Task Breakdown for Sentiment Analysis Project

Task 1: Import Necessary Libraries
* Import required libraries: pandas, numpy, matplotlib, seaborn, wordcloud, nltk, contractions, emoji, BeautifulSoup4

Task 2: Load and Explore Data
* Read the CSV file containing reviews.
* Perform initial data exploration:
    * Check data shape and types.
    * Identify missing values and duplicates.
    * Calculate summary statistics.

Task 3: Data Visualization
* Create a word cloud to visualize frequently occurring words.
* Plot the frequency of specific target words (e.g., "good", "service").

Task 4: Text Preprocessing
* Convert text to lowercase.
* Tokenize text into individual words.
* Remove stop words.
* Perform stemming or lemmatization.
* Remove numbers and special characters.
* Expand contractions.
* Remove emojis.
* Remove HTML tags.

Task 5: Feature Engineering
* Convert text data into numerical representation (e.g., using techniques like TF-IDF or word embeddings).

Task 6: Model Building and Training
* Split data into training and testing sets.
* Choose a suitable classification algorithm (e.g., Naive Bayes, Logistic Regression, SVM, Random Forest).
* Train the model on the training data.

Task 7: Model Evaluation
* Evaluate the model’s performance using metrics like accuracy, precision, recall, and F1-score.

Task 8: Sentiment Prediction
* Use the trained model to predict sentiment for new or unseen reviews.

Task 9: Model Improvement (Optional)
* Experiment with different preprocessing techniques, feature engineering approaches, and algorithms to enhance model performance.

Note: These tasks can be further divided into smaller subtasks depending on the project’s complexity and specific requirements.