**SENTIMENT ANALYSIS OF AIRLINE TWEETS**

1. **Business Understanding**
   1. **Overview**

In the highly competitive airline industry, leveraging sentiment analysis can provide a competitive edge. Sentiment analysis is the process of evaluating and categorizing opinions and emotions expressed in customer feedback. By analyzing various sources of customer feedback, such as social media posts, customer reviews, and survey responses, sentiment analysis can determine whether the sentiment of the feedback is positive, negative, or neutral. This can provide valuable insights into customers' opinions and preferences, aid in identifying common complaints, monitor trends in customer feedback, and enable the development of targeted strategies to enhance customer satisfaction and loyalty. Employing sentiment analysis can lead to data-driven decisions, improving overall performance and reputation, and resulting in increased profitability and success.

* 1. **Business Problem**

The airline industry faces various challenges, including high operational costs, security concerns, environmental impact, and fierce competition. In Kenya, for example, the national carrier Kenya Airways has been struggling financially as a result of some of these challenges. To remain competitive in such a challenging environment, airlines need to find innovative ways to improve their service offerings while keeping operational costs low. One way to achieve this is by leveraging sentiment analysis to better understand customer preferences and improve overall customer satisfaction. By analyzing customer feedback from various sources such as social media, customer review websites, and surveys, airlines can gain valuable insights into customer opinions and emotions. Categorizing this feedback using sentiment analysis allows airlines to identify common complaints, monitor trends in customer feedback, and develop targeted strategies to improve customer loyalty. This approach can lead to data-driven decisions that improve overall performance and reputation, resulting in increased profitability and success.

Therefore, we aim to leverage sentiment analysis through Natural Language Processing (NLP) to provide valuable insights into customer preferences and improve the overall customer experience which will consequently improve the airline's competitiveness.

* 1. **Objectives**

To build a machine learning model that accurately predicts the sentiment of airline passengers based on their tweets.

* + 1. **Specific Objectives**

1. To use NLP techniques to preprocess and clean the customer feedback dataset to prepare it for sentiment analysis.
2. To apply sentiment analysis techniques to categorize the customer feedback into positive, negative, or neutral sentiments.
3. To determine the model’s effectiveness in predicting the sentiment of airline passengers based on recall and F1-score metrics.
4. To interpret the model results and gain insights into the factors that influence customer satisfaction and dissatisfaction with the airline.
5. Top of Form
   1. **Success Criteria**

A model will be considered a success when it is able to correctly identify all tweets with positive, negative or neutral sentiment. For an airline, missing a tweet that expresses negative sentiment could lead to the airline missing an opportunity to address the issue and potentially lose a customer. Therefore, high recall is crucial in order to capture all relevant tweets expressing positive, negative or neutral sentiment, and take appropriate actions accordingly.

However, a model with high recall may also classify a large number of irrelevant tweets as positive, which would lower its precision and potentially mislead the interpretation of the sentiment analysis results. F1 score will be used as a secondary metric, to provide a balance between recall and precision, ensuring that the model accurately classifies relevant tweets as positive, negative, or neutral without falsely classifying irrelevant tweets, leading to a more accurate and reliable sentiment analysis.

1. **Data Understanding**
   1. **Overview**

The airline\_tweets dataset was obtained from [Kaggle](https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment). Data was scraped from Twitter and contributors were asked to first classify positive, negative, and neutral tweets, followed by categorizing negative reasons (such as "late flight" or "rude service").

The dataset consists of six airlines. It has 14640 rows and 15 columns.