**Project Title:** *Using Geographic Information System (GIS) to assess and identify factors leading to the increasing trend in drug overdose mortality in the United States*

**Goal:** The project's main objective is to apply data analytics and visualization tools, specifically GIS, to assess the relationship between socioeconomic factors and drug overdose mortality rates across different states and generate an interactive cartographic representation of the datasets. The project intends to answer the following:

1. Is there a correlation between socioeconomic status and drug abuse/drug overdose cases?
2. If so, which factors (i.e., socio-economic factors and educational attainment) significantly contribute to the drug overdose problem among the population?
3. Which state depicts the highest count of mortality due to drug overdose?
4. Which illicit drug has led to the highest number of deaths?

**Task-relevant data:** The project will utilize different datasets from the Centers for Disease Control and Prevention and the USDA Economic Research Service. The datasets from the CDC will include information about the drug overdose mortality rates in the US in years 2019 and 2023, while the datasets from USDA Economic Research Services will include information about the demographics and socioeconomic status of the general population in the US in years 2017, 2021, and 2022.

CDC Datasets:

1. PLACES: ZCTA Data (GIS Friendly Format), 2023 release dataset has 32 thousand rows and 77 columns. This dataset can be accessed and downloaded in different formats from <https://data.cdc.gov/500-Cities-Places/PLACES-ZCTA-Data-GIS-Friendly-Format-2023-release/kee5-23sr>.
2. **VSRR Provisional Drug Overdose Death Counts** dataset has 59 thousand rows and 12 columns. This dataset can be accessed and downloaded in different formats from <https://data.cdc.gov/NCHS/VSRR-Provisional-Drug-Overdose-Death-Counts/xkb8-kh2a>.

USDA Economic Research Service Datasets:

1. County-level datasets (Poverty, Unemployment, and Median Housing Income and Education) have 60 rows and seven columns. These datasets can be accessed and downloaded in an Excel format from <https://www.ers.usda.gov/data-products/county-level-data-sets/>.

**Clients:** The project's clients will be the general public, the policymakers, and the healthcare providers. This project will inform these clients about the current and existing problems with drug abuse and overdose. The general public will be informed about the problem's prevalence and its associated factors. The policymakers will be able to identify the factors contributing to this problem; hence, they can create policies to eliminate the problem and improve the existing approaches to cater to the affected individuals. Lastly, healthcare providers can advocate more effectively and efficiently about this matter.

**Kinds of knowledge:** The project will apply data wrangling to ensure that the information from different datasets is correctly formatted and free of unwanted and missing data, use various statistical tools to analyze and correlate the factors associated with drug overdose mortality rates, and effectively visualize/map the data.

**Data Science Tools:** The project will use ArcGIS software to analyze and visualize data.

**Conclusion:** Implementing this project will provide an effective and informative cartographic representation of the cases of drug overdose mortality in the US that will benefit the general public, policymakers, and healthcare providers to understand better the current and existing problems that people are currently facing with drug overdose.

**References:**

Centers for Disease Control and Prevention. (2023, August 8). *Understanding the Opioid*

*Overdose Epidemic*. Retrieved September 9, 2023, from <https://www.cdc.gov/opioids/basics/epidemic.html>.

Spencer, M. R., Miniño, A., Warner, M. (2022, December). *Drug Overdose Deaths in the United*

*States, 2001-2021*. NCHS Data Brief, No. 457. Retrieved September 9, 2023, from <https://www.cdc.gov/nchs/data/databriefs/db457.pdf>.

**Data Dictionary:**

1. VSRR Provisional Drug Overdose Death Counts Dataset

| Name | Description | Data Type | Sample |
| --- | --- | --- | --- |
| State |  | Plain Text | NY |
| Year |  | Number | 2015 |
| Month |  | Plain Text | April |
| Period |  | Plain Text | 12 month-ending |
| Indicator |  | Plain Text | Heroin (T40.1) |
| Data Value |  | Number |  |
| Percent Complete |  | Plain Text | 100 |
| Percent Pending Investigation |  | Number | 0 |
| State Name |  | Plain Text | New York |
| Footnote |  | Plain Text | … |
| Footnote Symbol |  | Plain Text | \*\* |
| Predicted Value |  | Number | 126 |

1. PLACES: ZCTA Data (GIS Friendly Format), 2023 release Dataset

| Name | Description | Data Type | Sample |
| --- | --- | --- | --- |
| ZCTA5 | 5-digit ZIP code | Plain Text | 01001 |
| TotalPopulation | Total population of Census 2010 | Number | 16,769 |
| ACCESS2\_CrudePrev | Model-based estimate for crude prevalence of current lack of health insurance among adults aged 18-64 years, 2021 | Number | 33.7 |
| … | … | … | … |