City

Toronto

NY

Paris

London

Websites:

Yelp

Google

Zomato

Data fields:

|  |  |  |  |
| --- | --- | --- | --- |
| Original fields | yelp | zomato | google |
| name | name |  | name |
| type | id |  | place\_id |
| city | price |  | price\_level |
| country | rating\_place |  | rating\_place |
| latitude | Review\_count |  | user\_rating\_total |
| longitude | latitude |  | lat\_place |
| price | longitude |  | lng\_place |
| Food rating | address |  | formatted\_address |
| Service rating | phone |  | formatted\_phone\_number |
| Nro reviews | rating |  | reviews |
| Business id | text |  | author\_name |
| url | user |  | rating |
| address | reviews |  | text |
| Phone number |  |  |  |
| Kids friendly |  |  |  |
| Outdoor sitting |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Original fields names | yelp | zomato | google |
| User name | rating |  | reviews |
| rating | text |  | author\_name |
| text | user |  | rating |
| buisnessid | reviews |  | text |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| User input filters | yelp | zomato | google | Consolidate profile |
| type |  |  |  |  |
| city |  |  |  |  |
| name |  |  |  |  |
| price |  |  |  |  |
| rating |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Method:

Develop google api script

Pull data from google

Address and phone number

Build database on google profile as a main

Create a table for each site

Use google table as a joint to the other two

One table for user input

I got 20,000 address coordinates for Toronto from an Airbnb dataset may 2019, randomly select 1,000 coordinates. 1 hour

Then rewrite the code in jupyter notebook to request the data points: 1 hour

Name

Address

Rating

Reviews

url

phone

coordinates