🚀 Setup Instructions

Follow these steps to set up and run the Data Engineering Project: Real-Time Top News Pipeline & Dashboard (USA & Canada) on your own machine.

1. Clone the Repository

```
git clone https://github.com/your-username/news-dashboard-pipeline.git
cd news-dashboard-pipeline
```

- *(Replace `your-username` with your GitHub username.)*

2. Set Up a Python Virtual Environment

```
python3 -m venv .venv
source .venv/bin/activate
```

3. Install Dependencies

```
pip install -r requirements.txt
```

4. Configure Environment Variables

- Create a `.env` file in the project root (if required).
- Add your NewsAPI key and any other required secrets:

```
NEWSAPI_KEY=your_newsapi_key
```

5. Set Up PostgreSQL Database

- Ensure PostgreSQL is installed and running.
- Create a database named `news_db`:

```
psql -U <your_username> -c "CREATE DATABASE news_db;"
```

- *(Replace `<your_username>` with your local PostgreSQL user, usually your macOS username.)*

6. Run the Pipeline and Dashboard

```
streamlit run news_pipeline_dashboard.py
```

- This will fetch news, clean and load it into PostgreSQL, and launch the interactive dashboard.
- Open the provided local URL in your browser (e.g., http://localhost:8501).

7. Troubleshooting

- If you see errors about missing columns or tables, try dropping the table in PostgreSQL and re-running the script:

```
psql -d news_db
DROP TABLE IF EXISTS news_headlines;
\q
```

- Ensure all environment variables and dependencies are set up correctly.

8. Project Structure

```
news-dashboard-pipeline/
 ├──── news_pipeline_dashboard.py
 ├──── requirements.txt
 ├──── README.md
 └──── .env (not committed)
```