Facebook Data Analysis

This project analyzes social media interactions on Facebook related to major global events and the Ukraine war. The goal is to visualize trends, events, and keywords to provide insights into public engagement over time.

Overview

The project includes:

1. **Total Interactions Timeline**: A line chart showing the total Facebook interactions (reactions, comments, and shares) from 2021 onwards.
2. **Global Events**: Scatter points marking significant global events that might have influenced public engagement.
3. **Ukraine War Events**: Scatter points highlighting key milestones in the Ukraine war.
4. **Combined Timeline**: A consolidated chart that overlays global and Ukraine war events onto the interaction timeline.
5. **Trending Keywords**: A word cloud showcasing frequently mentioned keywords related to the topic.

Features

- **Interactive Timeline**: View public engagement trends with scatter points marking key global and Ukraine-specific events.
- **Tabs for Navigation**: Switch between the original timeline and the combined timeline.
- **Dynamic Tooltips**: Hovering over events provides specific event titles.
- **Word Cloud**: Highlights popular keywords extracted from the data.
- **Python Data Cleaning**: Multiple Python scripts were used to clean, preprocess, and optimize the dataset for analysis.

Technologies Used

- **Frontend**:
    - HTML5
    - CSS3
    - JavaScript
    - Chart.js (for visualizing data)

    - WordCloud2.js (for rendering the word cloud)
- **Backend**:
    - Node.js with Express.js
    - MySQL (Database)
- **Python Scripts**:
    - Used for cleaning, preprocessing, and transforming the raw dataset.
    - Libraries: `pandas`, `numpy`, `re`, `langdetect`, `emoji`

Python Scripts for Data Cleaning

Several Python scripts were used to ensure the data is clean and ready for analysis:

1. **`text_cleaning.py`**:
    - Removed special characters, excessive whitespace, and emojis.
    - Tokenized text to prepare for keyword extraction.
2. **`language_filter.py`**:
    - Used the `langdetect` library to filter out non-relevant languages from the dataset.
3. **`duplicates_remover.py`**:
    - Identified and removed duplicate Facebook posts based on text similarity.
4. **`date_parser.py`**:
    - Standardized all date formats for consistency in analysis.
5. **`keywords_extraction.py`**:
    - Extracted trending keywords from Facebook posts and grouped them by frequency.

Each script contributed to refining the raw data into a usable format, allowing for seamless integration into the MySQL database and further analysis.

Setup Instructions

Follow these steps to set up the project locally:

Prerequisites

- Node.js installed on your machine.
- MySQL server set up with a database containing the required tables.
- Python 3.x installed with libraries: `pandas`, `numpy`, `langdetect`, `emoji`.

1. Clone the Repository

bash
Kopier kode

```
git clone <repository-link>
cd <repository-folder>
```

2. Install Dependencies

Navigate to the project folder and install the necessary packages:

bash
Kopier kode
```
npm install
```

3. Set Up the Database

1. Import the SQL schema for the required tables (`time`, `metrics`, `global_events`, `ukraine_war_events`, `keywords`) into your MySQL database.
2. Update the `.env` file with your database credentials:

plaintext
Kopier kode
```
DBHOST=localhost
DBUSER=root
DBPASSWORD=your_password
DBNAME=facebook_data_analysis
```

3.

Python scripts to clean and prepare raw data: bash

extractor.py finalCleaner.py importcsv.py keywordcleaner.py importpandasaspd.py

4. Run the Server

Start the Node.js server:

bash
Kopier kode
```
node server.js
```

The server will run on `http://localhost:3000`.

5. Open the Application

Open the `front.html` file in a browser or access it through a local server.

Endpoints

The backend provides the following API endpoints:

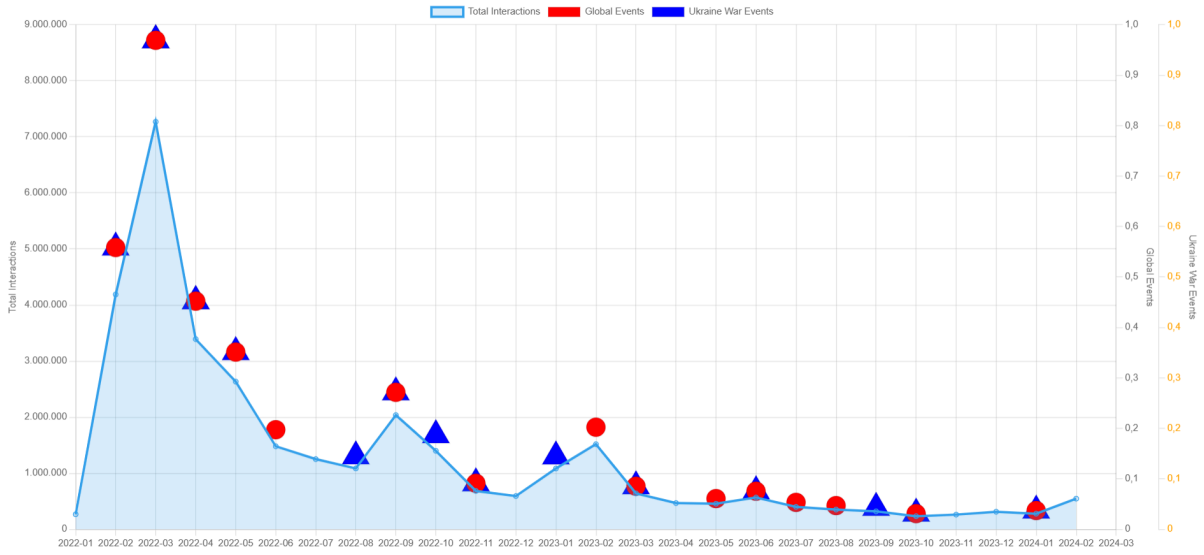
1. **`/api/interactionTimeline`**: Returns Facebook interaction data grouped by month.
2. **`/api/globalEvents`**: Fetches significant global events.
3. **`/api/ukraineWarEvents`**: Fetches major Ukraine war events.
4. **`/api/trendingKeywords`**: Retrieves trending keywords.

Usage

1. **Open the page in a browser**.
2. Switch between tabs:
    - **Original Timeline**: Displays the total interaction trends.
    - **Combined Timeline**: Shows global and Ukraine war events alongside interaction trends.
3. **Hover over scatter points** to see event titles.
4. **Scroll down** to view lists of events and the trending keywords word cloud.

Screenshots

Combined Timeline Chart



Trending Keywords Word Cloud

Future Improvements

- Add real-time data updates for more dynamic insights.
- Include filters for specific events or date ranges.
- Enhance the UI for mobile responsiveness.

License

This project is open-source

Contributors

- Dmikko, Mike
- MathiasHC99, Mathias
- Johanniemann, Johan
- DenStoerste, Esben

Acknowledgments

- **Chart.js**: For creating interactive charts.
- **WordCloud2.js**: For generating the word cloud.
- **Python Libraries**: `pandas`, `numpy`, `langdetect`, `emoji`
- Thanks to everyone who contributed to building and refining this project.