MAPPING STORIES IN THE WORLD

Sarah Chung, Sana Dadani, Chikodi Ohaya,
Eliska Peacock, and Sam Prieto Serrano

Exploring Computational Journalism
CS206/COMM281
Winter 2025

Introduction: Defining the Problem Space

Podcasts have become one of the most dynamic forms of modern storytelling, offering in-depth narratives, investigative journalism, and cultural insights in a highly accessible format. Unlike traditional news articles or television broadcasts, podcasts allow for immersive, long-form storytelling that often brings locations to life. Whether it's an episode about a historic landmark, a hidden gem in a city, or a personal story tied to a family-owned restaurant, there are many podcasts dedicated to location-based narratives. However, these location-based stories remain locked within saturated podcast platforms, making it difficult for listeners to discover, engage with, or explore these locations further.

Our project, Mapping Stories in the World, seeks to bridge this gap. We set out to build a platform that extracts locations from podcast transcripts and surfaces stories about those places based on where listeners are at any given moment. By leveraging Natural Language Processing (NLP), location entity extraction, and computational ranking methods, we aim to build a system that connects podcast content to physical locations, allowing listeners to experience journalism and storytelling as an interactive, place-based journey.

Project Goals

At its core, our project aims to enhance location-based storytelling by making podcast content more discoverable and interactive. We do this by:

- Extracting geographic references from podcast transcripts using NLP techniques
- Scoring location relevance based on factors such as frequency of mention and location cluster centrality

This system is particularly relevant for local and narrative-driven podcasts, such as KQED's *Bay Curious*, which answers listener-driven questions about Bay Area history, places, and culture. Our work aligns with broader efforts in geo-located storytelling, as seen in past projects, like Detour, Findery, and text-based news distribution systems, like Outlier Media.

Computational Journalism & Our Approach

Our work fits within the field of computational journalism by utilizing data science, AI, and digital storytelling techniques to enhance journalistic content. During the course of our work in Stanford's Exploring Computational Journalism (ECJ) class, we applied the fundamentals of computational journalism to our project by automating location extraction from unstructured podcast transcripts using LLMs and Python scripts. We also enhanced our project by using ECJ techniques to define the relevance of locations mentioned in podcasts and enhance audience interactivity.

By combining text analysis, geographic data, and audience engagement insights, our project builds upon the existing strengths of location-based journalism while addressing scalability challenges in automating geographic discovery.

Key Findings and Takeaways

We tackled our project goals by developing a system that extracts and ranks locations mentioned in podcast transcripts. Through iterative prototyping and computational methods, we identified several key insights:

1. *Defining Location Relevance* – A key breakthrough was developing a structured approach to ranking locations using four weighted factors: *frequency of mentions, presence in the episode title, geographic proximity to other locations, and importance of the location by transcript position*.

2. *Python Script* – Our rule-based Python prototype achieved 82% accuracy after refining our algorithm and dataset.

3. *LLM-Based Approach* – Testing a large-language model with industry best practices achieved accuracy of 95.2% but showed inconsistencies across runs.

4. *User Insights* – A survey of 68 *Bay Curious* listeners showed strong interest in location-based storytelling, with 57.4% wanting a tool to explore podcast locations, validating real-world applications of our project.

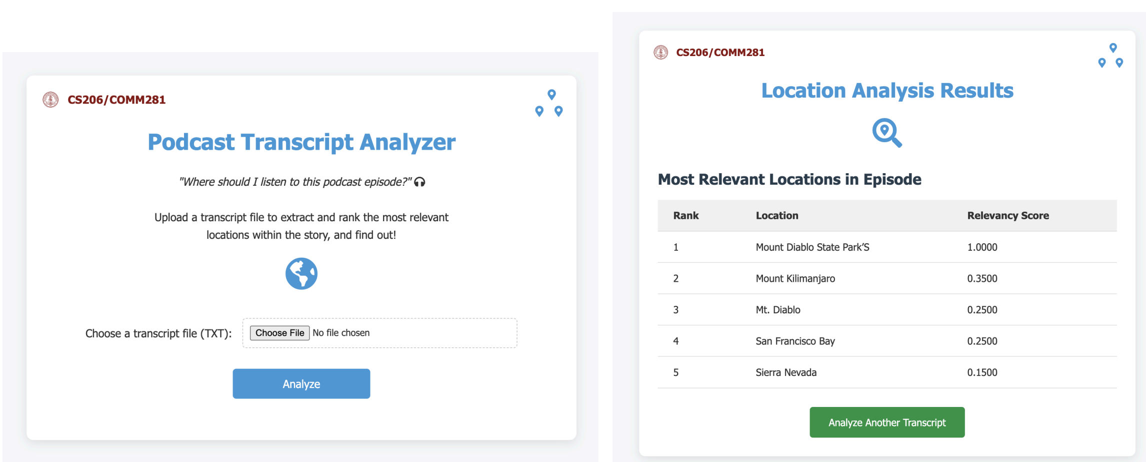
5. *Next Steps* – We propose a hybrid system integrating the adaptability of LLMs with the precision of Python methods. Future iterations may incorporate interactive mapping and location-based notifications to enhance user experience.

By combining computational journalism techniques with NLP advancements, our work helps bring location-based stories to life, making it easier to discover and explore places mentioned in podcasts at scale.

To access the code for the Python implementation please check out this public Github repo: https://github.com/samprietoserrano/cs206-mappingstories/tree/main. There you can find code for all three prototyping stages, along with the local website interface, the ground-truth dataset, and some evaluation/results logs.

Path to a Prototype: Rule-based Transcript Analyzer

Overview of our Final Tool



Landing page (input) and ranking page (output) from the Python tool

Our project concluded having developed a tool that analyzes podcast transcripts to determine the most relevant locations mentioned, which are defined as those the story is about. The tool, a “transcript analyzer,” is a deterministic script sequence in Python that makes use of open-source NLP function libraries and geolocator packages.

Accuracy and Results

Below are the accuracy results of our rule-based implementation through three stages of prototyping. We arrived at a final version that saw an increase in accuracy while using the most robust testing set.

It’s important to note that the prototypes were not all tested on the same dataset, as we iteratively adjusted the direction we wanted for our tool and, thus, sought new types of podcast and transcript structures.

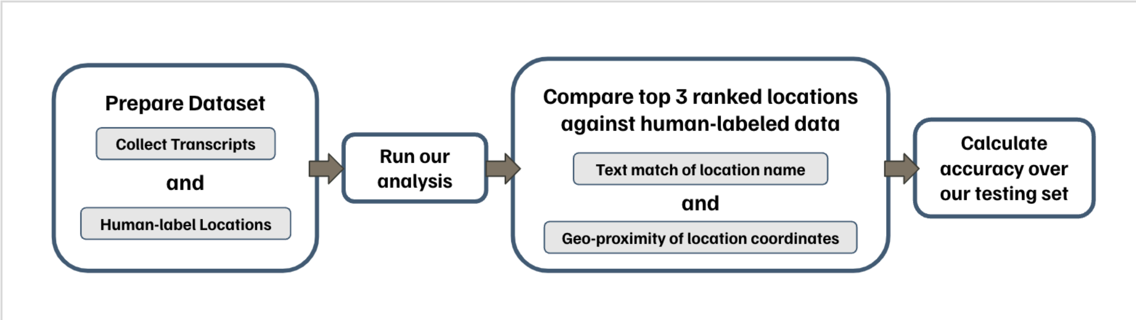
Version	Testing Count	Accuracy	Details
v1	9 episodes	100%	Only *BayCurious* episodes Limit locations to within the Bay Area
v2	22 episodes	64%	Expand sources to 10 podcasts No geographic limits on allowed locations
v3	28 episodes	82%	Decrease source to 8 No geographic limits on allowed locations

Evaluation Process

All our implementations were evaluated against human-labeled data for our test set. This data consisted of location names and coordinates determined by our team for the top location(s) in every episode in the set. To account for the limitations of the rule-based implementation (e.g., static & open-source geographic name database), we considered the top three locations for each episode for evaluation.


```
"episode_file": "transcript-082924.txt",
"genre": "History and Culture",
"source": "Bay Curious",
"data": [
    {
        "title": "Port Costa: Quirky, Historic, Cool",
        "location": ["Port Costa", "The Warehouse Cafe",
                        "The Burlington Hotel"],
        "coordinates": {
            "latitude": [38.0463, 38.0464, 38.0466],
            "longitude": [-122.1833, -122.1835, -122.1837]
        }
    }
```


Sample ground truth label



Representation of the evaluation process for our implementations

There were two ways in which an implementation output could get a match with the human-labeled "true" location:

1. If there was a fuzzy-match between the location names, using `fuzzy.partial_ratio` and ratio threshold=80%.
2. If the output location was within 2km of the truth location, using Haversine distance.

Total accuracy was calculated by the number of episodes that were matched by relevant locations divided by the number of episodes tested.

Datasets

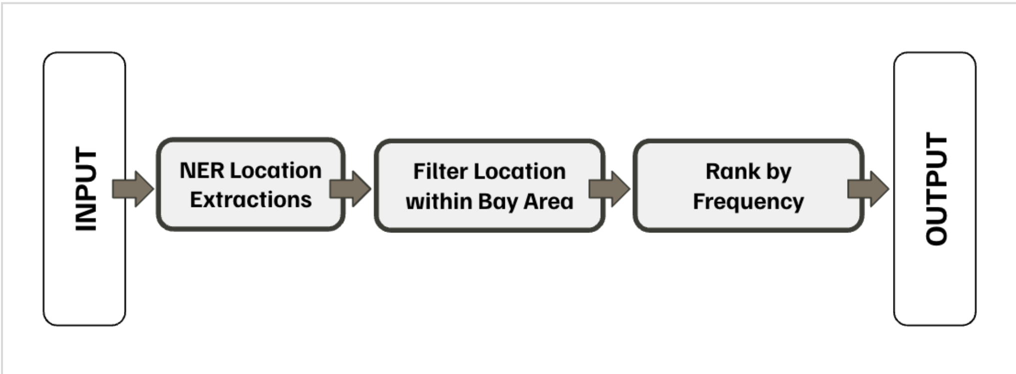
Below are details on the three datasets used during our prototype process.

Dataset	Episodes	Podcast Source
#1	9 episodes	All from *BayCurious*
#2	22 episodes	*99% Invisible, BayCurious, Bowery Boys, LA-ist, My Alaska Summer, On Location, Overheard, Stephenville, Reclaiming What's Been Lost, The Daily*
#3	28 episodes	*99% Invisible, BayCurious, Bowery Boys, LA-ist, My Alaska Summer, On Location, Stephenville, Reclaiming What's Been Lost*

Prototype in Progress: Phase 1

Overview

The processing pipeline for the first prototype of this tool is below.



The text extraction involved Named Entity Recognition (NER), an NLP process provided by the open-source spaCy module. We specifically use spaCy's "English Core Web Transformer" model.

At this prototype stage, we decided to evaluate five different open-source geolocator Python packages to identify which best handled the scale of locations we would be using (i.e., some might be good for cities or landmarks, but not neighborhoods or businesses). We decided to move forward with only the `geonames` API/package as we saw it was the fastest and tied for highest accuracy. The other packages were `geopy.geocoder`, `googlemaps`, `pgeocode`, and `geocoder`.

Feedback

The teaching team commended our technical execution in integrating geolocation filters, comparing methods to ensure accuracy, and our effort to engage with Bay Curious's existing audience.

Working with the teaching team to define and narrow our project goals, we were encouraged to focus on relevance and scalability. Recommendations included considering a deeper analysis of what makes a location "significant" beyond mention frequency and adapting our prototype to handle ambiguous, colloquial, and implied location mentions. Thinking beyond the Bay Area and podcasts in general to consider location-centered stories that took place in other areas and through different mediums was also proposed.

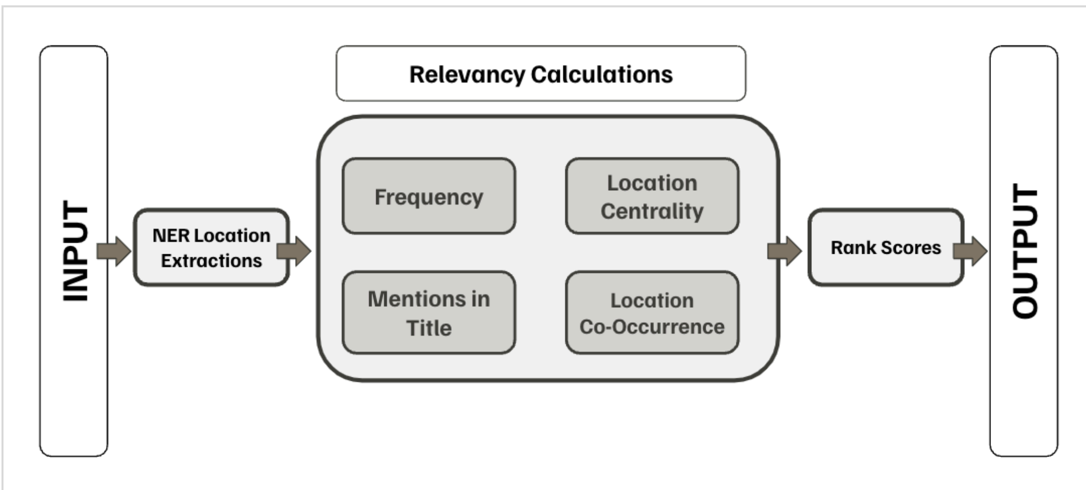
Specifically, we understood that the podcasts in our test cases included many locations in each episode, with varying levels of importance to the story and frequency of mention. We set out to investigate how we can measure a subjective term like relevance in order to consistently determine the most relevant location computationally.

With this expansion in scope, we also moved away from geographic filters within a "valid" region, since we no longer were going to work with only Bay Area locations.

Prototype in Progress: Phase 2

Overview

The processing pipeline for the second prototype of this tool is below.



Driven to develop a definition of "relevance" that applied to sources beyond just *BayCurious,* we studied other podcasts that provide "location-centered" stories. This allowed us to externalize the human mind's evaluation of "what location is relevant here" into a programmable process.

Ultimately, we arrived at a mathematical formulation of relevance. This formulation allowed us to include multiple factors identified in the relevance-determination process.

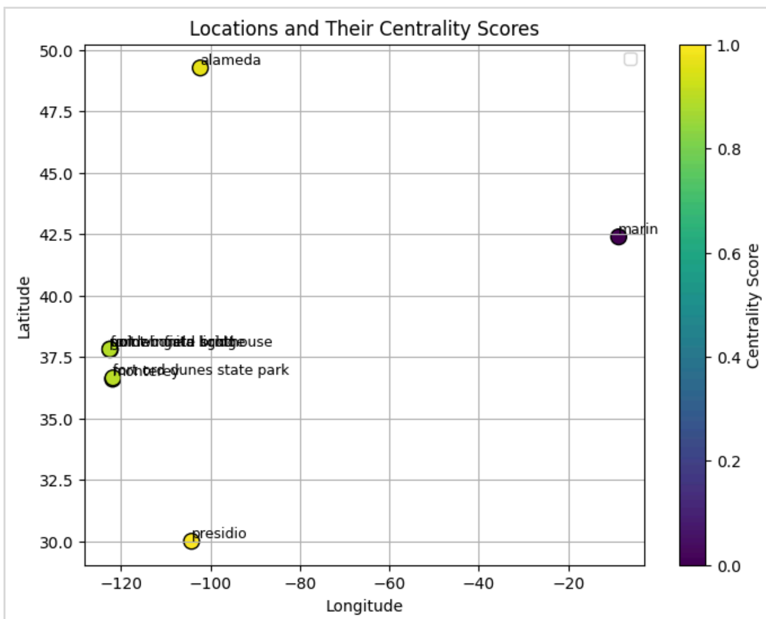
Explanation of Relevance Equation

The equation is made up of four factors. All are positive indicators of a location being relevant, but they do not necessarily fit with one another in a linear sequence. Therefore, we decided to calculate all four factors and combine their results into a composite "relevancy score," normalized from 0 to 1.

The first component is *frequency of mentions*. This is a score from 0 to 1 on how many times a location appears in the transcript proportionate to the others. The location mentioned most gets a score of 1.

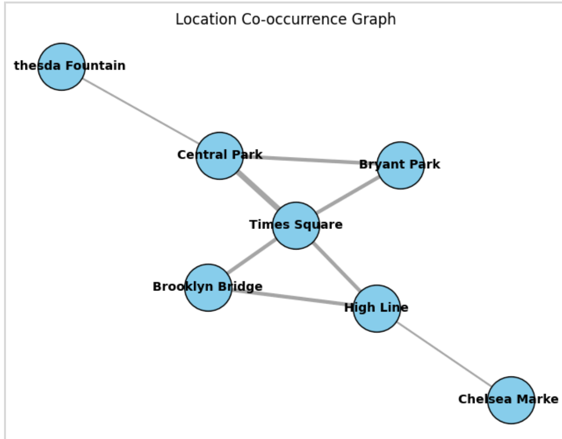
The second component is *mentions in the title*. This is a binary score, 0 or 1, of whether the location is mentioned in the title. A location mentioned in the title is likely relevant, but a relevant location is not always mentioned in the title.

The third component is *centrality to location clusters*. This is a score from 0 to 1 on how central a location is to the geographic cluster of all locations in the episode. A score close to 0 indicates the location is a strong outlier geographically (far away/removed) among all the locations mentioned. In our analysis, we found that transcripts usually have a central location and nearby supplemental locations. But far-outliers are usually in tangents or background information. This score allows us to lower the ranking of the outlier locations.



Sample cluster graph showing various locations plotted

The fourth component is: *pagerank from co-occurrence graph*. This is a score from 0 to 1 of the "importance" of a location based on (the existence and weight of) connections to other locations, using sentence co-occurrence to establish connections. A score of 1 indicates the location is the most “important” by being the one mentioned most in the same sentence as other locations. With that said, we are not as interested in the actual central node, because that is too black and white; instead we want to rank all nodes 0 to 1 by how well-connected they are. If we have a transcript with relevant location X and another location Y, if Y is rarely mentioned in the vicinity (same sentence) of X, then it's less likely to be related to the relevant location and should be scored accordingly. This approach allows us to rank locations relative to each other without knowing what X is yet.



Sample co-occurrence graph connecting some locations

Additionally, we decided to apply weights to each subscore so as to prioritize certain parts. Our evaluation tried different weights and arrived at the following the weight choice for maximum accuracy:

- 30% weight for frequency
- 40% weight for title
- 15% weight for centrality
- 15% weight for connectedness

Finally, we add all weighted subscores into a single "relevancy score" for each location.

Feedback

Our team's efforts in extending an applicable definition of relevance beyond frequency to include geographical and linguistic clustering solved earlier challenges and helped scale our prototype beyond *Bay Curious* podcasts. Surveying Bay Curious listeners on how they engage with podcasts and what types of podcasts they prefer helped us define user needs and prototype applicability.

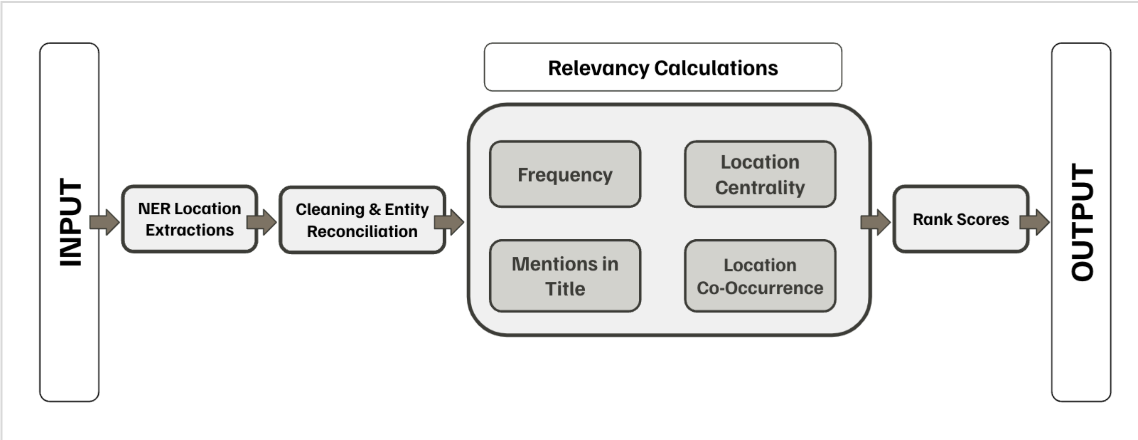
Turning toward a weighted system with subscores for different aspects of relevance increased our protoype's flexibility, but decreased accuracy. We were encouraged to refine our text processing pipeline to address specific fail cases (episodes that we failed) that we identified from testing. .

Our LLM-based approach was intriguing to the teaching staff, and we were asked to compare the outputs from both methods in order to see if there was a way to combine them.

Prototype in Progress: Phase 3

Overview

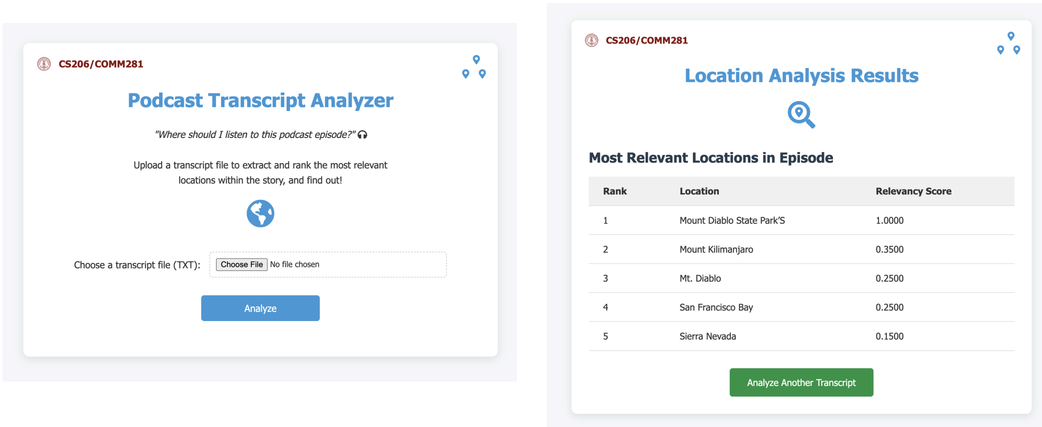
The processing pipeline for the third and final prototype of this tool is below.



To improve the accuracy of the tool in this prototype, we made two key changes to the code involving fuzzy text matching.

1. Entity Reconciliation – We noticed that some relevant location entities were not ending up at the top of our ranking despite being frequently found in the text. Upon inspection, we realized they had different formats across mentions. For example, the Mission District (in San Francisco) was mentioned and detected as “the mission district,” “the mission,” and “san francisco’s the mission district.” To tackle this issue we used fuzzy text matching with `token_set_ratio` to perform *entity reconciliation,* a process of consolidating items that are spelled differently but refer to the same entity.
2. Evaluating Against Truth Data – Similarly, we found that in some cases the correct locations were identified but generated false negatives because the text was not a character-by-character match with our truth labels. To adjust for this we used fuzzy text matching with `partial_ratio`.

Finally, we developed a local interface for the Python implementation to be run outside of the code terminal.



Landing page (input) and ranking (output) screens from the Python tool

LLM Prompt Testing

While our rule-based Python implementation approximated the human-evaluation of relevance to extract and rank locations from podcast transcripts, we recognized that the nuances of human language require understanding the context in which location terms are mentioned. To advance our goal of "connecting podcast content to physical locations," we evaluated whether large language models (LLMs) could complement our approach by capturing implicit location references that rule-based systems often miss.

While prototyping our Python script tool, we identified several instances where “the restaurant” or “the business” were not detected as location mentions, even though they were references to locations explicitly mentioned by name earlier in the transcript. The fact many implicit references weren’t counted towards relevance scores meant locations could be significantly under ranked.

Therefore, we tested relevant-location extraction using ChatGPT-4o, a variant of OpenAI’s GPT-4 large language model. Compared to models like Claude Opus or Gemini Ultra, GPT-4o offers a strong balance of cost and accuracy. While no LLM is fully deterministic, GPT-4o exhibited more consistency in extracting locations compared to older GPT models and open-source alternatives. With robust prompting, the model should extract the correct relevant location.

Approach: Few-Shot Learning

Despite GPT-4o’s strengths, running the API comes at a higher cost, is sometimes slower in performance compared to methods such as spaCy, and these models have been reported to hallucinate. After weighting these limitations, we pursued a practice called *few-shot learning* to give our LLM model the best shot at accurate results.

Instead of relying on the model to figure things out from scratch, we provided structured examples, starting with simple cases like a short transcript extract with one location mentioned before moving onto more complex transcripts with multiple references for the LLM to choose from but only one correct primary location. Few-shot learning helps the model learn patterns, improving its ability to extract locations based on our definition of relevance. We iterated on multiple versions of the prompt to see which yielded the best results. More advanced techniques included in the LLM prompt:

- Chain-of-thought prompting: encourage model to break down thought process
- Task decomposition into sequential subtasks
- Mandatory cross-checking of suggested locations against transcript text
- Business name prioritization over general locations
- Low-temperature setting (0) to reduce creativity of LLM and sticking to straightforward extraction

Using dataset #3 from our Python implementation, we created a simple prompt using OpenAI's API.

1. First, each podcast transcript was processed five times.
2. LLM-extracted primary locations were compiled and compared against ground truth.
    a. If at least one found location matches a correct location, count as a correct match.
3. **Accuracy = (Correct Podcasts / Total Podcasts) x 100**

Results and Analysis

- Average accuracy: 95.2%
- Standard deviation: 4.38% with a confidence interval of ± ~4%

Key Limitation: Across multiple runs, we saw that the model would sometimes extract different locations in each run, which poses a challenge for journalistic applications where precision and consistency may be key.

Takeaway: While the LLM results were promising, one of the biggest issues we encountered was running the same transcript multiple times didn't always yield the same primary locations. This lack of consistency may cause uncertainty about the trustworthiness of results. With more time, it would have been interesting to calculate our precision and recall scores. However, these results are enough to indicate that a hybrid model spells better outcomes for our goal of extracting relevant locations.

Insights from Our Bay Curious Listener Survey

To better understand how podcast listeners engage with location-based storytelling, we conducted a survey early in our prototyping process targeting listeners of KQED's *Bay Curious* podcast. We chose this as our survey audience because our initial prototype was built using many of its episodes for proof of concept. While we recognize that *Bay Curious* listeners are not representative of all podcast audiences, their responses provided valuable insights into the preferences and needs of listeners who engage with place-based storytelling.

Our survey was distributed through the *Bay Curious* monthly newsletter. We received a total of 68 responses, which provided insights into listening habits, interest in discovering locations from podcasts, and openness to using a location-based storytelling service.

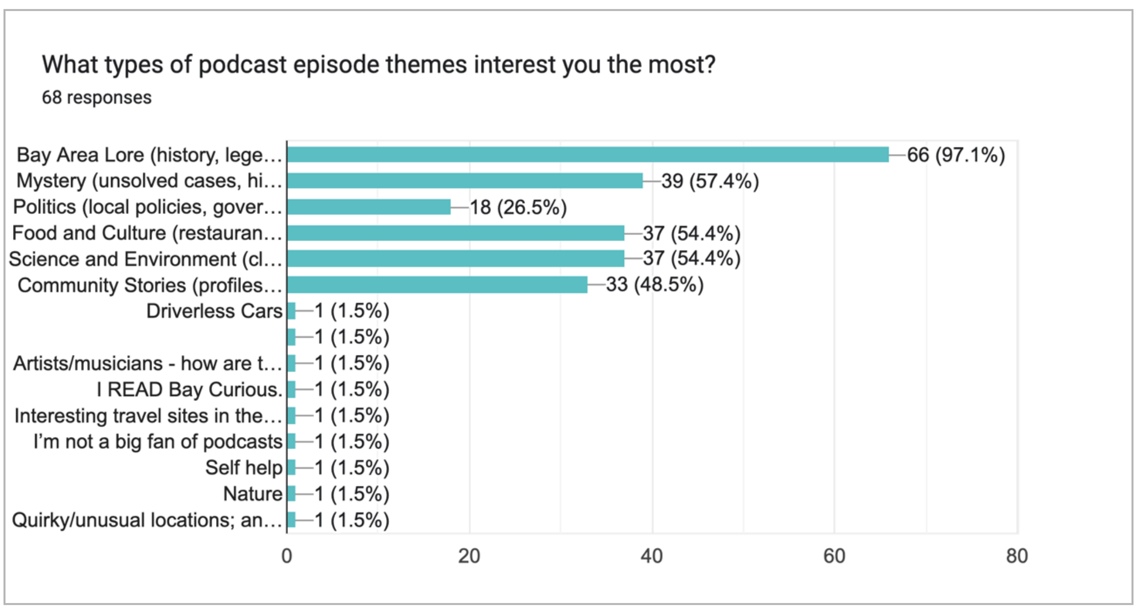
Listener Engagement with Bay Curious

We found that 77.9% of respondents listen to Bay Curious at least once per month, with 42.6% listening monthly and 35.3% tuning in weekly or daily. This demonstrates that the podcast has an engaged audience that regularly seeks out its content.

What Listeners Want from Podcast Storytelling

A critical part of our research goals was to identify which types of stories are most compelling to listeners among different types of location-based storytelling. Our survey results revealed a strong preference for historical narratives, local mysteries, and cultural exploration.

Nearly all respondents (97.1%) expressed an interest in "Bay Area lore," including history, legends, and cultural facts. Mystery-related content, such as unsolved cases and hidden gems, was also highly popular, with 57.4% of respondents selecting it as a preferred theme. Additionally, 54.4% expressed strong interest in stories related to food and culture, reinforcing the idea that listeners enjoy content tied to specific places, traditions, and community life. Overall, these metrics support the idea that our work can be particularly impactful when applied to storytelling that is deeply tied to place—whether through historical events, little-known locations, or culinary landmarks.

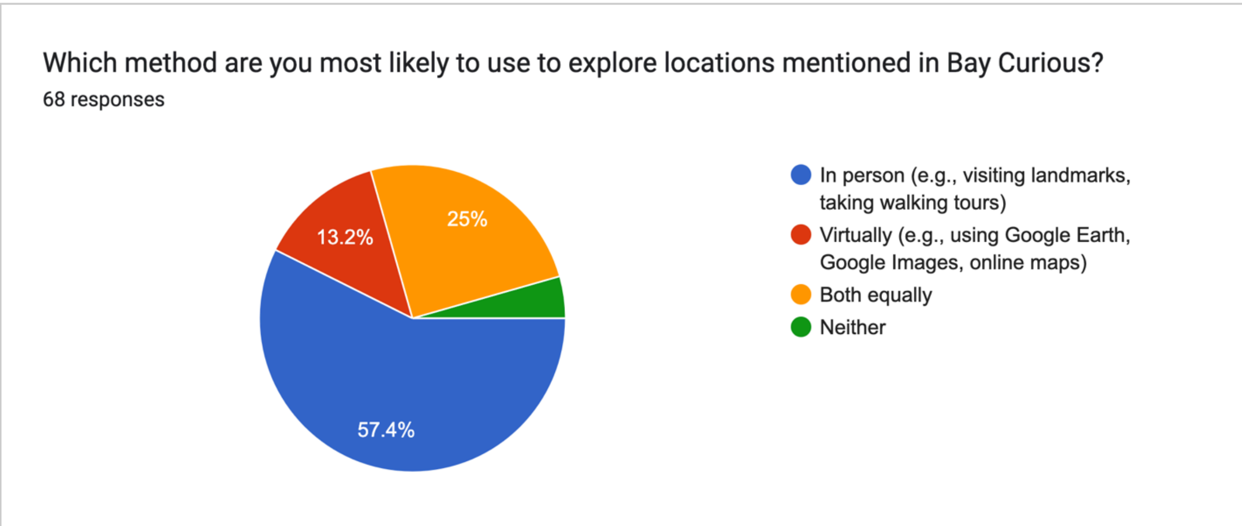


Interest in Exploring Locations from Bay Curious

We found that discovery of new places often happens organically and socially. Through our survey, we found that 79.4% of respondents rely on word of mouth when learning about new locations. This was followed by 54.4% who discover places by walking around and exploring, and 48.5% who rely on podcasts or local radio shows.

We also assessed listener interest in exploring locations mentioned in the *Bay Curious* podcast. Our results show that 57.4% of respondents would like to explore locations both in person and virtually.

An overwhelming percentage of respondents (82%) were interested in exploring locations in person or in person and virtually equally, which suggests that location-based storytelling should be flexible, offering both real-world and digital engagement options to accommodate different listener preferences.



Potential for a Location-Based Storytelling Tool

Our group also sought to determine the likelihood of listeners adopting a location-based storytelling tool that maps podcast stories to real-world locations. 54.4% of respondents rated their likelihood of downloading such an app at 7 or higher on a 1-10 scale.

From asking respondents what features would make them more likely to use a location-based storytelling tool, we found the most requested were interactive maps that allow users to see stories based on their current location. Other popularly requested features were personalized recommendations and notifications when they are near a relevant location. This confirms that users want a seamless, location-aware experience, where stories feel immediate, relevant, and personally curated.

Additional Insights from Open-Ended Responses

In the open-ended final comment section of our survey, many respondents compared the concept to existing audio experiences like Alcatraz Audio Tours, Apple Fitness' "Walk With…" feature, and Atlas Obscura.

Other themes that emerged from the comment sections were:

- Many listeners want a more immersive experience, such as an audio scavenger hunt or a tool that lets them track their exploration.

- Some respondents suggested partnering with established travel and history platforms like Atlas Obscura.
- A few respondents expressed concern about adding another app to their device, reinforcing the need for a lightweight, easy-access format like a web tool.
- Several emphasized that they would love to see diverse, inclusive storytelling that represents all cultures within the Bay Area.

Ultimately, these listener responses confirm that our project, if carried out further, should consist of more than extracting and mapping locations, and should consider how people want to engage with these stories in a meaningful and interactive way.

Conclusions & Next Steps

Challenges

Our project relied on a concrete understanding of what it means for a location to be relevant. During our prototyping process, we deconstructed how we determine relevance and translated these insights into a structured, programmable framework.

As we expanded beyond *Bay Curious* episodes, maintaining accuracy across diverse podcasts became challenging, highlighting the complexity of defining relevance. Realizing we may have scaled too quickly, we explored an LLM approach which addresses some limitations of the rule-based tool (e.g., implicit references). This new approach, though, raised its own issues with consistency.

Path Ahead

Our next step is refining a hybrid model that blends the adaptability of LLMs with the precision of Python scripting. GPT-based models struggle with word-for-word accuracy, sometimes misinterpreting or hallucinating locations, making them less reliable for precise data retrieval. Meanwhile, the Python script ensures accuracy but lacks flexibility and scalability. By combining both approaches, we can create a system that balances adaptability with reliability. One version of such a system might use Named Entity Recognition (NER) in Python to extract locations, and then apply structured few-shot LLM prompting for prominence and contextual relevance.

Based on our survey insights, we envision future developments that align with audience needs. To enhance immersion, we could embed or link to a podcast player that triggers episodes based on location. Mapping locations based on events or integrating with social media—like a Snapchat-style map or a TikTok heatmap showing popularity—could also make location-based storytelling more interactive and easily discoverable.

Conclusion

Our project connected podcast storytelling with geographic exploration by developing a system to identify and rank key locations in transcripts. Using rule-based extraction in Python, LLM prompting, and structured ranking, we balanced accuracy with adaptability. Iteration and user insights shaped a more flexible model. A hybrid Python/LLM system will further enhance scalability and reliability, highlighting the potential of location-based storytelling to transform listener engagement.