# Task 4: Sentiment140 – Category-wise Sentiment Utilization Summary

## Objective:

To perform sentiment analysis on tweets using the Sentiment140 dataset, analyzing distribution, emotional tone, and language patterns across sentiment categories.

## Dataset Overview:

The dataset includes 1.6 million labeled tweets, categorized into positive, negative, or neutral sentiment. For efficiency, a random 100,000 tweet sample was used.

## Analysis Conducted:

- Cleaned tweet text to remove URLs, mentions, hashtags, and punctuation.  
- Mapped sentiment targets to readable labels (Positive, Negative, Neutral).  
- Visualized sentiment distribution and tweet length by sentiment.  
- Generated WordClouds for each sentiment.  
- Extracted and analyzed most common hashtags.  
- Detected emotion categories using a basic emotion lexicon.

## Category-wise Summary:

Each sentiment category revealed unique language patterns and tweet lengths. Emotion detection showed that Joy and Anger were the most prominent emotions across the sample.  
  
Word clouds and bar plots further helped visualize key terms and emotional trends across sentiment labels.

## Conclusion:

The sentiment and emotion analysis provides a high-level overview of public emotional tone on Twitter. These findings can assist in refining models for emotion-aware NLP systems or social media monitoring.