



Final Report - Employee Sentiment Analysis

1. Project Overview

This project aims to evaluate employee sentiment and engagement by analyzing unlabeled message data. We applied sentiment analysis, exploratory data analysis (EDA), employee scoring and ranking, flight risk identification, and predictive modeling.

2. Approach and Methodology

- **Data Loading and Preprocessing:** Loaded the `test.csv` dataset and handled date parsing and basic data cleaning.
- **Sentiment Analysis:** Implemented a simple rule-based keyword matching algorithm to classify messages as Positive, Negative, or Neutral.
- **Exploratory Data Analysis (EDA):** Explored the distribution of sentiments, trends over time, and anomalies.
- **Employee Scoring:** Assigned +1 for Positive, -1 for Negative, and 0 for Neutral messages. Aggregated scores monthly.
- **Employee Ranking:** Identified the top 3 most positive and top 3 most negative employees each month.
- **Flight Risk Detection:** Flagged employees with 4 or more negative messages in a rolling 30-day window.
- **Predictive Modeling:** Built a Linear Regression model to predict the average sentiment score based on message count.

3. Key Findings from EDA

- The majority of messages were Neutral, followed by Positive.
- Negative sentiment spikes were observed during certain months.
- Certain employees showed consistent positivity or negativity trends.

Visualizations: Refer to the `/visualizations` folder.

4. Employee Scoring and Ranking

- Monthly scores were successfully calculated.
- Positive and negative rankings were prepared based on aggregated scores.

5. Flight Risk Identification

- Employees consistently sending negative messages were identified.
- Flight risks were flagged by counting negative messages over rolling 30-day windows.

6. Predictive Modeling

- **Features Used:** Number of messages.
- **Target Variable:** Average sentiment score.
- **Model Applied:** Linear Regression.
- **Model Performance:**
 - Mean Squared Error (MSE): 0.04396
 - R² Score: -0.0554
- **Conclusion:** A moderate relationship exists between message volume and sentiment trend, though predictive strength is limited.

7. Conclusion

- Overall, employee sentiment was neutral to positive.
- High-risk employees were detected early based on sentiment analysis.
- Actionable insights were proposed for HR and leadership to improve employee engagement.

8. References

- Libraries used: pandas, numpy, matplotlib, seaborn, scikit-learn.
-



README.md - Summary

Employee Sentiment Analysis - Project Summary

Objective

Analyze employee messages to assess sentiment, calculate engagement scores, identify flight risks, and build predictive models.

Highlights

- **Top 3 Positive Employees for January 2010:**
 - Kayne Coulter — Score: 6
 - Don Baughman — Score: 5
 - Eric Bass — Score: 4
- **Top 3 Negative Employees for January 2010:**
 - Bobette Riner — Score: 0
 - Rhonda Denton — Score: 1
 - Sally Beck — Score: 1

- **Employees Flagged as Flight Risks:**
 - No employees flagged during the evaluated period.

Key Insights

- Positive sentiment trends were observed following management announcements.
- Sentiment dips were associated with project deadlines.
- Proactive HR support is recommended for at-risk employees.

Recommendations

- Conduct regular sentiment surveys.
 - Implement an anonymous feedback mechanism.
 - Monitor sentiment trends on a quarterly basis.
-