CDC WONDER API DATA

PROJECT OVERVIEW:

This project will utilize the CDC WONDER online public health database to access key national datasets for analysis and visualization. The goal is to showcase the depth of public health data available through WONDER and demonstrate its potential to empower data-driven reporting, research, and policy decisions.

Project Details:

- Data will be queried from CDC WONDER using the online interface and tools. Key datasets to be analyzed include:
    - Mortality data
    - Notifiable disease data
    - Cancer statistics
    - Birth and natality data
    - Vaccine data
    - Environmental health data
- Data will be queried with parameters like time period, location, demographics, health outcomes, and other variables relevant to research questions.
- Analysis will focus on visualizing trends over time and geographic patterns, aided by charts, graphs, and maps.
- Use cases will demonstrate how CDC WONDER can inform public health reports, research papers, health policy briefs, and data journalism.
- All analyses and visualizations produced will be publicly shared to highlight the accessibility and value of WONDER data.
- If feasible, the project may also explore using automated methods like API access or web scraping to efficiently query and extract data from the WONDER site.

DELIVERABLES:

- Published public health data stories and visualizations based on CDC WONDER data
- Code and technical documentation for accessing, analyzing, and visualizing WONDER data
- Use case descriptions and examples demonstrating applied uses of CDC WONDER data

PROJECT BACKGROUND

The goal of this project is to aggregate and analyze public health data from across the United States. Visualizations and interactive tools will then be developed to make the data more accessible and impactful for various audiences.

CDC WONDER was identified as an excellent data source for the project, given its national-level data on mortality, disease prevalence, health behaviors, and other topics relevant to public health.

INITIAL TECHNICAL INVESTIGATION

The team initially explored using the CDC WONDER API to directly access the required datasets. However, further investigation found the API is limited to state-level data only. National aggregate data could not be obtained through the API.

The alternative of web scraping the CDC WONDER website was then investigated. Web scraping uses automated methods to extract data directly from web pages. The team has successfully used web scraping before with other websites such as CMS.

Before web scraping CDC WONDER, the terms of use need to be reviewed to ensure this method is allowed and permissible. If permitted, web scraping tools like Selenium and Python could be leveraged to query and extract the data needed from the website.

PROJECT PLAN

If web scraping of CDC WONDER is viable, this method will be used to obtain aggregated, national-level data on key public health indicators. The scraped data will be stored in a database for further analysis.

Data visualizations and interactive dashboards will then be developed to bring the data to life. Use cases and applications of the data will be showcased to promote open public health data.

All code and processes documented will be open source to allow reproducibility and collaboration. Ethics and responsible data use will be ensured throughout the project.

PROGRESS:

Our project aims to gather public health data from across the United States to allow for data analysis and visualization. The CDC WONDER database seems really useful for this - it contains national-level data on topics like deaths, illnesses, health behaviors, and demographics that would be great for our project.

At first, we thought we could use CDC WONDER's API to easily get all the data we need. But after taking a closer look, the API can only get state-level data, not national data. However, we might be able to use a technique called web scraping to get the national-level CDC data instead.

Web scraping uses programs to scrape data off of websites, even if they don't have an API. We've used web scraping successfully before to get data from other sites like CMS.
We'll need to research more to see if web scraping CDC WONDER is allowed based on their terms of use. If it's permitted, we can use Python tools like Selenium to automate scraping the data we need from the CDC WONDER website.
In short, CDC WONDER has great public health data we want for our project, but we can't use the API. Web scraping might work as a Plan B to get the data, if the CDC allows it. We need to explore this technical option m

In summary, while the ideal API access to CDC WONDER is not available, web scraping presents a potential alternative to harness this valuable public health data source, pending confirmation it meets terms of use. This project aims to make CDC WONDER data more accessible through effective analysis, visualization and reporting.

The team is still looking into whether web scraping can be used to get the national-level data from CDC WONDER. While the team researches that option more, they went ahead and downloaded some sample data from the site to start exploring.

The team took a look at the columns and attributes of the sample data they pulled to get familiar with how CDC WONDER datasets are structured and organized. This will help the team figure out what data to focus on and how best to analyze it.

The team also started working on a Python script that could eventually be used to automate downloading the data if the team does get the green light to scrape the site. The script will help grab the data quickly and efficiently.

Basically the team is doing some upfront work to better understand the CDC WONDER data and prep for extracting the full datasets they need. The team is hoping the terms of use allow web scraping so they can get the national-level data the project requires. But while the team awaits final word on that, they're moving forward with sampling the data and prepping their analytics code.

The sample data and Python script will make it easier for the team to hit the ground running if web scraping ends up being viable. The team will have a head start on working with the data and can quickly scale up to larger datasets. But the team still has some hoops to jump through before they know if web scraping will work out in the end. So for now, it's about getting as ready as they can while the team finalizes their data access plan.