Yelp Business Data Analysis

COP 5725

Spring 2019
Instructor: Dr. Markus Schneider

Group 24

Group Members:

Name:	UFID:	Email:
● Dashuai Qin	91493498	dqin@ufl.edu
● Shubham Kathuria	33939680	skathuria@ufl.edu
● Vikas Pathak	71984931	pathak.vikas@ufl.edu
● Pranav Mokal	68121781	mokalpranav@ufl.edu

INTRODUCTION:

The yelp business dataset has detailed information such as location, category, ratings, reviews, price, hours, and other attributes about various businesses. Using this database, we can recommend various businesses such as restaurants to users depending on their needs, i.e. restaurants good for late night dining, restaurants good for the family dinners, or even type of restaurants that be a good for starting as a business etc. Users can also use the data to do trip planning since we have queries like the total number of different category of restaurants and an average rating of restaurants. Even though our focus is mainly on the restaurants, we will also provide the same features to users in terms of hotels, shopping businesses etc.

MOTIVATION:

Whenever we visit a new place or go on a trip, we are unaware of all the available businesses or facilities such as restaurants, hotels, doctors or even shopping places that are around us. Even if we are able find such places around us, it is difficult to differentiate the good ones from the rest. Not only while visiting new places but we are sometimes unaware of such things in the city where we reside as well. Also, there are instances where because of one bad choice of a restaurant or a hotel, the entire trip gets spoiled
It is for this reason that we intend to use this database and ensure that any user gets the best possible experience while visiting any new place in terms of food, accommodations, shopping etc. The yelp database does not only provide reviews but also make sure that they are authentic thereby providing accurate results. This will help the users save time and the disappointment of trying out a new place and then come out dissatisfied.

EXPECTED USER FUNCTIONALITIES:

<u>User Types:</u>
All the data and the various functionalities will be available to all the users since there will only be a Single User Type.

- The user will be able to view all the data related to particular business

<u>Web/User Interface Functionality:</u>
Some of the possible functionalities that can be available on the Web Interface are:

- Provide a list of suggestion of businesses similar to the they are viewing.
- Provide all related data pertaining to a particular like type, ratings, hours, price-range and all the reviews etc.
- Visual representation such as charts to display various trends such as ratings, changes in pricings, number of locations etc. over the years.
- Provides photos of the business if they are available.

POSSIBLE QUERIES:

- Find out the average restaurant rating of each city(review_count > 50)
- Total number different category of restaurants
- Business purpose restaurants(RestaurantsPriceRange > 4, stars > 4, Business Parking: true, Noise Level: Low, review_count> 100)
- Family purpose Restaurants (4 > RestaurantsPriceRange> 1, stars > 3, GoodforKids: true, review_count > 100)
- Budget Restaurants(RestaurantsPriceRange = 1)
- Potential hidden gem restaurants(stars = 5, review_count < 100)
- Best Pizza places to eat in a particular city(stars >=3)
- Places to eat at and are open within a particular radius(say radius < 2 miles)
- Best Chinese restaurants to eat near me with parking available.
- Places that serve food as well as alcohol in a particular city
- Cheap and Decent hotels to stay at in a particular city (Price-Range <= 3 and Stars >= 3)
- Provide charts(if possible) of a possible study
    - Where the ratings of a particular type of restaurant has gone below in a particular city (For example: trying to find out if Pizzas in a particular city are becoming bad by comparing ratings from different years or months)
- Provide the list of businesses that will probably go under(be closed) in the next 1 or 2 years
- Restaurants with live music
- Search restaurants according to Cuisine
- Total number of restaurants that make deliveries in each city
- Category that has highest average rating in each city(What is popular in the city)
- Restaurants good for going on dates or romantic outings(RestaurantsPriceRange > 3, stars > 3, Noise Level: Average/Low, review_count > 100, OutdoorSeating: True)
- Late night snack restaurants(hours > 22:00)

DATA MANAGEMENT NEEDS:

Our database will include various data types such as integers, floats, strings, dates etc. It may also contain images. The queries will be executed on the database consisting of more the 150,000 tuple(probably more than 180,000). The data will be read and fetched from .json files.

1. Dataset Information:

We will use the database from either one or both of the following sources:

1. https://www.kaggle.com/yelp-dataset/yelp-dataset/home
2. https://www.yelp.com/dataset

The size of dataset contains over 150000 records and each record has more than 15 attributes.

2. Software Requirements:

- **Database**
  We will use CISE Oracle for the database and might use Kibana to help visualize data.
- **Interface**
  The potential languages we may use for our front end and back end services are HTML, CSS, Javascript, Python, and/or C++.

3. Database Security:

- There is no security needed for users to view the data available.
- Users need to login if they have post reviews or give ratings to a business.

References:

3. https://www.kaggle.com/yelp-dataset/yelp-dataset/home
4. https://www.yelp.com/dataset