**Group Project (Spring 2021)**

**Guidelines:**

* Choose one of the two topics and address empirical questions for that topic.
* Can use any software tool.
* A team of 2 or 3;
* Feel free to add additional data or use your own data if necessary.
* Submit a short **write-up** before you work intensively on the analysis (Mar. 18th 2021). Your write-up should briefly discuss your proposal, show the dataset, and explain the potential variables you use. Less than one single-spaced page.
* Submit a **final report** at the end of the semester (Apr. 29th 2021). Your final report should include the following two parts:
* The background: A brief description, one single-spaced page or less, describing what’s great about your analysis and your graphic.
  + Highlight how your analysis helped you arrive at your conclusions.
  + Highlight how your graphic illustrates your key findings.
* The story: A graphic that reflects the major findings of your data analysis.
  + It can be a static image or interactive (such as JavaScript). Interactive graphics must be self-contained and run without special software on any Windows 10 PC. Web-based graphics must run using the Chrome browser.
  + It should be in the form of an infographic.

**Your entry will be disqualified if…**

* Either of the write-up or the final report is submitted after the deadline.
* You don’t specify the project you are addressing in your final report.
* Your names and IDs are not on both the submissions.
* The attachments won’t open or are in the wrong file format.

**How entries will be evaluated**

* Graphic (10 points)
  + Clarity (how well the graphic stands on its own without additional explanation)
  + Novelty/creativity (originality of thought; surprising way of approaching the data)
  + Insight (graphic aids understanding of the data)
  + Utility (ability of the graphic to aid decision making)
* Analysis (10 points)
  + Relevance (analysis relates to the problem statement)
  + Completeness (degree to which the analysis answers the stated question)
  + Depth (sophistication of the analysis)
  + Consistency (conclusions consistent with the analysis)

**Topic 1: Motor Vehicle Collisions - Crashes**

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC from 2019. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least $1000 worth of damage. It should be noted that the data is preliminary and subject to change when the MV-104AN forms are amended based on revised crash details. For the most accurate, up to date statistics on traffic fatalities, please refer to the [NYPD Motor Vehicle Collisions page](https://www1.nyc.gov/site/nypd/stats/traffic-data/traffic-data-collision.page) (updated weekly) or [Vision Zero View](http://www.nycvzv.info/) (updated monthly).  
  
Questions:

1) Which factors do you think lead to high risks of death from motor vehicle crash? Why?

2) Use the data to validate your above guess. If results are different from your guess, what might be the reason?

3) Do you think your analysis suggests causality or just correlation? Why?

4) How can NYPD use your findings to reduce motor vehicle crash deaths?

Other data: https://data.cityofnewyork.us/Transportation/For-Hire-Vehicles-FHV-Active/8wbx-tsch

**Topic 2: Restaurants and Rodents**

You will use two datasets to examine whether more opening of restaurants leads to more rodent problems.

Dataset 1: https://data.cityofnewyork.us/Health/Rodent-Inspection/p937-wjvj

The Rat Information Portal (RIP) is a web-based mapping application where users can view rat inspection data. Notes on data limitations: Please note that if a property/taxlot does not appear in the file, that does not indicate an absence of rats - rather just that it has not been inspected. Similarly, neighborhoods with higher rates of active rat signs may not actually have higher rat populations but simply have more inspections.

Dataset 2: https://data.cityofnewyork.us/Transportation/Open-Restaurant-Applications/pitm-atqc

Open Restaurant Applications is a dataset of applications from food service establishments seeking authorization to re-open under Phase Two of the State’s New York Forward Plan, and place outdoor seating in front of their business on the sidewalk and/or roadway.

Questions:

1) Do you believe more restaurants lead to more rodent inspections? Why?

2) Use the data to validate your above guess. If results are different from your guess, what might be the reason?

3) Do you think your analysis suggests causality or just correlation? Why?

4) How can NYC Department of Health and Mental Hygiene (DOHMH) use your findings to manage rodent problems?

Other data: https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3