Write me an executive summary regarding The following are prescription drug from the Controlled Substance Monitoring Database (CSMD) for all years of available data. This dataset also indicates which of these data are currently featured on the Tennessee Drug Overdose Data Dashboard (the five most recent years of available data). This should include an overview of the project. It should briefly touch on the motivation, data question, data to be used, along with any known assumptions and challenges

**ChatGPT said:**

ChatGPT

**Executive Summary: Analysis of Controlled Substance Monitoring Database (CSMD) Data**

**Overview**

This project focuses on analyzing prescription drug data extracted from the Controlled Substance Monitoring Database (CSMD) over all years of available records. The dataset identifies prescription drug trends and includes information on which data points are featured on the Tennessee Drug Overdose Data Dashboard, which emphasizes the five most recent years of available data. The analysis aims to inform public health strategies, assess prescription trends, and monitor potential risks associated with controlled substances.

**Motivation**

The primary motivation for this project is to support efforts in addressing the opioid epidemic and other drug-related challenges by leveraging data-driven insights. The Tennessee Drug Overdose Data Dashboard provides a public resource for understanding overdose patterns, and this analysis extends its utility by evaluating prescription trends over a broader historical scope. This effort is critical for informing policymakers, healthcare providers, and public health organizations in their strategies to combat substance misuse and overdose-related fatalities.

**Data Questions**

The project aims to address several key questions:

1. What are the trends in the prescription of controlled substances over time?
2. How do prescription patterns align with overdose trends highlighted in the Tennessee Drug Overdose Data Dashboard?
3. Which drugs or categories have shown significant changes in prescription volume?
4. Are there notable correlations or gaps between prescription data and overdose metrics?

**Data to Be Used**

The analysis utilizes:

* **Controlled Substance Monitoring Database (CSMD):** Comprehensive historical data on prescription drug trends.
* **Tennessee Drug Overdose Data Dashboard:** A subset of the most recent five years of data, focusing on overdose-related metrics.

**Assumptions and Challenges**

**Assumptions:**

* The data accurately reflects all prescriptions for controlled substances within the reporting period.
* Trends in the dataset are representative of broader public health challenges in Tennessee.
* Integration of data between CSMD and the Overdose Dashboard aligns without significant discrepancies.

**Challenges:**

* Potential inconsistencies or gaps in data reporting across years.
* Difficulty in directly correlating prescription rates with overdose incidents due to external factors (e.g., illegal drug use, socioeconomic influences).
* Limited granularity in demographic or geographic details may obscure localized trends.

This analysis represents a vital step in utilizing historical and recent data to generate actionable insights into controlled substance trends and their implications for public health initiatives in Tennessee.

4o

**You said:**

Define your MVP. This should be a description of what your final capstone will look like, including visualizations, how the analysis will be presented, who the intended audience is, etc.

**ChatGPT said:**

ChatGPT

**Minimum Viable Product (MVP): Controlled Substance Monitoring Data Analysis Capstone**

**Description**

The final capstone project will provide a comprehensive analysis of prescription trends for controlled substances in Tennessee using the Controlled Substance Monitoring Database (CSMD) and data from the Tennessee Drug Overdose Data Dashboard. The goal is to produce a visually engaging, data-driven report that highlights key findings and provides actionable insights for stakeholders involved in public health, policy, and healthcare.

**Components of the MVP**

1. **Interactive Dashboard**:
   * A user-friendly, web-based interactive dashboard will showcase the analysis results. The dashboard will include:
     + Time-series visualizations of prescription trends for specific drugs or drug classes (e.g., opioids, benzodiazepines).
     + Geographic heat maps displaying prescription rates and overdose trends across Tennessee counties.
     + Overlap graphs showing correlations between prescription patterns and overdose incidents.
   * Filters for users to explore the data by drug type, year, and region.
2. **Key Visualizations**:
   * Line graphs illustrating long-term trends in the volume of prescriptions for controlled substances.
   * Stacked bar charts showing the breakdown of prescription types by year.
   * Comparative scatterplots of prescription rates versus overdose occurrences.
   * A side-by-side view of CSMD data and the Overdose Dashboard for the past five years.
3. **Written Analysis**:
   * A detailed report summarizing the findings, emphasizing major insights such as:
     + Drugs with the most significant changes in prescription patterns.
     + Regions or populations most affected by changes in controlled substance use.
     + Possible implications of prescription trends on overdose rates.
   * Discussion of any discrepancies or gaps between the two datasets.
4. **Presentation of Insights**:
   * The final deliverable will include an executive summary, detailed analysis, and actionable recommendations presented in both a formal written report and an oral presentation.
   * Highlighted sections will include policy recommendations, potential interventions for reducing overdose rates, and suggestions for future data collection or analysis.
5. **Supplementary Materials**:
   * Access to cleaned datasets (or links to data sources).
   * An appendix detailing methodology, assumptions, and limitations.

**Intended Audience**

* **Public Health Officials**: To inform statewide or regional interventions addressing substance misuse.
* **Policy Makers**: To shape legislation or guidelines aimed at curbing the opioid crisis.
* **Healthcare Providers**: To raise awareness about prescribing practices and their potential consequences.
* **Researchers and Data Analysts**: To provide a foundational dataset and analysis for further studies.
* **General Public**: To increase transparency and awareness about trends in controlled substance use and overdose.

**MVP Outcomes**

The final capstone will offer a polished, actionable deliverable that combines data visualizations, written insights, and an interactive component to effectively communicate findings and empower stakeholders to address controlled substance misuse in Tennessee.

4o

**You said:**

write out a schedule from now until the due date of 3/6 with internal demos due on 2/27

**ChatGPT said:**

ChatGPT

​

**You said:**

write out a schedule from now until the due date of 3/6 with internal demos due on 2/27 1. Get the Data (finish date) 2. Clean & Explore the Data (finish date) 3. Create Presentation of your Analysis (finish date) - Should be a presentation, but could include a Jupyter Notebook or dashboard in Excel, Tableau, or PowerBI 4. Internal demos (<date of internal demos) 5. Demo Day!! (<date of demo day>)

**ChatGPT said:**

ChatGPT

Here’s a detailed schedule to help you stay on track for your project, ensuring readiness for internal demos on 2/27 and Demo Day on 3/6:

**Schedule**

**1. Get the Data**  
**Finish Date**: **1/26**

* Collect all necessary data from the Controlled Substance Monitoring Database (CSMD) and Tennessee Drug Overdose Data Dashboard.
* Ensure data access is reliable and datasets are complete.

**2. Clean & Explore the Data**  
**Finish Date**: **2/9**

* Clean the data: Remove duplicates, handle missing values, and normalize the data for analysis.
* Conduct exploratory data analysis (EDA):
  + Generate summary statistics and visualizations.
  + Identify trends, anomalies, or gaps.
  + Develop insights into relationships between variables (e.g., prescription trends vs. overdose rates).
* Document assumptions, challenges, and preliminary findings.

**3. Create Presentation of Your Analysis**  
**Finish Date**: **2/23**

* Build initial visualizations (e.g., time-series graphs, heat maps, scatterplots) to communicate findings effectively.
* Develop the presentation structure, including:
  + Overview of the project and key objectives.
  + Methodology, insights, and major findings.
  + Actionable recommendations or conclusions.
* Create a prototype of your dashboard using Tableau, Power BI, or another tool.
* Refine content to align with the target audience (public health officials, policymakers, etc.).

**4. Internal Demos**  
**Date**: **2/27**

* Present your findings and dashboard to internal stakeholders for feedback.
* Gather input on areas for improvement, additional analysis, or clearer visualizations.
* Revise and polish based on feedback.

**5. Demo Day!!**  
**Date**: **3/6**

* Finalize all visualizations and the interactive dashboard.
* Complete and rehearse your presentation to ensure clarity and flow.
* Deliver the polished capstone presentation with actionable insights.

This schedule allows for adequate time to collect, clean, analyze, and present your data while incorporating feedback from internal demos.

4o

**You said:**

where will data come from for this project?

**ChatGPT said:**

ChatGPT

The data for this project will come from the following primary sources:

1. **Controlled Substance Monitoring Database (CSMD):**
   * The CSMD provides comprehensive data on prescription drug trends, including all controlled substances prescribed over the available years.
   * This dataset will be the foundation for analyzing long-term trends in prescription volumes and identifying patterns or anomalies.
2. **Tennessee Drug Overdose Data Dashboard:**
   * This dashboard includes data on overdose-related metrics for the five most recent years of available data.
   * It provides a subset of relevant information, such as overdose counts, locations, and trends, to complement the CSMD data.

**Additional Considerations:**

* Ensure proper data access through public or authorized channels for these sources.
* Verify metadata and documentation from these datasets to ensure accurate interpretation.
* Supplementary data (if needed): External data sources such as census information or healthcare utilization statistics may be considered for demographic or geographic context.

This combination of datasets will provide a robust foundation for analyzing controlled substance use and overdose patterns in Tennessee.

4o

**You said:**

Write out the steps to clean the data in excel

**ChatGPT said:**

ChatGPT

Here is a step-by-step process to clean the data in Excel:

**Steps to Clean the Data in Excel**

**1. Load and Inspect the Data**

* Open the dataset(s) in Excel.
* Inspect the data for structure, column headers, and general content.
* Identify missing values, duplicates, inconsistent formatting, or unnecessary columns.

**2. Remove Unnecessary Columns**

* Delete columns that are not relevant to the analysis, such as metadata or unrelated information.
  + Example: Remove columns like "Notes" or "Extra Fields" if they are not useful.

**3. Handle Missing Values**

* **Filter Missing Data:**
  + Use the filter function to identify rows with blank cells in important columns.
* **Decide on Treatment:**
  + For critical columns: Remove rows with missing data if they cannot be filled logically.
  + For non-critical columns: Fill missing values using appropriate methods:
    - **Numerical Data:** Replace with the mean, median, or other imputation methods.
    - **Categorical Data:** Replace with "Unknown" or the most common value.

**4. Remove Duplicates**

* Select all data columns.
* Go to the **Data** tab and click **Remove Duplicates**.
* Review the results to ensure duplicates were correctly removed.

**5. Standardize Data Formats**

* **Date Formatting:**
  + Ensure all dates are in a consistent format (e.g., "MM/DD/YYYY").
* **Text Consistency:**
  + Use the **TRIM()** function to remove extra spaces.
  + Use **UPPER()** or **LOWER()** functions to standardize case formatting for text values.
  + Correct spelling inconsistencies in categorical columns.

**6. Correct Data Types**

* Verify that numerical values are in number format and text fields are in text format:
  + Select the column and set the format under the **Home** tab.
  + Convert non-numeric values (e.g., "$1,234") into proper numbers by removing special characters.

**7. Address Outliers**

* Create a new column for key metrics (e.g., prescription count).
* Sort numerical data in ascending/descending order to identify unusually high or low values.
* Investigate and decide whether to exclude, modify, or keep outliers.

**8. Consolidate Data**

* If working with multiple datasets, ensure all datasets share consistent headers and formatting.
* Use **VLOOKUP**, **INDEX/MATCH**, or **Power Query** to merge datasets on common fields (e.g., drug name, county, or year).

**9. Create Validation Rules**

* Use **Data Validation** to restrict data entry errors:
  + Example: Limit values in a column to a specific range or list (e.g., valid drug names).

**10. Document Changes**

* Create a new tab called "Data Cleaning Log" to track changes, assumptions, and issues encountered during cleaning.

By following these steps, your dataset will be cleaned and ready for analysis while maintaining consistency and reliability