Data Table Schema

Counties

List of counties on the New York-Newark-Jersey City, NY-NJ-PA Metropolitan Statistical Area.
26 rows & 3 columns. Size: < 1MB per year. Source: Public.

Field	Type	Description
state_code	INTEGER	A 2-digit integral number representing the state 34 = NJ 36 = NY 42 = PA
county_code	INTEGER	A 3-digit integral number uniquely assigned to each county
county_name	STRING	The county's name

Census_YYYY

Census data for year YYYY (2009 to 2018) for each tract in the New York-Newark-Jersey City, NY-NJ-PA Metropolitan Statistical Area.
4700 rows & 17 columns. Size: < 1MB per year. Source: Census API.
More information on the fields can be found here.

The *generate_census_data.py* file can be used to recreate these files with more columns if you wish to do so. In order to do that, you'll need to create an API key using this link and install the Census python package. Before running, edit the Python file and set:

- Your API KEY
- The output file name
- The list of variables you want to request
- Which years you want to query the API for

The existing files have these variables:

Field	Type	Description
geoid	STRING	Geographical identifier which is concatenation of state, county, and tract
year	INTEGER	The year of the data
NAME	STRING	The name of the tract
B01001_001E	INTEGER	Total Population
B19013_001E	INTEGER	Median household income (in dollars)
B25077_001E	INTEGER	Median home value (in dollars)
B03002_003E	INTEGER	Number of non-Hispanic Caucasians

B03002_004E	INTEGER	Number of non-Hispanic blacks or African Americans
B02001_004E	INTEGER	Number of American Indians and Alaskans
B03002_006E	INTEGER	Number of non-Hispanic Asians
B03002_007E	INTEGER	Number of non-Hispanic Hawaiians or Pacific Islanders
B03002_008E	INTEGER	Number of non-Hispanic others
B03002_009E	INTEGER	Number of non-Hispanic multi-racials
B03002_012E	INTEGER	Number of Hispanics or Latinos
state	INTEGER	The state ID
county	INTEGER	The county ID
tract	INTEGER	The tract ID

311Calls_YYYY

Sample (25% per year) of “311 complaints” from New York City's open data portal from 2010 - 2018.
Size: ~60MB zipped, ~300MB unzipped per year. Source.

These files were created by splitting the original 12GB file by year (using the “Created Date” field). From each of the individual files, a random sample of 25% was taken with Python’s pandas library using the following code:

df.sample(frac =.25, random_state = 123456)

where “df” was the DataFrame containing all the data for each year.

Field	Type	Description
Unique Key	STRING	Unique identifier of a Service Request (SR) in the open data set
Created Date	STRING	Datetime that SR was created
Closed Date	STRING	Datetime that SR was closed by responding agency
Agency	STRING	Acronym of responding city government agency
Agency Name	STRING	Full agency name of responding city government agency
Complaint Type	STRING	This is the first level of a hierarchy identifying the topic of the incident or condition. Complaint Type may have a corresponding Descriptor (below) or may stand alone

Descriptor	STRING	This is associated with the Complaint Type, and provides further detail on the incident or condition. Descriptor values are dependent on the Complaint Type, and are not always required in SR
Location Type	STRING	Describes the type of location used in the address information
Incident Zip	STRING	Incident location zip code, provided by geo-validation
Incident Address	STRING	House number of incident address provided by submitter
Street Name	STRING	Street name of incident address provided by the submitter
Cross Street 1	STRING	First cross street based on the geo-validated incident location
Cross Street 2	STRING	Second cross street based on the geo-validated incident location
Intersection Street 1	STRING	First intersecting street based on geo-validated incident location
Intersection Street 2	STRING	Second intersecting street based on geo-validated incident location
Address Type	STRING	Type of incident location information available
City	STRING	City of the incident location provided by geo-validation
Landmark	STRING	If the incident location is identified as a landmark, the name of the landmark will display here
Facility Type	STRING	If available, this field describes the type of city facility associated to the SR
Status	STRING	Status of SR submitted
Due Date	Date & Time	Date when responding agency is expected to update the SR. This is based on the Complaint Type and internal Service Level Agreements (SLAs).
Resolution Description	STRING	Describes the last action taken on the SR by the responding agency. May describe next or future steps
Resolution Action Updated Date	STRING	Date when responding agency last updated the SR

Community Board	STRING	Provided by geo-validation. Refers to a specific neighborhood
BBL	STRING	Borough-Block-Lot (BBL) or parcel numbers identifying the location of buildings or properties
Borough	STRING	Provided by the submitter and confirmed by geo-validation
X Coordinate (State Plane)	FLOAT	Geo-validated X-coordinate of the incident location
Y Coordinate (State Plane)	FLOAT	Geo-validated Y-coordinate of the incident location
Open Data Channel Type	STRING	Indicates how the SR was submitted to 311; e.g. phone, online, mobile, other, or unknown
Park Facility Name	STRING	If the incident location is a parks department facility, the name of the facility will be displayed here
Park Borough	STRING	The borough of incident if it is a parks department facility
Vehicle Type	STRING	If the incident is a taxi, this field describes the type of TLC vehicle
Taxi Company Borough	STRING	If the incident is identified as a taxi, this field will display the borough of the taxi company.
Taxi Pick Up Location	STRING	If the incident is identified as a taxi, this field displays the taxi pick-up location
Bridge Highway Name	STRING	If the incident is identified as a bridge/highway, the name will be displayed here
Bridge Highway Direction	STRING	If the incident is identified as a bridge/highway, the direction where the issue took place would be displayed here
Road Ramp	STRING	If the incident location was bridge/highway this column differentiates if the issue was on the road or the ramp
Bridge Highway Segment	STRING	Additional information on the section of the bridge/highway where the incident took place
Latitude	FLOAT	Geo-based latitude of the incident location
Longitude	FLOAT	Geo-based longitude of the incident location
Location	STRING	Combination of the geo-based latitude & longitude of the incident location