Foodies' University Ranking
Project Report

Group Member
He Gao (hg296)
Meiqi Yang (my353)
Xi Wang (xw347)

I. Working Split

- **Collective Work - Information Collection and Database Relational Design**

  We all make efforts to collect the geographic information of 67 universities in 10 areas, design the BCNF tables including selecting and reforming attributes in the existing data files from Yelp and design the formula to calculate the ranking score.

- **Xi Wang**

  Responsible for pre-processing the university data and data from Yelp dataset https://www.yelp.com/dataset_challenge to generate the following BCNF tables: review, user, user_votes, business, business_attributes, business_category, area_state, state_country, university, university_business_distance tables.

- **He Gao**

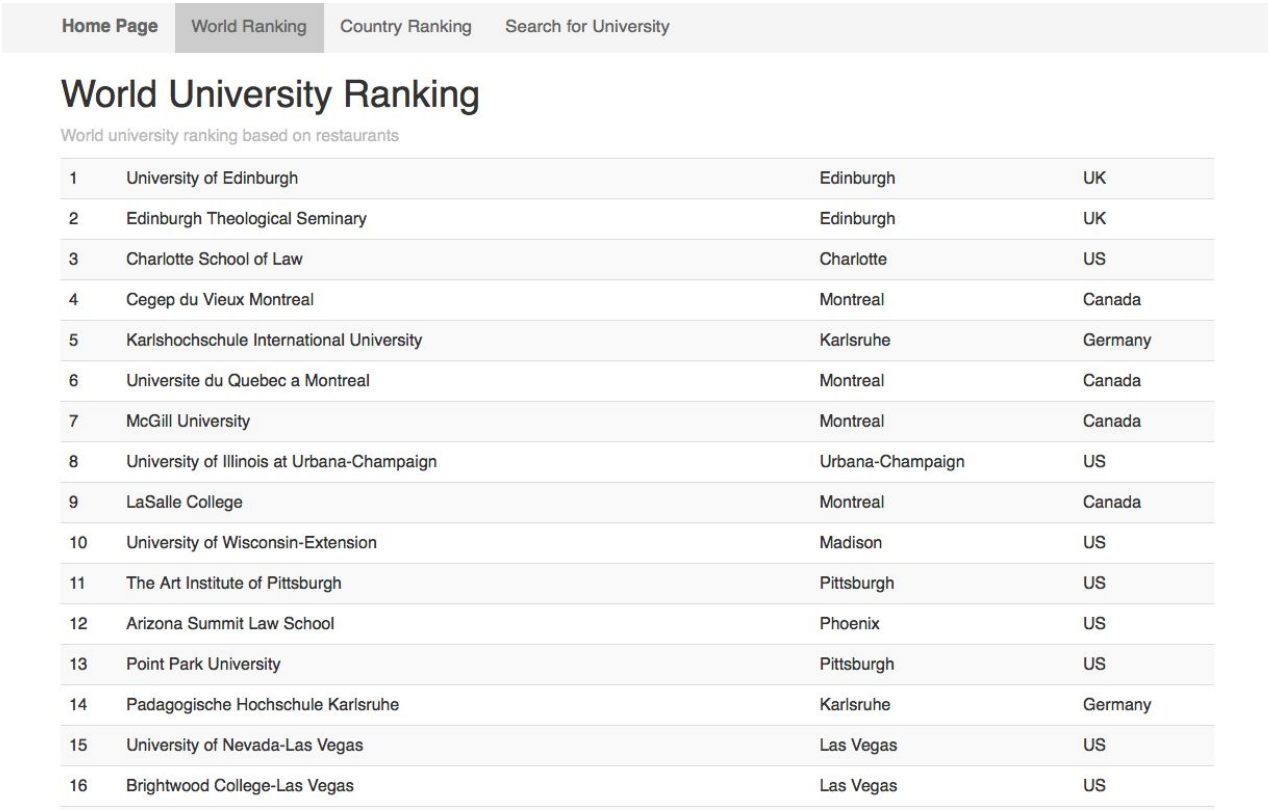
  Responsible for web application structural design (html & js building blocks), backend implementation (php), and SQL integration and writing dynamic SQL queries. Designed the ranking mechanism for restaurants near a specific university inside SQL queries.

- **Meiqi Yang**

  Responsible for UI design, web page implementation and deployment. Also responsible for loading data to database, building appropriate indexes to accelerate queries and writing SQL queries to provide data to the application.

II. Functionality Highlights

- **World university ranking based on great restaurants nearby**
    - **Logic:**
        - Scores are calculated based on number of nearby restaurants (<10 km), proximity, restaurant rating (stars) and popularity (# reviews)
            - Score = 70% norm(sum(star/dist)/numstore) + 30% norm(numStore)
        - Calculate a score for each university in DB
        - Rank all universities based on scores
    - **Example:**



Home Page | World Ranking | Country Ranking | Search for University

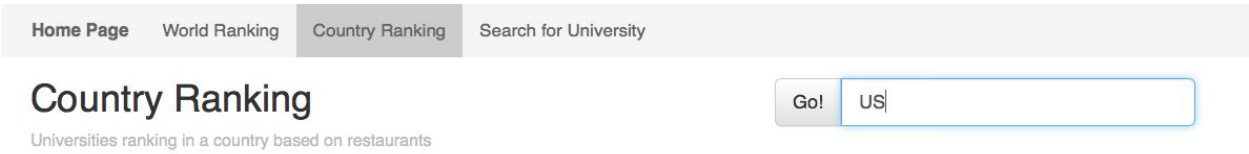
World University Ranking

World university ranking based on restaurants

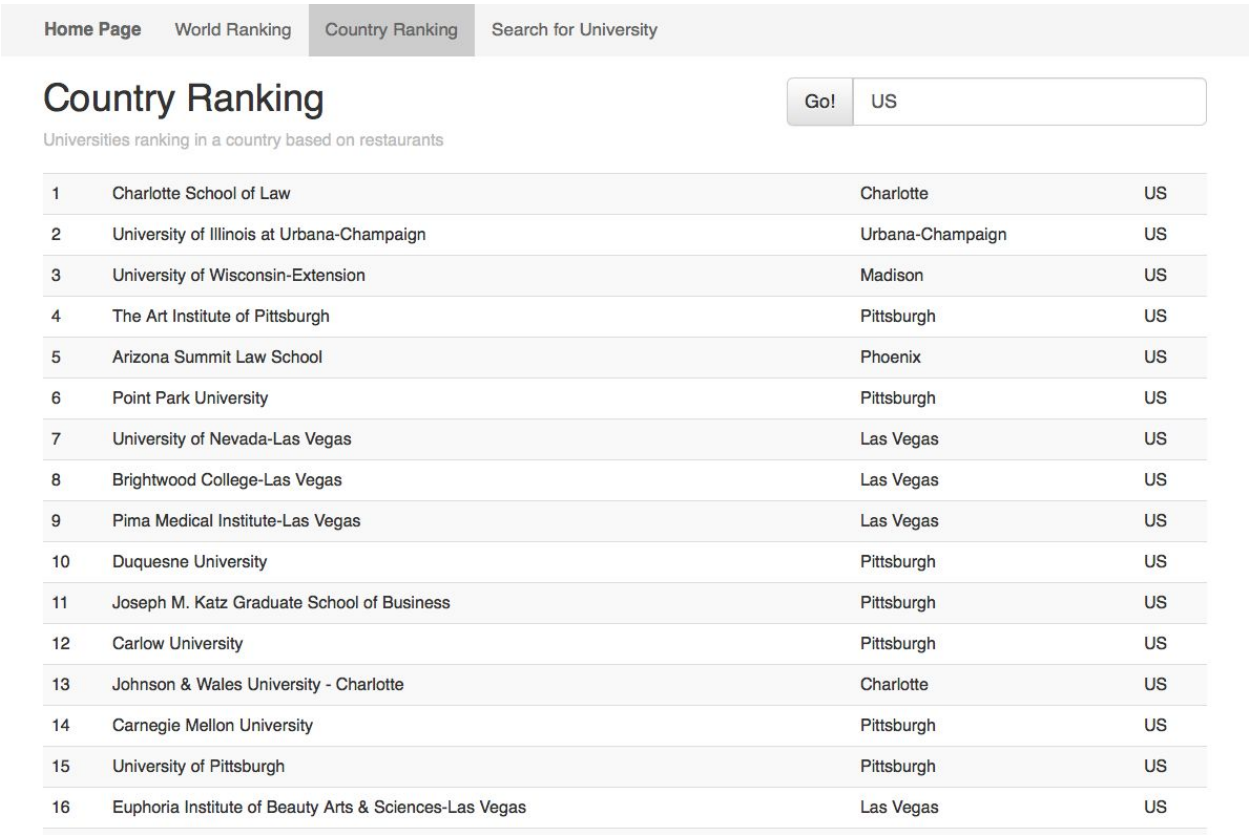
1	University of Edinburgh	Edinburgh	UK
2	Edinburgh Theological Seminary	Edinburgh	UK
3	Charlotte School of Law	Charlotte	US
4	Cegep du Vieux Montreal	Montreal	Canada
5	Karlshochschule International University	Karlsruhe	Germany
6	Universite du Quebec a Montreal	Montreal	Canada
7	McGill University	Montreal	Canada
8	University of Illinois at Urbana-Champaign	Urbana-Champaign	US
9	LaSalle College	Montreal	Canada
10	University of Wisconsin-Extension	Madison	US
11	The Art Institute of Pittsburgh	Pittsburgh	US
12	Arizona Summit Law School	Phoenix	US
13	Point Park University	Pittsburgh	US
14	Padagogische Hochschule Karlsruhe	Karlsruhe	Germany
15	University of Nevada-Las Vegas	Las Vegas	US
16	Brightwood College-Las Vegas	Las Vegas	US

- **Country-specific university ranking**
    - **Logic:**
        - Similar logic to world ranking
        - Users can search for a specific country and get university ranking in that country
    - **Example:**

- User first input a country name

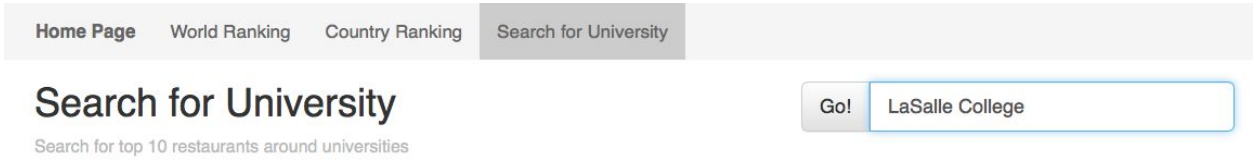


- Then user can confirm and view university country ranking

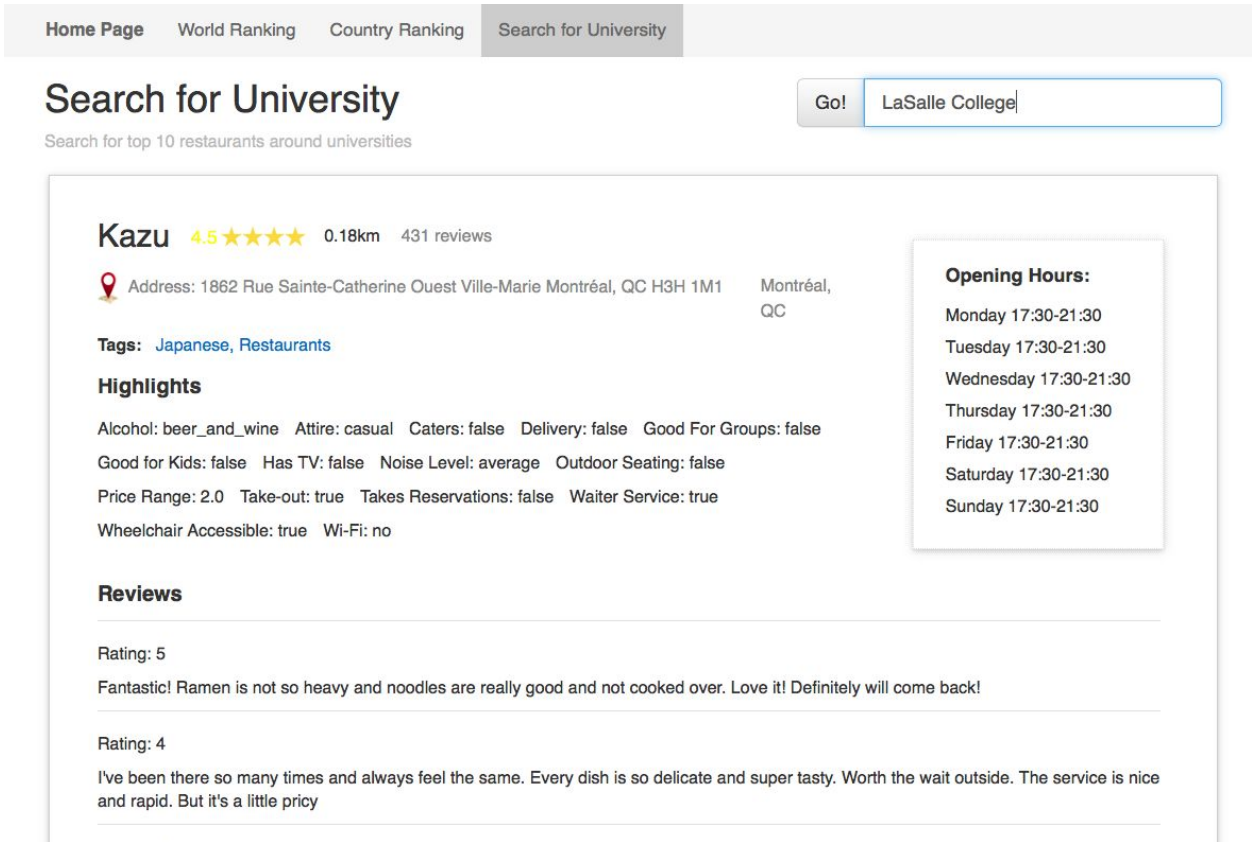


1	Charlotte School of Law	Charlotte	US
2	University of Illinois at Urbana-Champaign	Urbana-Champaign	US
3	University of Wisconsin-Extension	Madison	US
4	The Art Institute of Pittsburgh	Pittsburgh	US
5	Arizona Summit Law School	Phoenix	US
6	Point Park University	Pittsburgh	US
7	University of Nevada-Las Vegas	Las Vegas	US
8	Brightwood College-Las Vegas	Las Vegas	US
9	Pima Medical Institute-Las Vegas	Las Vegas	US
10	Duquesne University	Pittsburgh	US
11	Joseph M. Katz Graduate School of Business	Pittsburgh	US
12	Carlow University	Pittsburgh	US
13	Johnson & Wales University - Charlotte	Charlotte	US
14	Carnegie Mellon University	Pittsburgh	US
15	University of Pittsburgh	Pittsburgh	US
16	Euphoria Institute of Beauty Arts & Sciences-Las Vegas	Las Vegas	US

- **Recommendation of the best restaurants near a specific university**
    - **Logic:**
        - User can query a university
        - Scores for restaurants near the university are calculated dynamically based on distance, ratings, and popularity.
            - Score = rating * log2 (popularity)/log10 (distance)
        - Top 10 restaurants are recommended with all necessary attributes
    - **Example:**
        - User first input a university name



- Then user can see the best restaurants near the university, along with all information related to the restaurants



III. Limitations

- **Limited Data**

  In this exercise, our data mainly comes from Yelp datasets, therefore several limitations on data exist:
  - Yelp datasets only provided restaurant data of ten regions from 4 countries
  - Only universities from those covered regions are being ranked (67 universities)
  - Yelp restaurant info might not be comprehensive enough to evaluate restaurants in all regions discussed (e.g. there might be a strong bias towards restaurants in

USA)

- **Limited Restaurant Info**

  Our restaurant recommendations do not support advanced interactive functionalities such as checking reviewers info, showing restaurant location on the map, etc

- **Limited accuracy / fitness of ranking algorithms**

  Our score ranking mechanisms might be simplistic in more complex settings. Further tests and more data might be needed to make the ranking mechanism closer to optimal.

- Despite the limitations in our system, we still believe it works well in current scale and provides a reasonable ranking / recommendation feature.

IV. Challenges Conquered

- **Table design**

  - The available yelp dataset contains 6 files (business, review, user, check-in, tip, photos) in JSON format. There are a lot of redundant information and nested attributes, thus we need to get rid of unrelated information to our projects, and also flatten nested attributes to make it fit into database.

  - We want to make sure that every table in our database follows BCNF, however, some data in the database turned out not as we expected. For example, we previously assumed that each city belongs one state. However, we found that in the yelp dataset, city such as Edinburgh belongs to two states, EDH, FIF. So in the area_state(area, state) table, instead of making area the primary key, we assign the primary key to the combination of (area, state) to resolve this problem.

- **Data processing:**

  - we need to figure out and re-organize the important information from the massive Yelp dataset. For example, the “attributes” in business.json file from Yelp is a composite attribute, the format is like {"Happy Hour": true, "Accepts Credit Cards": true, "Good For Groups": true, "Outdoor Seating": false, "Price Range": 1}. However, we need to convert the complex attributes in a relational table friendly way in the form of business_id - attribute - value. One business can have multiple attributes. Another challenge is that we need to generate derived attributes such as distance between university and business. For example, we only have longitude and latitude attribute of each university and business and

have to use it to calculate the distance. In order to speed up the computation, we first select the business based on the state of the university (some universities are in the area which is in the middle of two states, we take both states into account.) Then we compute the distance by the following Haversine formula:

$$a = \sin^2(\Delta\varphi/2) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\Delta\lambda/2)$$
$$c = 2 \cdot \operatorname{atan2}(\sqrt{a}, \sqrt{(1-a)})$$
$$d = R \cdot c$$

where φ is latitude, λ is longitude, R is earth's radius (mean radius = 6,371km)

- **Ranking Mechanism Design**
    - Although main factors affecting ranking are pretty clear (distance, rating, popularity, etc), it is quite tricky to determine how important those factors are and the boundary point of those factors coming into play.
    - Therefore, we had to iteratively test and improve our ranking formula to arrive at a reasonable ranking / recommendation solution.

V. Main Contributions

- There is no other existing solution of university ranking based on the quality of restaurants in the neighborhood.
- Academic ranking such as US News and QS ranking is boring.
- Our system offers a new standard and a different perspective for students and teachers to evaluate universities. Our system can benefit users including
    - Students who plan to apply for universities and want to get the information about the surroundings of the universities.
    - Students/professors who have future research / exchange trips in those areas.
    - New-coming students/professors who try to get familiar with their campus.

VI. Selling Points

- Novel university ranking benefits a wide range of target users.
- Our database design follows BCNF.
- Professional web interface.
- Easy-to-use query mechanism.
- Logically useful and comprehensive information display.
- Fast response from the website.