Final Project: Proposal

* The group members (names and UNIs)

Han Bao (hb2699)

Jasmine Niu (jn2855)

Qing Zhou (qz2266)

* The tentative project title

NYC Childcare Center Inspections

* The motivation for this project

High-quality childcare is important for children safety and health. In our homework, we did some data visualization with the NYC restaurant inspection dataset. We find this very inspiring and the idea of analyzing child care center inspections’ data came to our mind. A violation in a health care center would be, to some extent, more serious than violation in the restaurant and we would consider our project as a meaningful and realistic topic. We aim to find out some possible variables that would influence the rate of violation in child care centers in NYC.

* The intended final products

1. Interactive maps and dashboard (See “Visualization”)
2. A report that fits a linear regression model and summarizes exploratory analysis results with corresponding plots.

* The anticipated data sources

Our research focuses on the DOHMH Childcare Center Inspections dataset and other related datasets, aiming to investigate and visualize the overall trends for key outcomes including childcare violation rate percent, public health hazardous violation rate, critical violation rate, etc. across NYC five boroughs over time.

The DOHMH Childcare Center Inspections dataset comes from the NYC Open Data (https://opendata.cityofnewyork.us/). It describes the results of DOHMH’s inspection on child care centers in NYC and any associated violations at active, city-regulated, center-based child care programs and summer camps over the past 3 years.

* The planned analyses / visualizations / coding challenges
* Planned analyses:

Prior to data analysis, we will search for information such as yearly city budget on childcare and health care challenges in 2019-2022 and review relevant literature. Firstly, we will establish regression models to explore the association between the outcomes and the potential predictors, such as major demographic variables, total educational workers, program type, average annual income per capita, as well as government funding in specific districts where the childcare centers are located.

Moreover, we will perform visualization to see if there are differences in key outcomes among different categories or levels of these predictor candidates, followed by t-test or ANOVA to check if the differences are significant. Chi-squared tests will also be employed if necessary for the initial testing of the independence or association between outcomes and predictors. Last, linear or ordinal regression models will be built to confirm or refute our hypotheses.

We expect to have analyses on critical violation rate influenced by variables such as maximum capacity, different boroughs, numbers of educational workers, maximum capacity of program, and age range of the program. We’ll choose variables whichever we find statistically significant for the model.

* Visualizations:

1. Interactive maps that include the following information: average critical violation rate by boroughs in the past 3 years (2019-2022), distribution of childcare centers by violation categories
2. A dashboard that allows users to search for violation centers according to the violation description, zip code and so on
3. Time trend graph reflecting average critical violation rate over time
4. Histograms or scatterplots showing relevant information, e.g. the number of different age-range program in each borough

* Coding challenges:

We need a clear understanding and their relations on the four columns: violation rate, average violation rate, critical violation rate, average violation rate. Some other useful datasets influencing childcare should be found to produce a more comprehensive outcome.

* The planned timeline

10/28 – 11/12 Team registration and proposal

11/15 – 11/18 Project review meeting

11/21 – 11/25 Discussion and adjustment based on the project review meeting

11/28 – 12/3 Import, tidy and manipulate the data; conduct planned analysis and visualization

12/4 – 12/8 Finish the report, webpage and screencast

12/9 – 12/10 Revise (if necessary) and submit the project