**Phase 1: Problem Definition and Design Thinking**

**Problem Statement**

The problem at hand is to perform sentiment analysis on customer feedback to gain insights into competitor products. By understanding customer sentiments, companies aim to identify strengths and weaknesses in competing products, with the ultimate goal of improving their own offerings. This project requires the application of various Natural Language Processing (NLP) methods to extract valuable insights from customer feedback data.

**Design Thinking**

**1. Data Collection**

**Objective: Identify a dataset containing customer reviews and sentiments about competitor products.**

* **Data Source:** We will use a dataset containing customer feedback data related to major U.S. airlines. This dataset was scraped from Twitter in February 2015 and includes tweets that have been classified into positive, negative, and neutral sentiments.

**2. Data Preprocessing**

**Objective: Clean and preprocess the textual data for analysis.**

* **Text Cleaning:** Remove special characters, HTML tags, and irrelevant information from the text data.
* **Tokenization:** Split the text into individual words or subword units.
* **Stop Word Removal:** Eliminate common words that do not carry significant meaning.
* **Stemming/Lemmatization:** Reduce words to their base form for consistency.

**3. Sentiment Analysis Techniques**

**Objective: Employ different NLP techniques like Bag of Words, Word Embeddings, or Transformer models for sentiment analysis**.

* **Bag of Words (BoW):** Use BoW representation to convert text data into numerical features by counting the frequency of words.
* **Word Embeddings (e.g., Word2Vec, GloVe):** Explore pre-trained word embeddings to capture semantic relationships between words.
* **Transformer Models (e.g., BERT):** Leverage transformer-based models for advanced sentiment analysis that considers context.

**4. Feature Extraction**

**Objective: Extract features and sentiments from the text data.**

* **Feature Engineering:** Create additional features like text length, punctuation count, and sentiment scores.
* **Sentiment Labeling:** Assign sentiment labels (positive, negative, neutral) to each customer review based on the analysis.

**5. Visualization**

**Objective: Create visualizations to depict the sentiment distribution and analyze trends.**

* **Sentiment Distribution Plot:** Visualize the distribution of sentiment labels (positive, negative, neutral) in the dataset.
* **Word Clouds:** Generate word clouds to highlight frequently occurring words in positive and negative sentiments.
* **Time Series Analysis (if applicable):** Analyze how sentiments change over time (e.g., monthly trends).

**6. Insights Generation**

**Objective: Extract meaningful insights from the sentiment analysis results to guide business decisions**.

* **Strengths and Weaknesses Identification:** Identify common themes or reasons associated with positive and negative sentiments.
* **Competitor Analysis:** Compare sentiments and feedback about competitor products to assess their market position.
* **Recommendations:** Provide actionable recommendations for improving the company's own products based on customer feedback.

**7. Dataset**

**Objective: Utilize the provided dataset for sentiment analysis.**

* **Dataset Description:** The dataset contains tweets related to major U.S. airlines and includes sentiment labels (positive, negative, neutral) as well as reasons for negative sentiments (e.g., "late flight," "rude service").

**Conclusion**

* The design thinking process outlines a structured approach to solving the problem of sentiment analysis for competitor product feedback. By following these steps, we will gather, preprocess, and analyze the provided dataset to gain valuable insights into customer sentiments. These insights will help companies identify areas for improvement in their own products and better understand their position in the market relative to competitors. The next phases of the project will involve implementing these steps and delivering actionable results.