Analysing Social Media Sentiment: Trends in Public Opinion Over Time

Project Overview

This project aims to analyse public sentiment on social media and track how it evolves. By examining large-scale sentiment trends on which the datasets can be accessed, we can identify patterns, correlate them with significant global events, and understand the broader societal impact of online discussions. Our goal is to provide insights into how people react to major occurrences such as political decisions, product launches, Geo politics and social movements in real-time.

Dataset Selection

To conduct this analysis, we are using two primary datasets:

- **Twitter Sentiment Dataset (from Kaggle):** A large dataset containing millions of tweets labelled as positive, negative, or neutral feelings. Covering a wide range of topics, it is ideal for tracking sentiment trends over time.
- **World Events Dataset (e.g., UCI Global Events or another reliable source):** A catalogue of significant global events, including elections, protests, and product launches, which will help in correlating sentiment changes with real-world incidents.

By integrating these datasets, we can gain a better understanding of how public sentiment shifts in response to major events.

Tools and Technologies:

Project 1: Foundational Analysis and Visualizations

- **Tableau** – Used for creating initial visualizations, helping to track sentiment trends over time and map sentiment distribution across different regions.
- **Excel/Google Sheets** – Handy for quick data analysis, generating basic statistical summaries, and identifying initial sentiment patterns.
- **Jupyter Notebook** – Acts as a coding workspace for data cleaning, testing sentiment models, and conducting exploratory analysis in an interactive manner.
- **SQL (PostgreSQL/MySQL)** – Efficiently manages large datasets, allowing seamless storage, retrieval, and structured sentiment analysis.
- **Matplotlib & Seaborn** – Helps create clear and insightful static charts, making it easier to spot trends and patterns in sentiment data before diving into advanced analytics.

Project 2: Advanced Analysis and Interactive Visualizations

Python Libraries for Data Processing & Sentiment Analysis

- **pandas** – Essential for handling data efficiently, including cleaning, merging, and structuring datasets for smooth analysis.
- **nltk & TextBlob** – Used for basic natural language processing (NLP) tasks like sentiment classification and extracting important keywords from text.

- **VADER (Valence Aware Dictionary and Sentiment Reasoner)** – Specifically designed for analyzing sentiment in social media posts, making it ideal for short-form text.
- **spaCy** – A powerful NLP tool for advanced tasks like entity recognition, tokenization, and analyzing linguistic patterns in tweets.
- **Scikit-learn** – Supports machine learning-based sentiment analysis, helping in clustering trends and training models for better classification.
- **Plotly & Altair** – Used for building interactive visualizations, allowing users to explore sentiment data dynamically.
- **NetworkX** – Helps in studying how sentiment spreads across social media, identifying key influencers and opinion leaders.

Visualization & Front-End Technologies

- **D3.js** – A powerful JavaScript library for creating custom, interactive visualizations. It helps users explore sentiment trends over time and understand shifts in public opinion.
- **Flask/Django** – Used for backend development, enabling smooth API integration and seamless connection with sentiment analysis models.
- **Streamlit/Dash** – Makes it easy to build interactive web applications, allowing users to explore sentiment trends with simple and intuitive interfaces.
- **Power BI** – Ideal for business intelligence dashboards, helping visualize sentiment insights in a structured and actionable format.
- **Google Data Studio** – A cloud-based tool that enables real-time reporting and easy sharing of sentiment analysis results in a visually appealing way.

Column	**Tool/Technology**	**How It Helps**
Tweet ID	SQL (PostgreSQL/MySQL)	Efficiently stores and retrieves tweet records, ensuring seamless data management and quick access.
Timestamp	pandas / Matplotlib / Seaborn	Organizes tweets by date and time, enabling trend analysis through visualizations and time-series graphs.
User Location	Tableau / D3.js / Plotly	Maps sentiment trends geographically, offering insights into regional variations in public opinion.
Hashtags	pandas / NetworkX	Identifies trending topics and examines relationships between hashtags, helping to track viral discussions.

Objectives

Project 1: Data Exploration & Visualizations

- Clean data (remove duplicates, handle missing values, and standardize text).
- Perform EDA to analyze sentiment and tweet frequency.
- Create trend graphs and sentiment heatmaps for regional/time-based insights.
- Filter out irrelevant tweets (spam, bots).
- Track tweet activity spikes linked to major events.
- Compare sentiment shifts with global events.
- Calculate sentiment averages over time.
- Identify influential accounts affecting sentiment trends.

Project 2: Advanced Analysis & Visualizations

- Integrate sentiment data with world events for context.
- Use time-series analysis to track sentiment over time.
- Apply machine learning for improved sentiment classification.
- Identify trending topics via sentiment shifts in hashtags/keywords.
- Build interactive visualizations (D3.js) for real-time sentiment exploration.
- Design a dashboard for live sentiment updates.
- Perform network analysis on sentiment propagation.
- Conduct linguistic analysis to extract key sentiment words.
- Automate reports to summarize insights.

Methodology and Implementation

Project 1: Data Exploration and Initial Insights

Stage 1: Understanding the Data

- Explore the Twitter sentiment dataset, focusing on key variables such as tweet text, sentiment labels, and timestamps.
- Generate line graphs to illustrate sentiment trends over time.
- Create heat maps to visualize geographical sentiment distribution.

Stage 2: Data Processing and Correlation Analysis

- Clean and pre-process the dataset by handling missing values and normalizing text data.
- Conduct correlation analysis between sentiment spikes and major global events.
- Example: Analysing sentiment shifts around political speeches, economic announcements, or viral trends.

Project 2: Advanced Analysis and Interactive Visualizations

Stage 3: Integrating Additional Data and Conducting Deeper Analysis

- Merge Twitter sentiment data with the world events dataset to assess sentiment changes before and after major events.
- Use natural language processing (NLP) to extract deeper insights from tweet content.

Stage 4: Developing Advanced Visualizations and Interactive Storytelling

- Build an interactive webpage using D3.js for dynamic sentiment exploration.
- Implement features that enable users to filter sentiment trends by date, event category, and region.

Interactive Webpage and User Experience

Main Page:

- Provides an overview of the project, emphasizing the importance of tracking public sentiment and its real-world impact.
- Highlights key insights from the analysis, such as major trends and geographical sentiment distributions.

Inner Pages:

- Offer detailed visualizations with interactive elements.
- Include interactive timelines, maps, and sentiment charts with filtering options for time and event types.

Key Data for Interactive Webpage

1. Twitter Sentiment Data

Essential Columns:

- **Tweet ID** – Backend use for unique identification.
- **Tweet Text** – Shown in pop-ups or tooltips.
- **Sentiment** – Filter options (Positive, Negative, Neutral).
- **Timestamp** – Allows tracking sentiment trends over time.
- **User Location** – Helps visualize sentiment geographically.
- **Hashtags** – Enables topic-based filtering and trend analysis.

Key Tweets to Highlight:

- Latest tweets tied to major events(Elon musk doge coin).
- Tweets with high engagement (likes, retweets, Celebrity marriage and birthdays).
- Tweets linked to sentiment shifts before, during, or after key events.

2. World Events Data

Essential Columns:

- **Event ID** – For linking sentiment data (not shown to users).
- **Event Title** – Displayed on visualizations.
- **Event Type** – Filters events by category (Politics, Entertainment, Tech, etc.).
- **Start & End Date** – Helps track sentiment before and after events.
- **Location** – Used for mapping and correlation with sentiment.

Key Events to Highlight:

- Events that trigger significant sentiment changes.
- Trending topics with high public discussion.
- Political, entertainment, and tech events with widespread reactions.

3. Interactive Features

- **Time Filter** – Select specific periods for sentiment analysis.
- **Sentiment Toggle** – View only positive, negative, or neutral discussions.
- **Geo Sentiment Map** – Displays sentiment distribution by location.
- **Hashtag Explorer** – Track trends based on hashtags.
- **Event Sentiment Link** – Connects world events to public reactions.

Timeline (12 Weeks)

Project 1 (Weeks 1-7):

- **Weeks 1-3:** Data exploration and cleaning; initial visualizations.
- **Weeks 4-5:** Correlation analysis between sentiment trends and major events.
- **Weeks 6-7:** Summarization and presentation of key insights.

Project 2 (Weeks 8-12):

- **Weeks 8-9:** Merge datasets and refine analysis techniques.
- **Weeks 10-11:** Develop advanced visualizations and interactive storytelling elements.
- **Week 12:** Deploy the final interactive webpage and document key findings.

Deliverables

Project 1:

- Cleaned dataset and exploratory analysis.
- Basic visualizations of sentiment trends and geographic sentiment distributions.
- Initial correlation analysis results.

Project 2:

- Integrated datasets and advanced sentiment analysis.
- Interactive visualizations for dynamic sentiment exploration.
- A fully developed interactive webpage for data storytelling.

Evaluation Criteria

- **Clarity:** Well-documented approach, clear explanations, and meaningful visualizations.
- **Feasibility:** Practical implementation of tools and adherence to the project timeline.
- **Creativity:** Effective use of interactivity and innovative visualization techniques.
- **Technical Proficiency:** Skilled utilization of Tableau, Python, and D3.js for insightful analysis.
- **Data Relevance:** Quality and reliability of datasets used in the project.

This project provides an in-depth exploration of social media sentiment, offering valuable insights into public opinion dynamics. Through interactive visualizations, users can engage with the data first hand, uncovering trends and correlations intuitively and compellingly.

1. Sample Data Tables

Below are example structures for the two datasets you'll be using: the **Twitter Sentiment Dataset** and the **World Events Dataset**.

1.1. Twitter Sentiment Dataset (Example)

Link for twitter dataset [Twitter Sentiment Database]

Tweet ID	Tweet Text	Sentiment	Timestamp	User Location	Hashtags
12345	"Loving the new product release!"	Positive	2025-01-10 12:34:56	New York, USA	#newproduct
12346	"Feeling disappointed with the service."	Negative	2025-01-10 13:00:00	Los Angeles, USA	#disappointed
12347	"Can't wait for the concert tonight!"	Positive	2025-01-11 08:30:21	London, UK	#concert
12348	"Politics are just getting worse."	Negative	2025-01-11 14:15:45	Berlin, Germany	#politics
12349	"Such a great vacation destination!"	Positive	2025-01-12 09:45:30	Paris, France	#vacation

Columns Explanation:

- **Tweet ID**: Unique identifier for each tweet.
- **Tweet Text**: The content of the tweet.
- **Sentiment**: Sentiment classification (Positive, Negative, Neutral).
- **Timestamp**: Date and time when the tweet was posted.
- **User Location**: The geographical location of the user posting the tweet.
- **Hashtags**: Any hashtags included in the tweet, which can be used to categorize topics.

1.2. World Events Dataset (Example)

Link for the dataset events datasets

Event ID	Event Title	Event Type	Start Date	End Date	Location
001	"Product Launch: New iPhone"	Technology	2025-01-10	2025-01-10	Worldwide
002	"Global Election: USA Presidential Race"	Politics	2025-01-11	2025-01-11	United States
003	"World Music Concert"	Entertainment	2025-01-12	2025-01-12	Global
004	"Political Crisis in Venezuela"	Politics	2025-01-12	2025-01-14	Venezuela

Columns Explanation:

- **Event ID**: Unique identifier for each world event.
- **Event Title**: A brief description or title of the event.
- **Event Type**: Categorization of the event (e.g., Technology, Politics, Entertainment).
- **Start Date**: The start date of the event.
- **End Date**: The end date of the event (if applicable).
- **Location**: The geographical location of the event.

2. Workflow Diagram

Below is a simplified **workflow diagram** illustrating the stages of the project we will follow from data collection to the creation of the final interactive visualization. The workflow outlines how data will flow through the system and how each stage of the project fits together.