This is to inform you about the details for the upcoming group assignment submission for **Track 1**. You are required to submit **two Colab notebooks** as part of this assignment. The guidelines for each notebook are as follows:

1. **Dataset Preparation for Fine-Tuning:**
    - Prepare the dataset up to the data-loading stage.
    - Output relevant intermediate results and include detailed comments about the following:
        - The type of dataset used.
        - Cleansing techniques applied to the dataset.
    - Perform **Exploratory Data Analysis (EDA)** on the cleaned data. Include any insights or analyses that you consider relevant.
    - Print a few sample instances from your cleaned dataset to demonstrate the quality of the preprocessing.
2. **Word Embedding Visualization Using Co-occurrence Matrix:**
    - Construct a co-occurrence matrix for a dataset of sentences. (Suggested dataset size: approximately **5,000–10,000 sentences** for meaningful results.)
    - Apply **Principal Component Analysis (PCA)** to reduce the dimensionality of the matrix to 2D.
    - Visualize the 2D word embeddings for a selected set of words (e.g., **"king," "queen," "man," "woman," "apple," "orange"**).
    - Compare your results with existing 2D embeddings (e.g., from **pre-trained models like Word2Vec, GloVe, or FastText**) and analyze the differences.