding.