**MIDTERM PROJECT - ANALYZING NEW YORK CITY DATA WITH SQL, PYTHON, AND VERSION CONTROL**

**SCALABLE DATABASES – CS673**

**TEAM MEMBERS:**

**ROHAN KALAPAVRUKSHA – U01911326**

**ROSHAN KALAPAVRUKSHA – U01929541**

**SRIMANTH MADIRA – U01885136**

**VAISHNAVI KOYA – U01884267**

**DATA SETS:**

1. [**NYC Water Consumption Dataset - Housing Development**](https://data.cityofnewyork.us/Housing-Development/Water-Consumption-And-Cost-2013-Feb-2023-/66be-66yr)

The "NYC Water Consumption Dataset - Housing Development" is a collection of data that provides insights into water consumption patterns specifically related to housing developments in New York City. The dataset likely includes information such as water usage metrics, trends, and relevant variables specific to residential areas within the city. This dataset could be valuable for researchers, policymakers, and urban planners interested in understanding water consumption dynamics in the context of housing development, which can have implications for resource management, infrastructure planning, and sustainability initiatives within the urban environment of New York City.

1. [**NYPD Arrest Dataset - Public Safety**](https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc)

The "NYPD Arrest Dataset - Public Safety" is a compilation of data that sheds light on arrests made by the New York City Police Department (NYPD). This dataset likely contains information about the nature of arrests, demographics of individuals involved, time and location of incidents, and possibly additional contextual details. Analyzing this dataset can provide valuable insights into public safety trends, law enforcement activities, and the distribution of arrests across various demographic and geographic factors. Researchers, law enforcement agencies, and policymakers can leverage this dataset to better understand crime patterns, assess the effectiveness of policing strategies, and inform decisions related to public safety initiatives in New York City.

1. [**NYC Air Quality Dataset - Environment**](https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r)

The "NYC Air Quality Dataset - Environment" is a dataset that compiles information related to air quality in New York City. This dataset likely includes measurements of various air pollutants, such as particulate matter, ozone, nitrogen dioxide, and others, along with corresponding temporal and spatial data. Researchers, environmentalists, and policymakers can utilize this dataset to analyze trends, assess the impact of human activities on air quality, and monitor compliance with air quality standards. Understanding the dynamics of air quality in NYC is crucial for addressing public health concerns, formulating environmental policies, and implementing measures to mitigate pollution, contributing to the overall well-being of the city's residents and the urban environment.

**REFERENCES & EXTERNAL SOURCES:**

1. [**Geek Culture for Data Visualizations**](1.%09https:/medium.com/geekculture/48-data-visualizations-that-load-in-a-single-line-of-code-a0be5bea903b)
2. [**Data Camp for Data Visualization Techniques**](https://www.datacamp.com/blog/data-visualization-techniques)
3. [**Stack Overflow for Data Wrangling with Python**](https://stackoverflow.com/questions/28823371/data-wrangling-with-python-pandas/28823417)
4. [**Pro Git book by Scott Chacon and Ben Straub for Version Control**](https://git-scm.com/book/en/v2)

**STEPS TO ACCESS PROJECT FILES:**

1. **Open Terminal or Command Prompt:**

Open the terminal or command prompt on your local machine. You can typically find this by searching for "Command Prompt" on Windows or using the Terminal application on macOS or Linux.

1. **Clone the Repository:**

Use the following command to clone the project repository to your local machine. This command will download all files, including datasets, from the remote repository to your computer:

git clonehttps://github.com/RohanNK12345/midterm\_project\_scalable\_database.git

1. **Navigate to the Project Directory:**

Move into the directory that was created by the clone operation:

Command to Change Directory:

cd “path to directory”

Example:

C:\Users\SrimanthMadira\midterm\_project\_scalable\_database\midterm\_project

\_scalable\_database

1. **Extend Future Processes and Implement Automation:**

With the datasets and project files now on your local machine, you can extend future processes. Utilize the data for automation, implement new analyses, or uncover new findings based on the available datasets.

**CONCLUSION:**

In summary, by following these steps, you've successfully cloned the project repository to your local machine, navigated to the project directory, and accessed the datasets. This process empowers you to explore and analyze the datasets locally, providing the foundation for future processes, automation, and the discovery of valuable insights. The local setup allows for seamless collaboration and individual contributions to the project.