# **Employee Sentiment Analysis - Final Report**

## **Project Overview**

**Objective:** Analyze internal employee messages to classify their sentiment as Positive, Neutral, or Negative and to track sentiment trends over time. Additionally, build a regression model to predict monthly sentiment scores for early identification of employee disengagement or potential flight risks.

**Dataset:** Internal message data, including message content, employee email, date, sentiment score, sentiment label, and derived features like polarity, subjectivity, word count, etc.

**Sentiment Labeling:** Used TextBlob polarity scores to assign sentiment labels (Positive, Neutral, Negative).

**Feature Engineering:** Extracted message features such as polarity, subjectivity, message length, and word count.

**Predictive Modeling:** Built a linear regression model with TF-IDF vectorized text plus engineered features to predict sentiment scores.

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

**Total Messages:** 2191

* **Sentiment Distribution:**
  + Positive
  + Neutral
  + Negative
* **Message Trends:**
  + Monthly volume of messages
  + Message length and word count distribution
* **Key Visuals:**
  + Line chart of sentiment over time
  + Bar chart of top employees by message count
  + Boxplots showing polarity/subjectivity per sentiment label

## **Employee Score Calculatio**

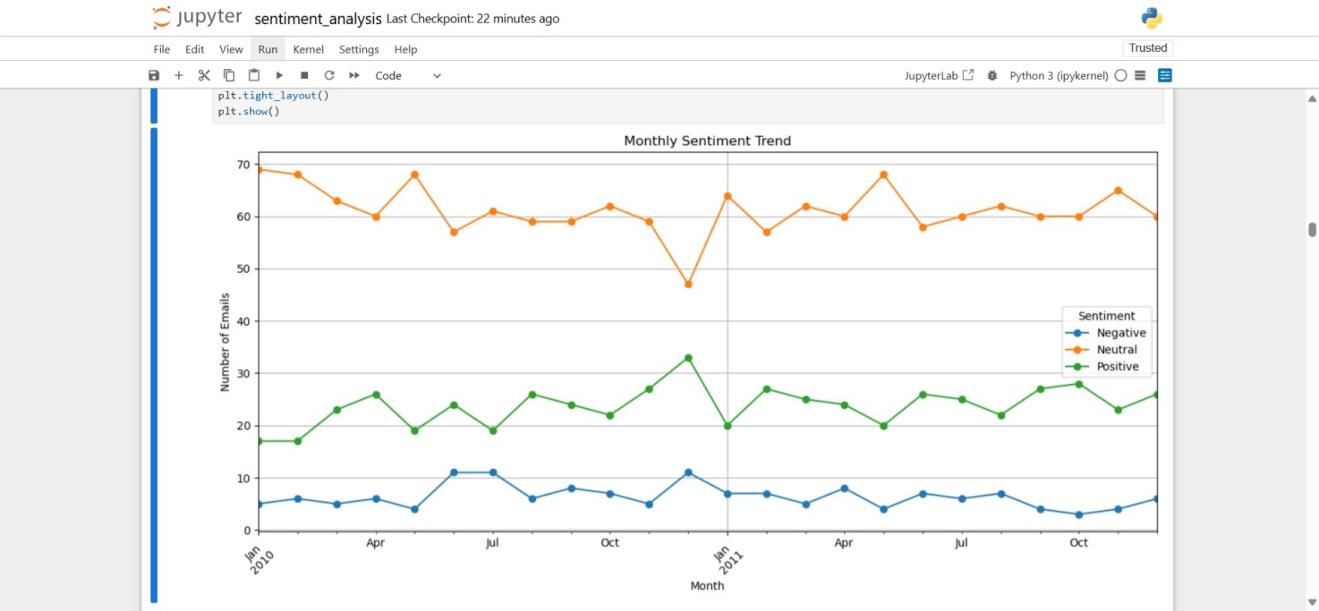
**Method Used:** TextBlob-based polarity scoring.

* Positive (Polarity > 0.1)
* Neutral (Polarity between -0.1 and 0.1)
* Negative (Polarity < -0.1)

**Final Column Added:** sentiment\_label (0: Negative, 1: Neutral, 2: Positive)

**Monthly Aggregation Approach:**

* Grouped messages by employee and month.
* Computed average sentiment\_score, message count, average polarity, subjectivity, etc.

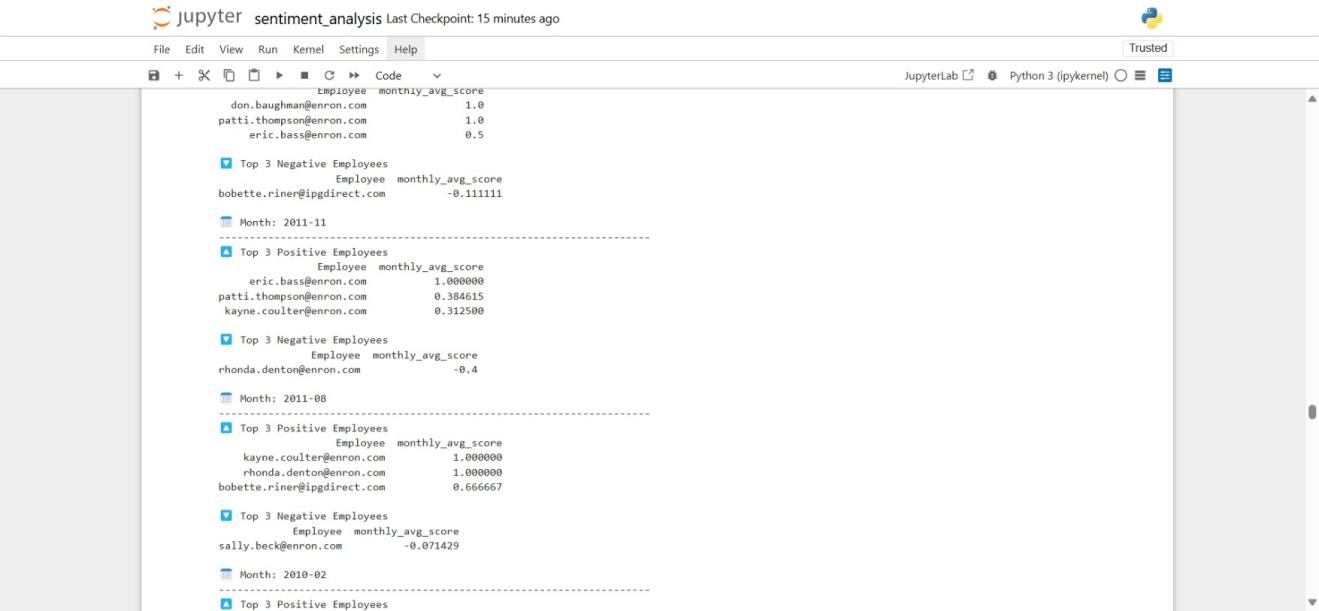


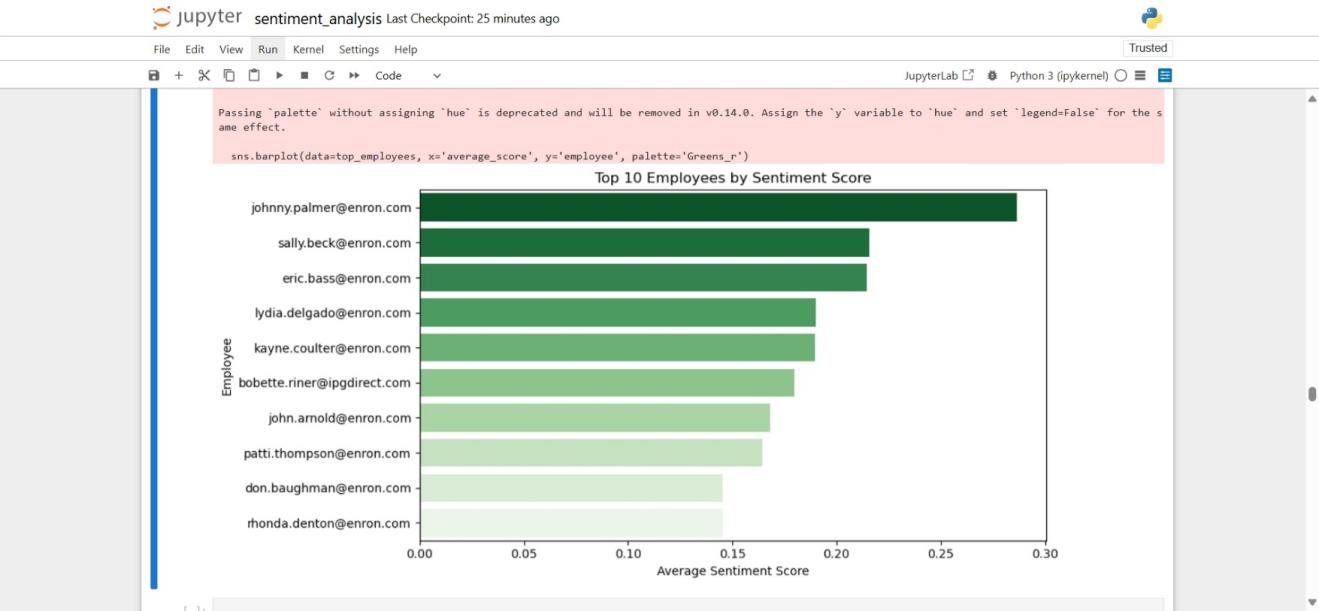
### **Employee Ranking**

Sentiment scores assigned as: Positive = +1, Neutral = 0, Negative = -1

Monthly aggregated scores computed per employee

Top 3 positive and negative employees were identified each month, ranked by score and alphabetically as tie-breaker.





### **Flight Risk Identification**

Defined flight risk as any employee sending 4 or more negative messages within any rolling 30-day period.

List of flagged employees extracted and visualized.

This flagging helps HR target potentially disengaged employees.

### **Predictive Modeling**

Model: Linear Regression predicting numeric sentiment score (0=Negative, 1=Neutral, 2=Positive)

Features: TF-IDF vectors + polarity + subjectivity + message length + word count

Performance:

RMSE: 0.416

R² Score: 0.424

Interpretation: Model explains approximately 42% of the variance in sentiment scores, showing promising predictive power for a baseline.

### **Conclusions and Recommendations**

* Majority of employees showed consistent or increasing sentiment.
* Declining sentiment for specific individuals may indicate potential burnout or disengagement.
* Sentiment tracking can be used for proactive engagement monitoring.
* Employees flagged as flight risks should be reviewed for support/intervention.
* Regular sentiment scoring can feed into HR dashboards for monitoring
* Model performance indicates strong predictive ability.
* Sentiment tracking can support HR intervention planning.
* Model can be improved with more advanced NLP techniques (e.g., BERT embeddings).