**ASSIGNMENT ANALYZING DATA REPORT**

**Introduction**

The objective of this project was to analyze feedback data collected from various sources (Teachers, Facilitators, and M&E) to understand the sentiments expressed in the feedback and categorize the responses using clustering techniques. The analysis involved preprocessing the text data, conducting sentiment analysis, and visualizing the feedback data to derive actionable insights. The feedback categories ranged from 0 to 5, indicating varying levels of engagement and response from participants.

Additionally, this analysis was conducted using both Python and Power BI to compare and enhance the insights derived from the feedback data.

**Data and Methodology**

**Data**

The dataset comprised 27 feedback entries, each annotated with the source of the feedback (Teacher, Facilitator, M&E) and a category ranging from 0 to 5. The categories were labeled as follows:

0: Nobody

1: Very Few

2: Half

3: Most

4: All

5: Not Sure

**Methodology**

The project was carried out in several key steps:

**1. \*\*Data Preprocessing\*\*:**

- The feedback data was converted into a pandas DataFrame for structured analysis.

- Numeric category values were mapped to descriptive labels for better interpretability.

**2. \*\*Sentiment Analysis\*\*:**

- Defined sets of positive and negative words to evaluate the sentiment of the feedback.

- A function was created to tokenize the feedback text and calculate the counts of positive and negative words.

- Feedback was classified into Positive, Negative, or Neutral sentiments based on the counts of these words.

**3. \*\*Visualization\*\*:**

- Various plots were created using seaborn and matplotlib to visualize the distribution of feedback categories, sentiment distribution, and other relevant metrics.

- Power BI was used to create interactive dashboards for a more dynamic and user-friendly analysis.

**Detailed Steps**

* **Data Preprocessing**

The initial step involved converting the feedback data into a structured format using a pandas DataFrame. The feedback entries were then mapped to their respective category labels. This step ensured that the data was in a format suitable for analysis and visualization.

* **Sentiment Analysis**

Sentiment analysis was performed using predefined sets of positive and negative words. A custom function was used to tokenize the feedback text and count the occurrences of these words. The sentiment of each feedback entry was determined based on the relative counts of positive and negative words:

- If positive word count > negative word count, the sentiment was classified as Positive.

- If negative word count > positive word count, the sentiment was classified as Negative.

- If positive word count = negative word count, the sentiment was classified as Neutral.

* **Visualization**

Visualization played a crucial role in summarizing and interpreting the feedback data. Various types of plots were used to convey insights effectively:

- Bar plots to show the count of feedback entries in each category.

- Stacked bar plots to visualize the distribution of sentiments by source.

- Box plots to analyze feedback length by category and sentiment.

- Count plots to display sentiment distribution across different sources.

Additionally, Power BI was used to create interactive dashboards that allowed for dynamic exploration of the data. These dashboards provided a user-friendly interface for stakeholders to interact with and derive insights from the feedback data.

**Results and Insights**

**Feedback Category Distribution**

The distribution of feedback categories was visualized through bar plots. The majority of feedback fell into categories 3 (Most) and 4 (All), indicating a high level of participation and engagement among the feedback providers. The detailed distribution is as follows:

- Category 0 (Nobody): Very few entries.

- Category 1 (Very Few): Few entries.

- Category 2 (Half): Moderate number of entries.

- Category 3 (Most): Significant number of entries.

- Category 4 (All): Highest number of entries.

- Category 5 (Not Sure): Few entries.

**Sentiment Analysis**

Sentiment analysis revealed that most feedback entries were classified as Positive. This suggests a generally positive sentiment among the feedback providers. The detailed sentiment distribution is as follows:

- Positive: Majority of entries.

- Neutral: Moderate number of entries.

- Negative: Few entries.

**Source-Specific Sentiment**

A stacked bar plot was used to visualize the sentiment distribution by source. The analysis revealed that:

- Teachers and M&E sources provided more Positive feedback compared to Facilitators.

- Facilitators had a relatively higher proportion of Neutral and Negative feedback.

This variation highlights the different perspectives and experiences of the stakeholders involved.

**Feedback Length Analysis**

Box plots were used to analyze the length of the feedback entries by category and sentiment. The analysis showed that more detailed feedback entries tend to have more variation in sentiment, providing deeper insights into specific areas of concern or success. For example:

- Positive feedback entries generally had longer descriptions, indicating more detailed and constructive feedback.

- Negative and Neutral feedback entries were shorter, suggesting less detailed responses.

**Power BI Analysis**

Power BI dashboards were created to provide interactive visualizations and enhance the analysis. The following insights were derived from the Power BI dashboards:

**1. \*\*Interactive Category Distribution\*\*:**

- Users could dynamically filter the feedback categories to see how different categories performed over time or by specific sources.

**2. \*\*Sentiment Trends\*\*:**

- Interactive line charts allowed users to observe trends in sentiment over time, helping identify periods of particularly positive or negative feedback.

**3. \*\*Source Comparison\*\*:**

- The dashboards enabled a detailed comparison of feedback from different sources, highlighting which sources consistently provided more positive or negative feedback.

**4. \*\*Detailed Feedback Exploration\*\*:**

- Drill-down capabilities allowed users to explore individual feedback entries in detail, providing context to the aggregated data and helping identify specific areas for improvement.

**Conclusion**

This project successfully utilized text preprocessing, sentiment analysis, and data visualization techniques to analyze feedback data. The findings provide valuable insights into the effectiveness of the initiatives and help identify areas for improvement. The key insights from the analysis include:

- A predominance of Positive feedback, indicating general satisfaction and engagement.

- Variations in sentiment based on the source of the feedback, with Teachers and M&E providing more positive assessments compared to Facilitators.

- Detailed feedback entries tend to offer more nuanced insights, suggesting the importance of encouraging detailed responses in future feedback collection.

The integration of Power BI added significant value to the analysis by providing interactive and user-friendly dashboards. These dashboards allowed stakeholders to explore the data dynamically, uncovering deeper insights and facilitating more informed decision-making.

To further enhance the analysis, the sentiment analysis model could be refined to include more nuanced sentiment indicators and advanced natural language processing techniques. This would provide even more accurate and actionable insights, helping stakeholders to better understand and address the needs and concerns of the participants.