**Yelp Dataset**

* the public dataset does not contain New York area so I used their API to pull ~12K restaurants and bars in Manhattan, Queens and Brooklyn
* the code is attached in the notebook
* This is what the schema looks like
* Graphical user interface, text, application

  Description automatically generated
* Table

  Description automatically generated
* **Additionally, we can pull reviews via the API, Yelp only limits to three reviews per business. (I haven’t pulled this yet but we can make use of this as an additional table). I’ve included a snippet of code is in the notebook.**

**NYC health data**

* I downloaded the latest dataset from the NYC website
* This dataset contains 200K rows but spans many many years
* I checked and most restaurants can be linked either by name or by phone number
* Graphical user interface

  Description automatically generated
* A picture containing table

  Description automatically generated

Four Square API

<https://location.foursquare.com/developer/reference/place-search>

* Four square is similar to Yelp but provides additional reviews and check-in details
* Data would need to be pulled via API

NYC income by zip-code

<https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a>

* We can use this data to potential answer questions such as is there a correlation between income level and restaurant sanitation

I stumbled upon this data visualization which is very cool and utilizes similar approach <http://nycfoodiverse.com/>

<https://github.com/Jiahao01121/restaurant-data-visualization>