Analysis of Yelp Dataset

Group 5

Harshraj Mahesh
Uddesh Karda
Sharath Chandra
Sowrabh Chandraa

Dataset

Yelp Open Dataset

An all-purpose dataset for learning

The Yelp dataset is a subset of our businesses, reviews, and user data for use in personal, educational, and academic purposes. Available as JSON files, use it to teach students about databases, to learn NLP, or for sample production data while you learn how to make mobile apps.

The Dataset

6,685,900 reviews

192,609 businesses

200,000 pictures

10 metropolitan areas

1,223,094 tips by 1,637,138 users
Over 1.2 million business attributes like hours, parking, availability, and ambience
Aggregated check-ins over time for each of the 192,609 businesses

- **business.json**

- review.json

- user.json

- tip.json

```json
{
    "business_id": "tnhfDv5Il8EaGSXZGiuQGg",
    "name": "Garaje",
    "address": "475 3rd St",
    "city": "San Francisco",
    "state": "CA",
    "postal code": "94107",
    "latitude": 37.7817529521,
    "longitude": -122.39612197,
    "stars": 4.5,
    "review_count": 1198,
    "is_open": 1,
    "attributes": {
        "RestaurantsTakeOut": true,
        "BusinessParking": {
            "garage": false,
            "street": true,
            "validated": false,
            "lot": false,
            "valet": false
        },
    },
    "categories": [
        "Mexican",
        "Burgers",
        "Gastropubs"
    ],
    "hours": {
        "Monday": "10:00-21:00",
        "Tuesday": "10:00-21:00",
        "Friday": "10:00-21:00",
        "Wednesday": "10:00-21:00",
        "Thursday": "10:00-21:00",
        "Sunday": "11:00-18:00",
        "Saturday": "10:00-21:00"
```

- business.json

- review.json

- user.json

- tip.json

```
{
    "review_id": "zdSx_SD6obEhz9VrW9uAWA",
    "user_id": "Ha3iJu77CxlrFm-vQRs_8g",
    "business_id": "tnhfDv5Il8EaGSXZGiuQGg",
    "stars": 4,
    "date": "2016-03-09",
    "text": "Great place to hang out after work: the prices are
      decent, and the ambience is fun. It's a bit loud, but very
      lively. The staff is friendly, and the food is good. They
      have a good selection of drinks.",
    "useful": 0,
    "funny": 0,
    "cool": 0
}
```
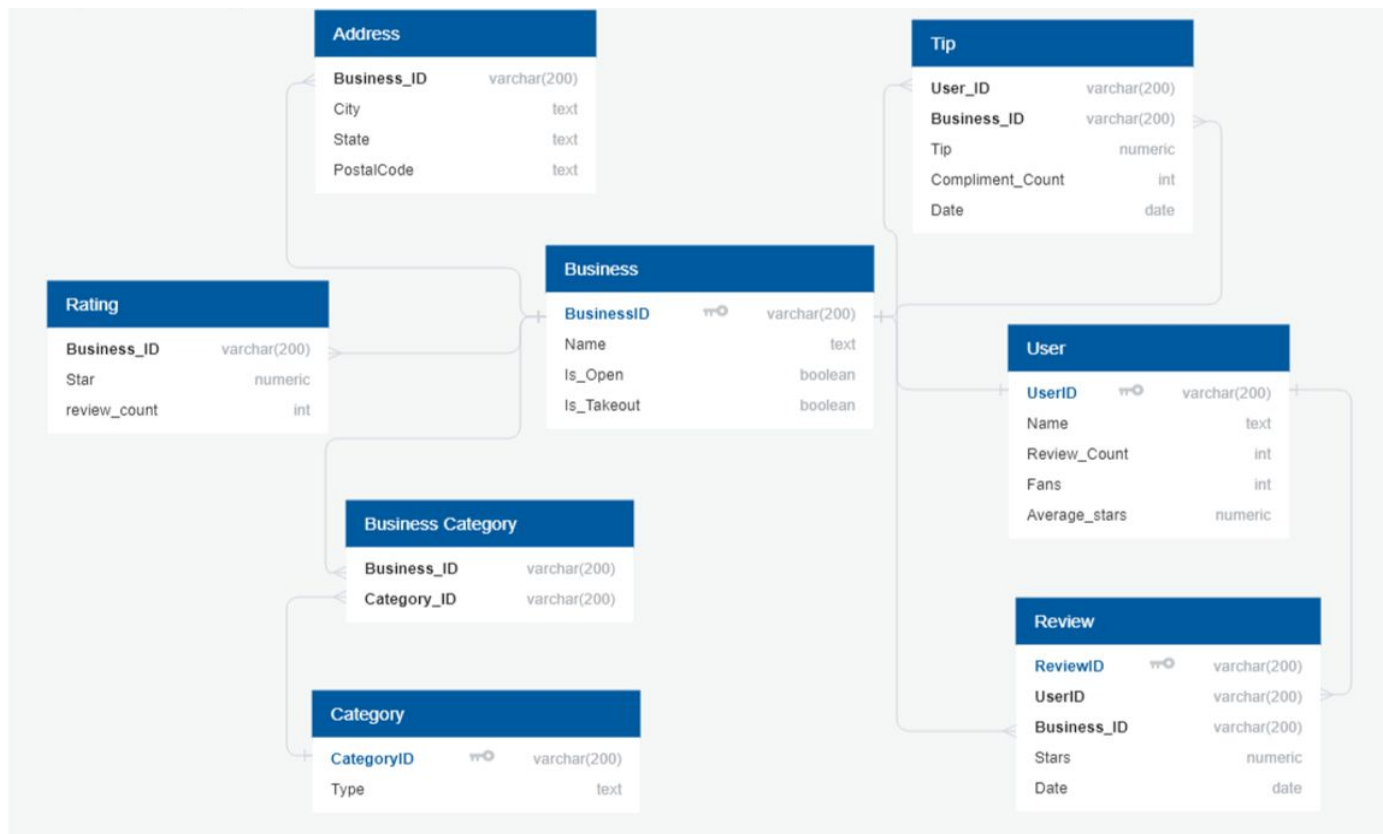
- business.json

- review.json

- user.json

- tip.json

```
{
    "user_id": "Ha3iJu77CxlrFm-vQRs_8g",
    "name": "Sebastien",
    "review_count": 56,
    "yelping_since": "2011-01-01",
    "friends": [
        "wqoXYLWmpkEH0YvTmHBsJQ",
        "KUXLLiJGrjtSsapmxmpvTA",
        "6e9rJKQC3n0RSKyHLViL-Q"
    ],
    "useful": 21,
    "funny": 88,
    "cool": 15,
    "fans": 1032,
    "elite": [
        2012,
        2013
    ],
    "average_stars": 4.31,
    "compliment_hot": 339,
    "compliment_more": 668,
    "compliment_profile": 42,
    "compliment_cute": 62,
    "compliment_list": 37,
    "compliment_note": 356,
    "compliment_plain": 68,
    "compliment_cool": 91,
    "compliment_funny": 99,
    "compliment_writer": 95,
    "compliment_photos": 50
}
```

- business.json

- review.json

- user.json

- tip.json

```
{
    "text": "Secret menu - fried chicken sando is da bombbbbb
      Their zapatos are good too.",
    "date": "2013-09-20",
    "compliment_count": 172,
    "business_id": "tnhfDv5Il8EaGSXZGiuQGg",
    "user_id": "49JhAJh8vSQ-vM4Aourl0g"
}
```

ER Diagram

Functional Dependencies

Business table   business_id -> name,is_open ,is_takeout

Review table   review_id -> user_id, business_id ,stars

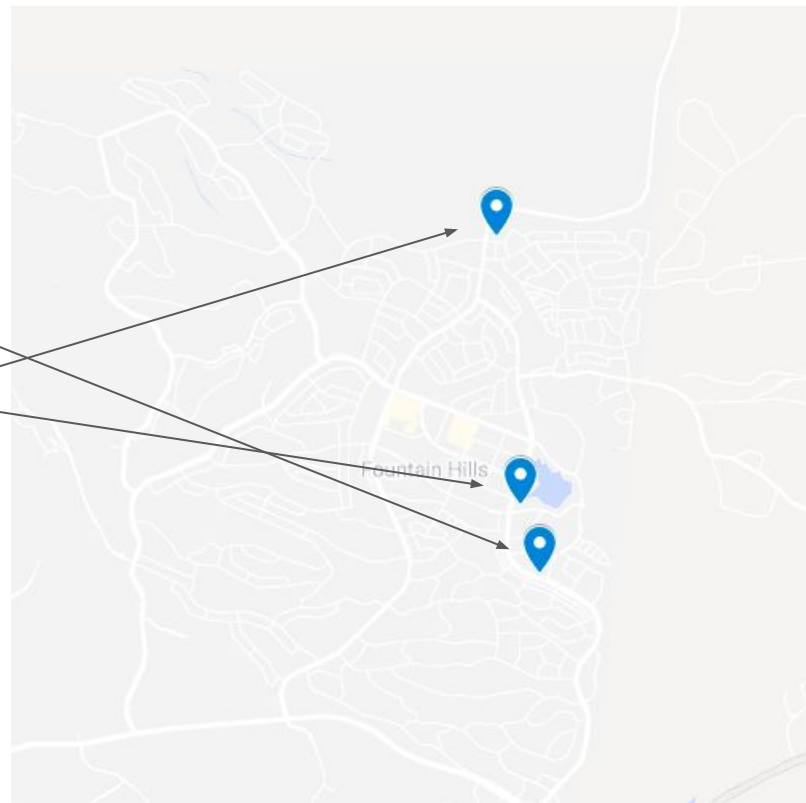**Tip table**   user_id,business_id -> tip,complement_count, tip_date

User Table   user_id -> name,fans,average_stars

Interesting Queries

Restaurants by state that have rating > 4 and with a review count > 100

Chodang	Chandler	AZ
Otaku	Chandler	AZ
Chon Thai Food	Chandler	AZ
Pesto's	Chandler	AZ
Grimaldi's Pizzeria	Chandler	AZ
Barrelhouse American Kitchen...	Chandler	AZ
Sapori D'Italia	Fountain Hills	AZ
Maskadores Taco Shop	Fountain Hills	AZ
Arrivederci Cucina Italiana	Fountain Hills	AZ
Curry Garden	Gilbert	AZ
Not Your Typical Deli	Gilbert	AZ
Four Seasons Sushi Bar & Grill	Gilbert	AZ
Waffle Love - Gilbert	Gilbert	AZ
Guedo's Taco Shop	Gilbert	AZ
Arizona BBQ Company	Gilbert	AZ
The Keg Steakhouse + Bar	Gilbert	AZ

5 Star Pizza places in Las Vegas

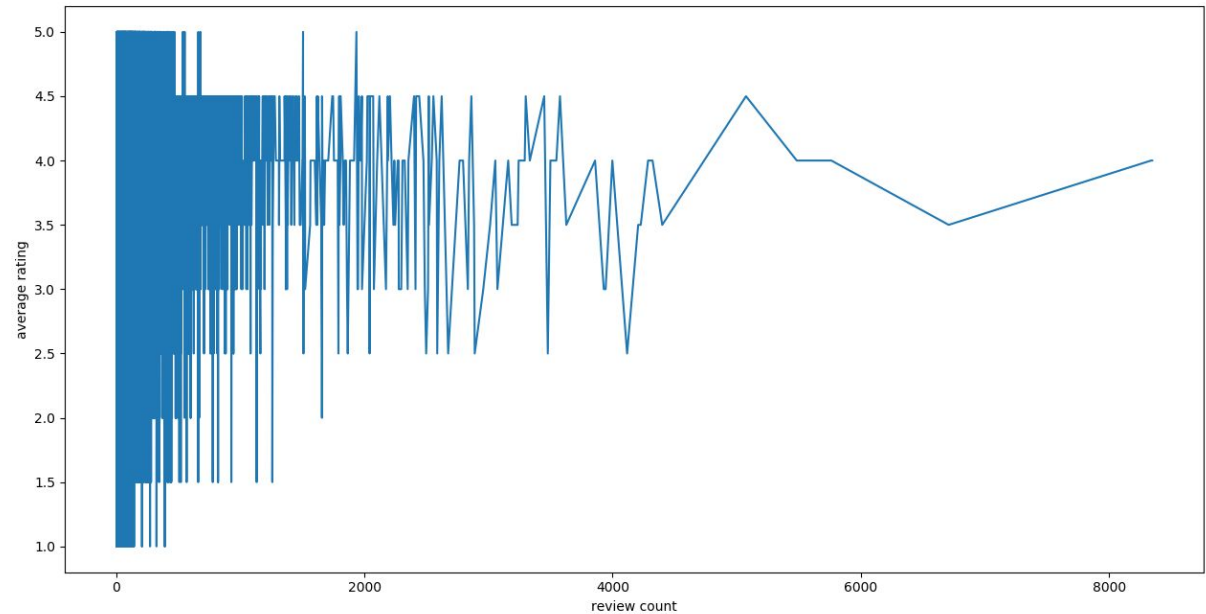# Top indian restaurants in toronto based on ratings and reviews

Best Beer Spots in Each City

city	array_agg
Akron	{"Lizardville Beer Store & Whiskey Bar","Havana House"}
Aurora	{LCBO,Lcbo}
Beachwood	{"The Pub Beachwood"}
Bethel Park	{"Colonial Beer"}
Braddock	{"Brew Gentlemen"}
Calgary	{Co-op,"Safeway Wine & Spirits","Co-op Liquor","Bin 905 Wine & Spirits","5 Vines Wine, Craft Beer…
Canonsburg	{"Guntown Beer"}
Carefree	{"Brix Wine"}
Carnegie	{"Fine Wine & Good Spirits"}
Chandler	{BevMo!}
Charlotte	{"Reid's Fine Food","Noda Bodega","Lincoln's Haberdashery","Davidson Street Public House","The…
Chesterland	{"Colonial Wine and Beverage"}
Cheswick	{"Allegheny 6 Pack & Doghouse"}
Clairton	{"Clairton Beer Distributors"}
Cleveland	{"Ohio City Brew Stop","Coquette Patisserie","21 Lounge & Kitchen","Crazy Horse","McNultys Bier…
Concord	{"Cabo Winery"}
Cornelius	{"Bottled & Tapped"}
Cottage Grove	{"Village Cafe"}
East Mc Keesport	{"R-Time At Rene's"}
Export	{"Rooz Brews"}
Fairport Harbor	{"The Pompadour"}

73 row(s) affected, took 62.1ms
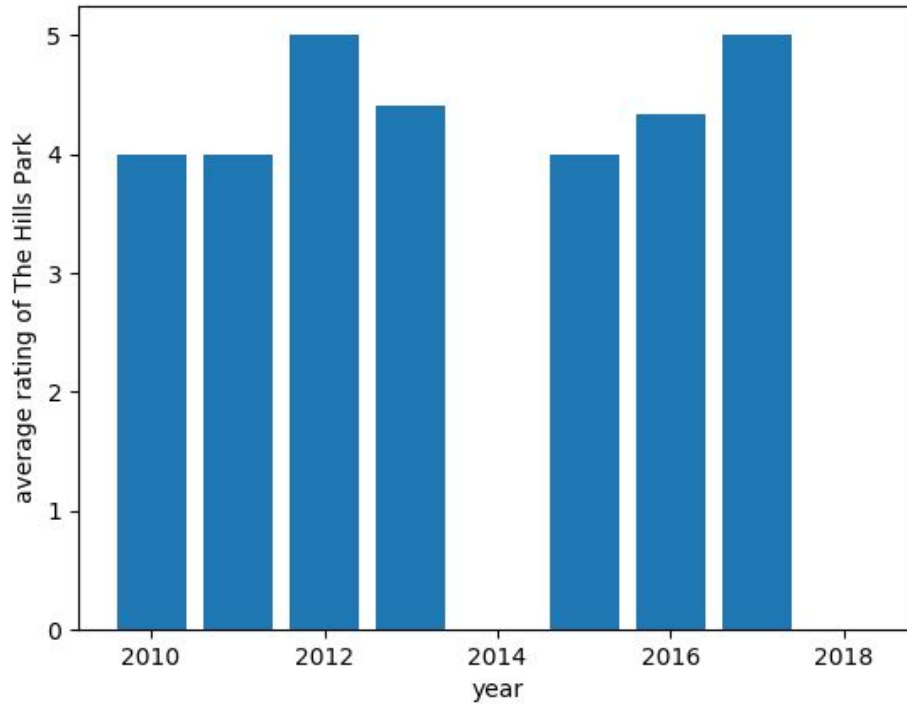
Top 3 police departments belonging to different states

Review count v/s average rating for business

Top 5 Businesses

Average rating of "The Hills Park, Nevada" by year

Data Integration

- **Business_tips**

- Number_of_restaurants

- Business_name_category

name	tip
1st Pet Veterinary Centers	1st Pet was very kind in regards to helping with my b...
346 Patisserie	Just found out on their Instagram (@346patisserie) t...
A & K Chinese Restaurant	If XLB (Xiao Long Bao / Shanghai soup dumplings) $...
A Peaceful Farewell	Ending your pet's life is devastating enough, it's nice...
Aldi	Open during construction until May 25! Grand reope...
Aloha Kitchen	Aired on Diners, Drive-ins and Dives on 3/2/2018 for...
Aloha Specialties	As Sunday 6-12-16 they will be closed for remodelin...
Animal Hospital of Cornelius	Thank you Dr. Kevin and staff for the great care and...
Bacchanal Buffet	No more waiting in line for hours. Self check-in to put...
Bacchanal Buffet	The definition of Bacchanal:
Baguette Cafe	Blown away again today by Chef Olivier's fresh, veg...
Baladie Café	Heads up.... The sweet older couple who owned this...
Baladie Café	The owner of Baladie Cafe (Richie A) sent me a nast...
Basil & Garlic Bistro	Was gonna write "this is our new favorite pizza in the...
Bistro Buffet	Friends don't let friends eat here...unless it's free. Als...
Brooklyn Bowl	Opens at 9am for NFL Sundays. $1 blue ribbon wing...
Camelback Mountain	Lovely hike but my bottom line question. Why are pe...
Capriotti's Sandwich Shop	Wow! What a surprise. This Capriotti's had two relat...
Carnevino	05.2018 - LVS closing all Batali restaurants 07.27.20...

93 row(s) affected, took 8.2ms

- Business_tips

- Number_of_restaurants

- Business_name_category

city	count
Toronto	2509
Las Vegas	1840
Montréal	1242
Phoenix	1184
Calgary	863
Pittsburgh	746
Charlotte	739
Mississauga	491
Cleveland	439
Scottsdale	400
Mesa	371
Madison	318
Tempe	313
Chandler	278
Henderson	271
Markham	258
Glendale	230
Gilbert	179
Scarborough	174

- Business_tips

- Number_of_restaurants

- Business_name_category

business_name	category
Arizona Biltmore Golf Club	Golf
Options Salon & Spa	Beauty & Spas
Nucleus Information Service	Local Services
The Kilted Buffalo Langtree	Bars
Fremont Arcade	Arcades
Bruster's Real Ice Cream	Food
Tomkins Park	Parks
Gap	Men's Clothing
Bank of America	Financial Services
Apteka	Nightlife
M Catering by Michael's	Caterers
Dunn-Edwards Paints	Interior Design
Complete Auto Glass	Home Services
Kirsty Blattner	Professional Services
All Star Donuts & Kolaches	Food
Ume Fashion Sushi	Sushi Bars
Holiday Inn Express & Suites Pi…	Hotels
Artisan Fine Dining Room	Restaurants
Olson Eye Care	Shopping

192127 row(s) affected, took 712.8ms

Relational v/s NoSQL?

Itemset Mining

Building Our Lattices

1. Build the best_tips table from the tip table by pulling out user_id, business_id where the compliment count was greater than 0.

2. From the best_tips we built the L1 lattice table by taking frequent itemsets of size 1 from best_tips with a minimum support that the users in the itemsets have at least greater than 1 business in common.

3. Using this minimum support we find the lattice tables upto the nth level until the lattice level at the nth level is empty.

4. In the last step we take the n-1 th lattice which will be non empty and join it with users table in order to display user names.

```
users : Bella, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Mike, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Vi, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Mimi, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Terri, Christie, Joanna, Kevin, Amanda, visited frequency :2
users : Jordan, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Joe, Grace, MarVy, Sharr, Christalle, visited frequency :2
users : Jennifer, Grace, MarVy, Sharr, Christalle, visited frequency :2
```

Thank You