[ CS 838 - Spring 2017 ] Stage 4 Report

Tarun Bansal, Ayush Gupta, Rohit Damkondwar

1. Combination Algorithm

**Schema Matching:**

Both tables A and B have the almost same schema.

A(Yelp) Schema:

1. ID
2. Restaurant Name
3. Address
4. City
5. Zipcode
6. Latitude
7. Longitude
8. Review\_count
9. Rating
10. Yelp\_id

B(Zomato) Schema:

1. ID
2. Restaurant Name
3. Address
4. City
5. Zipcode
6. Latitude
7. Longitude
8. Review\_count
9. Rating
10. Zomato\_id

Since, schema is almost same (Yelp has extra Yelp\_id and Zomato has extra Zomato\_id columns), Schema was pretty easy.

**Data Merging:**

1. Table E schema

Table E contains following columns:

1. Auto generated ID
2. Restaurant Name
3. Address
4. City
5. Zipcode
6. Latitude
7. Longitude
8. Rating
9. Source ID

Total Number of tuples in E: **730**

1. Sample Tuples

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1464 | Bonfyre American Grille | 2601 West Beltline Highway$\*$ Madison$\*$ WI 53713 | Madison | 53713 | 43.03 | -89.42 | 754 | 4.0 | 17503464 | 2YlUn3s132hNq5ueGeIiJg |
| 1937 | Presti's Bakery & Caf | 12101 Mayfield Rd$\*$ Cleveland$\*$ OH 44106 | Cleveland | 44106 | 41.50 | -81.59 | 779 | 4.27 | 16962390 | orrrhqRRUORIzUSxWTveKg |
| 11357 | 131 Main | 9886 Rea Road$\*$ Charlotte$\*$ NC 28277 | Charlotte | 28277 | 35.0341674 | -80.80 | 811 | 4.14648582 | 17148614 | 110iMPMPEEjFlf8HKVq84g |
| 11882 | Ilios Noche | 11508 Providence Road$\*$ Suite I$\*$ Charlotte$\*$ NC 28277 | Charlotte | 28277 | 35.05361881 | -80.77 | 890 | 4.078089888 | 17147444 | c-l4nDPZcEwapEiV-Xf08w |

1. Python code