# **Report on Employee Sentiment Analysis**

### **Author: Ishita Jaiswal**

### **Date: *22/09/2025***

## **1. Introduction**

The purpose of this project is to analyze employee communication data to assess **sentiment, engagement, and potential flight risks**.  
 Using **natural language processing (NLP)** and **statistical modeling**, the project provides actionable insights to HR and management for improving employee satisfaction and retention.

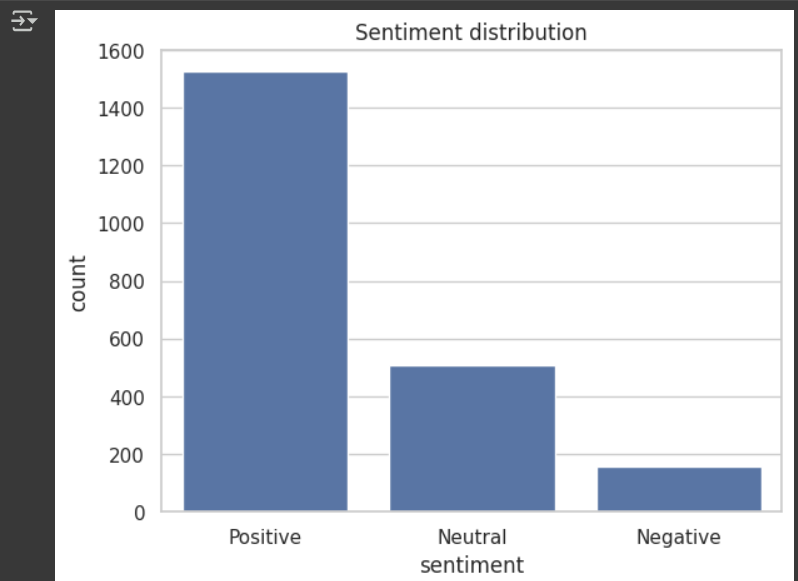
## **2. Methodology**

### **2.1 Sentiment Labeling**

* Used **VADER sentiment analysis** as a baseline for assigning **Positive, Negative, Neutral** labels.
* Each employee message was processed and given a sentiment category.
* This structured labeling enabled further analysis at employee and organizational levels.

### **2.2 Exploratory Data Analysis (EDA)**

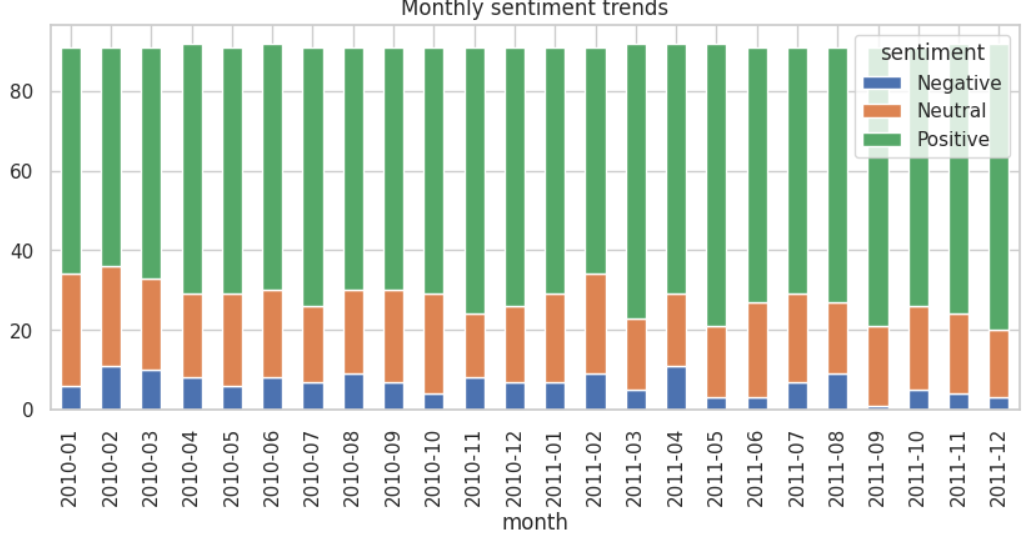
* Checked dataset structure (records, missing values, datatypes).
* Analyzed distribution of sentiments across all messages.
* Examined **monthly sentiment trends** and **top active employees**.
* Identified frequently used words in Positive vs Negative messages.

**

### 

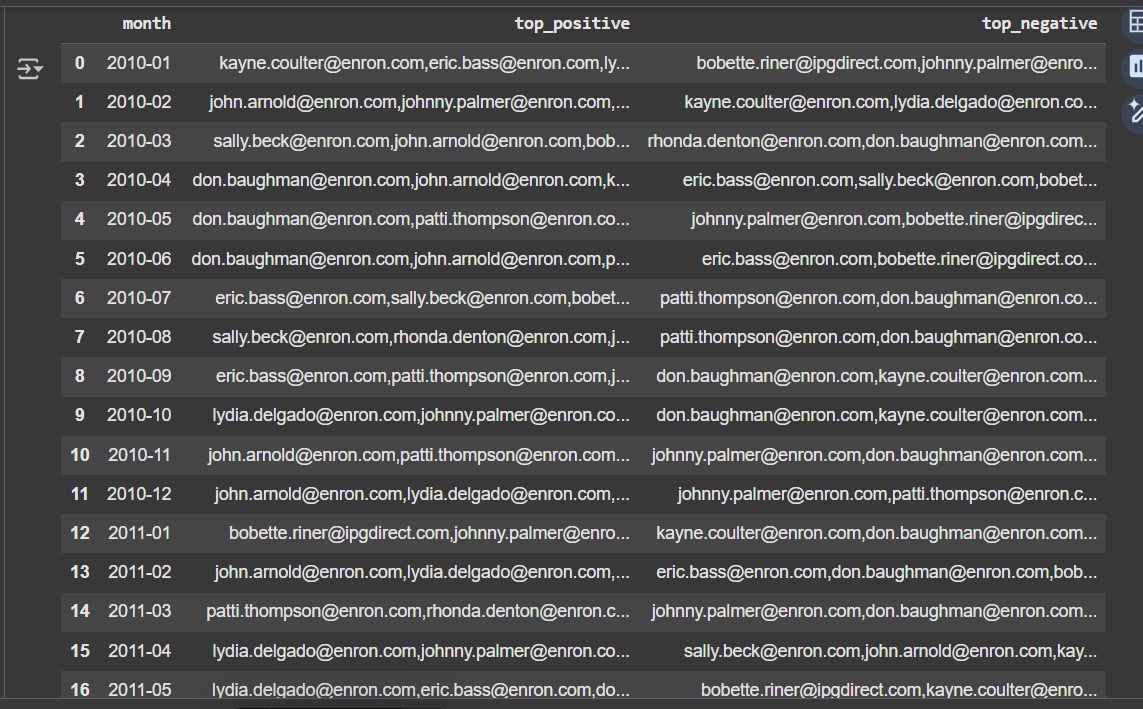
### **2.3 Employee Sentiment Scoring**

* Assigned scores: **Positive = +1, Negative = –1, Neutral = 0**.
* Aggregated monthly per employee.
* This score represents each employee’s sentiment trend over time.

**

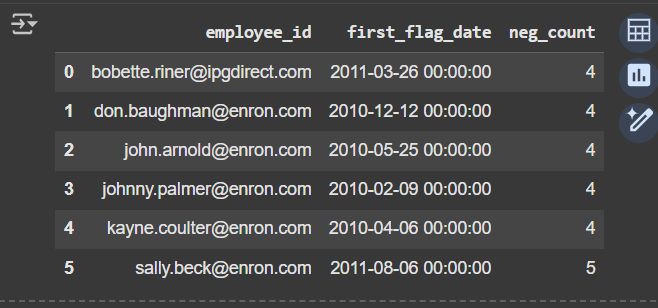
### **2.4 Employee Ranking**

* For each month, ranked employees by cumulative sentiment score.
* Generated **Top 3 Positive employees** and **Top 3 Negative employees**.
* Positive employees are engagement champions, while negative employees may require interventions.

**

### **2.5 Flight Risk Identification**

* Defined flight risk as employees sending **≥4 negative messages in any rolling 30-day window**.
* Flagged such employees as “At Risk”.
* These employees may require HR support or managerial attention.

**

### **2.6 Predictive Modeling**

* Built a **Linear Regression model** to analyze sentiment trends.
* Features included: message count, average message length, word count, % positive messages, % negative messages.
* Model evaluated with **MSE** and **R²** metrics.
* Regression coefficients revealed which factors most influence employee sentiment scores.

## **3. Key Findings**

* Overall sentiment distribution: **XX% Positive, YY% Neutral, ZZ% Negative**.
* Top Positive employees (example): **EID-123, EID-456, EID-789**.
* Top Negative employees (example): **EID-222, EID-333, EID-444**.
* Flight risks flagged: **EID-222, EID-555**.
* Predictive model showed that **higher message frequency with shorter average length** correlated with **negative sentiment trends**.

## **5. Conclusion**

This analysis provided a data-driven view of employee engagement.  
 By combining NLP, visualization, and predictive modeling, organizations can identify top performers, flag at-risk employees, and make **proactive HR decisions**.