**Introduction**

Since COVID-19 has not disappeared, we still need to coexist with it for a long time in the future. Although the probability of serious illness and death after vaccination is greatly reduced, some studies still point out that there is a high probability of sequelae after COVID-19, as well as the high risk of the elderly population, and there are obvious symptoms of pneumonia, cough, pharyngitis and other painful symptoms after the disease. These factors remind people that although COVID-19 has been controlled, its threat has not been completely eliminated And it is still mutating, so we want to study this related problem and choose to predict the epidemic time by the virus concentrations measured in NYC wastewater.

**Research Questions**

We assume that the virus concentrations measured in NYC wastewater should have a certain relationship with the number of people tested. What is the time relationship between them? In addition, we can explore the time relationship between the number of infections detected and the number of hospitalizations and deaths Applying this analysis in practice can guide hospitals to predict in advance the number of patients in COVID-19 be admitted to hospital, and enable people to strengthen the prevention of the pandemic of COVID-19 even predict the pandemic of COVID-19 in advance

**Data to be Used**

Our data are both from NYC open data. One of them is planned to use the downloaded csv file, and the other is to use the API provided by the website to obtain json text and process it.

**COVID-19 Outcomes by Testing Cohorts: Cases, Hospitalizations, and Deaths**

<https://data.cityofnewyork.us/Health/COVID-19-Outcomes-by-Testing-Cohorts-Cases-Hospita/cwmx-mvra>

**SARS-CoV-2 concentrations measured in NYC Wastewater**

[SARS-CoV-2 concentrations measured in NYC Wastewater | NYC Open Data (cityofnewyork.us)](https://data.cityofnewyork.us/Health/SARS-CoV-2-concentrations-measured-in-NYC-Wastewat/f7dc-2q9f)

**Approach**

We plan to store it in postgresql. First, use the ER graph to draw the structure of the database, then process the obtained data, and import the obtained dataframe object into the database. We plan to use Exploratory Data Analysis (EDA) and Descriptive Statistical Analysis to research problems, and try to use Predictive Analysis. Using EDA can quickly find potential correspondences between data. Descriptive Statistical Analysis allows us to summarize some general rules through the analysis of data characteristics in different periods, while predictive analysis allows us to predict future conditions. We will mainly use line chart to reflect the trend of different data, and use histogram to show the change rate of data. We plan to complete data acquisition, data cleaning and data storage in the 10th week. Carry out some analysis in week 11. Generate new data according to new requirements Continue analysis and report in week 12. Complete the report in week 13 Pinxue Lin is responsible for obtaining data processing data and writing them into the database. Aayush Bhetuwal and Siddanta K C are responsible for the design database and data analysis.