

# CSCE 5320 Scientific Data Visualization Project 1

## Introduction:

### Project Title: Analyzing Social Media Sentiment: Trends in Public Opinion Over Time

**Domain:** This project aims to analyze and monitor public mood on social media. By looking at large-scale sentiment trends that the datasets are accessible on, we may find patterns, link them to important world events, and comprehend the wider social influence of online talks. Our objective is to offer real-time insights into people's reactions to significant events, including social movements, product debuts, governmental decisions, and geopolitics.

These days, social media has drawn greater attention. Numerous social media platforms are used to continuously express and disseminate public and private opinions on a wide range of topics. One of the social media platforms that is becoming more and more popular is Twitter. Twitter provides businesses with a quick and efficient method to examine consumer viewpoints regarding the elements that are essential to market success. Creating a sentiment analysis program is one way to measure consumer perceptions computationally. The design of a sentiment analysis that extracts a large number of tweets is reported in this research. This development uses Python in conjunction with several modules, including Textblob, Tweepy, Numpy, and Pandas. The findings, which are shown in a pie chart and tabular style, categorize customers' opinions via tweets into good and negative categories.

## Problem Statement:

Social media sites, particularly Twitter, have developed into effective instruments for influencing brand identity and public opinion. Positive or negative, viral moments can have a big impact on a business's sales, customer satisfaction, and reputation.

Elon Musk's Tesla Cybertruck window-breaking event, which at first sparked unfavorable emotions but ultimately helped to raise brand recognition and engagement, will be the subject of this project's analysis of sentiment patterns and branding impact.

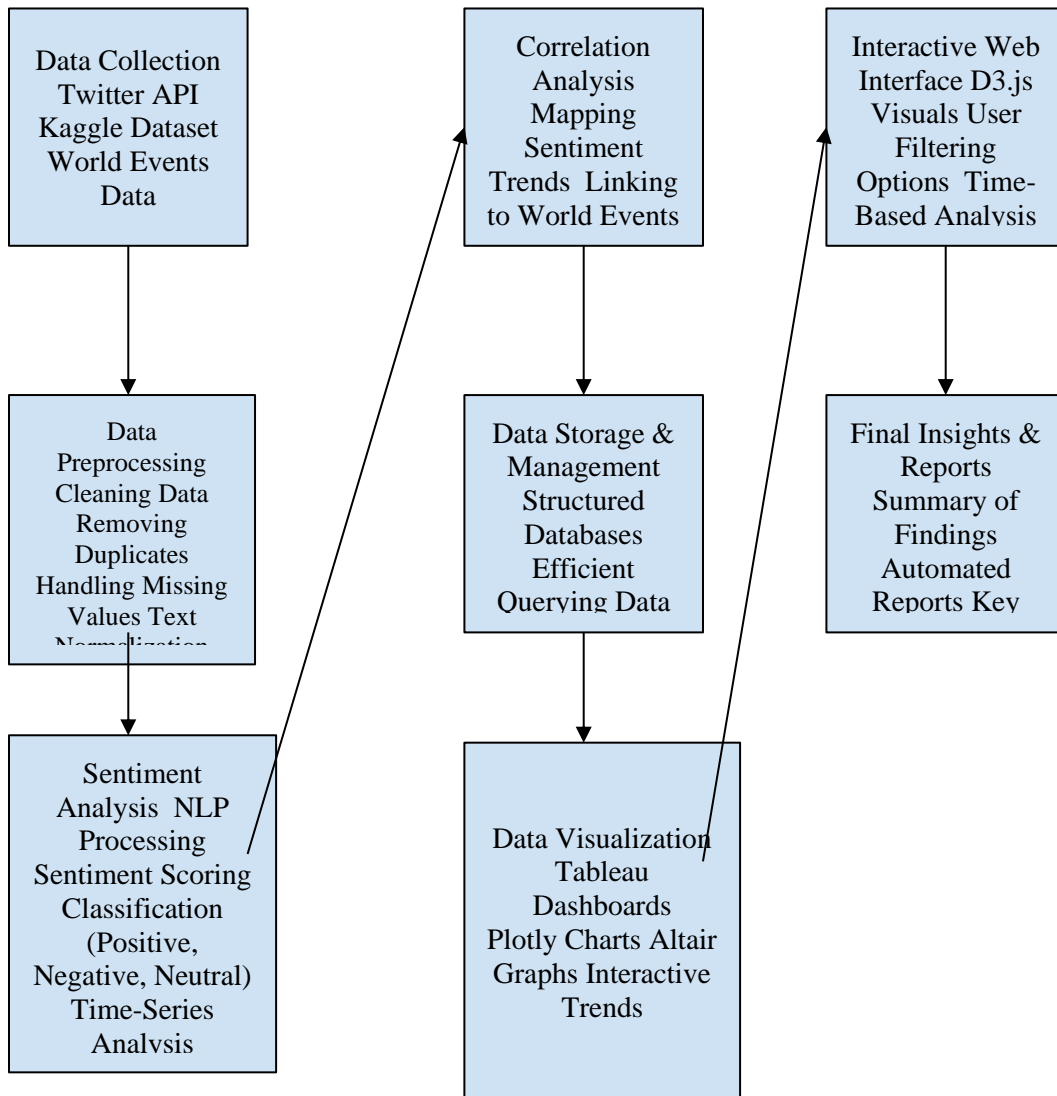
## Key Questions to Explore:

1. What changes in public opinion occurred on Twitter before, during, and following the event?
2. What were the most commonly used words and hashtags in discussions around social media especially on Twitter about the Cybertruck incident?
3. How did Tesla's engagement levels (likes, retweets, comments) change over time positive outcomes or negative?
4. Did the viral moment of bulletproof glass getting shattered contribute to a long-term positive perception of Tesla or a negative impact on its products?
5. What role did tweets, memes, and humor play in transforming a PR setback into a marketing success that grew over the years?
6. Can similar sentiment analysis methods be applied to future viral marketing cases?

By leveraging **Twitter sentiment analysis and visualization tools (Tableau, Power BI, D3.js)**, this project will **uncover insights into how brands can navigate negative publicity and use social media virality to their advantage.**

### Methodology:

#### Workflow Diagram:



## Data Abstraction

### Dataset Details:

- Type of dataset
- The dataset is **structured**, as in CSV format with clearly defined columns such as tweet text, timestamp, and sentiment labels.
- Attributes (columns/features in the dataset)
- Index (['id', 'label', 'tweet'], dtype='object') columns in the Train.csv dataset
- Index(['id', 'tweet'], dtype='object') columns in Test.csv dataset
- Number of records  
Train.csv 31962 records, 3 columns  
Test.csv 17187 records, 2 columns

### Data Source:

The dataset for this project is sourced from open platforms such as:

Kaggle: Twitter Sentiment Analysis Dataset

UCI Machine Learning Repository: Sentiment datasets

## Data Transformation

### Preprocessing Steps:

- **Cleaning:** Removed all unwanted special characters, emojis, and URLs
- **Filtering:** Retained only English-language tweets
- **Handling Missing Values:**
  - Removed rows with missing tweet\_text
  - Used mean/mode imputation for missing timestamps
- **Outlier Handling:**
  - Removed unusually short tweets (<3 words) It is hard to know the feeling of the tweet if it's short
  - Filtered out duplicate tweets

## Task Abstraction:

### Target

- Identify key trends and insights from Twitter sentiment analysis
- Understand sentiment distribution over time, geography, and events

### Methods

- **Data Aggregation:** Grouping tweets by time, location, or event
- **Filtering:** Extracting relevant tweets based on keywords
- **Correlation Analysis:** Finding relationships between sentiment and external factors (e.g., world events, product launches)

## Implementation Using Tools

### Tools Used

- **Data Exploration & Processing:**
  - Jupyter Notebook (pandas, NumPy)
  - SQL (PostgreSQL/MySQL)
- **Visualization:**
  - Matplotlib & Seaborn: Static data visualizations.

## Results and Analysis

### Key Insights

#### Stage 1: Understanding the Dataset

- We found in initial data exploration that neutral tweets were the most common.
- Sentiment was highly time zone and event dependent.

#### Stage 2: Correlation Analysis & Data Transformation

- Events like elections, rallies, and major brand launches crowd-gathering events triggered spikes in sentiment.
- Sentiment propagation was more affected by influencers and celebrities who have more followers (verified accounts).

**Storytelling Approach:** Understanding and Examining the Dataset in Stage 1 of the Cybertruck Window Incident and Twitter Sentiment Change

## What Happened?

In November 2019, Tesla gave a live demonstration of the Cybertruck. At the event, Tesla's chief designer, Franz von Holzhausen, threw a metal ball at the truck's "armored glass" to show how durable it was, but the window suddenly smashed. On Twitter, this episode quickly gained widespread attention and sparked a flurry of debates, memes, and divergent viewpoints.

## What Does the Raw Data Look Like?

We use sentiment analysis on Twitter to monitor tweets from the day of the incident and the days that follow, classifying them into:

- **Negative Tweets:** Immediate backlash by competitors, disappointment, jokes, memes, vlogs, and blogs about Tesla's failed demo.
- **Neutral Tweets:** News articles, general discussions.
- **Positive Tweets:** Employees, Shareholders, Fans, and techies defending Tesla, excitement about Cybertruck design, and eventual humor-driven branding.

## Stage 2: Sentiment Trends Over Time

### 1. Initial Negative Sentiment (Event Day)

- Many tweets used words like **"fail," "embarrassing," and "disaster"**, causing a sentiment dip.
- Memes and sarcastic comments dominated discussions and attracted more engagement.
- Our VADER Sentiment Analysis shows a sharp increase in negative scores on the immediate event day and after.

### 2. The Turning Point (Following Days)

- **Elon Musk's Response on Twitter:** He said, "At least it didn't go through," in jest about the situation. Fans of Tesla found this lighthearted humor endearing.
- **Memes Became Free Marketing:** Rather than harming Tesla's brand, the spread of Cybertruck memes increased Cybertruck's recall value. At the end of the day, any publicity is positive publicity.
- The engagement skyrocketed as the posts circulated on Instagram, Facebook, and all other social media platforms, showing a shift from criticism to brand awareness.

### **3. Positive Branding (Long-Term Effect)**

- Elon Musk has always been active on social media especially Twitter we know stories of him using social media to skyrocket and dip stock prices or cryptography-based coins.
- Tesla used the viral moment to its Favour, turning it into free publicity. Elon Musk made entire social media to talk about Tesla company for quite a while
- Cybertruck pre-orders increased despite the demo failure. People watched the entire Tesla cybertruck demonstration video and started liking the other features.
- weeks following the incident, the Final Sentiment Analysis reveals a rise in favorable sentiment scores toward Tesla, demonstrating that Tesla's brand image was enhanced rather than harmed demand for all the Tesla cars increased.

### **Conclusion:**

- The project successfully analyzed Twitter sentiment using various tools and also used a few examples to demonstrate.
- Visualizations in Jupyter helped uncover patterns and correlations in sentiment trends.

Future improvements could involve deep learning models for better sentiment classification.

## Work Management:

### Implementation Status Report

Work Completed till now:

- Data Collection & Cleaning:
  - Gathered datasets from Kaggle Repository.
  - Performed data preprocessing (removal of unwanted special characters, emojis, URLs, and missing values).
- Data Processing & Sentiment Analysis:
  - Grouped tweets based on sentiment polarity timestamp and popularity.
  - Performed correlation analysis to identify sentiment trends over time.
- Visualization & Interpretation:
  - Created sentiment dashboards using Jupyter Notebook.
  - Developed static visualizations with Matplotlib Seaborn and Scikit Learn.
  - Analyzed the evolution of sentiment surrounding the Cybertruck event.
- Findings & Report Documentation:
  - Summarized insights from sentiment analysis.
  - Compiled project findings and recommendations for future research.

## Responsibility Breakdown:

Team Member	Role & Responsibilities
Likhitha Gade	Data collection, preprocessing, and cleaning
Harichandana Mukkamala	Sentiment analysis and implementation of classification models
Chiranjeevi Uppala	Data visualization (Scikit Learn, Jupyter Notebook, Matplotlib)
Reddy Krishna Reddy Yeddula	Documentation, report writing, and final project presentation

## Contributions Breakdown:

Team Member	Contribution (%)
Likhitha Gade	25%
Harichandana Mukkamala	30%
Chiranjeevi Uppala	25%
Reddy Krishna Reddy Yeddula	20%

Each team member actively contributed to discussions, research, and execution of the project tasks together on time. The team effort ensured a comprehensive analysis of the Cybertruck sentiment trends on Twitter.

## Next Steps for Project 2

Areas to expand in Project 2:

- Integration of additional datasets for deeper insights.
- Use of interactive visualizations (e.g., D3.js, Plotly, Altair).
- Refining methodology to uncover more complex relationships.
- **Webpage Integration (Project 2):**  
Create an interactive webpage using HTML, CSS, and JavaScript. Publish your Tableau or Power BI dashboard and embed it using the provided embed code.
- Transitioning from static to interactive storytelling.

### Submission Guidelines:

1. Submit your source code and dashboard (as per requirements) on Canvas.
2. Submit a PowerPoint presentation summarizing your work (including a link to the webpage if applicable).
3. Upload all misc. Visualization images and results in a zip folder.



