U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

MDSE MASTER IN DATA SCIENCE AND ENGINEERING

Restaurants beyond Philadelphia

Harnessing reviews, ratings and locations

"A multi-modal approach to restaurant recommendations around Philadelphia"

Group A:
Carolina Madaleno, up2024098098
Jorge Gamonal, up202502660
Ricardo Jorge Correia, up201202477
Vitor Souza, up202400084

Porto, 11th October 2025

Table of Contents

01 Introduction

Problem Statement

User's reviews across different restaurants, both texts and ratings, are crucial to address different key points:

- ❑ dining preferences
- ❑ locate rising food spots
- ❑ uncover hidden patterns in customer behaviour

We propose to develop a recommender system that captures users dining preferences from Philadelphia, PA, USA, and use these insights to recommend suitable restaurants in neighboring cities, where users may have little to no prior activity.

01 Introduction

Objectives

Main Objective:

To explore how ratings, reviews and locations can be combined to recommend restaurants in Philadelphia's surroundings.

Specific goals :

Phase I (RS + NLP): Evaluate how restaurant's category topic features obtained from user's reviews increase recommendation accuracy.

Phase II (SNA + Time-Series Forecasting)

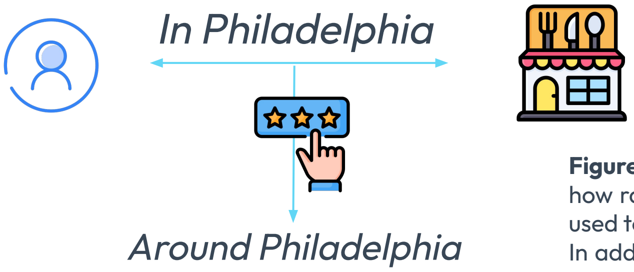


Figure 1: Schematic representation of the main goal for this project, namely of how ratings and reviews provided by users in restaurants in Philadelphia may be used to develop a recommender system for other restaurants in the surroundings. In addition, user preference communities and preferences forecasting over time were performed.

01 Introduction

Motivation & Significance

- **Local tourism:** Encourage the discovery of new restaurants around Philadelphia, improving visibility for smaller businesses.

- **Economic boost:** Promote local businesses by recommending them to new users, increasing regional growth.

- **User experience:** Creates a smoother experience, providing more personalized dining suggestions.

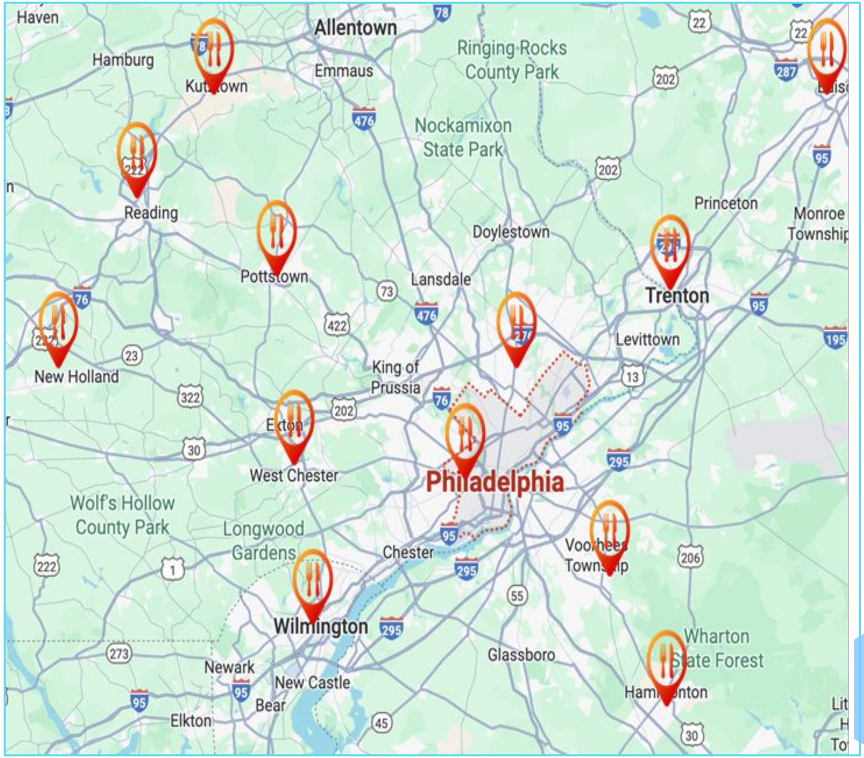


Figure 2: Map representation of Philadelphia and surrounding cities.

02 Dataset Overview

Raw Dataset

```
Dataset_User_Agreement.pdf 0.08 MB
yelp_academic_dataset_business.json 113.36 MB
yelp_academic_dataset_checkin.json 273.67 MB
yelp_academic_dataset_review.json 5094.4 MB
yelp_academic_dataset_tip.json 172.24 MB
yelp_academic_dataset_user.json 3207.52 MB
```

- **Source:** yelp

- **Datasets used:**

  Business dataset: address, city, ZIP code, categories, review count, avg stars

  Review dataset: review ID, user ID, business ID, text, date, stars given

  User dataset: user ID, review count, avg stars

	review_id	user_id	business_id	stars_x	text	date	review_count_x	average_stars	name_y	address	city	state	postal_code	stars_y	review_count_y	is_open	attributes	categories
0	KU_O5udG6zpxOg-VcAEodg	mh_-eMZ6K5RLWhZyISBhwA	XQfwVwDr-v0ZS3_CbbE5Xw	3	If you decide to eat here, just be aware it is...	2018-07-07 22:09:11	33	4.06	Turning Point of North Wales	1460 Bethlehem Pike	North Wales	PA	19454	3.0	169	1	{'AcceptsInsurance': None, 'AgesAllowed': None...	Restaurants, Breakfast & Brunch, Food, Juice B...
2	AqPFMleE6RsU23_auESxiA	_7bHUi9Uuf5__HHc_Q8guQ	kxX2SOes4o-D3ZQBkiMRfA	5	Wow! Yummy, different, delicious. Our favo...	2015-01-04 00:01:03	9	4.78	Zaika	2481 Grant Ave	Philadelphia	PA	19114	4.0	181	1	{'AcceptsInsurance': None, 'AgesAllowed': None...	Halal, Pakistani, Restaurants, Indian
6	oyaMhzBSwfGgemSGuZCdwQ	Dd1jQj7S-BFGqRbApFzCFw	YtSqYv1Q_pOltsVPSx54SA	5	Tremendous service (Big shout out to Douglas) ...	2013-06-24 11:21:25	2	5.00	Rittenhouse Grill	1701 Locust St	Philadelphia	PA	19103	3.5	290	1	{'AcceptsInsurance': None, 'AgesAllowed': None...	Wine Bars, Restaurants, Nightlife, Steakhouses...
7	Xs8Z8lmKkosqW5mw_sVAoA	IQsF3Rc6IgCzjVV9DE8KXg	eFvzHawVJofxSnD7TgbZtg	5	My absolute favorite cafe in the city. Their b...	2014-11-12 15:30:27	182	3.41	Good Karma Cafe	928 Pine St	Philadelphia	PA	19107	4.0	249	1	{'AcceptsInsurance': None, 'AgesAllowed': None...	Food, Cafes, Coffee & Tea, Restaurants
9	G_5UczbCBJriUAbxz3J7Tw	cIWLI5OZP2ad25ugMVI8gg	x4XdNhp0Xn8lOivzc77J-g	5	Best thai food in the area. Everything was au...	2013-08-15 15:27:51	17	3.24	Thai Place Restaurant	700 Nutt Rd, Ste 730	Phoenixville	PA	19460	4.5	222	1	{'AcceptsInsurance': None, 'AgesAllowed': None...	Thai, Restaurants

Figure 3: Example of the columns in the business, review and user concatenated dataset

02 Dataset Overview

Exploring the Dataset

1. Filter for location with **highest number of reviews**: **Pennsylvania, USA**
2. Filter for business ID including **“Restaurants”**
3. Filter for **open businesses** only

03 Part I: Natural Language Processing

Based on business categories, how is the general landscape of restaurants in Pennsylvania?

Dataset: business IDs unique, open and categorised as “Restaurants” in Pennsylvania

Methodology:

1. **Category text normalization:** cleaning and standardizing category names: lowercase and replace symbols
2. **Merge by business:** ensure each restaurant is represent only once
3. **Custom tokenizer:** split categories by commas only
4. **Define stopwords:** removing generic, non-food related words
5. **TF-IDF vectorization:** convert category text into weighted numerical features.
6. **Build feature matrix** (restaurants x category terms)

03 Part I: Natural Language Processing

Key Points:

- Pennsylvania's restaurant landscape is dominated **by diverse options**.
- **Highest TF-IDF weights:** casual, everyday food categories such as *pizza*, *sandwiches*, *coffee and tea*, and *breakfast and brunch*.
- Smaller terms like *sushi bars*, *mexican*, *vegan*, and *seafood* reflect more **niche culinary options.**

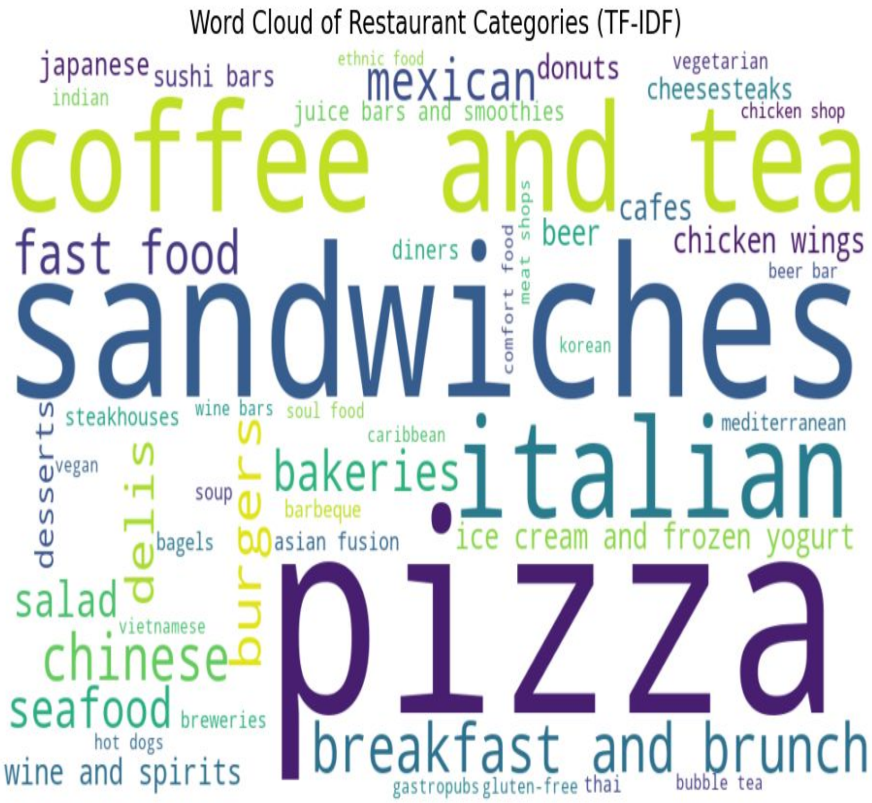


Figure 4: Word cloud with word sizes reflecting the relative TF-IDF weight, showing how characteristic or distinctive a category is among all restaurants.

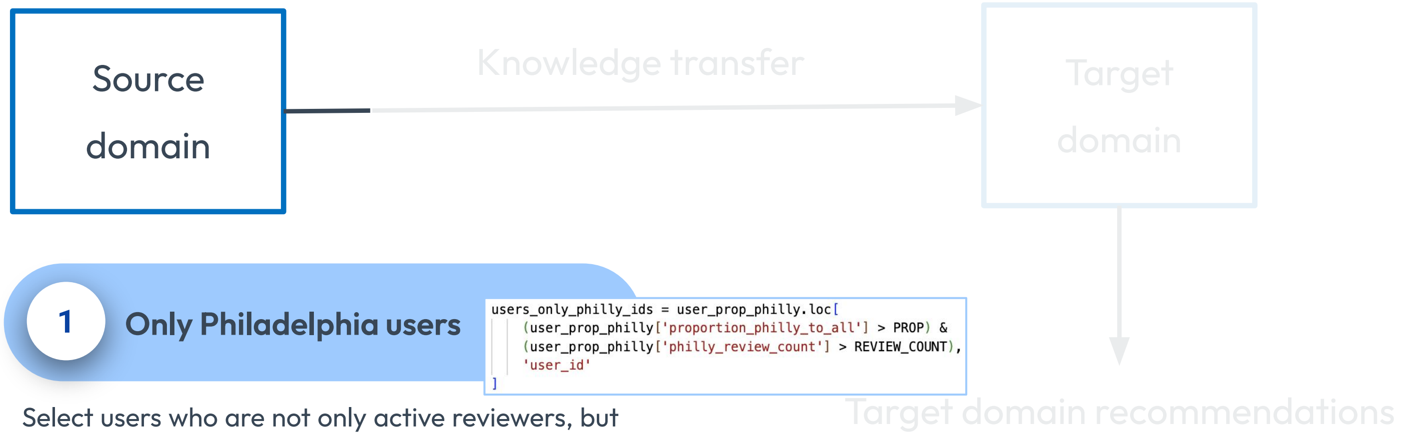
03 Part I: Recommender System

Rationale to a Recommender System

Cross-domain (content-based) recommender system

03 Part I: Recommender System

Rationale to a Recommender System



1 **Only Philadelphia users**

```
users_only_philly_ids = user_prop_philly.loc[
    (user_prop_philly['proportion_philly_to_all'] > PROP) &
    (user_prop_philly['philly_review_count'] > REVIEW_COUNT),
    'user_id'
]
```

Select users who are not only active reviewers, but who primarily review restaurants in **Philadelphia**.

2 **Philadelphia users' taste profiles**

Analyze the restaurant reviews of **Philadelphia users** to generate a taste **vector** representing their preferences (e.g., high values for attributes like 'spicy' and 'sushi', and low values for 'buffets').

03 Part I: Recommender System

Recommender System

Methodology

1. For **each user**, all the restaurants they have reviewed in **Philadelphia** are gathered, and represented by their **TF-IDF category vectors**, which describe each restaurant's category.

2. The **user's taste profile** is created by computing a **weighted average** of these vectors, where the weights correspond to the user's ratings (stars - 3):
   - Positive weights for highly rated restaurants (4-5 ⭐)
   - Neutral weights for average ratings (3 ⭐)
   - Negative weights for poorly rated restaurants (1-2⭐)

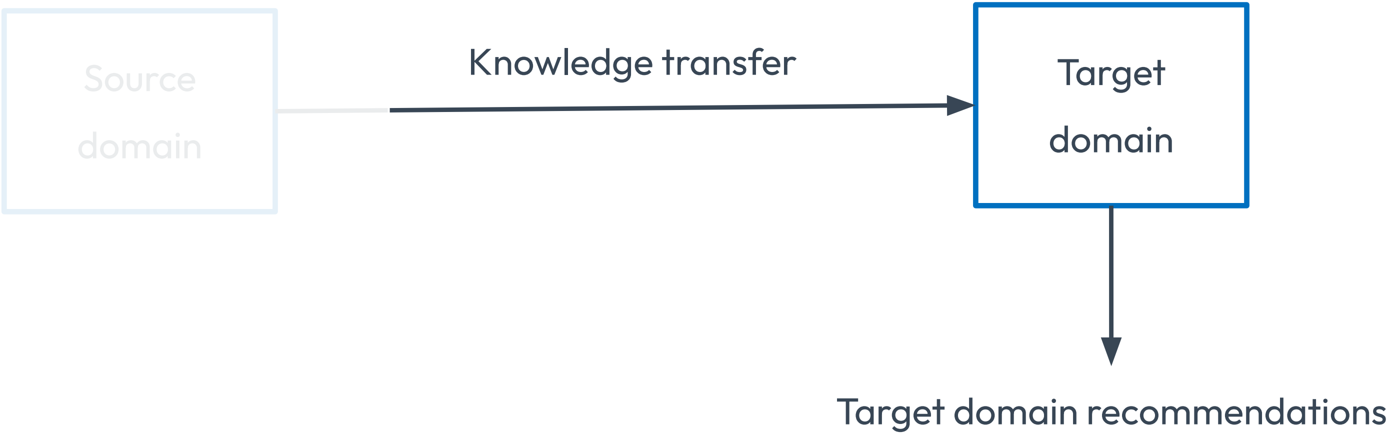
3. The resulting vector is normalized, allowing user profiles to be compared on a common scale.

	acai bowls	afghan	african	arabic	argentine	armenian	asian fusion	australian	bagels	bakeries	...
-33OBttPeyt_pUozPxkFEg	0.0	0.0	0.0	0.0	0.000000	0.000000	-0.082844	0.0	0.000000	0.000000	...
-bWX0mGixU6a64fqLWpMXg	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	-0.123010	...
0DTzZXI163l6wcvA4cbBYw	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...
1w1x35h0V9Nc59lpeEAn9Q	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...
2OlIQugbCFJTPuKD872ETA	0.0	0.0	0.0	0.0	0.000000	0.138118	0.143200	0.0	0.118563	0.206029	...
...	...	...	...	...	...	...	...	...	...	...	...
z5Ub2FeTWpCJEZZbvagoRw	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.528590	0.000000	...
zAQ9hl0DBBJvGQUf8CBZYA	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.102156	...
zURCJsyweyzqcPUzJHxttg	0.0	0.0	0.0	0.0	0.000000	0.000000	0.295903	0.0	0.029411	0.022612	...
znJdrKZVvLEUnpur9hQLVw	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...
zoS5f0JWThXQnpkxzzUUhQ	0.0	0.0	0.0	0.0	0.059441	0.000000	-0.028929	0.0	0.000000	0.161892	...

We successfully modeled the tastes of 78 active users across all restaurant types.

03 Part I: Recommender System

Rationale to a Recommender System



03 Part I: Recommender System

Generating recommendations

Methodology

Hybrid approach

1. Similarity score: calculates the **cosine similarity** between the **user's taste profile** vector and the **feature vectors of all candidate restaurants**. This score represents how well each restaurant matches the user's personal taste.

2. Credibility score: a **Bayesian average** that prevents restaurants with only a few 5-star reviews from dominating. It shrinks a restaurant's rating towards the average rating of all restaurants, with the effect being stronger for restaurants with fewer reviews. This score represents general quality and popularity:

$$\textit{Blending score} = \alpha \times \textit{Similarity} + (1 - \alpha) \times \textit{Credibility}$$

A higher α gives more weight to personal taste (similarity), while a lower α gives more weight to general quality/popularity (credibility).

03 Part I: Recommender System

Rankings

Example: 0DTzZXI163I6wcvA4cbBYw user reviewed these restaurants in Philadelphia:

name_y	stars_x	categories
Broad Street Diner	5	Diners, American (Traditional), Restaurants
Devil's Den	5	Nightlife, Bars, American (Traditional), Pubs, Breakfast & Brunch, Restaurants, Beer Bar
Station Bar & Grill	5	Bars, Arts & Entertainment, Nightlife, Sports Bars, Restaurants, American (Traditional), Gastropubs, Music Venues
Sam's Morning Glory Diner	5	Breakfast & Brunch, Restaurants, Diners
Soul Boat	5	Restaurants, Seafood, Soul Food
Ristorante Pesto	5	Restaurants, Italian
Green Eggs Café	4	Restaurants, American (New), Diners, Breakfast & Brunch
Pat's King of Steaks	3	Italian, American (Traditional), Sandwiches, Fast Food, Restaurants, Local Flavor, Cheesesteaks
Scannicchio's	2	Seafood, Italian, Restaurants, Mediterranean
Domino's Pizza	2	Sandwiches, Pizza, Restaurants, Chicken Wings
Isabella Pizza	1	Burgers, Italian, Restaurants, Pizza, Breakfast & Brunch

03 Part I: Recommender System

Rankings – comparing α weights

Recommendations (α=0.8)

name_y	city	stars_y	review_count_y	categories	blended_score
The Metropolitan American Diner & Bar	North Wales	4.0	506	American (Traditional), Diners, Bars, Restaura...	0.944696
Metro Diner	Bensalem	4.0	491	Restaurants, Diners, Breakfast & Brunch, Ameri...	0.944438
Classic Diner	Malvern	4.0	414	Diners, American (New), American (Traditional)...	0.942923
Cross Keys Diner	Doylestown	4.0	204	Restaurants, Breakfast & Brunch, Diners	0.935994
Folcroft Diner	Folcroft	4.0	120	Restaurants, Diners, American (Traditional), B...	0.930826
Pat's Colonial Kitchen	Newtown	4.0	87	Restaurants, Breakfast & Brunch, Diners	0.928022
Sunrise Diner	Croydon	4.0	72	Diners, Food, Restaurants, Breakfast & Brunch	0.926538
Coffee Cup Restaurant	Downingtown	4.0	70	Diners, Breakfast & Brunch, Restaurants	0.926329
Double D Diner	Coatesville	4.0	38	Diners, Restaurants, Breakfast & Brunch	0.922535
Fran Keller's Eatery	Kennett Square	4.0	35	Breakfast & Brunch, Restaurants, Diners	0.922131

Recommendations (α=0.2)

name_y	city	stars_y	review_count_y	categories	blended_score
El Limon	Conshohocken	4.5	1001	Mexican, Restaurants	0.854127
Grumpy's Handcarved Sandwiches	Pottstown	5.0	170	Restaurants, Breakfast & Brunch, Sandwiches	0.836725
The Salt House	New Hope	4.5	326	Restaurants, Gastropubs	0.825726
Station Taproom	Downingtown	4.5	267	Gastropubs, Beer Bar, Restaurants, Nightlife, ...	0.824772
The Couch Tomato Bistro	Manayunk	4.5	681	Italian, Tapas/Small Plates, Gluten-Free, Bars...	0.820338
Annamarie's Place	Royersford	4.5	326	Restaurants, Breakfast & Brunch, American (Tra...	0.818626
Bittersweet Kitchen	Media	4.5	402	Restaurants, Coffee & Tea, Food, Breakfast & B...	0.814488
Blue Sage Vegetarian Grille	Southampton	4.5	572	Restaurants, Vegan, Vegetarian	0.814411
Carlucci's Grill Yardley	Yardley	4.5	434	Restaurants, Italian	0.806393
DuBu	Elkins Park	4.5	509	Restaurants, Vegetarian, Korean, Comfort Food,...	0.804643

03 Part I: Recommender System

Evaluation

Offline evaluation: “Hold-out domain” simulation

- To find users (**“bridge users”**) who are active in both **Philadelphia** and **another city** (in this case, King of Prussia, which has the most shared users with Philadelphia with more than 10 reviews in each).

- For these bridge users, the **"ground truth"** is based on their King of Prussia reviews. Restaurants in KoP rated 4 stars or higher are considered **"good” recommendations**.

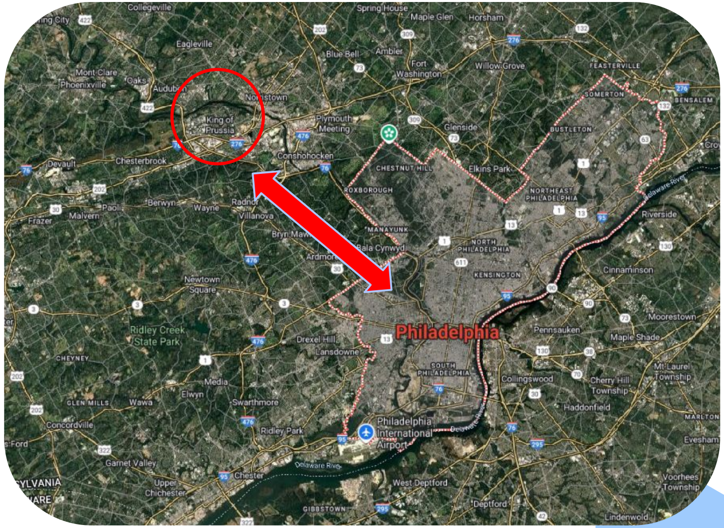


Figure 5: Satellite view representation of Philadelphia and King of Prussia.

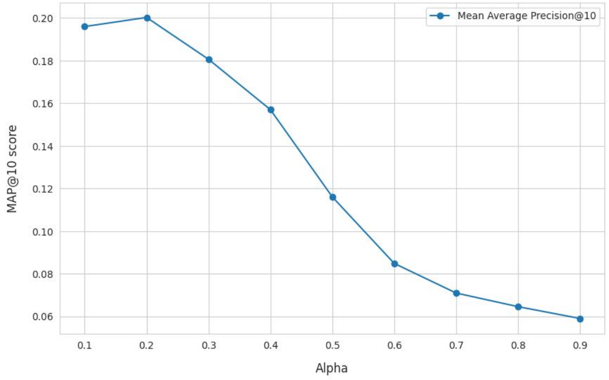
03 Part I: Recommender System

Evaluation

- For this cross-domain task, **a restaurant's credibility is more powerful predictor of user preference than a direct match of its content features** (categories).

- At its peak performance and without robust fine-tune implementations, the model achieves a solid result. It serves as a **good baseline** and proves that the methodology of building taste profiles from one domain and applying them to another is valid.

Average Precision @ 10



Precision and Recall @ 10

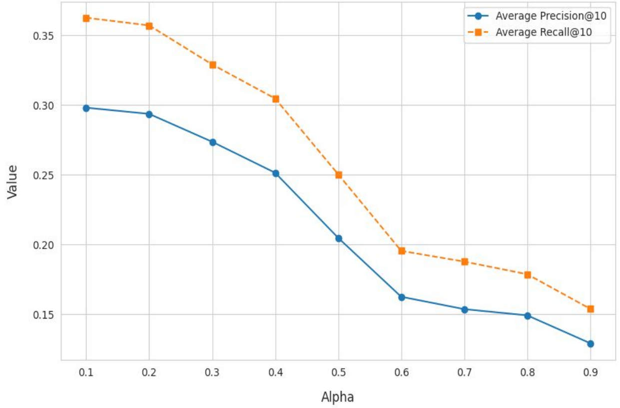


Figure 6: Graphical representation of average precision (upper graph) and precision and recall (bottom graph).

03 Part I: Recommender System

Evaluation - Baseline models

Similarity model (α = 1)

Blending score = α × Similarity + (1 − α) × Credibility

Credibility model (α = 0)

Blending score = α × Similarity + (1 − α) × Credibility

	Baseline models				
	Similarity model	Δ	Credibility model	Δ	Our model (α = 0.2)
Precision @ 10	0.129	+127.13%	0.300	-2.33%	0.293
Recall @ 10	0.155	+130.32%	0.364	-1.92%	0.357
Average Precision @ 10	0.067	+198.51%	0.204	-1.96%	0.200

04 Part II: Social Network Analysis

How are users connected to one another, considering their preferred food taste?

Dataset: user profiles built (restaurant features x user ID)

Methodology:

1. Pairwise cosine similarities between user profiles matrix computed
2. Symmetrized and cleaned matrix: ensures all connections are mutual and diagonal values are zero
3. Base graph built and fixed layout
4. Iterated over selected threshold: 0.5 *

* Comparison across different thresholds performed (see *Supplementary Data*).

04 Part II: Social Network Analysis

Two distinct components emerge, suggesting groups with similar but internally cohesive culinary interests:

- **Right-hand cluster** is **dense**, indicating **stronger shared preferences** among those users.
- **Left-hand cluster is smaller but tightly interconnected — a niche community**.
- Scattered isolated nodes represent users with unique or less overlapping preferences, who do not strongly connect with others at this threshold.

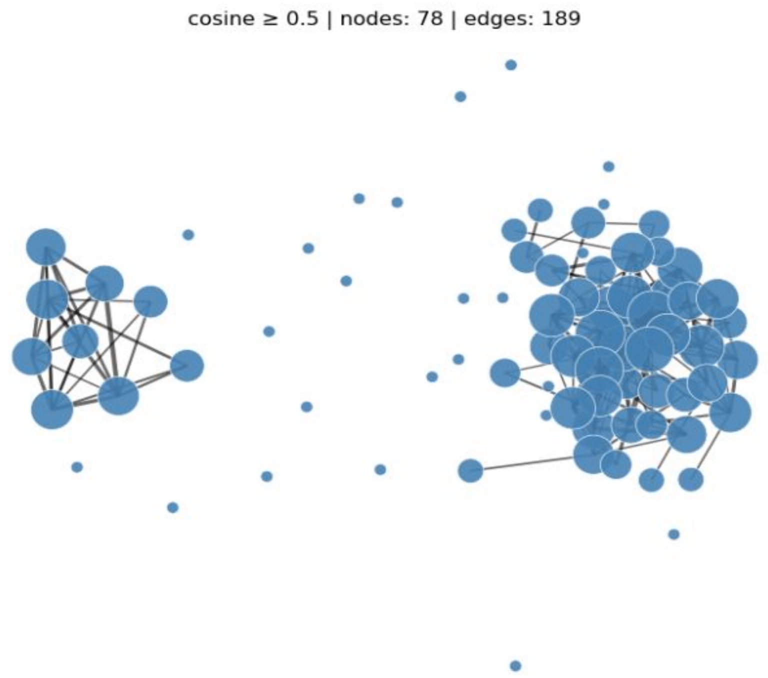


Figure 7: User Affinity Network at Cosine ≥ 0.5. Each node represents a user; edges connect users with a cosine similarity ≥ 0.5 in restaurant preferences.
Larger nodes indicate higher connectivity (degree).

04 Part II: Social Network Analysis

What communities are formed, considering the users taste profiles?

Dataset: user profiles built (restaurant features x user ID)

Methodology:

1. Compute pairwise users profiles cosine similarities matrix

2. **Communities** were detected using the Louvain algorithm:
   - **Modularity (Q) based** algorithm
   - **Bottom-up** agglomerative (builds communities)
   - More **efficient and scalable** for large networks
   - Generates stable and interpretable modular structures
   - **Optimal for dense or weighted graphs** (such as the cosine -based user similarity)

* Modularity per threshold assessment performed (see *Supplementary Data*).

04 Part II: Social Network Analysis

Key points:

Profile communities. For each community, the top TF-IDF categories were computed (mean score per category):

- **Community 0 (N=8)**

*breakfast and brunch: 0.322. sandwiches: 0.266. italian: 0.240. mexican: 0.182. **asian fusion**: 0.161. sushi bars: 0.159. pizza: 0.147 salad: 0.129.*

- **Community 1 (N=13)**

*breakfast and brunch: 0.469 mexican: 0.198 sandwiches: 0.164 **seafood**: 0.136 coffee and tea: 0.130 **burgers**: 0.128 **diners:** 0.116 **desserts**: 0.112*

- **Community 2 (N=11)**

***italian**: 0.380 pizza: 0.273 breakfast and brunch: 0.227 sandwiches: 0.143 gastropubs: 0.142 cafes: 0.139 wine bars: 0.130 desserts: 0.113*

- **Community 3 (N=9)**

***sandwiches:** 0.271 **vegan**: 0.234 pizza: 0.222 delis: 0.156 sushi bars: 0.149 seafood: 0.148 breakfast and brunch: 0.144 coffee and tea: 0.131*

- **Community 4 (N=4):**

***coffee and tea**: 0.328 **bubble tea**: 0.279 **juice bars and smoothies**: 0.276 cafes: 0.249 taiwanese: 0.192 ethnic food: 0.170 japanese: 0.169 sushi bars: 0.152*

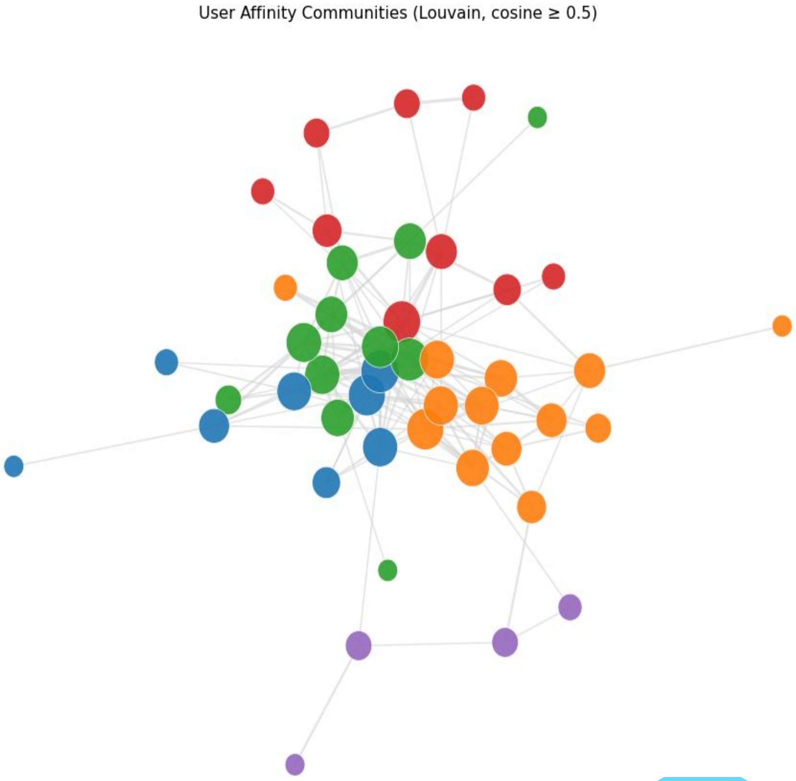


Figure 8: User Affinity Communities determined using Louvain algorithm at Cosine ≥ 0.5. Each node represents a user; edges connect users with a cosine similarity ≥ 0.5 in restaurant preferences. Different colors represent the different communities found.

04 Part II: Social Network Analysis

Key points:

- Community 0 (N=8), Community 1 (N=13) and Community 2 (N=11) appear **well-defined** and **internally compact**, confirming that the graph partition reflects real preference similarity.

- **Community 3 (N=9) and Community 4 (N=4) more peripheral**, suggesting smaller or niche taste groups

- The overall spatial distribution verifies that the **network-based grouping matches the underlying data geometry**

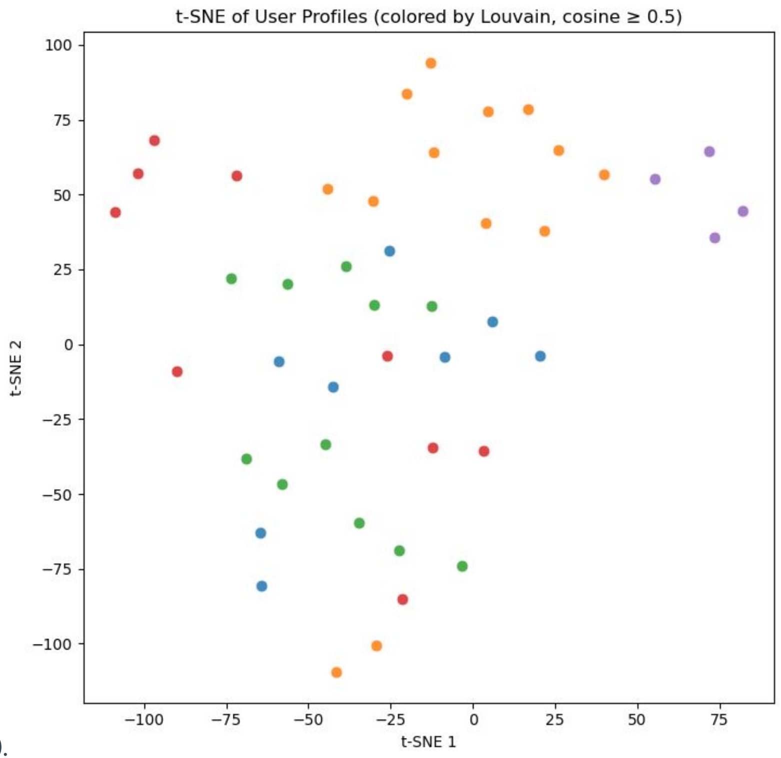


Figure 9: t-SNE graph os user profiles colored by its correspondent determined Louvain community attributed.

Note: Assessment of Top 15% influencers evaluated (see *Supplementary Data*).

04 Part II: Time series Forecasting

How does the number of reviews vary per month, for users that do not belong to any community?

Users not belonging to any community were selected for time series forecasting because they represent independent behavioural patterns, free from network influence or collective bias.

Dataset: Number of reviews each month for uses that did not belong to any community

Methodology:

1. Count the number of reviews aggregated by month
2. Line plot of the series
3. Checking for missing dates
4. Series decomposition Using STL with Log Values
5. Grid Search for the ETS model
6. Residual Analysis

04 Part II: Time series Forecasting

Figure 10: Average YELP reviews per month between 2011 and 2019, for users that did not belong to any community .

Key Points:

- **Variance increase** with time
- **Clear** trend
- **Hard to see a seasonality** component
- **No missing dates** in the time series
- Straight linear **decomposition** is not best suited for this kind of time series, therefore **log transformed data** was used

04 Part II: Series Decomposition

Key Points:

- Applied **Log transformation to the series**, in order to remove changes in Variance
- **Changes in seasonal pattern through time**
- **Residual centered around zero** and a small scale
- **STL decomposition handled well changes over time in seasonal** component
- STL only works with **additive decomposition**, that why the log transformation

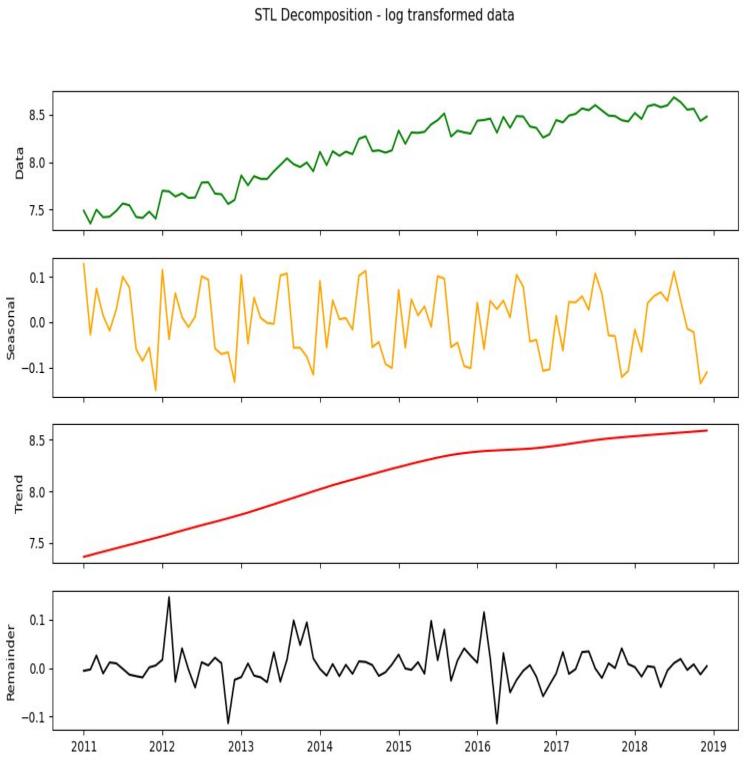
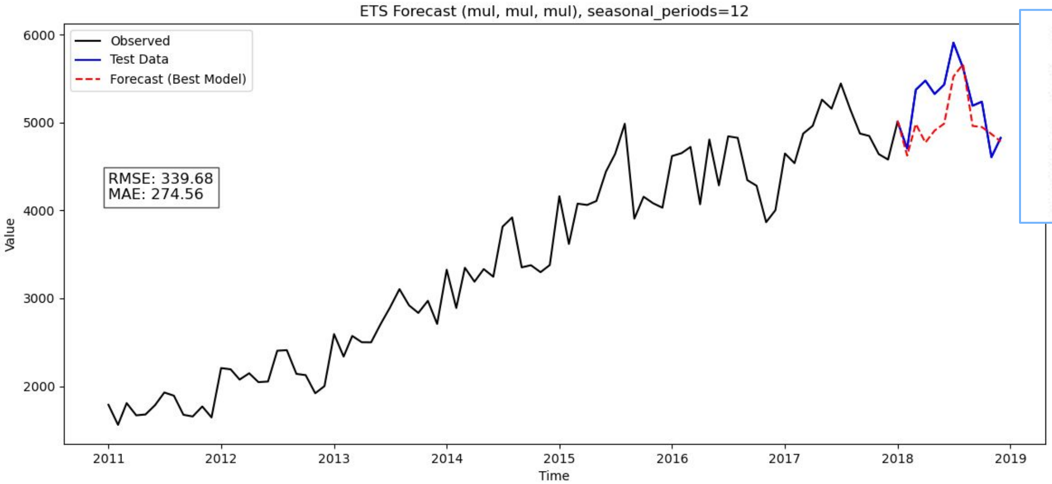


Figure 11: Series decomposition of the average YELP reviews per month between 2011 and 2019, for users that did not belong to any community .

04 Part II: Forecasting




```
Best configuration:
error              mul
trend              mul
seasonal           mul
periods             12
MAE         274.561156
RMSE        339.679741
```

Figure 12: ETS forecast of the average YELP reviews per month between 2011 and 2019, for users that did not belong to any community .

Key Points:

- Grid Search to tune the best hyperparameters
- Since the series showed a great change over time an ETS model has been seen to be a good alternative
- ETS performs a exponentially decreasing weights to past observations,

04 Part II: Model Fit

Key Points:

- The **Portmanteu test** indicates that the residual is likely **white noise**
- The **model** appears to have **good fit** to the original series
- The **ADF test** shows that the residuals are likely **stationary**

Figure 13: Model fit of the ETS forecast of the average YELP reviews per month between 2011 and 2019, for users that did not belong to any community .

05 Conclusions

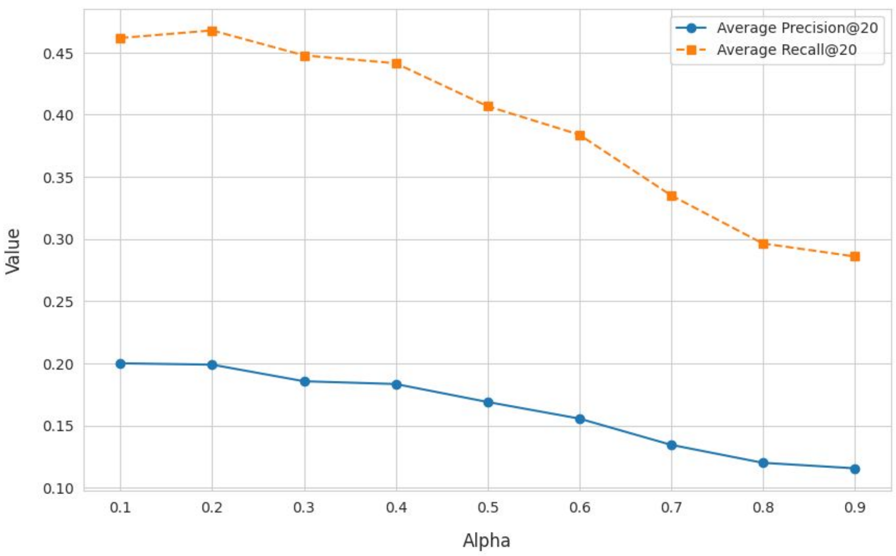
Take-Home Messages

- **NLP** of open restaurants in Pennsylvania reveal a **high-diversity** of categories, with **casual**, everyday food categories such as *pizza*, *sandwiches*, *coffee and tea*, and *breakfast and brunch*.
- The **Recommender System** reveals an acceptable evaluation. As future perspectives, some approaches could be implemented:
  - ★ **One-hot encode** restaurants' `attributes` column and concatenate them with the existing TF-IDF feature vectors, allowing the model to learn preferences for price, ambiance, services.
  - ★ Implementing a **pre-trained sentence transformer model** (e.g., BERT) in the review text for each restaurant, as this embeddings capture **semantic meaning** and can lead to much better similarity matching.
- Our **Social Network** analysis reveal that the users selected may be grouped in 5 communities of food preferences. Interestingly, for 4 of these communities at least one top **15% influencer** was found.
- Considering the **Time Series** forecasting, we observed an **increase variance** with time of the number of reviews of users that were not attributed to any community, with a **clear trend** but hard to see **seasonality**. For forecasting, using the ETS model resulted in a good forecasting.

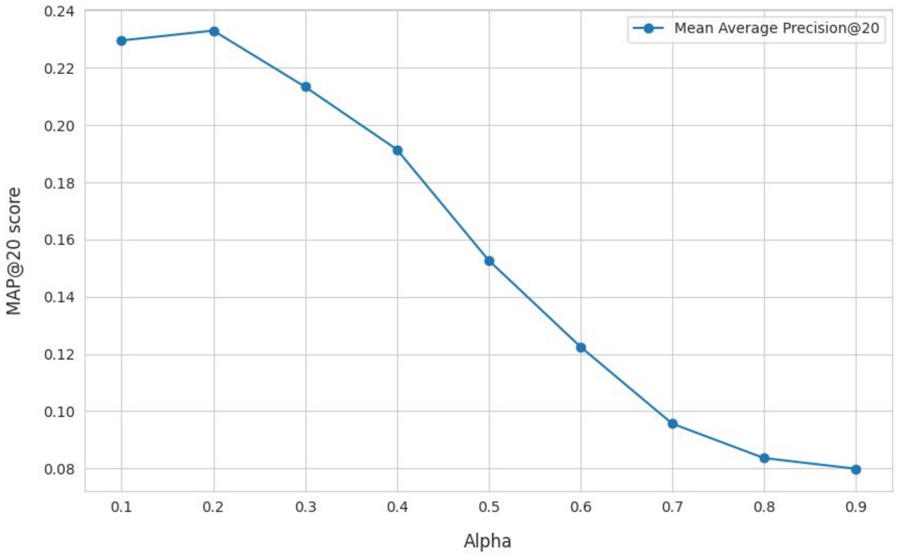
Supplementary Slides

03 Part I: Recommender System

Precision and Recall @ 20

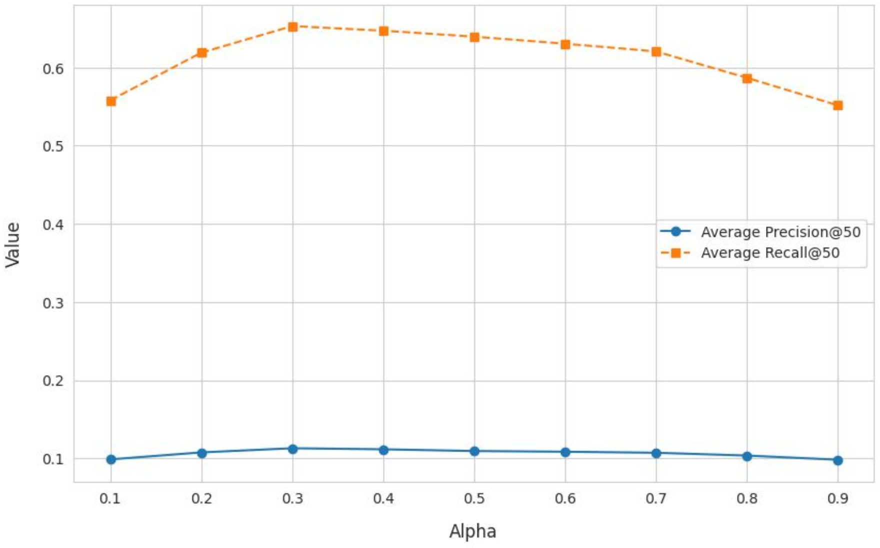


Average Precision @ 20

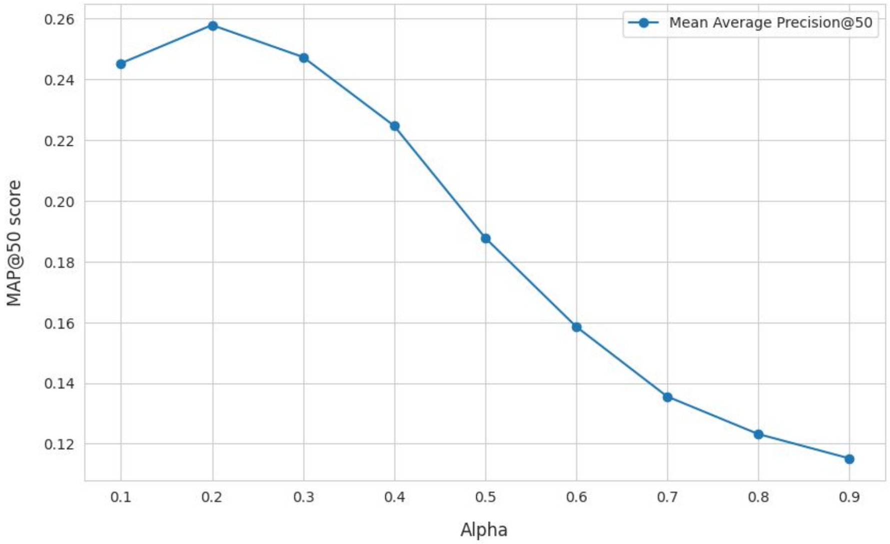


03 Part I: Recommender System

Precision and Recall @ 50



Average Precision @ 50



04 Part II: Social Network Analysis

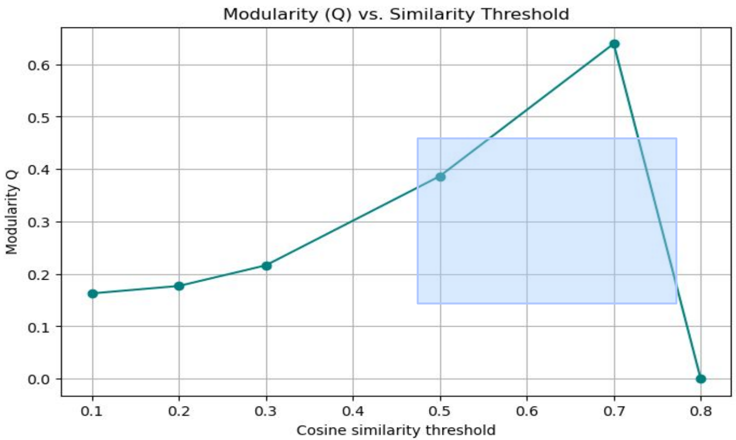


Figure S1: Modularity (Q) vs. Similarity threshold to define optimal threshold.

=== Network metrics by threshold ===

threshold	nodes	edges	density	avg_degree	avg_clustering	n_components	largest_component
0.1	78	1509	0.502498	38.692308	0.305839	1	78
0.2	78	1118	0.372294	28.666667	0.343365	1	78
0.3	78	735	0.244755	18.846154	0.366213	6	73
0.5	78	189	0.062937	4.846154	0.287811	26	45
0.7	78	13	0.004329	0.333333	0.032635	67	6
0.8	78	0	0.000000	0.000000	0.000000	78	1

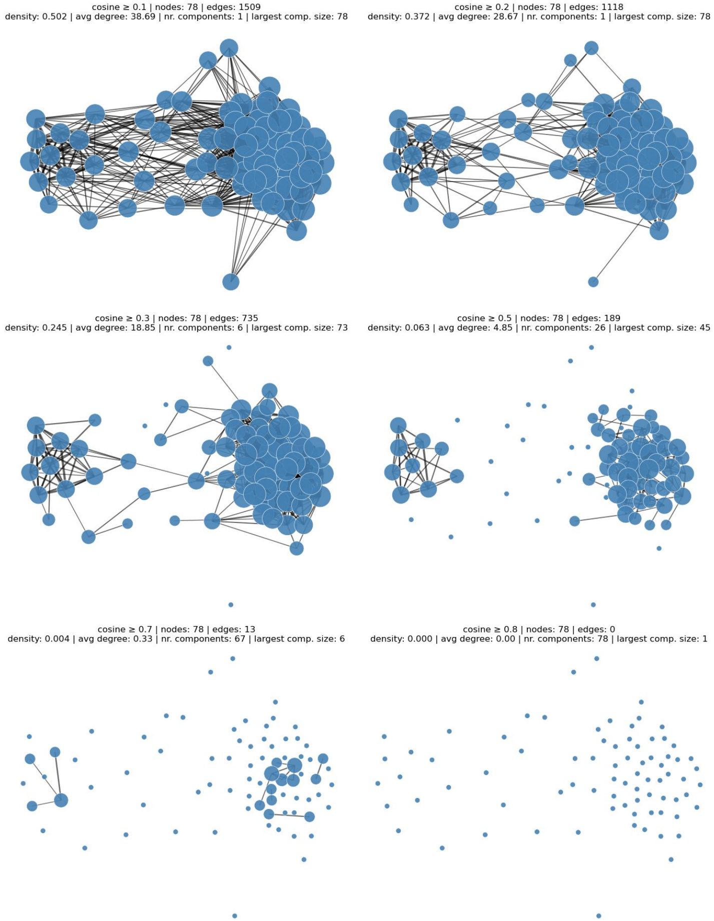


Figure S2: User Affinity Network at several cosine thresholds. Each node represents a user; edges connect users with the respective cosine similarity in restaurant preferences.
Larger nodes indicate higher connectivity (degree).

04 Part II: Social Network Analysis

```python
# Compute composite influence score (normalized ranks)
centrality_df["composite_score"] = (
    0.4 * centrality_df["degree"].rank(pct=True) +
    0.3 * centrality_df["betweenness"].rank(pct=True) +
    0.2 * centrality_df["eigenvector"].rank(pct=True) +
    0.1 * centrality_df["closeness"].rank(pct=True)
)
```

	degree	betweenness	closeness	eigenvector	community	composite_score
T4Uk_zyBFvIUsBVninUqRg	0.477273	0.231501	1.599208	0.346836	0	0.995556
_cRADNksh9fa_lrm_UnOuQ	0.477273	0.159619	1.575204	0.333369	2	0.975556
6kJFLAHV-tNsBEZaRTqEWQ	0.409091	0.187104	1.511497	0.272148	3	0.953333
YBUQi7JTik4_SRbFBv4GXQ	0.386364	0.053911	1.301385	0.229593	1	0.886667
zoS5f0JWThXQnpkxzzUUhQ	0.386364	0.045455	1.483924	0.304422	2	0.868889
XMHRH9_8T8HhjQMS81cdCw	0.272727	0.087738	1.299124	0.216488	0	0.860000
vS8H4lgp0AvtPSjrmKOQkg	0.295455	0.047569	1.271967	0.219857	2	0.837778

Figure S3: Top 15% influencers, considering the highest composite score, according to weights given by the degree, betweenness, closeness and eigenvector (right) and code to generate these metrics (left). Community 2 seems to have more influencers compared to the remaining communities. No influencer was detected for Community 4.