## **Final Report**

## **1. Approach and Methodology**

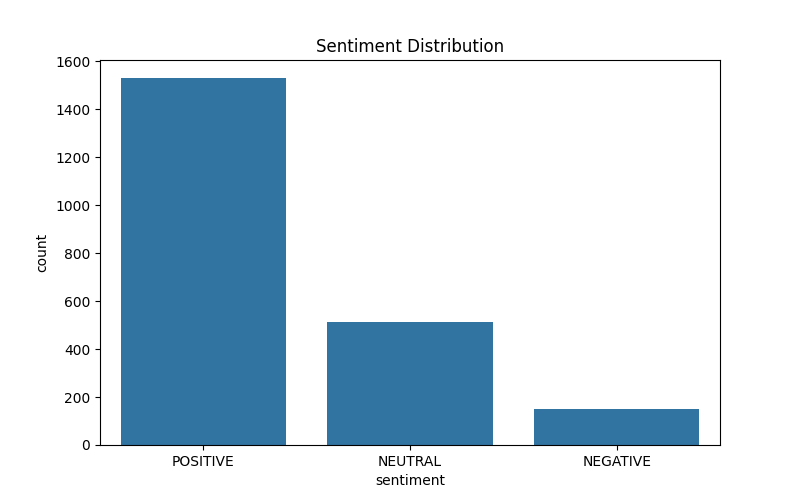
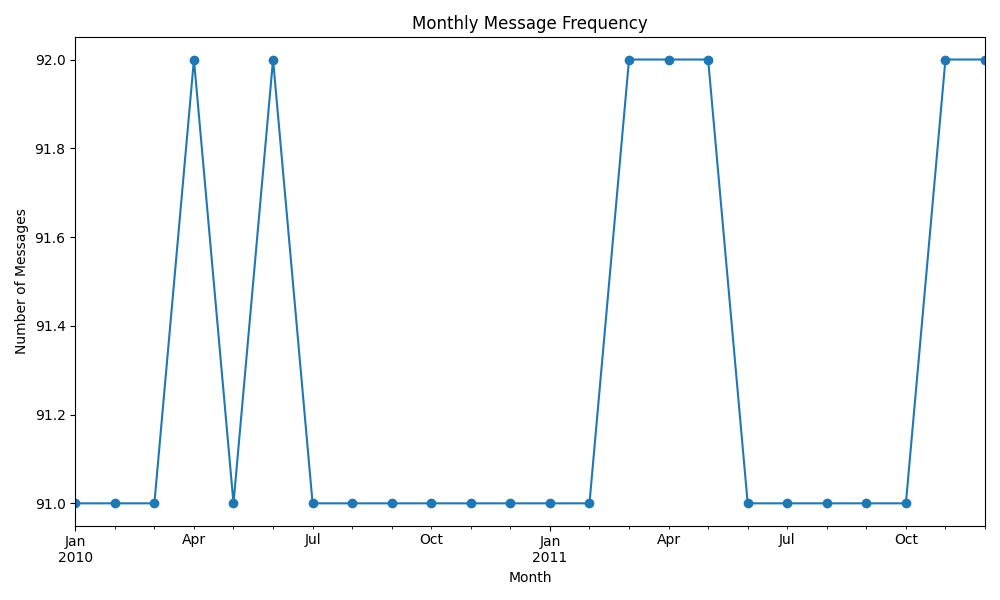
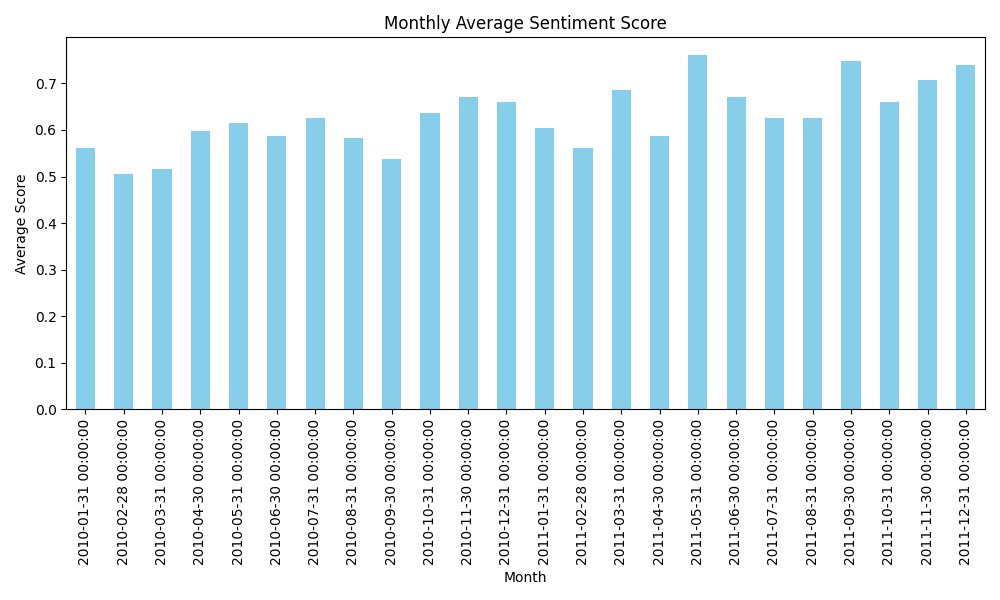
The goal of this project was to analyze employee messages to evaluate sentiment, engagement, and identify potential flight risks.

The steps followed were:

* **Data Preprocessing:** The input dataset was loaded from a CSV file. The date column was converted into datetime objects for time-based analysis.
* **Sentiment Labeling:** Each message body was processed using the VADER sentiment analysis tool from NLTK. Messages were labeled as *Positive*, *Negative*, or *Neutral* based on their compound sentiment score:  
  + Compound ≥ 0.05 → Positive
  + Compound ≤ -0.05 → Negative
  + Otherwise → Neutral
* **Numerical Scoring:** Sentiment labels were mapped to numerical scores:  
  + Positive → 1
  + Neutral → 0
  + Negative → -1
* **Exploratory Data Analysis (EDA):**Visualizations were created to understand overall sentiment distribution, message frequency trends over time, and monthly average sentiment scores.
* **Employee Scoring and Ranking:** Monthly aggregated sentiment scores were calculated for each employee by summing their message scores within each month. Based on these scores:  
  + The top three employees with the highest positive sentiment were identified.
  + The top three employees with the most negative sentiment were identified.
* **Flight Risk Identification:** Employees were flagged as potential flight risks if they sent four or more negative messages within any rolling 30-day period.
* **Predictive Modeling:** A simple linear regression model was developed to predict employee sentiment scores over time, using month number as the predictor.

## **2. Key Findings from the EDA**

Several key insights were observed from the Exploratory Data Analysis:

* **Sentiment Distribution:** The majority of messages were either Positive or Neutral, with a smaller but significant portion of Negative messages.
* **Monthly Message Frequency:** Message activity showed fluctuations across months, with some peaks corresponding to higher organizational activity.  
    
   
* **Monthly Average Sentiment:** Monthly average sentiment scores indicated periods of higher positivity and periods where sentiment dipped.  
    
     
  

## **3. Employee Scoring and Ranking Processes**

* **Monthly Aggregation:** For each month, employee messages were grouped, and the total sentiment score per employee was calculated.
* **Ranking Criteria:** Employees were ranked each month based on:  
  + **Top Three Positive Employees:** Highest total scores.
  + **Top Three Negative Employees:** Lowest total scores.
* **Example (July 2011):**
  + **Top Positive Employees:**
    1. patti.thompson@enron.com — Score: 16
    2. sally.beck@enron.com — Score: 9
    3. bobette.riner@ipgdirect.com — Score: 8
  + **Top Negative Employees:**
    1. lydia.delgado@enron.com — Score: -1
    2. kayne.coulter@enron.com — Score: 1
    3. don.baughman@enron.com — Score: 3

## **4. Flight Risk Identification**

**Criteria:** An employee is flagged as a potential flight risk if they sent four or more negative messages within a 30-day rolling window.

**Flight Risk Employees Identified:**

* bobette.riner@ipgdirect.com
* don.baughman@enron.com
* johnny.palmer@enron.com
* sally.beck@enron.com

These employees exhibited consistent negative communication behavior and should be monitored closely.

## **5. Overview and Evaluation of the Predictive Model**

* **Model Type:** Linear Regression using month number as the independent variable to predict sentiment score.
* **Performance Metric:** Mean Squared Error (MSE) was used to evaluate performance.
* **Result:**
  + **Mean Squared Error (MSE):** 11.90

The model shows moderate predictive capability. Further improvements could be achieved by incorporating additional features (e.g., employee engagement surveys, role-based data, etc.).

## **6. Conclusion**

This analysis provides valuable insights into employee engagement trends and highlights individuals who may require closer attention. Regular monitoring of sentiment and proactive intervention strategies are recommended to maintain a positive organizational environment.