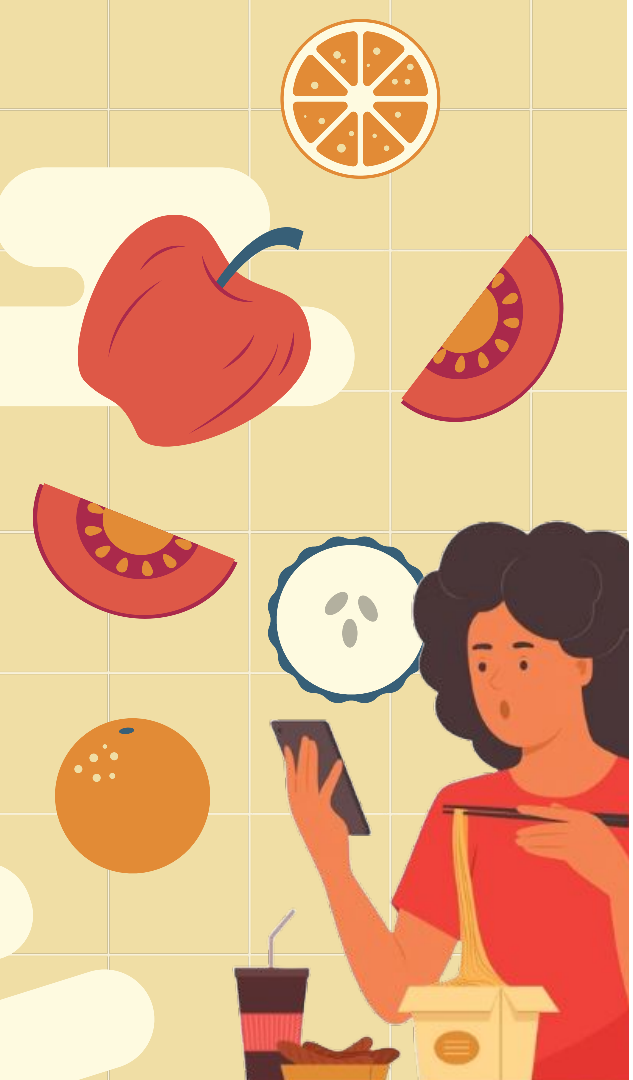


# RESTAURANT RECOMMENDATION SYSTEM

By Team 424 (Nexus Quad):
Hannah Shen, Nilesh Rakhecha,
Sai Karun Gaddam, Sravya Velamuri

# Problem Statement

- As culinary diversity continues to flourish, we find ourselves overwhelmed with an ever-expanding array of dining options. This increasing number of new restaurants poses a conundrum:
  **‘Where should one dine amidst this abundance of choices ?’**.

- Concurrently, restaurant owners are grappling with the challenge of expanding their businesses in the digital era. Beyond traditional promotions, they must now **cater to individual preferences** to remain competitive in an increasingly discerning market landscape.

- Given that we reside in AZ, we wanted to focus on: ‘**how we can help local restaurant businesses in AZ connect/serve customers who are inclined to enjoy their offerings based on their past dining history and preferences (personalization).**

# Proposed Solution

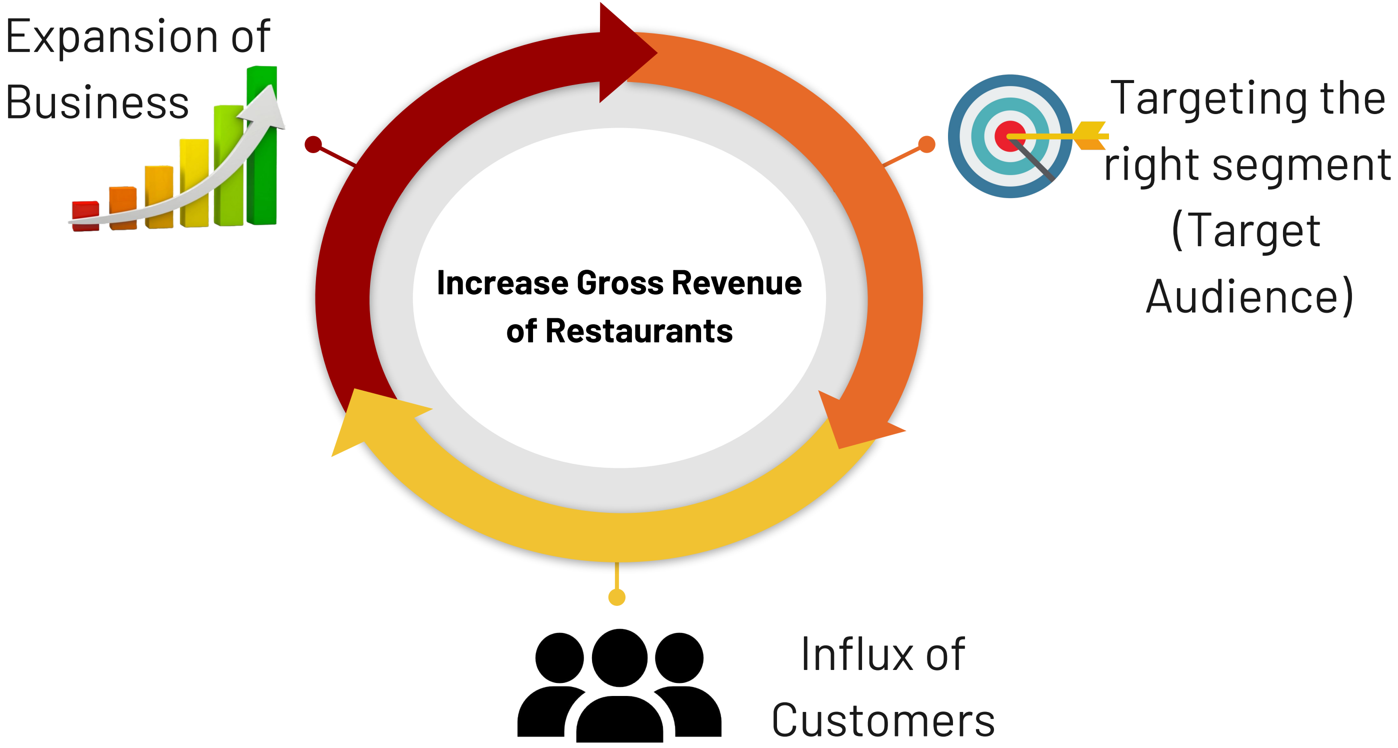
A '**Restaurant Recommendation System'** that leverages:

- Star Ratings of Restaurants
- User Reviews
- Location

to deliver **tailored** dining suggestions within a **15-mile radius of Tucson.**

Our system analyzes past reviews alongside aggregated restaurant data to construct personalized recommendations that align with the customer's culinary preferences.

# How are we serving different types of businesses?



# Project Overview

**Part 1** EDA (Exploratory Data Analysis)

**Part 2** Topic Modeling (LDA)

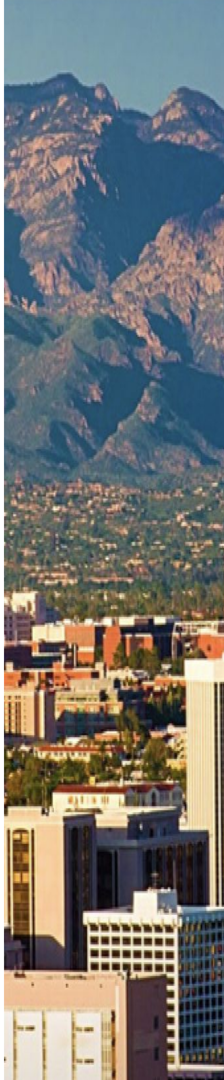
**Part 3** Recommender System (Content based)

# EDA

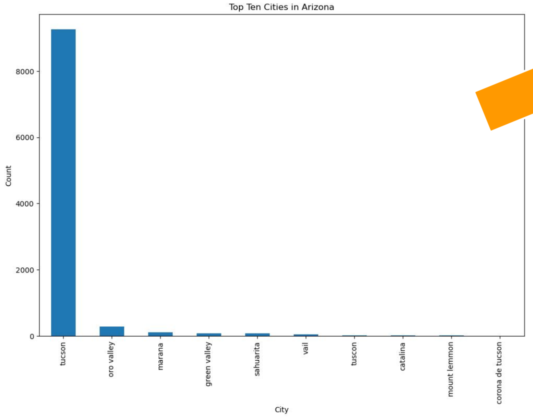
Bettering our understanding of the  data

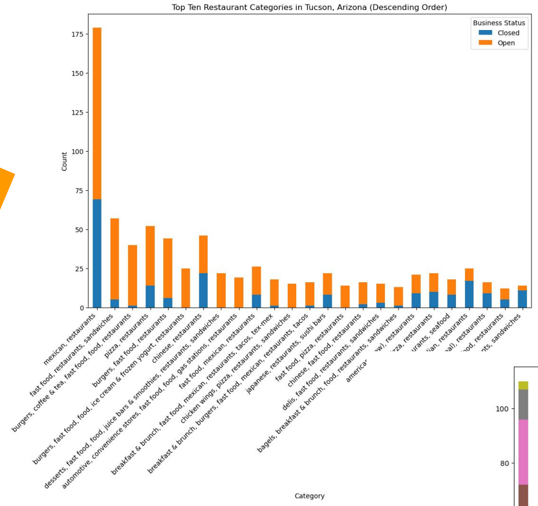
# Yelp Dataset Filtering - Choosing our Target Audience

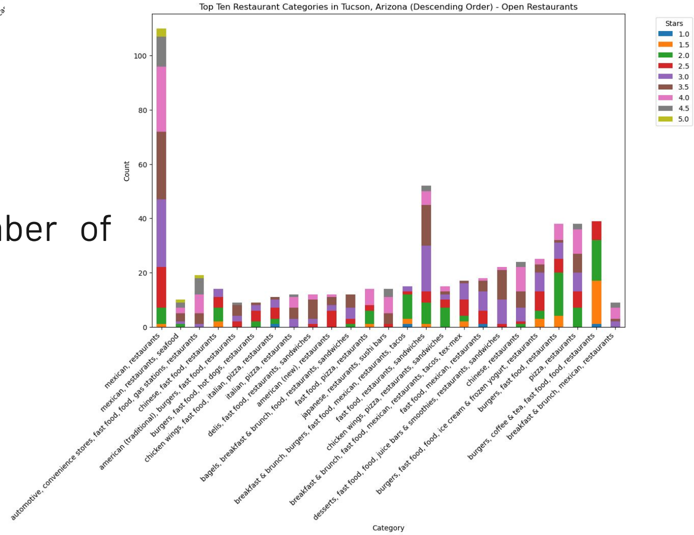
- Initial dataset:
    - 150,346 Restaurants.
    - 1,987,929 Users.
    - 6,990,280 Reviews.
- Focus Area: **Tucson** (highest no. of restaurants in AZ)
    - 1,807 Restaurants
    - 75,999 Users
    - 216,461 Reviews
- Included reviews from **2005 to 2022**.



# EDA used for Filtering



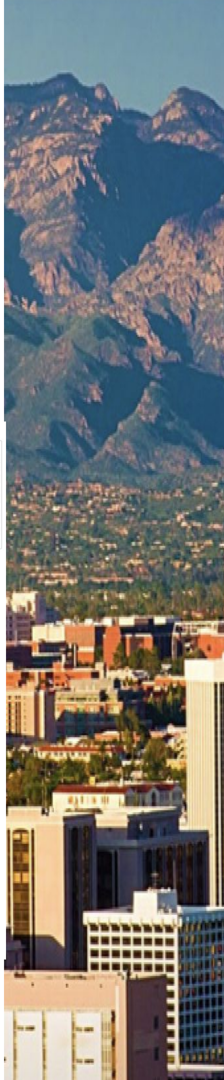



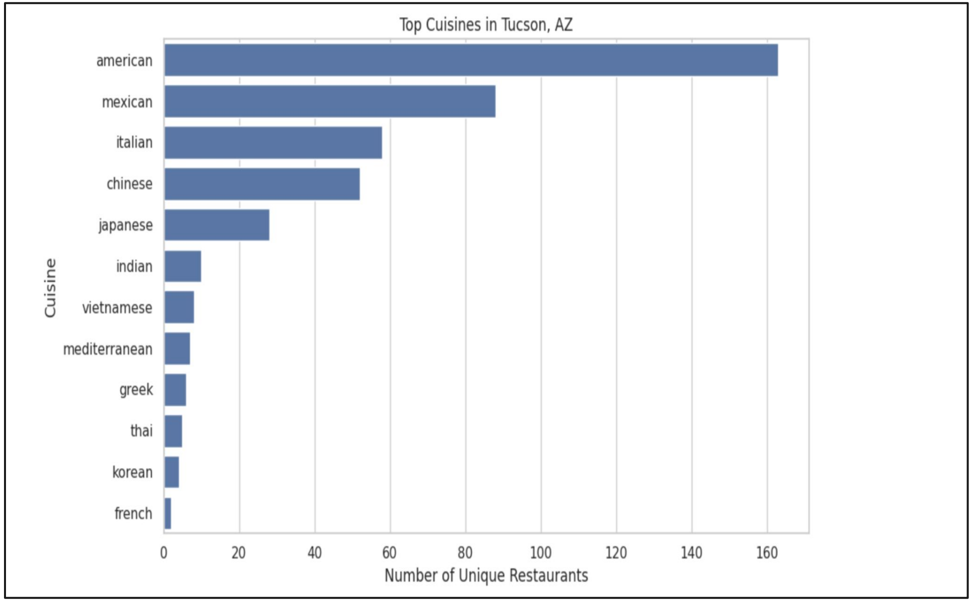

1. Did not take Oro Valley, Marana and Green valley into consideration despite having above 100 reviews is due to their location being over a 15 mile radius from Tucson.

2. Checked for number of open restaurants.

3. Considered restaurants has highly rating only (>=3)



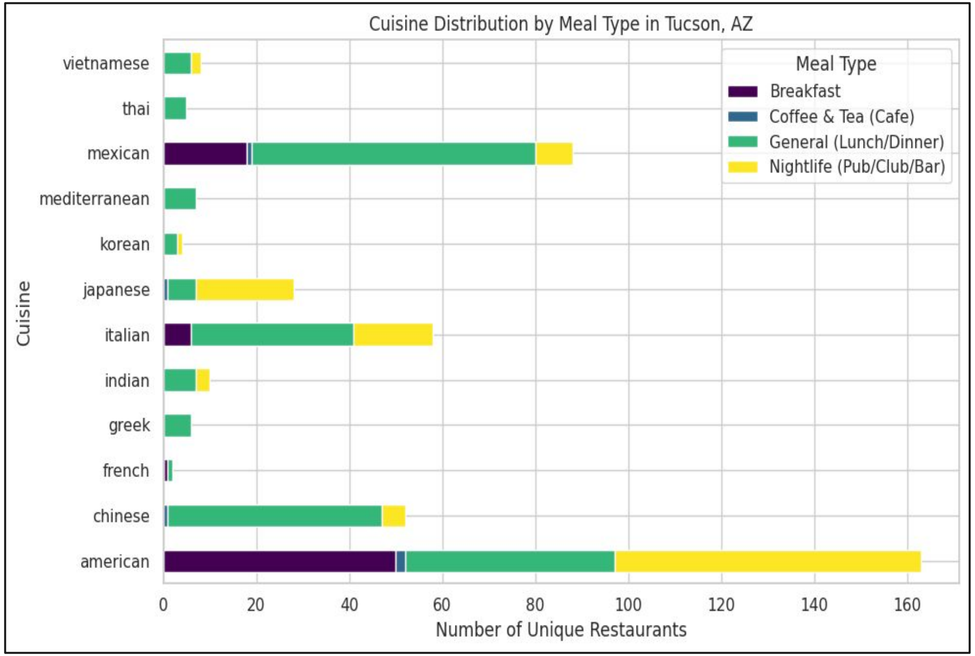
# EDA: Restaurants by Cuisine Type



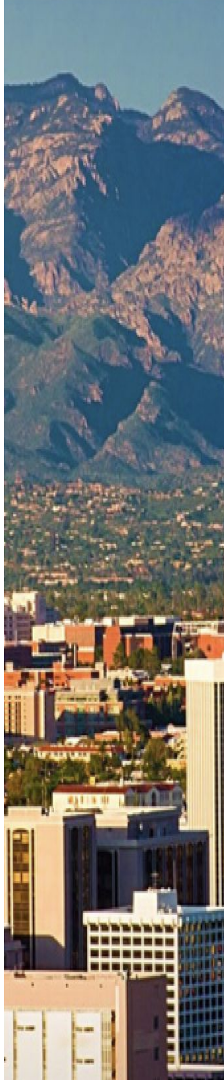

Top 5 Cuisines in Tucson:

- **American** (163 Restaurants)
- Mexican (88 Restaurants)
- Italian (58 Restaurants)
- Chinese (52 Restaurants)
- Japanese (28 Restaurants)

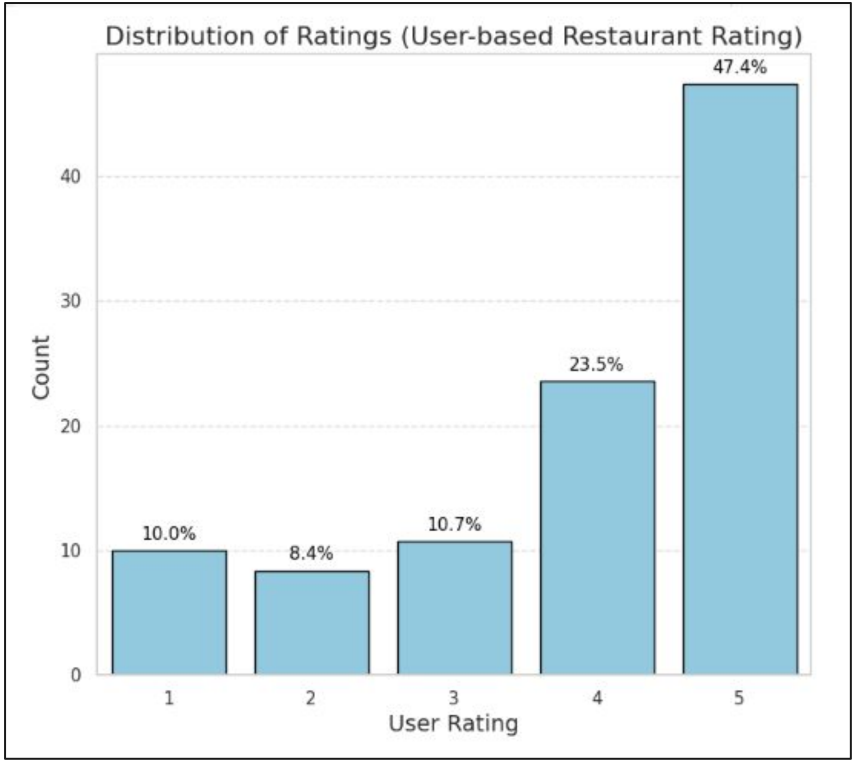
# EDA: Restaurants by Cuisine Type and Meal Type




- Each Cuisine type is further classified by Meal type
- Meal type  is classified into 4 categories:
    - Breakfast
    - Coffee & Tea
    - General
    - Nightlife



# EDA: Restaurants Reviews

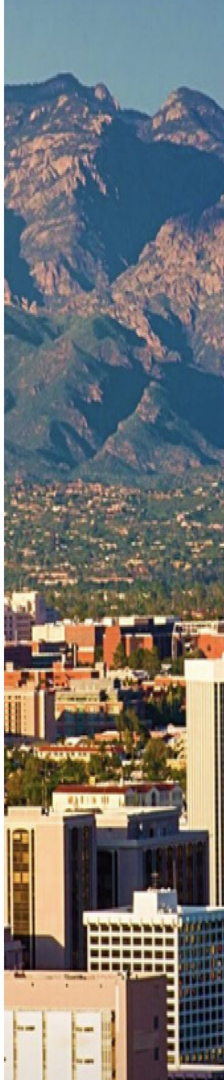



**80,172** Positive Reviews

- Prep & Pastry
- Most # of Positive Reviews (1,443)

**31, 001** Negative Reviews

- El Charro Cafe
- Most # of Negative Reviews (416)



# Topic Modeling

What Dominates Our Data?
Exploring the Most Discussed Topics

# Text Mining Techniques for Enhanced Restaurant Recommendations:

- **Removing Stop Words:** Streamlines review text analysis, ensuring focus on meaningful words for more accurate sentiment and preference detection.
- **Lemmatization:** Enhances text consistency, improving the system's ability to match user preferences with relevant restaurant features.
- **LDA:** Identifies prevalent topics in reviews, such as food quality or service, allowing for recommendations based on specific user interests.

# Text Mining Techniques for Enhanced Restaurant Recommendations:

- **Converting Words in Reviews to Lists:** Prepares text for analysis, facilitating efficient processing and extraction of insights from user reviews.
- **Generating Keywords for Reviews:** Highlights critical aspects of reviews, enabling quick identification of user preferences and restaurant attributes.
- **Identifying Dominant Topics for Each Review:** Categorizes reviews by topics, helping to tailor recommendations to align with user's dining experiences and expectations.


# Results of Topic Modeling

Word cloud representing top words in each topic

- ➢ Topic 0: Food items and flavors in cuisine.
- ➢ Topic 1: Beverage selection and nightlife atmosphere.
- ➢ Topic 2: General restaurant ambiance and service quality.
- ➢ Topic 3: Breakfast and Mexican food focused options.
- ➢ Topic 4: Dining wait times and service experience.

Useful to categorize customer reviews and menu descriptions into coherent themes. This helps in understanding customer preferences and identifying key aspects they care about, allowing for more personalized and accurate restaurant recommendations.

# Recommender System

STEPS TO HOW WE BUILT IT

# Content-Based Recommendation

As mentioned earlier, our recommender system employs a content-based approach. SO, How does it work?

- It extracts key phrases from reviews and evaluates sentiment to construct a user-specific profile.

| | Document_No | Dominant_Topic | Perc_Contribution | Keywords | Text |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 0.4317 | food, place, great, service, restaurant, love,... | [favorite, indian, restaurant, fancy, absolute... |
| 1 | 1 | 2 | 0.6044 | food, place, great, service, restaurant, love,... | [food, amazing, service, great, price, reasona... |
| 2 | 2 | 2 | 0.4774 | food, place, great, service, restaurant, love,... | [oily, weekend, buffet, well] |
| 3 | 3 | 2 | 0.3661 | food, place, great, service, restaurant, love,... | [favorite, indian, place, tucson, hand, always... |
| 4 | 4 | 0 | 0.4117 | order, sauce, taste, salad, flavor, dish, chic... | [tear, star, decide, generous, offering, place... |
| 5 | 5 | 2 | 0.5103 | food, place, great, service, restaurant, love,... | [lab, fan, food, different, part, world, glad,... |
| 6 | 6 | 2 | 0.4382 | food, place, great, service, restaurant, love,... | [service, pleasant, outstanding, mean, slightl... |
| 7 | 7 | 4 | 0.3223 | order, wait, time, table, food, ask, give, ser... | [become, fan, sher, punjab, last, couple, year... |
| 8 | 8 | 2 | 0.6372 | food, place, great, service, restaurant, love,... | [indian, die, definitely, hide, gem] |
| 9 | 9 | 2 | 0.6661 | food, place, great, service, restaurant, love,... | [indian, food, ever, tikka, masala, crazy, cut... |

# Content-Based Recommendation

- Each restaurant in the dataset is represented as a 'bag of words', where the bag contains unique words found in the restaurant's content.

- Frequency of words may influence similarity, similarity measures like cosine similarity are employed to quantify the similarity between restaurants.

| name | bag_of_words |
|---|---|
| 4 seasons | {"'taste'", "'order'", "'bar'", "'chinese'", "... |
| 47 scott | {"'comfort food'", "'taste'", "'order'", "'bar... |
| 5 points market & restaurant | {"'delis'", "'taste'", "'order'", "'salsa'", "... |
| agustin kitchen | {"'french'", "'taste'", "'order'", "'bar'", "'... |
| amigos burgers and beer | {"'taste'", "'order'", "'bar'", "'salsa'", "'l... |

# Content-Based Recommendation

- This profile is then compared against a pre-processed matrix of restaurant features, using cosine similarity to identify the best matches.

```
array([[1.        , 0.91767731, 0.89077397, ..., 0.57876617, 0.87266012,
        0.6140346 ],
       [0.91767731, 1.        , 0.86349619, ..., 0.54608172, 0.83604608,
        0.5963984 ],
       [0.89077397, 0.86349619, 1.        , ..., 0.63011647, 0.82517941,
        0.54696603],
       ...,
       [0.57876617, 0.54608172, 0.63011647, ..., 1.        , 0.53571547,
        0.72063664],
       [0.87266012, 0.83604608, 0.82517941, ..., 0.53571547, 1.        ,
        0.60179348],
       [0.6140346 , 0.5963984 , 0.54696603, ..., 0.72063664, 0.60179348,
        1.        ]])
```

# Content-Based Recommendation

- Recommendations are generated by identifying restaurants with the highest similarity scores to the user's preferences, ensuring a personalized dining suggestion that closely aligns with their unique taste and previous positive dining experiences.

```
Enter the name of a restaurant you've visited: Casa Valencia

Recommendations based on similarity:
1. la hacienda restaurant
2. taco giro mexican grill - w valencia
3. goyita's
4. el cisne restaurant
5. villa peru
```

# Content-Based Recommendation (Screen Grab)

```
Enter the name of a restaurant you've visited: [          ]
```

```
Enter the name of a restaurant you've visited: 4 Seasons

Recommendations based on similarity:
1. sawmill run restaurant
2. pastiche modern eatery
3. harvest oro valley
4. bob dobbs
5. signature grill


Enter the name of a restaurant you've visited: Casa Valencia

Recommendations based on similarity:
1. la hacienda restaurant
2. taco giro mexican grill – w valencia
3. goyita's
4. el cisne restaurant
5. villa peru


Enter the name of a restaurant you've visited: Taco Bell


Restaurant 'taco bell' not found.


No recommendations found.

Enter the name of a restaurant you've visited: quit


Bye!
```

# Conclusion

Limitations and What's Next?

# Limitations

- The model fails to provide accurate recommendations if a user misspells the restaurant's name.
- The '**cuisine preferences**' of users is overlooked when a a  restaurant recommendation is made.
- The '**user's budget'** is overlooked as well when a restaurant recommendation is made.
- While we can offer suggestions to active, existing customers, we encounter limitations in providing recommendations to new or inactive users.

# What's Next?

- **Improving User-centricity:** Users will be prompted to provide their unique ID, allowing retrieval of their last visited restaurant. Leveraging this data, our recommender system suggests suitable dining spots. Given that users may not review every visited restaurant, we verify if the pulled review corresponds to their last visit. If not, users are prompted to input their latest dining experience.

- **Personalizing Selection:** Users can specify their preferred cuisine, enabling retrieval of restaurant(s) within the chosen category. Our recommender system then suggests suitable dining spots. In cases where users haven't visited any restaurant within the selected cuisine category, we employ collaborative filtering. By comparing their past reviews to those of users with similar reviewing patterns, we recommend a spot accordingly.

# What's Next?

- **Location-Based Recommendations for New Users**: For users with no prior dining history, our system employs location-based recommendations to suggest nearby highly rated restaurants.

- **Enhancing the Model:** Using BERT for topic modeling to provide a more refined suggestion that align closely with users' tastes and preferences.

- **Superscore:** Currently we rely solely on the base rating as our Superscore, however, we aim to enhance our scoring system by integrating sentiment scoring and polarity scoring to provide a more comprehensive evaluation of our available reviews.

By leveraging our recommendation system, businesses gain a strategic edge by offering tailored suggestions to customers. This not only enhances competitiveness in the market but also fosters growth and expansion opportunities.

By catering to the specific wants and desires of customers, businesses can effectively meet market demands and drive success.