

# Lufthansa Twitter Analysis

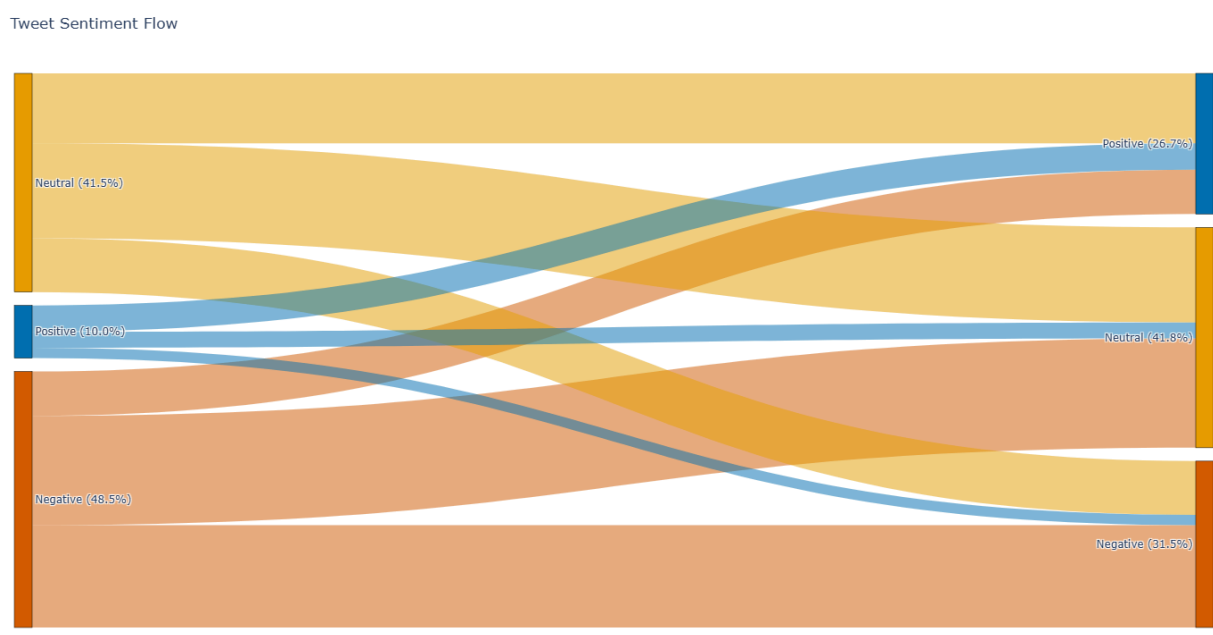
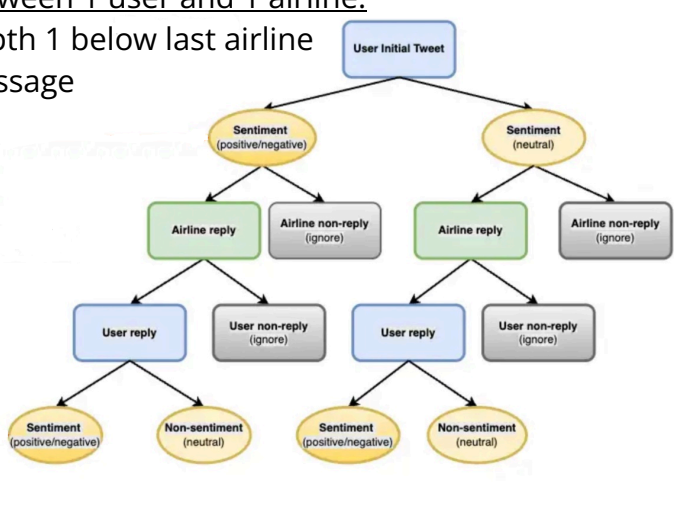
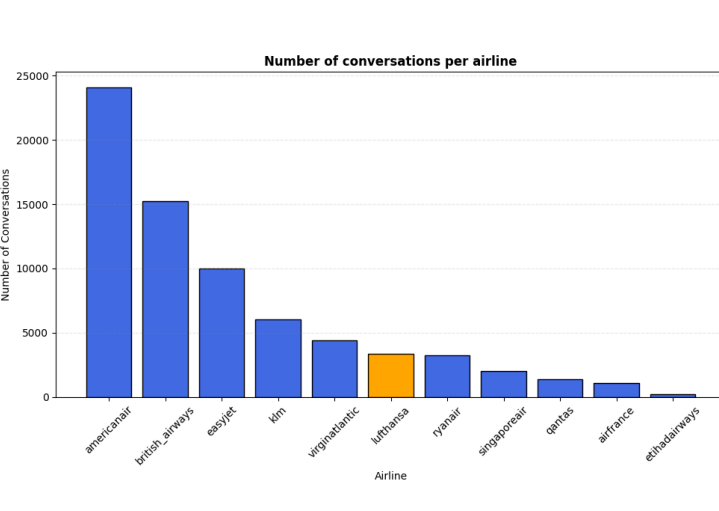
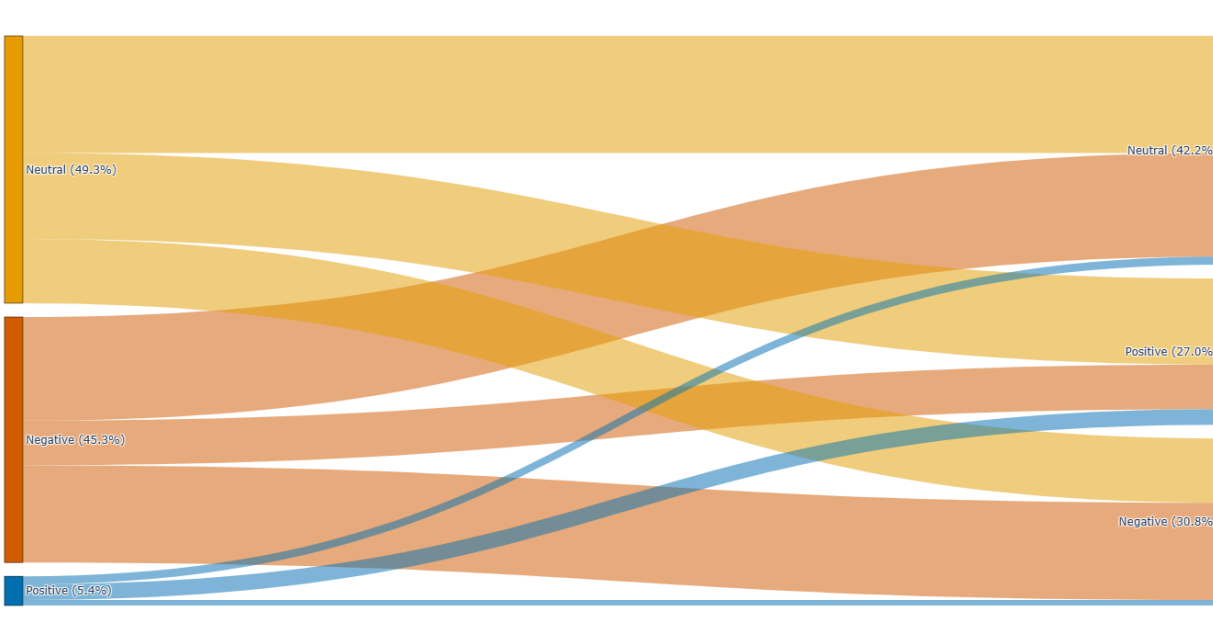


DBL Cork

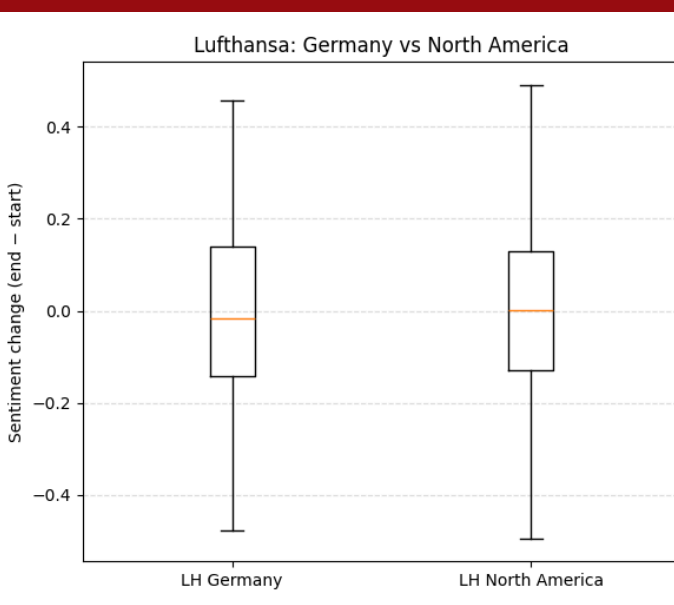
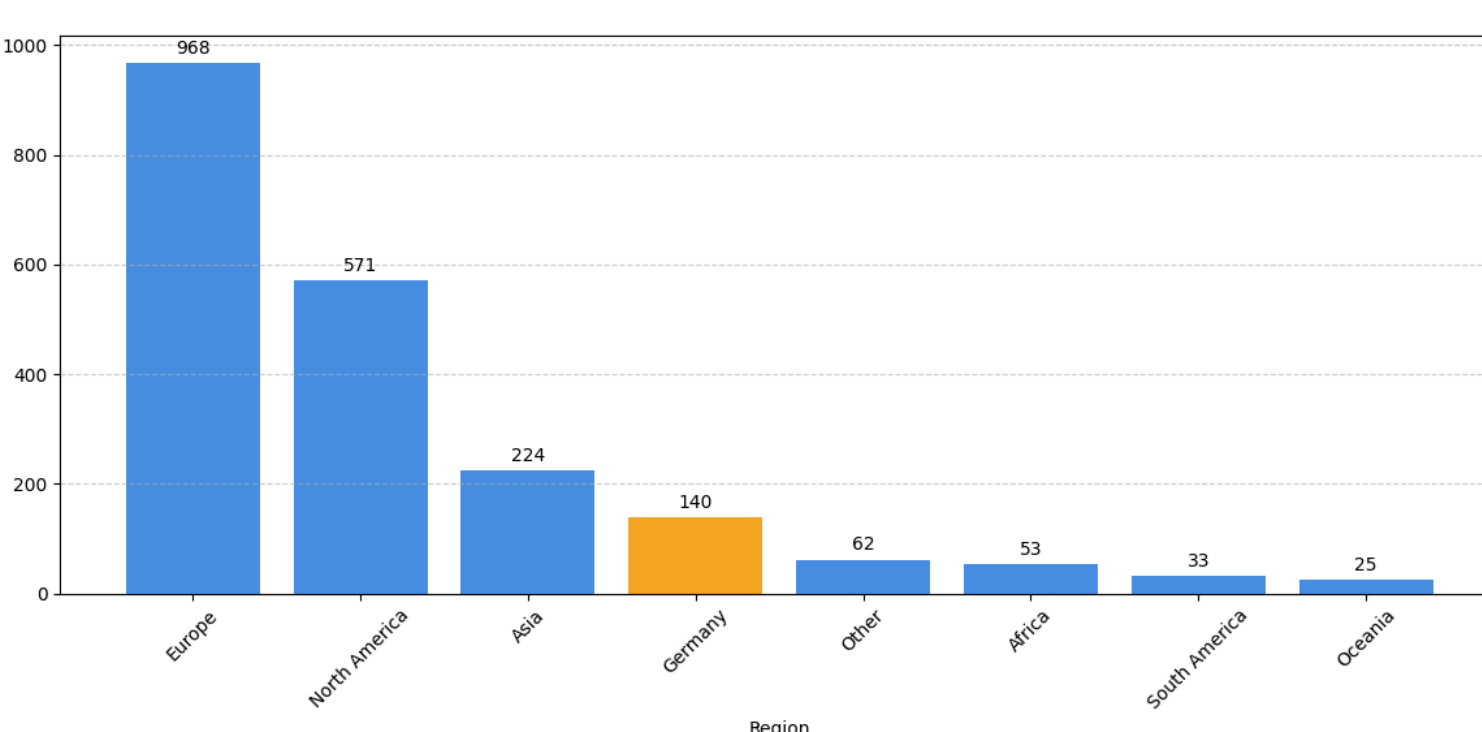
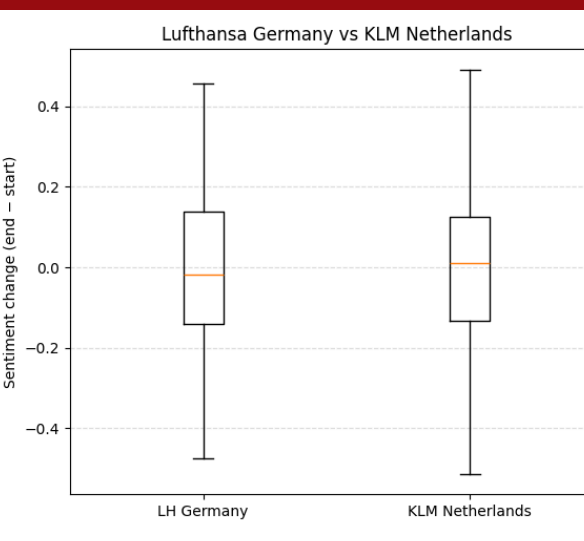
Presenting: D.J Fulton, N. Botas Bernardo





Introduction and Recap		Sentiment Evolution	Sankey Diagrams
<p>Social media provides valuable customer feedback in real time. In this project, we analyzed Lufthansa-related tweets to uncover sentiment trends and identify opportunities for service improvement.</p> <ul style="list-style-type: none"><li>6 million tweets provided</li><li>5.45 million tweets used after data cleaning</li><li>Grouped 70,000 conversations for sentiment analysis</li><li>3,379 conversations analyzed for Lufthansa</li></ul>		<p><u>Evolution of All Airlines</u></p> <ul style="list-style-type: none"><li>Compared to Lufthansa's, it has a higher share of initially Negative tweets and a slightly lower of Neutral ones.</li><li>The dominant flow of Negative sentiment either remains Negative or shifts to Neutral, with minimal conversion to Positive.</li><li>While sentiment transitions occur, a large portion of Negative tweets across airlines tends to persist or shift to Neutral rather than improve.</li></ul>	
<p><u>Conversation Tree</u></p> <p>Between 1 user and 1 airline: Depth 1 below last airline message</p>  <p><u>Conversations Per airline</u></p> 		<p><u>Evolution of Lufthansa</u></p> <ul style="list-style-type: none"><li>Tweets initially characterized as Neutral were redistributed across all categories.</li><li>Interestingly, a significant proportion of tweets that started as Negative remained Negative or shifted to Neutral.</li><li>This indicates persistent dissatisfaction or a neutralization of negative sentiment.</li><li>Only a small fraction transitions to Positive, reflecting limited positive engagement.</li></ul>	

## Business Idea

Geoclustering		Lufthansa - Region Comparison	
<p><u>Explanation of Geoclustering</u></p> <ul style="list-style-type: none"><li>Used to determine Lufthansa's performance per region.</li><li>Took user location information, and turned it into coordinates and from there confident guesses into regions.</li></ul>		<p><u>Hypothesis Test</u></p> <ul style="list-style-type: none"><li><math>H_0</math>: There is no difference in sentiment between Lufthansa North America and domestic.</li><li><math>H_a</math>: There is a difference in sentiment.</li><li>Test statistic: <math>t = -0.320</math>, <math>p = 0.749</math>.</li><li>Since <math>p &gt; 0.05</math>, we fail to reject <math>H_0</math>. Thus, there is no significant sentiment difference.</li></ul>	<p><u>Boxplot For Region Comparison</u></p> 
<p><u>The Process of Geoclustering</u></p> <ul style="list-style-type: none"><li>Took location data from user profiles.</li><li>Connected to <code>user_id</code> of those in conversations.</li><li>Used <code>geopy.geocoders</code> to turn location data into longitudinal and latitudinal coordinates.</li><li>Used <code>reverse_geocoder</code> to translate those coordinates into regions.</li></ul>		<h2>Lufthansa and KLM - Domestic Comparison</h2>	
<p><u>Conversations by Region Lufthansa</u></p> 		<p><u>Hypothesis Test</u></p> <ul style="list-style-type: none"><li><math>H_0</math>: There is no difference in sentiment between domestic Lufthansa tweets and KLM tweets.</li><li><math>H_a</math>: There is a difference in sentiment.</li><li>Test statistic: <math>t = -0.414</math>, <math>p = 0.679</math>.</li><li>Since <math>p &gt; 0.05</math>, we fail to reject <math>H_0</math>. Thus, there is no significant sentiment difference.</li></ul>	<p><u>Boxplot For Airline Comparison</u></p> 
<h2>Recommendation</h2> <ul style="list-style-type: none"><li>Lufthansa's performance internationally does not indicate the need to make adjustments for foreign markets.</li><li>While their performance is not worse than their main competitor, it also fails to outperform them.</li><li>Their overall performance in conversations does little to change people's opinions.</li><li>It is the opinion of the group that Lufthansa needs to invest more in their Twitter team to improve overall response results.</li></ul>		<h2>Potential Future Analysis</h2> <p><u>Topic Clustering</u></p> <p>Grouping conversations by the topic being discussed (e.g. baggage loss, flight delay etc).</p> <ul style="list-style-type: none"><li>Is there an area for which Lufthansa needs significant improvement?</li><li>Is there an area for which they excel?</li><li>Why might this be?</li></ul>	