Capwell Well Database

Overview

This repository contains the following three components:

1. **National Wells Database.xlsx**: A dataset containing detailed information about wells across the country.
2. **DatabaseSearcher.ipynb**: A Jupyter Notebook for analyzing and searching through the National Wells Database.
3. **CompanySpecificNews.ipynb**: A Jupyter Notebook for analyzing company-specific news data.

This README will guide you through understanding the contents of these files, how they interact, and how you can run and explore them on your own system. It is intended for beginners with no coding experience, so don't worry if this is your first time working with this type of project!

File Descriptions

1. National Wells Database.xlsx

- **What is it?** This is a spreadsheet file containing comprehensive data about wells, including locations, operational status, and other relevant attributes. The data has been curated from publicly available resources.
- **Where did the data come from?** The data originates from publicly accessible government databases, such as geological and environmental agencies. It has been compiled into a single file for ease of use.
- **How is it structured?** The file includes columns such as:
    - Well ID
    - Location (State, County)
    - Depth
    - Operational Status (Active/Inactive)

2. DatabaseSearcher.ipynb

- **What does this file do?** This is a Python Jupyter Notebook designed to help users analyze and search through the National Wells Database. You can filter wells based on various criteria such as depth, location, or operational status.
- **Features:**
    - Load the dataset into a structured format.
    - Perform advanced searches, e.g., finding wells deeper than a certain threshold.

    - Visualize data trends (e.g., a bar chart showing the number of active wells per state).
- **Why is it useful?** The notebook allows you to easily explore large datasets without having to use complicated tools or software.

3. CompanySpecificNews.ipynb

- **What does this file do?** This is another Jupyter Notebook designed to analyze news data for specific companies. It can help you find trends, sentiment, and other useful insights from a dataset of company-related news articles.
- **Features:**
    - Import a dataset of news articles (not provided in this repository).
    - Filter articles based on company names or keywords.
    - Perform sentiment analysis to determine if the news is positive, negative, or neutral.
- **Why is it useful?** This tool is helpful for market analysts, investors, or anyone interested in tracking public sentiment or trends around a specific company.

Requirements

To use these files, you will need the following:

- A computer with Python installed (version 3.8 or higher).
- Access to Jupyter Notebook (part of the Anaconda distribution or available separately).
- The following Python libraries:
    - pandas
    - matplotlib
    - seaborn
    - openpyxl
    - nltk (for the CompanySpecificNews notebook)

If you don’t already have these libraries, don’t worry! Instructions for installing them are provided below.

Installation Instructions

1. **Install Python and Jupyter Notebook:**
    - Download and install Anaconda (https://www.anaconda.com/products/distribution).
    - Open the Anaconda Navigator and launch Jupyter Notebook.
2. **Install the required Python libraries:**
    - Open a terminal or command prompt and type the following commands:
      pip install pandas matplotlib seaborn openpyxl nltk

3. **Download the files:**
    - Download the files from this repository onto your computer.
    - Place all files in a single folder for easy access.

How to Run the Notebooks

General Instructions for accessing:

Go to the [Google colab website](https://colab.research.google.com) select "Upload" at the right side of the orange menu bar, and select the required `.ipynb` file. This will open up the required file inside of Google's own environment. After running the file, there will be text prompts on the screen.
If you enter the wrong information and get an error, refresh the notebook and try again.

Running DatabaseSearcher.ipynb

1. Open Jupyter Notebook from the Anaconda Navigator.
2. Navigate to the folder where you saved the files.
3. Click on **DatabaseSearcher.ipynb** to open it.
4. Follow the step-by-step instructions provided in the notebook to load and analyze the National Wells Database.

Running CompanySpecificNews.ipynb

1. Open Jupyter Notebook.
2. Click on **CompanySpecificNews.ipynb** to open it.
3. If you have a dataset of news articles, make sure it is in the same folder as the notebook.
4. Follow the steps in the notebook to analyze company-specific news.

Common Issues and Troubleshooting

- **Jupyter Notebook won't open:** Ensure that Anaconda is installed correctly, and try launching Jupyter Notebook again.
    - If using the google collab notebook, try refreshing the page.
- **Missing libraries:** If you see an error about a missing library, make sure you've installed it using the `pip install` command.
- **Dataset not found:** Double-check that the `National Wells Database.xlsx` file is in the same folder as the notebooks.