COMP 370 Homework 5 – 311 Question Formulation

Assigned   Oct 2, 2023
Due   Oct 11, 2023 @ 11:59 PM

In this homework, we continue with our engagement with the New York City data division. We're working on refining questions. You'll be using the same derivate of the following dataset:

- https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9

As before, for the purpose of this assignment:

1. Download `nyc_311.csv.tgz` from MyCourses.
2. Trim it down to only include the incidents that occurred in 2020 (for an added challenge, see if you can trim the dataset down using exactly one call to the grep command line tool).

For the remainder of this assignment, you should only work with the trimmed down dataset.

For each task, below you have 2 objectives.

Objective 1: Formalize the question in a way that you can actually answer. When you do this, follow the example given in class (online political violence) in which the question was iteratively refined. For each question, show at least four versions of the question. Version 1: the question as posed by the stakeholder. Version 2: one refinement of the question that is more measurable and maps better onto the available data. Version 3: a second refinement that maps even better onto the available data. Version 4: the final version of the question that is very quantifiable and maps onto the available data.

Objective 2: Use python (Jupyter notebooks, CLIs, or any other pythonic approach) to build a visualization that credibly answers the question.

Task 1: Noise

The mayor wants to know if noise issues tend to stem from different causes across the year.

Task 2: Urban Rodents

The Departments of Sanitation and Health would like to know the where in the city rats and mice are most likely to create sanitation issues. In discussion with them, you determine that they aren't thinking in terms of geography, but more in terms of the kinds of buildings/properties/structures we find around a city.

Submission Instructions

- question_formalizations.md – a document with two sections, one for each task. Each section contains an itemized list of the versions of the question, ending with the fully refined question.
- task1_plot.png/jpg – the plot generated for task 1 that addresses the question in Task 1.
- task2_plot.png/jpg – the plot generated for task 2 that addresses the question in Task 2.