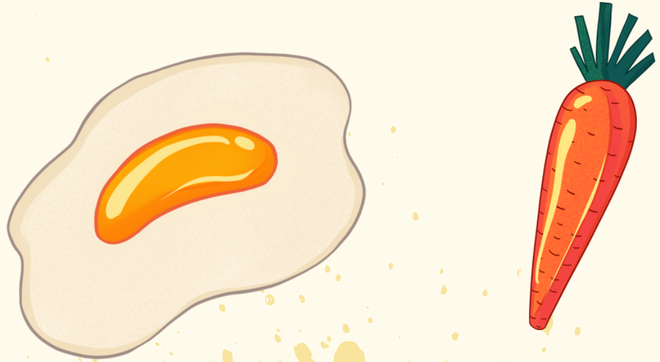
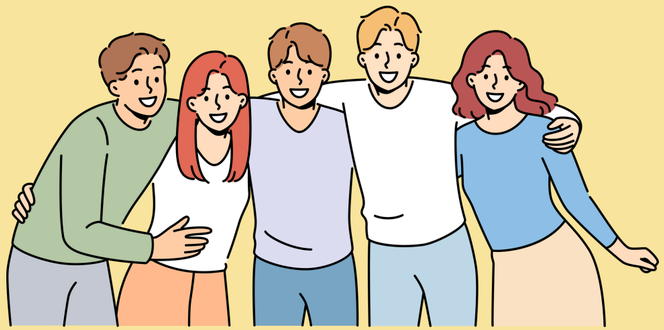


The Foodies

by Anouk de Brouwer
and Lisa Levasseur

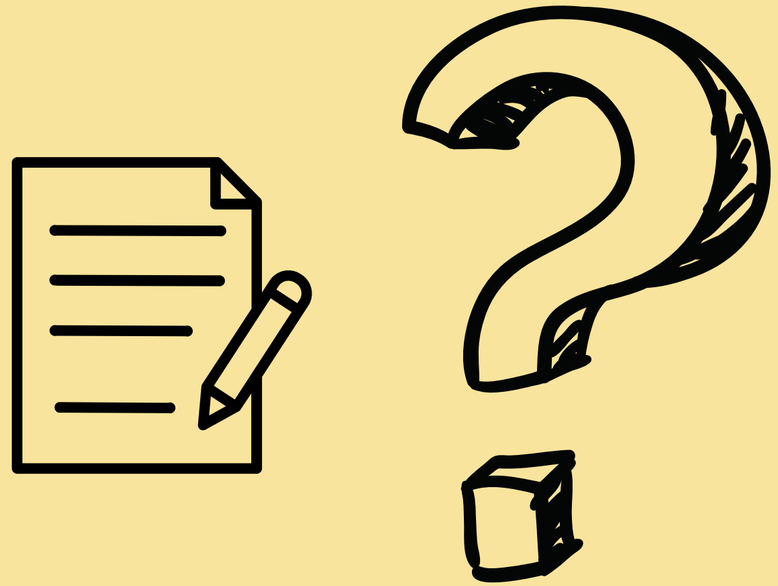
Why Recipes?

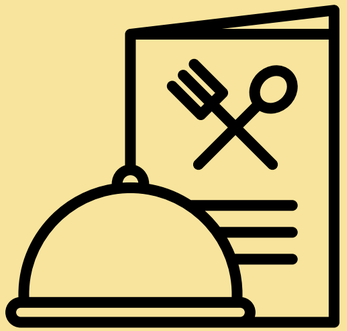
Everyone is eating food
important everyday topic



Dein Abschnittstext

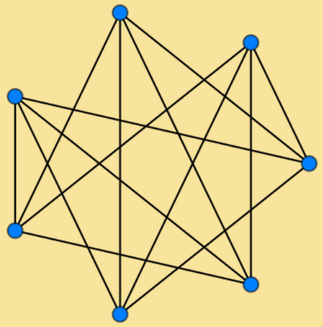
Why Recipes?





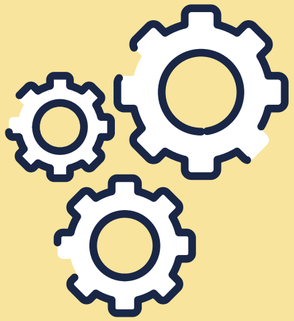
Why Recipes?

Why Graph?



Interconnection

Scalability



Customization

Dataset

Food.com → Kaggle

Groups of data :

- raw data
- user interaction
- preprocessed data

Dataset

Files :

Raw_recipes.csv

name	id	minutes	contributo...	submitted	tags	nutrition	n_steps	steps	description	ingredients	n_ingredie...
arriba baked winter squash mexican style	137739	55	47892	2005-09-16	['60-minutes-or-less', 'time-to-make', 'course', 'main-ingredient', 'cuisine', 'preparation', 'occas...	[51.5, 0.0, 13.0, 0.0, 2.0, 0.0, 4.0]	11	['make a choice and proceed with recipe', 'depending on size of squash , cut into half or fourths', ...	autumn is my favorite time of year to cook! this recipe can be prepared either spicy or sweet, you...	['winter squash', 'mexican seasoning', 'mixed spice', 'honey', 'butter', 'olive oil', 'salt']	7

Raw_interactions.csv

user_id User ID	recipe_id Recipe ID	date Date of interaction	rating Rating given	review Review text
1533 2.00b	38 538k	2000-01-25 2018-12-20	0 5	**1125284** unique values
38094	40893	2003-02-17	4	Great with a salad. Cooked on top of stove for 15 minutes.Added a shake of cayenne and a pinch of sa...

Dataset

Files :

interactions_train.csv

interactions_validation.csv

interactions_test.csv

user_id	recipe_id	date	# rating	# u	# i
8937	44551	2005-12-23	4.0	2	173538
56680	126118	2006-10-07	4.0	16	177847
349752	219596	2008-04-12	0.0	26	89896
628951	82783	2007-11-13	2.0	45	172637
92816	435013	2013-07-31	3.0	52	177935
280271	228179	2007-07-29	5.0	55	178179
345569	186470	2008-10-05	0.0	57	177482

Dataset

Files :

PP_recipes.csv

id	i	name_toke...	ingredient...	steps_toke...	techniques	calorie_level	ingredient...
424415	23	[40480, 37229, 2911, 1019, 249, 6878, 6878, 2839, 1781, 40481]	[[2911, 1019, 249, 6878], [1353], [6953], [15341, 3261], [2056, 857, 643, 1631, 20480]]	[40480, 40482, 21662, 481, 6878, 500, 246, 1614, 1911, 10757, 240, 674, 9933, 8400, 40478,	[0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	0	[389, 7655, 6270, 1527, 3406]

ingr_map.pkl

```
ingr_map = pd.read_pickle('archive/ingr_map.pkl')

print(ingr_map.head(10))
print("\Columns availables :")
print(ingr_map.columns)
```

```
\Columns availables :
Index(['raw_ingr', 'raw_words', 'processed', 'len_proc', 'replaced', 'count',
       'id'],
```

raw_ingr	raw_words	processed	len_proc	replaced	count	id
medium heads bibb or red leaf lettuce, washed,...	13	medium heads bibb or red leaf lettuce, washed,...	73	lettuce	4507	4308
mixed baby lettuces and spring greens	6	mixed baby lettuces and spring green	36	lettuce	4507	4308
romaine lettuce leaf	3	romaine lettuce leaf	20	lettuce	4507	4308
iceberg lettuce leaf	3	iceberg lettuce leaf	20	lettuce	4507	4308
red romaine lettuce	3	red romaine lettuce	19	lettuce	4507	4308
head romaine lettuce	3	head romaine lettuce	20	lettuce	4507	4308
curly endive lettuce	3	curly endive lettuce	20	lettuce	4507	4308
romaine lettuce hearts	3	romaine lettuce heart	21	lettuce	4507	4308
baby leaf lettuce	3	baby leaf lettuce	17	lettuce	4507	4308
head of lettuce	3	head of lettuce	15	lettuce	4507	4308

PP_users.csv

u	techniques	items	n_items	ratings	n_ratings
0	[8, 0, 0, 5, 6, 0, 0, 1, 0, 9, 1, 0, 0, 0, 1, 0, 5, 0, 0, 0, 3, 1, 0, 0, 0, 0, 0, 2, 2, 2, 0, 0, 0, ...	[1118, 27680, 32541, 137353, 16428, 28815, 65803, 148743, 129939, 99377, 33025, 173660, 46706, 10051...	31	[5.0, 5.0, 5.0, 5.0, 5.0, 5.0, 5.0, 4.0, 4.0, 4.0, 5.0, 5.0, 4.0, 5.0, 5.0, 5.0, 5.0, 4.0, 4.0, 5.0,...	31

Graph Structure

Nodes

Recipes

Edges

Recipes sharing same ingredients

Weights

of shared ingredients

Graph Structure

Nodes

Recipes

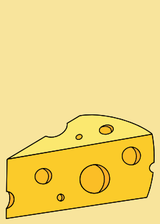
Edges

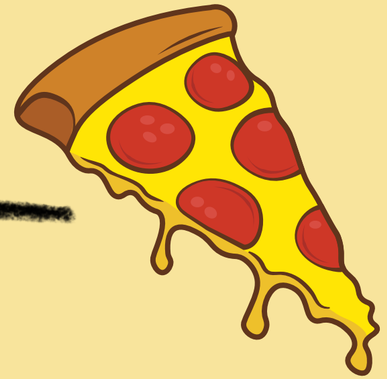
Recipes sharing same ingredients

Weights

of shared ingredients

2





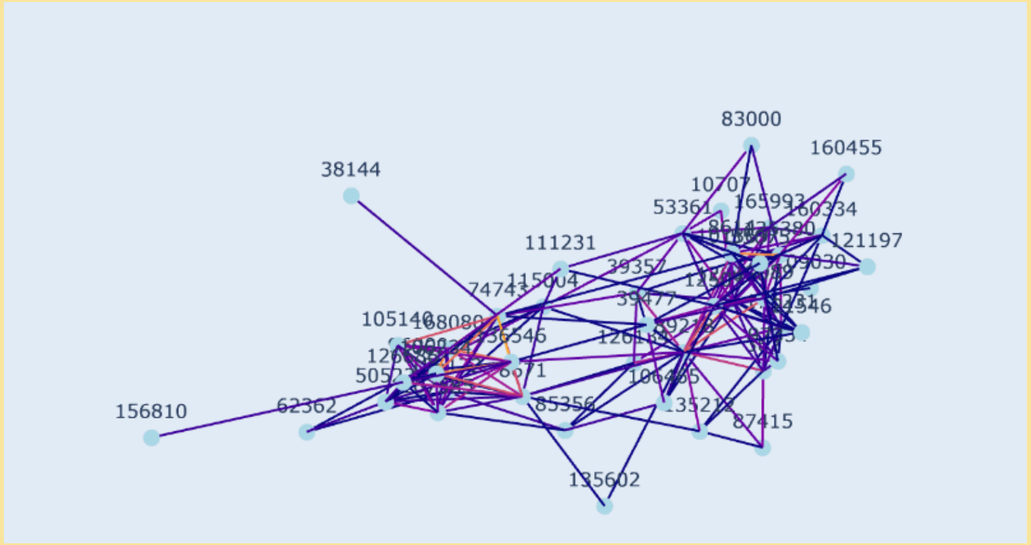
Number of Recipes

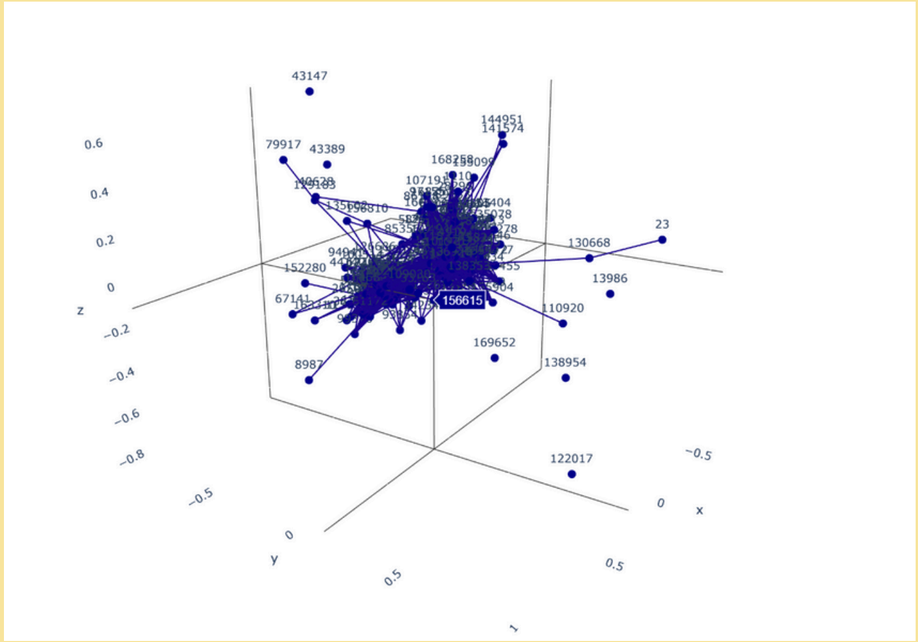
~~>200.000~~

200

Visualization

~ 200 Recipes





Optimization

Inverted index

```
# Build the ingredient-to-recipes mapping
#    {ingredient: {recipe_id1, recipe_id2, ...}}
ingredient_to_recipes = defaultdict(set)

for idx, row in df.iterrows():
    recipe_id = row['i']
    ingredients = decode_ingredients(row['ingredient_tokens'], token_to_replaced)
    for ing in ingredients:
        ingredient_to_recipes[ing].add(recipe_id)
```

Edge weights

```
# Filter edges based on weight
for (r1, r2), weight in edge_weights.items():
    if weight >= 5:
        G.add_edge(r1, r2, weight=weight)
```

Removed most frequent ingredients

(salt, water, oil...)

Analytics

Analytics

Nr of nodes & edges

Density

Average connections

Centers

Cliques

Small World phenomenom

Analytics

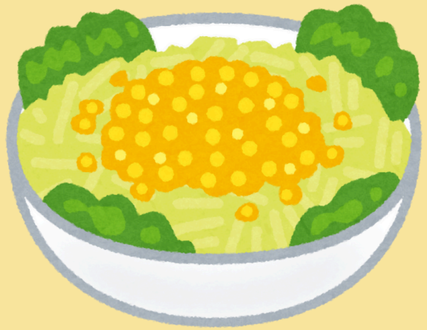
200 recipes

3000 edges

15 avg. connections

6.6 avg. shared ingredients

Analytics



Simple Corn Salad:

101 neighbours

Sugar Glazed Walnuts:

718 sum of shared ingredients

Analytics

Many Neighbours

Many shared Ingredients

Analytics

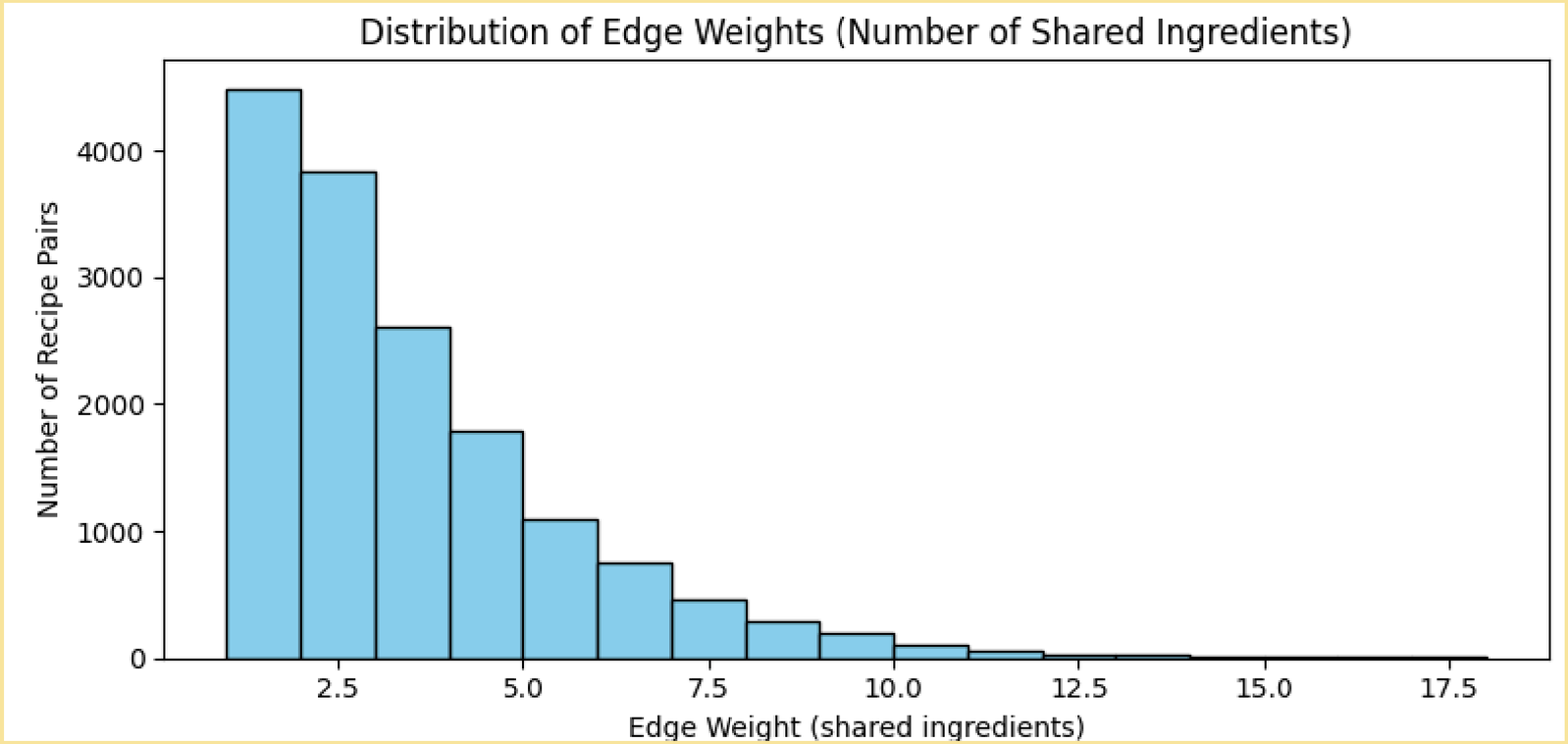
7 components

1–194 recipes

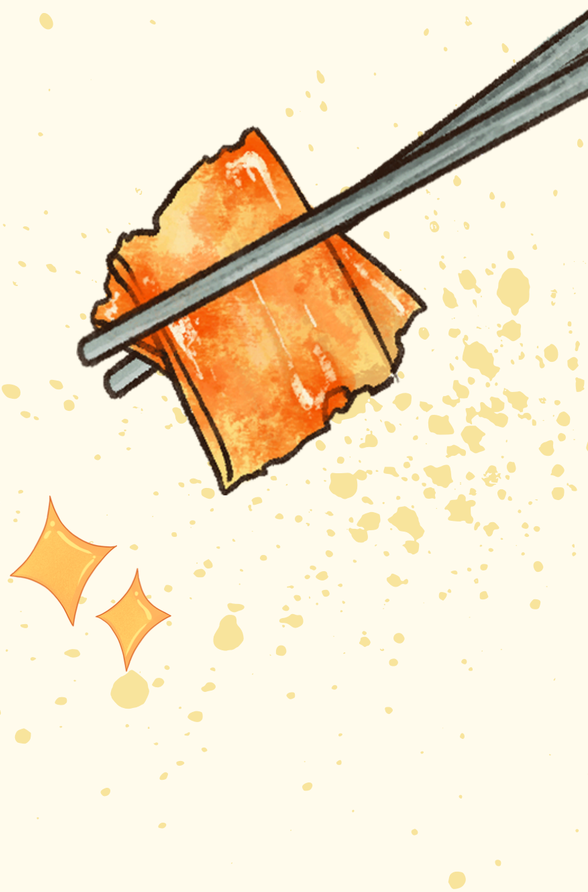
2.15 avg. shortest path

0.581 avg. clustering coefficient

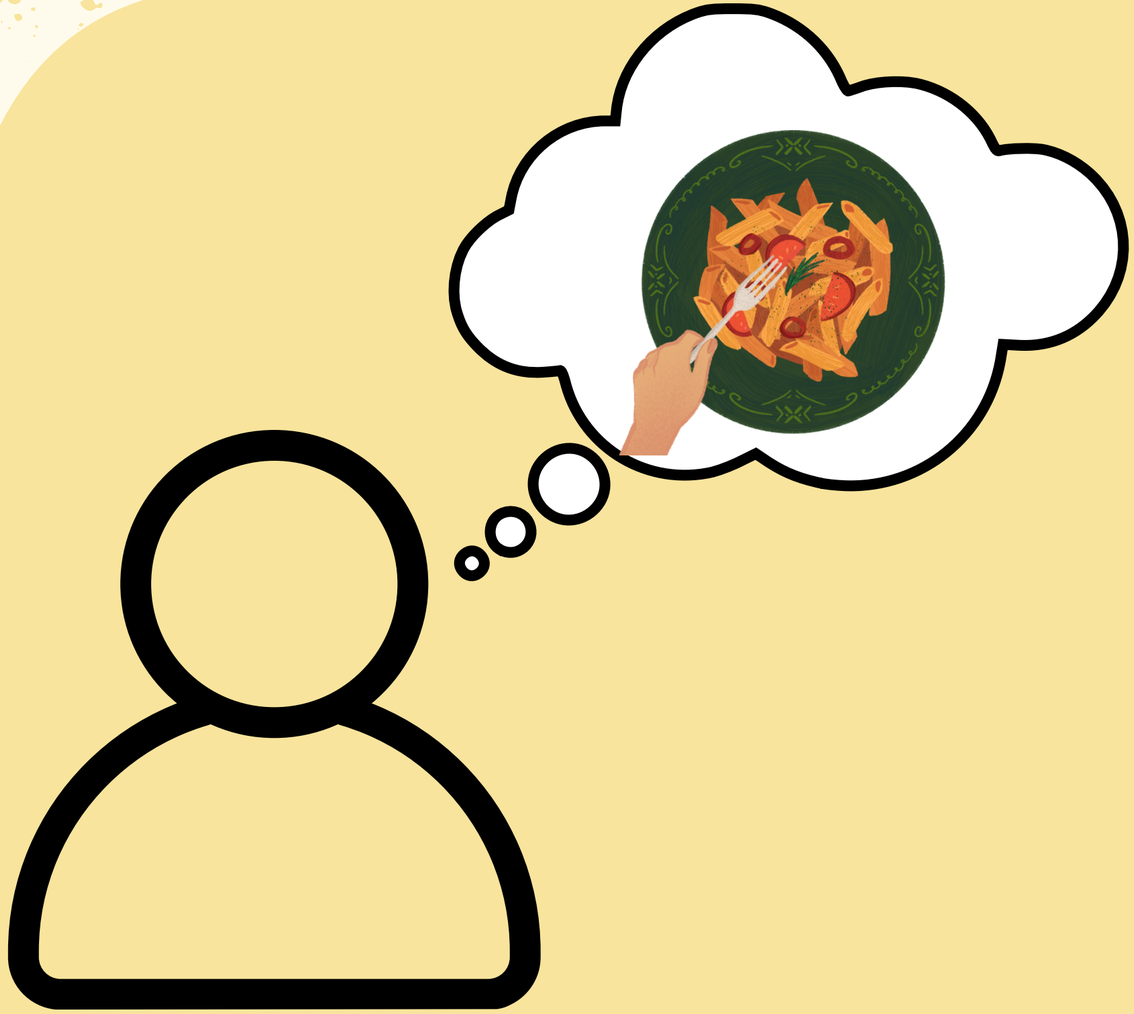
Statistics


Distribution of Edge Weights (Number of Shared Ingredients)
Number of Recipe Pairs
0
1000
2000
3000
4000
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Edge Weight (shared ingredients)

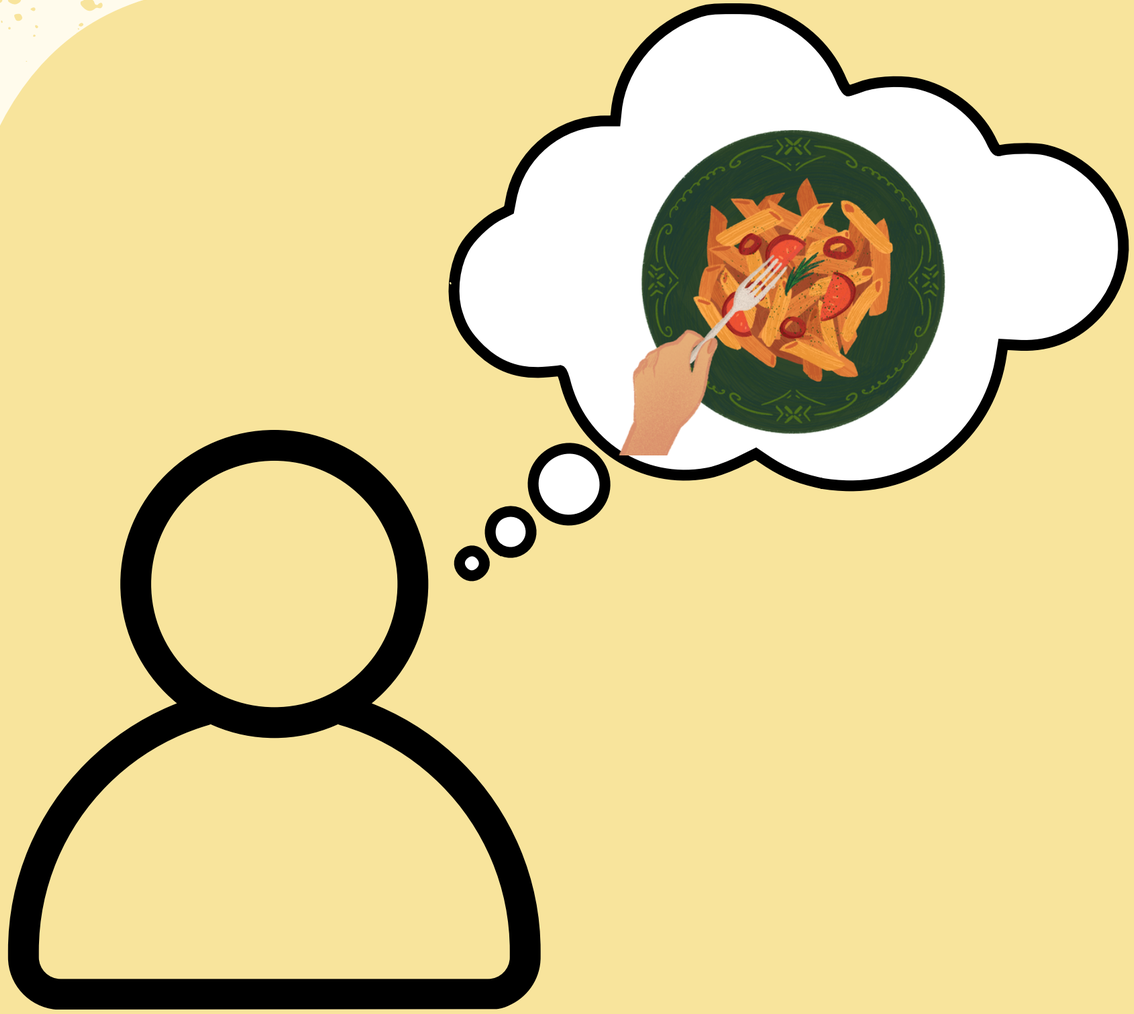
Recipe Recommendation Tool



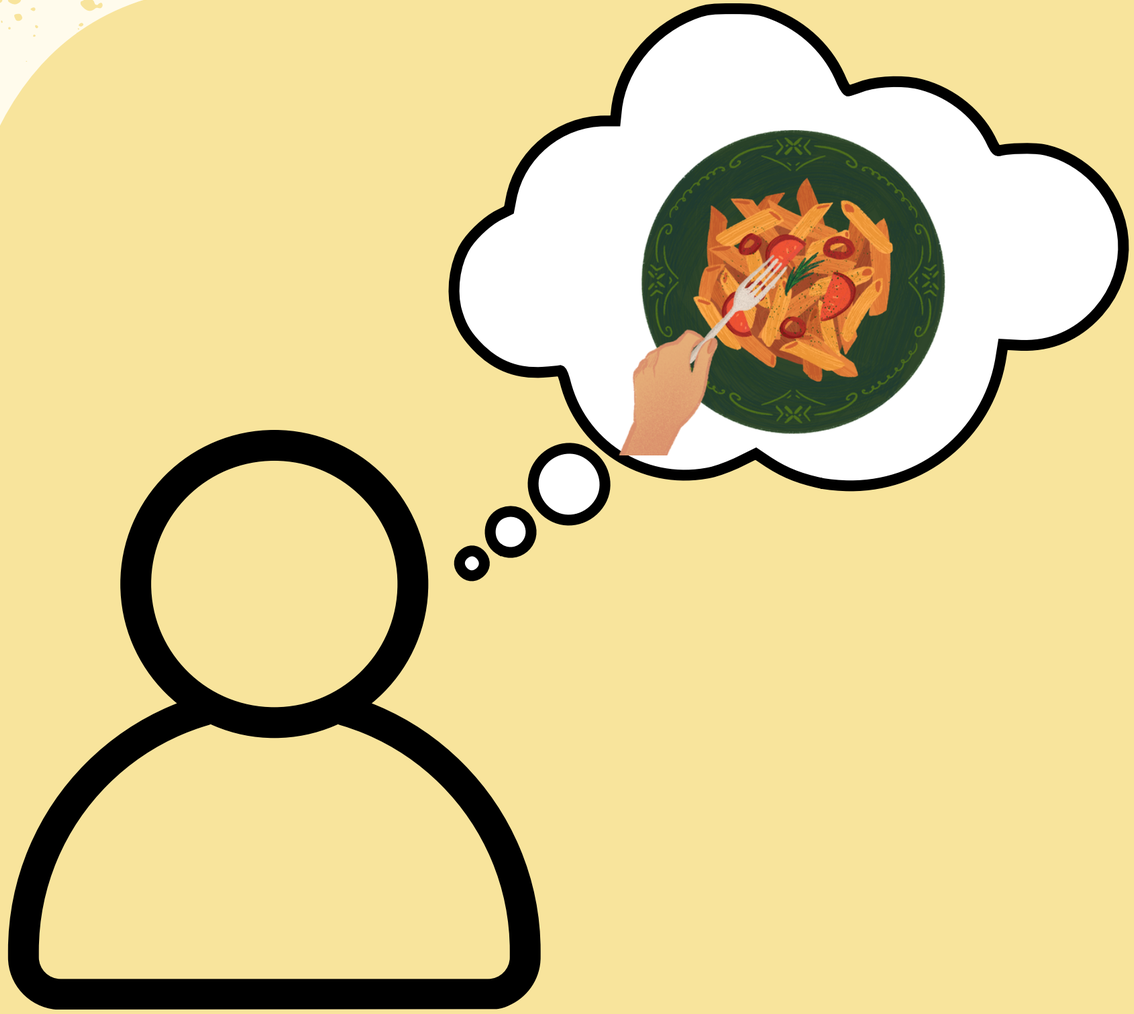
Recipe Recommender



Recipe Recommender



Recipe Recommender

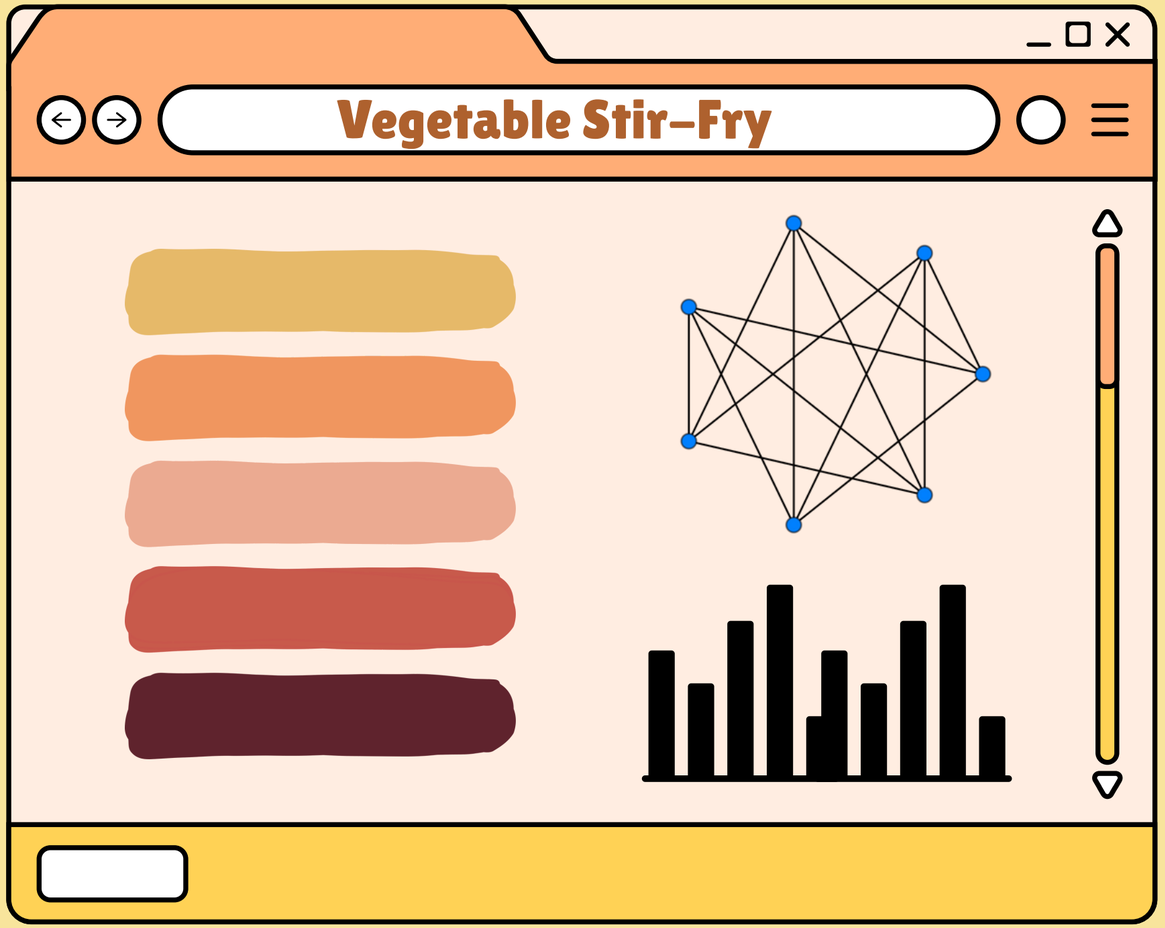


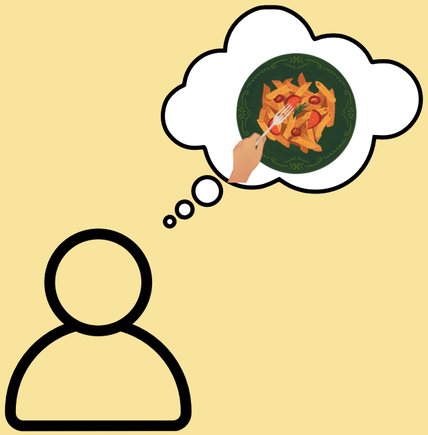
Enter any recipe:

Vegetable Stir-Fry

→ Recommended Recipes:
1. Chickpea Salad (4 shared ingredients)
2. Tofu Curry (3 shared ingredients)
3. Quinoa Bowl (3 shared ingredients)
4. Vegetable Soup (2 shared ingredients)
5. Lentil Stew (2 shared ingredients)
6. Grilled Zucchini (2 shared ingredients)
7. Couscous Salad (2 shared ingredients)
8. Roasted Veggie Wrap (1 shared ingredients)
9. Avocado Toast (1 shared ingredients)
10. Miso Soup (1 shared ingredients)

Recipe Recommender





Conclusion

Project Planning

1st week

Brainstorming

- Topic
- Dataset

2nd week

Graph

- Visualization
- Statistics

next weeks...

Recommendation feature

Distribution of tasks

References

Kaggle Dataset https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions/data?select=PP_recipes.csv

ChatGPT : https://chatgpt.com/

Wikipedia : https://en.wikipedia.org/wiki/Graph_(abstract_data_type)

Python library for visualization : https://plotly.com/python/
Python package Graph: https://networkx.org/

Food in Society (Peter Atkins, Ian Bowler): https://doi.org/10.4324/9781315824819

Thank you for your attention