

Avi Bauer  
MESA8413  
Dec 10, 2025

## Final Project, Part 0: Proposal & ER Diagram

### Data Source:

NYC Office of Technology and Innovation (2025). *311 Service Requests from 2010 to Present* [Data set]. NYC OpenData.

[https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9/about\\_data](https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9/about_data)

### Synopsis:

For this project, I will explore the 311 Service Requests from 2010 to Present dataset published by NYC OpenData. This data includes information on the time, location, and nature of all service requests made to the 311 non-emergency help line in New York City from 2010 to the present. The full dataset is updated daily, and contains (as of December 8, 2025) 41.9 million rows and 41 data columns. The data may be downloaded directly from the NYC OpenData portal as a CSV, or as JSON using an API.

In order to downsize the dataset to a manageable size, on Nov 25, 2025 I used the NYC OpenData API to download the most recent 300 thousand rows of data, and saved these locally as a series of ten CSV files. The combined size of these files is 229 MB. The code used for this download is included in the notebook `MESA8413_Bauer_FinalProject_PART1.ipynb`, but please note before running it yourself that the full download took several hours to complete. The CSV files are also included in the folder `Data/0_NYC311_raw`.

Below is the draft ERD for the database built from the NYC 311 dataset. The final logical design is available in the file `MESA8413_Bauer_FinalProject_PART1.ipynb`. The original 41 columns are split into six tables (`agency`, `complaint_type`, `complaint_incident`, `complaint_status`, `taxi_details`, and `road_details`), and an additional two tables are populated with synthetic data (`call_operator` and `call_center`). These data are generated and organized in the file `MESA8413_Bauer_FinalProject_PART2.ipynb`.

In addition to creating and using a database using Postgresql (see files `MESA8413_Bauer_FinalProject_PART2.ipynb` and `MESA8413_Bauer_FinalProject_PART3.ipynb`) and analyzing the data in Python (see `MESA8413_Bauer_FinalProject_PART4.ipynb`), I will extend this project by migrating a subset of data to MongoDB and comparing query performance, as well as creating and testing

two roles with different access permissions. These analyses are available in the file MESA8413\_Bauer\_FinalProject\_PART5.ipynb.

### Draft ERD:

