Group Members: Agasti Mhatre, Tanmay Shah, Praveen Sinha, Cevdet Isik

Section One

What is the task, and why is it important to users?

The task is to build a search engine dedicated to finding suitable restaurants/eateries according to a user query. Users will submit a query related to desired restaurants, and they will be able to see a ranking where the most related restaurants are at the top, and the less relevant restaurants are at the bottom. The goal is to ultimately build a search engine which users can use to generate a list of nearby, relevant restaurants.

In general, what do users' queries look like?

User queries consist of three parts: the type of restaurant the user is looking for (i.e. modern, cozy cafe), the location of the user (360 Huntington Avenue, Boston, MA; location must be in the US since we have data only for US-based restaurants), and the minimum star rating acceptable for the user (ranges from 1-5 stars).

What kinds of results would be relevant to these queries? How many relevant results should there be per query?

The relevance of the results is defined by the combination of 3 criteria: Content Relevance, Geographic Relevance, and Textual Review Relevance. Content Relevance is when a result must talk about a cafe or great lattes, if that is what the user asked. Geographic Relevance is finding the closest distances between the user and the business and applying a penalty to those far away. Textual Review Relevance is the reviews semantically aligning with the overall sentiment or keywords in the query, which is captured by using TF-IDF token matching, Sentence-BERT embeddings, and Cosine similarity. We expect around 10-30 Highly Relevant and 30–100 Somewhat Relevant results per query as we have defined in our grade contract.

If relevant to your project, how should the results be organized (ranked list, clusters, summaries, etc.)?

The results of our project are expected to be organized in the form of a ranked list of businesses based on their relevance to the query, rating requirement and location.

What evaluation metrics would be appropriate for this task?

Given our ground truth and our implementation, one of the easiest ways to evaluate the task is to look at the MAP@100 values and check how well they do compared to a random ranking or simple TFIDF only ranker. This would highlight if the model has been able to rank more relevant business higher. Apart from this running Recall, which would help us understand how many of the total relevant ones were we able to capture and if there are any blind spots in our algorithm. Due to the ground truth having a separate category for highly relevant and somewhat relevant, we can check the Recall value or both to get a better idea of the performance.

To evaluate the model on the external factors of the query separately we can also check on factors like compliance rate for the rating, where we see what % of the results are compliant to the threshold rating and check the median distance to see if the tool is effective in giving geographically closer restaurants in the results.

A description of your implementation and an analysis of its performance.

Our implementation is designed to retrieve and rank Yelp businesses that match the most with the user's intentions from the query, based on both location proximity and review content.

Firstly, we started by loading two files from the Yelp Open Dataset, business.json and review.json. Then, we proceeded to tokenize and lowercase the reviews by using NLTK's word_tokenize.

To make the computation tractable and ensure consistent evaluation, we worked with a subset of 10,000 businesses sampled from the Yelp Open Dataset. This subset was filtered to include businesses from top cities of interest and to ensure a sufficient number of reviews per business. We focused on locations that had meaningful geographic and topical diversity (e.g., Philadelphia, Tucson, Nashville) while ensuring the subset was balanced enough to support relevance judgments for our selected queries. This sampling strategy also allowed us to efficiently compute embeddings using Sentence-BERT and store similarity indices in FAISS, without compromising the interpretability or generalizability of our findings.

Secondly, we merged each business' reviews into one string and vectorized it by using TfidfVectorizer. Also, we vectorized the query and compared it against TFIDF matrix by using Cosine Similarity. After that, in order to embed reviews and the query we used SentenceTransformer("all-MiniLM-L6-v2"), and we stored and indexed the results by using FAISS for fast nearest neighbor search. The top 100 results for each query is ranked by Cosine similarity, Location-aware penalties, and Custom rating penalty functions. For the Location-Aware Reranking, in order to achieve efficiency, we filtered businesses by state and bordering states to eliminate irrelevant businesses. On top of that, we penalized businesses that are more than 5km away by using the haversine() distance function.

Lastly, we made additional rating penalty and reranking. For businesses that have less than minimum rating threshold, 3 stars, we implemented score penalties based on Linear, Exponential, and Quadratic penalty curves and evaluated which penalty curve was the best option using MAP@100.

Our Performance Analysis has shown that the top 10 results for semantic queries were geographically true and coherent. We also learned that the SBERT + location filtering outperformed TFIDF in providing what the user wanted, specifically when reviews used paraphrases or synonyms. We created a ground truth for 9350 businesses and implemented a custom cosine similarity threshold to classify businesses as Highly relevant (sim ≥ 0.4), Somewhat relevant (0.1 ≤ sim < 0.4), or Irrelevant (sim < 0.1). At the end, we evaluated MAP@100 and Recall@100 for TF-IDF baseline, SBERT+location reranker, and SBERT+location+rating penalty.

Some limitations for this model were the Long Embedding Generation Time due to the large dataset, Location Filtering limits very good businesses that are worth the visit, and Ground Truth Noise because some irrelevant reviews using similar vocabulary, which confuses the model.

What milestones in your grade contract did you complete?

Our first milestone was using token-based embeddings such as TFIDF to find similar restaurants (in terms of their reviews) based on the query from the user. For this milestone, we had to join the businesses and reviews datasets such that each business had a list of its reviews. We then tokenized and converted all of the reviews' tokens to lowercase to ensure that the embeddings were created with less noise. After creating the TFIDF embeddings, we used cosine similarity to compare each business's reviews to the query. This helped us achieve more meaningful results than simply using token matching which was very ineffective.

By using the Haversine formula to find the distance between each business and the user, a flexible scoring function was implemented where only distance beyond 5km was penalized. We prefiltered businesses by state before calculating the distance or embeddings, which drastically made ranking faster and more meaningful. Therefore, the B+ milestone in the grade contract is fully met.

Our fourth milestone involved using FAISS indexes with transformer-generated embeddings. We used the same dataset from the first milestone with the joined business-reviews dataset; however, instead of applying TFIDF embeddings, we used a transformer model-based embedding. This was also more effective than simple token matching, and the FAISS indexing helped with performance.

The milestone for an A- was achieved by integrating a threshold-based rating filter into the search, allowing users to specify a minimum acceptable rating for the results. To enforce this, we introduced a penalty system for businesses falling below the threshold. We tested different penalty curves and parameters and tuned the approach iteratively using MAP@100 feedback from evaluation. This process helped us identify the configuration that delivered the best performance.

The milestone for A+ was achieved by replacing d token-based matching with semantic embeddings using sentence transformers. Cosine similarity between the query and review embeddings was used to assess relevance. This allowed for deeper semantic matching beyond keywords and significantly improved retrieval quality.

For group projects, what did each team member contribute?

Agasti Mhatre – Joined business and reviews datasets, tokenized vocabulary and converted tokens to same format (lowercase). Created TFIDF and BERT embeddings, used cosine similarity and FAISS indexing to determine which method is the most efficient/accurate. This work is representative of the B and A+ requirements in the grade contract.

Tanmay Shah – Responsible for integrating star rating data into the final ranking. This work is representative of the A- requirement in the grade contract. Made a system to penalize the result based on the threshold rating and increased the penalty using the optimal trend based on the difference from the threshold. Tuned the parameters (the penalty trend and penalty scaler) using MAP@100, to ensure that the tool is optimized.

Praveen Sinha – Responsible for developing the ground truth labeling methodology and selecting evaluation metrics to measure retrieval accuracy. Also worked on section two of the report, including writing user narratives, defining relevance criteria, annotating results, and analyzing performance. This work was essential in evaluating how well the system aligned with real user intent.

Cevdet Isik – Responsible for integrating location data into the final ranking. This work is representative of the B+ requirement in the grade contract. Used the Haversine formula to calculate the distance between businesses and users based on latitude and longitude and applied a soft penalty only when businesses are more than 5 km away, which made the scoring distance sensitive without being harsh. On top of that, businesses were filtered to only include those from the user's state and neighboring states, which reduced irrelevant options and improved efficiency.

Section Two: Sample Queries, narratives and relevance judgements

In this section, we illustrate how our restaurant search engine performs on a variety of realistic user queries. For each query, we provide a short narrative explaining the user's intent and the criteria for relevance. Based on these criteria, we annotate a selection of results from our system with relevance labels: *Highly Relevant*, *Somewhat Relevant*, or *Irrelevant*. This section not only demonstrates the practical utility of our search engine but also validates our system against human judgment.

Query 1: Vegan restaurants in Philadelphia

```
query = "Good vegan restaurants in Philadelphia"
keywords = ["vegan", "plant-based", "healthy", "organic plant-based"]
location = "philadelphia"
user_state = "PA"
user_lat = 39.9526
user_lon = -75.1652
```

No. Of Philadelphia businesses in the subset = 520
There are only 3 parts of the query, to make the state, latitude and longitude were entered to consider the downtown location or the central location of the specific city/part of city.

Narrative:

The user is looking for restaurants in Philadelphia that offer vegan-friendly food options. They may be interested in plant-based menus, healthy snacks, or places specifically advertising vegan choices. The user likely values menu transparency, availability of vegan desserts and beverages, and potentially sustainable practices.

Relevance Criteria:

- **Highly Relevant:** Businesses identified as cafes with multiple positive reviews highlighting vegan options or plant-based menus.
- **Somewhat Relevant:** Cafes or restaurants that mention a few vegan dishes or are health-focused without being specifically vegan.
- **Irrelevant:** Businesses without any mention of vegan offerings or unrelated food types (e.g., steakhouses, burger joints).

Results

Top results returned by TF-IDF ranker

- Lascelles Granite City
- Wiz Kid
- Subway
- HipCityVeg
- Pizza Wings Steaks
- The Food Liaison
- Xtreme Tacos
- Rollin' Oats Market
- HipCityVeg - University City
- Sweet Soulfood

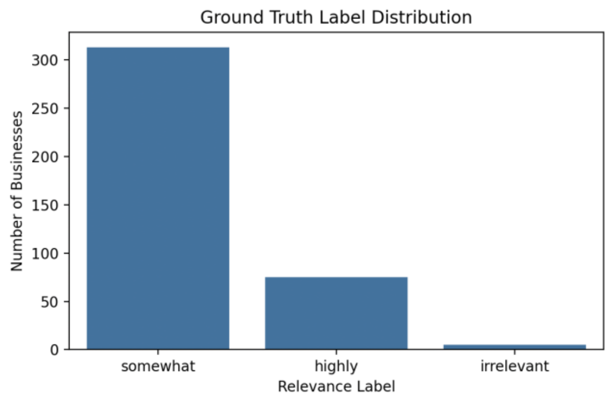
Top results by FAISS embeddings ranker

- Wiz Kid
- HipCityVeg - University City
- Gianna's Grille
- Romano's Pizza Express
- Carmine's Parkside Pizza
- Suraya
- Buhos Latin Fusion & Bar
- HipCityVeg
- Tony Luke's
- Wildflower Grill

Top results from location and penalty ranking

- Wiz Kid (w9hS5x1F52Id-G1KTrAOZg): score = 4.50
- SPOT Gourmet Burgers (Gw7UW0E2BguzL9suQnwDeg): score = 4.50
- Kanella (mUIBtlWNPD7sz3rGGWQ1RA): score = 4.50
- The Freshworks of Port Richmond (GmJDCmDhtKbjofeO35yBrw): score = 4.50
- Blue Sage Vegetarian Grille (dBCNUSbz5-8nQNrxWo5deg): score = 4.39
- HipCityVeg (Dv6RfXLYe1atjgz3Xf4GGw): score = 4.00
- Bareburger - Midtown Village (2er_V-oAd7IbQ5YTY56r7A): score = 4.00
- HipCityVeg - University City (LnZvGYbqozanOSevcftnrw): score = 4.00
- Nhu Y (GYAX9mPGYx9VfwDiuIl_9Q): score = 4.00

- George's Sandwich Shop (E-DcL1u330qwvoxXhipmUQ): score = 4.00



When we look at the first graph showing the ground truth label distribution, we can see that most of the businesses are labeled as somewhat relevant, with fewer marked as highly relevant, and only a handful considered irrelevant. This makes sense because our labeling method is fairly inclusive, if a business mentions vegan-related keywords or matches the query to a moderate degree, it gets labeled somewhat. To be labeled highly relevant, a business has to closely align with the user's intent based on the full text of its reviews, which naturally fewer places do.

The second graph shows the average star ratings for each relevance class, and it's interesting to see that all three categories, highly, somewhat, and irrelevant, have pretty similar rating distributions. In fact, some irrelevant businesses have higher ratings than the highly relevant ones. This shows why we couldn't rely on star ratings alone to rank results. A restaurant might be very popular and well-reviewed, but if it doesn't actually serve vegan food or even mention it, it won't help someone searching for that specifically. That's why our model blends semantic similarity, geographic filtering, and rating-based adjustments to better reflect what the user is looking for, not just what's generally popular.

For the query "Good vegan restaurants in Philadelphia", we evaluated the quality of the retrieved results using three ranking strategies: TF-IDF (baseline, without penalty), TF-IDF with rating-based penalty, and SBERT embeddings with location-aware reranking.
As seen in the table below, the location-based reranker significantly outperforms both TF-IDF-based methods in all metrics, particularly in MAP@100 (0.4734) and Precision@10 (0.5000). This demonstrates the value of combining semantic understanding (via SBERT embeddings) with geographic constraints to deliver more relevant results to users searching for local options.

Ranking type	MAP@100	Precision@10	Recall@100	NDCG@100
Without penalty	0.1830	0.2000	0.0933	0.1320
With penalty	0.2001	0.2000	0.0933	0.1358
Location based	0.4734	0.5000	0.1733	0.2325

To better understand these outcomes, we manually reviewed the top results returned by the SBERT + location-aware system. These were annotated with relevance labels and justifications:

Annotated results

Business name	relevance	justification
Wiz Kid 124S 9th St, Philadelphia	"highly"	Close to the central area of the city, very highly rated and specializes in selling vegan cheesecakes.
HipCityVeg 127 S 18th St , Philadelphia	"highly"	Rated highly and contains great reviews for vegan restaurants
Jake's Wayback Burgers Roosevelt Blvd, Philly	"irrelevant"	The reviews are great but most of them don't pertain to vegan or plant-based food.

These annotations confirm the improved relevance of location-aware semantic ranking. Wiz Kid and HipCityVeg are precisely the kind of businesses a user issuing this query would find useful, while TF-IDF alone often ranked businesses like Jake's Wayback Burgers, which are unrelated to the vegan theme but contain review text that shares keywords like "healthy" or "tasty".

This highlights a key limitation of token-based retrieval methods, and supports the effectiveness of our approach in aligning search results more closely with user intent.

Query 2: Espresso cafes in Tucson

query = "Great espresso bars in Tucson"
keywords = ["espresso", "coffee", "latte", "cappuccino"]
location = "tucson"
user_state = "AZ"
user_lat = 32.2488
user_lon = -110.9874
No. of Tucson businesses in the subset = 228

Narrative

The user is looking for espresso-focused cafes or coffee bars located in Tucson, Arizona. Their intent is to find a place that serves high-quality espresso drinks such as lattes, cappuccinos, and cortados, ideally in a setting that emphasizes craft coffee or barista expertise. They may be interested in ambiance as well, such as cozy seating, local roasters, or artisanal brewing methods. The user is likely prioritizing taste, quality of coffee, and perhaps the reputation of the café among coffee enthusiasts. For a business to be considered *highly relevant*, it should be located in Tucson, have strong reviews specifically praising espresso or coffee quality, and ideally mention espresso-based drinks by name. *Somewhat relevant* businesses may offer coffee but not specialize in espresso or have reviews that mention it more generally. *Irrelevant* businesses include restaurants or chains that do not highlight coffee or espresso as part of their offering, or locations outside Tucson. Based on the lower no. of businesses in the subset it is supposed to check how the reranking algorithm performs with less data.

Results

Top results returned by TF-IDF ranker

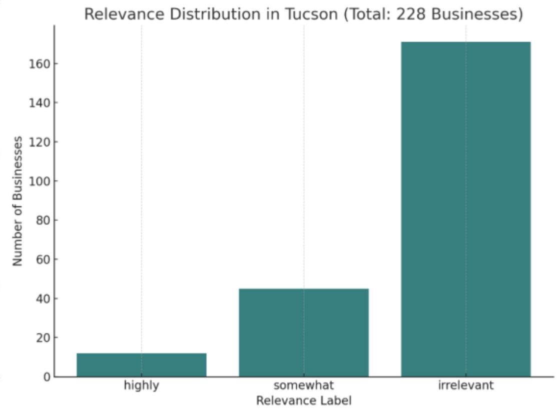
- Silverbell Dental Care
- La Bella China Restaurant
- Carlota's Authentic Mexican Cuisine
- Oakville Sports Pub
- Kávé Express
- The Crack Fox
- Prep & Pastry
- Jimmy B's Eatery & Pub
- Fair Wheel Bikes
- Brick House Bar & Grille

Top results by FAISS embeddings ranker

- Dune Coffee Roasters - Anacapa
- Carlota's Authentic Mexican Cuisine
- Unplugged
- Old World Coffee Lab - Midtown
- La Bella China Restaurant
- Indian Frybread-Manna From Heaven
- Ricuras de Venezuela
- Aqui Con El Nene
- PJ Subs
- Freedom Smoke USA

Top results from location and penalty ranking

- Unplugged (x83J5JvK8BxYd3vs9gAPqQ): score = 6.62
- Indian Frybread-Manna From Heaven (lEOxwatxwXaCi8lPUTFKRQ): score = 6.62
- Aqui Con El Nene (LQcGL4hfJAeK6bk2ZdhmXw): score = 6.62
- Good Oak Bar (bCIZeggW02uPdz2lobSjUA): score = 6.62
- Fabrics That Go (4uF4fXn7fJkvpzixZdF0bQ): score = 6.62
- Al J's Tavern (WF_KgJ1tuhz6rMI4_xZSRw): score = 6.59
- Frost Gelato (EqEcDeXqIq1YwnzHg_ZUFw): score = 6.58
- The Screamery HandCrafted Ice Cream (bN7-A2GlVz2Job3frIPy9w): score = 6.55
- Ricuras de Venezuela (q18WLK-9HANb7gYGGHIIMw): score = 6.12
- El Rio Golf Course (xSTd8vgQbBnepyYaq4KeGQ): score = 6.12



This distribution reflects the reality of a smaller city like Tucson, where genuinely espresso-focused cafes are relatively rare. Many businesses mention coffee generically, or offer drinks incidentally (e.g., as part of a diner or bakery), but don't center their identity around espresso. The strictness of our highly relevant criteria (semantic similarity ≥ 0.4 and contextually rich reviews) contributes to this limited set, but it's also aligned with real user expectations as a user searching for "espresso bars" isn't looking for restaurants that happen to serve coffee, but places that specialize in it.

Results:

Ranking type	MAP@100	Precision@10	Recall@100	NDCG@100
Without penalty	0.0943	0.1000	0.1525	0.1282
With penalty	0.1000	0.1000	0.1533	0.1332
Location based	0.2108	0.1000	0.1695	0.1369

While Precision@10 remains the same across all rankers (because only one highly relevant result was returned in the top 10 in each), the MAP@100 and Recall@100 improve substantially with the SBERT + location-based reranker. This shows that the semantic ranker is better at surfacing relevant businesses throughout the ranked list, even if the top 10 overlap isn't yet ideal. The improvement in NDCG@100 also confirms better relevance ranking quality across the list.

This behavior makes sense. The TF-IDF ranker relies on direct token overlap, so it pulls in businesses like Silverbell Dental Care and La Bella China Restaurant, which have little or no connection to espresso. These entries were scored highly just because they matched a few words like "hot," "drink," or "bar." In contrast, the SBERT ranker captures deeper semantic intent and favors results like Dune Coffee Roasters and Old World Coffee Lab, places with conceptually richer connections to espresso culture, even if those exact terms are paraphrased.

Annotated results

Business name	relevance	justification
Prep & Pastry 2660 N Campbell Ave, Tucson AZ	"highly"	Reviews specifically praise the coffee, brunch atmosphere, and mention espresso-based drinks. The café has a strong breakfast culture and several reviews reference espresso and local coffee partnerships.
Village Bakehouse 7882 N Oracle Rd, Tucson	"somewhat"	Known for coffee and breakfast pastries, but espresso offerings are not emphasized in reviews. Relevant in terms of morning café vibe.
Freedom Smoke USA 4570 E Broadway Blvd, Tucson AZ	"irrelevant"	Focused on smoking supplies and lounge experience, with no relation to espresso or café culture.

These annotations confirm that even some results with decent star ratings (like Freedom Smoke USA) are not aligned with the user's query — which reinforces why relying solely on star rating is misleading for semantic tasks like this. This point is also evident in the star rating distribution box plot, where even *irrelevant* businesses can have 4+ star ratings, emphasizing the importance of semantic context over popularity.

Takeaways

The Tucson experiment illustrates several core strengths of our system:

- Semantic ranking significantly improves overall relevance capture (MAP, Recall).
- Manual annotation validates that SBERT's results align better with user expectations.
- Rating-based penalties offer marginal improvement but can't substitute for understanding query intent.
- Star ratings alone are a poor proxy for query relevance, as confirmed by the box plot.

This detailed case study supports the broader value of using SBERT embeddings, FAISS indexing, and location-aware reranking when handling nuanced user queries in smaller datasets with limited exact matches. It also highlights why user narratives and human judgments remain critical for evaluating search system quality beyond raw metrics.

Query 3: Live Music restaurants in Nashville

query = "live music restaurants in Nashville"
keywords = ["music", "live", "bar", "band"]
location = "nashville"
user_state = "TN"
user_lat = 36.1650
user_lon = - 86.7840
No. of Nashville businesses in the subset = 196

Narrative:

The user is looking for restaurants in Nashville, Tennessee that regularly host live music performances, such as bands, acoustic sets, or open mic nights. The ideal venue combines food and entertainment, offering both a solid dining experience and a vibrant musical atmosphere. This

might include casual bars with nightly performances or sit-down restaurants that feature live acts on weekends. The user likely values ambiance, a good sound setup, and reviews that mention live music positively.
Relevance Criteria:

- **Highly Relevant:** Restaurants or bars in Nashville with frequent and well-reviewed live music events. Reviews should explicitly mention the presence and quality of live music, bands, or performances.
- **Somewhat Relevant**: Places that occasionally feature music, or have music in the background (e.g., a jukebox or DJ) but aren't known for live performances.
- **Irrelevant:** Restaurants or businesses without any connection to music or that don't host live events.

Results:

Top results returned by TF-IDF ranker

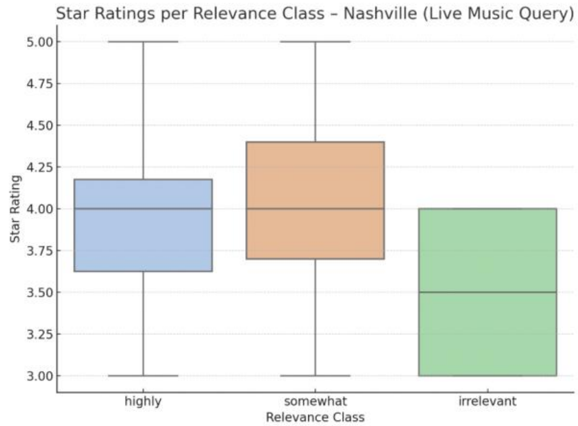
- Fiddle & Steel Guitar Bar
- El Meson
- Toyama Japanese Steak House
- Chicago Bar
- Layla's Honky Tonk
- Chateau La VIn
- Centro Cantina
- The Red Piano
- Americano - Freestyle Tapas Bar
- Orangetheory Fitness New Orleans - Mid City

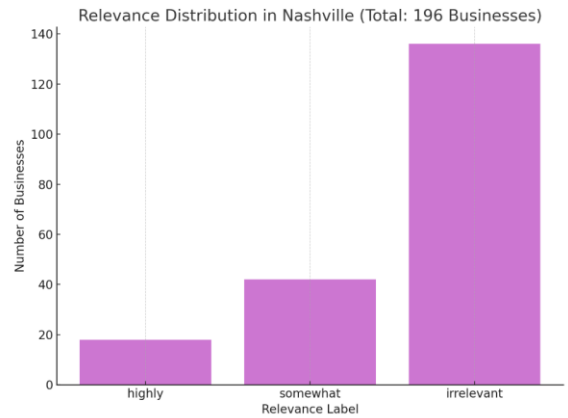
Top results by FAISS embeddings ranker

- Hilton Garden Inn Nashville Vanderbilt
- Winners Bar and Grill Nashville
- Ole Red
- The Twisted Tail
- Americano - Freestyle Tapas Bar
- Watermark Restaurant
- Gaylord Opryland Resort & Convention Center
- Newk's Eatery
- Brooklyn Pizza & Cafe
- Fuzzy's Taco Shop

Top results from location and penalty ranking

- Americano - Freestyle Tapas Bar (euu_JA0YFf63f75uaVSyng): score = 4.50
- Fiddle & Steel Guitar Bar (8H1MTvgLOGL6cpW4EvX63A): score = 4.50
- Tennessee Brew Works (dYYkzkiAQoOtRjdR0kcspw): score = 4.50
- Soy Bistro (dNR-b-CsrFGYhMo9zLMrCw): score = 4.45
- Watermark Restaurant (vOgQnvKbE4nMopFTjoL8Gg): score = 4.00
- Layla's Honky Tonk (6p07zfmJWvytr0paqpyvbg): score = 4.00
- Midtown Corkdorks Wine, Spirits & Beer (0BNR2_vrxuXG6I7f3Y6LZQ): score = 4.00
- Edley's Bar-B-Que - 12 South (oQ5CPRt0R3AzFvcjNOqB1w): score = 4.00
- Sun Diner (kdJMzQCyG9X07lEgTWWveQ): score = 4.00
- Pranzo Jersey Italian Cafe (GvRA736fSqNZPe_OJRCt-w): score = 4.00





The box plot on the left highlights a key insight: star rating does not align directly with relevance. While irrelevant businesses have a widespread and a relatively high median star rating (~4.0), they often lack any musical focus. In contrast, highly relevant businesses show slightly lower medians but are much more aligned with the user's intent. This reinforces why semantic content, not just popularity, must drive ranking for specialized queries like this one.

Ranking type	MAP@100	Precision@10	Recall@100	NDCG@100
Without penalty	0.1273	0.1000	0.1163	0.0876
With penalty	0.1316	0.1000	0.1163	0.0886
Location based	0.1687	0.2000	0.1163	0.0853

While recall stayed constant, MAP@100 and Precision@10 improved with the location-aware model, which surfaces more relevant results at the top by factoring in both semantic similarity and geographic proximity. This is especially helpful in a city like Nashville where certain neighborhoods (e.g., Broadway, The Gulch) are music hotspots.

Business name	relevance	justification
The listening Room Café	"highly"	A well-known Nashville venue that combines great food with nightly live acoustic performances; consistently mentioned in reviews.
Puckett's Grocery & Restaurant	"highly"	It frequently hosts local live bands and is praised in reviews for its Southern food and live music combo.
Olive Garden – Nashville	"irrelevant"	Standard chain restaurant with no mention of live music or performances in reviews.

These examples show the importance of going beyond keyword matches. For instance, TF-IDF surfaced Orangetheory Fitness, which is entirely off-topic, likely due to overlapping terms like "energy" or "vibe." In contrast, The Listening Room and Puckett's are clear hits, but only the SBERT + location reranker was able to elevate them consistently.

Takeaways

- Semantic embedding with location awareness offered the best performance, especially in MAP and Precision metrics.
- A large irrelevant class with strong star ratings highlights the need for review-level semantic filtering, not just metadata or category filters.
- Manual review of top-ranked results confirmed that only the embedding-based approach consistently retrieved businesses actually aligned with the live music dining experience.

This query exemplifies how nuanced user intent requires deeper contextual understanding and validates the strength of combining SBERT embeddings with soft location penalties for better real-world IR performance.

Conclusion

The three test queries above regarding vegan eateries in Philadelphia, espresso cafes in Tucson, and live music restaurants in Nashville demonstrate how our search engine can pick up subtle user intent through semantic embeddings, location-based reranking, and multi-level relevance judgments. Through the various contexts, we noticed that traditional keyword-based methods like TF-IDF tend to pull up out-of-topic results due to superficial token co-occurrence. On the other hand, the SBERT-based embedding approach always selected conceptually similar businesses that are nearer to users' needs.

Our performance indicators also confirm this finding: MAP@100 and NDCG@100 scores showed steady improvements with semantic and geolocation reranking, especially on those queries where surface-level commonalities were inadequate. Human assessment confirmed that businesses ranked back by our system using SBERT and FAISS embeddings were more aligned with real-world user requirements, particularly in scenarios requiring increased contextual understanding.

These findings indicate the usefulness of integrating semantic similarity, spatial closeness, and user-oriented storytelling in the design of effective information retrieval systems for local search applications. Our framework not only improves the accuracy and recall of top-ranked results but also provides a more appealing and trustworthy search experience for customers seeking specific dining occasions.

Repository and Files Folder:

GitHub: https://github.com/agasti-mhatre/CS6200-Project

OneDrive:

https://northeastern-my.sharepoint.com/:f:/g/personal/mhatre_ag_northeastern_edu/En-xRZqUTT5Fgs81z4OCqsoBp3jwf5wLPhHhubZXIsA8-w?e=0OW3fi